INSERTION/DELETION (INDEL) BASED APPROACH FOR THE DETECTION OF *ESCHERICHIA COLI* O157:H7 IN FRESHWATER ENVIRONMENTS

INSERTION/DELETION (INDEL) BASED APPROACH FOR THE DETECTION OF *ESCHERICHIA COLI* O157:H7 IN FRESHWATER ENVIRONMENTS

BY SHIRLEY Y. WONG, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2014) McMaster University

(Biology) Hamilton, Ontario

TITLE: Insertion/deletion (indel) based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments

AUTHOR: Shirley Y. Wong, B.Sc. (McMaster University, Hamilton, ON)

SUPERVISOR: Dr. Herb E. Schellhorn

NUMBER OF PAGES: x, 103

ABSTRACT

Though pathogenic strains represent a small portion of the total variety of existing *Escherichia coli* strains, they contribute extensively to human morbidity and mortality. Disease outbreaks caused by enterohaemorrhagic *E. coli* of the serotype O157:H7 and the "Big Six" serotypes (i.e., O26, O45, O103, O111, O121 and O145) have driven the development of assays for pathogen detection. From culture-based assays requiring several days for confirmation of target organisms, to quantitative PCR (qPCR) tests that provide pathogen identification in several hours' time, the sensitivity, specificity and speed of bacterial diagnostics have seen improvements that increased the efficacy of assays used to detect pathogens at clinically relevant levels. One relatively unexplored field of diagnostics is the use of conserved signature insertion/deletions (CSIs) as stable genetic markers for pathogen detection. This thesis presents two qPCR assays that target an *E. coli* O157:H7-specific insertion in a CSI. In a more preliminary study, an EvaGreen-based qPCR assay was developed that had a detection limit of 16 *E. coli* O157:H7 genome equivalents. An improved format of the O157:H7-specific CSI assay, using TaqMan probes, was later established. TaqMan probes are sequence-specific, while DNA-intercalating EvaGreen dye is sequence-independent. Though the TaqMan probe-based assay had a higher detection limit of 100 genome equivalents, the assay maintained detection sensitivity in presence of genetically similar (*E. coli* K-12) and dissimilar (fish sperm) DNA in excess amounts (1000-fold and 800-fold excess of target DNA, respectively), demonstrating its potential for pathogen detection in environmental samples where the presence of background flora may influence detection. These assays thus

represent an exploration into the use of CSIs as diagnostic tools. This thesis also provides a guide for future developments of pathogen detection using CSIs, such as those that may be present in toxigenic species of Cyanobacteria and human pathogens, including *Vibrio* and *Campylobacter*.

ACKNOWLEDGEMENTS

During my time here, the Schellhorn lab has been my second home. I am very grateful to Dr. Herb Schellhorn for accepting me as a graduate student in his lab, and for giving me the necessary space and tutelage to grow academically and as a person. I have learned important lessons from him on how to write, how to present and, most importantly, how to think critically. Thank you, Dr. Schellhorn, for the opportunity to be your student and for all the guidance you have given me.

I also thank Dr. Radhey Gupta for his advice these past few years. My project would not be successful if not for his expertise on indels. His input has been invaluable and I deeply appreciate his guidance during my time here. Thank you, Dr. Gupta.

And of course, I would like to thank all the friends I have made during my time here. In no particular order, I'd like to thank current and past members of the Schellhorn lab: Deepinder, Hamoody, Lingzi, Sarah, Steve, Mahi, Myra, Pardis, Rachel, Rosemary, Ryan and Victoria. You guys have been integral to my academic life, and oftentimes outside of it as well, when we go out for all-you-can-eat sushi or for some pho at Saigon. I have laughed so much with you all and made many wonderful memories. Thanks, guys!

(Also, I actually put the list in alphabetical order.)

Lastly, a humongous thank you to my family for their unending support throughout these years: their continuous encouragement during the troubleshooting phase of my experiments, their joyful celebration when I told them that I had results and their endless amusement when I geek out about bacteria. You guys have been amazing, and I hope I never let you guys down. Thank you, and thank you again.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

Diagnostic assays for enterohaemorrhagic *E. coli*

1.1 Introduction

Disease outbreaks due to bacterial contamination of the food and drinking water supply pose a danger to public health. The pathogenic *Escherichia coli* O157:H7 strain EDL933 was first isolated in United States in 1975 and became known as a serious foodborne pathogen in 1982, when it was associated with outbreaks of hemorrhagic colitis in Michigan and Oregon (Riley et al., 1983). The O157:H7 serotype was subsequently declared a food adulterant by the United States Department of Agriculture (USDA) in 1994 (FSIS, 1999). Later, this serotype became recognized as a waterborne pathogen when it was associated with large-scale outbreaks of hemorrhagic colitis in Missouri in 1989 (Swerdlow et al., 1992) and in Walkerton, Ontario, in 2000 (CCDR, 2000). These high-profile cases, along with others, highlight the need for food and drinking water quality monitoring, as well as the development of assays that facilitate these processes.

This literature review will focus on detection assays that have been developed to monitor pathogenic *E. coli* strains in the food and drinking water supply. Chromogenic media, antibody-based and DNA-based methods will be discussed.

1.2 *Escherichia coli*

Most *E. coli* strains are commensal bacteria found in the gastrointestinal tract of several mammals, including humans. However, certain strains have virulence factors that enable them to cause human morbidity and mortality.

Enterohaemorrhagic *E. coli* (EHEC) are a clinically-relevant subset of Shiga-toxigenic *E. coli* (STEC), a class of *E. coli* capable of secreting Shiga toxins. EHEC strains, such as strains in the O157:H7 serogroup, can cause outbreaks of bloody diarrhea and hemolytic uremic syndrome (HUS), particularly in children (Karmali et al., 1985), and as such, are a major focus of food quality monitoring programs. Other EHEC serogroups, such as O26, O45, O103, O111, O121 and O145, are also emerging as pathogens capable of causing outbreaks (Brooks et al., 2005), and have been declared by the USDA as food adulterants in 2011 (Almanza, 2011). Subsequent sections that discuss non-O157 strains will focus on these six serogroups.

1.3 Virulence factors and modes of pathogenesis

1.3.1 *E. coli* O157:H7

E. coli O157:H7 strains, such as EDL933, can cause disease by creating attaching and effacing (A/E) lesions on epithelial cell microvilli, destroying the host intestinal epithelium (DeVinney et al., 2001). Genes facilitating this process include *eae* (*Escherichia coli* attaching and effacing) and *tir* (translocated intimin receptor), which are encoded on the LEE pathogenicity island in the EDL933 genome (Tatsuno et al., 2000). O157:H7 can also produce Shiga toxins, which induce the production of pro-inflammatory cytokines in the renal tube epithelial cells of the host, leading to kidney damage and HUS (Lee et al., 2002). The production of Shiga toxins is dependent on the expression of *stx1* and *stx2* genes encoded by prophages CP-933V and BP-933W, respectively (Perna et al., 2001). More recently, enterohemolysin from the EDL933 strain

was found to induce a cytotoxic, pro-inflammatory release of the cytokine IL-1β, though the mechanism has yet to be fully elucidated. This virulence factor is encoded by *ehx* (Zhang et al., 2012).

1.3.2 Non-O157 EHEC

The six non-O157 EHEC serogroups most commonly associated with human disease also create A/E lesions on the intestinal wall, facilitated by *eae* gene products (Boerlin et al., 1999; Eklund et al., 2001). The enterohemolysin gene *ehly*, which enables the destruction of red blood cells, has been found in non-O157 EHEC strains (Eklund et al., 2001). In addition, the six non-O157 EHEC strains also express *stx1* and/or *stx2*, and produce Shiga toxins (Eklund et al., 2001). These virulence factors in non-O157 EHEC strains suggest similar modes of pathogenesis to those exhibited by O157:H7 strains (Boerlin et al., 1999).

1.4 Animal reservoirs and sources of infection

1.4.1 *E. coli* O157:H7

E. coli O157:H7 is carried asymptomatically in domesticated animals, particularly dairy cows (Dunn et al., 2004; Mechie et al., 1997; Wells et al., 1991) and beef cattle (Gannon et al., 2002; Oporto et al., 2008; Reinstein et al., 2009). Other species that harbour the strain include sheep (Kudva et al., 1996; Oporto et al., 2008), swine (Eriksson

et al., 2003; Feder et al., 2003), goats (Dontorou et al., 2004; Keen et al., 2006) and deer (Garcia-Sanchez et al., 2007; Renter et al., 2001).

In the US, ground beef was the most common source of O157 infection (i.e., infections caused by the O157:H7 and the Shiga-toxigenic O157:NM strains) in foodborne outbreaks that occurred from 1982-2002 (Rangel et al., 2005). In Canada, contaminated ground beef has led to outbreaks of O157:H7 infection in Winnipeg (Macdonald et al., 2000), as well as Calgary and surrounding rural areas (Currie et al., 2007).

Other food products, such as dairy and leafy green vegetables, can also be sources of infection. Contaminated raw milk was the source of multiple outbreaks in the US (CDC, 2007; Keene et al., 1997; Wells and Varel, 2008). Contaminated lettuce has been associated with outbreaks of *E. coli* O157 infections in the US (Ackers et al., 1998; CDC, 2006 Dec 14; Hilborn et al., 1999) and Canada (PHAC, 1997).

Waterborne *E. coli* O157:H7 infections are also a concern. As mentioned previously, outbreaks occurred in Missouri (Swerdlow et al., 1992) and Walkerton, Ontario (CCDR, 2000). The combination of manure use, irrigation and the presence of livestock increases the pathogen load in agricultural runoff that reaches surface waters, which are then used for drinking and recreational purposes (Jokinen et al., 2012).

1.4.2 Non-O157 EHEC

Non-O157 EHEC strains have reservoirs in ruminants, particularly cattle, and swine. Table 1 outlines specific animal reservoirs that harbour the six non-O157 EHEC serogroups discussed in this section.

Documented sources of non-O157 EHEC infections include primarily meat and dairy products. Sausage made from beef, mutton and pork were the sources of disease outbreaks caused by EHEC strains O26 in Denmark (Ethelberg et al., 2009), O103 in Norway (Schimmer et al., 2008) and O111 in Australia (Paton et al., 1996), respectively. Venison was also associated with EHEC O103 infections in US (Rounds et al., 2012). These outbreaks were due to undercooked or improperly cured meat. In addition, contaminated ice cream produced on a farm in Belgium was the source of an outbreak of EHEC O145 infections (De Schrijver et al., 2008). For waterborne infections, a case of HUS caused by EHEC O121 was traced back to drinking contaminated well water in Japan (Yatsuyanagi et al., 2002). However, most non-O157 EHEC infections have been associated with animal products and infections associated with drinking water are rare.

Table 1. Six non-O157 EHEC serogroups and their animal reservoirs.

Serogroups	Reservoirs (Reference(s))
O26	Cattle (Fukushima and Seki, 2004; Ojo et al., 2010; Wells et al., 1991), sheep (Blanco et al., 2003; Ojo et al., 2010), goat (Ojo et al., 2010), pig (Rios et al., 1999), roe deer (Mora et al., 2012), wild boar (Mora et al., 2012)
O45	Cattle (Wells et al., 1991)
O103	Cattle (Fukushima and Seki, 2004; Ojo et al., 2010; Wells et al., 1991), sheep (Ojo et al., 2010; Zweifel et al., 2004), goat (Duhamel et al., 1992)
O111	Cattle (Fukushima and Seki, 2004; Ojo et al., 2010; Wells et al., 1991), sheep (Ojo et al., 2010), goat (Ojo et al., 2010), pig (Ojo et al., 2010)
O121	Cattle (Fukushima and Seki, 2004), sheep (Zweifel et al., 2004), roe deer (Mora et al., 2012)
O145	Cattle (Ojo et al., 2010; Wells et al., 1991), roe deer (Mora et al., 2012)

1.5 Detection methods

1.5.1 Chromogenic media

Chromogenic media were developed to initially detect and isolate EHEC strains prior to further characterization by molecular methods, such as immunomagnetic separation and PCR, to enrich for and confirm the identity of targeted EHEC strains.

1.5.1.1 Modified MacConkey medium

MacConkey agar with sorbitol (SMAC) was first used as a method of detecting non-sorbitol-fermenting *E. coli* O157:H7, which form colourless colonies on the medium, unlike the sorbitol-fermenting background flora that produce pink colonies (March and Ratnam, 1986). Later, cefixime and tellurite were added to SMAC media (CT-SMAC) to reduce the growth of other non-sorbitol-fermenting organisms, such as *Proteus*, *Providencia* and *Aeromonas* species (Zadik et al., 1993), thus improving the original SMAC medium for O157:H7 isolation.

With the discovery of sorbitol-fermenting O157 strains in patients with HUS (Gunzer et al., 1992), new media were developed to detect both types of O157. One assay further modified the MacConkey medium using novobiocin, tellurite, sorbitol, X-gal and IPTG to create a selective chromogenic medium (Posse et al., 2008). The colour gradient produced by differential pH of colonies based on sorbitol fermentation ability is enhanced by the dark colour from X-gal hydrolyzation, which is indicative of lactose fermentation in the MacConkey medium (Posse et al., 2008). Resultant sorbitol positive colonies are confirmed with a pH indicator medium, phenol red broth, supplemented with L-

rhamnose, which O157 strains do not ferment (Posse et al., 2008). This assay allows for differentiation between sorbitol-fermenting and non-fermenting strains of O157.

To create a selective chromogenic agar for the differentiation of non-O157 EHEC serogroups O26, O103, O111 and O145, a similar modification to the MacConkey medium is made by the addition of novobiocin, tellurite, sucrose, sorbose, X-gal and IPTG (Posse et al., 2008). Similarly, the different abilities of the non-O157 EHEC strains to ferment sucrose and/or sorbose create a pH-dependent colour gradient that is enhanced by the dark colour from X-gal hydrolyzation (Posse et al., 2008). To avoid false positives, putative non-O157 EHEC strains are then assessed with a confirmation medium made of phenol red broth and a combination of D-arabinose, L-rhamnose, D-raffinose and dulcitol. The different abilities among the strains to ferment these carbohydrates produce different pH ranges, which are indicated as colour changes in the broth (Posse et al., 2008). Thus, these modifications to the MacConkey agar have enabled the isolation of both O157 and non-O157 EHEC strains, though laborious preparation of the different media required for these two-step isolation systems limits their widespread use.

1.5.1.2 Rainbow® Agar O157

Rainbow® Agar O157 (Biolog, USA) contains two chromogenic substrates: one for β-galactosidase (blue-black) and one for β-glucuronidase (red). O157 isolates appear as grey or black colonies due to their inability to produce functional β-glucuronidase. Other non-O157 EHEC strains, including O26 and O111, also produce colonies that range from dark blue to black (Bettelheim, 1998b). However, non-Shiga-toxin-producing strains

of those same serogroups can also produce black colonies (Bettelheim, 1998b), leading to false positives. Thus, this media is often used in conjunction with other methods to increase specificity.

1.5.1.3 CHROMagar™ O157 and CHROMagar™ STEC

CHROMagar™ O157 (CHROMagar Microbiology, France) has a proprietary chromogenic substrate that detects O157 strains, which produce pink or mauve colonies while other bacteria appear as steel blue or colourless. This medium offers higher sensitivity and better diagnostic ability than SMAC for the detection of O157 strains (Church et al., 2007). However, clinically-relevant sorbitol-fermenting O157 strains may appear as blue colonies, while non-Shiga-toxin-producing O157 strains may appear as pink colonies (Bettelheim, 1998a). To avoid misleading results, putative O157 isolates are further characterized using molecular methods.

Another media formulated by the same company, CHROMagar™ STEC is used to isolate Shiga-toxin producing strains of *E. coli*. STEC grow as pink or mauve colonies on the media, while other bacteria grow as blue or colourless. Though the media does not exclusively target the six major non-O157 EHEC strains involved in human disease outbreaks, five of the six major non-O157 EHEC strains (O45 was excluded from analysis) successfully grew on the media after 16 h of incubation (Hirvonen et al., 2012). Thus, this media can be used as an initial isolation step prior to subsequent analyses that identify isolates of interest.

1.5.2 Antibody-based techniques

1.5.2.1 Enzyme-linked immunosorbent assays (ELISAs)

ELISAs use two types of antibodies to capture the intended target protein. A sample with the particles of interest is washed over a solid substrate. The targeted particles are bound to primary antibodies adhered to the solid substrate. Any unbound particles are washed away, and a secondary antibody that is labelled with a reporter enzyme, such as horseradish peroxidase (HRP), is added. These labelled antibodies bind to the captured targets, and a substrate solution, such as tetramethylbenzidine, is added. The presence of the reporter enzyme induces a chemical change in the solution, indicating a positive result (Porstmann and Kiessig, 1992).

Antibodies for use in ELISAs have been developed to detect the presence of O157:H7 in samples. Monoclonal antibodies used for detecting the flagellar H7 antigen were developed in hybridomas (He et al., 1996). Recombinant antibodies were later created that target both the O157 and H7 antigens (Kanitpun et al., 2004). As well, bispecific, hybrid antibodies that recognized both *E. coli* O157:H7 and horseradish peroxidase were developed for use in single-enzyme ELISA assays (Guttikonda et al., 2007). In immunomagnetic ELISA format, these antibodies are capable of detecting 10 colony forming units (CFU) per mL (Guttikonda et al., 2007), creating a sensitive assay for O157:H7 detection.

1.5.2.2 High-throughput assays

With increasing interest in other clinically-relevant EHEC serogroups, recent methods aim to detect non-O157 EHEC strains in addition to O157. This section will outline high-throughput serotyping assays that can identify O157 and non-O157 EHEC serogroups.

Antibody microarray

Designed to detect the six major non-O157 EHEC serogroups, the antibody microarray uses a glass slide spotted redundantly with six anti-rabbit polyclonal antibodies against O26, O45, O103, O111, O121 and O145 (Hegde et al., 2013). Detection antibodies are labeled with Zenon Rabbit IgG labeling kit (Molecular Probes, OR, USA), which uses Alexa Fluor®-labeled Fab fragments that are anti-rabbit IgG. In tests with ground beef that was artificially inoculated with strains from the six serogroups, high fluorescent intensities were seen for each target serogroup, with no cross reactions (Hegde et al., 2013). Thus, this assay shows high specificity, and its microarray format can be scaled up for high-throughput serotyping of *E. coli* strains.

7-plex microbead-based immunoassay

Another high-throughput immunoassay is the 7-plex microbead-based immunoassay that allows for simultaneous detection of O157 and the six major non-EHEC O157 serogroups by Luminex xMAP® technology (Clotilde et al., 2013). Magplex magnetic microbeads (Luminex Corp., TX, USA) are conjugated with capture antibodies

raised against the O26, O45, O103, O111, O121, O145 and O157 serogroups. To produce signal, a reaction between biotinylated detector antibodies and streptavidin labeled with the protein-bound chromophores R-phycoerythrin is used. The reactions are done in a 96-well microtitre plate and the Luminex 100 flow analyzer reads fluorescent signals from the wells to rapidly serotype multiple samples. Using enriched bacterial cultures, this immunoassay was able to recover 98.7% (78/79) of strains from the seven tested serogroups (Clotilde et al., 2013). This multiplexed immunoassay enables efficient high-throughput serotyping of close to 100 analytes, though its applicability to recovery of EHEC strains from food matrices has yet to be assessed.

1.5.3 DNA-based methods

1.5.3.1 Multiplex PCR

As an EHEC detection tool, multiplex PCR (mPCR) was first developed to amplify O157 virulence genes (*stx* and *eae*) (Gannon et al., 1992), with later additions of the flagellar H7 gene (Gannon et al., 1997) and the gene encoding the O-antigen, rfb_{O157} (Paton and Paton, 1998a), as an improvement on specific identification of O157:H7 strains. To detect non-O157 EHEC serogroups, more recent mPCR tests use primers designed to amplify the O-antigen of serogroups O26, O45, O103, O111, O121 and O145 (Paddock et al., 2012), as well as the virulence genes associated with EHEC strains (Bai et al., 2012). These assays showed higher sensitivity at detecting non-O157 EHEC strains in cattle feces when compared with culture-based methods that used immunomagnetic separation to distinguish between serogroups. With a pre-enrichment step (6 h) in growth

media, the mPCR assays were able to detect 10-100 colony forming units (CFU) per gram of sample (Bai et al., 2012; Paddock et al., 2012).

1.5.3.2 Multiplex PCR – commercial systems

Commercial mPCR assays are also used for rapid detection of O157:H7 and non-O157 EHEC contamination in food products. The following are two commercial kits that are currently adopted by Canadian and American governments for food quality monitoring.

DuPont Qualicon BAX® System

A diagnostic test kit from DuPont, the BAX® System PCR Assay for *E. coli* O157:H7 MP (BAX-MP) uses multiplex PCR to detect two DNA targets specific to the O157:H7 strain. The sample must be enriched for 8-24 hours using the BAX® System MP media prior to running the multiplex PCR assay (DuPont, 2012). The BAX® System has shown 100% sensitivity to low levels of O157:H7 in beef samples, but this sensitivity is lowered by the presence of background bacteria, such as other Enterobacteriaceae (Bosilevac et al., 2010).

In the US, this test has been adopted by the Food Safety and Inspection Service as a part of commercial screening for the presence of O157:H7 in manufactured beef products (FSIS, 2010). Meanwhile, in the Compendium of Analytical Methods, Health

Canada has stated that using the BAX-MP kit in combination with the BAX® System MP media has been validated only on ground beef and beef trim.

BioControl Assurance GDS® (Gene Detection System)

Produced by BioControl, the Assurance GDS® MPX is a pathogen detection system that tests for the top 7 EHEC serogroups that are implicated in disease outbreaks. The sample is enriched in BioControl mEHEC® media for 10-18 hours. Using magnetic particles coated with antibodies, *E. coli* with the targeted serogroups (i.e., O26, O45, O103, O111, O121, O145 and O157) are then extracted from the enriched sample. After resuspension in a buffer solution, a multiplex PCR assay is performed that is specific to the *stx1*, *stx2* and *eae* genes that are carried by the targeted EHEC strains (BioControl, 2012). The Assurance GDS® *E. coli* O157:H7 kit has a tested sensitivity of over 90% (Bosilevac et al., 2010; Feldsine et al., 2005) even in low concentrations (Bosilevac et al., 2010). However, the presence of background organisms greatly reduces the sensitivity of the assay (Bosilevac et al., 2010).

Recently, this commercial method has become a part of government policies on food quality monitoring. Starting from June 2012, the Canadian Food Inspection Agency will require testing beef trimmings for export from Canada to USA using the Assurance GDS® MPX Top 7 STEC testing system (Elmufti, 2012). This is done as a part of its Hazard Analysis and Critical Control Point (HACCP) verification testing program for beef products (Elmufti, 2012).

