Rapid Identification and Classification of *Escherichia coli* and *Shigella* by Attenuated Total Reflectance - Fourier Transform Infrared Spectroscopy

Tien My Lisa Lam

Department of Food Science and Agricultural Chemistry
McGill University, Montréal, Québec

April 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

SUGGESTED SHORT TITLE

Rapid Identification and Classification of *E. coli* and *Shigella* by ATR-FTIR Spectroscopy

For my parents

ABSTRACT

Rapid identification of microorganisms is a trending topic in research today. By comparison with other methods of microorganism identification like polymerase chain reaction (PCR), pulsed field gel electrophoresis (PFGE) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy is much quicker in terms of time of analysis. In addition, ATR-FTIR spectroscopy requires no reagents, is cost effective and has potential to identify microorganisms down to the pathotype level. Most foodborne illnesses are due to improper food handling which can lead to the contamination of food products with pathogenic bacteria such as some strains of *Escherichia coli* (*E. coli*) and *Shigella* species. Differentiating between *E. coli* and *Shigella* species. is challenging because they are genetically similar and was investigated in this thesis by using ATR-FTIR spectroscopy. Various strains of *Escherichia* (n=190) and *Shigella* (n=145) of fecal and blood origin were obtained from the McGill University Health Center (MUHC), Laboratoire de Santé Publique du Québec (LSPQ) and Health Canada (HC). For preparation of the samples for analysis, the samples were taken from frozen cultures, plated onto culture media, and incubated at 37°C for 18-24 h. After incubation, the sample was sub-cultured and incubated using the same parameters as in the first culture. After sub-culturing, a single isolated colony was taken and smeared onto the ATR crystal of the ATR-FTIR instrument to acquire a spectrum. By using principal component analysis (PCA) and hierarchical cluster analysis (HCA) of the ATR-FTIR spectral data, *E. coli* was successfully discriminated from *Shigella* species based on their spectral differences at the regions of 1478-1411 and 1070-1040 cm^{-1}. Moreover, successful discrimination between *Shigella sonnei* and *Shigella flexneri* was achieved by using the spectral regions of 1136-1113 and 1218-1207 cm^{-1}. For *E. coli* O157:H7 and non-O157:H7 Shiga-toxin-producing *E. coli* (STEC), the separation between the two groups was successful using the regions of 1248-1212 and 1356-1344 cm^{-1}. In conclusion, ATR-FTIR spectroscopy has potential for identifying *E. coli* and *Shigella* at the genus, species, pathotype and serotype levels.

RÉSUMÉ

L'identification rapide des microorganismes est une tendance dans la recherche contemporaine. Par comparaison avec d'autres méthodes d'identification de microorganismes telles que la réaction en chaîne par polymérase (RCP), l'électrophorèse sur gel en champ pulsé (EPCP) et la spectrométrie de masse à temps de vol par ionisation / désorption laser assistée par matrice (MALDI-TOF MS), la spectroscopie infrarouge à transformée de Fourier par réflectance totale atténuée (ATR-FTIR) est beaucoup plus rapide en termes de temps d'analyse. À part d'être rapide, la spectroscopie ATR-FTIR n'utilise pas de réactifs et est rentable, et sa capacité d'identifier des microorganismes jusqu'au niveau du pathotype a été démontré. La plupart des maladies d'origine alimentaire sont dues à une mauvaise manutention des aliments qui engendre la contamination des produits alimentaires par des bactéries pathogènes, telles que certaines souches de l'*Escherichia coli* (*E. coli*) et le *Shigella*. La différenciation de ces deux microorganismes pose des problèmes particuliers dû à leur similarité génétique. Cette étude veut relever ce défi de différenciation par l'utilisation de la spectroscopie ATR-FTIR. Des souches diverses d'*Escherichia coli* (n=190) et *Shigella* (n=145) provenant des selles et du sang ont été obtenues du Centre universitaire de santé McGill (MUHC), de la Laboratoire de santé publique du Québec (LSPQ) et de Santé Canada (HC). Pour la préparation des échantillons pour l'analyse, les échantillons prélevés des cultures congelées ont été étalés sur des milieux de culture et incubés à 37°C pendant 18 à 24 heures. Après incubation, l'échantillon a été sous-cultivé et incubé à 37°C pendant 18 à 24 heures. De cette deuxième incubation, une seule colonie isolée a été prise et étalée sur le cristal ATR du spectromètre ATR-FTIR pour en générer le spectre infrarouge. En appliquant ensuite l'analyse en composantes principales (PCA) et l'analyse de groupement hiérarchique (HCA) sur les données spectrales, l'*E. coli* a été bien différencié des espèces de *Shigella* basé sur leurs différences spectrales dans la région de 1478-1411 cm^{-1}. De plus différenciation entre le *Shigella sonnei* et le *Shigella flexneri* a été réalisé en utilisant les régions spectrales de 1136-1113 et 1218-1207 cm^{-1}. La spectroscopie ATR-FTIR a aussi réussi à séparer *E. coli* O157:H7 du *E. coli* non-O157:H7 producteur de la toxine Shiga (STEC). La séparation de ces deux groupes a été réussie à l'aide des régions de 1248-1212 et 1356-1344 cm^{-1}. En conclusion, la spectroscopie ATR-FTIR a la capacité d'identifier l'*E. coli* et le *Shigella* aux niveaux du genre, de l'espèce, du pathotype et du sérotype.

ACKNOWLEDGMENTS

Without the help and collaboration of many institutions and individuals, my thesis would not have been completed till this date. First and foremost, I would like to give my greatest appreciation for the support and understanding of all the hardship I put on my parents Van Thi Thu Nguyen and Son Van Lam, my siblings, Nancy, Jimmy and Annie throughout my studies. I would also like to give a special thanks to my brother-in-law John for the mental support and encouragement for the fulfillment of my degree.

Thank you to all the helping hands of my family and friends, I am forever grateful and appreciative for all the patience, understanding, love, financial support and comfort that was given at no cost.

My thesis could not have been accomplished without the collaborations, mentorship and contribution from Dr. Pierre Lebel of McGill University Health Center, Dr. Irene Iugovaz of Health Canada, and Dr. Sadjia Bekal and Dr. Simon Lévesque of Laboratoire de Santé Publique du Québec. I am forever thankful for all the patience they had for me, providing me all the bacterial strains and facility to complete my research. I would also like to thank Dr. Jacqueline Sedman for keeping the projects on track, guidance and most importantly dedicating long hours editing my thesis. Without Dr. Sedman's thoroughness and dedication, my final thesis would not be whole. And of course, I would also like to extend appreciation all the laboratory fellows and colleagues for their time, patience, assistance and conversations during those long and late hours.

Without a doubt, I would like to give my greatest and sincerest appreciation and gratitude to my supervisor Dr. Ashraf Ismail. Without his supervision, dedication, patience, motivation and direction, I would have been disoriented. Thank you for allowing me to appreciate scientific research and understand all the excitement and importance it holds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATR-FTIR	Attenuated total reflectance - Fourier transform infrared
BA	Blood agar
CDC	Center for Disease Control and Prevention
DAEC	Diffusely adherent *Escherichia coli*
EAEC	Enteroaggregative *Escherichia coli*
E. coli	*Escherichia coli*
E. faecalis	*Entrococcus faecalis*
E. faecium	*Entrococcus faecium*
EHEC	Enterohaemoragic *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EPEC	Enteropathogenic *Escherichia coli*
GN	Gram-negative
GP	Gram-positive
HC	Health Canada
HCA	Hierarchical cluster analysis
Ipa	Invasion plasmid antigen
LPS	Lipopolysaccharide
LSPQ	Laboratoire de Santé Publique du Québec
MAC	MacConkey agar
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MLVA	Multiple locus variable-number tandem repeat analysis
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass spectroscopy
MSSA	Methicillin sensitive *Staphylococcus aureus*
MUHC	McGill University Health Center
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
VRE	Vancomycin-resistant *Enterococcus*
S. flexneri	*Shigella flexneri*
SMAC	Sorbitol MacConkey agar
S. sonnei	*Shigella sonnei*
STEC	Shiga-toxin-producing *Escherichia coli*
Stx	Shiga toxin
WGS	Whole genome sequencing

Chapter 1 – GENERAL INTRODUCTION

Bacterial resistance to antibiotics is becoming a concern in today's modern world. With the increase in population and need for rapid food production, food safety has also become an increasing concern. By increasing the quantity of food production while producing product at a quicker rate, there will become more room for errors. With the potential for more error in food safety, research on methods for rapidly detecting pathogens is of interest. In addition, the public health sector would benefit from rapid microorganism identification as it can potentially prevent resistance in antibiotics and make diagnostics more efficient.

By exploring attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy as a tool for rapid bacterial identification, it can be used in the food industry as well as public and private health and research laboratories. Compared to other current methods, ATR-FTIR spectroscopy is quick, cost effective and easy to use.

The main objective of the thesis is to determine the effectiveness of ATR-FTIR spectroscopy for the separation, determination and classification of *Escherichia coli* (*E. coli*) and *Shigella* spp. These microorganisms are foodborne pathogens that have genetic and biochemical similarities. *E. coli* O157:H7 has been linked to foodborne outbreaks which have left many ill and in worst case scenarios, led to death. In addition, *E. coli* O157:H7 produces Shiga toxins similarly to *Shigella*.

Gram-positive and Gram-negative bacteria are biochemically significantly different from one another. As a preliminary study in the given thesis, ATR-FTIR spectroscopy will be used for the separation of the two groups. In relation to foodborne pathogens, the study of the spectral differences between *E. coli* and *Shigella* species will then be examined by using ATR-FTIR spectroscopy as a method of discrimination. The spectral differences between *Shigella sonnei* and *Shigella flexneri* will also be examined and discussed to understand their biochemical differences. Given that *E. coli* O157:H7 is one of the leading causes of serious foodborne illnesses, it too will be examined alongside other Shiga-toxin-producing *E. coli*. Among all Shiga-toxin-producing *E. coli*, *E. coli* O157:H7 is the only serotype that is presently screened for as a benchmark for potential outbreaks.

As the toxins produced by both Shiga-toxin-producing *E. coli* and *Shigella* are responsible for infections, the separation between the toxin producers will be explored. Moreover, *E. coli* produces either Stx1 or Stx2 or it can express both. Each Shiga toxin also has variants, such as Stx2f which is a variant of Stx2. In the thesis, the possibility of discriminating between STEC strains producing different Shiga toxins by ATR-FTIR spectroscopy will also be examined. In *E. coli*, the *stx2* gene is the most common trigger for infections and it is therefore of interest to identify STEC strains producing Stx2.

Finally, the results of a validation study conducted by identifying unknown samples that were not in the spectral database developed in this work will be presented. Moreover, to verify that the results are reproducible between different systems of the same model, an instrument-to-instrument validation will also be presented.

Chapter 2 – LITERATURE REVIEW

2.1 Overview of Pathogenic Bacteria Found in Hospitals and Food

Rapid detection of pathogens in hospitals is important for public health and research for improving the current health care system. *Clostridium difficle* (*C. difficle*) and methicillin-resistant *Staphylococcus aureus* (MRSA) are pathogens causing infections and death within hospitals (Kramer et al., 2006). Nosocomial infections such as those caused by MRSA can affect the infected patient's blood, lungs, bones and even become life-threatening. Those who are immunocompromised, elderly or young are most susceptible. Recently in Canada, nosocomial infections within hospitals have been increasing due to incorrect prescription of antibiotics in addition to the misuse of antibiotics (PHAC, 2013). Current methods of detection of pathogenic bacteria are not as efficient as they could be for proper diagnosis and rapid detection.

Similarly, aside from nosocomial infections, there are other infections external from the hospital that can become fatal if incorrectly treated. Foodborne pathogens are bacteria, viruses and parasites that cause illnesses due to the consumption of contaminated foods and water. In Canada, norovirus is the cause of most foodborne illnesses. According to statistics from the Government of Canada, 1 in 8 Canadians is affected by foodborne illnesses each year. Moreover, 6% of all hospitalizations from foodborne illnesses in Canada are due to *Escherichia coli* (*E. coli*) and 8% of all deaths from foodborne illnesses are also due to *E. coli*. Infection with *E. coli* from contaminated foods can cause bacillary dysentery (Goebel, 2012). Similar to *E. coli, Shigella*, is another pathogen that invades the body following the consumption of contaminated food. *Shigella*, like *E. coli* will also cause bacillary dysentery, an infection which causes watery and bloody diarrhea, loss of appetite, muscle aches, nausea, vomiting, fever and dehydration (Van den Beld, 2012). There are many similarities between the two bacteria but there are also many small differences between the two. Given that there are slight genetic and biochemical differences, there is potential for them to be discriminated from one another. More importantly, the difference in virulence is a major aspect in classifying the two pathogens for proper and fast diagnosis.

E. coli O157:H7 and *Shigella sonnei* (*S. sonnei*) are the leading causes of foodborne illness, making them key pathogens for detection (Yang et al., 2014). Infection with *E. coli* O157:H7 is commonly referred to as the "hamburger disease" (Goebel, 2012), as it is associated with the consumption of contaminated undercooked meat where the contamination comes from, cattle feces or any other fecal matter originating from poultry, swine, lamb and turkey (Black et al., 1978). Shiga-toxin-producing *E. coli* (STEC) and *S. sonnei* are naturally found in the gut of animals, including humans (Santos, 2015). The illness can also be transmitted from person-to-person or by the fecal-to-oral route. The bacterial infection will cause nausea, bloody diarrhea, abdominal cramps, headaches, vomiting, haemorrhagic colitis and haemolytic uraemic syndrome (HUS) (Rangel et al., 2005). *S. sonnei* also causes the same symptoms as well as originating from the same sources as *E. coli* (Black et al., 1978). Both pathogens are found in contaminated water and food contaminated with fecal matter (since is it present in the gut of mammals and poultry) which is then consumed by the consumer. Depending on the consumer's health and immunity, the virulence of the pathogen and the severity of illness can vary from person to person.

Like MRSA and other antibiotic-resistant pathogenic bacteria, *Shigella* has also exhibited antibiotic resistance (Zhang et al. 2015). Over the years, the increase in antibiotic resistance among bacteria has proven that more research is needed to prevent resistance. In most cases, the current methods for detecting antibiotic-resistant bacteria are not cost effective, are labor intensive, require reagents and are time-consuming (Yang et al., 2014). Generally, what one method has, another method lacks. In today's world, it would be ideal to have a single method that is capable of having all the necessary positive components for the rapid detection of bacteria to better diagnose bacterially infected patients.

2.2 Introduction to *Escherichia coli*

Escherichia coli is one of the most common bacteria in humans and animals alike. The bacteria can be found in various locations; the environment, human gut, animals and contaminated food products. Given their abundance in nature, most strains are harmless with a handful of being pathogenic. There are six different groups of *E. coli* that are pathogenic which are characterized by their pathotype. Pathogenic *E. coli* causes diarrhea and vomiting when orally ingested while

non-pathogenic *E. coli* aids in the wellbeing of the gastrointestinal tract. The six pathotypes of pathogenic *E. coli* are, enterohemorrhagic *E. coli* (EHEC) (strains of this group are also known as Shiga-toxin producing *E. coli* (STEC) with the most common serotype being O157:H7), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enteroaggregative *E. coli* (EAEC), enteroinvasice *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Figure 2.1) (Donnenberg, 2013).

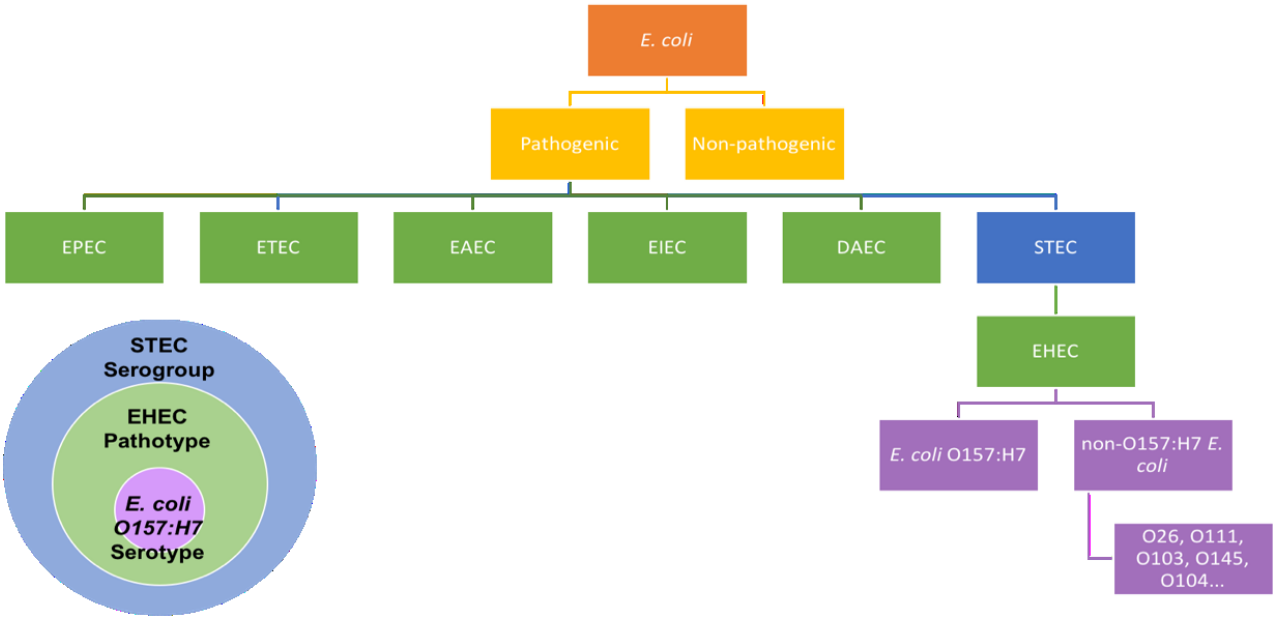


Figure 2.1. Understanding the sub-division of pathogenic *E. coli.*

E. coli infections result predominately from STEC, which is also known as verocytotoxic *E. coli* (VTEC). STEC is composed of two groups, *E. coli* O157:H7 (O157 or *E. coli* O157) and non-O157 STEC consisting of other serotypes of *E. coli* that can produce Shiga toxins and cause infections.

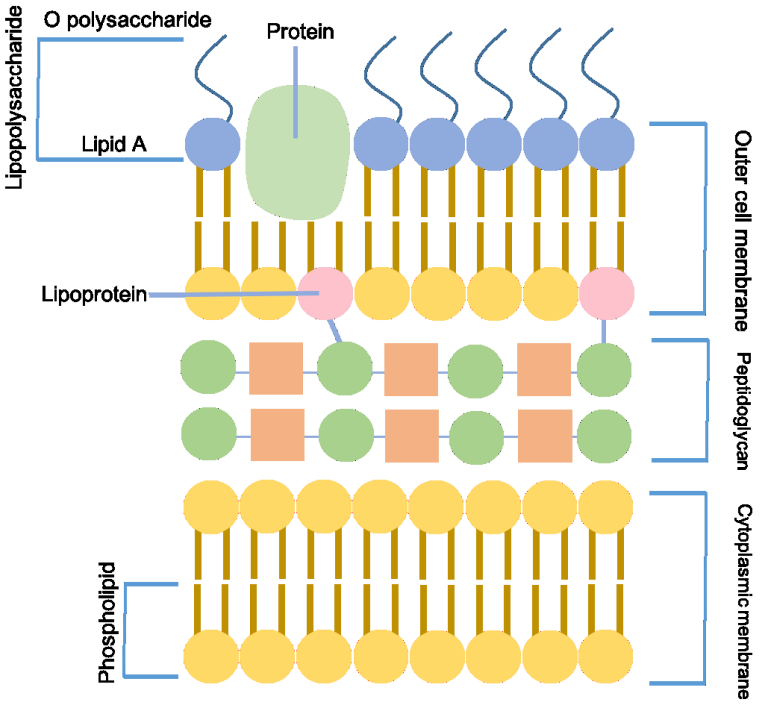


Figure 2.2. Overview of Gram-negative bacteria wall composition.

E. coli is a rod-shaped bacterium with fimbriae and a flagella (Rangel, Sparling, Crowe, Griffin, & Swerdlow, 2005). It has a thin peptidoglycan cell wall layer which classifies it as Gram-negative (Figure 2.12). In a bacterial Gram-stain test, Gram-positive bacteria retain the crystal violet stain and appear violet whereas Gram-negative bacteria do not.

2.3 General Overview of Zoonosis

Many viral and bacterial infections are transmitted from animals to humans, termed zoonosis. With increasing contact with wild animals (animal care, fishing, hunting and more), more zoonotic cases have been observed. Among the 1415 known pathogenic species that can infect humans, 61% of them are zoonotic in origin and this number has been increasing in recent years (Lin, 2013).

Considering the increase of zoonotic cases in recent years, surveillance studies have demonstrated that food is the principal conduit for zoonotic infections. *Salmonella* species are the leading cause

of foodborne illnesses followed by *E. coli* and *Shigella*. Infection from *Salmonella* is instigated by the consumption of the pathogen, for which farm animals such as cattle, chickens and pigs are reservoirs. The bacterium flourishes in water and when the water is being used during food processing, the food becomes contaminated as well. Moreover, after human consumption of the contaminated water or food, the bacterium adheres to and enters the intestinal cells, causing salmonellosis. The infection is associated with gastroenteritis and a high fever (Lin, 2013).

E. coli is the second leading cause of foodborne illnesses, with zoonotic EHEC strains including *E. coli* O157:H7 causing bloody diarrhea. Among known cases, cattle have been the primary source of *E. coli* that causes contamination of food for human consumption. The pathogen is not limited to cattle but is also found in other common farm animals and wildlife (Oporto, Esteban, Aduriz, Juste, & Hurtado, 2008). Since the bacterium is found in the gut of these animals, it is essential for slaughterhouses to be kept sanitary to avoid contamination and cross contamination.

Colibacillosis is the infection caused by *E. coli* and can become dangerous depending on the strain and whether or not it produces Shiga toxins. Similarly, *Shigella* causes shigellosis, giving the same symptoms as colibacillosis. Both species are derived from the same ancestors.

In a study conducted by Noori and Alwan (2016), 100 randomly selected samples of raw poultry meat were analyzed for the prevalence of zoonotic bacteria. Among these, 49 samples were taken locally while the other 51 were imported samples. They reported that 85 out of the 100 samples tested positive for zoonotic bacteria, of which 29 were positive for *E. coli* and 39 were positive for *Salmonella* (Noori & Alwan, 2016).

Considering that foodborne illnesses are on the rise, a 24-hour bacterial identification and classification period is extremely time-consuming. Using API 20 as a method, it can become very cost ineffective and time-consuming considering the requirement of performing 20 tests per sample. Accordingly, quicker methods for identification of bacteria are needed as well as increased food safety procedures and surveillance.

2.4 Shiga-Toxin-Producing *E. coli* (STEC)

E. coli 0157:H7 produces Shiga toxin (Stx), which is a toxin that is responsible for symptoms relating to food poisoning. Foods and water which are contaminated with feces from cattle and other livestock have been the main cause for contracting STEC infections (Thorpe, 2004). 5-10% of cases, STEC infections lead to hemolytic-uremic syndrome (HUS), which can have long-term implications in the kidneys (Thorpe, 2004). Other serotypes of *E. coli* may also produce the toxin and have also been found to cause bloody diarrhea (Jelacic et al., 2003). STEC has 3 virulence factors, *stx1*, *stx2* and *eae*, where *stx2* is more common and virulent to humans (Thorpe, 2004, Woods et al., 2016).

E. coli O157:H7 is a Stx producer that can be easily identified by its inability to ferment sucrose. Culturing *E. coli* O157:H7 on the selective medium sorbitol-MacConkey (SMAC) will result in bacterial growth to be colorless and classified as *E. coli* O157:H7. Other serotypes of STEC do not have the inability to ferment sucrose and will cause the growth to be positive for sorbitol fermentation (pink in color). With STEC being a heterogeneous group of strains, it becomes difficult to identify all STEC (Thorpe, 2016). To identify non-O157:H7 STEC strains, other methods are then performed such as enzyme-linked immunosorbent assay (ELISA) and latex agglutination. These methods detect Stx proteins and are not serotype dependent but become exceedingly cost ineffective (Thorpe, 2004). Ideally, a single method for STEC detection and identification is needed.