1.5.3.3 Multiplex quantitative PCR

For quantitative detection, probe-based quantitative PCR (qPCR) can detect multiple gene targets in reactions to be multiplexed, allowing for simultaneous detection of genes for virulence factors, O-antigens and H-antigens, as well as other genetic markers of interest.

TaqMan® assays (Life Technologies) use oligonucleotide probes complementary to the target sequence and will bind to target amplicons. These probes contain a reporter dye on the 5' end and a quencher dye on the 3' end, and the fluorescence signal of the reporter is quenched by the quencher when it is in close proximity. As *Taq* DNA polymerase extends PCR products, 5'-3' exonuclease activity of the enzyme cleaves the reporter from the oligo probe, and the fluorescence signal of the reporter increases as it is separated from the quencher (Heid et al., 1996). Different coloured fluorophores can be used for probes, enabling multiplex reactions.

The first multiplex TaqMan® assays designed to detect EHEC strains used probes for *stx* and O157:H7-specific *eae* (Sharma et al., 1999) or rfb_{O157} (Hsu et al., 2005). To detect the clinically relevant non-O157 EHEC strains, recent assays use *wzx* genes in the O-antigen gene clusters in addition to virulence genes to detect the six major non-O157 EHEC serogroups (Fratamico et al., 2011), or a combination of *wzy* and *wzx* genes to detect up to nine non-O157 EHEC serogroups (Lin et al., 2011). These assays are highly specific, with no false positives when tested against 120 *E. coli* strains and 23 other bacterial species (Lin et al., 2011). They also have low detection limits (50 CFU per reaction) and, with prior enrichment, can detect 1-2 CFU/25 g of food sample (Fratamico

et al., 2011). Though these assays offer high sensitivity and specificity, they can take up to 2 days to obtain results due to enrichment steps.

1.6 Summary

Since its discovery as a food- and water-borne pathogen in the 1980s, the O157:H7 *E. coli* serotype has been the focus for the development of pathogen detection systems. However, with the emergence of non-O157 EHEC strains as agents capable of causing disease outbreaks, traditional assays that rely primarily on chromogenic media for isolation and detection are now insufficient for the accurate isolation of these strains. Thus, molecular-based assays, such as the 7-plex microbead-based immunoassay and TaqMan® probe-based detection systems, provide additional steps to characterize and confirm putative isolates from samples. These assays offer highly sensitive and specific pathogen detection for food and water quality monitoring purposes.

CHAPTER 2:

E. coli O157:H7-specific assay, EvaGreen dye format

2.1 Abstract

Enterohaemorrhagic *E. coli* (EHEC) are serious biological contaminants found in food and water. Traditional methods for detecting EHEC strains, which involve different combinations of chromogenic media, immunogenic tests for serotyping and PCR assays for virulence genes, are well-established, but they are time-consuming and require pure cultures of putative isolates. Here, we propose a novel detection method for pathogenic strains of *E. coli* O157:H7 based on conserved insertions/deletions (indels) (CSIs) in widely distributed proteins. A quantitative PCR assay was developed for an *E. coli* O157:H7-specific insertion, and its specificity was tested against a panel of 13 *E. coli* laboratory strains. The assay is linear from 16 to 1.6 x 10^6 genome equivalents and is specific to *E. coli* O157:H7, as no detectable product was seen with non-O157:H7 *E. coli* strains (12 non-O157:H7 pathogenic strains and 1 K-12 strain). This O157:H7-specific indel assay contributes to a growing body of research on molecular-based tests for pathogen detection in the environment by providing one-step identification and quantification of clinically relevant pathogenic *E. coli*.

2.2 Introduction

Enterohaemorrhagic *E. coli* O157:H7 has been the cause of foodborne-illnesses over the years, with outbreaks of strains with this serotype most commonly associated with contaminated hamburger meat (Rangel et al., 2005). Modes of pathogenesis of O157:H7 strains include the production of virulence factors Tir (DeVinney et al., 1999) and intimin (Donnenberg et al., 1993) that facilitate the formation of attaching and effacing lesions on the host intestinal epithelium, and Shiga toxins that induce the production of pro-inflammatory cytokines in the renal tube epithelial cells of the host, which leads to kidney damage and hemolytic uremic syndrome (HUS) (Lee et al., 2002). Because of this serotype's ability to cause widespread damage to human health, as well as economic losses associated with recalls of contaminated food products, this human pathogen has been a pertinent focus in the development of detection assays.

Detection systems based on conserved *E. coli* genetic markers can provide sensitive and specific assays for pathogen detection. Conserved genetic markers, such as conserved signature insertions/deletions (indels) (CSIs), allow for stable identification of evolutionary clades. These CSIs have been found in the *Gammaproteobacteria* class, where four indels generated by rare genomic events were found in signature proteins that are specific to *Gammaproteobacteria* and to species in that class within the orders *Enterobacteriales*, *Pasteurellales*, *Vibrionales*, *Aeromonadales* and *Alteromonadales* (Gao et al., 2009). Additionally, CSIs can also be used to further characterize evolutionary clades to the species level, as seen by research done in Cyanobacteria (Gupta

and Mathews, 2010). CSIs that can classify organisms to a low taxonomic level, such as species, are applicable to the development of diagnostics.

Quantitative PCR (qPCR) assays provide more rapid methods of pathogen identification than traditional media-based methods, by selectively amplifying genetic markers of pathogens, rather than relying on the culturability of target organisms. EvaGreen (Biotium, California), a fluorescent DNA-intercalating dye that is a more stable and sensitive alternative to SYBR Green I, is used in qPCR to provide sensitive detection. The EvaGreen dye format of the indel-based *E. coli* O157:H7-specific qPCR assay targets an O157:H7-specific insertion in the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily, also known as the iron uptake protein (IUP) family. The objective of the project was to explore the use of CSIs as diagnostic tools by assaying the IUP indel assay for its limit of detection, as well as its specificity.

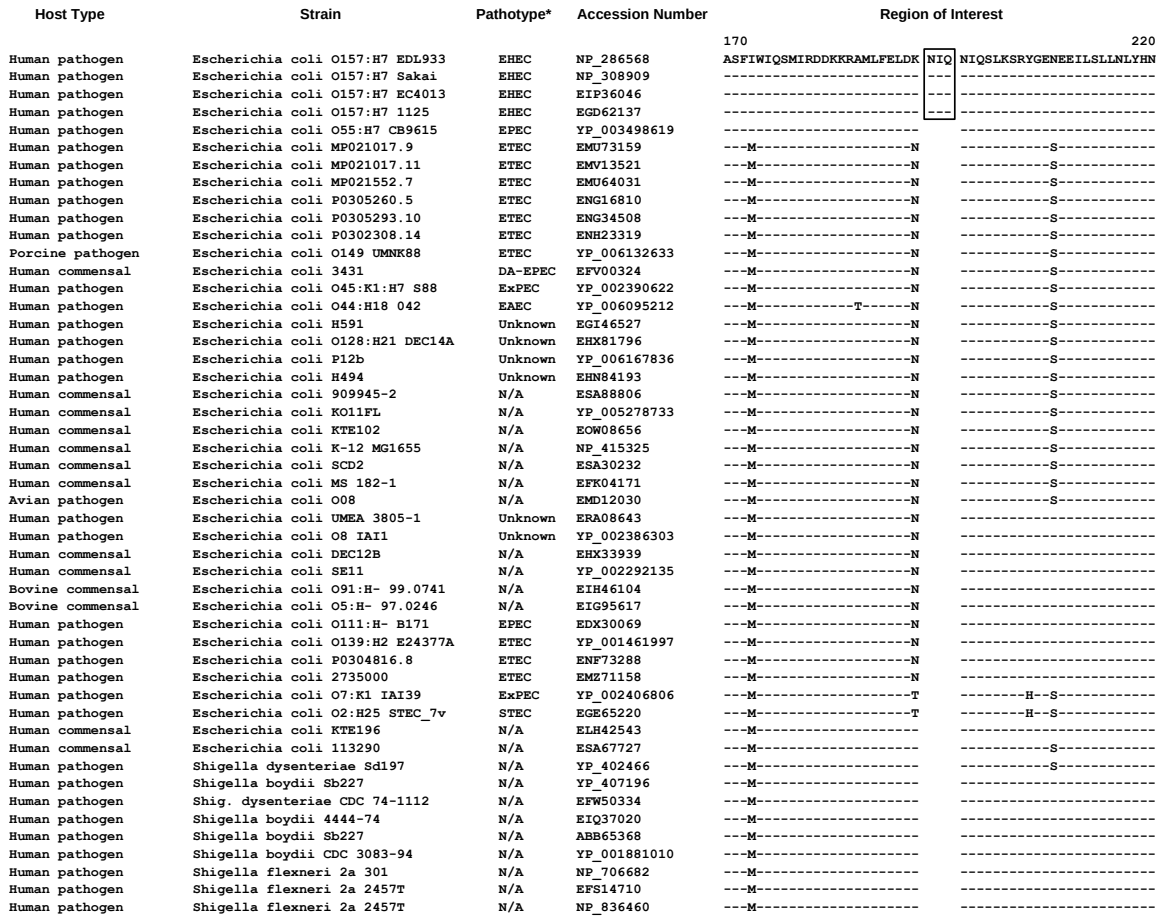
2.3 Methods

2.3.1 Primer sequences

Control and *E. coli* O157:H7-specific primers for the assay were designed using a multiple nucleotide sequence alignment of the CSI found in the 2-oxoglutarate and Fe(II)-dependent oxygenase protein superfamily (Figure 1). The O157:H7-specific assay primers were designed to produce a 153 bp amplicon from *E. coli* strains of the O157:H7 serotype, while the control primers produced amplicon sizes of 294 or 303 bp, depending on indel status, in all *E. coli* (Table 2).

2.3.2 Genomic DNA extractions of *E. coli* strains

An O157:H7 reference strain (EDL933), 12 non-O157:H7 pathogenic strains and a K-12 strain (MG1655) were used (Table 3). *E. coli* strains were streaked out on LB plates. For overnight cultures, a single colony from each were inoculated into 10 mL of LB and shaken at 200 rpm overnight at 37°C. DNA extractions were performed with a kit (Bacterial Genomic DNA Isolation Kit, cat. no. 17900, Norgen Biotek Corp., Ontario) using 1 mL of overnight culture per strain. DNA extracts were quantified with the NanoDrop 2000 (ThermoScientific, Massachusetts).



Host Type	Strain	Pathotype*	Accession Number	Region of Interest (170–220)
				ASFIWIQSMIRDDKKRAMLFELDK NIQ NIQSLKSRYGENEEILSLLNLYHN
Human pathogen	Escherichia coli O157:H7 EDL933	EHEC	NP_286568	
Human pathogen	Escherichia coli O157:H7 Sakai	EHEC	NP_308909	
Human pathogen	Escherichia coli O157:H7 EC4013	EHEC	EIP36046	
Human pathogen	Escherichia coli O157:H7 1125	EHEC	EGD62137	
Human pathogen	Escherichia coli O55:H7 CB9615	EPEC	YP_003498619	
Human pathogen	Escherichia coli MP021017.9	ETEC	EMU73159	
Human pathogen	Escherichia coli MP021017.11	ETEC	EMV13521	
Human pathogen	Escherichia coli MP021552.7	ETEC	EMU64031	
Human pathogen	Escherichia coli P0305260.5	ETEC	ENG16810	
Human pathogen	Escherichia coli P0305293.10	ETEC	ENG34508	
Human pathogen	Escherichia coli P0302308.14	ETEC	ENH23319	
Porcine pathogen	Escherichia coli O149 UMNK88	ETEC	YP_006132633	
Human commensal	Escherichia coli 3431	DA-EPEC	EFV00324	
Human pathogen	Escherichia coli O45:K1:H7 S88	ExPEC	YP_002390622	
Human pathogen	Escherichia coli O44:H18 042	EAEC	YP_006095212	
Human pathogen	Escherichia coli H591	Unknown	EGI46527	
Human pathogen	Escherichia coli O128:H21 DEC14A	Unknown	EHX81796	
Human pathogen	Escherichia coli P12b	Unknown	YP_006167836	
Human pathogen	Escherichia coli H494	Unknown	EHN84193	
Human commensal	Escherichia coli 909945-2	N/A	ESA88806	
Human commensal	Escherichia coli KO11FL	N/A	YP_005278733	
Human commensal	Escherichia coli KTE102	N/A	EOW08656	
Human commensal	Escherichia coli K-12 MG1655	N/A	NP_415325	
Human commensal	Escherichia coli SCD2	N/A	ESA30232	
Human commensal	Escherichia coli MS 182-1	N/A	EFK04171	
Avian pathogen	Escherichia coli O08	N/A	EMD12030	
Human pathogen	Escherichia coli UMEA 3805-1	Unknown	ERA08643	
Human pathogen	Escherichia coli O8 IAI1	Unknown	YP_002386303	
Human commensal	Escherichia coli DEC12B	N/A	EHX33939	
Human commensal	Escherichia coli SE11	N/A	YP_002292135	
Bovine commensal	Escherichia coli O91:H- 99.0741	N/A	EIH46104	
Bovine commensal	Escherichia coli O5:H- 97.0246	N/A	EIG95617	
Human pathogen	Escherichia coli O111:H- B171	EPEC	EDX30069	
Human pathogen	Escherichia coli O139:H2 E24377A	ETEC	YP_001461997	
Human pathogen	Escherichia coli P0304816.8	ETEC	ENF73288	
Human pathogen	Escherichia coli 2735000	ETEC	EMZ71158	
Human pathogen	Escherichia coli O7:K1 IAI39	ExPEC	YP_002406806	
Human pathogen	Escherichia coli O2:H25 STEC_7v	STEC	EGE65220	
Human commensal	Escherichia coli KTE196	N/A	ELH42543	
Human commensal	Escherichia coli 113290	N/A	ESA67727	
Human pathogen	Shigella dysenteriae Sd197	N/A	YP_402466	
Human pathogen	Shigella boydii Sb227	N/A	YP_407196	
Human pathogen	Shig. dysenteriae CDC 74-1112	N/A	EFW50334	
Human pathogen	Shigella boydii 4444-74	N/A	EIQ37020	
Human pathogen	Shigella boydii Sb227	N/A	ABB65368	
Human pathogen	Shigella boydii CDC 3083-94	N/A	YP_001881010	
Human pathogen	Shigella flexneri 2a 301	N/A	NP_706682	
Human pathogen	Shigella flexneri 2a 2457T	N/A	EFS14710	
Human pathogen	Shigella flexneri 2a 2457T	N/A	NP_836460	

Figure 1. Protein alignment showing the 3 amino acid insertion (boxed) located in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily.

Sequences are shown with the species name, followed by the GenBank accession number. The dashes (-) indicate amino acid identity to *E. coli* O157:H7 str. EDL933, and the numbers at the top indicate amino acid positions of the sequence from EDL933. *EHEC = enterohaemorrhagic *E. coli*; EPEC = enteropathogenic *E. coli*; ETEC = enterotoxigenic *E. coli*; DA-EPEC = diffuse adhering enteropathogenic *E. coli*; ExPEC = extraintestinal pathogenic *E. coli*; STEC = Shiga toxigenic *E. coli*; N/A = not applicable.

Table 2. Primers used in the IUP indel assay.

Name	Sequence (5'-3')	Amplicon size (bp)	Target	Reference
IUP_O157F	TTG TAT CCC TCC AGC AGC C	153	*E. coli* O157:H7	This study
IUP_O157R2	CAG CGA CTG AAT ATT CTG AAT ATT TTT GTC			
IUP_F	GGC GAA CTG GTC GTT AAT GAC	294/303[a]	All *E. coli*	This study
IUP_R	TCA GAT CTC CGA CCA TTC CC			

[a] 294 bp produced by non-O157:H7 *E. coli* strains; 303 bp produced by *E. coli* O157:H7 strains

Table 3. Strains used in the development of the IUP indel assay.

Seropathotype[a]	Serotype	Strain	Pathotype[b]	Source[c]
A	O157:H7	EDL933	EHEC	(Riley et al., 1983)
B	O26:H11	CL1	EHEC	(Karmali et al., 2003)
	O111:NM	R82F2	EHEC	(Karmali et al., 2003)
	O121:H19	CL106	EHEC	(Karmali et al., 2003)
	O145:NM	N00-6496	EHEC	(Karmali et al., 2003)
C	O5:NM	N00-4067	EHEC	(Karmali et al., 2003)
	O113:H21	CL3	EHEC	(Karmali et al., 2003)
	O121:NM	N99-4390	EHEC	(Karmali et al., 2003)
D	O103:H25	N00-4859	EHEC	(Karmali et al., 2003)
	O172:NM	EC6-484	EHEC	(Karmali et al., 2003)
E	O84:NM	EC2-044	EHEC	(Karmali et al., 2003)
	O98:H25	EC3-377	EHEC	(Karmali et al., 2003)
----	O6:K2:H1	CFT073	UPEC	(Mobley et al., 1990)
----	OR:H48:K-	MG1655	K-12	CGSC

[a] seropathotypes classified according to Karmali et al. (Karmali et al., 2003)
[b] EHEC = enterohaemorrhagic *E. coli*; UPEC = uropathogenic *E. coli*
[c] CGSC = Coli Genetic Stock Center

2.3.3 Optimization of annealing/extension temperature for IUP indel assay

Optimization of the IUP indel assay was performed using 10 µL reaction volumes in 96-well plates. The following components were added: 2X SsoFast™ EvaGreen® Supermix (cat. no. 172-5201, BioRad, California), 0.2 µM of IUP_O157F forward primer, 0.2 µM of IUP_O157R2 reverse primer, 10 ng of DNA template (EDL933 and MG1655) and double-distilled sterile water (up to 10 µL of reaction volume). For positive controls, control primers (IUP_F and IUP_R) were used. Reactions were done in triplicate. A gradient qPCR program was used: initial denaturation at 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 5 seconds and annealing/extension at 61°C, 63.2°C, 65°C, 66°C or 67°C for 10 seconds. A melt curve was then performed with a temperature range from 55-95°C with increments of 0.2°C for 5 seconds at each step. Fluorescence was read after every annealing/extension step and after every increment of 0.2°C for the melt curve. The CFX96 Touch™ Real-Time PCR Detection System (BioRad, California) was used. Data analysis was done with the BioRad CFX Manager™.

2.3.4 Determining the sensitivity of IUP indel assay

To calculate the number of bacterial genomes per dilution, the following formula was used (Whelan et al., 2003):

$$\text{Genome equivalent} = \frac{(6.02 \times 10^{23}\ \text{copy/mol}) \times (\text{DNA amount in grams})}{(\text{DNA length in bp}) \times (660\ \text{g/mol/bp})} \qquad [1]$$

The EDL933 genome size of 5.62 Mb was used, accounting for both the bacterial chromosome (5.5 megabases) (Perna et al., 2001) and the pO157 virulence plasmid (92

kilobases) (Burland et al., 1998). Six 10-fold serial dilutions of EDL933 DNA extract were made, from 1.6 x 10^6 to 16 genome equivalents/µL.

For the standard curve, 10 µL qPCR reaction mixtures were used. The components of the reaction mixtures were the same as those used for the optimization of the annealing/extension temperature of the IUP indel assay, except for the addition of 1 µL of 10-fold dilutions of EDL933 DNA extract for template. For the qPCR program, the steps were: initial denaturation at 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 5 seconds and annealing/extension at 66°C for 10 seconds. A melt curve analysis was then performed as outlined previously.

2.3.5 Determining the specificity of the IUP indel assay

Specificity test of the assay was done with 10 µL reaction volumes. The components were the same as those used in the optimization of the annealing/extension temperature of the IUP indel assay, except for the addition of 10 ng of DNA extract from different pathogenic strains (Table 3). Reactions were also done in duplicate. For the qPCR program, the same program was used as the one for determining the sensitivity of the IUP indel assay. Positive controls were run at 65°C for annealing/extension steps.

In addition to using melt curves to determine the specificity of the assay, PCR products were run on an agarose gel. A 2% TAE gel made with 1 µg/mL of ethidium bromide was run at 130 V for 45 minutes. Amounts loaded were 5 µL for samples, 2 µL for controls and 1.5 µL of GeneRuler 100 bp DNA ladder (ThermoScientific, USA).

2.4 Results

2.4.1 Optimization of annealing/extension temperature

The IUP indel primers amplified the *E. coli* K-12 str. MG1655 at annealing temperatures of 61°C, 63.2°C and 65°C, as seen by visible amplification curves that crossed the signal threshold (570 RFU) from cycle 37 and onwards (Figure 2). At 66°C, no amplification of *E. coli* K-12 was detected. *E. coli* O157:H7 was amplified at all tested temperatures. Melt curves indicated specific amplification using control and O157:H7-specific assay primers (Figure 3).

2.4.2 Sensitivity and linear range

For the standard curve, the IUP indel assay was linear from 16 to 1.6 x 10^6 genome equivalents, with a reaction efficiency (E) of 99.7% and a coefficient of determination (R^2) of 0.999 (Figure 4). Melt curve analysis of the reactions indicated specific amplification (data not shown).

2.4.3 Specificity of the assay

Only the tested O157:H7 pathogenic strain had fluorescence signals above the signal threshold (Figure 5). On the agarose gels, the 153 bp amplicon was seen for *E. coli* O157:H7, whereas the 153 bp product was not seen for the 12 tested non-O157 pathogenic strains (Figure 6). Amplification was seen for all positive controls (Figure 6).