In a recent study, multiplex-PCR was used to screen different STEC serotypes including the common O157:H7 serotype. For each of the tested serotypes (O26, O45, O103, O111, O121, O145), specific and optimal gene sequences for detecting the Shiga toxins were used. Among a total of 223 STEC samples tested, there were 6 false positives for *stx1* and 4 for *stx2* (Woods et al., 2016). The study was aimed at detecting both serotype and virulence genes of STEC in a single assay.

Polymerase chain reaction (PCR), ELISA and other currently employed methods are time-intensive and costly and require skilled technologists (Cook, 2016). FTIR spectroscopy has been

a promising alternative method for identification and classification of microorganisms (Davis & Mauer, 2010), with advantages of simplicity, rapidness, and cost effectiveness. An FTIR spectrum contains information about all the biochemical components (lipids, DNA, proteins and carbohydrates) of microbial cells. Each microorganism has a specific biochemical composition and therefore a unique FTIR spectrum. Given that FTIR spectroscopy is specific to the composition of a given bacterium, it has the prospect of classifying the different serotypes of STEC and possibly differentiating between the Stx types.

2.5 Current Methods of Detection of *E. coli* and *Shigella* in Hospitals

Traditionally, bacteria are spread on plates using various culture media depending on the growth conditions required by the microorganism of choice. Depending on the organism, specific culture media are used and in some instances the media will be selective to the organism. Culturing bacteria on growth media is very efficient for testing if there are bacteria present in a sample or not. However, if many bacteria are present in one sample (shown through many different sizes, morphologies and colors of the colonies), then specific classification of the bacteria cannot be done as the sample is mixed and further analysis or sub-culturing of each different colony is required.

With that being said, there are many selective media that can be used to determine whether the bacterial sample is Gram-negative or Gram-positive. For example, MacConkey agar is selective for Gram-negative lactose fermenters, whereas most nutrient agar is used to grow all types of bacteria without specific selection. Nutrient agar is generally used for non-pathogenic microorganisms (Murray et al., 2015).

More interestingly, phenylethyl alcohol agar is used to grow *Staphylococcus* while inhibiting Gram-negative bacilli like *E. coli*. Furthermore, there are no media that are selective for *Shigella* or *E. coli*. There is, however, sorbitol MacConkey (SMAC) which is selective for *E. coli* O157:H7 (Amrita, 2011).

Generally, bacterial growth on agar plates is the preliminary step for the analytical methods that will be discussed in further detail in the following sections. The methods for culturing bacteria are

important in demonstrating the absence of mixed colonies and providing enough biomass for analysis.

2.5.1 Polymerase Chain Reaction (PCR)

Although *E. coli* and *Shigella* are genetically similar, they can easily be separated by using PCR. PCR deals with the amplification of DNA and the usage of varying reagents specific to which location of DNA is needed for amplification. To proceed with the amplification, DNA polymerase, an enzyme, is required in addition to primer sequences of DNA. The reagents are mixed with the bacterial sample or specimen from an infected patient into a vial. From the vial, in about 30 seconds at a high temperature, the hydrogen bonds from the DNA in the sample will be cleaved by the action of the enzymes. After heating the sample to facilitate the enzymatic reaction, the sample is then cooled to allow DNA synthesis at the primers. After completion of the latter step, the sample is heated again for a minute during which the replication of the DNA strand is completed. After the process is finished, the targeted DNA strand of the sample is amplified for better analysis (Metzker and Caskey, 2009).

Given that PCR provides information on DNA fragmentation of the sample, results can be compared to standards and other results for identification. Following amplification, the samples are placed into sample analysis plates, mixed with chemicals and connected to capillary electrophoresis. Using electrophoresis, the DNA fragments are separated by size. The results are computed by a computer software where an electropherogram is generated and compared to standards. This entire process using PCR is called the multiple locus variable-number tandem repeat analysis (MLVA process) (CDC, 2016).

In a reported study completed by Sethabutr et al. in 1993, a specific antigen, invasion plasmid antigen H (IpaH), was employed for the detection of pathogenic *Shigella* and *E. coli*. The Ipa genes (*A*, *B* and *C*) are present in both *Shigella* and enteroinvasive *E. coli* (EIEC). Twenty patients who tested positive for shigellosis or *E. coli* infection provided stool samples for PCR analysis. The patients' stool samples were collected before and after antibiotic treatment to determine if the pathogen was still present after treatment. The samples were then treated with various reagents to

facilitate amplification of the targeted gene. As a control of the PCR analysis, the DNA of *Shigella flexneri* and *E. coli* K 12 Xac were used as the positive and the negative control, respectively (Sethabutr et al., 1993). The method was successful for the detection of *E. coli* and *Shigella* by detecting IpaH in the stool samples and invasion plasmid locus *(ial)*. After the patients had undergone antibiotic treatment, new stool samples were analyzed to prove the function of the antibiotics. The stool samples taken from the patients subsequent to ciprofloxacin treatment lacked the *A*, *B* and *C* genes of Ipa in both *Shigella* and *E. coli*. Specific genes for *Shigella* and *E. coli* need to be determined through whole-genome sequencing to clearly differentiate the two pathogens from one another. While this PCR method proved to be sensitive and accurate (Sethabutr et al., 1993), PCR requires reagents that are specific to the gene of interest for amplification, and it is also time-consuming; taking roughly 6 to 48 hours for a complete analysis (Drake et al., 2011).

2.5.2 Pulsed-Field Gel Electrophoresis (PFGE)

Pulsed-field gel electrophoresis (PFGE) is another method used for the identification of microbial pathogens, including *E. coli* and *Shigella*. PFGE has been adopted by the Centers for Disease Control and Prevention (CDC) for the surveillance of foodborne illnesses. The method is used to subtype bacteria using various materials and reagents. It takes ~24 hours for the full analysis and up to 3 or 4 days for more robust results. Unlike PCR, PFGE can provide the subtype of bacteria (Ribot et al., 2006). It is the "gold standard" for many agencies like CDC, Health Canada and others (Dalmasso and Jordan, 2015). One of the disadvantages of PFGE is that the results are not reproducible from laboratory to laboratory. For example, the results of PFGE analysis done in a laboratory in China cannot be compared to the ones done in a laboratory in France, so that databases cannot be shared (Ribot et al., 2006, Leclair et al., 2006). However, many protocols are being researched to obtain reproducible results between laboratories (Durmaz et al., 2009).

In a hospital example, contaminated human specimens are collected and grown on an agar plate. After culturing, the bacteria is mixed with agarose and poured into plug molds (a mold that holds the samples and will be plugged into the wells of the gel) and lysed or broken up by chemicals to

obtain fragments of the DNA. Once the DNA is isolated into the plug mold, it is treated with enzymes and plugged into the gel. An electric field is placed on the gel and separates the DNA fragments by size. As a result, the gel image (DNA fingerprint) can be scanned and stored into a database and physical analysis under a UV light can be completed (CDC, 2016).

Although the results are not yet reproducible from laboratory to laboratory, the protocol for performing PFGE is being standardized so that the results from different laboratories can be comparable. Currently, PulseNet has standardized protocols for participating laboratories where all their results from the subtyping of STEC, *Shigella* and *Salmonella* are documented into a database and can be compared with one another without difficulty. Similarly to PCR, PFGE requires extensive training, reagents and time. The advantage of PFGE over PCR is the capability to subtype microorganisms by a standardized method (Ribot et al., 2006). The method is very good for detecting and preventing foodborne outbreaks on a national level, but in a hospital setting, it is too costly and time consuming.

2.5.3 Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

One of the newest technologies that is currently being used in many healthcare laboratories is matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This technology is costly but does not require a lot of time or training (Bourassa et al., 2015). The two companies that currently manufacture MALDI-TOF MS systems are BioMérieux (VITEK MS) and Bruker (Biotyper). These systems can rapidly identify bacterial and fungal isolates from clinical and food samples (Dingle et al., 2013).

For analysis, bacterial or fungal samples that have been cultured on agar plates are placed onto a specific MALDI-TOF MS target plate made of conductive metal with an overlay of organic acid solution. After being placed into the instrument, laser pulses and charged particles going through an electric field are targeted at the sample. The charged proteins produced by the MALDI process then travel through a field-free vacuum. The smaller proteins will travel faster than the larger ones (with equivalent mass-to-charge ratio). The time of flight (TOF) of the proteins reaching the detector is recorded and then the computer generates a spectral fingerprint unique to the microorganism (Bourassa et al., 2015).

With the Bruker system, MALDI-TOF MS has been proven reliable with over 83% accuracy (Lévesque et al., 2015). In a direct comparison of the two systems using this technology, it was found that VITEK MS had more difficulties with identifying microorganisms not included in the database than the Bruker Biotyper (Lévesque et al., 2015). The application of MALDI-TOF MS for identifying bacteria and fungi has proven its worth in numerous hospitals. Not only is it accurate, it is much faster than the other current methods and the analysis time can be reduced by ~ 2 days (Santos et al., 2015). The analysis using MALDI-TOF MS can be done in a single day (24 hours) (Wenning et al., 2014). MALDI-TOF MS is capable of distinguishing the genus and species of the tested bacterial samples but differentiating between microorganisms at the sub-species level proved difficult (Santos et al., 2015, Drake et al., 2011). Since the system is connected to a network, all results are collected into a database, and each sample's spectrum is compared with the reference spectra in the database to identify which microorganism it is. Moreover, MALDI-TOF MS has difficulties differentiating between *E. coli* and *Shigella* (Bourassa et al., 2015). MALDI-TOF MS has been reported to have the potential to detect vancomycin resistance and susceptibility in *Enterococcus faecium* but did not give reproducible results with MRSA (Griffin et al., 2012, Dingle et al., 2013).

Because MALDI-TOF MS is automated, works for both bacteria and fungi, and is simple, rapid, and cost effective, it is one of the most popular microbial discrimination and identification methods in hospitals and government laboratories today.

2.6 Introduction to Fourier Transform Infrared (FTIR) Spectroscopy

In the past four decades, multiple researchers have evaluated the potential of using Fourier transform infrared spectroscopy (FTIR) for the discrimination and identification of foodborne pathogens and nosocomial pathogens.

FTIR spectroscopy has various applications that are currently being used in various fields such as forensics, petrochemical engineering, and food analysis among others (Zarnowiec, 2015), and applications of FTIR spectroscopy in other fields such as microbiology and cancer diagnosis are

under active investigation (Romeo et al., 2003). The IR region of the electromagnetic spectrum spans a wavenumber range of 12,800 to approximately 10 cm^{-1} and can be further subdivided into the near-IR (12,800 – 4000 cm^{-1}), mid-IR (4000 – 200 cm^{-1}) and far-IR (200 – 10 cm^{-1}) regions (Zarnowiec, 2015). The principle of IR spectroscopy is that molecules absorb IR radiation at different wavelengths that correspond to the vibrational frequencies of their bonds. In an IR spectrometer, a sample is placed between an infrared source and an infrared detector, and the energy reaching the detector is measured as a function of wavelength or wavenumber to obtain the IR spectrum of the sample. The drop in energy at a specific wavelength(s) (or wavenumber(s)) resulting from absorption of the IR radiation by the sample will depend on the chemical composition of the sample, yielding a unique spectrum for each compound (Zarnowiec, 2015). From the spectrum, many details about the sample can be deduced, including its composition, the concentrations of components and how the molecules interact.

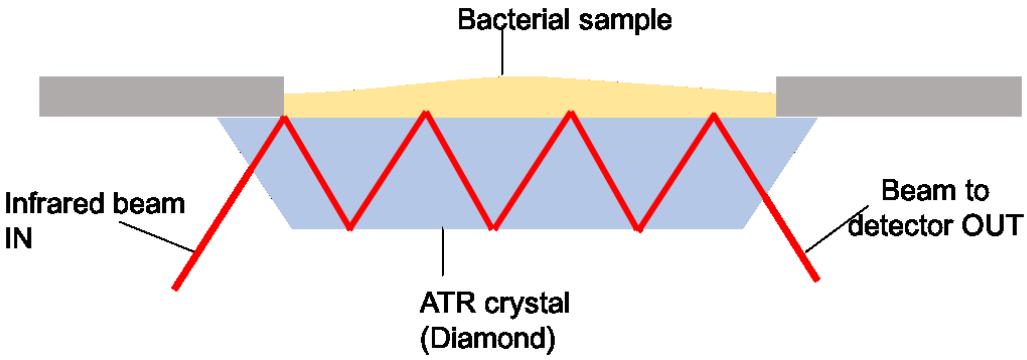


Figure 2.3. Basic principle of ATR-FTIR spectroscopy. An evanescent wave is produced at the surface of the ATR crystal by total internal reflection of the infrared beam and interacts with the bacterial sample placed in contact with the ATR crystal.

An FTIR spectrometer can be paired with an attenuated total reflectance (ATR) accessory abbreviated to ATR-FTIR. In an ATR-FTIR instrument (Figure 2.3), the sample is placed in contact with an ATR crystal, which is made from a material with a high refractive index (commonly zinc selenide (ZnSe) or diamond is used) and generally has a small surface area. The IR beam undergoes total internal reflection as it passes through the crystal, giving rise to an evanescent wave that penetrates a short distance into the sample and is attenuated at the

wavelengths at which the sample absorbs IR radiation (Dole et al., 2011). As a result, the IR spectrum of the sample is obtained by measuring the energy reaching the detector as a function of wavelength (or wavenumber).

For the practical application of FTIR analysis for bacteria identification, a simple protocol is required. The first step, as in the other methods mentioned above, is the sample preparation step. Bacterial samples for FTIR analysis are grown on culture media to increase the amount of biomass and obtain pure colonies. A simple method of acquiring the FTIR spectrum of the sample in the next step is provided by ATR-FTIR spectroscopy, as isolated colonies harvested from culture media can be transferred directly to the surface of the ATR crystal. Accordingly, no sample preparation steps is required beyond the initial culture. Acquisition of the ATR-FTIR spectrum of the sample then takes only minutes to complete, and the identification of the microorganism can then be obtained by comparing the spectrum of the sample to those of reference strains in an infrared spectral database. (Petibois and Desbat, 2010).

FTIR spectroscopy is not currently being used like MALDI-TOF MS, PCR or PFGE, but with continual studies, it may become the new "gold standard". Analysis by ATR-FTIR spectroscopy is rapid, reagent-free, and easy to learn, small amounts of sample are required and the machine itself has the potential to become smaller in size to save laboratory space. Not only can ATR-FTIR spectroscopy be used with microbial samples, it is very versatile as it can also analyze blood, fungi, pus, tissue, and other bodily fluids and samples (Petibois and Desbat, 2010). A drawback of ATR-FTIR spectroscopy is that while MALDI-TOF MS is automated, ATR-FTIR spectroscopy is currently not, but it has potential to be automated.

2.7 Using ATR-FTIR Spectroscopy for the Identification of *E. coli* and *Shigella*

As mentioned before, MALDI-TOF MS is currently being used in hospitals for the detection and identification of pathogens from patients' samples. However, MALDI-TOF MS has difficulties in differentiating between *E. coli* and *Shigella*. ATR-FTIR spectroscopy shows the potential to differentiate between *E. coli* and *Shigella* and among strains of the same species of bacteria as the spectra contain information about all the biomolecules the bacteria are comprised

of (Wenning et al., 2014). This tool can become very useful for surveillance and detection purposes.

2.7.1 Limitation for the Discrimination between *E. coli* and *Shigella*

Difficulties with the discrimination between *Shigella* and *E. coli* arise because they have similar biochemical compositions and physiology (Van den Delb et al., 2012). In various studies in which the relationship between *Shigella* and *E. coli* was examined, it was proven that the two are related species within the genus *Escherichia* (Zuo et al., 2013, Lan and Reeves, 2002). Even with evidence to prove that *Shigella* and *E. coli* belong to the same genus, today, *Shigella* is still categorized as a genus with four species (*flexneri*, *sonnei*, *boydii* and *dysenteriae*). It is said that in the past centuries, *Shigella* evolved divergently from *E. coli*. Although they are genetically related, the distinction between them is needed for the diagnosis of dysentery (varying severity, different diagnosis) (Van den Beld et al., 2012).

All four species of *Shigella* are non-motile while over 80% of *E. coli* strains are motile. Also, *Shigella* is always lysine decarboxylase (LDL) negative while *E. coli* is always LDL positive. Another distinction between the two is that when tested with D-glucose, *E. coli* will form a gas while *Shigella* will test negative for the formation of gas.

2.7.2 Using MALDI-TOF MS as a Tool for Differentiating *Shigella* from *E. coli*

Many studies using MALDI-TOF MS found that the differentiation between *Shigella* and *E. coli* is difficult because of their biochemical and physiological similarities (Paauw et al., 2015). However, in a recent study in which a custom MALDI-TOF MS reference library was developed for this differentiation, 94.4% correct classification of *E. coli* and *Shigella* was obtained (Paauw et al., 2015). In another MALDI-TOF MS study, also focusing specifically on differentiation between *Shigella* and *E. coli,* only 3% of isolates were misclassified (Khot and Fisher, 2013).

2.7.3 Considerations for Potential Use of FTIR Spectroscopy for the Discrimination of *E. coli* and *Shigella*

Currently there are many ongoing studies on the use of FTIR spectroscopy for the identification, discrimination and detection of microorganisms. There has not been a specific study on the discrimination of *E. coli* and *Shigella* using FTIR spectroscopy. Other studies have been completed for spores, lactic acid producing bacteria and foodborne pathogens, including *E. coli* and *Shigella* (but not solely for the two) (Wenning et al., 2014). While MALDI-TOF MS has difficulties in differentiating between *Shigella* and *E. coli*, ATR-FTIR spectroscopy proves to have potential in differentiating between them. Unlike MALDI-TOF MS, which detects the ribosomal proteins, FTIR spectroscopy detects molecular vibrations stemming from all biochemical components of the organisms (Wenning et al., 2014), generating a complete microorganism fingerprint. FTIR spectroscopy not only can discriminate down to the strain level but also does not require reagents, is non-destructive to the samples, is very fast and easy to use, and is very cost effective (Maity et al., 2013). Moreover, once a spectral database has been generated, the identification of microorganisms can be done almost in an instant (Maity et al., 2013).

There are reasons why FTIR spectroscopy is not currently being used in hospitals and laboratories, such as the method has not been established yet, a large database is needed for classification and identification, pure or purified cultures are needed, and the spectra of microrganisms are sensitive to their growth conditions as well as spectral acquisition conditions (Mauer, 2008). It is essential to employ a standard protocol, and conditions for culturing bacterial samples (media, temperature, time and sources of bacteria) must be strictly controlled; otherwise the performance of ATR-FTIR spectroscopy will be decreased compared to MALDI-TOF MS (Wenning et al., 2014). Although there are some disadvantages, with future modification and standardization, there is a possibility that ATR-FTIR spectroscopy will become a certified method as there are many advantages. Some of the advantages are that it does not require reagents, a small amount of sample is needed, it takes about 2 minutes to acquire data, and each microorganism has a unique spectrum (Mauer, 2008).

2.8 General Conclusion

To protect the general public, it is important to detect foodborne illness before it reaches the potential to become an outbreak. Fast detection methods that are accurate, precise, inexpensive, easy to use and sensitive are highly desirable. With ongoing research in ATR-FTIR spectroscopy related to the identification, characterization, detection and discrimination of microorganisms, more data are being generated that demonstrate the potential of FTIR spectroscopy as a method that can complement MALDI-TOF MS for use in reference and public health laboratories. Foodborne illness outbreaks can be prevented with rapid pathogen screening and the health of many can be maintained.

Many hospital laboratories have already adopted MALDI-TOF MS, and the addition of ATR-FTIR spectroscopy could be also desirable and may allow patients who suffer from *Shigella* and *E. coli* infections to be diagnosed faster and more accurately. Furthermore, ATR-FTIR instrumentation has the potential to become more portable and could be suitable for point-of-care testing which aims at the detection of pathogens directly from patients once samples have been extracted by the health care providers. With new upcoming methods and technologies, the improvement of the health care system can be made possible.

Chapter 3 – MATERIALS AND METHODS

The following sections of the chapter will discuss the methods and growth media employed for culturing bacteria to prepare samples for ATR-FTIR spectroscopic analysis, the procedure followed to acquire ATR-FTIR spectra of the bacterial samples, and the analysis of the spectral data.

3.1 Culturing Technique

McGill University Health Center (MUHC), Health Canada (HC) and Laboratoire de Santé Publique du Québec (LSPQ) generously provided all the strains of *E. coli* and *Shigella* employed in this research. The samples had been isolated from stool, blood or urine. The bacterial samples provided were in 3-mL vials and had been suspended in glycerol and frozen at -80°C for long-term storage. All samples used for the thesis research were cultured from frozen samples on agar plates. The first culture was plated via the 4-quadrant streak plate method. A single isolated colony was used for sub-culturing, likewise via the quadrant streak plate method (Figure 3.1). The procedure entailed using sterile wooden sticks to scrape the surface of the frozen glycerol-bacterial suspension and transfer it onto the plating media where sterile disposable streaking loops were used to collect the single isolated colony for sub-culturing (Figure 3.1).

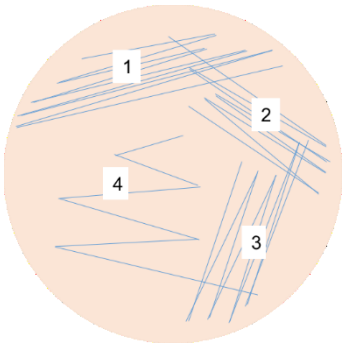


Figure 3.1. Quadrant streak plate method for the isolation of a single colony and to determine pure cultures.

Using a disposable loop, the sample was streaked 4 times in the 4 different quadrants of the agar plate as depicted in Figure 3.1. This initial culturing method facilitates the identification of mixed or contaminated bacterial cultures. For the research, only pure cultures are of interest for the development and validation of an infrared spectral database of microorganisms. After the sample was streaked on the agar plate, the plates were incubated at 37ºC for 18-24 h. Once the initial incubation was completed, the sample was sub-cultured and was again incubated at 37ºC for 18-24 h. Sub-culturing the frozen samples will ensure that the shock of being frozen is depleted and the bacteria are fully hydrated after being sub-cultured (Iizuka and Hasegawa, 1968).

3.2 Sample Preparation

Using the culturing technique described above, different sample sets required different culture media. The culture media used are summarized in Table 1. After obtaining bacterial growth, the samples are ready for ATR-FTIR spectroscopic analysis.

Table 1 - Culture Media Used in Preparation of Bacterial Samples for ATR-FTIR Analysis

Level of identification	Sample set	Media
Gram positive/Gram negative	MRSA, MSSA, *E. faecalis*, *E. faecium*	Blood agar
Genus	*E. coli* & *Shigella*	MacConkey agar
Species	*Shigella sonnei* & *Shigella flexneri*	MacConkey agar
Serotype	*E. coli* O157:H7 and Non-O157:H7	MacConkey agar
Toxin production	STEC	Soribol MacConkey agar

3.2.1 Growth Media

3.2.1.1 Blood Agar

For the purpose of this research, blood agar (BA) was used to grow most of the bacteria. BA is a nutrient-rich agar which contains 5% sheep blood to aid in the growth of bacteria. Generally, most microorganisms can grow on BA. BA is one of the most commonly used enriched media for cultivating microorganisms. In surveillance laboratories, it is used to isolate pathogenic

bacteria from non-pathogenic bacteria. Additionally, because the media contains blood, it can differentiate between α, ß and γ hemolytic bacteria (Murray, Rosenthal, & Pfaller, 2015).

3.2.1.2 MacConkey Agar

MacConkey agar (MAC) is not an enriched medium like BA but it is a selective medium that grows only Gram-negative bacteria such as *E. coli* and *Shigella*. Since MAC has a pH indicator that will select for lactose fermenters, those bacteria that are lactose fermenters will change the color of the media to pink upon incubation. The media will turn pink because when lactose is being fermented, an acid is produced in the process, lowering the pH and turning the colonies and media pink (Murray et al., 2015).