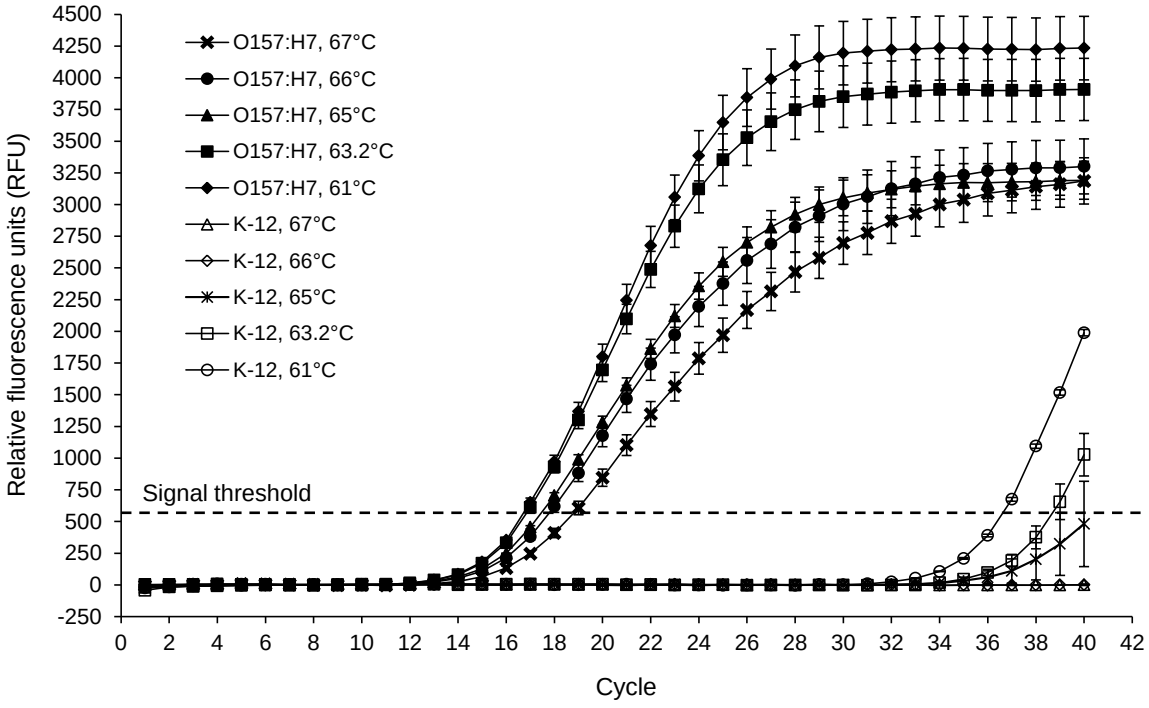


Figure 2. IUP indel assay is specific for *E. coli* O157:H7 at 66°C.

Amplification curves for reactions performed at 61°C, 63.2°C, 65°C, 66°C and 67°C annealing/extension temperatures with *E. coli* O157:H7 and K-12 strains are shown. The signal threshold is indicated by the dotted line. Standard error values were calculated based on three technical replicates and are shown as error bars.

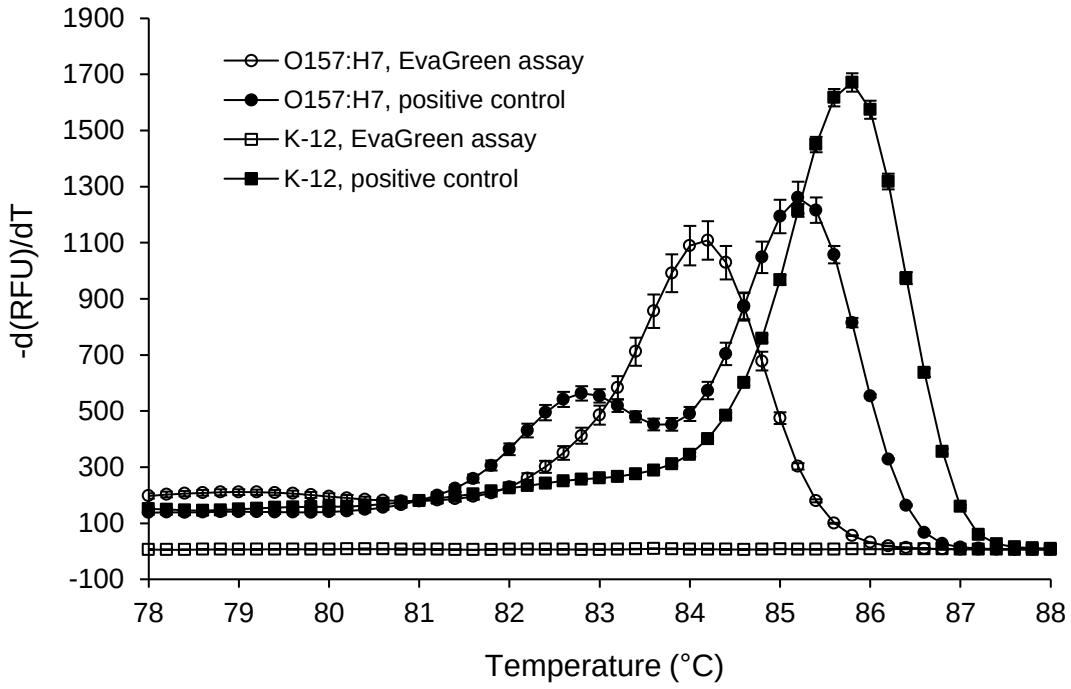


Figure 3. Specific amplification for the *E. coli* O157:H7-specific EvaGreen-based assay and positive controls indicated by melt curves.

Sample melt curves for *E. coli* O157:H7 and K-12 strains are shown. The negative derivative of the relative fluorescence unit measurements (-d(RFU)/dT) was calculated at each temperature reading. Standard error values were calculated based on three technical replicates and are shown as error bars.

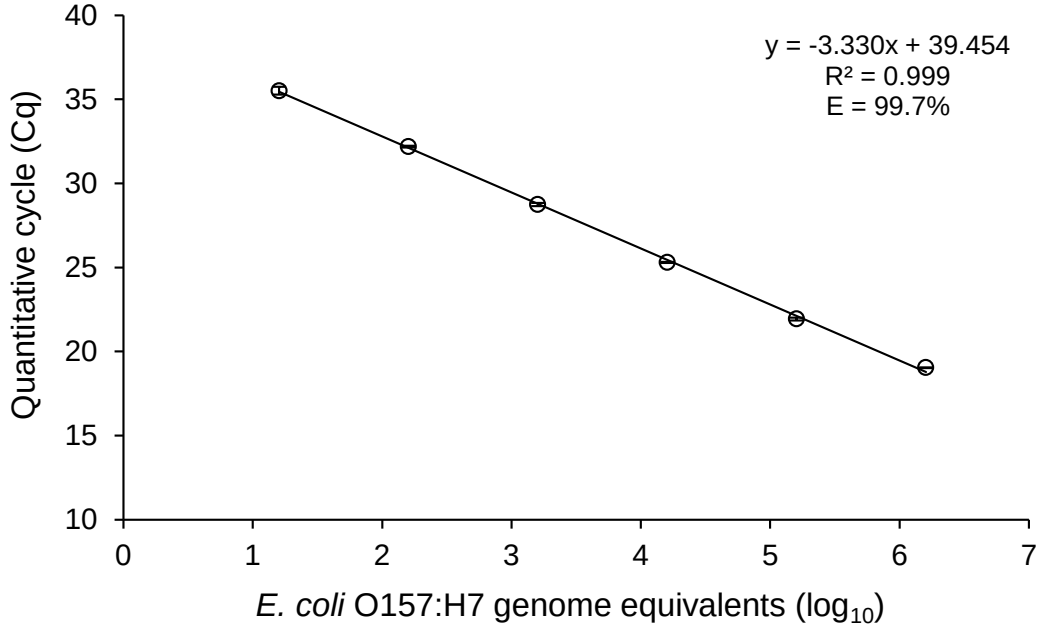


Figure 4. IUP indel assay is linear from 16 to 1.6 x 10^6 genome equivalents.

The Cq (quantitative cycle) is shown as a function of the $\log_{10}$ transformation of the starting quantity (i.e., genome equivalents). The coefficient of determination (R^2) and reaction efficiency (E) are shown. The 10-fold dilutions were done in triplicate.

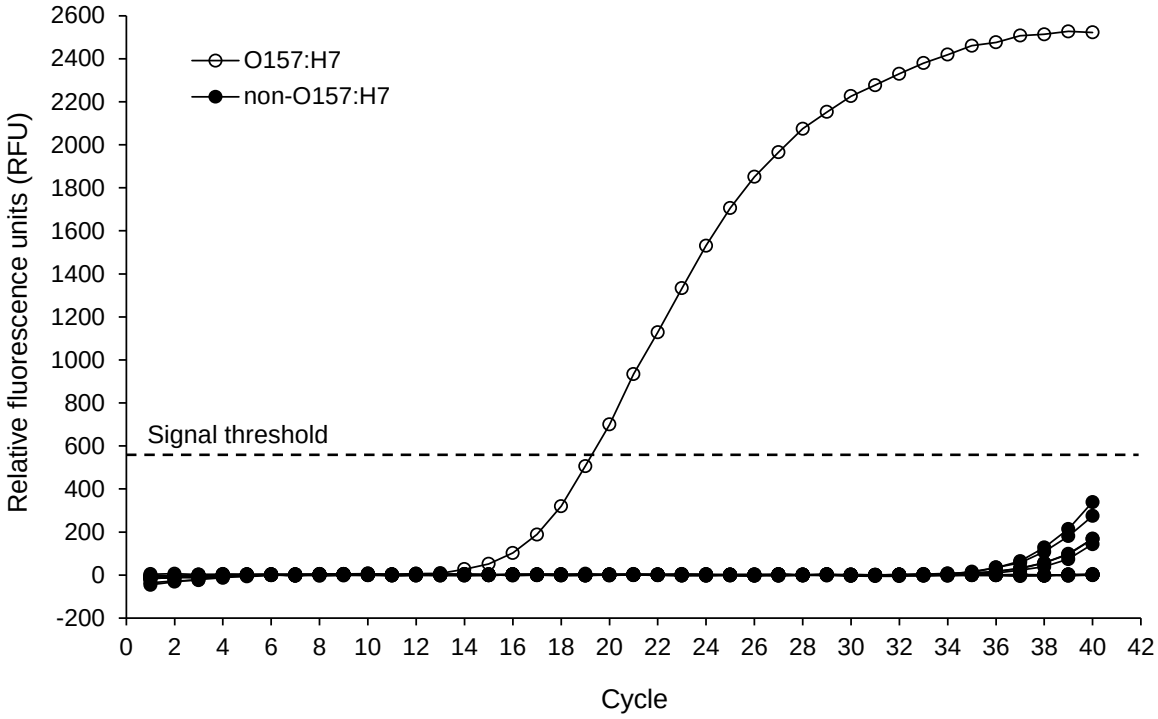


Figure 5. Only *E. coli* O157:H7 produces amplification signals above threshold.

Amplification curves are shown for 12 non-O157:H7 pathogenic *E. coli* strains and one *E. coli* O157:H7 strain. The average of two replicates was used for each data point. The dotted line across the graph indicates the signal threshold.

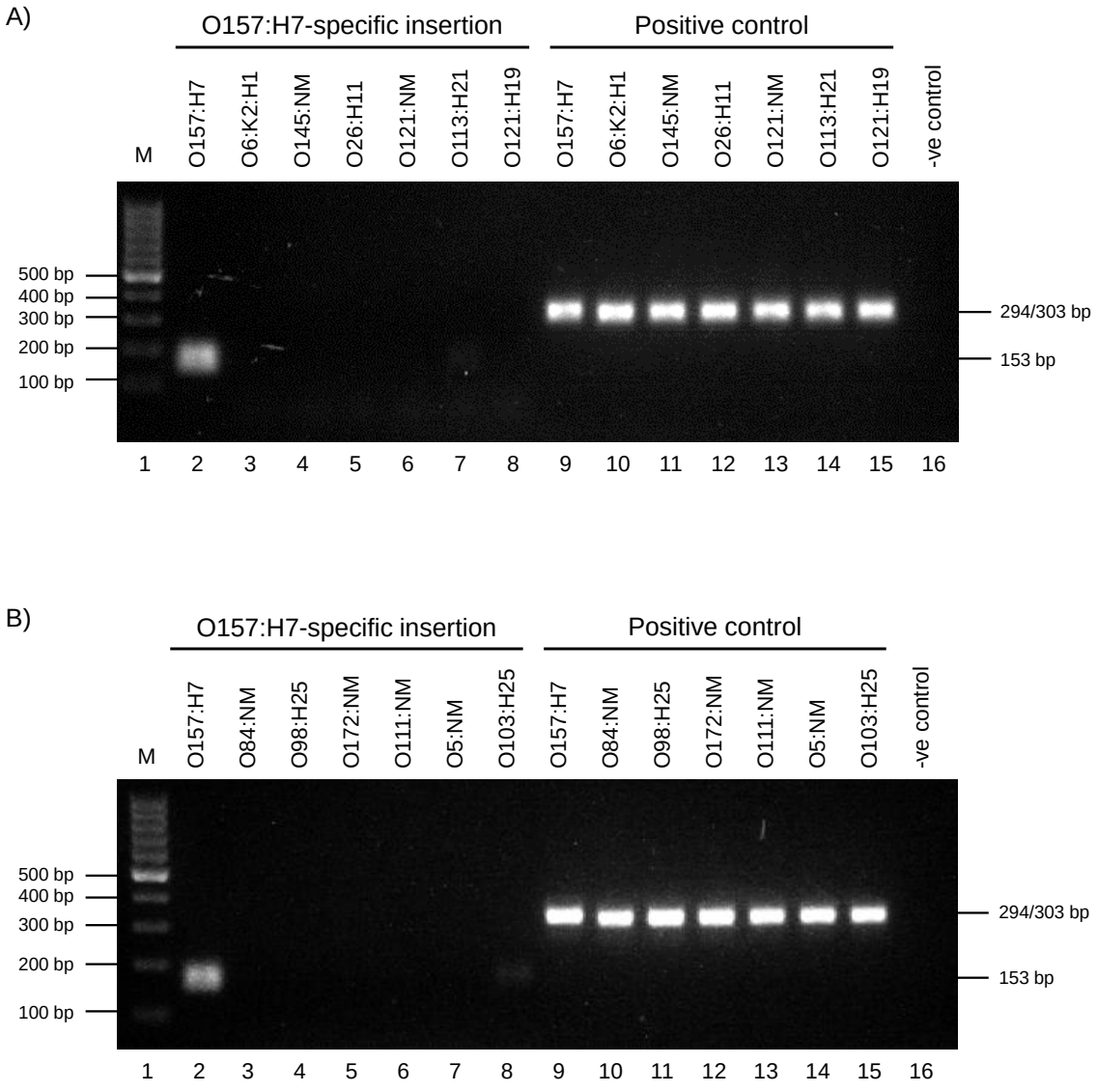


Figure 6. Confirmation of qPCR products in specificity assays.

Results for the IUP indel assay (lanes 2-8), positive controls (lanes 9-15) and negative control (lane 16), and a 100 bp marker (lane 1) are shown. The diagnostic band of the IUP indel assay is 153 bp in size. The positive control primers amplify a region surrounding the IUP indel, producing a product band of 303 bp (O157:H7) or 294 bp (the 12 non-O157:H7 pathogenic strains).

2.5 Discussion

Compared with PCR-based assays, DNA-intercalating dyes, such as SYBR Green I and EvaGreen, provide more sensitive and quantitative detection of *E. coli* O157:H7 when used in quantitative PCR (qPCR) tests. Assays have been developed that target the O- and H-antigen-encoding genes, *rfbE* and *fliC* (Yang et al., 2013), as well virulence genes, such as *eae* (Park et al., 2011). While these assays improve on existing PCR-based detection strategies, detection with multiple targets relies on positive signals from all targets involved for conclusive pathogen identification, while tests that target virulence genes may not target *E. coli* O157:H7 strains directly, as serotypes within the Shiga-toxigenic *E. coli* (STEC) pathotype also share certain virulence genes, such as *stx* and *eae* (Eklund et al., 2001). In contrast, using a conserved signature insertion/deletion (CSI) as a molecular marker enables unambiguous, single-target identification of the target organism.

This study presented an EvaGreen-based O157:H7-specific indel assay that targets a CSI found in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily (also known as iron uptake protein or "IUP"). The assay was found to be specific to the O157:H7 strain through specificity testing with a panel of 13 non-O157:H7 *E. coli* strains. The assay also had a detection limit (16 genome equivalents) comparable to published qPCR-based O157:H7 detection tests that use DNA-intercalating dye, such as one targeting *rfbE* and *fliC* that had a detection limit of 2.95×10^1 genome equivalents (Yang et al., 2013), and one that uses *eae* as a genetic marker and has a limit of 7.33×10^1 genome equivalents (Park et al., 2011). However, limit of detection of the IUP indel assay

was not established in environmental samples or from extracts of spiked food matrices. Thus, the currently-established limit of detection does not reflect on the practical detection limit of the assay, though this reported parameter does demonstrate the possibility of using CSIs for sensitive detection.

A future development of this assay includes the conversion of this O157:H7-specific CSI into a TaqMan-probe based qPCR test. While EvaGreen-based qPCR does offer increased sensitivity over end-point PCR, DNA-intercalating dyes are insensitive to the sequence of the molecular target and can bind to amplification artefacts, artificially inflating results or creating high background fluorescence and reducing the dynamic range of the assay (Higgins et al., 2003). TaqMan probes circumvent this problem, as the probes are designed to a specific target sequence. This sequence-specificity allows for the use of the IUP indel assay in environmental samples that may contain high concentrations of non-target genetic material, such as water samples from more polluted areas or primary effluent samples from wastewater treatment plants.

CHAPTER 3:

E. coli O157:H7-specific assay, TaqMan probe format

3.1 Abstract

Enterohaemorrhagic *E. coli* O157:H7 has been implicated in outbreaks of gastrointestinal illness and other severe health problems, including hemolytic uremic syndrome, worldwide. Monitoring its occurrence in the food and water supply is thus an important public health issue. Currently-used molecular methods used to detect strains of this serotype target multiple virulence genes, and O- and H-antigens. However, as virulence genes can be readily transferred among bacterial strains such markers may be imperfect identifiers. Alternatively, highly conserved genetic markers, which are closely linked to specific strains, may provide direct identification of target pathogenic strains. In this study, using TaqMan-probe-based quantitative PCR (qPCR), we examine a new detection strategy for pathogenic strains of *E. coli* O157:H7 serotype based on a conserved signature insertion/deletion (CSI) located in the *ybiX* gene. The qPCR assay was linear from 1.0×10^2 to 1.0×10^7 genome equivalents and was specific to O157:H7 when tested against a panel of 15 non-O157:H7 *E. coli*. The assay was also maintained detection sensitivity in the presence of competing *E. coli* K-12 and fish sperm DNA spiked in at 1000-fold and 800-fold excess of target DNA, respectively, demonstrating the assay's ability to detect *E. coli* O157:H7 in the presence of high levels of background DNA. This study thus explores the use of strain-specific CSIs as diagnostic markers for pathogen detection, and future work will focus on application of the *E. coli* O157:H7-specific qPCR assay to environmental water and food matrices.

3.2 Introduction

A serotype of clinical importance, enterohaemorrhagic *Escherichia coli* (EHEC) O157:H7 has been the focus of diagnostic detection since the isolation of the *E. coli* O157:H7 strain EDL933 in the United States in 1975, and its association with outbreaks of hemorrhagic colitis in Michigan and Oregon in 1982 (Riley et al., 1983). This facultative anaerobe is found in the gastrointestinal tract of mammals and has reservoirs primarily in cattle (Gannon et al., 2002; Wells et al., 1991), sheep (Kudva et al., 1996) and swine (Feder et al., 2003). Its modes of transmission are varied, with pathogenic loads carried in contaminated beef (Macdonald et al., 2000), leafy green vegetables (Ackers et al., 1998) and drinking water (CCDR, 2000), leading to outbreaks of gastrointestinal illness and, in more severe cases, bloody diarrhea and HUS (Karmali et al., 1985). The presence of *E. coli* O157:H7 in the food and drinking supply is thus a public health concern.

DNA-based tests can be used to detect *E. coli* O157:H7 and other pathogens. The first PCR tests targeted virulence genes *stx* (Gannon et al., 1992) and *eae* (Gannon et al., 1993) in STEC. Since then, multiplex PCR assays have been developed that target genes encoding O- and H-antigens in addition to virulence genes (Hu et al., 1999; Nagano et al., 1998), allowing for the identification of *E. coli* O157:H7 strains in one PCR reaction. The use of fluorogenic TaqMan probes targeting virulence genes in quantitative PCR (qPCR) assays further increased the sensitivity and specificity of *E. coli* O157:H7 detection systems (Sharma and Dean-Nystrom, 2003; Sharma et al., 1999), achieving detection limits of 10 colony forming units (CFU) per gram with enrichment (Sharma and Dean-

Nystrom, 2003). However, the use of enrichment increases assay time. Furthermore, virulence genes can be horizontally transferred among pathogenic *E. coli* strains, potentially confounding results of virulence-gene-based assays used to detect presumptive *E. coli* O157:H7 strains. This was illustrated by the 2011 outbreak of gastroenteritis and HUS in Germany that was caused by an *E. coli* O104:H4 strain that was found to possess virulence properties of both enteroaggregative and Shiga-toxin-producing *E. coli* (Bielaszewska et al., 2011; Frank et al., 2011). Though classified as enteroaggregative *E. coli*, the strain contains a mobile Stx2-phage that enables the production of Shiga toxins and has shown to be closely related to the Stx2-phage from an enterohaemorrhagic *E. coli* O111:H- strain, suggesting a lateral gene transfer event that occurred in the evolutionary history of the *E. coli* O104:H4 outbreak strain (Laing et al., 2012).

Alternative assays that use conserved genetic markers specific to *E. coli* O157:H7 can circumvent lateral gene transfer by directly detecting the presence of target bacteria with non-mobile molecular markers. For example, the +93 mutation in the *uidA* gene that is unique to *E. coli* O157:H7/H- strains renders β-glucuronidase non-functional and serves as a stable genetic marker for qPCR assays (Yoshitomi et al., 2003, 2006). Another qPCR assay for *E. coli* O157:H7 detection uses ORF (open reading frame) Z3276, which encodes a putative fimbrial protein and is unique to *E. coli* O157:H7 (Li and Chen, 2012).

Similarly, conserved signature insertions or deletions (CSIs) are also valuable molecular targets for diagnostic assays. These molecular markers are caused by rare genetic changes that occurred in a common ancestor of the members within a certain taxonomic clade (Gupta, 1998). Clade-specific CSIs are unique to many bacterial

taxonomic groups (Bhandari et al., 2012), and thus enable the elucidation of evolutionary history among bacterial groups. Though CSIs are often found in higher taxonomic groups, a species-specific CSI has been used for the development of a pyrosequencing assay to distinguish *Bacillus anthracis*, the etiological agent of anthrax, from genetically similar members of the *B. cereus* group (Ahmod et al., 2011). Because of their conserved nature, species- or strain-specific CSIs represent stable genetic markers for pathogen detection.

In this study, a TaqMan-probe-based qPCR assay was developed to target an *E. coli* O157:H7-specific CSI located in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily, encoded by the *ybiX* gene. The specificity and dynamic range of the assay, as well as its utility in detecting target *E. coli* O157:H7 in the presence of background DNA, were established with *E. coli* strains and fish sperm DNA. These data indicate that an *E. coli* O157:H7-specific CSI may be used as a diagnostic marker for *E. coli* strains of the O157:H7 serotype, facilitating the direct and sensitive detection of targeted pathogenic *E. coli*. In principle, this strategy can be used for many other pathogens that possess unique CSIs.

3.3 Materials and methods

3.3.1 Identification and confirmation of *E. coli* O157:H7-specific insertion in a conserved signature indel (CSI)

The CSI with the *E. coli* O157:H7-specific insertion was provided and located by Dr. R. S. Gupta using previously published methods (Bhandari and Gupta, 2012; Gao et al., 2009). Briefly, BLASTp searches of all *E. coli* K-12 ORFs were performed against the non-redundant (nr) NCBI Protein database. Protein sequences from different *E. coli* strains resulting from the searches were further aligned with Clustal X version 2.0 (Larkin et al., 2007) and alignments were then visually inspected for conserved flanking regions. Those without conserved flanking regions were not considered usable molecular markers (Gupta, 1998). For proteins with conserved flanking regions, further BLASTp searches using a shorter protein sequence surrounding the CSI as query were done to retrieve a list of organisms with the CSI.

Specificity of the insertion in the CSI found in the *ybiX* gene to *E. coli* O157:H7 strains was then confirmed to be specific with a BLASTn search against *E. coli* strains in the non-redundant Nucleotide collection (nr/nt) NCBI database. A tBLASTn analysis against NCBI database of *E. coli* O157:H7 draft genomes was performed to examine the presence of the insertion in *E. coli* O157:H7 strains. An additional tBLASTn analysis against draft genomes of EHEC strains available in the NCBI database was also performed.

3.3.2 Primer and probe sequences

Primers were designed for a smaller 82 bp amplicon to allow for efficient amplification. The specificity of the primer and the probe sequences to the *ybiX* gene in *E. coli* was checked with a Primer-BLAST analysis against the NCBI Nucleotide collection (nr) database. A BLASTn analysis was also performed using the target amplicon from *E. coli* O157:H7 str. EDL933, and the resulting Expect values (E values) for *E. coli* O157:H7 strains in the NCBI database were 3×10^{-34}. Sequences of the primers and TaqMan probe used in this study are outlined in Table 4.

3.3.3 Bacterial strains

E. coli strains used in this study are listed in Table 5. A total of 16 *E. coli* strains were tested, of which 14 are non-O157:H7 *E. coli* pathogenic strains, one *E. coli* K-12 strain (MG1655) and one *E. coli* O157:H7 strain (EDL933).

3.3.4 Growth conditions, DNA extraction and DNA quantification

Stocks of *E. coli* strains were streaked onto LB (Luria-Bertani) agar plates and grown overnight at 37°C. Single colonies were picked and grown overnight in LB broth at 37°C in a shaking incubator (200 rpm). DNA extractions were performed using a kit (Bacterial Genomic DNA Isolation Kit, cat. no. 17900, Norgen Biotek Corp., Ontario) and the DNA extracts were quantified with the Qubit® dsDNA BR Assay Kit (cat. no. Q32850) and a Qubit® fluorometer (Invitrogen, California).

Table 4. Primers and probe used in the assay and the expected amplicon sizes.

Name	Sequence (5'-3')	Amplicon size (bp)	Target	Reference
O157ybiXF	5CGC CAT GCT GTT TGA ACT GG	82	*E. coli* O157:H7	This study
O157ybiXP	(FAM)-ATT CAG AAT ATT CAG TCG CTG AAA AGC-(BHQ-1)[a]			
O157ybiXR	CAG GAT CTC TTC ATT TTC AC			
ECybiXF	GGC GAA CTG GTC GTT AAT GAC	294/303[b]	All *E. coli*	This study
ECybiXR	TCA GAT CTC CGA CCA TTC CC			

[a] FAM = 6-carboxyfluorescein; BHQ-1 = Black Hole Quencher 1

[b] 294 bp produced by non-O157:H7 *E. coli* strains; 303 bp produced by *E. coli* O157:H7 strains

Table 5. *E. coli* strains used in this study.

Seropathotype[a]	Serotype	Strain	Pathotype[b]	Source[c]
A	O157:H7	EDL933	EHEC	(Riley et al., 1983)
B	O26:H11	CL1	EHEC	(Karmali et al., 2003)
	O26:H11	CL9	EHEC	(Karmali et al., 2003)
	O111:NM	R82F2	EHEC	(Karmali et al., 2003)
	O121:H19	CL106	EHEC	(Karmali et al., 2003)
	O145:NM	N00-6496	EHEC	(Karmali et al., 2003)
	O145:NM	N02-5149	EHEC	(Karmali et al., 2003)
C	O5:NM	N00-4067	EHEC	(Karmali et al., 2003)
	O113:H21	CL3	EHEC	(Karmali et al., 2003)
	O121:NM	N99-4390	EHEC	(Karmali et al., 2003)
D	O103:H25	N00-4859	EHEC	(Karmali et al., 2003)
	O172:NM	EC6-484	EHEC	(Karmali et al., 2003)
E	O84:NM	EC2-044	EHEC	(Karmali et al., 2003)
	O98:H25	EC3-377	EHEC	(Karmali et al., 2003)
----	O6:K2:H1	CFT073	UPEC	(Mobley et al., 1990)
----	OR:H48:K-	MG1655	K-12	CGSC

[a] seropathotypes classified according to Karmali et al. (Karmali et al., 2003)
[b] EHEC = enterohaemorrhagic *E. coli*; UPEC = uropathogenic *E. coli*
[c] CGSC = Coli Genetic Stock Center

3.3.5 Amplifying, sequencing and aligning the region of interest in *E. coli* strains

Control primers ECybiXF and ECybiXR (Table 4) were used to amplify the region of interest located in the *ybiX* gene in *E. coli* strains used in this study (Table 5). PCR was performed in 50 µL reactions, consisting of 5 µL 10X *Pfu* Buffer with $MgSO_4$, 1.25 U *Pfu* DNA polymerase (cat. no. EP0501, Thermo Fisher Scientific, Massachusetts), 0.2 M dNTPs (cat. no. R0192, Thermo Fisher Scientific, Massachusetts), 0.5 µM ECybiXF forward primer, 0.5 µM ECybiXR reverse primer, 10 ng DNA template and double-distilled sterile water (to 50 µL of reaction volume). Experiments were performed on the Mastercycler Gradient (Eppendorf, Germany). The PCR program was as follows: 95°C for 3 min, followed by 30 cycles of 95°C for 30 s, 62°C for 30 s and 72°C for 45 s, and then 72°C for 5 min. Amplicons were then visualized using agarose gel electrophoresis and extracted with a kit (NucleoSpin® Gel and PCR Clean-up, cat. no. 740609.250, Macherey-Nagel, Germany). Sequencing was performed on the 3730 DNA Analyzer (Applied Biosystems, California). Sequencing of PCR products were performed in duplicate using two separate PCR reactions for each strain to account for amplification artefacts. Both strands were sequenced. Sequences of the sequenced region of interest in *E. coli* strains were aligned using Clustal X version 2.0 (Larkin et al., 2007).

3.3.6 Construction of positive and negative control plasmids

For the positive control plasmid (pCR4-O157), a 303 bp gene fragment of *ybiX* containing the *E. coli* O157:H7-specific insertion was cloned into the pCR4™-TOPO® cloning vector using a kit (TOPO® TA Cloning® Kit, cat. no. 450030, Invitrogen,

California). For the negative control plasmid (pCR4-EC), a 294 bp gene fragment of *ybiX* generated from *E. coli* K-12 that does not contain the *E. coli* O157:H7-specific insertion was cloned into the pCR4™-TOPO® cloning vector using the same kit. Both types of plasmids were introduced into the *E. coli* DH5α strain using TSS transformation (Chung et al., 1989) for propagation.

3.3.7 qPCR conditions

For specificity testing (Table 5), 0.2 ng of DNA extract (approximately 4.0×10^4 genome equivalents) was used as template. To establish the range of linearity and limit of detection of the assay, serial 10-fold dilutions of *E. coli* O157:H7 DNA extract were made, from 1.0×10^2 to 1.0×10^7 genome equivalents. The genome equivalents were calculated from the amount of DNA using the following equation (Whelan et al., 2003):

$$\text{Genome equivalent} = \frac{(6.02 \times 10^{23}\ \text{copy/mol}) \times (\text{DNA amount in grams})}{(\text{DNA length in bp}) \times (660\ \text{g/mol/bp})} \qquad [1]$$

qPCR experiments were performed in 10 µL reaction volumes in 96-well plates. The following components were added to reactions that used the TaqMan probe: 5 µL 2X TaqProbe qPCR Mastermix (cat. no. Mastermix-PS, Applied Biological Materials Inc., British Columbia), 1 µM O157ybiXF forward primer, 1 µM O157ybiXR reverse primer, 0.3 µM O157ybiXP TaqMan probe (Integrated DNA Technologies, Iowa), 1 µL DNA template and double-distilled sterile water (to 10 µL of reaction volume). For EvaGreen dye-based external amplification controls, the 10 µL reactions consisted of 5 µL 2X

SsoFast™ EvaGreen® Supermix (cat. no. 172-5201, Bio-Rad, California), 0.2 µM ECybiXF forward primer, 0.2 µM ECybiXR reverse primer, 1 µL DNA template and double-distilled sterile water (to 10 µL of reaction volume). Experiments were performed on the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, California). Samples, positive and negative controls, and external amplification controls were performed in triplicate, while no-template controls were performed in duplicate. The qPCR program included an initial denaturation step at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s (denaturation step) and 62°C for 60 s (annealing/extension step). Amplicons were then visualized using agarose gel electrophoresis.

3.3.8 Competition experiments with *E. coli* K-12 DNA extract

Competing *E. coli* K-12 DNA was spiked into samples with 1.0×10^4 genome equivalents of *E. coli* O157:H7 DNA. Concentrations of *E. coli* K-12 DNA ranged from 0 to 1.0×10^7 genome equivalents, representing 0- to 1000-fold excess amount of target *E. coli* O157:H7 DNA. Experimental controls included plasmid (pCR4-O157 and pCR4-EC), unspiked (*E. coli* O157:H7 DNA only), negative (*E. coli* K-12 DNA only), external amplification and no-template controls. A standard curve was also performed. Reaction conditions were as outlined above under qPCR conditions. Size and specificity of amplification products were verified using agarose gels.

3.3.9 Competition experiments with fish sperm DNA

Fish sperm DNA (cat. no. 11467140001, Roche Applied Science, Germany) was spiked into samples with 1.0 x 10^4 genome equivalents of *E. coli* O157:H7 DNA, and amounts of fish sperm DNA ranged from 0 to 6.2 x 10^1 ng (1000-fold excess of target *E. coli* O157:H7 DNA). Experimental controls and the standard curve were performed as outlined above in spiking experiments with *E. coli* K-12 DNA. Size and specificity of amplification products were verified using agarose gel electrophoresis.

3.3.10 qPCR data analysis

To assess the efficiency of the assay and determine the R^2 value (coefficient of determination), a linear regression analysis was performed, with Cq (quantitative cycle) values plotted as a function of the logarithmic transformation of the starting DNA template quantity (genome equivalents). The signal thresholds used to determine the Cq value were established empirically, at 170 RFU (relative fluorescence units) for TaqMan probe-based reactions, and at 500 RFU for EvaGreen-based reactions. qPCR experiments were performed in accordance with MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines (Bustin et al., 2009). The CFX Manager™ Software (Bio-Rad, California) was used for qPCR data analysis.

3.4 Results

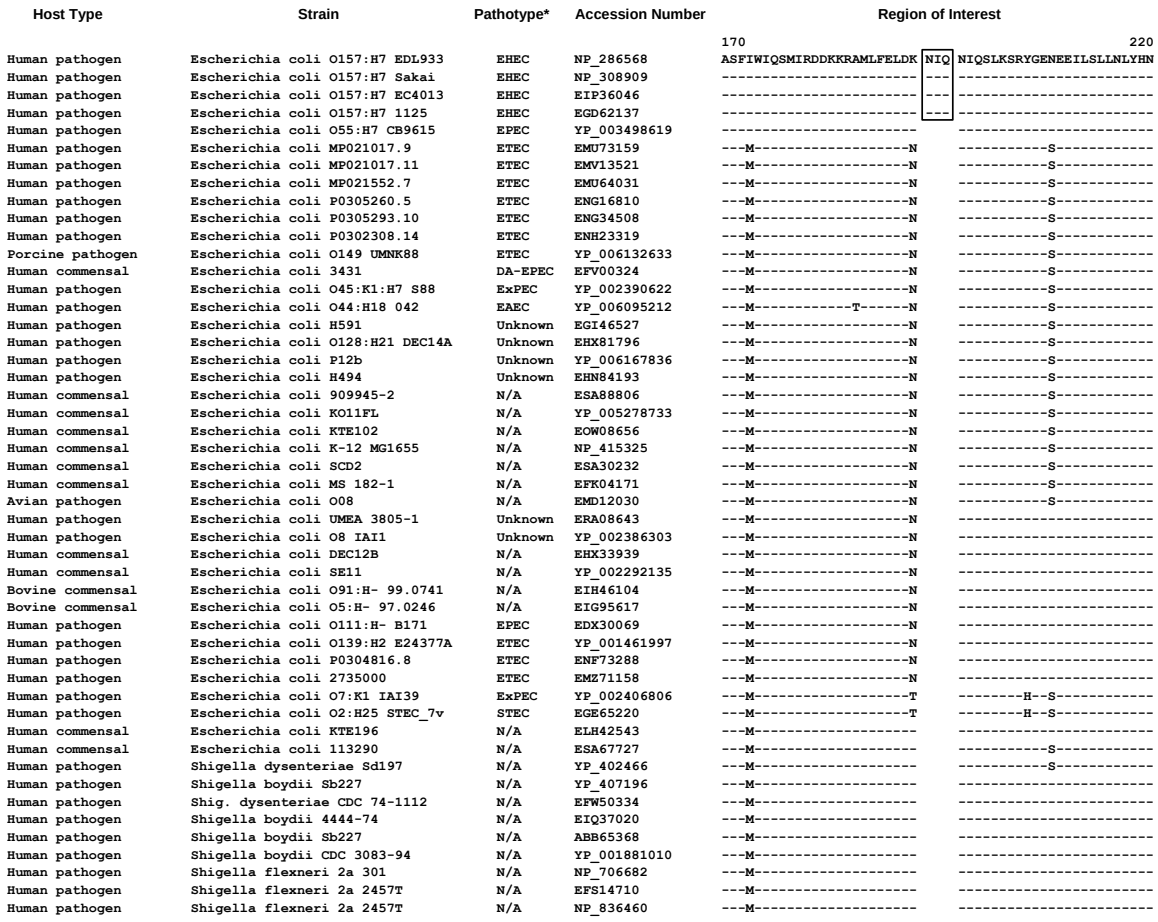
3.4.1 Location and specificity of the CSI in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily, and positions of the primers and TaqMan probe in *ybiX*

The CSI located in the *ybiX* gene product of *E. coli* strains and *Shigella* spp. is 3 amino acids in size, and the insertion was found only in *E. coli* O157:H7 strains (Figure 7). The insertion was present in draft genomes of 97 sequenced *E. coli* O157:H7 isolates (Figure S2) and absent in the draft genomes of non-O157 EHEC strains (Figure S3). The insertion was present in draft genomes of non-motile *E. coli* O157:NM, a Shiga-toxin-producing serotype that has been isolated from patients with hemolytic uremic syndrome (Gunzer et al., 1992).

The primers were designed around the CSI in the *ybiX* gene (Figure 8A). Using these primers, the amplification product sizes produced were 82 bp for *E. coli* O157:H7 and 73 bp for non-O157:H7 *E. coli* (Figure 8B). The 6-carboxyfluorescein-labelled TaqMan probe spans 6 bp of the insertion on the 5' end of the probe (Figure 8A).

3.4.2 Linear range and sensitivity of the qPCR assay

The qPCR assay was linear from 1.0×10^2 to 1.0×10^7 genome equivalents of *E. coli* O157:H7 (Figure 9). The assay had an amplification efficiency of 100.5% and a coefficient of determination of 0.991 (Figure 9). Specific amplification was verified by the presence of the expected 82 bp amplicon on an agarose gel (Figure S4).



Host Type	Strain	Pathotype*	Accession Number	Region of Interest (170–220)
				170 220
Human pathogen	Escherichia coli O157:H7 EDL933	EHEC	NP_286568	ASFIWIQSMIRDDKKRAMLFELDK NIQ NIQSLKSRYGENEEILSLLNLYHN
Human pathogen	Escherichia coli O157:H7 Sakai	EHEC	NP_308909	---------- --- ----------
Human pathogen	Escherichia coli O157:H7 EC4013	EHEC	EIP36046	---------- --- ----------
Human pathogen	Escherichia coli O157:H7 1125	EHEC	EGD62137	---------- --- ----------
Human pathogen	Escherichia coli O55:H7 CB9615	EPEC	YP_003498619	---------- ----------
Human pathogen	Escherichia coli MP021017.9	ETEC	EMU73159	---M----------N ----------S----------
Human pathogen	Escherichia coli MP021017.11	ETEC	EMV13521	---M----------N ----------S----------
Human pathogen	Escherichia coli MP021552.7	ETEC	EMU64031	---M----------N ----------S----------
Human pathogen	Escherichia coli P0305260.5	ETEC	ENG16810	---M----------N ----------S----------
Human pathogen	Escherichia coli P0305293.10	ETEC	ENG34508	---M----------N ----------S----------
Human pathogen	Escherichia coli P0302308.14	ETEC	ENH23319	---M----------N ----------S----------
Porcine pathogen	Escherichia coli O149 UMNK88	ETEC	YP_006132633	---M----------N ----------S----------
Human commensal	Escherichia coli 3431	DA-EPEC	EFV00324	---M----------N ----------S----------
Human pathogen	Escherichia coli O45:K1:H7 S88	ExPEC	YP_002390622	---M----------N ----------S----------
Human pathogen	Escherichia coli O44:H18 042	EAEC	YP_006095212	---M----------T------N ----------S----------
Human pathogen	Escherichia coli H591	Unknown	EGI46527	---M----------N ----------S----------
Human pathogen	Escherichia coli O128:H21 DEC14A	Unknown	EHX81796	---M----------N ----------S----------
Human pathogen	Escherichia coli P12b	Unknown	YP_006167836	---M----------N ----------S----------
Human pathogen	Escherichia coli H494	Unknown	EHN84193	---M----------N ----------S----------
Human commensal	Escherichia coli 909945-2	N/A	ESA88806	---M----------N ----------S----------
Human commensal	Escherichia coli KO11FL	N/A	YP_005278733	---M----------N ----------S----------
Human commensal	Escherichia coli KTE102	N/A	EOW08656	---M----------N ----------S----------
Human commensal	Escherichia coli K-12 MG1655	N/A	NP_415325	---M----------N ----------S----------
Human commensal	Escherichia coli SCD2	N/A	ESA30232	---M----------N ----------S----------
Human commensal	Escherichia coli MS 182-1	N/A	EFK04171	---M----------N ----------S----------
Avian pathogen	Escherichia coli O08	N/A	EMD12030	---M----------N ----------S----------
Human pathogen	Escherichia coli UMEA 3805-1	Unknown	ERA08643	---M----------N ----------
Human pathogen	Escherichia coli O8 IAI1	Unknown	YP_002386303	---M----------N ----------
Human commensal	Escherichia coli DEC12B	N/A	EHX33939	---M----------N ----------
Human commensal	Escherichia coli SE11	N/A	YP_002292135	---M----------N ----------
Bovine commensal	Escherichia coli O91:H- 99.0741	N/A	EIH46104	---M----------N ----------
Bovine commensal	Escherichia coli O5:H- 97.0246	N/A	EIG95617	---M----------N ----------
Human pathogen	Escherichia coli O111:H- B171	EPEC	EDX30069	---M----------N ----------
Human pathogen	Escherichia coli O139:H2 E24377A	ETEC	YP_001461997	---M----------N ----------
Human pathogen	Escherichia coli P0304816.8	ETEC	ENF73288	---M----------N ----------
Human pathogen	Escherichia coli 2735000	ETEC	EMZ71158	---M----------N ----------
Human pathogen	Escherichia coli O7:K1 IAI39	ExPEC	YP_002406806	---M----------T --------H--S----------
Human pathogen	Escherichia coli O2:H25 STEC_7v	STEC	EGE65220	---M----------T --------H--S----------
Human commensal	Escherichia coli KTE196	N/A	ELH42543	---M---------- ----------
Human commensal	Escherichia coli 113290	N/A	ESA67727	---M---------- ----------S----------
Human pathogen	Shigella dysenteriae Sd197	N/A	YP_402466	---M---------- ----------S----------
Human pathogen	Shigella boydii Sb227	N/A	YP_407196	---M---------- ----------
Human pathogen	Shig. dysenteriae CDC 74-1112	N/A	EFW50334	---M---------- ----------
Human pathogen	Shigella boydii 4444-74	N/A	EIQ37020	---M---------- ----------
Human pathogen	Shigella boydii Sb227	N/A	ABB65368	---M---------- ----------
Human pathogen	Shigella boydii CDC 3083-94	N/A	YP_001881010	---M---------- ----------
Human pathogen	Shigella flexneri 2a 301	N/A	NP_706682	---M---------- ----------
Human pathogen	Shigella flexneri 2a 2457T	N/A	EFS14710	---M---------- ----------
Human pathogen	Shigella flexneri 2a 2457T	N/A	NP_836460	---M---------- ----------

Figure 7. Protein alignment showing the 3 amino acid insertion (boxed) located in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily.

Sequences are shown with the species name, followed by the GenBank accession number. The dashes (-) indicate amino acid identity to *E. coli* O157:H7 str. EDL933, and the numbers at the top indicate amino acid positions of the sequence from EDL933. *EHEC = enterohaemorrhagic *E. coli*; EPEC = enteropathogenic *E. coli*; ETEC = enterotoxigenic *E. coli*; DA-EPEC = diffuse adhering enteropathogenic *E. coli*; ExPEC = extraintestinal pathogenic *E. coli*; STEC = Shiga toxigenic *E. coli*; N/A = not applicable.