3.2.1.3 Sorbitol MacConkey Agar

Sorbitol MacConkey agar (SMAC) is a selective medium that is used to select *E. coli* O157:H7. It is a modified version of MAC where instead of selecting for lactose fermenters, it is selecting for sorbitol fermenters. *E. coli* O157:H7 does not ferment sorbitol and therefore will render a negative SMAC result. This method is used in surveillance laboratories to detect *E. coli* O157:H7 contaminations in food samples or individual human specimens (Murray et al., 2015).

3.3 Data Acquisition

After bacterial samples have been cultured and sub-cultured, infrared spectra of the bacteria were acquired as follows. First, the diamond sampling surface (~2 mm) of the ATR-FTIR spectrometer was cleaned with ethanol and dried, and the background spectrum was collected. Bacteria from 2-3 isolated colonies were then lightly scraped off the agar plate, ensuring that no agar was taken. The biomass was smeared onto the sampling surface of the spectrometer, and a spectrum was collected immediately (Figure 3.2). Spectral acquisition was carried out by using Cognisolve[1] software (Cognisolve Inc., Montréal, Canada). All spectra were acquired by co-addition of 64 scans at a resolution of 8 cm^{-1}.

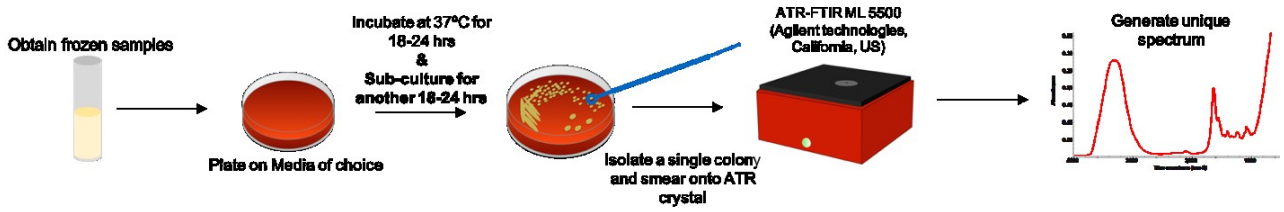


Figure 3.2. Flow diagram of obtaining sample from frozen storage, culturing, sub-culturing, and collecting sample for ATR-FTIR spectral acquisition.

Each strain was sampled in triplicate to evaluate spectral reproducibility. The three spectra acquired were combined to give a better representation of the sample (providing all three spectra were nearly identical).

3.4 Data Analysis

In order to ensure that interference from water vapour was minimized in the ATR-FTIR spectra of bacteria, a background spectrum was recorded before each spectral acquisition of a bacterial sample. Figure 3.3 shows representative residual water vapour absorptions that would result from improper spectral acquisition protocol (i.e., spectrum of sample recorded a couple of minutes after recording the background spectrum).

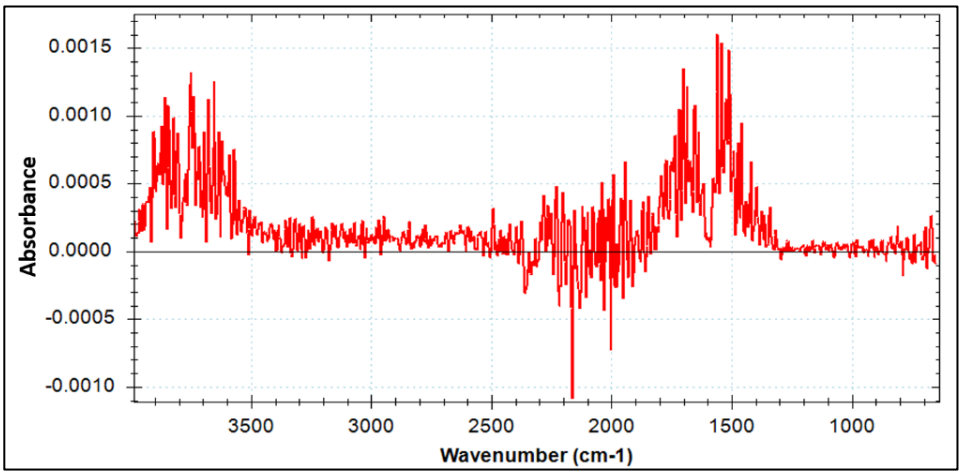


Figure 3.3. Residual water vapour absorptions in a background spectrum ratio against another background spectrum.

Examination of the bacterial spectrum (Figure 3.4) reveals that the spectral region between 1480 and 980 cm^{-1} contains information from all the biomolecules in an intact bacterial cell. These include infrared absorptions from polysaccharides, proteins, lipids and DNA.

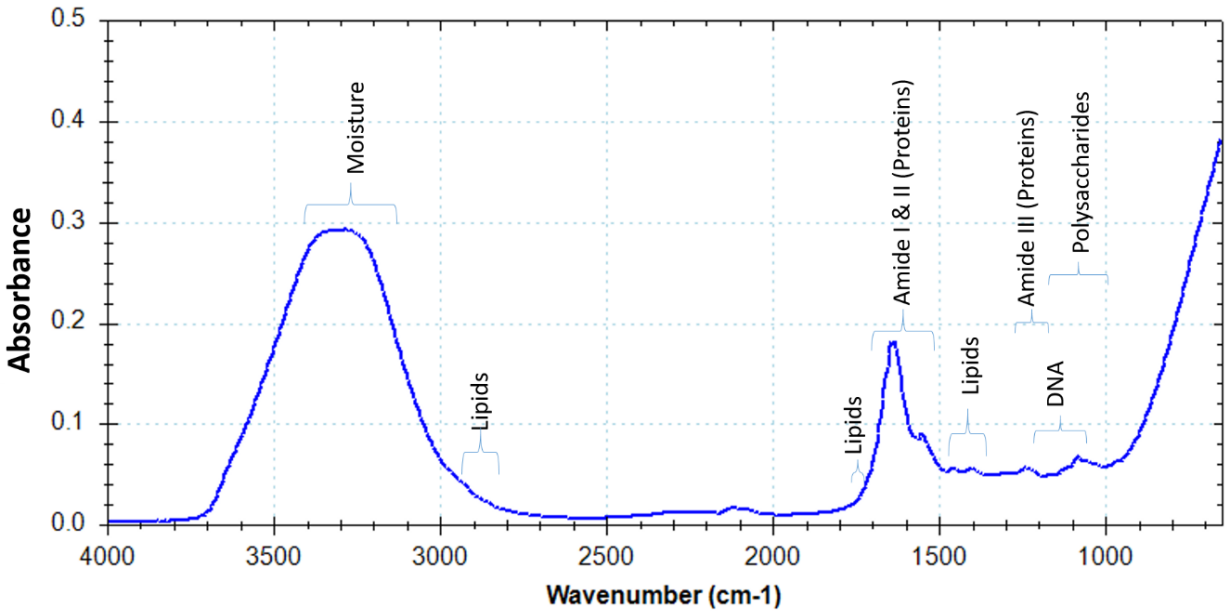


Figure 3.4. Typical bacterial spectrum. The specific regions in which water and the major biochemical components of bacterial cells absorb are indicated.

It is necessary to acquire the spectrum of a bacterial sample immediately after removing it from the agar plate, as there is a large variability between the spectra recorded over time in terms of the moisture content (see Figure 4.3). Moreover, each bacterial species retains moisture differently. The aim is to have relatively constant moisture content in all the spectra in the database. Fortunately, *Escherichia* and *Shigella* species share similar physiology and retain moisture to a similar extent.

In order to eliminate any baseline deviation from the infrared spectra, all spectra were transformed by taking the first derivative. Secondly, normalization of the spectrum was carried out to compensate for biomass variability among different samples.

3.4.1 Principal Component Analysis

Principal component analysis (PCA) reduces the dimensionality of the spectral dataset to condense the information into a more valuable and concise representation. The largest variance within the dataset would be in the first principal component (PC), and the second largest variance would be in the second PC, so on and so forth (Kirkwood, 2007).

By removing redundant spectral information and maintaining important information, PCA has been widely used for spectral analysis. Using PCA, there is a great reduction in computational time for spectral data analysis.

3.4.2 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) is used to determine the relationship between the components of a dataset. Using the divisive method of HCA, given a dataset, the data will be clustered into smaller sets where the relationship between the clusters will be determined. The data in this thesis will be represented by dendrograms where the relationships between the clusters are represented by branches. The relatedness of the clustering groups will be determined by the length of the connecting branches of the dendrogram (Kirkwood, 2007).

Chapter 4 – RESULTS AND DISCUSSION

4.1 Discrimination between Gram-Positive and Gram-Negative Bacteria Using ATR-FTIR Spectroscopy

The possibility of differentiation between microorganisms using ATR-FTIR spectroscopy is driven by the information content in the infrared spectrum that reflects the diversity of the biochemical composition of the microorganisms.

Typically, a Gram-positive (GP) bacterial cell would have a thick peptidoglycan layer composed of stacked chains of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) residues that are cross-linked by short peptide chains. Gram-negative (GN) bacteria, on the other hand, have a linear polysaccharide chain of alternating, NAG and NAM residues, and hence the peptidoglycan layer is much thinner than that of GP bacteria (Seltmann & Holst, 2013). The peptidoglycan layer of GN bacteria is so thin that it only accounts for 15-20% of the cell wall whereas in GP bacteria it accounts for 60-90% (Harisha, 2005). Moreover, another difference between the two is that GP bacteria have teichoic acids linking the different layers via covalent bonds, including the peptidoglycan layer. The outer membrane of GN bacteria is composed of lipopolysaccharides and phospholipids. The outer lipopolysaccharide (LPS) layer is composed of different sugars like ketodeoxyoctonate, glucose, glucosamine sugars, heptose and others (Seltmann & Holst, 2013). The presence of LPS is also one of the characteristics of GN bacteria; it is not present in GP bacteria (Knirel & Valvano, 2011). Depending on the microorganism, the arrangement and composition of the sugars define as well the antigenic structures (Seltmann & Holst, 2013).

In the current study, a database comprising ATR-FTIR spectra of various clinical strains was developed. A summary of the strains used, which included both GN and GP bacteria, is presented in Table 2.

Table 2 - Summary of Bacterial Samples Used in Preliminary Studies for Distinguishing between Gram-Positive and Gram-Negative Bacteria

	Genus	Number of samples
Gram Negative	*Escherichia*	369
	Shigella	
	Acinetobacter	
	Aeromonas	
	Campylobacter	
	Citrobacter	
	Enterobacter	
	Klebsiella	
	Salmonella	
	Pseudomonas	
	Neisseria	
	Stenotrophomonas	
	Serratia	
Gram Positive	*Bacillus*	210
	Enterococcus	
	Listeria	
	Staphylococcus	
	Streptococcus	
	Lactococcus	

PCA of the ATR-FTIR spectra of the bacterial samples in Table 2 resulted in complete separation of the GP and GN bacteria on the PC scores plot for the second and third PCs (Figure 4.2). The spectral region employed, which was selected by using a forward-search algorithm included in the software, was 1000-980 cm^{-1}.

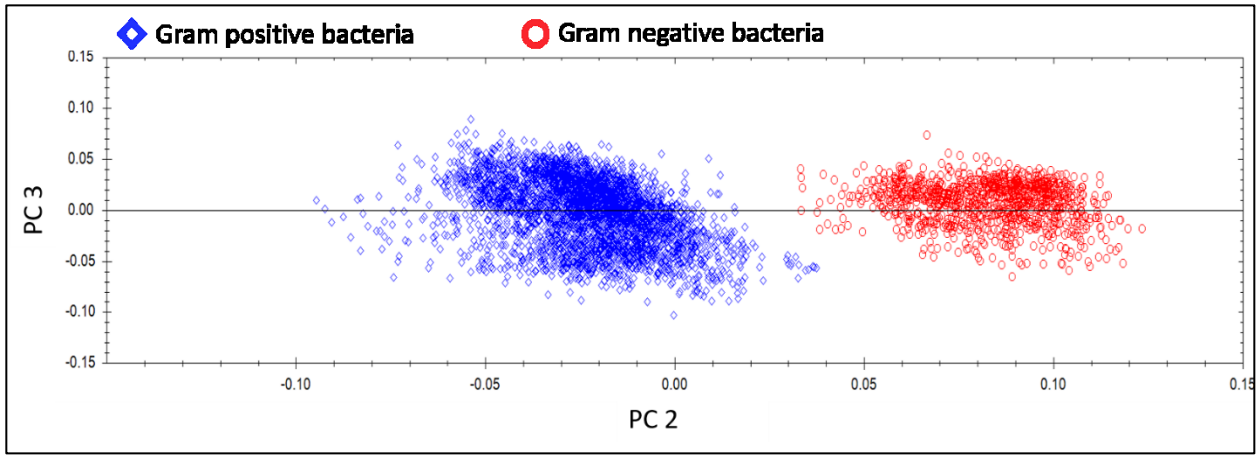


Figure 4.1. Principal component scores plot showing separation of Gram-positive and Gram-negative bacteria using principal components 3 and 2.

In Figure 4.2, the average spectra of 50 GN bacteria and of 50 GP bacteria are superimposed. There are obvious differences in the relative peak intensities stemming from protein, polysaccharide and phosphate absorptions.

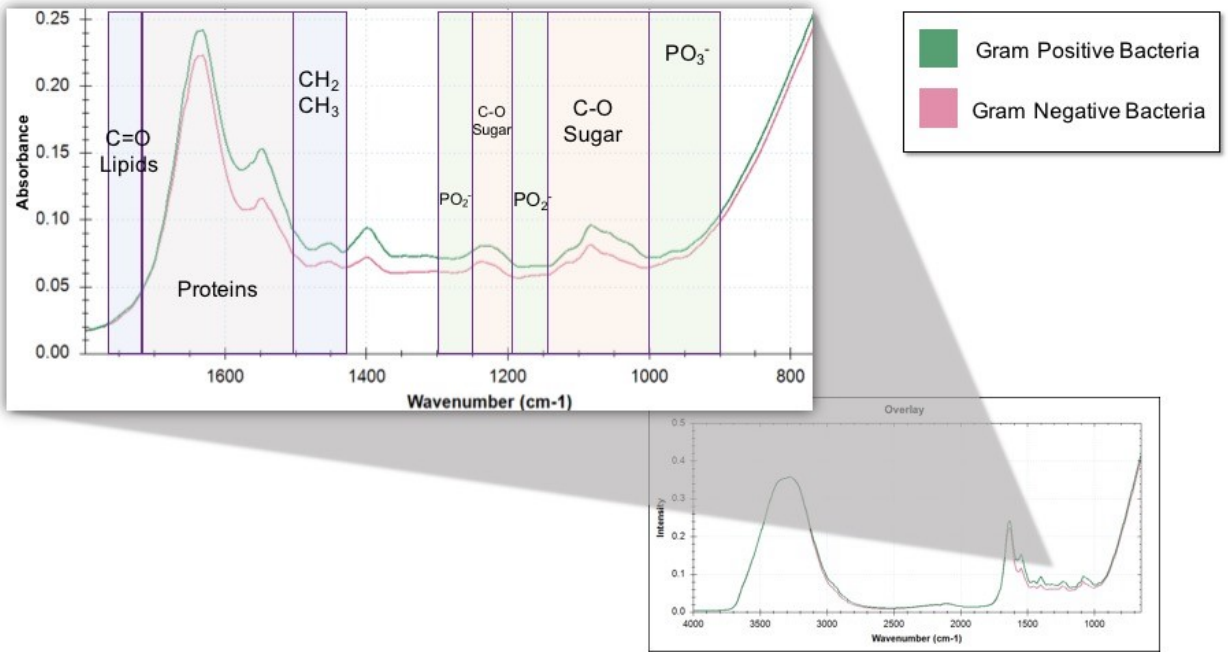


Figure 4.2. Superimposed average spectra of Gram-positive and Gram-negative bacteria and a close-up of spectral differences in relation to their biochemical compositions.

Previous studies have exhibited clustering of the infrared spectra of GN and GP bacteria based on the presence or absence of teichoic acid (Alvarez-Ordóñez & Prieto, 2012). After the investigation of GN and GP bacteria in the present study, it is evident that the separation is again possible

4.2 ATR-FTIR Spectroscopic Examination of *E. coli* and *Shigella* Species

Within the food industry, GN bacteria such as *Escherichia coli* and *Shigella* are more important in terms of detection. The following sections will examine the possible differentiations between them using ATR-FTIR spectroscopy.

4.2.1 Discrimination between *E. coli* and *Shigella* Using ATR-FTIR Spectroscopy

Shigella and *E. coli* were plated on BA with 5% sheep blood to facilitate the growth of bacteria. Generally, in surveillance labs, BA and MAC are used as a preliminary test to select for the lactose fermenters or non-lactose fermenters. Being a lactose fermenter would be associated with Gram-negative bacteria such as *E. coli* and *Shigella* and, as mentioned before, MAC does not support the growth of GP bacteria.

Both *E. coli* and *Shigella* have been closely linked with foodborne illnesses. *E. coli* and *Shigella* are closely related genetically. Based on genome sequencing, they have at least 88% genetic similarities in the DNA backbone (Primrose & Twyman, 2013). Some of the differences between them are that *Shigella* does not ferment lactose, is lysine decarboxylase negative, is non-motile and has no gas formation when in contact with D-glucose.

The objective of this study is to utilize ATR-FTIR spectroscopy to determine if there are spectral differences between the two very similar genera. Given that *Shigella* and *E. coli* are not completely identical, their biochemical composition is different and may be detected via ATR-FTIR spectroscopy and can be further classified using PCA.

Table 3 - Summary of Spectra Used for the Analysis of *E. coli* and *Shigella* Species

	Total spectra collected	**Removed spectra***	**Outlier spectra**	**Total spectra analyzed**
E. coli	160	10	21	129
Shigella	160	28	21	111
TOTAL	320	38	42	**240**

*Spectra showing too low sample moisture content based on pre-set absorbance criteria were removed from the dataset.

The potential utility of ATR-FTIR spectroscopy for discrimination between *E. coli* and *Shigella* was investigated with a total of 108 samples, of which 54 were *E. coli* and 54 were *Shigella*. Spectra were collected in triplicate for each sample, resulting in a total of 320 spectra. Putting the spectra into a spectral analysis software and setting rejection criteria based on absorbance values as described below, 38 spectra showing too low sample moisture content were rejected. An

additional 42 spectra were removed as outliers based on the results of PCA. In Section 4.3 of the thesis, more will be discussed on outliers. After removing all rejected spectra and outliers, a total of 240 spectra were used for the following evaluation.

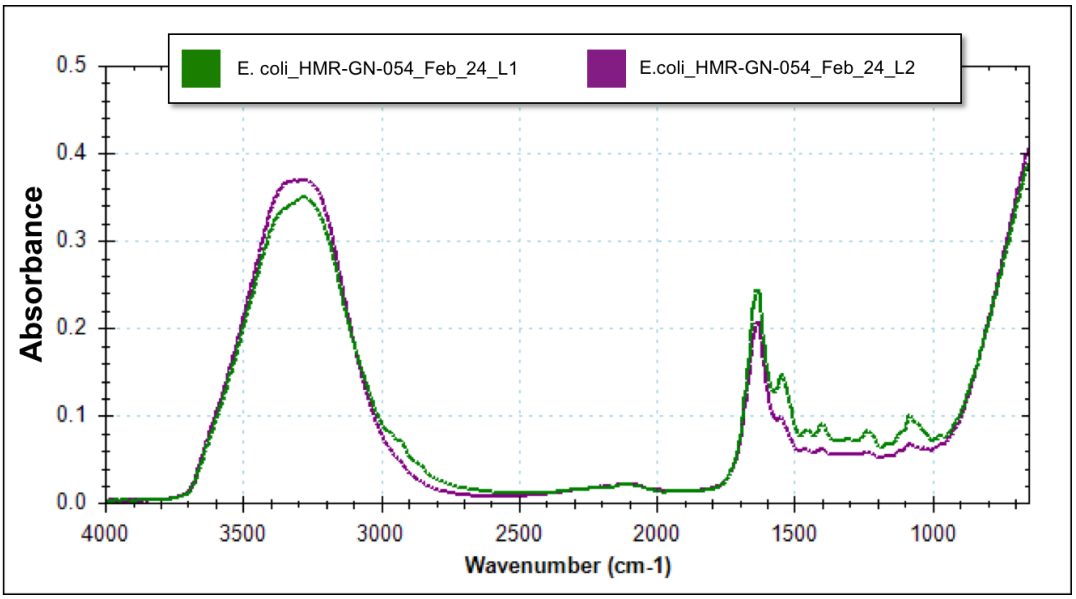


Figure 4.3. Superimposed spectra showing the effects of differences in sample moisture content on the spectrum of an *E. coli* sample.

To maintain consistency between spectra, all samples should have comparable moisture contents. From Table 3, 38 spectra were removed because of their lower moisture content using criteria based on the water absorption band centered at 3333 cm^{-1}. Figure 4.3 depicts representative duplicate spectra of an *E. coli* sample (E.*coli*_HMR-GN-054_Feb_24_Lx). At 3333 cm^{-1}, the purple spectrum has an absorbance value of 0.378, while the green spectrum has an absorbance value of 0.345. In this study, any spectra having an absorbance at 3333 cm^{-1} below 0.375 were removed to allow consistency and facilitate a reliable comparison between strains. For the given example in Figure 4.3, the green spectrum would be removed. By taking three spectra of each sample, it generally allowed for at least one good spectrum per sample. Removing spectra that do not meet the benchmark moisture absorbance eliminates some of the variability between samples. With the moisture band standardized, the region of 1800-800 cm^{-1} can be compared among spectra without the moisture content affecting the comparison.

Other studies have used drier samples where the absorbance of the moisture band is below 0.300. By using lower moisture content samples, there was great variability in the moisture content, making spectrum-to-spectrum comparisons difficult (Cook, 2016; Mauer & Reuhs, 2008). The purpose of using the maximum moisture content possible is again to limit the variation of the spectra, to obtain an average spectrum that represents the bacterial group, and to acquire a better understanding of the spectral differences that can be translated to differences in biochemical composition of the microorganisms.

If MAC is used as the culture medium, *Shigella* species will give a negative result for lactose fermentation whereas *E. coli* will give a positive result. Since MAC is a selective medium, it does change the biochemical composition of the bacteria. To minimize the effect of varying culture media on the infrared spectra, BA was therefore used.

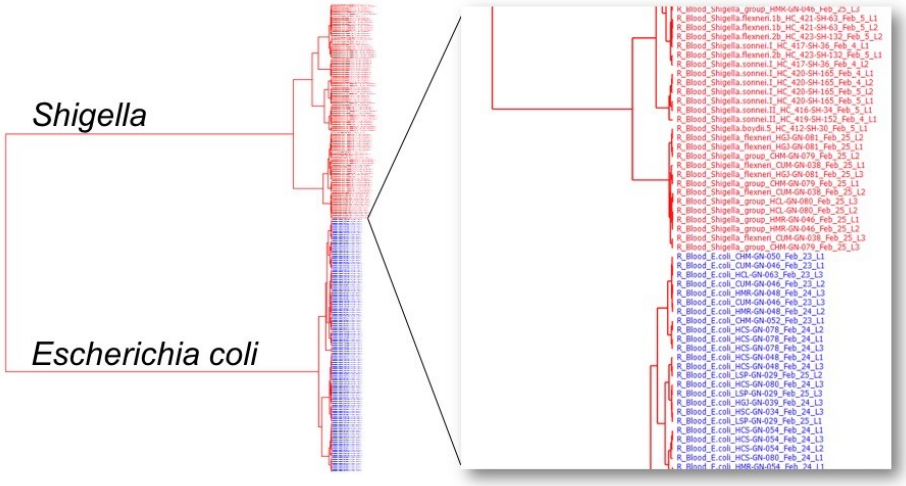


Figure 4.4. Dendrogram showing the complete discrimination between *Shigella* and *E. coli* by hierarchical cluster analysis of ATR-FTIR spectral data.