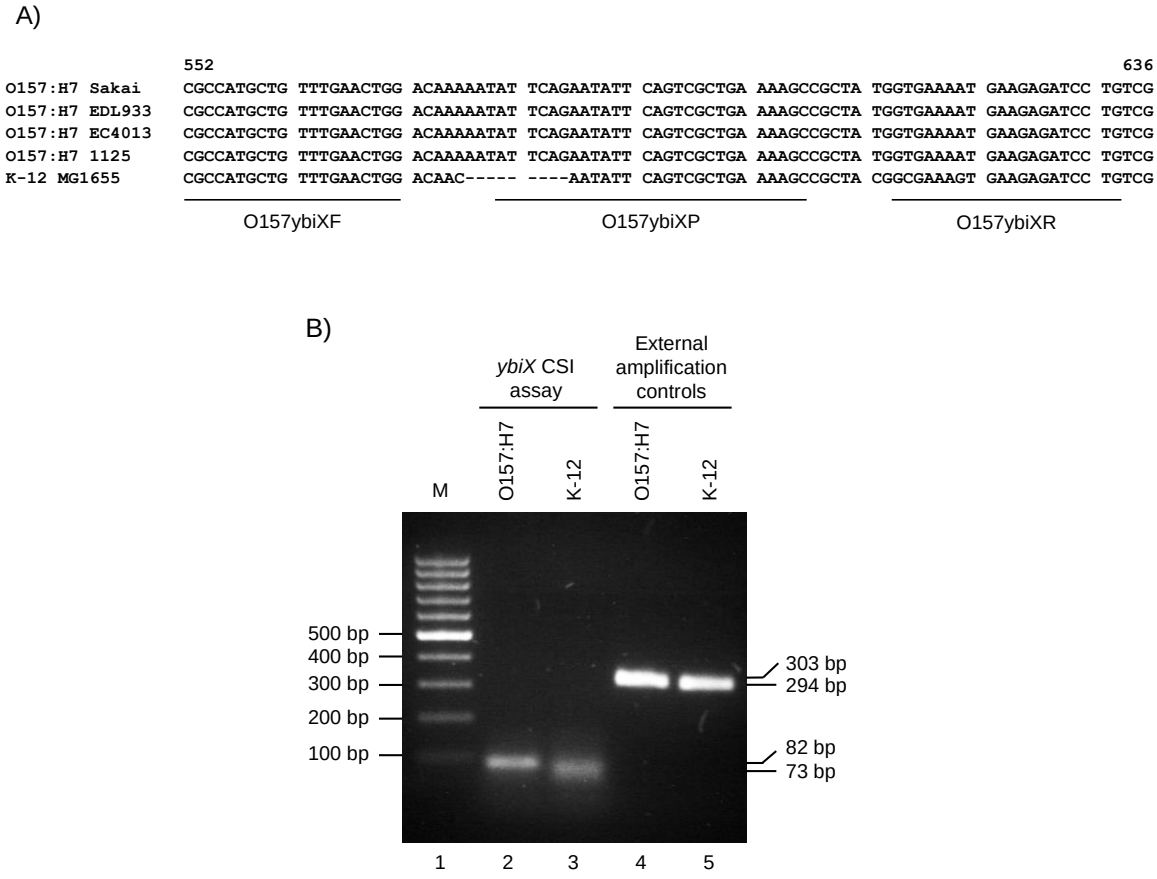


Figure 8. Location of *ybiX* CSI assay primers and TaqMan probe, and expected amplification products for both *ybiX* CSI assay and external amplification controls.

A) Partial nucleotide alignment showing *E. coli* O157:H7-specific insertion located in the *ybiX* gene. *E. coli* O157:H7 strains and a sample strain containing the deletion (*E. coli* K-12) are shown. The dashes (-) indicate a deletion. Positions of the primers and probes in the gene are indicated by the base-pair positions at the top and the labels. B) In samples with *E. coli* O157:H7 DNA, 82 bp amplicons were produced for the *ybiX* CSI assay (lane 2) and 303 bp amplicons were produced for external amplification controls (lane 4). With *E. coli* K-12 DNA, 73 bp amplicons were produced for the *ybiX* CSI assay (lane 3) and 294 bp amplicons were produced for external amplification controls (lane 5). A 100 bp marker (M) was run (lane 1).

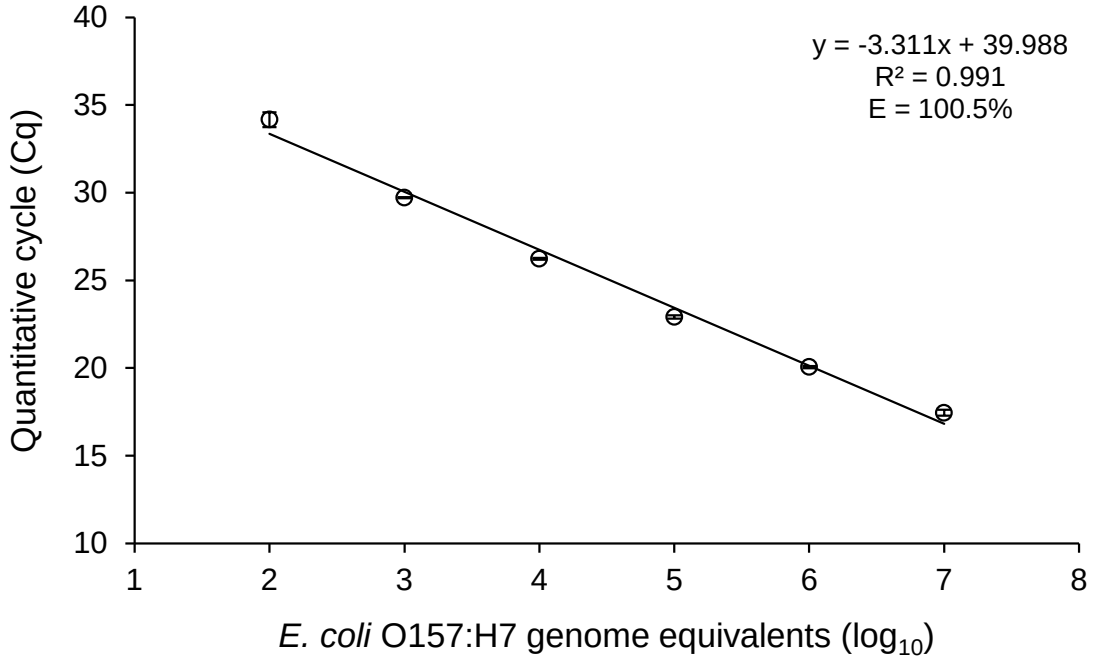


Figure 9. Linear range of *E. coli* O157:H7-specific assay.

A linear regression analysis of quantitative cycle (Cq) as a function of the logarithmic transformation of the starting quantity of *E. coli* O157:H7 is shown. Ten-fold dilutions of *E. coli* O157:H7 genomic DNA extract used were from 1.0 x 10^2 to 1.0 x 10^7 genome equivalents. The efficiency (E) and the coefficient of determination (R^2) are shown. Standard error was calculated for three technical replicates.

3.4.3 Specificity of the assay – *E. coli* strains

To determine the specificity of the assay, 15 non-O157:H7 *E. coli* strains and one *E. coli* O157:H7 strain were used. Only *E. coli* O157:H7 produced amplification signals above threshold (Figure 10), while external amplification controls for all strains produced signals above threshold. Reactions with *E. coli* O157:H7 DNA extract had a Cq value of 24.58 ± 0.17 (Table S1). The expected 82 bp amplicon for *E. coli* O157:H7 was visualized using agarose gel electrophoresis (Figure S5).

The insertion was present only in *E. coli* O157:H7 from an alignment of the sequenced region of interest in the 16 *E. coli* strains used in this study (Figure 11). There were no differences between duplicate sequences, and sequencing of both strands produced complementary sequences (data not shown).

3.4.4 Sensitivity of the assay – competition with *E. coli* K-12 DNA

To demonstrate the assay's ability to specifically amplify target *E. coli* O157:H7 DNA in the presence of genetically similar DNA, competing *E. coli* K-12 DNA was spiked into qPCR reactions with *E. coli* O157:H7 DNA. The assay showed no decrease in sensitivity in the presence of *E. coli* K-12 DNA spiked in reactions at up to a 1000-fold excess amount of target DNA (Figure 12A).

Negative controls with 1.0×10^7 genome equivalents of *E. coli* K-12 produced a signal above threshold for one technical replicate, with a Cq value of 37.82 (Table S2). Positive amplification controls showed amplification signals above threshold (Table S3).

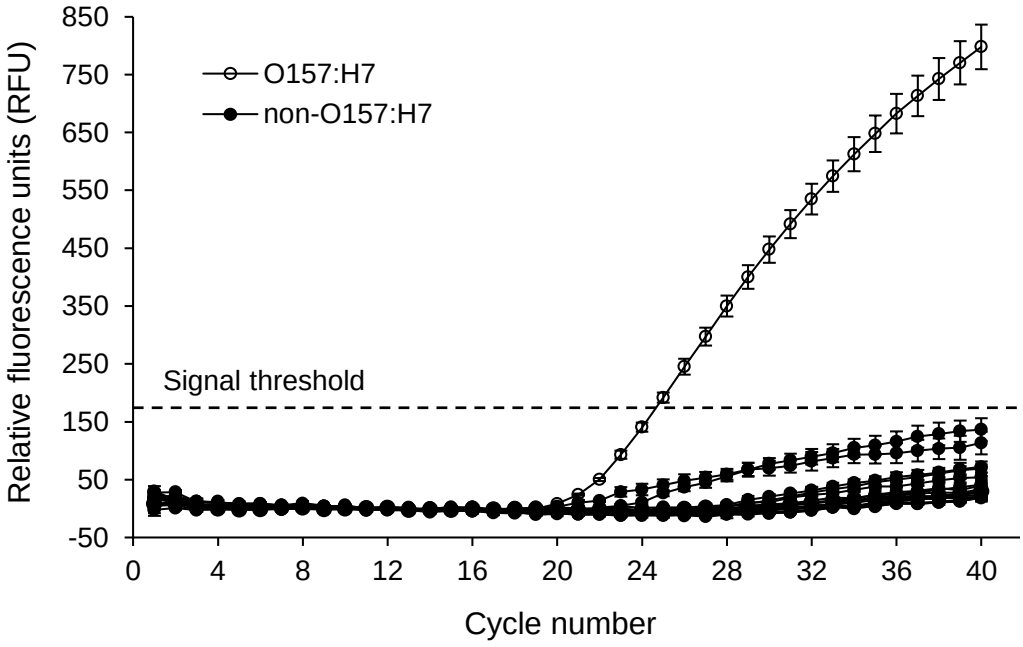


Figure 10. Only *E. coli* O157:H7 produced amplification signals above threshold.

Amplification curves for 15 non-O157:H7 *E. coli* strains and one *E. coli* O157:H7 strain are shown. Standard error was calculated from three technical replicates and shown as error bars. The dotted line across the graph indicates the signal threshold.

Seropathotype	Serotype	Strain	Region of Interest
A	O157:H7	EDL933	TGCTGTTTGAACTGGACAAA AATATTCAG AATATTCAGTCGCTGAAAAG
B	O26:H11	CL1	----------C----------
B	O26:H11	CL9	----------C----------
B	O111:NM	R82F2	----------C----------
B	O121:H19	CL106	----------C----------
B	O145:NM	N00-6496	----------C----------
B	O145:NM	N02-5149	----------C----------
C	O5:NM	N00-4067	----------C----------
C	O113:H21	CL3	----------C----------
C	O121:NM	N99-4390	----------C----------
D	O103:H25	N00-4859	----------C----------
D	O172:NM	EC6-484	----------C----------
E	O84:NM	EC2-044	----------C----------
E	O98:H25	EC3-377	----------C----------
N/A	O6:K2:H1	CFT073	----------C----------
N/A	OR:H48:K-	MG1655	----------C----------

Figure 11. Insertion in ybiX is present only in *E. coli* O157:H7.

Nucleotide alignment of the sequenced region of interest in the *ybiX* gene from 15 non-O157:H7 *E. coli* strains and one *E. coli* O157:H7 strain is shown. Strains are classified into seropathotypes A to E (Karmali et al., 2003). The 9 bp insertion is indicated with a black box.

A standard curve was included and was linear from 1.0×10^3 to 1.0×10^7 genome equivalents (data not shown).

The presence of the expected 82 bp amplification product was confirmed using agarose gel electrophoresis (Figure S6).

3.4.5 Sensitivity of the assay – competition with fish sperm DNA

Fish sperm DNA was used as a source of DNA with low genetic similarity to target *E. coli* O157:H7 DNA to demonstrate the assay's ability to detect *E. coli* O157:H7 in high levels of molecular noise. At a spiking amount of 1000-fold excess of target DNA, the Cq value for the amplification of 1.0×10^4 genome equivalents of *E. coli* O157:H7 was higher than at other spiking amounts (Figure 12B). The assay showed no decrease in sensitivity when fish sperm DNA was spiked into reactions at concentrations up to an 800-fold excess amount of target DNA.

Positive amplification controls showed amplification signals above threshold (Table S3). No positive amplification controls were done for fish sperm DNA, as the control primers were designed for *E. coli* strains.

A standard curve with a linear range of 1.0×10^3 to 1.0×10^7 genome equivalents (data not shown) was included.

Similarly to the competition experiments with spiked *E. coli* K-12 DNA, specificity of amplification products were verified using agarose gel electrophoresis and the expected 82 bp amplicon was seen (Figure S7).

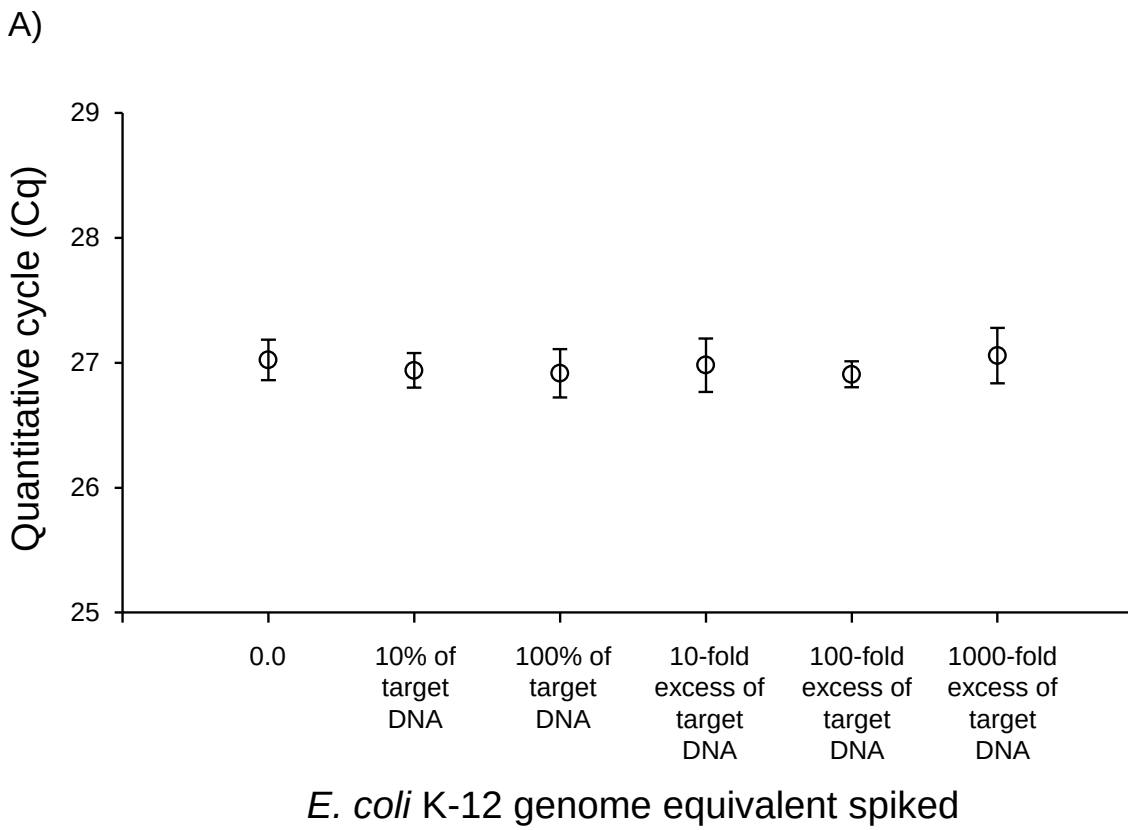


Figure 12. Detection sensitivity of *E. coli* O157:H7 DNA is not affected by competing *E. coli* K-12 DNA at tested spiking amounts and lowered by competing fish sperm DNA spiked in 1000-fold excess of target DNA.

Cq values for the amplification of 1.0 x 10^4 genome equivalents of *E. coli* O157:H7 are plotted as a function of increasing amounts of competing A) *E. coli* K-12 DNA or B) fish sperm DNA that were spiked into reactions. Reactions were performed in triplicate. Standard error values calculated from technical replicates are shown as error bars.

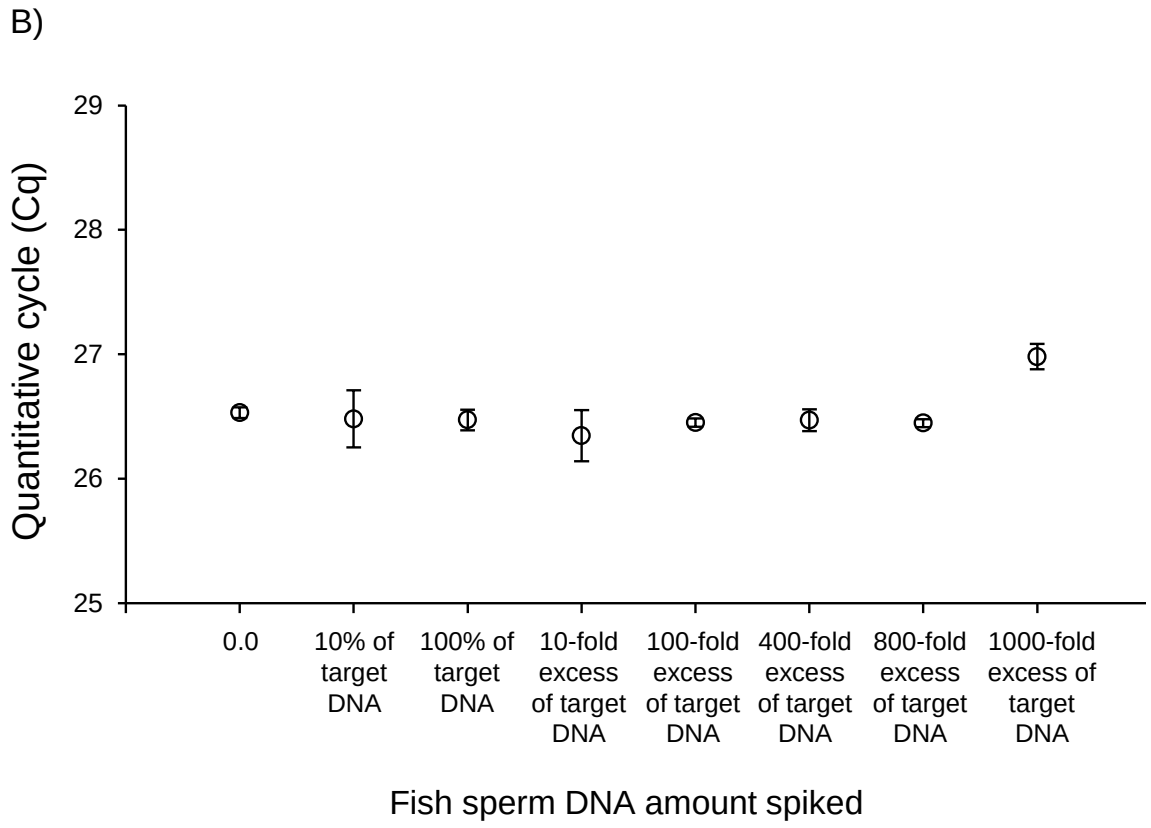


Figure 12. (continued)

Cq values for the amplification of 1.0 x 10^4 genome equivalents of *E. coli* O157:H7 are plotted as a function of increasing amounts of competing A) *E. coli* K-12 DNA or B) fish sperm DNA that were spiked into reactions. Reactions were performed in triplicate. Standard error values calculated from technical replicates are shown as error bars.

3.5 Discussion

As a causative agent of outbreaks of gastrointestinal illness, enterohaemorrhagic (EHEC) *Escherichia coli* O157:H7 is a major focus of molecular assay development to monitor the presence of this pathogen in food and water. Though PCR-based diagnostic assays are very sensitive, many current pathogen detection assays target virulence genes that can be horizontally-transferred to other *E. coli* strains reducing both the specificity and utility of assays. The importance of horizontal transfer was recently exemplified in the 2011 Germany outbreak of gastroenteritis caused by an enteroaggregative *E. coli* O104:H4 strain that had acquired a mobile Stx2-phage (Laing et al., 2012), enabling the production of Shiga toxins that is more characteristic of Shiga-toxigenic *E. coli*. Identification based on more-conserved genetic targets may yield assays that are more reliable than virulence factor-targeted assays. In this study, we explored an alternative qPCR assay that uses a conserved signature indel (CSI) for direct detection of *E. coli* O157:H7 strains. Using competition experiments with *E. coli* K-12 and fish sperm DNA, we demonstrated the effectiveness of this assay for detection of *E. coli* O157:H7 in the presence of large amounts of non-target DNA.

CSIs present at different taxonomic levels provide a reliable means for identifying members of evolutionary clades. These have been used extensively to elucidate evolutionary histories of members within different bacterial groups, such as Gammaproteobacteria (Gao et al., 2009) and Cyanobacteria (Gupta and Mathews, 2010). CSIs at a lower taxonomic level, such as species-specific CSIs, are also useful for the development of diagnostic assays (Ahmod et al., 2011), as they are stable molecular

markers with high resolving power even at the species level. In this study, we focused on a subspecies-specific CSI that distinguishes between serotypes of *E. coli*. This CSI is a 3-amino-acid indel found in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily, with an insertion found specifically in *E. coli* O157:H7 strains (Figure 7).

The TaqMan-probe-based qPCR assay developed with this genetic marker offers a specific and sensitive method for directly identifying *E. coli* O157:H7 strains, rather than inferring their presence with virulence markers that are also encoded in different strains of enterohaemorrhagic *E. coli*. Through specificity tests using *E. coli* strains, the assay was able to distinguish between targeted *E. coli* O157:H7 and non-target *E. coli* using a small 9 base-pair difference of the CSI, similar to the O157:H7-specific assay developed against the single base mismatch at the +93 position of the *uidA* gene (Yoshitomi et al., 2003). The assay also had a detection limit of 1.0×10^2 genome equivalents of *E. coli* O157:H7 (Figure 9), within the upper limits of the minimum infectious dose of 100 colony forming units (Paton and Paton, 1998b). Though the detection limit of the assay has yet to be determined in environmental water samples, these results provisionally demonstrate the utility of this CSI for *E. coli* O157:H7 detection.