Using hierarchical cluster analysis (HCA), a dendrogram showing the complete discrimination between *Shigella* and *E. coli* was obtained (Figure 4.4). This dendrogram is based on spectral distances, computed using the cosine metric, within the region of 1478-1411 cm^{-1} selected through the use of a forward-search feature selection algorithm. The spectral bands in this region may be assigned in part to carboxyl side-chain absorptions of amino acids and $–CH_2$ bending vibrations

of lipids (Kirkwood, 2007, Mauer & Reuhs, 2008, Beekes et al., 2007). The *Shigella* and *E. coli* spectra split into two different branches of the dendrogram. Within each branch, other variations between the species are observed via the smaller branches. Discrimination between *Shigella* and *E. coli* was also obtained by PCA employing the spectral region between 1478 and 1411 cm^{-1}as depicted in Figure 4.5. By examining the spread of the points on the PC scores plot, we can better comprehend how diverse each group of spectra is. The spectra of *E. coli* cluster closer together in comparison to those of *Shigella*, which are more dispersed. However, the two groups of spectra are not diverse enough to overlap with one another. Again, PCA has proven that the spectra of *Shigella* and *E. coli* are different enough for separation.

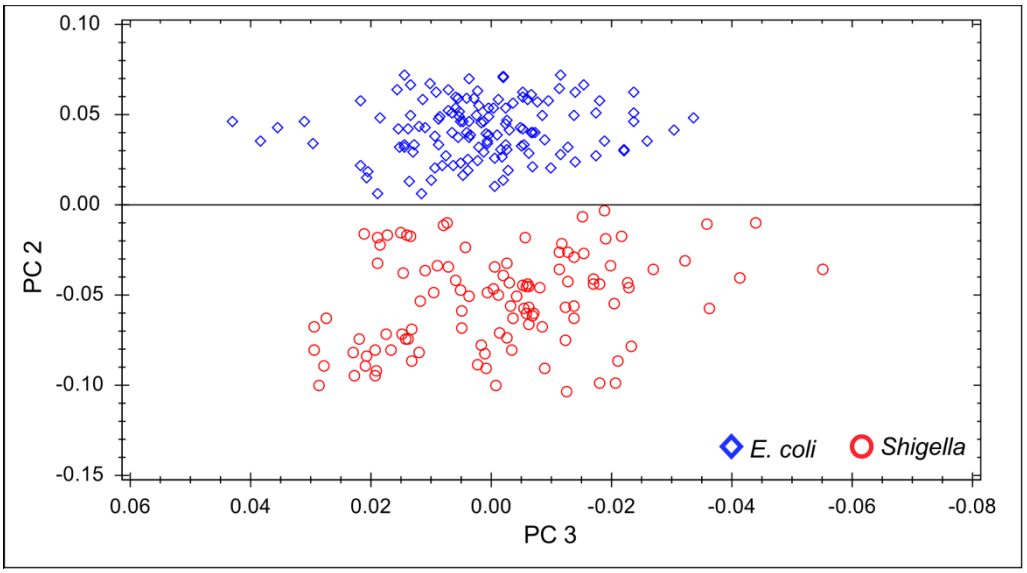


Figure 4.5. Principal component scores plot obtained using the spectral region 1478-1411 cm^{-1} showing separation of *E. coli* and *Shigella* based on principal component 2.

Figure 4.6 depicts the average ATR-FTIR spectra of the *E. coli* strains (in red) and the *Shigella* strains (in blue). The average spectra were computed from the final spectral dataset, which consisted of 129 and 111 spectra of *E. coli* and *Shigella*, respectively. The top panel of Figure 4.6 depicts the spectra in the full mid-IR region. From the literature, *E. coli* and *Shigella* are similar and this is reflected by the average spectra of the organisms.

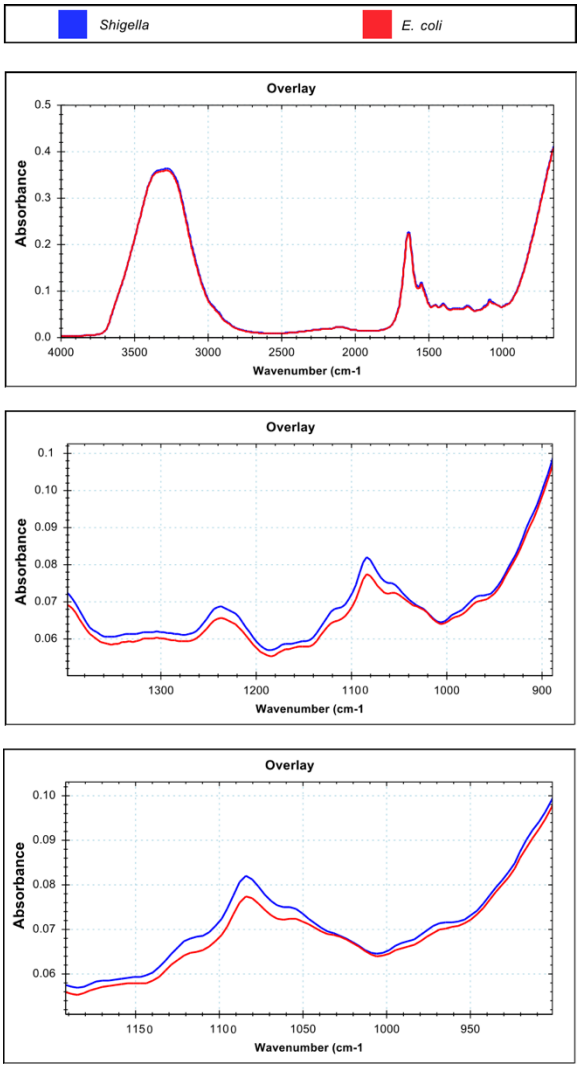


Figure 4.6 Average ATR-FTIR spectra of *E. coli* (red) and *Shigella* (blue).

A close-up of the 1400-900 cm^{-1} region is depicted in the middle panel of Figure 4.6. Absorption between 1120 and 1105 cm^{-1} is associated with DNA and RNA backbone vibrations and C-O-C and C-O ring vibrations of carbohydrates (Kirkwood, 2007, Beekes et al., 2007). The average spectra indicate slight differences in DNA/RNA composition between the two genera. In the middle panel of Figure 4.6, at 1027-998 cm^{-1}, *Shigella* has a relatively lower absorption band compared to *E. coli*. From the literature, the absorption peaks in this region correspond to C-O-H bending vibrations (Mauer & Reuhs, 2008).

Additionally, over the ~1070-1040 cm^{-1} region (bottom panel of Figure 4.6), higher absorbance is observed for *Shigella* than for *E. coli*. The region is associated with backbone vibrations of DNA/RNA, carbohydrates and proteins (Kirkwood, 2007). In terms of differences in protein content between *Shigella* species and *E. coli*, it has been observed that *Shigella* species have higher protein content on their cell membrane than *E. coli.* IcsA is a protein that is found on the cell membrane of *Shigella* species and helps in facilitating infection of the host cells in the human body (Goldberg & Theriot, 1995). In the case of *E. coli*, the protein is present within the cell but does not easily move to the cell membrane as a higher O-antigen length suppresses the protein from reaching the surface. Accordingly, IcsA is highly concentrated on the cell membrane of *Shigella* species compared to the cell membrane of *E. coli* (Robbins et al., 2001). It is possible that ATR-FTIR spectroscopy has detected the higher amounts of IcsA found in *Shigella* species than in *E. coli*. The spectral differences allowing for discrimination between *Shigella* species and *E. coli* may also be related to infection rates and virulence since toxins are proteins.

If differences in protein content are responsible for the discrimination between *Shigella* species and *E. coli* by ATR-FTIR spectroscopy, then discriminatory information should be contained in the spectral ranges of 1700-1600 cm^{-1} and 1575-1485 cm^{-1} which are associated with the strongest protein absorption bands, the amide I and amide II bands (Beekes et al., 2007). PCA was performed using the above ranges to investigate the separation of *Shigella* and *E. coli* based on their protein profiles (Figure 4.7).

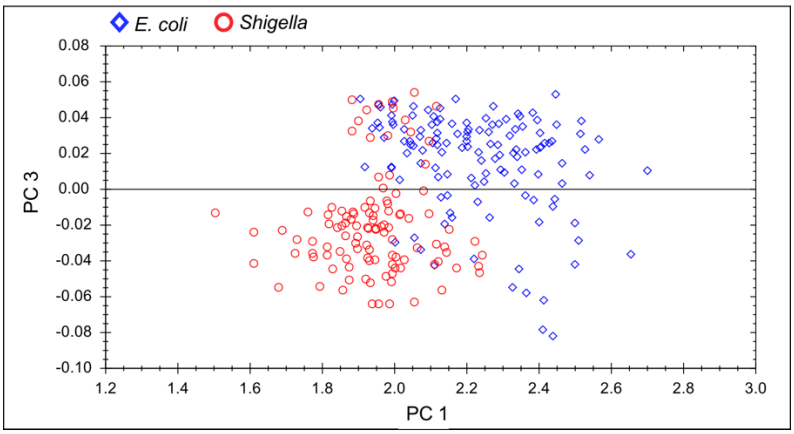


Figure 4.7. Scores plot for principal components 1 and 3 from principal component analysis of amide I and amide II bands in the ATR-FTIR spectra of *E. coli* and *Shigella* species.

As observed in the above figure, the separation between *E. coli* and *Shigella* was not distinct as there is an overlap between the two groups. This proves that there is some separation due to proteins like IcsA, but the distinction is not absolute. Moreover, using only a portion of the amide I band, 1695-1685 cm^{-1}, a clearer separation is observed (Figure 4.8).

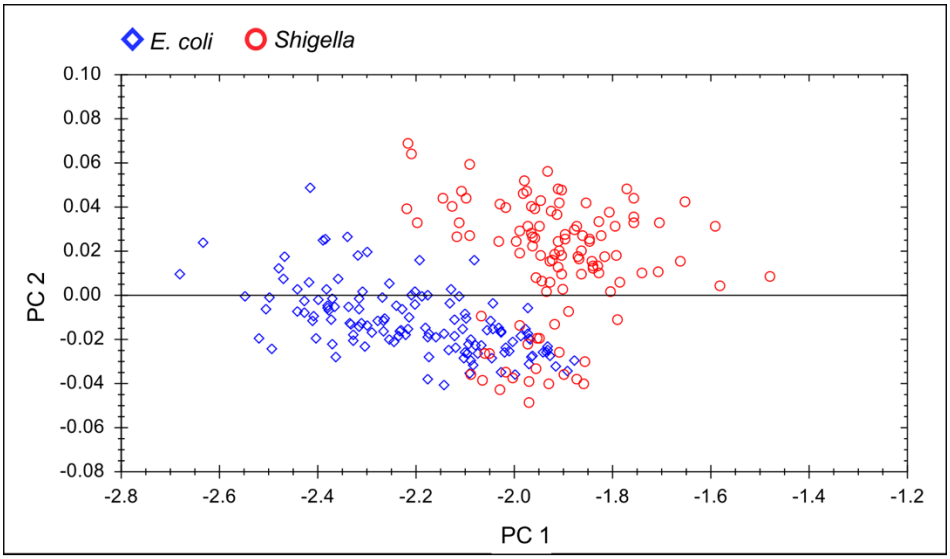


Figure 4.8. Scores plot for principal components 1 and 2 from principal component analysis of the region 1695-1685 cm^{-1} in the ATR-FTIR spectra of *E. coli* and *Shigella* species.

As seen in the above figure, the separation between *E. coli* and *Shigella* species is almost possible using only the portion of the amide I band between 1695 and 1685 cm^{-1}, however, some *Shigella* samples were overlapping with *E. coli* samples. Upon further investigation, the *Shigella* samples overlapping with *E. coli* in Figure 4.8 were all cultivated from the MUHC. Moreover, the MUHC *Shigella* samples were overlapping with *E. coli* samples from the MUHC. Perhaps the methods of freezing samples or sample preparation were different from laboratory to laboratory and affected the protein contents. Nonetheless, PCA has proved that there are significant variations between *Shigella* and *E. coli* in terms of their protein content.

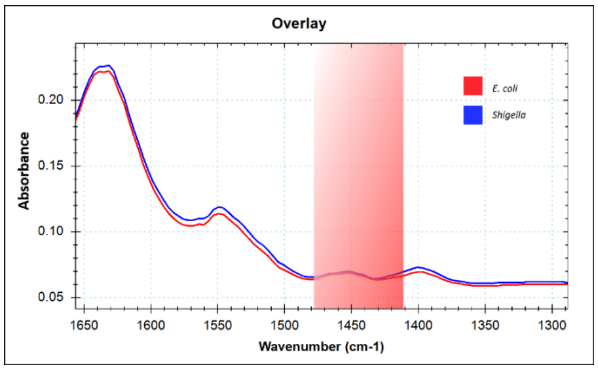


Figure 4.9. Average spectra of *E. coli* and *Shigella* in the region 1478-1411 cm^{-1}.

The spectral region that was selected by the forward-search algorithm for the separation of the two groups (1478-1411 cm^{-1}) shows subtle variation between the two average spectra (region highlighted in red in Figure 4.9). These differences allowed for complete separation between *E. coli* and *Shigella* (Figure 4.5). Using regions with maximum variance (e.g., 1120-1105, 1027-998, 1070-1040 cm^{-1}) resulted in a poorer separation between the two genera (Figure 4.10). This may be attributed to differences in the metabolic profiles of different strains within each genus. Accordingly, region selection based on large spectral differences may not be an effective means of optimizing bacteria differentiation.

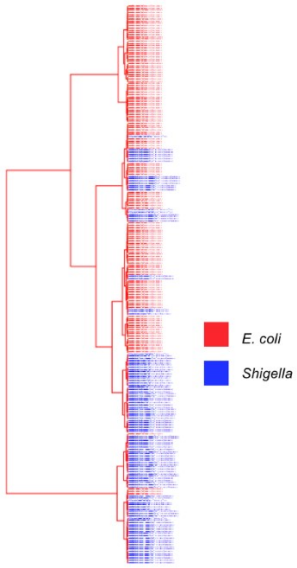


Figure 4.10 Dendrogram generated from HCA using 1120-1105, 1027-998, and 1070-1040 cm^{-1} regions for the separation of *E. coli* and *Shigella.*

ATR-FTIR spectroscopy was successfully used for the separation of *E. coli* and *Shigella* through both HCA and PCA. The average spectra of *E. coli* and *Shigella* were examined to show the differences in biochemical composition between the two. From literature and comparison to the two average spectra, it is apparent that ATR-FTIR spectroscopy can detect slight biochemical differences that allow for the *E. coli* and *Shigella* discrimination from one another by both HCA and PCA.

4.2.2 Discrimination at the Species Level of *Shigella flexneri* and *Shigella sonnei*

In addition to successful discrimination between genera, as shown in the previous section, ATR-FTIR spectroscopy may also be used for the separation between two species of the same genus. In the present study, discrimination between *Shigella flexneri* and *Shigella sonnei* by ATR-FTIR spectroscopy was investigated.

Table 4 - Summary of Spectra Collected for Discrimination between *Shigella flexneri* and *Shigella sonnei*

	Total spectra collected	**Removed spectra***	**Outlier spectra**	**Total spectra analyzed**
Shigella flexneri	246	24	9	213
Shigella sonnei	230	8	1	221
TOTAL	476	32	10	**434**

*Spectra showing too low sample moisture content based on pre-set absorbance criteria were removed from the dataset.

As mentioned in earlier sections of the thesis, *S. flexneri* and *S. sonnei* are the two most prevalent species of the genus *Shigella* that cause foodborne illnesses. A total of 71 samples of *Shigella flexneri* and 74 samples of *Shigella sonnei* were included in this study. The samples were plated onto MAC agar. Since both species are from the *Shigella* genus, both will not be lactose fermenters and will give a negative result on MAC. Although they are both negative, *Shigella sonnei* has been reported to be a slow lactose fermenter (Ito, 1991). Considering this difference in lactose metabolism, the biomolecule compositions of the two species after growth on MAC will have significant variations for discrimination.

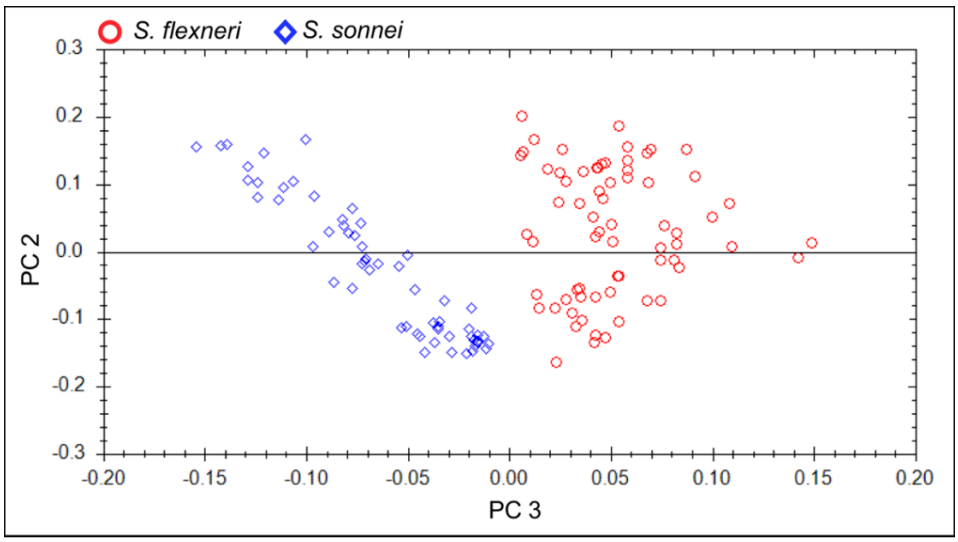


Figure 4.11. Scores plot for principal components 3 and 2 from principal component analysis of the ATR-FTIR spectra of *Shigella flexneri* and *Shigella sonnei* using the whole spectral region of interest (1480-900 cm^{-1}).

The two species displayed enough spectral variations to allow discrimination between them by PCA (Figure 4.11) using the whole spectral region of interest (1480-900 cm^{-1}). Upon processing with a forward-search algorithm (Ghetler, 2010) to remove the non-relevant spectral features in the spectra, the regions of 1136-1113 and 1218-1207 cm^{-1} enhanced the separation between the two groups and identified an outlier (Figure 4.12). According to the literature, absorptions in these regions stem from nucleic acids as well as carbohydrates (Kirkwood, 2007). The three spectral points in the bottom left corner of Figure 4.12 originate from one sample of *Shigella sonnei*, which is considered an outlier as it does not cluster with the rest of the group. Upon further investigation, the sample was found to be a mixed culture.

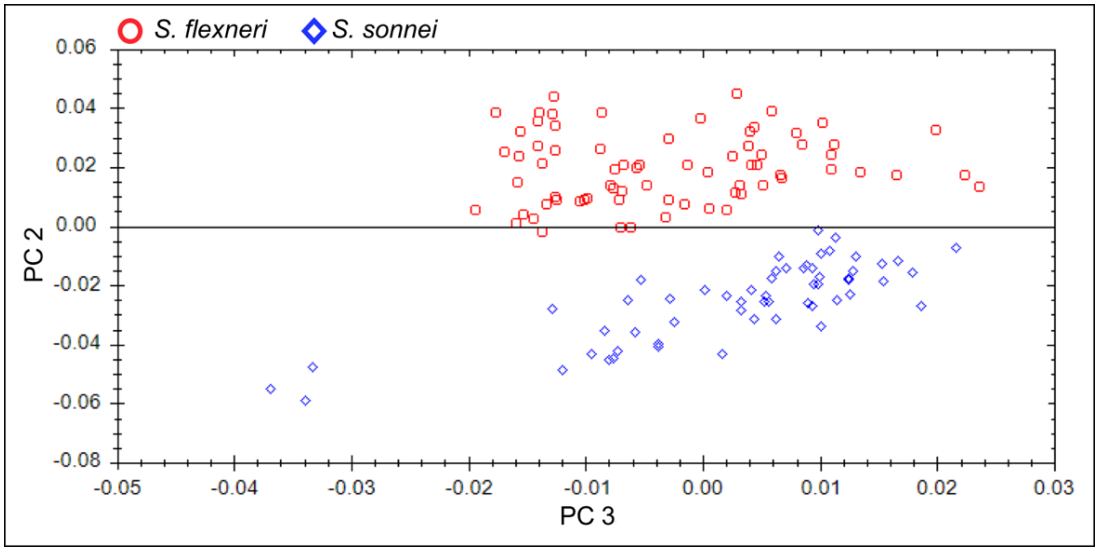


Figure 4.12. Scores plot for principal components 2 and 3 from principal component analysis of the ATR-FTIR spectra of *Shigella flexneri* and *Shigella sonnei* using the regions of 1136-1113 and 1218-1207 cm^{-1}. The three spectral points in the bottom left corner of the plot originate from one sample of *Shigella sonnei*, which is considered an outlier.

The spectral differences between the two species may result from their difference in lactose fermenting abilities. *Shigella flexneri* does not ferment lactose whereas *Shigella sonnei* does, but at a very slow rate. Although *Shigella sonnei* ferments lactose at a slow rate, during the 18-24 h of incubation, the samples were observed to be negative on MAC. When plating the samples on MAC, upon visual observation the plates were all lactose negative for both *Shigella flexneri* and *Shigella sonnei*. Perhaps ATR-FTIR spectroscopy detected slight lactose fermentation by *Shigella sonnei* which is causing the separation. Another possible cause for separation is differences in the lipopolysaccharides (LPS) that form part of the outer membrane of the cell wall. The outer membrane is broken down into three components, the Lipid A, core polysaccharide and the O-antigen, which is part of the polysaccharide component (Knirel & Valvano, 2011). *Shigella flexneri* has been more studied than *Shigella sonnei*, but in comparison to *E. coli* there are some slight variations and differences between the two. *E. coli* has a Lipid A that is composed of beta-D-glucosaminyl-(1-6)-alpha-D-glucosamine disaccharide where both *Shigella flexneri* and *Shigella sonnei* have the same $\beta(1\rightarrow6)$ disaccharide. The polysaccharide core is also similar among *Shigella flexneri*, *Shigella sonnei*, *E. coli* and *Salmonella*. The innermost layer of the core region has 2-keto-3-deoxy-D-manno-octulosonic acid (Kdo) and L-glycero-D-manno-heptose. For *Shigella* species, the outer region of the core is similar in the hexose region. Between *Shigella*

flexneri and *Shigella sonnei*, there are slight differences in the hexose region (Lindberg, Karnell, & Weintraub, 1991). Since there is a slight difference in LPS between the two species, it is possible that it contributes to the separation.

The averages of the spectra of each species were generated and are superimposed in Figure 4.13. The full spectral range is displayed in the top panel of Figure 4.13, and the average spectra of the two species show strong similarities. However, moving towards the second panel, there is an obvious difference in spectral absorbance within the 1173-953 cm^{-1} region in addition to the 1657-1623 cm^{-1} and 1575-1508 cm^{-1} regions in the bottom panel. Again, bands around 1160 cm^{-1} and 1076 cm^{-1} correspond to DNA and RNA backbone vibrations, which indicates that the two species have a different genetic make-up. Bands in the 1173-1000 cm^{-1} region are associated with C-O-C bending and C-O stretching vibrations of carbohydrates (Kirkwood, 2007). The difference in hexose composition between *Shigella sonnei* and *Shigella flexneri* may be reflected in this spectral region. At 1085 cm^{-1}, there is a P=O stretching band which is again associated with DNA and RNA.