External factors affect the limit of detection for pathogen detection assays when applied to environmental samples and food matrices. One factor is the use of a growth enrichment step, which increases the number of target sequences in the sample and thus improves the limit of detection of the assay. Such enrichment steps have been used in different sample types to detect *E. coli* O157:H7. An 18-hour enrichment step can lead to a 5-6 log increase of stressed *E. coli* O157:H7 spiked into environmental water, lowering

detection limit to 4 CFU/100 mL (Sen et al., 2011). As well, an 8-hour enrichment period, combined with propidium monoazide, which enters injured or dead cells with compromised membranes to intercalate into DNA and inhibit PCR amplification, allowed for the detection of 80 CFU of viable *E. coli* O157:H7 per gram of beef (Li and Chen, 2012). Another factor that may reduce a given assay's limit of detection is the presence of inhibitors, such as organic compounds. These inhibitors can be removed by activated charcoal coated with bentonite (Lee and Levin, 2011) or *Pseudomonas fluorescens* (Abolmaaty et al., 2007), enabling a lowered detection limit from 1.0×10^3 CFU/g to 5 CFU/g in lettuce spiked with *E. coli* O157:H7 (Lee and Levin, 2011), and an-over-4-log decrease in the limit of detection to 1.0×10^3 CFU/g in spiked oyster samples (Abolmaaty et al., 2007). Thus, a theoretical limit of detection in complex samples for the *E. coli* O157:H7-specific assay shown in this study can be established by taking into consideration the influences of enrichment and removal of PCR inhibitors, and may be different from a limit of detection of 1.0×10^2 genomes demonstrated in this study.

Though the utility of assays are often framed in the context of extraction efficiencies of protocols (Mull and Hill, 2009; Sanchez et al., 2012), the elimination of organic inhibitors and the addition of enrichment periods, other confounding factors may play a role in influencing an assay's ability to sensitively detect target pathogens, such as background genetic material co-extracted with samples. This aspect was examined in a PCR assay targeting the gene encoding the O-antigen of O157:H7, *rfbE* (Timmons et al., 2013). The test was assessed by using produce washes containing 10^6-10^7 CFU of bacteria spiked with 10-fold dilutions of *E. coli* O157:H7 ranging from 10 to 10^9

CFU/mL, and the authors concluded that background flora had no effect on detection sensitivity (Timmons et al., 2013). However, similar studies have not yet been done for *E. coli* O157:H7-specific qPCR assays. To further explore the use of the indels as diagnostics, the *ybiX* assay's detection sensitivity to *E. coli* O157:H7 in the presence of background DNA was examined. *E. coli* K-12 DNA spiked into samples with target *E. coli* O157:H7 DNA provided a measure of the assay's performance in samples with genetically similar, non-target template that could potentially lead to non-specific binding of the TaqMan probe. The detection sensitivity of the assay was not affected at *E. coli* K-12 spiking concentrations of up to 1000-fold excess amount of target DNA, potentiating its use in monitoring environmental water samples to trace potential agricultural run-off or sewage contamination. Bodies of freshwater around watersheds are seeded by *E. coli* communities sustained in the soils, with persistent *E. coli* that can survive for months-long periods (Byappanahalli et al., 2006; Whitman et al., 2006). The long-term persistence of indigenous *E. coli* leads to a constant background microbial presence in water samples collected at these sites and, depending on the bacterial load, may disrupt detection of targeted pathogenic strains. To further examine the influence of background DNA, fish sperm DNA spiked into reactions with *E. coli* O157:H7 DNA also created a reaction environment that contained high levels of heterologous DNA that is complex in nature, thus serving as a proxy for environmental DNA. The assay's detection of target *E. coli* O157:H7 DNA was unaffected by excess DNA (up to an 800-fold excess). For detection of *E. coli* O157:H7 in environmental samples, the assay's effectiveness is dependent on specific detection of target DNA among background DNA. Water sample

types from urban settings, such as storm water run-off, contain a multitude of microorganisms, including *Enterococcus*, *Campylobacter*, *Salmonella* and human-specific adenovirus (Sidhu et al., 2012). Thus, the assay's ability to detect *E. coli* O157:H7 in the presence of different types of excess non-target DNA has demonstrated its potential use on samples from environmental waters.

In this study, we developed a qPCR assay that directly detects *E. coli* O157:H7 using a novel CSI-based strategy. The ability of the assay to amplify target *E. coli* O157:H7 DNA in the presence of excess competing background DNA and the assay's utility in different simulated sample types was also explored. Though the applicability of this assay to environmental water samples and different food matrices has yet to be assessed in future studies, this proof-of-principle assay represents a possible use for a conserved *E. coli* O157:H7-specific insertion in a CSI that enables direct identification of a clinically relevant pathogenic *E. coli* strain capable of causing severe human morbidity and mortality. In principle, the basic strategy employed in this study can be further used to develop CSI-based assays for other important human pathogens, including *Campylobacter* and *Enterococcus*, as well as toxigenic bacteria, such as those found in the Cyanobacteria phylum. Clade-specific indels have been identified for these important water microorganisms and CSI-based assays can be readily developed that will help to identify these organisms in water and wastewater samples.

References

Abolmaaty, A., Gu, W., Witkowsky, R., and Levin, R.E. (2007). The use of activated charcoal for the removal of PCR inhibitors from oyster samples. Journal of microbiological methods *68*, 349-352.

Ackers, M.L., Mahon, B.E., Leahy, E., Goode, B., Damrow, T., Hayes, P.S., Bibb, W.F., Rice, D.H., Barrett, T.J., Hutwagner, L., *et al.* (1998). An outbreak of *Escherichia coli* O157:H7 infections associated with leaf lettuce consumption. The Journal of infectious diseases *177*, 1588-1593.

Ahmod, N.Z., Gupta, R.S., and Shah, H.N. (2011). Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. Journal of microbiological methods *87*, 278-285.

Almanza, A.V. (2011). Shiga Toxin-Producing *Escherichia coli* in Certain Raw Beef Products U.S.D.o. Agriculture, ed. (Washington), pp. 58157-58165.

Bai, J., Paddock, Z.D., Shi, X., Li, S., An, B., and Nagaraja, T.G. (2012). Applicability of a multiplex PCR to detect the seven major Shiga toxin-producing *Escherichia coli* based on genes that code for serogroup-specific O-antigens and major virulence factors in cattle feces. Foodborne Pathog Dis *9*, 541-548.

Bettelheim, K.A. (1998a). Reliability of CHROMagar O157 for the detection of enterohaemorrhagic *Escherichia coli* (EHEC) O157 but not EHEC belonging to other serogroups. J Appl Microbiol *85*, 425-428.

Bettelheim, K.A. (1998b). Studies of *Escherichia coli* cultured on Rainbow Agar O157 with particular reference to enterohaemorrhagic *Escherichia coli* (EHEC). Microbiol Immunol *42*, 265-269.

Bhandari, V., and Gupta, R.S. (2012). Molecular signatures for the phylum Synergistetes and some of its subclades. Antonie van Leeuwenhoek *102*, 517-540.

Bhandari, V., Naushad, H.S., and Gupta, R.S. (2012). Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution. Frontiers in cellular and infection microbiology *2*, 98.

Bielaszewska, M., Mellmann, A., Zhang, W., Kock, R., Fruth, A., Bauwens, A., Peters, G., and Karch, H. (2011). Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. The Lancet infectious diseases *11*, 671-676.

BioControl (2012). Assurance GDS MPX for Top 6 or 7 STEC.

Blanco, M., Blanco, J.E., Mora, A., Rey, J., Alonso, J.M., Hermoso, M., Hermoso, J., Alonso, M.P., Dahbi, G., Gonzalez, E.A., *et al.* (2003). Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from healthy sheep in Spain. J Clin Microbiol *41*, 1351-1356.

Boerlin, P., McEwen, S.A., Boerlin-Petzold, F., Wilson, J.B., Johnson, R.P., and Gyles, C.L. (1999). Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. J Clin Microbiol *37*, 497-503.

Bosilevac, J.M., Kalchayanand, N., Schmidt, J.W., Shackelford, S.D., Wheeler, T.L., and Koohmaraie, M. (2010). Inoculation of beef with low concentrations of *Escherichia*

coli O157:H7 and examination of factors that interfere with its detection by culture isolation and rapid methods. J Food Prot *73*, 2180-2188.

Brooks, J.T., Sowers, E.G., Wells, J.G., Greene, K.D., Griffin, P.M., Hoekstra, R.M., and Strockbine, N.A. (2005). Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. The Journal of infectious diseases *192*, 1422-1429.

Burland, V., Shao, Y., Perna, N.T., Plunkett, G., Sofia, H.J., and Blattner, F.R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Nucleic Acids Res *26*, 4196-4204.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., *et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry *55*, 611-622.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Sadowsky, M.J., and Ishii, S. (2006). Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. Environmental microbiology *8*, 504-513.

CCDR (2000). Waterborne outbreak of gastroenteritis associated with a contaminated municipal water supply, Walkerton, Ontario, May-June 2000. Canada communicable disease report = Releve des maladies transmissibles au Canada *26*, 170-173.

CDC (2006 Dec 14). Multistate Outbreak of *E. coli* O157 Infections, November-December 2006.

CDC (2007). *Escherichia coli* O157:H7 infection associated with drinking raw milk--Washington and Oregon, November-December 2005. MMWR Morb Mortal Wkly Rep *56*, 165-167.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences of the United States of America *86*, 2172-2175.

Church, D.L., Emshey, D., Semeniuk, H., Lloyd, T., and Pitout, J.D. (2007). Evaluation of BBL CHROMagar O157 versus sorbitol-MacConkey medium for routine detection of *Escherichia coli* O157 in a centralized regional clinical microbiology laboratory. J Clin Microbiol *45*, 3098-3100.

Clotilde, L.M., Bernard, C.t., Salvador, A., Lin, A., Lauzon, C.R., Muldoon, M., Xu, Y., Lindpaintner, K., and Carter, J.M. (2013). A 7-plex microbead-based immunoassay for serotyping Shiga toxin-producing *Escherichia coli*. Journal of microbiological methods *92*, 226-230.

Currie, A., NacDonald, J., Ellis, A., Siushansian, J., Chui, L., Charlebois, M., Peermohamed, M., Everett, D., Fehr, M., and Ng, L.K. (2007). Outbreak of *Escherichia coli* O157:H7 infections associated with consumption of beef donair. J Food Protect *70*, 1483-1488.

De Schrijver, K., Buvens, G., Posse, B., Van den Branden, D., Oosterlynck, O., De Zutter, L., Eilers, K., Pierard, D., Dierick, K., Van Damme-Lombaerts, R., *et al.* (2008). Outbreak of verocytotoxin-producing *E. coli* O145 and O26 infections associated with the consumption of ice cream produced at a farm, Belgium, 2007. Euro Surveill *13*.

DeVinney, R., Puente, J.L., Gauthier, A., Goosney, D., and Finlay, B.B. (2001). Enterohaemorrhagic and enteropathogenic *Escherichia coli* use a different Tir-based mechanism for pedestal formation. Mol Microbiol *41*, 1445-1458.

DeVinney, R., Stein, M., Reinscheid, D., Abe, A., Ruschkowski, S., and Finlay, B.B. (1999). Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. Infect Immun *67*, 2389-2398.

Donnenberg, M.S., Tzipori, S., McKee, M.L., O'Brien, A.D., Alroy, J., and Kaper, J.B. (1993). The role of the *eae* gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. The Journal of clinical investigation *92*, 1418-1424.

Dontorou, A., Papadopoulou, C., Filioussis, G., Apostolou, I., Economou, V., Kansouzidou, A., and Levidiotou, S. (2004). Isolation of a rare *Escherichia coli* O157:H7 strain from farm animals in Greece. Comp Immunol Microbiol Infect Dis *27*, 201-207.

Duhamel, G.E., Moxley, R.A., Maddox, C.W., and Erickson, E.D. (1992). Enteric infection of a goat with enterohemorrhagic *Escherichia coli* (O103:H2). J Vet Diagn Invest *4*, 197-200.

Dunn, J.R., Keen, J.E., and Thompson, R.A. (2004). Prevalence of Shiga-toxigenic *Escherichia coli* O157:H7 in adult dairy cattle. J Am Vet Med Assoc *224*, 1151-1158.

DuPont (2012). BAX System PCR Assay *E. coli* O157:H7 MP.

Eklund, M., Scheutz, F., and Siitonen, A. (2001). Clinical isolates of non-O157 Shiga toxin-producing *Escherichia coli*: serotypes, virulence characteristics, and molecular profiles of strains of the same serotype. J Clin Microbiol *39*, 2829-2834.

Elmufti, M., and Gagnon, S. (2012). Laboratory requirements for testing of non O157:H7 Shiga Toxin-Producing *Escherichia coli* (STEC) in raw beef trimmings for export to the United States of America, C.F.I. Agency, ed.

Eriksson, E., Nerbrink, E., Borch, E., Aspan, A., and Gunnarsson, A. (2003). Verocytotoxin-producing *Escherichia coli* O157:H7 in the Swedish pig population. Vet Rec *152*, 712-717.

Ethelberg, S., Smith, B., Torpdahl, M., Lisby, M., Boel, J., Jensen, T., Nielsen, E.M., and Molbak, K. (2009). Outbreak of non-O157 Shiga toxin-producing *Escherichia coli* infection from consumption of beef sausage. Clin Infect Dis *48*, e78-81.

Feder, I., Wallace, F.M., Gray, J.T., Fratamico, P., Fedorka-Cray, P.J., Pearce, R.A., Call, J.E., Perrine, R., and Luchansky, J.B. (2003). Isolation of *Escherichia coli* O157:H7 from intact colon fecal samples of swine. Emerging infectious diseases *9*, 380-383.

Feldsine, P.T., Green, S.T., Lienau, A.H., Stephens, J., Jucker, M.T., and Kerr, D.E. (2005). Evaluation of the assurance GDS for *E. coli* O157:H7 method and assurance GDS for shigatoxin genes method in selected foods: collaborative study. J AOAC Int *88*, 1334-1348.

Frank, C., Werber, D., Cramer, J.P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., *et al.* (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. N Engl J Med *365*, 1771-1780.

Fratamico, P.M., Bagi, L.K., Cray, W.C., Jr., Narang, N., Yan, X., Medina, M., and Liu, Y. (2011). Detection by multiplex real-time polymerase chain reaction assays and isolation of Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis *8*, 601-607.

FSIS (1999). FSIS Policy on Non-intact Raw Beef Products Contaminated with *E. coli* O157:H7, U.S.D.o. Agriculture, ed. (Washington).

FSIS (2010). FSIS Procedure for the Use of *Escherichia coli* O157:H7 Screening Tests.

Fukushima, H., and Seki, R. (2004). High numbers of Shiga toxin-producing *Escherichia coli* found in bovine faeces collected at slaughter in Japan. FEMS Microbiol Lett *238*, 189-197.

Gannon, V.P., D'Souza, S., Graham, T., King, R.K., Rahn, K., and Read, S. (1997). Use of the flagellar H7 gene as a target in multiplex PCR assays and improved specificity in identification of enterohemorrhagic *Escherichia coli* strains. J Clin Microbiol *35*, 656-662.

Gannon, V.P., Graham, T.A., King, R., Michel, P., Read, S., Ziebell, K., and Johnson, R.P. (2002). *Escherichia coli* O157:H7 infection in cows and calves in a beef cattle herd in Alberta, Canada. Epidemiol Infect *129,* 163-172.

Gannon, V.P., King, R.K., Kim, J.Y., and Thomas, E.J. (1992). Rapid and sensitive method for detection of Shiga-like toxin-producing *Escherichia coli* in ground beef using the polymerase chain reaction. Appl Environ Microbiol *58*, 3809-3815.

Gannon, V.P., Rashed, M., King, R.K., and Thomas, E.J. (1993). Detection and characterization of the *eae* gene of Shiga-like toxin-producing *Escherichia coli* using polymerase chain reaction. J Clin Microbiol *31,* 1268-1274.

Gao, B., Mohan, R., and Gupta, R.S. (2009). Phylogenomics and protein signatures elucidating the evolutionary relationships among the Gammaproteobacteria. International journal of systematic and evolutionary microbiology *59*, 234-247.

Garcia-Sanchez, A., Sanchez, S., Rubio, R., Pereira, G., Alonso, J.M., Hermoso de Mendoza, J., and Rey, J. (2007). Presence of Shiga toxin-producing *E. coli* O157:H7 in a survey of wild artiodactyls. Vet Microbiol *121*, 373-377.

Gunzer, F., Bohm, H., Russmann, H., Bitzan, M., Aleksic, S., and Karch, H. (1992). Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. J Clin Microbiol *30*, 1807-1810.

Gupta, R.S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiology and molecular biology reviews : MMBR *62*, 1435-1491.

Gupta, R.S., and Mathews, D.W. (2010). Signature proteins for the major clades of Cyanobacteria. BMC evolutionary biology *10*, 24.

Guttikonda, S., Tang, X.L., Yang, B.M., Armstrong, G.D., and Suresh, M.R. (2007). Monospecific and bispecific antibodies against *E. coli* O157 for diagnostics. J Immunol Methods *327*, 1-9.

He, Y., Keen, J.E., Westerman, R.B., Littledike, E.T., and Kwang, J. (1996). Monoclonal antibodies for detection of the H7 antigen of *Escherichia coli*. Appl Environ Microbiol *62*, 3325-3332.

Hegde, N.V., Praul, C., Gehring, A., Fratamico, P., and Debroy, C. (2013). Rapid O serogroup identification of the six clinically relevant Shiga toxin-producing *Escherichia coli* by antibody microarray. Journal of microbiological methods *93*, 273-276.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996). Real time quantitative PCR. Genome research *6*, 986-994.

Higgins, J.A., Nasarabadi, S., Karns, J.S., Shelton, D.R., Cooper, M., Gbakima, A., and Koopman, R.P. (2003). A handheld real time thermal cycler for bacterial pathogen detection. Biosensors & bioelectronics *18*, 1115-1123.

Hilborn, E.D., Mermin, J.H., Mshar, P.A., Hadler, J.L., Voetsch, A., Wojtkunski, C., Swartz, M., Mshar, R., Lambert-Fair, M.A., Farrar, J.A., *et al.* (1999). A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. Arch Intern Med *159*, 1758-1764.

Hirvonen, J.J., Siitonen, A., and Kaukoranta, S.S. (2012). Usability and performance of CHROMagar STEC medium in detection of Shiga toxin-producing *Escherichia coli* strains. J Clin Microbiol *50*, 3586-3590.

Hsu, C.F., Tsai, T.Y., and Pan, T.M. (2005). Use of the duplex TaqMan PCR system for detection of Shiga-like toxin-producing *Escherichia coli* O157. J Clin Microbiol *43*, 2668-2673.

Hu, Y., Zhang, Q., and Meitzler, J.C. (1999). Rapid and sensitive detection of *Escherichia coli* O157:H7 in bovine faeces by a multiplex PCR. J Appl Microbiol *87*, 867-876.

Jokinen, C.C., Edge, T.A., Koning, W., Laing, C.R., Lapen, D.R., Miller, J., Mutschall, S., Scott, A., Taboada, E.N., Thomas, J.E., *et al.* (2012). Spatial and temporal drivers of zoonotic pathogen contamination of an agricultural watershed. J Environ Qual *41*, 242-252.

Kanitpun, R., Wagner, G.G., and Waghela, S.D. (2004). Characterization of recombinant antibodies developed for capturing enterohemorrhagic *Escherichia coli* O157:H7. Southeast Asian J Trop Med Public Health *35*, 902-912.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K., and Kaper, J.B. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J Clin Microbiol *41*, 4930-4940.

Karmali, M.A., Petric, M., Lim, C., Fleming, P.C., Arbus, G.S., and Lior, H. (1985). The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. The Journal of infectious diseases *151,* 775-782.

Keen, J.E., Wittum, T.E., Dunn, J.R., Bono, J.L., and Durso, L.M. (2006). Shiga-toxigenic *Escherichia coli* O157 in agricultural fair livestock, United States. Emerging infectious diseases *12,* 780-786.

Keene, W.E., Hedberg, K., Herriott, D.E., Hancock, D.D., McKay, R.W., Barrett, T.J., and Fleming, D.W. (1997). A prolonged outbreak of *Escherichia coli* O157:H7 infections caused by commercially distributed raw milk. The Journal of infectious diseases *176,* 815-818.

Kudva, I.T., Hatfield, P.G., and Hovde, C.J. (1996). *Escherichia coli* O157:H7 in microbial flora of sheep. J Clin Microbiol *34,* 431-433.

Laing, C.R., Zhang, Y., Gilmour, M.W., Allen, V., Johnson, R., Thomas, J.E., and Gannon, V.P. (2012). A comparison of Shiga-toxin 2 bacteriophage from classical enterohemorrhagic *Escherichia coli* serotypes and the German *E. coli* O104:H4 outbreak strain. PLoS One *7,* e37362.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23,* 2947-2948.

Lee, J.E., Kim, J.S., Choi, I.H., Tagawa, M., Kohsaka, T., and Jin, D.K. (2002). Cytokine expression in the renal tubular epithelial cells stimulated by Shiga toxin 2 of *Escherichia coli* O157:H7. Renal failure *24*, 567-575.

Lee, J.L., and Levin, R.E. (2011). Detection of 5 CFU/g of *Escherichia coli* O157:H7 on lettuce using activated charcoal and real-time PCR without enrichment. Food Microbiol *28*, 562-567.

Li, B., and Chen, J.Q. (2012). Real-time PCR methodology for selective detection of viable *Escherichia coli* O157:H7 cells by targeting Z3276 as a genetic marker. Appl Environ Microbiol *78*, 5297-5304.

Lin, A., Sultan, O., Lau, H.K., Wong, E., Hartman, G., and Lauzon, C.R. (2011). O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol *28*, 478-483.

Macdonald, C., Drew, J., Carlson, R., Dzogan, S., Tataryn, S., Macdonald, A., Ali, A., Amhed, R., Easy, R., Clark, C., *et al.* (2000). Outbreak of *Escherichia coli* O157:H7 leading to the recall of retail ground beef--Winnipeg, Manitoba, May 1999. Canada communicable disease report = Releve des maladies transmissibles au Canada *26*, 109-111.

March, S.B., and Ratnam, S. (1986). Sorbitol-MacConkey medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis. J Clin Microbiol *23*, 869-872.

Mechie, S.C., Chapman, P.A., and Siddons, C.A. (1997). A fifteen month study of *Escherichia coli* O157:H7 in a dairy herd. Epidemiol Infect *118*, 17-25.

Mobley, H.L., Green, D.M., Trifillis, A.L., Johnson, D.E., Chippendale, G.R., Lockatell, C.V., Jones, B.D., and Warren, J.W. (1990). Pyelonephritogenic Escherichia coli and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. Infect Immun *58*, 1281-1289.

Mora, A., Lopez, C., Dhabi, G., Lopez-Beceiro, A.M., Fidalgo, L.E., Diaz, E.A., Martinez-Carrasco, C., Mamani, R., Herrera, A., Blanco, J.E., *et al.* (2012). Seropathotypes, Phylogroups, Stx subtypes, and intimin types of wildlife-carried, shiga toxin-producing *Escherichia coli* strains with the same characteristics as human-pathogenic isolates. Appl Environ Microbiol *78*, 2578-2585.

Mull, B., and Hill, V.R. (2009). Recovery and detection of *Escherichia coli* O157:H7 in surface water, using ultrafiltration and real-time PCR. Appl Environ Microbiol *75*, 3593-3597.

Nagano, I., Kunishima, M., Itoh, Y., Wu, Z., and Takahashi, Y. (1998). Detection of verotoxin-producing *Escherichia coli* O157:H7 by multiplex polymerase chain reaction. Microbiol Immunol *42*, 371-376.

Ojo, O.E., Ajuwape, A.T., Otesile, E.B., Owoade, A.A., Oyekunle, M.A., and Adetosoye, A.I. (2010). Potentially zoonotic shiga toxin-producing *Escherichia coli* serogroups in the faeces and meat of food-producing animals in Ibadan, Nigeria. Int J Food Microbiol *142*, 214-221.

Oporto, B., Esteban, J.I., Aduriz, G., Juste, R.A., and 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 Public Health *55*, 73-81.

Paddock, Z., Shi, X., Bai, J., and Nagaraja, T.G. (2012). Applicability of a multiplex PCR to detect O26, O45, O103, O111, O121, O145, and O157 serogroups of *Escherichia coli* in cattle feces. Vet Microbiol *156*, 381-388.

Park, S.H., Hanning, I., Jarquin, R., Moore, P., Donoghue, D.J., Donoghue, A.M., and Ricke, S.C. (2011). Multiplex PCR assay for the detection and quantification of *Campylobacter* spp., *Escherichia coli* O157:H7, and *Salmonella* serotypes in water samples. FEMS Microbiol Lett *316*, 7-15.

Paton, A.W., and Paton, J.C. (1998a). Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfbO111*, and *rfbO157*. J Clin Microbiol *36*, 598-602.

Paton, A.W., Ratcliff, R.M., Doyle, R.M., Seymour-Murray, J., Davos, D., Lanser, J.A., and Paton, J.C. (1996). Molecular microbiological investigation of an outbreak of hemolytic-uremic syndrome caused by dry fermented sausage contaminated with Shiga-like toxin-producing *Escherichia coli*. J Clin Microbiol *34*, 1622-1627.

Paton, J.C., and Paton, A.W. (1998b). Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. Clinical microbiology reviews *11*, 450-479.

Perna, N.T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., *et al.* (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature *409*, 529-533.

PHAC (1997). Hospital Outbreak of *Escherichia coli* O157:H7 Associated with a Rare Phage Type - Ontario.

Porstmann, T., and Kiessig, S.T. (1992). Enzyme immunoassay techniques. An overview. J Immunol Methods *150*, 5-21.

Posse, B., De Zutter, L., Heyndrickx, M., and Herman, L. (2008). Novel differential and confirmation plating media for Shiga toxin-producing *Escherichia coli* serotypes O26, O103, O111, O145 and sorbitol-positive and -negative O157. FEMS Microbiol Lett *282*, 124-131.

Rangel, J.M., Sparling, P.H., Crowe, C., Griffin, P.M., and Swerdlow, D.L. (2005). Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerging infectious diseases *11*, 603-609.

Reinstein, S., Fox, J.T., Shi, X., Alam, M.J., Renter, D.G., and Nagaraja, T.G. (2009). Prevalence of *Escherichia coli* O157:H7 in organically and naturally raised beef cattle. Appl Environ Microbiol *75*, 5421-5423.

Renter, D.G., Sargeant, J.M., Hygnstorm, S.E., Hoffman, J.D., and Gillespie, J.R. (2001). *Escherichia coli* O157:H7 in free-ranging deer in Nebraska. J Wildl Dis *37*, 755-760.

Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., *et al.* (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med *308*, 681-685.

Rios, M., Prado, V., Trucksis, M., Arellano, C., Borie, C., Alexandre, M., Fica, A., and Levine, M.M. (1999). Clonal diversity of Chilean isolates of enterohemorrhagic *Escherichia coli* from patients with hemolytic-uremic syndrome, asymptomatic subjects, animal reservoirs, and food products. J Clin Microbiol *37*, 778-781.

Rounds, J.M., Rigdon, C.E., Muhl, L.J., Forstner, M., Danzeisen, G.T., Koziol, B.S., Taylor, C., Shaw, B.T., Short, G.L., and Smith, K.E. (2012). Non-O157 Shiga toxin-producing *Escherichia coli* associated with venison. Emerging infectious diseases *18*, 279-282.

Sanchez, G., Elizaquivel, P., and Aznar, R. (2012). A single method for recovery and concentration of enteric viruses and bacteria from fresh-cut vegetables. Int J Food Microbiol *152*, 9-13.

Schimmer, B., Nygard, K., Eriksen, H.M., Lassen, J., Lindstedt, B.A., Brandal, L.T., Kapperud, G., and Aavitsland, P. (2008). Outbreak of haemolytic uraemic syndrome in Norway caused by *stx2*-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. BMC Infect Dis *8*, 41.

Sen, K., J, L.S., Boczek, L., and Rice, E.W. (2011). Development of a sensitive detection method for stressed *E. coli* O157:H7 in source and finished drinking water by culture-qPCR. Environmental science & technology *45*, 2250-2256.

Sharma, V.K., and Dean-Nystrom, E.A. (2003). Detection of enterohemorrhagic *Escherichia coli* O157:H7 by using a multiplex real-time PCR assay for genes encoding intimin and Shiga toxins. Vet Microbiol *93*, 247-260.

Sharma, V.K., Dean-Nystrom, E.A., and Casey, T.A. (1999). Semi-automated fluorogenic PCR assays (TaqMan) forrapid detection of *Escherichia coli* O157:H7 and other shiga toxigenic *E. coli*. Mol Cell Probes *13*, 291-302.

Sidhu, J.P., Hodgers, L., Ahmed, W., Chong, M.N., and Toze, S. (2012). Prevalence of human pathogens and indicators in stormwater runoff in Brisbane, Australia. Water research *46*, 6652-6660.

Swerdlow, D.L., Woodruff, B.A., Brady, R.C., Griffin, P.M., Tippen, S., Donnell, H.D., Jr., Geldreich, E., Payne, B.J., Meyer, A., Jr., Wells, J.G., *et al.* (1992). A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. Ann Intern Med *117*, 812-819.

Tatsuno, I., Kimura, H., Okutani, A., Kanamaru, K., Abe, H., Nagai, S., Makino, K., Shinagawa, H., Yoshida, M., Sato, K., *et al.* (2000). Isolation and characterization of mini-Tn*5*Km2 insertion mutants of enterohemorrhagic *Escherichia coli* O157:H7 deficient in adherence to Caco-2 cells. Infect Immun *68*, 5943-5952.

Timmons, C., Dobhal, S., Fletcher, J., and Ma, L.M. (2013). Primers with 5' flaps improve the efficiency and sensitivity of multiplex PCR assays for the detection of Salmonella and *Escherichia coli* O157:H7. J Food Prot *76*, 668-673.

Wells, J.E., and Varel, V.H. (2008). Viability of zoonotic pathogens *Escherichia coli* and *Salmonella* in swine manure slurries with and without a urease inhibitor and thymol. Lett Appl Microbiol *46*, 477-482.

Wells, J.G., Shipman, L.D., Greene, K.D., Sowers, E.G., Green, J.H., Cameron, D.N., Downes, F.P., Martin, M.L., Griffin, P.M., Ostroff, S.M., *et al.* (1991). Isolation of *Escherichia coli* serotype O157:H7 and other Shiga-like-toxin-producing *E. coli* from dairy cattle. J Clin Microbiol *29*, 985-989.

Whelan, J.A., Russell, N.B., and Whelan, M.A. (2003). A method for the absolute quantification of cDNA using real-time PCR. J Immunol Methods *278*, 261-269.

Whitman, R.L., Nevers, M.B., and Byappanahalli, M.N. (2006). Examination of the watershed-wide distribution of *Escherichia coli* along Southern Lake Michigan: an integrated approach. Appl Environ Microbiol *72*, 7301-7310.

Yang, X., Cheng, H.W., Chen, L., Zhao, J., Chang, H.T., Wang, X.W., Liu, H.Y., Yao, H.X., Zhang, L.X., and Wang, C.Q. (2013). A duplex SYBR Green I real-time quantitative PCR assay for detecting *Escherichia coli* O157:H7. Genetics and molecular research : GMR *12*, 4836-4845.

Yatsuyanagi, J., Saito, S., and Ito, I. (2002). A case of hemolytic-uremic syndrome associated with shiga toxin 2-producing *Escherichia coli* O121 infection caused by drinking water contaminated with bovine feces. Jpn J Infect Dis *55*, 174-176.

Yoshitomi, K.J., Jinneman, K.C., and Weagant, S.D. (2003). Optimization of a 3'-minor groove binder-DNA probe targeting the *uidA* gene for rapid identification of *Escherichia coli* O157:H7 using real-time PCR. Mol Cell Probes *17*, 275-280.

Yoshitomi, K.J., Jinneman, K.C., and Weagant, S.D. (2006). Detection of Shiga toxin genes *stx1*, *stx2*, and the +93 *uidA* mutation of *E. coli* O157:H7/H-using SYBR Green I in a real-time multiplex PCR. Mol Cell Probes *20*, 31-41.

Zadik, P.M., Chapman, P.A., and Siddons, C.A. (1993). Use of tellurite for the selection of verocytotoxigenic *Escherichia coli* O157. Journal of Medical Microbiology *39*, 155-158.

Zhang, X., Cheng, Y., Xiong, Y., Ye, C., Zheng, H., Sun, H., Zhao, H., Ren, Z., and Xu, J. (2012). Enterohemorrhagic *Escherichia coli* specific enterohemolysin induced IL-1beta in human macrophages and EHEC-induced IL-1beta required activation of NLRP3 inflammasome. PLoS One *7*, e50288.

Zweifel, C., Blanco, J.E., Blanco, M., Blanco, J., and Stephan, R. (2004). Serotypes and virulence genes of ovine non-O157 Shiga toxin-producing *Escherichia coli* in Switzerland. Int J Food Microbiol *95*, 19-27.

Appendix A: Supplementary figures and tables

A.1 Chapter 2 – supplementary information

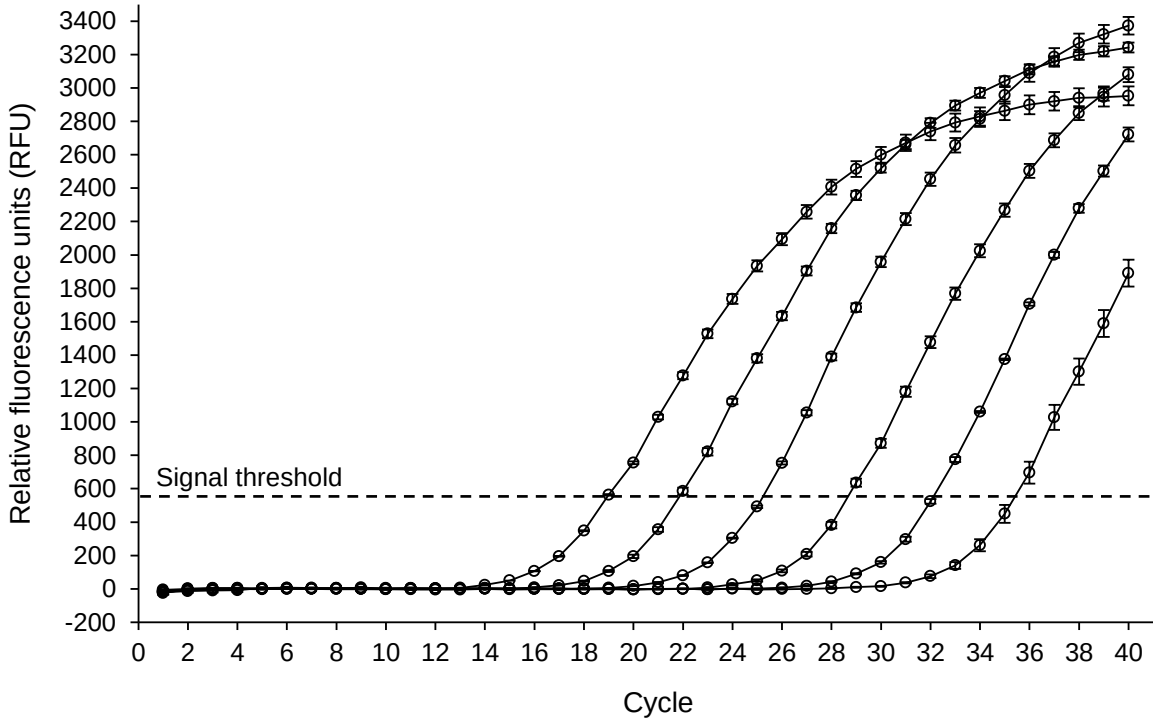


Figure S1. Standard curve of EvaGreen-based indel assay – amplification curves.

The 10-fold dilution curve ranged from 16 to 1.6 x 10^6 genome equivalents. Standard error was calculated using three technical replicates and shown as error bars. The signal threshold is indicated by the dotted line.

A.2 Chapter 3 – supplementary information

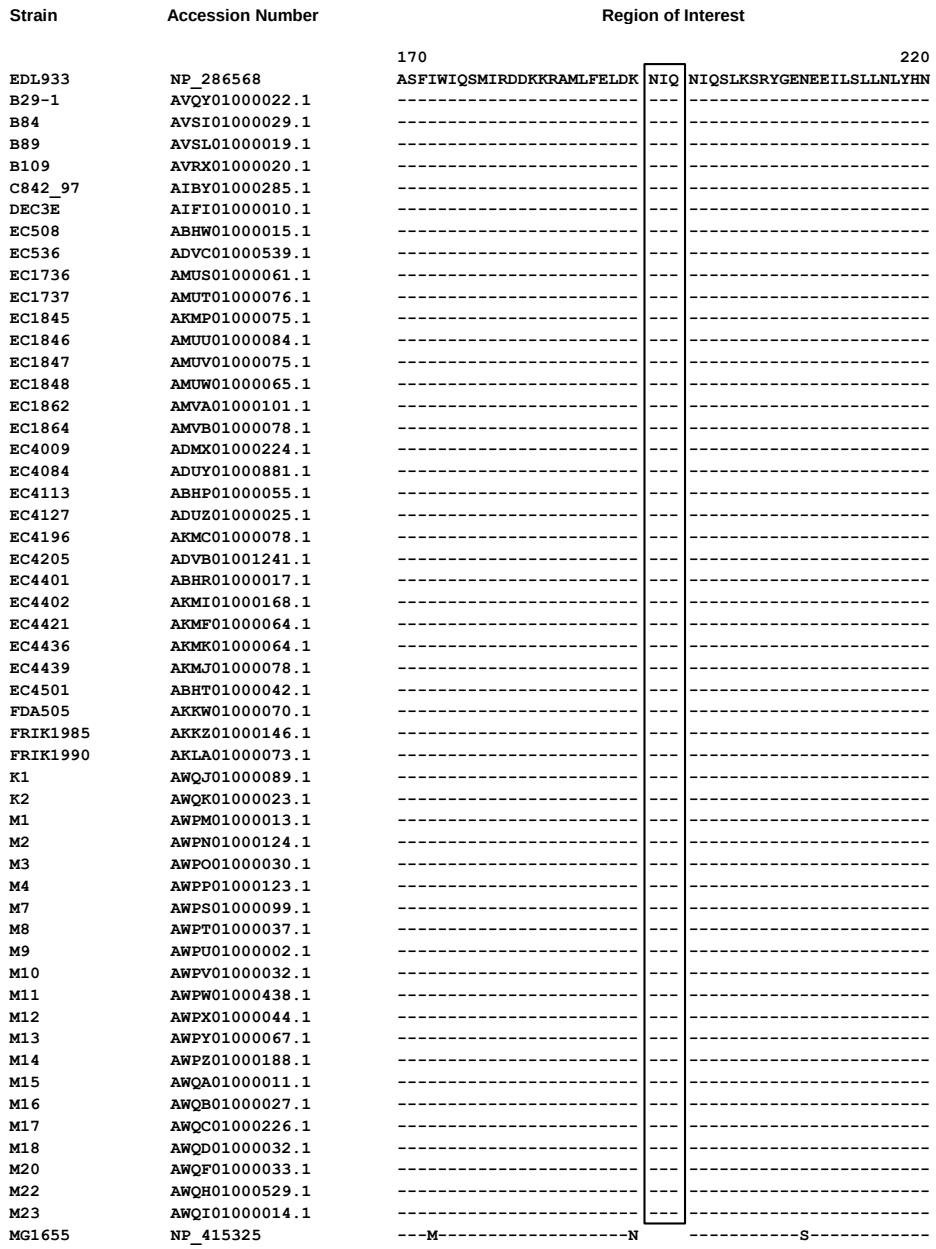


Figure S2. Translated protein alignment showing the insertion present in draft genomes of *E. coli* O157:H7 strains.

Strain names are listed on the left, followed by the accession number. All strains, except for MG1655, have an O157:H7 serotype. The 3 amino acid insertion in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily is indicated by a black box. The dashes (-) indicate amino acid identity to *E. coli* O157:H7 reference strain EDL933, and the numbers at the top indicate amino acid positions for EDL933.

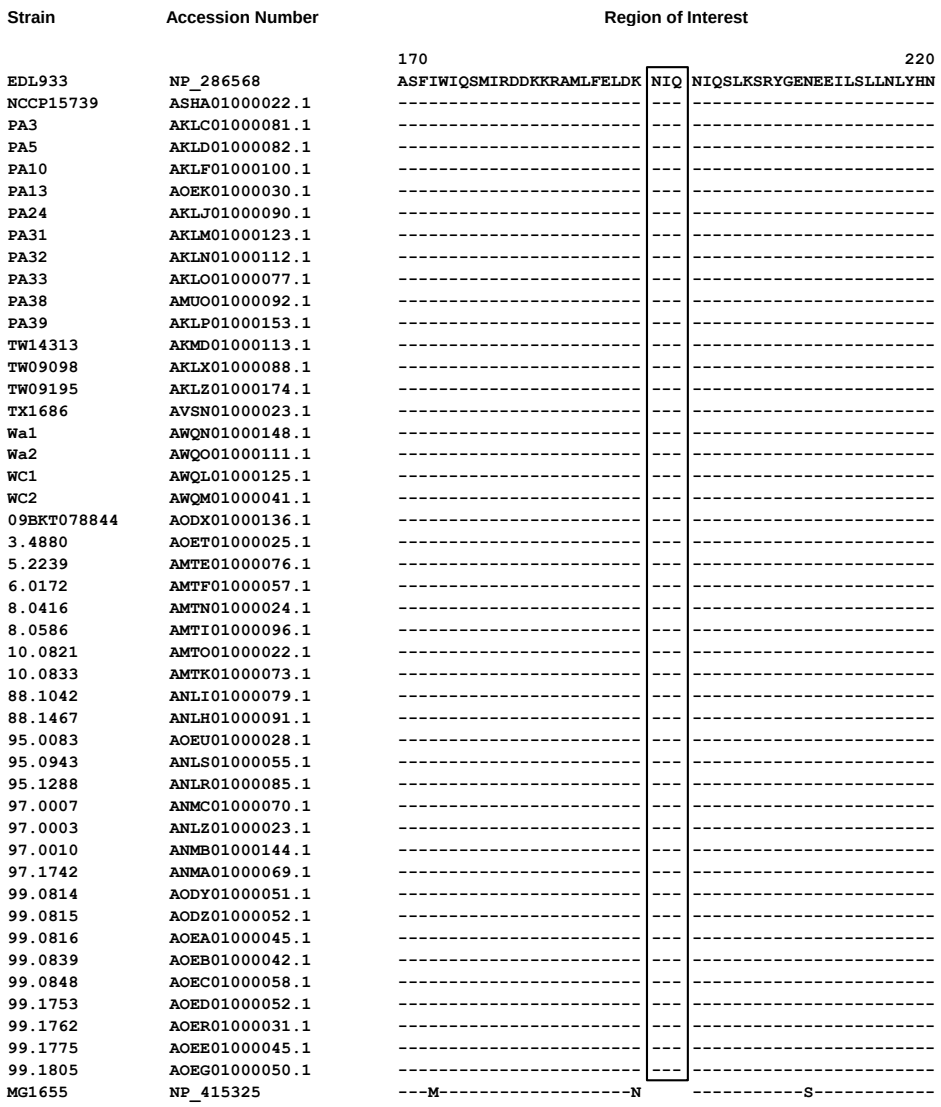


Figure S2. (continued)

Strain names are listed on the left, followed by the accession number. All strains, except for MG1655, have an O157:H7 serotype. The 3 amino acid insertion in the 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily is indicated by a black box. The dashes (-) indicate amino acid identity to *E. coli* O157:H7 reference strain EDL933, and the numbers at the top indicate amino acid positions for EDL933.

Seropathotype	Serotype	Strain	Accession Number	Region of Interest	
				170	220
A	O157:H7	EDL933	NP_286568	ASFIWIQSMIRDDKKRAMLFELDK	NIQ NIQSLKSRYGENEEILSLLNLYHN
A	O157:NM	493-89	AETY01000156.1	------------------------	--- ------------------------
A	O157:NM	H 2687	AETZ01000161.1	------------------------	--- ------------------------
B	O103:H2	CVM9450	AJVR01000224.1	---M------------------N	------------------------
B	O103:H2	E22	AAJV02000017.1	---M------------------N	------------------------
B	O26:H11	CVM10030	AKBA01000447.1	---M------------------N	----------S------------
B	O26:H11	CVM9942	AJVW01000549.1	---M------------------N	----------S------------
B	O26:H11	CVM10224	AKBB01000324.1	---M------------------N	----------S------------
B	O26:H11	CFSAN001629	AMXO01000070.1	---M------------------N	----------S------------
B	O26:H11	CVM10026	AJVX01000468.1	---M------------------N	----------S------------
B	O26:H11	CVM10021	AKAZ01000455.1	---M------------------N	----------S------------
B	O26:H11	CVM9952	AKBC01000425.1	---M------------------N	----------S------------
B	O121:H19	MT#2	AGTJ01000090.1	---M------------------N	----------S------------
C	O91:H21	B2F1	AGTI01000034.1	---M------------------N	----------S------------
C	O113:H21	CL-3	AGTH01000003.1	---M------------------N	----------S------------
D	O103:H25	NIPH-11060424	AGSG01000066.1	---M------------------N	----------S------------
D	O103:H25	CVM9340	AJVQ01000252.1	---M------------------N	----------S------------

Figure S3. Translated protein alignment showing the CSI in draft genomes of EHEC strains.