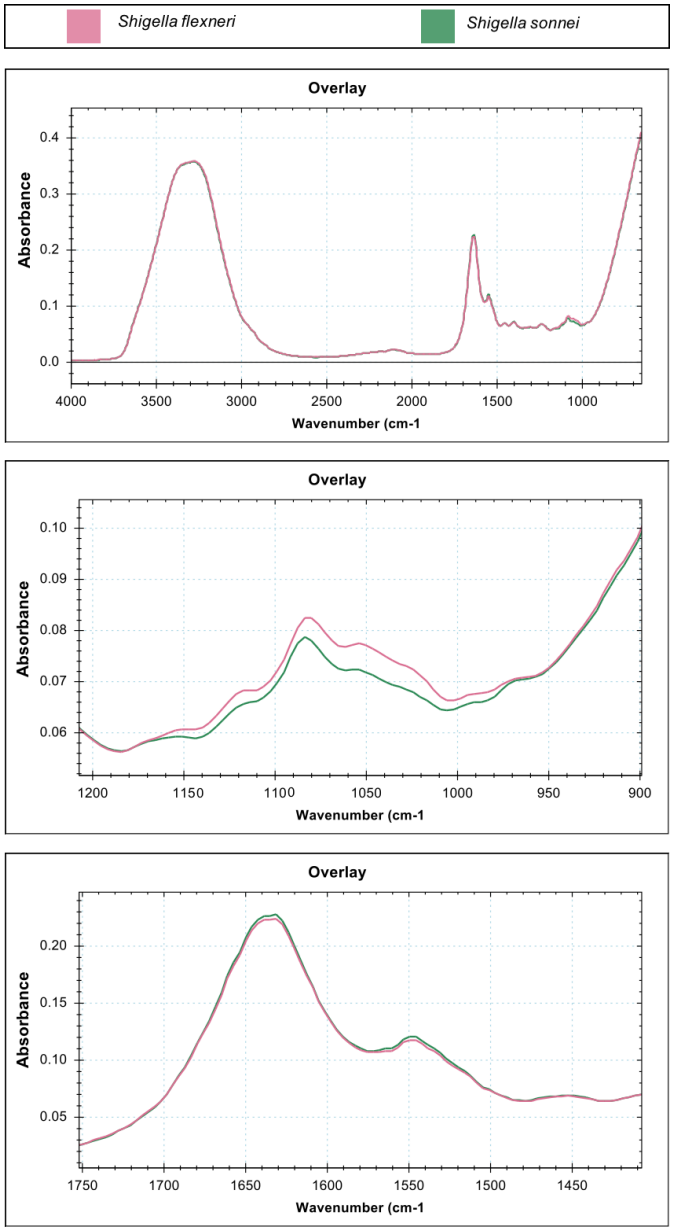


Figure 4.13 Average ATR-FTIR spectra of *Shigella flexneri* and *Shigella sonnei*.

The two main bands seen in the bottom panel of Figure 4.13 are the amide I and amide II bands associated with proteins. More interestingly, the band between 1515 and 1505 cm^{-1} can be assigned to the C=C stretch of tyrosine (Mauer & Reuhs, 2008, Beekes et al., 2007).What is more interesting is that tyrosine-phosphorylated proteins have been reported to regulate the virulence factors of *Shigella flexneri* (Standish et al., 2016). *Shigella flexneri* is one of the major causes of bacillary dysentery in humans, is linked with bloody diarrhea in children, and causes at least 14,000 deaths

in Asia per year (Standish et al., 2016). Figure 4.13 shows that the spectrum of *Shigella sonnei* has a higher tyrosine band than the spectrum of *Shigella flexneri*. Although there are not many studies on tyrosine-phosphorylated proteins in *Shigella sonnei*, there are recent studies in *Shigella flexneri* (Standish et al., 2016, Kuehl et al., 2014)

In general, for pathogenic bacteria, the ability to invade host cells is based on the virulence plasmid, where pathogenic bacteria have a large plasmid of about 220 kb (Standish et al., 2016). Different genes in the virulence plasmid encode how the invasion occurs, gene expression, and responses of the host. Most, if not all, pathogens have a type 3 secretion system (T3SS) encoded in their genes that allows the pathogens to inject proteins into the host cells until expression occurs (Standish et al., 2016). *Shigella flexneri* has been reported to have at least 905 sites of tyrosine phosphorylation in 573 of its proteins. Conversely, *E. coli* O157:H7 has 342 proteins with 512 sites of tyrosine phosphorylation. From that study, it was concluded that *Shigella flexneri* is more virulent than *E. coli* O157:H7 (Standish et al., 2016).

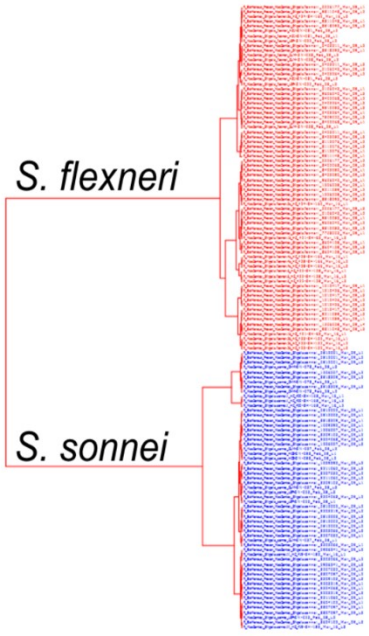


Figure 4.14. Dendrogram showing the separation of *Shigella flexneri* and *Shigella sonnei* obtained by hierarchical cluster analysis of the ATR-FTIR spectral data in the regions 1136-1113 and 1218-1207cm^{-1}.

Figure 4.14 shows the separation between *Shigella flexneri* and *Shigella sonnei* obtained by HCA using two spectral regions selected by the forward-search feature selection algorithm, 1136-1113 and 1218-1207cm^{-1}. Two distinct branches for the two species are observed, indicating that the variability among the spectra of each species is small relative to the spectral differences between the two species.

4.2.3 Discrimination between *E. coli* O157:H7 and non-O157:H7 *E. coli* Using ATR-FTIR Spectroscopy

Food safety has been a big concern for the last decade as there has been an increase in population and a higher demand for processed foods. With more food being in production at a faster rate, food safety has become better regulated. *E. coli* O157:H7 is one of the most common serotypes of *E. coli* causing outbreaks of foodborne illness, most notoriously the *E. coli* O157:H7 outbreak case in 1982. This study aims at investigating the potential of ATR-FTIR spectroscopy as a method of distinguishing between *E. coli* O157:H7 and non-O157:H7 *E. coli*.

Table 5 - Summary of Spectra Collected for Discrimination between *E. coli* O157:H7 and Non-O157:H7 *E. coli*

	Total spectra collected	Removed spectra*	Outlier spectra	Total spectra analyzed
Non-O157:H7 *E. coli*	183	33	9	141
***E. coli* O157:H7**	189	57	0	132
TOTAL	372	90	9	**273**

*Spectra showing too low sample moisture content based on pre-set absorbance criteria were removed from the dataset.

Pathogenic *E. coli* are separated into non-O157:H7 and O157:H7 (Figure 2.1). For this study, non-O157:H7 *E. coli* and *E. coli* O157:H7 were used, where non-O157:H7 *E. coli* belonged to the pathotypes UPEC, EPEC and STEC. The STEC pathotype did not include *E. coli* O157:H7, even though *E. coli* O157:H7 is classified as STEC. The samples that were provided for the study were collected by Health Canada and were all analyzed by PCR to validate their pathotypes. A total of 124 isolates were obtained, of which 63 were *E. coli* O157:H7, and 3 spectra were acquired for each isolate, giving a total of 372 spectra. Ninety spectra were removed based on moisture criteria described in Section 4.2.1. A total of 3 non-O157:H7 samples were removed as outliers. From an

initial set of 372 spectra (Table 5), only 273 were used to obtain the results presented in the figures below.

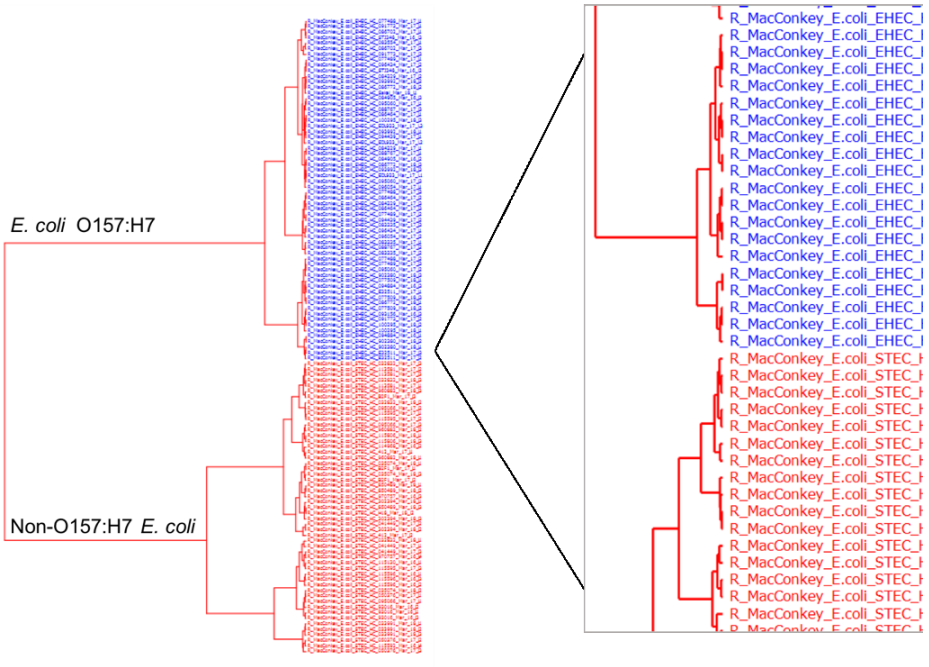


Figure 4.15. Dendrogram showing the perfect separation of *E. coli* O157:H7 and non-O157:H7 *E. coli* by hierarchical cluster analysis.

Using HCA, the separation of *E. coli* O157:H7 from non-O157:H7 *E. coli* based on their spectral differences was successful. Again, there is a significant separation between the non-O157:H7 *E. coli* pathotypes, which can be further investigated in a more detailed study. For the purpose of this study, the separation between EHEC and other *E. coli* pathotypes is of interest. This study is aimed at determining whether or not the common *E. coli* O157:H7 serotype can be distinguished from other Shiga-toxin-producing *E. coli*. It is important to identify *E. coli* O157:H7 to eliminate the process of serotyping all of the samples.

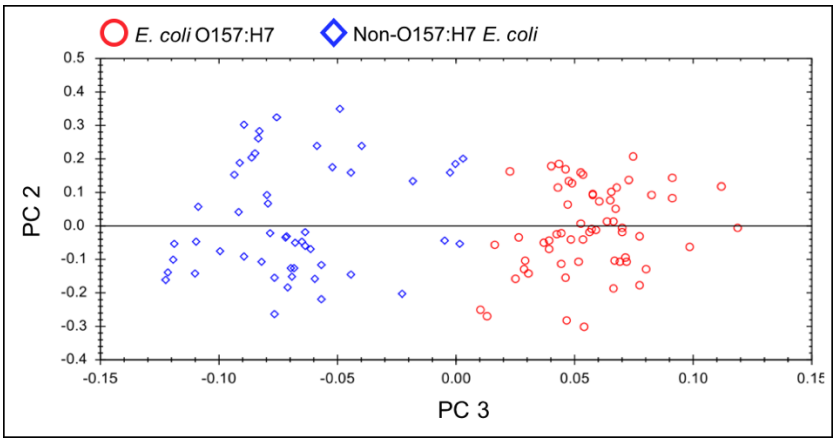


Figure 4.16. Scores plot for principal components 2 and 3 obtained by principal component analysis of the ATR-FTIR spectra of non-O157:H7 *E. coli* and *E. coli* O157:H7.

Figure 4.16 shows the results obtained by PCA using two spectral regions selected by the forward-search region selection algorithm, above, 1248-1212 and 1356-1344 cm^{-1}. As visually observed, there is a separation between non-O157:H7 *E. coli* and *E. coli* O157:H7, although some points in the non-O157:H7 group are close to the O157:H7 group. With the regions stated above, the separation was a success but not definite. The selected spectral regions indicate that there are differences between the two classified groups in RNA/DNA backbone vibrations and in polysaccharides (Kirkwood, 2007). Although the separation of the groups on the PC scores plot in Figure 4.16 is not as distinct as that seen in Figure 4.11, the results obtained from HCA, using the same regions as for PCA, show that the two groups are dissimilar enough to separate into two branches (Figure 4.15).

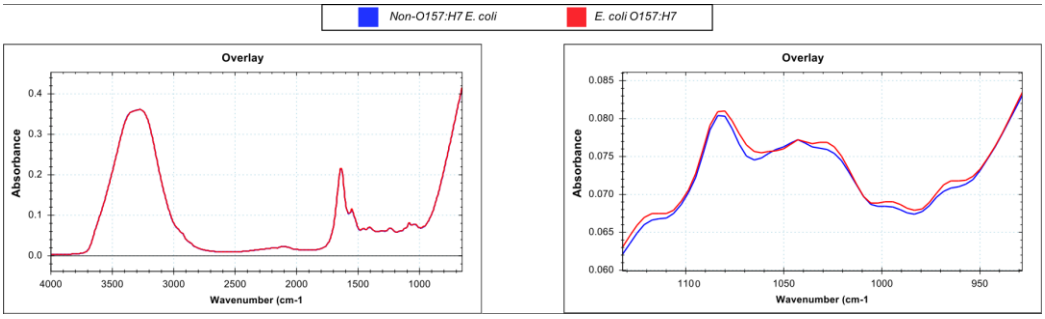


Figure 4.17. Average spectra of non-O157:H7 *E. coli* and *E. coli* O157:H7 for the elucidation of biochemical differences.

Examination of the average spectra of *E. coli* O157:H7 and non-O157:H7 *E. coli* revealed a few differences in the region 1140-950 cm^{-1} (Figure 4.17). Major differences between the two groups were found in the following regions: 1247-1213 cm^{-1}, 1143-1102 cm^{-1}, 1087-1053 cm^{-1}, 1042-1012 cm^{-1}, 1005-975 cm^{-1} and 968-949 cm^{-1}. These may be attributed to variations in PO_2 stretching of phosphodiester bonds in DNA and RNA and phospholipids and variations in carbohydrate composition.

The main difference between *E. coli* O157:H7 (EHEC) and non-O157:H7 *E. coli* is that EHEC is more virulent and is more likely to cause an outbreak. Because EHEC produces more Stx2 than Stx1 in vitro, the virulence of EHEC is higher than that of other STEC (Baker et al., 2007). Also, Stx2 is an AB toxin, which is a protein that consists of two components and is responsible for causing the internal cell infection (Mantis, 2012). The amide II band (1586-1482 cm^{-1}) in the average spectrum of EHEC has a slightly higher absorbance, which may be associated with increased Stx2 production (Figure 4.17).

4.2.4 Discrimination between Shiga-Toxin Producers: STEC and *Shigella*

The aim of this study was to determine if STEC and *Shigella* species can be differentiated by ATR-FTIR spectroscopy given that they both produce Shiga toxins. The spectra employed to investigate discrimination between STEC, including *E. coli* O157:H7, and *Shigella* species are summarized in Table 6. The *Shigella* species that were used included *Shigella flexneri*, *Shigella sonnei*, *Shigella dysenteriae*, *Shigella boydii* and some unknown *Shigella* species that are listed in the appendix.

Table 6 - Summary of Spectra Collected for Shiga-Toxin-Producing *E. coli* (STEC) and *Shigella* Species

	Total spectra collected	**Removed spectra***	**Outlier spectra**	**Total spectra analyzed**
STEC	154	31	3	120
Shigella	218	41	7	170
TOTAL	372	72	10	**290**

*Spectra showing too low sample moisture content based on pre-set absorbance criteria were removed from the dataset.

The set of 125 samples, of which 52 were STEC, was collected from Health Canada and McGill University Health Center. Upon analysis of the 372 spectra collected for these samples, 72 spectra were removed due to absorbance criteria based on moisture content, and there were about 10 outlier spectra (~ 4 outlier isolates), leaving a total of 290 spectra.

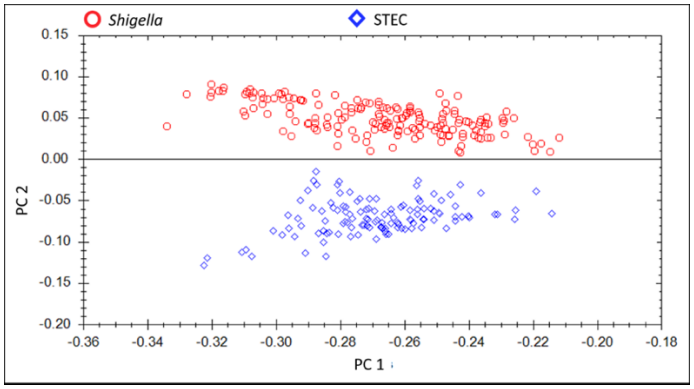


Figure 4.18. Scores plot for principal components 1 and 2 obtained by principal component analysis of the ATR-FTIR spectra of *Shigella* species and STEC.

Figure 4.18 illustrates a PC scores plot that demonstrates the complete separation between *Shigella* and STEC by PCA employing the spectral regions 1288-1270, 1367-1348, and 1426-1408 cm^{-1}. These regions were selected by the forward-search feature selection algorithm employed throughout this research.

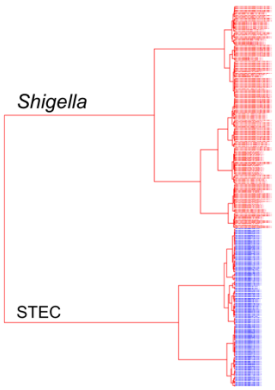


Figure 4.19. Dendrogram showing the separation of STEC and *Shigella* by hierarchical cluster analysis.

Figure 4.19 depicts the separation between *Shigella* and STEC obtained by HCA employing the same spectral regions. The difference between comparing *Shigella* and STEC versus *E. coli* and *Shigella* is that not all *E. coli* strains produce Shiga toxin whereas *Shigella* and *E. coli* O157:H7 both produce the toxin that causes illnesses upon consumption. Accordingly, Figures 4.18 and 4.19 illustrate the separation between two groups of toxin producers.

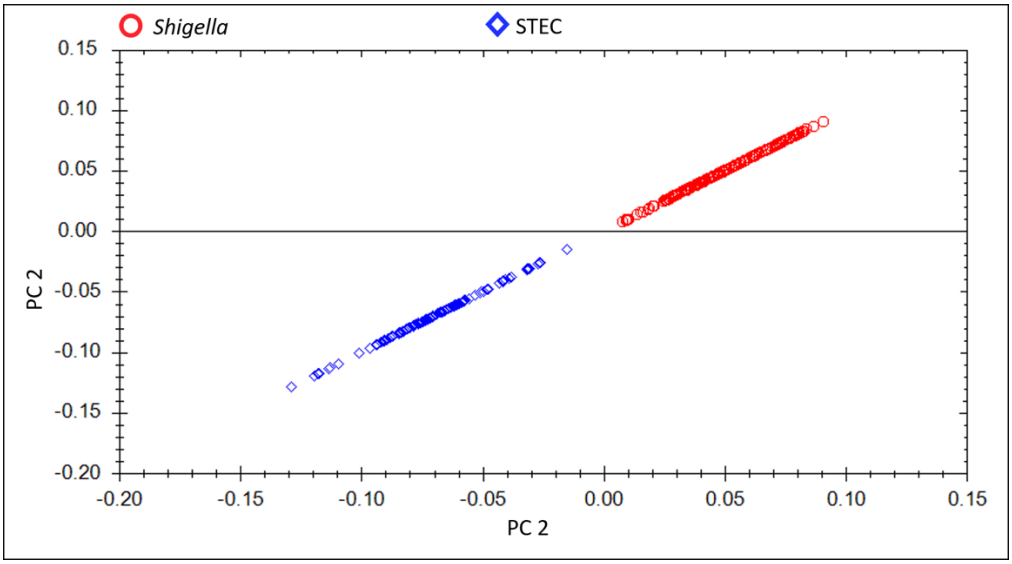


Figure 4.20. Principal component scores plot for *Shigella* and STEC using principal component 2 on both axis.

Figure 4.20 shows the discrimination between *Shigella* and STEC based only on the scores of principal component 2 from PCA of the ATR-FTIR spectral data in the regions of 1288-1270, 1367-1348, and 1426-1408 cm^{-1}. The points representing *Shigella* species in Figure 4.20 are clustered closer together than those representing STEC. The greater diversity among STEC could be because EHEC is more virulent compared to other Shiga-toxin-producing *E. coli*.

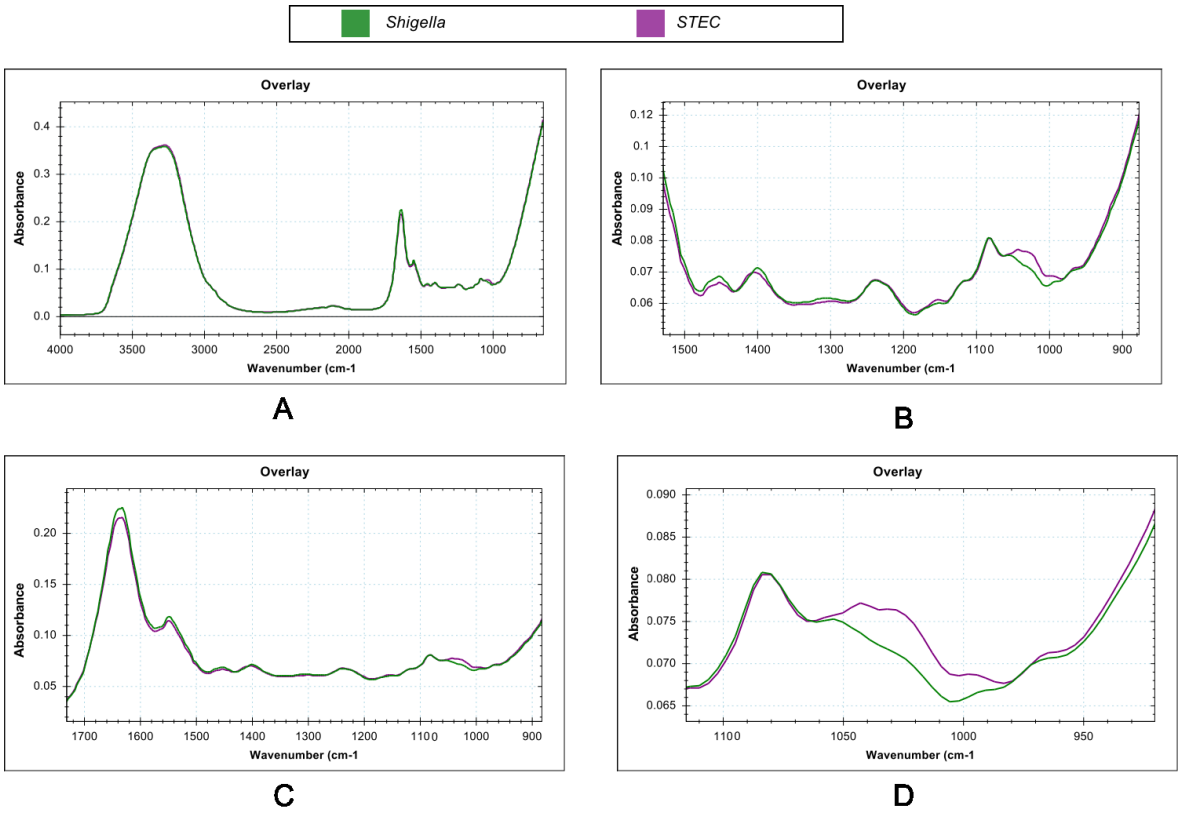


Figure 4.21. Analysis of average spectra of *Shigella* and STEC for the elucidation of biochemical differences.

Comparison of the average spectra of *Shigella* species (*S. flexneri* and *S. sonnei*) and STEC revealed some interesting differences between them. In panel A of Figure 4.21, the average spectra over the full spectral range are superimposed. The overlaid spectra in panel B of the same figure show that the major spectral differences between the two are in the region of 1061-975 cm^{-1}, which can also be seen in panels C and D of the same figure. This region contains absorption bands of DNA at 970 cm^{-1}, assigned to the O-P-O bending vibration, and at 1053 cm^{-1}, assigned to the C-O stretching vibration of deoxyribose (Mello & Vidal, 2012). Other sugar moieties also have absorption bands in this region. In panel C of the above figure, it is shown that the bands in the region 1675-1433 cm^{-1} are higher in the *Shigella* spectrum than in the STEC spectrum. This higher spectral absorbance is associated primarily with the amide I and amide II bands of proteins. More specific information about the biochemical differences that give rise to the spectral differences between *Shigella* species and STEC cannot be elucidated by visual examination of the average

spectra of the two groups. Accordingly, visual spectral analysis does not provide full information about the differences in biochemical composition that are the basis for the successful discrimination between the two groups by PCA and HCA

4.2.5 Discrimination among STEC by Type of Shiga Toxin Produced: Stx1, Stx2, and Stx2f

Going into more depth, the potential use of ATR-FTIR spectroscopy to discriminate between Shiga-toxin types produced in STEC is examined below.

There are different types of Shiga toxins produced by STEC where the most virulent toxin to cause foodborne illness is Stx2. Although there are Stx1- or Stx2-producing *E. coli* strains, some *E. coli* have exhibited the presence of both toxigenic genes, *stx1* and *stx2*. For the purpose of this study, ATR-FTIR spectroscopy will be used to classify STEC strains according to type(s) of toxin produced.

Table 7 - Summary of Spectra Collected for STEC Strains Classified According to the Shiga Toxin Gene(s) Detected by a Multiples-PCR

Shiga toxin gene(s) detected by PCR	**Total spectra collected**	**Removed spectra***	**Total spectra analyzed**
stx1	111	0	111
stx2	47	1	46
stx2f	9	0	9
stx1* & *stx2	19	0	19
TOTAL	187	1	186

*Spectra showing too low sample moisture content based on pre-set absorbance criteria were removed from the dataset.

Determining the toxigenicity of bacteria is complicated, time-consuming and costly. Only few samples were provided by LSPQ that had been tested by multiplex PCR to detect the genes that encode for Stx1 and Stx2. Although the sample set for this study is significantly smaller than the other sets studied in this research, it is still valid for preliminary evaluation.

The samples that were provided by LSPQ were grown on SMAC. SMAC is selective to *E. coli* O157:H7 where the agar will render a negative result for sorbitol fermentation. Upon growth on

SMAC, the 9 samples of *E. coli* that had tested positive for the presence of *stx2f* all gave a negative result on SMA, indicating that they are *E. coli* O157:H7 strains. Contrary to this finding, various studies in the literature have linked other serotypes like O63:H6, O113:H6 and O125:H6 with *stx2f* (Friesema et al., 2014). More research is needed to account for this discrepancy.

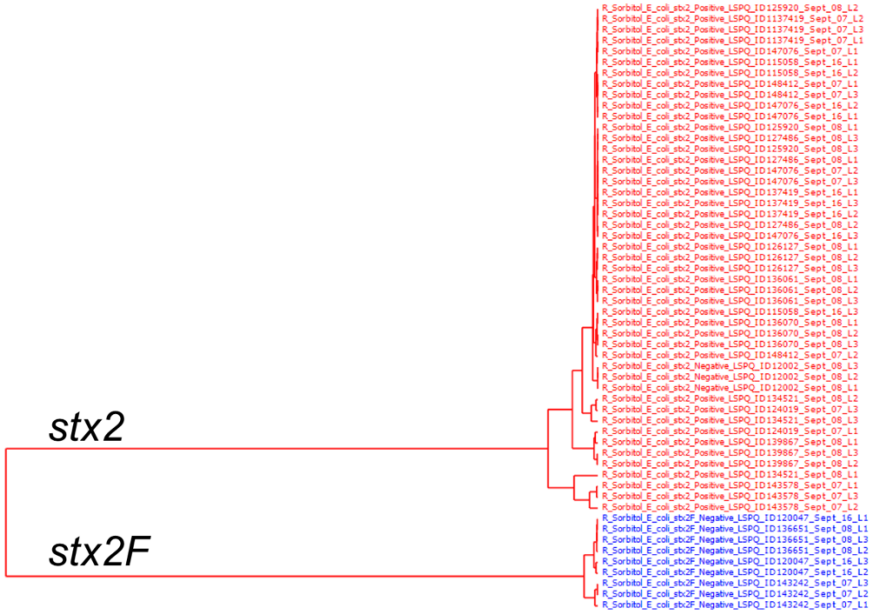


Figure 4.22. Hierarchical cluster analysis dendrogram showing the separation between *E. coli* producing Stx2 and Stx2f.

In Figure 4.22, the dendrogram obtained by HCA using the regions of 1061-1050, 1311-1296, and 1341-1329 cm^{-1} shows the separation between *E. coli* producing Stx2 and *E. coli* producing Stx2f. Again, since all the samples in the *stx2f* group gave a negative result on SMAC whereas those in the *stx2* group gave a positive result, the separation may be due to the different biochemical composition of the two groups in terms of growth on SMAC. Therefore, other variants of *stx2* are needed to confirm whether the separation is due to the different variants or to the difference in SMAC result. The samples could be grown on blood agar for confirmation of the separation, but again, LSPQ uses SMAC to determine *E. coli* O157:H7 first for PCR analysis, and therefore blood agar was not used.

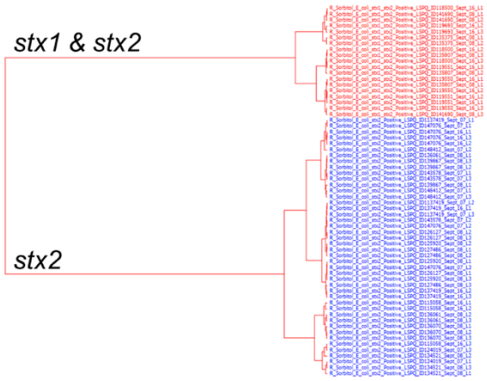


Figure 4.23. Hierarchical cluster analysis dendrogram showing the separation between *E. coli* expressing both *stx1* & *stx2* and *E. coli* expressing only *stx2*.

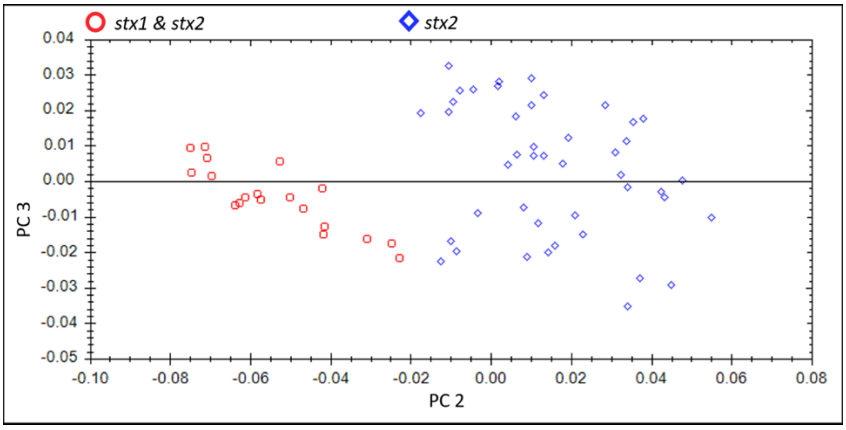


Figure 4.24. Scores plot for principal components 2 and 3 obtained by principal component analysis of ATR-FTIR spectra of *E. coli* strains expressing both *stx1* and *stx2* and those expressing only the *stx2* gene.

For the separation of *E. coli* strains expressing both *stx1* and *stx2* from those only expressing the *stx2* gene, the regions selected by the forward-search feature selection algorithm were 1140-1121, 1225-1214, and 1329-1318 cm^{-1}. Separation between the two groups was obtained by HCA using these regions (Figure 4.23) as well as by PCA (Figure 4.24). Although there is a separation between the two groups on the PC scores plot in Figure 4.24, the points belonging to each are not clustered closely together. As seen in Figure 4.25, HCA and PCA of the ATR-FTIR spectra of *E. coli* strains producing only Stx1 and those producing only Stx2 did not produce a distinct separation between the two groups.

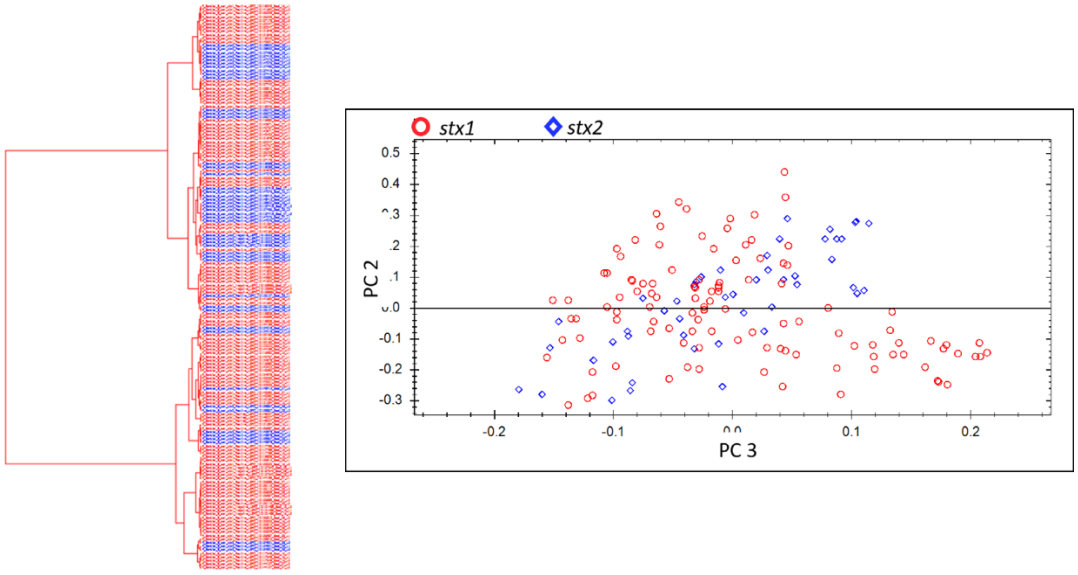


Figure 4.25. Clustering of ATR-FTIR spectra of *E. coli* strains producing only Stx1 and those producing only Stx2 by HCA (left) and PCA (right), showing that there are not enough spectral differences between the two groups to separate them.

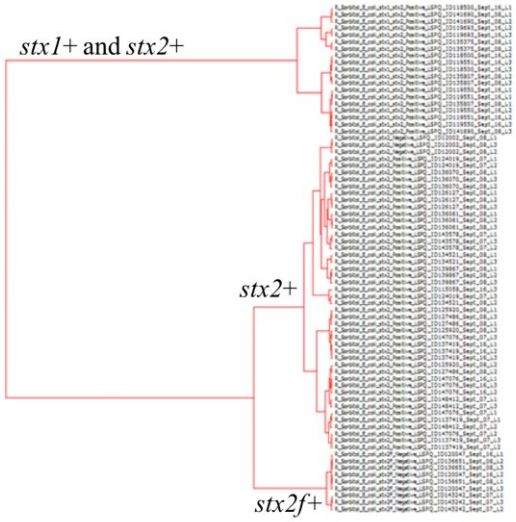


Figure 4.26. Dendrogram based on hierarchical cluster analysis of ATR-FTIR spectral data, showing discrimination between non-O157 STEC strains that are *stx1*+ and *stx2*+ and those that are *stx1*- and *stx2*+ as well as separate clustering of strains that are *stx2f*+.

HCA of the ATR-FTIR spectra of the *E. coli* strains expressing *stx1 & stx2, stx2* only, and *stx2f* was completed to show the relationships among the three groups. As seen in Figure 4.26, the dendrogram showed the relationship between the *stx2* and *stx2f* groups whereas the two groups had a distant relationship to *E. coli* expressing both *stx1* and *stx2* genes.

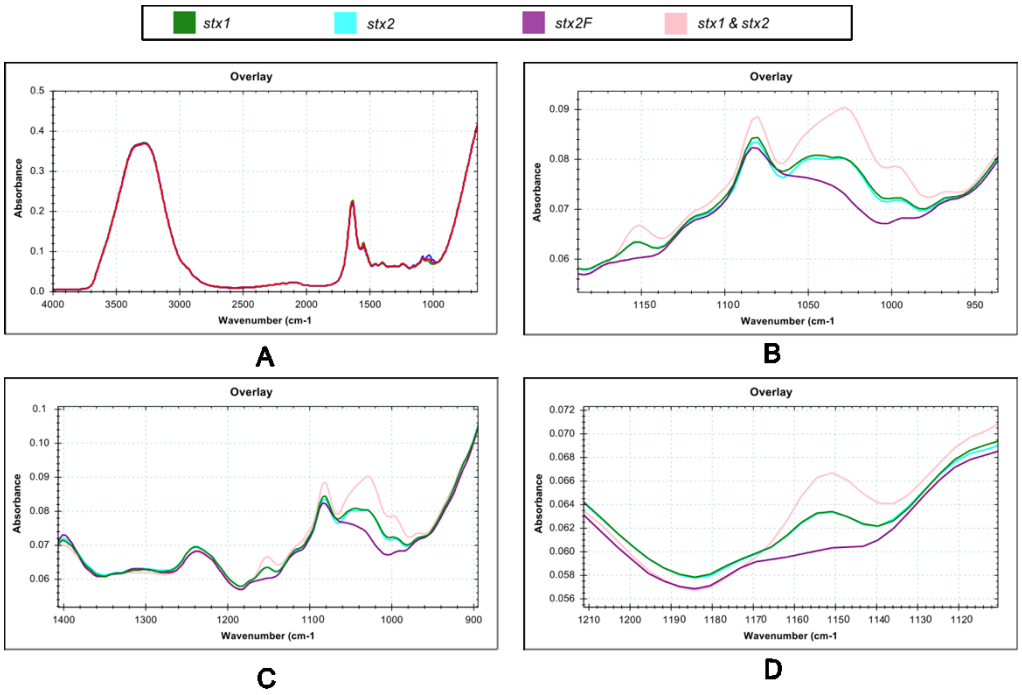


Figure 4.27. Average spectra of *stx1*, *stx2*, *stx2f* and *stx1 & stx2* groups of strains for the elucidation of biochemical differences.

The average spectra of the *stx1*, *stx2*, *stx1 & stx2* and *stx2f* groups of strains are superimposed in Figure 4.27. From the figure, overlay A is the general overlay of The whole mid-IR spectral range is shown in panel A whereas spectral regions showing variations are presented in panels B, C, and D. There is a visual similarity between the average spectra of the *stx1*and *stx2* groups, so that there will be difficulties distinguishing between them, as was shown in Figure 4.25. Interestingly, *E. coli* strains that have both *stx1* and *stx2* present show a very different spectral profile in the 1165-957 cm^{-1} region where the peaks at 1027, 1076, and 1150 cm^{-1} show significantly higher absorbance. The spectral range corresponds to various DNA and RNA backbone vibrations as well as some vibrations of carbohydrates.

Based on the literature, *stx2* is more commonly related with EHEC, which causes serious infections and illnesses. The present study investigating the potential of ATR-FTIR spectroscopy to distinguish between important toxins which cause harm to humans shows inconclusive results. Although some separation is exhibited, better understanding of toxins is required as well as a requirement for additional samples.

Given that Shiga toxin is an AB toxin which is composed of two protein structures, if ATR-FTIR spectroscopy has detected the toxins, then there should be a difference in protein spectral absorbance in the region of 1695-1515 cm^{-1} between STEC and non-STEC samples. Given that these samples were provided on SMAC, from the database collection, there are no other SMAC samples of non-STEC strains for comparison. The average spectra in Figure 4.27 all show similar absorbance in the protein region. Furthermore, the forward-search region selection algorithm did not find discriminatory information in the protein region.

For the preliminary study of using ATR-FTIR spectroscopy as a means of detecting toxin type of *E. coli*, more research is required. The present study has shown that there are limitations to the method in terms of specificity and sensitivity at the present time.

4.3 Outliers

Outlier in the thesis refers to those spectra that do not cluster with the majority of the other spectra. Spectral outliers in the previous sections were removed from the final analysis. Their removal was based on HCA and PCA, which showed that those spectra were different from the majority. Upon further investigation, the samples were further analyzed in the laboratory to re-identify the organism. By culturing the samples on BA and visually inspecting for more than one colony, the samples were identified as mixed cultures. Although the given samples were supposed to be pure, contamination from improper handling of the frozen samples may have occurred. As the samples are in fact mixed cultures, they were removed from the analysis as the purpose of the research is to identify pure cultures only.

4.4 Validation of ATR-FTIR Predictions

4.4.1 Validation of Discrimination by Using Two Different ATR-FTIR Spectrometers of the Same Model

ATR-FTIR spectroscopy has shown potential in discriminating between microorganisms based on differences in their infrared spectra. Potentially, public health laboratories will want to use this technology to rapidly identify bacteria. To use this method for identifying bacteria, databases would be developed and would be continuously updated as more laboratories contribute spectra to the database. In such cases, instrument-to-instrument validation is necessary to ensure there is little to no variation between instruments. In the present study, two ATR-FTIR spectrometers of the same model (ML 5500 ATR FTIR spectrometer, Agilent Technologies, Santa Clara, California) were used side-by-side to determine whether or not they would provide the same results. For this instrument-to-instrument validation study, 20 strains of *S. sonnei* and 20 strains of *S. flexneri* were used.

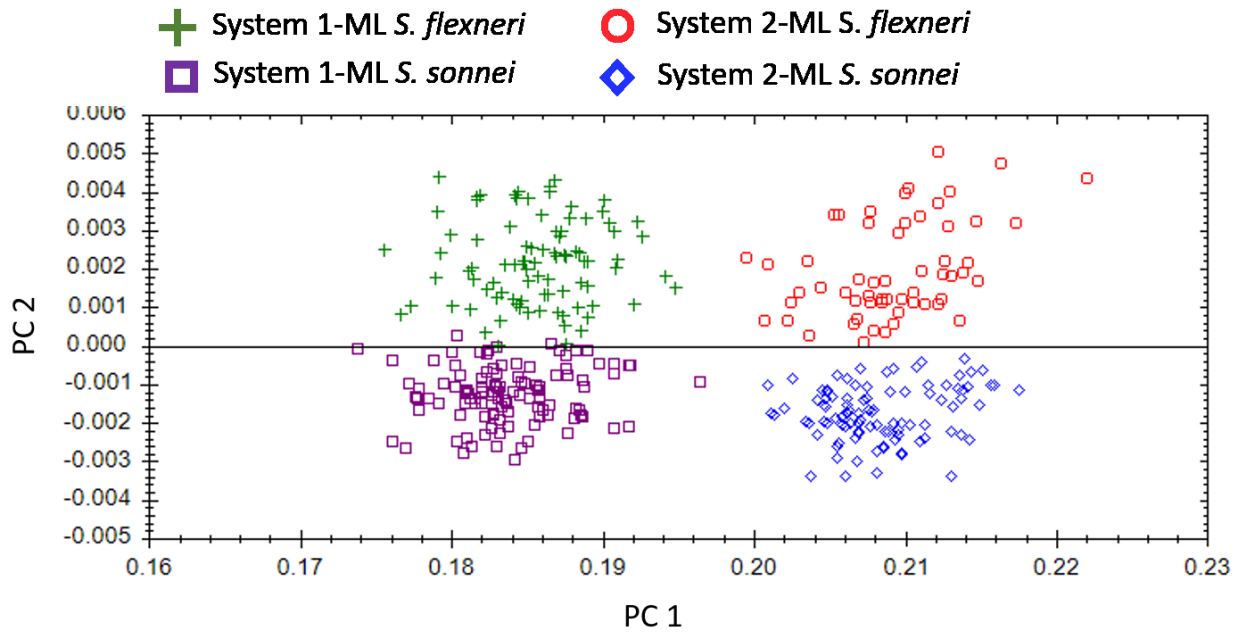


Figure 4.28. Scores plot for principal components 1 and 2 obtained by principal component analysis of ATR-FTIR spectra of two *Shigella* species acquired on two spectrometers of the same model.

Figure 4.28 shows that good discrimination between the two *Shigella* species based on the ATR-FTIR spectra acquired on each instrument was achieved. However, the separation between the clusters representing the spectra acquired on the different instruments demonstrates that the instrument-to-instrument variability would preclude transferring of the databases created on one instrument to the other instrument. As a consequence, a new database would need to be created for each spectrometer, which would seriously limit the practical utility of the technique. Accordingly, advancements in instrumental design and manufacture to produce identical ATR-FTIR spectra of the same strains on multiple systems are required.

4.4.2 Validation Using Database to Predict Unknown Samples

The instrument-to-instrument validation study described above showed that databases created on one instrument could not be transferred to the other for use in the prediction of unknowns because the spectra acquired on the two instruments were different. Accordingly, in the validation studies presented below, the spectra of the unknowns were acquired by using the same spectrometer that the database employed for prediction was generated on.

4.4.2.1 Validation for the Identification of *E. coli* and *Shigella* Species

For the first validation study, 5 *E. coli* and *5 Shigella* strains that were not included in development of the spectral database were treated as unknowns. Within the database used for the identification of the 10 unknown samples, there are over 300 spectra of each genus. In Table 8, there is a summary of the samples used in addition to the predictions based on the closest spectral match in the database Only the spectral region 1478-1411 cm^{-1} selected for discrimination between *E. coli* and *Shigella* species (Section 4.2.1) was employed in matching the spectra of unknowns to those in the database and un calculating the percent similarity to the closest spectral match.

Table 8 - Predictions for *E. coli* and *Shigella* Validation Samples and Percent Similarities to Closest Spectral Match in the Database*

Identification label	Known identification	ATR-FTIR prediction	% Similarity
Unknown 1	*E. coli* x1	*E. coli*	96.66%
Unknown 2	*E. coli* O157:H7 y1	*E. coli*	97.73%
Unknown 3	*E. coli* z1	*E. coli*	95.85%
Unknown 4	*E. coli* AA1	*E. coli*	96.96%
Unknown 5	*E. coli* BB1	*E. coli*	97.42%
Unknown 6	*Shigella flexneri* CC1	*Shigella*	97.06%
Unknown 7	*Shigella sonnei* DD1	*Shigella*	97.45%
Unknown 8	*Shigella sonnei* EE1	*Shigella sonnei*	97.29%
Unknown 9	*Shigella flexneri* FF1	*Shigella flexneri*	96.92%
Unknown 10	*Shigella sonnei* GG1	*Shigella*	96.52%

*Spectral region employed for prediction and calculation of % similarity: 1478-1411 cm^{-1}.

As shown in Table 8, all unknown samples of *E. coli* and *Shigella* were correctly identified with at least 95.85% similarity to their spectral match in the database and at most 97.73%, giving an average of 96.99%. Moreover, unknown samples 8 and 9 were correctly identified to the species level.

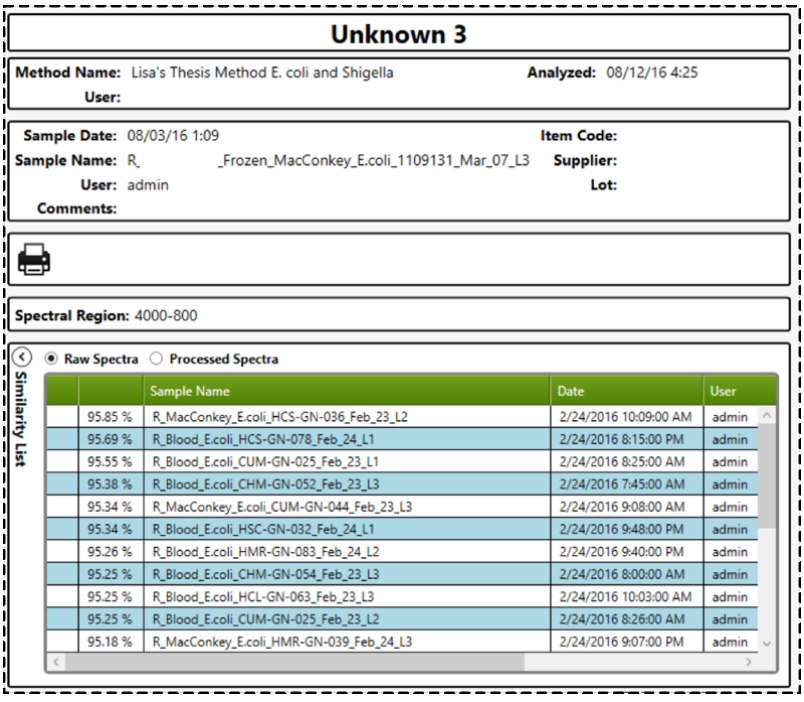


Figure 4.29. Computer screenshot of results for Unknown 3 showing percent similarities to 11 closest spectral matches in the database.

As an illustration, the results for identification of Unknown 3 are presented in Figure 4.29. The top 10 matches to the unknown samples were all *E. coli*. The percent similarity to the closest spectral match (95.85%) is lower than the values for the rest of the unknowns in Table 8 but this unknown is still correctly identified. Given more samples in the database representing the diversity of *E. coli*; the percent similarities may be expected to increase.

On another note, the database used for the validation contained the spectra of Gram negative bacteria grown on both blood agar and MacConkey agar. For all 10 unknown samples, the closest matches were grown on the same growth media as the unknown. This result also validates that the growth media is a factor in ATR-FTIR spectroscopic identification of microorganisms.

4.4.2.2 Validation for the Identification of *S. sonnei* and *S. flexneri* Species

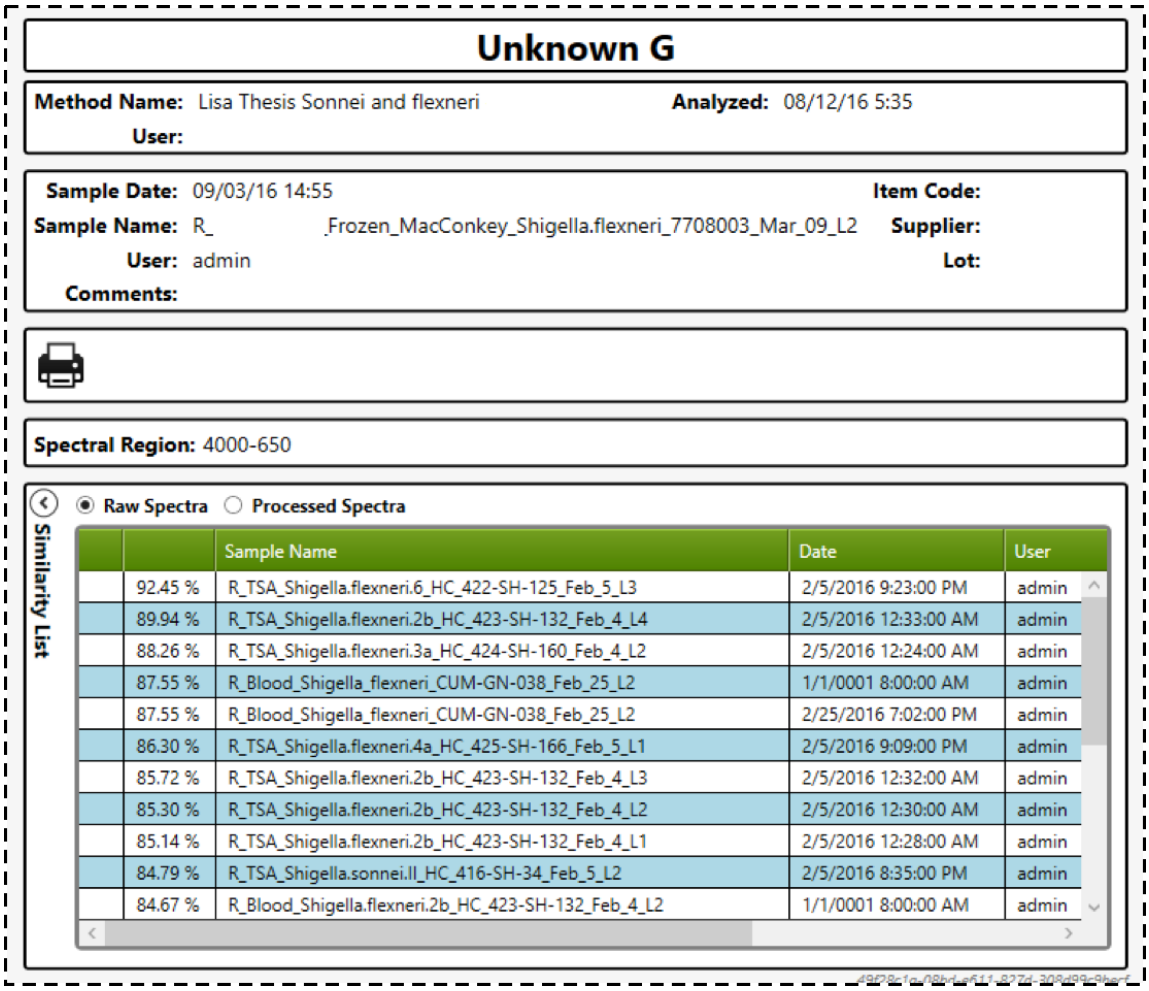
Aside from being able to identify the genus of a microorganism, the validation of species-level identification is of interest. The validation study for the identification of species was performed with 5 *Shigella sonnei* and 5 *Shigella flexneri* strains that had not been used in the development of the database. For the prediction of the unknowns, the spectral regions 1136-1113 and 1218-1207 cm^{-1} selected for the discrimination between *S. sonnei* and *S. flexneri* (Section 4.2.2) were employed.

Table 9 - Species Predictions for *Shigella* Validation Samples and Percent Similarities to Closest Spectral Match in the Database*

Identification label	Known identification	ATR-FTIR species prediction	% Similarity
Unknown A	*Shigella sonnei* xxx4068	*Shigella sonnei*	94.52%
Unknown B	*Shigella sonnei* xxx1060	*Shigella sonnei*	95.14%
Unknown C	*Shigella sonnei* xxx2003	*Shigella sonnei*	94.56%
Unknown D	*Shigella sonnei* xxx7097	*Shigella sonnei*	96.06%
Unknown E	*Shigella sonnei* xxx0003	*Shigella sonnei*	96.03%
Unknown F	*Shigella flexneri* xxx3211	*Shigella flexneri*	96.38%
Unknown G	*Shigella flexneri* xxx8003	*Shigella flexneri*	92.45%
Unknown H	*Shigella flexneri* xxx2076	*Shigella flexneri*	94.11%
Unknown I	*Shigella flexneri* xxx9020	*Shigella flexneri*	94.87%
Unknown J	*Shigella flexneri* xxx6025	*Shigella flexneri*	94.64%

*Spectral regions employed for prediction and calculation of % similarity: 1136-1113 and 1218-1207 cm^{-1}.

All 10 unknowns were correctly identified at the species level, with at least 92.45% similarity to the closest match in the spectral database and at most 96.38%, averaging to 94.88% similarity. Comparing to the identification of genera, there is a lower average percent similarity.



Unknown G

Method Name: Lisa Thesis Sonnei and flexneri **Analyzed:** 08/12/16 5:35
User:

Sample Date: 09/03/16 14:55 **Item Code:**
Sample Name: R_ _Frozen_MacConkey_Shigella.flexneri_7708003_Mar_09_L2 **Supplier:**
User: admin **Lot:**
Comments:

Spectral Region: 4000-650

Raw Spectra Processed Spectra

Similarity List

	Sample Name	Date	User
92.45 %	R_TSA_Shigella.flexneri.6_HC_422-SH-125_Feb_5_L3	2/5/2016 9:23:00 PM	admin
89.94 %	R_TSA_Shigella.flexneri.2b_HC_423-SH-132_Feb_4_L4	2/5/2016 12:33:00 AM	admin
88.26 %	R_TSA_Shigella.flexneri.3a_HC_424-SH-160_Feb_4_L2	2/5/2016 12:24:00 AM	admin
87.55 %	R_Blood_Shigella_flexneri_CUM-GN-038_Feb_25_L2	1/1/0001 8:00:00 AM	admin
87.55 %	R_Blood_Shigella_flexneri_CUM-GN-038_Feb_25_L2	2/25/2016 7:02:00 PM	admin
86.30 %	R_TSA_Shigella.flexneri.4a_HC_425-SH-166_Feb_5_L1	2/5/2016 9:09:00 PM	admin
85.72 %	R_TSA_Shigella.flexneri.2b_HC_423-SH-132_Feb_4_L3	2/5/2016 12:32:00 AM	admin
85.30 %	R_TSA_Shigella.flexneri.2b_HC_423-SH-132_Feb_4_L2	2/5/2016 12:30:00 AM	admin
85.14 %	R_TSA_Shigella.flexneri.2b_HC_423-SH-132_Feb_4_L1	2/5/2016 12:28:00 AM	admin
84.79 %	R_TSA_Shigella.sonnei.II_HC_416-SH-34_Feb_5_L2	2/5/2016 8:35:00 PM	admin
84.67 %	R_Blood_Shigella.flexneri.2b_HC_423-SH-132_Feb_4_L2	1/1/0001 8:00:00 AM	admin

Figure 4.30. Computer screenshot of results for Unknown G showing percent similarities to 11 closest spectral matches in the database.

As an illustration, the results from prediction of Unknown G are presented in Figure 4.30. The closest match to the unknown sample is identified as *Shigella flexneri* at 92.45% similarity and the next 10 closest matches are identified as *Shigella flexneri* too. Again, the percent similarities are based solely on the spectral regions employed for the predictions. They indicate that there are differences in these regions between the unknown sample spectrum and the spectra of the closest

matches in the database but the unknown sample was still correctly identified. The overall results of this validation study indicate the potential of ATR-FTIR spectroscopy for bacteria identification at the species level.

Chapter 5 – SUMMARY AND CONCLUSIONS

Throughout the course of this research, ATR-FTIR spectroscopy has been demonstrated to be a very powerful technique for the discrimination of intact bacteria in a matter of minutes without the use of any consumables past the culturing step. The information content embedded in the ATR-FTIR spectra of the bacteria has been successfully mined with the use of multivariate techniques such as principal component analysis and hierarchical cluster analysis. Using such techniques, the differentiation between Gram-positive and Gram-negative bacteria, between closely related genera, and between two species of the same genus was conducted successfully. In addition, pathotype differentiation and toxin typing have also been demonstrated. The results obtained for each of these levels of discrimination are summarized below.

Gram-positive – Gram-negative bacteria

Examination of a large number of spectra of Gram-positive and Gram-negative bacteria revealed significant differences in the spectral region between 1000 and 980 cm^{-1}. The differences are the outcome of the difference in polysaccharide composition of the two groups. This work confirmed that the spectral differences observed in the infrared spectra can be effectively exploited to differentiate between Gram-positive and Gram-negative bacteria by ATR-FTIR spectroscopy. Indeed, this spectroscopic approach is 3 times faster than the time required to carry out a Gram stain (~10-15 minutes). Furthermore, the ATR-FTIR-based method does not require any reagents and can analyze the bacteria from a single colony removed from an agar plate or an isolate from a positive blood culture.

E. coli and *Shigella*

Through multivariate analysis, the separation between *E. coli* and *Shigella* was achieved solely based on differences in the 1478-1411 cm^{-1} region of their ATR-FTIR spectra. The absorption bands in this region may be attributed to lipids and carboxyl side chains of amino acids. Additionally, spectral differences were also observed between 1070 and 1040 cm^{-1}, which may be

attributed to differences in carbohydrate composition. In a validation study conducted with 5 *E. coli* and 5 *Shigella* strains, correct predictions were made for all strains.

Shigella **species separation**

ATR-FTIR spectroscopy was successfully employed for the discrimination between *S. sonnei* and *S. flexneri* by applying PCA to the entire mid-IR region. Better separation between the two species was achieved by utilizing selected regions of 1136-1113 and 1218-1207 cm^{-1}. The spectral differences in these regions may be attributed to differences in DNA/RNA and carbohydrate composition. Moreover, by analyzing the average spectra of each species, additional differences were ascertained in the 1173-953, 1657-1623, and 1575-1508 cm^{-1} regions. In a subsequent validation study conducted with 5 *S. sonnei* and 5 *S. flexneri* strains, correct predictions were made for all strains. Accordingly, this work demonstrates the potential of ATR-FTIR spectroscopy for the discrimination of two species from the same genus.

E. coli **O157:H7 and non-O157:H7**

E. coli O157:H7 and non-O157:H7 *E. coli* pathotypes were separated by PCA based on their spectral differences in the regions of 1248-1121 and 1356-1344 cm^{-1}. Additional spectral differences were also observed in the regions of 1247-1213 cm^{-1} and 1140-950 cm^{-1} by visual inspection of the average spectra of the two groups. The cause for separation between the two groups may be attributed primarily to differences in nucleic acid and polysaccharide regions.

Toxin producers STEC and ***Shigella***

Differentiation between Shiga-toxin-producing *E. coli* (STEC) and toxin-producing *Shigella* based on differences in their respective ATR-FTIR spectra was demonstrated. The key spectral regions responsible for the separation were 1288-1270, 1367-1348, and 1426-1408 cm^{-1}. The absorption bands in these regions may stem from proteins (amide III band and absorption bands of amino acid side-chain carboxylate groups). Upon visual inspection of the overlaid average spectra of the two groups, significant differences in relative intensities were observed in

the 1061-975 cm^{-1} region, in which bands due to C-O-C and C-O-H bending vibrations of sugar moieties are observed.

Typing of Shiga-toxin-producing *E. coli*

Application of ATR-FTIR spectroscopy for typing of Shiga-toxin-producing *E. coli* proved to be promising but inconclusive at the present time, possibly due to the limited number of strains that were available for this study. *E. coli* strains expressing both *stx1* and *stx2* genes and those producing only the *stx2* gene could be separated from each other based on differences in the spectral regions 1140-1121, 1225-1214, and 1329-1318 cm^{-1}. The separation between the two sets may be based on differences in the DNA/RNA composition. *E. coli* strains possessing only the *stx1* gene could not be separated from those possessing only the *stx2* gene, possibly due to the greater intra strain variability of the *E. coli* possessing the *stx1* gene or possibly as a result of lack of toxin expression by the *stx1*-producing strains. Further investigation is required to investigate this. For example, quantitative determination of Stx concentration for each stain could be carried out by immunoassay or liquid chromatography – mass spectrometry (LC-MS). Genetic diversity of the strains could be ascertained through whole genome sequencing (WGS) of Stx-producing strains.

In conclusion, ATR-FTIR spectroscopy has been applied effectively for the differentiation between Gram-positive and Gram-negative bacteria, between closely related genera, between species of the same genus, and between pathotypes. The method employs a very simple analysis protocol that requires no reagents and little to no sample handling, yielding results within minutes. The capability to identify unknowns by this method was shown in two validation studies, which gave at least 95% similarities between the spectra of the unknowns and their closest matches in the spectral database. All unknown samples were correctly identified and down to the media used as well, showing the media effect on the spectra.

The results obtained by using ATR-FTIR spectroscopy were found to be consistent between the two instruments employed in this research. However, when spectra of the same sample were acquired on the two instruments, significant spectral differences were discernible. Accordingly,

one drawback which is currently being addressed by our research group is related to instrument-to-instrument cross validation. Although different ATR-FTIR instruments show the same performance, the database generated on one instrument is not transferable to a second instrument due to spectrometer manufacturer limitations. Discussion with one of the ATR-FTIR manufacturers is underway to overcome this hurdle. Once this issue is resolved, ATR-FTIR spectroscopy may find a role in microbial analysis, competing alongside MALDI-TOF MS-based techniques.

REFERENCES

Alvarez-Ordóñez, A., & Prieto, M. (2012). *Fourier transform infrared spectroscopy in food microbiology*: Springer.

Baker, D. R., Moxley, R. A., Steele, M. B., LeJeune, J. T., Christopher-Hennings, J., Chen, D.-G., . . . Francis, D. H. (2007). Differences in virulence among *Escherichia coli* O157: H7 strains isolated from humans during disease outbreaks and from healthy cattle. *Applied and environmental microbiology, 73*(22), 7338-7346.

Beekes, M., Lasch, P., & Naumann, D. (2007). Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. *Veterinary microbiology*, *123*(4), 305-319.

Black, R., Craun, G. F., & Blake, P. A. (1978). Epidemiology of common-source outbreaks of shigellosis in the United States, 1961–1975. *American journal of epidemiology, 108*(1), 47-52.

Bourassa, L., & Butler-Wu, S. M. (2015). MALDI-TOF mass spectrometry for microorganism identification. *Methods in microbiology, 42*, 37-85.

CDC (2016). *PFGE and MLVA Images | Pathogens and Protocols | PulseNet | CDC*. Retrieved 26 January 2016, from http://www.cdc.gov/pulsenet/pathogens/protocol-images.html#pfge

Clark, A. E., Kaleta, E. J., Arora, A., & Wolk, D. M. (2013). Matrix-assisted laser desorption ionization–time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clinical microbiology reviews, 26*(3), 547-603.

Cook, C. N. (2016). *Development of a Rapid Method for Detection and Differentiation of Escherichia coli Serotypes and Strains Using Attenuated Total Reflectance Fourier Transform Infrared Microspectroscopy.* Ph.D. thesis, The Ohio State University.

Dalmasso, M., & Jordan, K. (2015). PFGE as a tool to track Listeria monocytogenes in food processing facilities: Case studies. *Pulse field gel electrophoresis: Methods and protocols*, 29-34.

Davis, R., & Mauer, L. J. (2010). Fourier transform infrared (FT-IR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. *Current research, technology and education topics in applied microbiology and microbial biotechnology*, *2*, 1582-1594. Jelacic, J.

Dingle, T. C., & Butler-Wu, S. M. (2013). MALDI-TOF mass spectrometry for microorganism identification. *Clinics in laboratory medicine, 33*(3), 589-609.

Dole, M. N., Patel, P. A., Sawant, S. D., & Shedpure, P. S. (2011). Advance applications of Fourier transform infrared spectroscopy. *Int. J. Pharm. Sci. Rev. Res, 7*, 159-166.

Donnenberg, M. (2013). *Escherichia coli: pathotypes and principals of pathogenesis*: Academic Press.

Drake, R. R., Boggs, S. R., & Drake, S. K. (2011). Pathogen identification using mass spectrometry in the clinical microbiology laboratory. *Journal of mass spectrometry, 46*(12), 1223-1232.

Durmaz, R., Otlu, B., Koksal, F., Hosoglu, S., Ozturk, R., Ersoy, Y.,. ..& Caliskan, A. (2009). The optimization of a rapid pulsed-field gel electrophoresis protocol for the typing of *Acinetobacter baumannii*, *Escherichia coli* and *Klebsiella* spp. *Jpn J Infect Dis*, *62*(5), 372-7.

Friesema, I., van der Zwaluw, K., Schuurman, T., Kooistra-Smid, M., Franz, E., van Duynhoven, Y., & van Pelt, W. (2014). Emergence of *Escherichia coli* encoding Shiga toxin 2f in

human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. *Euro Surveill, 19*, 26-32.
Ghetler, A. (2010). Development of an expert system for the identification of bacteria by focal plane array Fourier transform infrared spectroscopy. Ph.D. thesis, McGill University.
Ghuysen, J. M., & Hakenbeck, R. (Eds.). (1994). *Bacterial cell wall* (Vol. 27). Elsevier.
Goebel, W. (2012). *Genetic approaches to microbial pathogenicity* (Vol. 118): Springer Science & Business Media.
Goldberg, M. B., & Theriot, J. A. (1995). *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proceedings of the National Academy of Sciences of the United States of America, 92*(14), 6572-6576.
Griffin, P. M., Price, G. R., Schooneveldt, J. M., Schlebusch, S., Tilse, M. H., Urbanski, T., ... & Venter, D. (2012). Use of matrix-assisted laser desorption ionization–time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *Journal of clinical microbiology*, *50*(9), 2918-2931.
Harisha, S. (2005). *An introduction to practical biotechnology*. Firewall Media.
Holmes, B., Willcox, W., & Lapage, S. (1978). Identification of Enterobacteriaceae by the API 20E system. *Journal of clinical pathology, 31*(1), 22-30.
Iizuka, H., & Hasegawa, T. (Eds.). (1970). *Culture collections of microorganisms: proceedings of the International Conference on Culture Collections, Tokyo, Oct. 7-11, 1968*. University Park Press.
Ito, H., Kido, N., Arakawa, Y., Ohta, M., Sugiyama, T., & Kato, N. (1991). Possible mechanisms underlying the slow lactose fermentation phenotype in *Shigella* spp. *Applied and environmental microbiology*, *57*(10), 2912-2917.
Jelacic, K., Damrow, T., Chen, G. S., Jelacic, S., Bielaszewska, M., Ciol, M., ... & Tarr, P. I. (2003). Shiga toxin–producing *Escherichia coli* in Montana: bacterial genotypes and clinical profiles. *Journal of infectious diseases*, *188*(5), 719-729.
Khot, P. D., & Fisher, M. A. (2013). Novel approach for differentiating *Shigella* species and *Escherichia coli* by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of clinical microbiology, 51*(11), 3711-3716.
Kirkwood, J. P. (2007). Identification of bacteria by infrared imaging with the use of focal plane array Fourier transform infrared spectroscopy. Ph.D. thesis, McGill University.
Knirel, Y. A., & Valvano, M. A. (2011). *Bacterial lipopolysaccharides: structure, chemical synthesis, biogenesis and interaction with host cells*: Springer Science & Business Media.
Kramer, A., Schwebke, I., & Kampf, G. (2006). How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases, 6*, 130-130. doi: 10.1186/1471-2334-6-130
Kuehl, C. J., Dragoi, A. M., & Agaisse, H. (2014). The *Shigella flexneri* type 3 secretion system is required for tyrosine kinase-dependent protrusion resolution, and vacuole escape during bacterial dissemination. *PloS one*, *9*(11), e112738.
Lan, R., & Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes and infection, 4*(11), 1125-1132.
Leclair, D., Pagotto, F., Farber, J. M., Cadieux, B., & Austin, J. W. (2006). Comparison of DNA fingerprinting methods for use in investigation of type E botulism outbreaks in the Canadian Arctic. *Journal of clinical microbiology*, *44*(5), 1635-1644.
Lévesque, S., Dufresne, P. J., Soualhine, H., Domingo, M.-C., Bekal, S., Lefebvre, B., & Tremblay, C. (2015). A side by side comparison of Bruker Biotyper and VITEK MS: utility

of MALDI-TOF MS technology for microorganism identification in a public health reference laboratory. *PloS one, 10*(12), e0144878.
Lim JY, Yoon J, Hovde CJ. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. J Microbiol Biotechnol. 2010;20:5–14.
Lin, C.-N. (2013). Zoonosis and Human Health: Review. Toxinology: Biological Toxins and Bioterrorism. P. Gopalakrishnakone. Dordrecht, Springer Netherlands**:** 1-16.
Lindberg, A. A., Karnell, A., & Weintraub, A. (1991). The lipopolysaccharide of *Shigella* bacteria as a virulence factor. *Review of infectious diseases, 13*(Supplement 4), S279-S284.
Maity, J. P., Kar, S., Lin, C.-M., Chen, C.-Y., Chang, Y.-F., Jean, J.-S., & Kulp, T. R. (2013). Identification and discrimination of bacteria using Fourier transform infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 116*, 478-484.
Majcherczyk, P. A., McKenna, T., Moreillon, P., & Vaudaux, P. (2006). The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus* aureus. *FEMS microbiology letters, 255*(2), 233-239.
Mantis, N. (2012). *Ricin and Shiga toxins: Pathogenesis, immunity, vaccines and therapeutics* (Vol. 357): Springer Science & Business Media.
Mauer, L. J., & Reuhs, B. L. (2008). Mid-infrared sensors for the rapid analysis of select microbial food borne pathogens. *Wiley handbook of science and technology for homeland security*.
Mello, M. L. S., & Vidal, B. (2012). Changes in the infrared microspectroscopic characteristics of DNA caused by cationic elements, different base richness and single-stranded form. *PloS one, 7*(8), e43169.
Metzker, M. L., & Caskey, C. T. (2009). Polymerase chain reaction (PCR). *eLS*.
Murray, P. R., Rosenthal, K. S., & Pfaller, M. A. (2015). *Medical microbiology*: Elsevier Health Sciences.
Mullis, K. B., Erlich, H. A., Arnheim, N., Horn, G. T., Saiki, R. K., & Scharf, S. J. (1987). Process for amplifying, detecting, and/or-cloning nucleic acid sequences: Google Patents.
Noori, T., & Alwan, M. (2016). Isolation and identification of zoonotic bacteria from poultry meat. *Int. J. Adv. Res. Biol. Sci, 3*(8), 57-66.
Oporto, B., Esteban, J., Aduriz, G., Juste, R., & Hurtado, A. (2008). *Escherichia coli* O157: H7 and non-O157 Shiga toxin-producing *E. coli* in healthy cattle, sheep and swine herds in northern Spain. *Zoonoses and public health, 55*(2), 73-81.
Paauw, A., Jonker, D., Roeselers, G., Heng, J. M., Mars-Groenendijk, R. H., Trip, H., . . . de Jong, A. L. (2015). Rapid and reliable discrimination between *Shigella* species and *Escherichia coli* using MALDI-TOF mass spectrometry. *International Journal of Medical Microbiology*.
Petibois, C., & Desbat, B. (2010). Clinical application of FTIR imaging: new reasons for hope. *Trends in biotechnology, 28*(10), 495-500.
PHAC,. (2013). *The Chief Public Health Officer's Report on the State of Public Health in Canada, 2013 Infectious Disease—The Never-ending Threat* Retrieved 26 January 2016, from http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2013/resistance-eng.php
Primrose, S. B., & Twyman, R. (2013). *Principles of gene manipulation and genomics*: John Wiley & Sons.
Rangel, J. M., Sparling, P. H., Crowe, C., Griffin, P. M., & Swerdlow, D. L. (2005). Epidemiology of *Escherichia coli* O157: H7 outbreaks, United States, 1982–2002.

Ribot, E. M., Fair, M., Gautom, R., Cameron, D., Hunter, S., Swaminathan, B., & Barrett, T. J. (2006). Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157: H7, Salmonella, and *Shigella* for PulseNet. *Foodbourne pathogens & disease, 3*(1), 59-67.

Robbins, J. R., Monack, D., McCallum, S. J., Vegas, A., Pham, E., Goldberg, M. B., & Theriot, J. A. (2001). The making of a gradient: IcsA (VirG) polarity in *Shigella flexneri*. *Molecular microbiology*, *41*(4), 861-872.

Romeo, M. J., Wood, B. R., Quinn, M. A., & McNaughton, D. (2003). Removal of blood components from cervical smears: implications for cancer diagnosis using FTIR spectroscopy. *Biopolymers*, *72*(1), 69-76.

Santos, T., Capelo, J. L., Santos, H. M., Oliveira, I., Marinho, C., Gonçalves, A., . . . Igrejas, G. (2015). Use of MALDI-TOF mass spectrometry fingerprinting to characterize *Enterococcus* spp. and *Escherichia coli* isolates. *Journal of proteomics*.

Seltmann, G., & Holst, O. (2013). *The bacterial cell wall*: Springer Science & Business Media.

Shimizu, T., Ohta, Y., & Noda, M. (2009). Shiga toxin 2 is specifically released from bacterial cells by two different mechanisms. *Infection and immunity, 77*(7), 2813-2823.

Sethabutr, O., Venkatesan, M., Murphy, G. S., Eampokalap, B., Hoge, C. W., & Echeverria, P. (1993). Detection of *Shigella*e and enteroinvasive *Escherichia coli* by amplification of the invasion plasmid antigen H DNA sequence in patients with dysentery. *Journal of infectious diseases, 167*(2), 458-461.

Standish, A. J., Teh, M. Y., Tran, E. N. H., Doyle, M. T., Baker, P. J., & Morona, R. (2016). Unprecedented abundance of protein tyrosine phosphorylation modulates *Shigella flexneri* virulence. *Journal of molecular biology, 428*(20), 4197-4208.

Thorpe, C. M. (2004). Shiga toxin—producing *Escherichia coli* infection. *Clinical infectious diseases*, *38*(9), 1298-1303.

Van den Beld, M., & Reubsaet, F. (2012). Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*. *European journal of clinical microbiology & infectious diseases, 31*(6), 899-904.

Wenning, M., Breitenwieser, F., Konrad, R., Huber, I., Busch, U., & Scherer, S. (2014). Identification and differentiation of food-related bacteria: A comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry. *Journal of microbiological methods, 103*, 44-52.

Woods, T. A., Mendez, H. M., Ortega, S., Shi, X., Marx, D., Bai, J., ... & Deshpande, A. (2016). Development of 11-plex MOL-PCR assay for the rapid screening of samples for Shiga toxin-producing *Escherichia coli*. *Frontiers in cellular and infection microbiology*, *6*.

Yang, L. J., Wang, J., Li, Z. J., Song, X. H., Liu, Y. M., & Hu, Q. R. (2014). *Rapid Differentiation and Identification of Shigella sonnei and Escherichia coli O157: H7 by Fourier Transform Infrared Spectroscopy and Multivariate Statistical Analysis.* Paper presented at the Advanced Materials Research.

Zarnowiec, P., Lechowicz, L., Czerwonka, G., & Kaca, W. (2015). Fourier transform infrared spectroscopy (ftir) as a tool for the identification and differentiation of pathogenic bacteria. *Current medicinal chemistry, 22*(14), 1710.

Zhang, J., Wang, F., Jin, H., Hu, J., Yuan, Z., Shi, W., . . . Xu, X. (2015). Laboratory monitoring of bacterial gastroenteric pathogens Salmonella and *Shigella* in Shanghai, China 2006–2012. *Epidemiology and infection, 143*(03), 478-485.

Zuo, G., Xu, Z., & Hao, B. (2013). *Shigella* strains are not clones of *Escherichia coli* but sister species in the genus *Escherichia*. *Genomics, proteomics & bioinformatics, 11*(1), 61-65.

APPENDIX

Table 1a. *Escherichia coli* Isolates Obtained from Health Canada and Grown on MacConkey Agar for 24 Hours: Observations

SAMPLE	SAMPLE ID	PATHOTYPE	SEROTYPE	LACTOSE	COMMENTS	SIZE	MORPHOLOGY	MOBILITY
1	x15	UPEC	O4:H27	POS		5mm	matte	yes
2	x96	UPEC	O4:K12:H51	POS		3mm		semi-motile
3	xxx592	STEC	O111:NM	POS		3-5mm	flat matte	yes
4	xxx808	UPEC	O2:H6	POS			flatish	yes
5	xxx876	UPEC	O18:H5	POS		5mm	flatish	yes
6	xxx348	EHEC	O157:H7	POS		3mm	non-round	semi-motile
7	xxx884	EHEC	O157:H7	POS		3mm	non-round	semi-motile
8	xxx156	EHEC	O157:H7	POS		3mm	non-round	semi-motile
9	xxx905	EHEC	O157:H7	POS		5mm		semi-motile
10	xxx956	EPEC	O128:H2	POS		3-4mm	non-round	non-motile
11	xxx952	UPEC	O2:H31	POS		3-5mm		yes
12	xxx858	UPEC	O22:H5	POS/NEG	Gradiant color change	3-5mm	non-round	
13	x19	STEC	O26:H11	POS		2mm		semi-motile
14	xxx130	STEC	O26:H21	POS		2mm	semi-round	non-motile
15	xxx63B	EPEC	O26:H11	POS		3mm	non-round	non-motile
16	xxx757	STEC	O46:H38	POS		3mm	semi-flat	non-motile
17	xxx596	STEC	O69:H11	POS		3mm	semi-flat	non-motile
18	xxx16	STEC	O113:H21	POS/NEG	Gradiant color change	3mm	semi-flat	non-motile
19	xxx68	STEC	O123:H2	POS		2-3mm	non-round	non-motile
20	xxx051	EPEC	O128:H2	POS		2-3mm	non-round	non-motile
21	xxx56A	EPEC	O128:H35	POS	matte film	3mm	flatish	
22	xxx832	STEC	O121:H19	POS	matte film	3mm	flatish	non-motile
23	xxx054	EHEC	O157:H7	POS	matte film	3-4mm	roundish	non-motile
24	xxx554	EHEC	O157:H7	POS		5mm	non-round	non-motile
25	xxx434	EHEC	O157:H7	POS	matte film	3mm		non-motile
26	xxx702	EHEC	O157:H7	POS		3mm	roundish	non-motile
27	xxx773	EHEC	O157:H7	POS		2-3mm	flatish	non-motile
28	xxx767	EHEC	O157:H7	POS		2-3mm	flatish	non-motile
29	xxx489	EHEC	O157:H7	POS		5mm	flatish	non-motile
30	xxx488	EHEC	O157:H7	POS		3-4mm	flatish	non-motile
31	xxx111	STEC	O111:H11	POS	matte film	2mm	flat	non-motile
32	xxx877	EPEC	O119:K69:NM	NEG	Slight positive	2mm	round	non-motile
33	xxx581	STEC	O26:H11	POS		3mm	roundish	non-motile
34	xxx01A	EPEC	O?:H31	POS	shiney film	2-3mm	convexish	non-motile
35	xxx953	EPEC	OUT:H1	POS	SMALL NEG/ mix?	5mm	non-round	non-motile
36	xxx767	UPEC	O6:H19	POS	looks mixed	3mm	flatish	non-motile
37	xxx066	STEC	O121:H10	POS	matte film	2mm	semi-round	
38	xxx595	STEC	O103:H21	POS		3mm	non-round	non-motile
39	xxx925	STEC	O121:H19	NEG		5mm	non-flat	motile
40	xxx933	EHEC	O157:H7	POS		3-5mm	non-flat	non-motile
41	xx69	UPEC	O2:H4	POS/NEG	Mostly neg	5mm	flat	motile
42	xxx762	EPEC	O114:H19	POS	has a film	2-3mm		non-motile
43	xxF1	STEC	O91:H21	POS		2-3mm	non-flat	non-motile
44	xxx985	EPEC	O128:H2	POS		3mm	non-round	non-motile
45	xxx037	UPEC	O6:H31	POS	On spec, did not label pathotype	3mm	non-flat	non-motile
46	xxx464	EHEC	O157:H7	POS		3mm	non-round	non-motile
47	xxx511	EHEC	O157:H-(unk)	POS/NEG	Small neg/ mix?	5mm	flatish	motile
48	xxx439	EHEC	O157:H7	POS		3-5mm	semi-flat	non-motile
49	xxx329	EHEC	O157:H7	POS		3mm	non-flat	non-motile

50	xxx192	EPEC	O128:H2	POS		2-3mm	roundish	
51	xxx402	UPEC	O4:H5	POS		3-5mm	non-flat	non-motile
52	xx2-1	UPEC	O2:H1	POS		3mm	flat	non-motile
53	xxx399	UPEC	O25:H2	NEG		3-5mm	flat	motile
54	xxx070	EPEC	O128:H2	POS		3mm	convexish	non-motile
55	xx2	STEC	O139:K82	POS		3-5mm	flat	non-motile
56	xxx449	STEC	O145:NM	POS		3mm	convexish	non-motile
57	xxx376	STEC	O55:H7	POS		5mm	flatish	non-motile
58	xxx335	EHEC	O157:H7	POS		3mm	convexish	non-motile
59	xxx060	EHEC	O157:H7	POS		3mm	convexish	non-motile
60	xxx986	EPEC	O128:H2	POS		1-3mm	convex	non-motile
61	xxxkai	EHEC	O157:H7	POS		2-3mm	convexish	non-motile
62	xxx991	EPEC	O55:H19	POS		2-3mm		non-motile
63	xxx89A	EPEC	O18:NM	POS	bumpy	2-3mm		
64	xxI.8	STEC	O128:NM	POS		3mm	flat	non-motile
65	xxx178	UPEC	O2:H4	POS		3-5mm	flatish	non-motile
66	xxxx074	STEC	O91:H28	POS		3mm	flatish	non-motile
67	xxx772	EHEC	O157:H7	POS		1-2mm	flat	non-motile
68	xxx295	EHEC	O157:H7	POS		3mm	jagged	non-motile
69	xxx277	UPEC	O2:H7	POS		2-3mm	flat	non-motile
70	xxx194	EPEC	O55:H7	POS		3mm	convexish	non-motile
71	xxx979	EPEC	O86:H8	POS	mix?	2-3mm	non-round	
72	xxx955	EPEC	O128:H2	POS/NEG	Mix?		roundish	
73	xxx832	STEC	O121:H19	POS	shiney film	2-3mm	convexish	
74	xxx991	STEC	O111:NM	POS		3mm	convexish	
75	xxx380	EHEC	O157:H7	POS		2-3mm	non-round	
76	xxx992	EHEC	O157:H7	POS		3mm	non-flat	non-motile
77	xxx806	STEC	O103:H2	POS		3mm	flatish	non-motile
78	xxx419	UPEC	O157:H45	POS		3-4mm	convexish	non-motile
79	xxx073	UPEC	O6:K2:H1	POS		2-3mm	convexish	
80	xxx489	STEC	O91:H2	POS		3mm	roundish	non-motile
81	xxx417	UPEC	O166:H15	POS	shiney film	2-3mm	oval	
82	xxx509	EHEC	O157:H7	POS		3-4mm	convexish	non-motile
83	xxx737	UPEC	O6:H31	POS		3-5mm	roundish	non-motile
84	xxx92			POS	shiney film	2-3mm	convexish	
85	xxx741	UPEC	O8:H15	POS		3-5mm	roundish	non-motile
86	xx6	UPEC	O6:H31	POS		2-3mm	non-round	non-motile
87	xxx424	UPEC	O75:H5	POS		3-5mm		non-motile
88	xxx858	UPEC	O22:H5	POS	SmaLL NEG/ mix?	3-5mm	flat	
89	Exxx8/69	EPEC	O127:H6	POS		pinpoint	convex	non-motile
90	xxx988	UPEC	O6:H31	POS	matte film	3mm		non-motile
91	xxx394	STEC	O104:H21	POS		2-3mm		non-motile
92	xxx882	STEC	O103:H2	POS		3mm	convexish	non-motile
93	xxx860	UPEC	O6:H19	NEG			flatish	semi-motile
94	xxx57A	UPEC	O?:H31	POS		3mm	flatish	non-motile
95	xxx806	EPEC	O132:H34	POS		2-3mm		non-motile
96	xxx875	UPEC	O6:H31	POS		3-5mm		non-motile
97	xxx740	EPEC	O15:H18	POS	Small NEG/mix?, one colony is mucoid		flat	motile
98	xxx069	EPEC	O26:NM	POS			convexish	non-motile
99	xxx999	EPEC	O128:H2	POS		2-3mm	convexish	non-motile
100	xxx716	UPEC	O75:H9	NEG	semi-mucoid	3-5mm		
101	xxx189	UPEC	O6:H31	POS		2-3mm	convexish	non-motile

Table 2a - *Escherichia coli* Isolates Obtained from McGill University Health Center and Grown on Blood Agar for 24 Hours: Observations

SAMPLE	SAMPLE ID	HEMOLYTIC	COMMENTS	SIZE	MORPHOLOGY	MOBILITY
1	CHM-GN-050	Alpha				
2	CHM-GN-051	Beta				
3	CHM-GN-052	Alpha				
4	CHM-GN-053	Alpha				
5	CHM-GN-054	Alpha				
6	CHM-GN-055	Alpha				
7	CUM-GN-024	Alpha				Motile
8	CUM-GN-025	Non				
9	CUM-GN-026	Strong Beta				
10	CUM-GN-027	Beta				
11	CUM-GN-028	Alpha	Mucoid		concave	
12	CUM-GN-029		Non-pure	4 colonies (2 mucoid 3 and 2mm, pinpoint 2mm, flat)	can be polymorphism	
13	CUM-GN-030		Not an *E. coli*, possibly Cytrobactor	2 colonies		
14	CUM-GN-044	Alpha				
15	CUM-GN-045					
16	CUM-GN-046		Non-pure			
17	HCL-GN-038	Alpha				
18	HCL-GN-049	Alpha				
19	HCL-GN-054	Alpha				
20	HCL-GN-061	Alpha				
21	HCL-GN-063	Beta				
22	HCL-GN-070	Alpha				
23	HCS-GN-036					
24	HCS-GN-040	Non				
25	HCS-GN-048	Alpha				
26	HCS-GN-049					
27	HCS-GN-053	Beta				
28	HCS-GN-054	Non			Different	Slight mobility
29	HCS-GN-055	Alpha				
30	HCS-GN-056	Alpha				
31	HCS-GN-078	Alpha				Hyper motile
32	HCS-GN-080	Beta				
33	HGJ-GN-039	Alpha				
34	HGJ-GN-040	Beta				
35	HGJ-GN-041	Alpha				Hyper motile
36	HGJ-GN-068	Non				
37	HMR-GN-036		Mix? 2 colonies	one small and one midsize		
38	HMR-GN-039	Beta				
39	HMR-GN-048	Alpha				
40	HMR-GN-053		Pure BUT has 1 othe tiny colony			
41	HMR-GN-054	Alpha			not-typical morphology	motile
42	HMR-GN-058	Alpha				
43	HMR-GN-078					
44	HMR-GN-083	Beta				
45	HSC-GN-032	Beta	Looks contaminated			

46	HSC-GN-034	Alpha				
47	HSC-GN-035	Alpha	Non-pure	2 colonies (1 is big fatty and convex, other is 1mm flat)		
48	HSC-GN-050	Beta		3mm	flat	Not very motile
49	HSC-GN-057		Non-pure	3 colonies (small, convex/fatty, doughnut)		
50	HSC-GN-060	Alpha	mucoid	3mm		
51	HSC-GN-082	Alpha		3mm	convex	
52	LSP-GN-018	Alpha		3mm	convex	
53	LSP-GN-029	Beta				

Table 3a – *Escherichia coli* Isolates Obtained from McGill University Research Center and Grown on MacConkey Agar for 24 Hours: Observations

SAMPLES	SAMPLE ID	LACTOSE	COMMENTS	SIZE	MORPHOLOGY	MOBILITY
1	CHM-GN-050	POS				
2	CHM-GN-051	POS				
3	CHM-GN-052	POS				
4	CHM-GN-053	POS				
5	CHM-GN-054	POS				
6	CHM-GN-055	POS				
7	CUM-GN-024	NEG				
8	CUM-GN-025	POS				
9	CUM-GN-026	POS				
10	CUM-GN-027	POS				
11	CUM-GN-028	POS				
12	CUM-GN-029	POS				
13	CUM-GN-030	POS				
14	CUM-GN-044	NEG				
15	CUM-GN-045	NEG	Film former			
16	CUM-GN-046	POS				
17	HCL-GN-038	POS				
18	HCL-GN-049	POS				
19	HCL-GN-054	POS				
20	HCL-GN-061	POS				
21	HCL-GN-063	POS				
22	HCL-GN-070	POS				
23	HCS-GN-036	NEG				
24	HCS-GN-040	NEG				
25	HCS-GN-048	POS				
26	HCS-GN-049	NEG				
27	HCS-GN-053	POS				
28	HCS-GN-054	POS				
29	HCS-GN-055	POS				
30	HCS-GN-056	POS				
31	HCS-GN-078	POS				
32	HCS-GN-080	NEG				
33	HGJ-GN-039	POS				
34	HGJ-GN-040	POS				
35	HGJ-GN-041	POS				
36	HGJ-GN-068	POS	Not very pink		mucoid	
37	HMR-GN-036	POS				
38	HMR-GN-039	NEG				
39	HMR-GN-048	POS				
40	HMR-GN-053	POS				
41	HMR-GN-054	POS				
42	HMR-GN-058	POS				
43	HMR-GN-078	NEG				
44	HMR-GN-083	POS				
45	HSC-GN-032	NEG/POS	Mutation? Or Contamination?			
46	HSC-GN-034	POS				
47	HSC-GN-035	POS				
48	HSC-GN-050	POS				
49	HSC-GN-057	POS				
50	HSC-GN-060	POS				
51	HSC-GN-082	POS				Non-motile
52	LSP-GN-018	POS		4mm	Flat	motile
53	LSP-GN-029	NEG		8mm		very motile

Table 4a – *Shigella* Isolates Obtained from Health Canada and Grown on MacConkey Agar for 24 Hours: Observations

SAMPLE	SAMPLE ID	SEROTYPE	LACTOSE	COMMENTS	SIZE	MORPHOLOGY	MOBILITY	TSA
1	411-SH-29	*S. boydii 4*		No growth, grew on TSA				1-2mm, convex, non-motile, fatty
2	412-SH-30	*S. boydii 5*	NEG		4mm	non-round	non-motile	
3	413-SH-31	*S. boydii 6*	POS	4 colonies, mixed, non-pure	small			
4	414-SH-41	*S. boydii 9*	NEG	looks mixed, few POS colonies	3mm			
5	415-SH-49	*S. boydii 14*	NEG		3mm	non-round	semi-motile	
6	416-SH-34	*S. sonnei II*	NEG		3-5mm	flat	motile	
7	417-SH-36	*S. sonnei I*		No growth, grew on TSA				1-2mm, convex, non-motile
8	418-SH-37	*S. sonnei II*	NEG		3-5mm	flat	motile	
9	419-SH-152	*S. sonnei II*	NEG		5mm	flat	motile	
10	420-SH-165	*S. sonnei I*	NEG			non-round	semi-motile	
11	421-SH-63	*S. flexneri 1b*	NEG		1-2mm	convex	non-motile	
12	422-SH-125	*S. flexneri 6*	NEG	Looks mixed, 2 colonies, 1 is pinpoint	3mm	non-flat	non-motile	
13	423-SH-132	*S. flexneri 2b*	NEG		4mm	semi-flat	motile	
14	424-SH-160	*S. flexneri 3a*	NEG		5mm	flat	semi-motile	
15	425-SH-166	*S. flexneri 4a*	NEG		3mm	non-round	semi-motile	
16	426-SH-01	*S. dysenteriae 1*	NEG		3mm	non-flat	motile	
17	427-SH-143	*S. dysenteriae 2*	NEG		3mm	non-flat	motile	
18	428-SH-144	*S. dysenteriae 4*	NEG		3-5mm	non-flat	motile	
19	429-SH-145	*S. dysenteriae 7*	NEG		pinpoint	convex	non-motile	
20	430-SH-158	*S. dysenteriae 14*	NEG		5mm	non-flat	motile	

Table 5a – *Shigella* Isolates Obtained from McGill University Health Center and Grown on MacConkey Agar for 24 Hours: Observations

SAMPLE	SAMPLE ID	SPECIES	LACTOSE	COMMENTS	SIZE	MORPHOLOGY	MOBILITY
1	CHM-GN-078	*Shigella sonnei*	NEG				motile
2	CHM-GN-079	*Shigella group*	NEG				motile
3	CHM-GN-080	*Shigella group*	NEG		big		motile
4	CUM-GN-037	*Shigella sonnei*	NEG		big		motile
5	CUM-GN-038	*Shigella flexneri*	NEG		10mm		motile
6	HCL-GN-080	*Shigella group*	NEG				motile
7	HCL-GN-083	*Shigella group*	NEG		big		motile
8	HCS-GN-058	*Shigella sonnei*	NEG				motile
9	HCS-GN-076	*Shigella dysenteriae*	NEG				motile
10	HGJ-GN-081	*Shigella flexneri*	NEG				motile
11	HMR-GN-046	*Shigella group*	NEG				motile
12	HSC-GN-061	*Shigella group*	NEG		big		motile
13	LSP-GN-022	*Shigella flexneri*	NEG				motile
14	LSP-GN-023	*Shigella sonnei*	NEG				motile
15	LSP-GN-031	*Shigella boydii*	NEG				motile

Table 6a – *Shigella* Isolates Obtained from McGill University Health Center and Grown on Blood Agar for 24 Hours: Observations

SAMPLE	SAMPLE ID	SPECIES	HEMOLYTIC	COMMENTS	SIZE	MORPHOLOGY	MOBILITY
1	CHM-GN-078	*Shigella sonnei*	Alpha	grey	10mm	flat	motile
2	CHM-GN-079	*Shigella group*	Alpha		4mm	convex	
3	CHM-GN-080	*Shigella group*	Alpha	grey	10mm		
4	CUM-GN-037	*Shigella sonnei*	Alpha	grey	10mm		
5	CUM-GN-038	*Shigella flexneri*	Alpha		3mm	convex	
6	HCL-GN-080	*Shigella group*	Alpha	grey		convex	
7	HCL-GN-083	*Shigella group*	Alpha	grey	10mm		
8	HCS-GN-058	*Shigella sonnei*	Alpha		10mm		motile
9	HCS-GN-076	*Shigella dysenteriae*	Alpha	grey	3mm	convex	
10	HGJ-GN-081	*Shigella flexneri*	Alpha	Mix? 2 colonies	2mm and 1mm	convex	
11	HMR-GN-046	*Shigella group*	Alpha	grey	3mm	convex	
12	HSC-GN-061	*Shigella group*	Alpha	grey	3mm		
13	LSP-GN-022	*Shigella flexneri*	Alpha	grey	3mm	convex	
14	LSP-GN-023	*Shigella sonnei*	Alpha		10mm	flat	motile
15	LSP-GN-031	*Shigella boydii*		grey	7mm	viscous	