Available draft genomes for EHEC serotypes listed in (Karmali et al., 2003) were used in this alignment. Strain names are listed on the left, followed by the accession number. Strains are classified into different seropathotypes (Karmali et al., 2003). The 3 amino acid insertion located in 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily is indicated by a black box. The dashes (-) indicate amino acid identity to *E. coli* O157:H7 reference strain EDL933, and numbers at the top indicate the amino acid positions of the sequence.

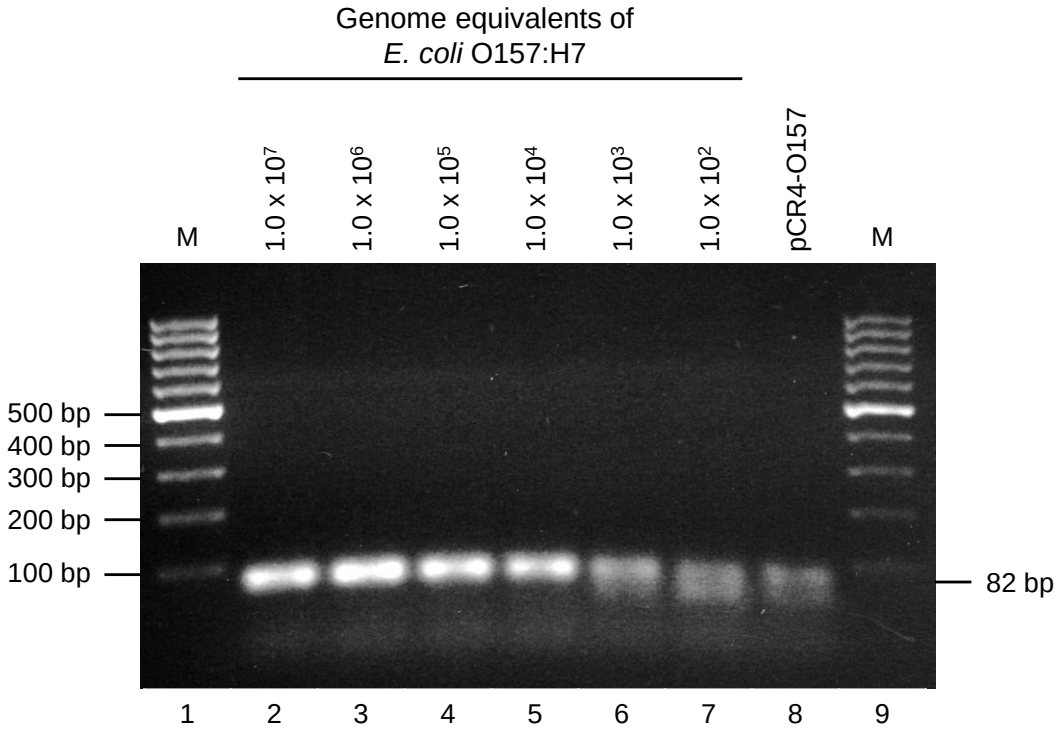


Figure S4. Specific amplification of 10-fold dilutions of *E. coli* O157:H7 DNA extract.

Six 10-fold dilutions of the standard curve, representing 1.0×10^7 to 1.0×10^2 genome copies (lanes 2-7), and the positive control plasmid, pCR4-O157 (lane 8), are shown. The expected 82 bp amplicon of the *E. coli* O157:H7-specific TaqMan assay is seen. A 100 bp marker (M) was run in lanes 1 and 9.

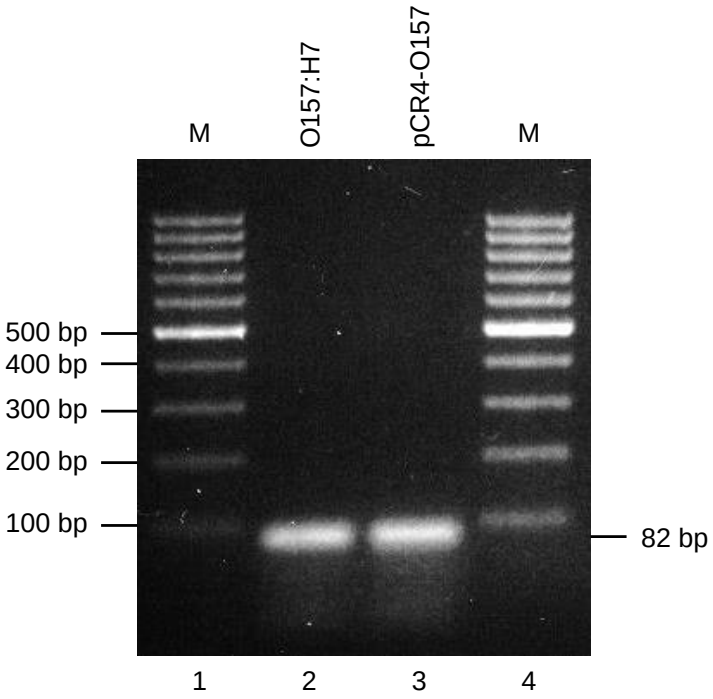


Figure S5. Specific amplification of *E. coli* O157:H7 DNA extract in specificity tests using *E. coli* strains.

The expected 82 bp amplicon of the *E. coli* O157:H7-specific TaqMan assay is seen for *E. coli* O157:H7 (lane 2) and the positive control plasmid, pCR4-O157 (lane 3). A 100 bp marker (M) was run in lanes 1 and 4.

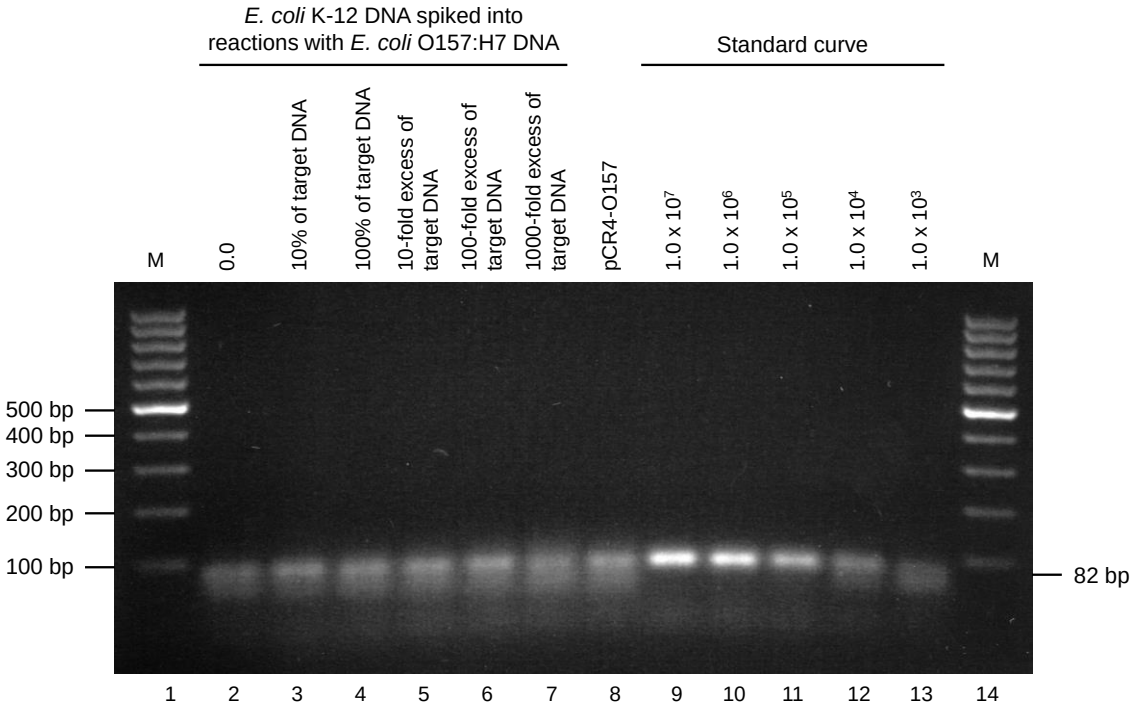


Figure S6. Specific amplification of *E. coli* O157:H7 DNA in spiking experiments with *E. coli* K-12.

Amplification products of reactions with increasing amounts of *E. coli* K-12, from 0 to 1000-fold excess of target DNA, that were spiked into 1.0 x 10^4 genome equivalents of *E. coli* O157:H7 (lanes 2-7), the positive control plasmid, pCR4-O157 (lane 8) and standard curve dilutions (lanes 9-13) are shown. The expected 82 bp amplicon of the *E. coli* O157:H7-specific TaqMan assay is seen. A 100 bp marker (M) was run in lanes 1 and 14.

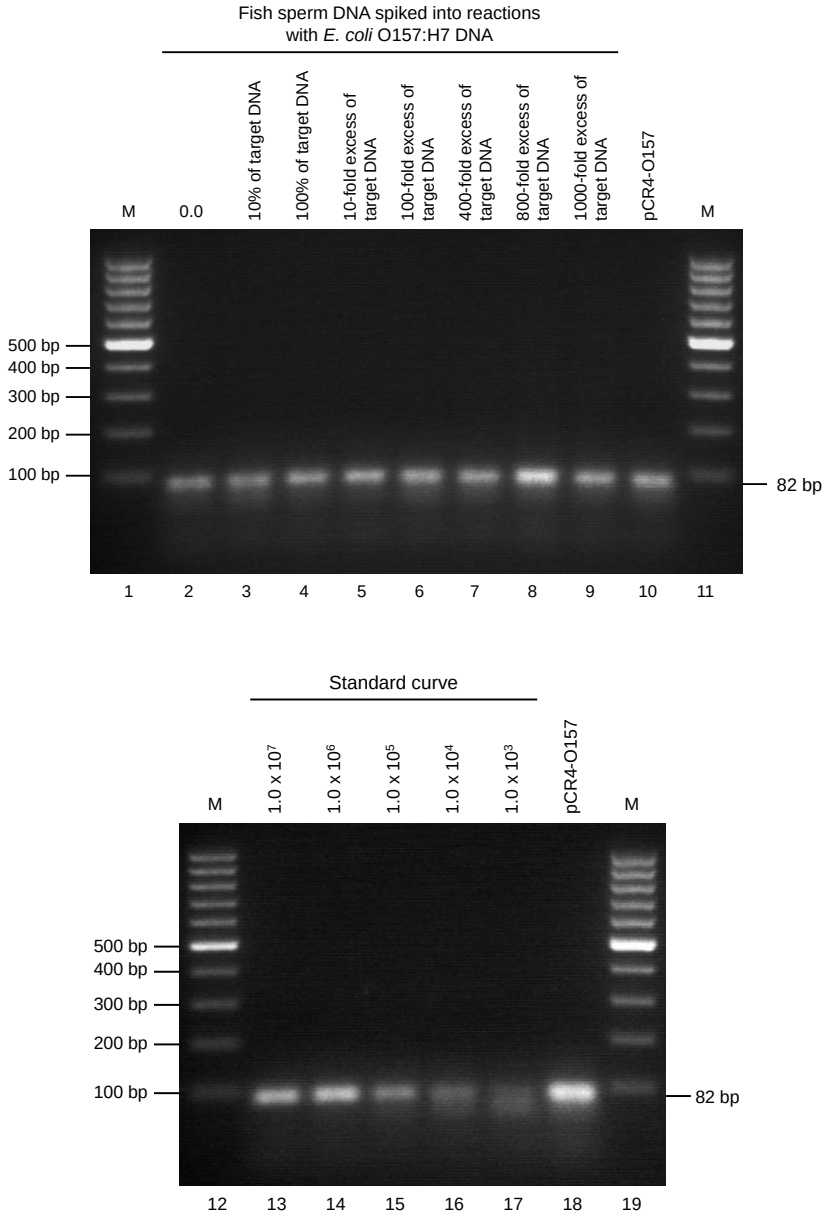


Figure S7. Specific amplification of *E. coli* O157:H7 DNA in spiking experiments with fish sperm DNA.

Amplification products from reactions with increasing amounts of fish sperm DNA, from 0 to 1000-fold excess of target DNA, spiked into 1.0 x 10^4 genome equivalents of *E. coli* O157:H7 (lanes 2-9), the positive control plasmid, pCR4-O157 (lanes 10 and 18) and standard curve dilutions (lanes 13-17) are shown. The expected 82 bp amplicon of the *E. coli* O157:H7-specific TaqMan assay is seen. A 100 bp marker (M) was run in lanes 1, 11, 12 and 19.

Table S1. Cq values of *E. coli* strains used in specificity testing.

Seropathotype[a]	Serotype	Strain	Pathotype[b]	Cq[c]
A	O157:H7	EDL933	EHEC	24.58 ± 0.17[d]
B	O26:H11	CL1	EHEC	>40
	O26:H11	CL9	EHEC	>40
	O111:NM	R82F2	EHEC	>40
	O121:H19	CL106	EHEC	>40
	O145:NM	N00-6496	EHEC	>40
	O145:NM	N02-5149	EHEC	>40
C	O5:NM	N00-4067	EHEC	>40
	O113:H21	CL3	EHEC	>40
	O121:NM	N99-4390	EHEC	>40
D	O103:H25	N00-4859	EHEC	>40
	O172:NM	EC6-484	EHEC	>40
E	O84:NM	EC2-044	EHEC	>40
	O98:H25	EC3-377	EHEC	>40
----	O6:K2:H1	CFT073	UPEC	>40
----	OR:H48:K-	MG1655	K-12	>40

[a] seropathotypes classified according to Karmali et al. (Karmali et al., 2003)
[b] EHEC = enterohaemorrhagic *E. coli*; UPEC = uropathogenic *E. coli*
[c] Cq = quantitative cycle
[d] standard error calculated from three technical replicates

Table S2. Mean Cq values for *E. coli* K-12 (negative controls).

E. coli K-12 (genome equivalents)	Mean Cq[a] (quantitative cycle)
1.0×10^3	ND
1.0×10^4	ND
1.0×10^5	ND
1.0×10^6	ND
1.0×10^7	37.82[b]

[a] ND = not detected; no amplification signal above threshold
[b] one technical replicate with amplification signal above threshold

Table S3. Mean Cq values for positive amplification controls.

Template	Mean Cq[a] (quantitative cycle)	
	E. coli K-12 DNA spiking	Fish sperm DNA spiking
E. coli O157:H7 DNA extract	23.20 ± 0.05	23.25 ± 0.04
E. coli K-12 DNA extract	22.81 ± 0.06	N/A[c]
pCR4-O157[b]	21.97 ± 0.07	21.85 ± 0.08
pCR4-EC[b]	22.69 ± 0.23	21.76 ± 0.20

[a] standard error calculated from three technical replicates
[b] pCR4-O157 = positive control plasmid; pCR4-EC = negative control plasmid
[c] N/A = not applicable

Appendix B: Standard operating procedures (SOPs)

B.1 Genomic DNA extraction from bacterial cultures and DNA quantification

1. Streak *E. coli* bacteria from stocks stored at -80°C on to an LB plate.
2. Incubate overnight at 37°C.
3. Inoculate a single colony from the LB plate into 10 mL of liquid 1X LB.
4. Incubate overnight at 37°C with shaking (200 rpm).
5. Take 1.5 mL of overnight liquid culture and use for DNA extraction by kit (Bacterial Genomic DNA Isolation Kit, cat. no. 17900, Norgen Biotek Corp., Ontario).
6. Check the quality of DNA by agarose gel electrophoresis (i.e., run samples on a 1% TAE agarose gel containing 0.1 ng/mL of ethidium bromide at 100 V for 1 h).
7. Check the quantity of DNA in the following ways:
    a. NanoDrop 2000 (ThermoScientific, Massachusetts)
        i. Run the NanoDrop 2000/2000c Spectrophotometer software.
        ii. Select "Nucleic Acid."
        iii. Blank the spectrophotometer with 1 µL of sterile water by dropping it on to the surface of the measurement pedestal, lowering the sampling arm and pressing "Blank."
        iv. Gently wipe the surface clean with a KimWipe.
        v. Dispense 1 µL of sample on to the surface, lower the sampling arm and press "Measure."
        vi. After measurement, gently wipe the surface clean with a KimWipe and dispense the next sample.
    b. Qubit® fluorometer (Invitrogen, California)
        i. Use the Qubit® dsDNA BR Assay Kit (cat. no. Q32850) for double-stranded DNA quantification. Follow the online manual, but for each sample, use 90 µL of working solution and 10 µL of sample (100 µL total volume).
8. For long-term storage, keep DNA extracts at -20°C. DNA can be stored in concentrated form at 4°C for a few days.

B.2 Polymerase chain reaction (PCR)

1. For sequencing purposes, set up 50 µL reaction volumes on ice by adding the following components to thin-walled 0.2 mL PCR tubes:

Reagent	Volume (µL)	Final concentration
Sterile H_2O	38.5	----
10X *Pfu* Buffer with $MgSO_4$	5.0	1X
10 mM dNTPs	1.0	0.2 mM
25 µM forward primer	2.0	1 µM
25 µM reverse primer	2.0	1 µM
2.5 U/µL *Pfu*	0.5	1.25 U
DNA template (10 ng/µL)	1.0	10 ng

2. Prepare a negative control by adding sterile water in place of DNA template.
3. Program the following into a thermocycler:

Step	Temperature (°C)	Duration (min)
a. Initial denaturation	95	3
b. Denaturation	95	0.5
c. Annealing	5°C below primer melting temperature	0.5
d. Extension	72	2 min per 1 kb
e. Repeat steps (b) to (d) 29 times.		
f. Final extension	72	5

4. Place tubes into the thermocycler and start the program.
5. After finishing the PCR program, mix 5 µL of product in 1 µL of 6X loading dye and load it into a 1.5% TAE agarose gel containing 0.1 ng/mL of ethidium bromide.
6. Run the gel at 100 V for 1 h.
7. Visualize PCR products on a UV illuminator. If the amplification was specific (only one band visible), clean up the rest of the PCR reaction volume by using a NucleoSpin® Gel and PCR Clean-up Kit (cat. no. 740609, Macherey-Nagel, Germany). If multiple bands are seen, but the target band is still present, cut out the PCR band and follow the gel clean-up protocol of the NucleoSpin® kit.
8. Quantify PCR product concentration by Qubit, dilute the PCR reactions to recommended concentrations as listed on the Mobix facility's sequencing request form and send samples for Sanger sequencing.

B.3 Quantitative polymerase chain reaction (qPCR)

1. Calculate the genome equivalents present in DNA extracts using the following equation (Whelan et al., 2003):

$$\text{Genome equivalent} = \frac{(6.02 \times 10^{23} \text{ copy/mol}) \times (\text{DNA amount in grams})}{(\text{DNA length in bp}) \times (660 \text{ g/mol/bp})}$$

2. Dilute template to desired concentrations. For 10-fold quantification curves, use 10 µL of total volume for dilutions and start the first dilution with 1×10^7 or 1×10^6 genome equivalents/µL. Serially dilute by adding 1 µL of the previous dilution to 9 µL of sterile water. Use at least 6 dilutions for the curve.
3. The following components are for each 10 µL reaction. Add the components in order (i.e., from top to bottom in the table). Two qPCR chemistries are below:

SsoFast EvaGreen dye
(BioRad, California)

Reagent	Volume (µL)	Final concentration
Sterile water	3.6	----
10 µM forward primer	0.2	0.2 µM
10 µM reverse primer	0.2	0.2 µM
2X SsoFast EvaGreen	5.0	1X
DNA template	1.0	----

TaqMan probe
(Applied Biological Materials Inc., British Columbia)

Reagent	Volume (µL)	Final concentration
Sterile water	1.7	----
10 µM forward primer	1.0	1.0 µM
10 µM reverse primer	1.0	1.0 µM
2X TaqProbe Mastermix	5.0	1X
10 µM TaqMan probe	0.3	0.3 µM
DNA template	1.0	----

4. Prepare a master mix in for the number of reactions. All reactions, except for no-template control (NTC), are in triplicate. NTC should be done in at least duplicate. Allow for 5-10% pipetting error when calculating volumes for the master mix.
5. Dispense 9 µL of reaction mix into the wells of a 96-well plate using a 10 µL digital pipette. Ensure that the whole volume has been dispensed by pressing the tip to the side of the well.

6. Dispense 1 µL of sample to the reaction mix. Touch the pipette tip to the side of the well when dispensing to avoid drawing up the reaction mix.
7. Roll an adhesive seal over the wells and press down.
8. Place the plate in the CFX96 qPCR thermocycler and run the following program:

SsoFast EvaGreen dye

Step	Temperature (°C)	Duration
a. Enzyme activation/ initial denaturation	95	2 min
b. Denaturation	95	5 s
c. Annealing/extension	55-60	5 s
d. Repeat steps (b) and (c) 39 times.		
e. Melt curve	55-95 (in 0.2°C increments)	5 s per incremental step

TaqMan probe

Step	Temperature (°C)	Duration
a. Enzyme activation/ initial denaturation	95	10 min
b. Denaturation	95	15 s
c. Annealing/extension	60-65	60 s
d. Repeat steps (b) and (c) 39 times.		

9. To visualize PCR products for reactions with TaqMan probes, mix 10 µL of reaction with 2 µL of 6X loading dye, load the mixes on to an agarose gel (2% TAE containing 0.1 ng/mL of ethidium bromide) and run the gel at 100 V for 1.5 h or 130 V for 45 minutes (if short on time). If primer pairs are being tested with SsoFast EvaGreen-based reactions for the first time, it is recommended that the reactions are visualized on an agarose gel (in addition to the melt curve).
10. After the qPCR run, export data on to a USB flash drive and open the data file in the BioRad CFX Manager software. Designate a plate file for the data file or manually attribute properties to each well (e.g., sample type = Standard, Unknown, NTC). Save the optical file. The data is now ready to be analyzed.

Reference: Whelan, J.A., Russell, N.B., and Whelan, M.A. (2003). A method for the absolute quantification of cDNA using real-time PCR. J Immunol Meth *278*, 261-269.

B.4 Spiking experiments

1. Determine a range of competing DNA that will be spiked into reactions containing target sequences. For example, competing DNA can be spiked into reactions at 10-fold excess to 1000-fold excess of target DNA. Remember to include negative controls (i.e., only non-target, competing DNA present in the mixture).
2. Calculate the genome equivalents of DNA (Whelan et al., 2003):

$$\text{Genome equivalent} = \frac{(6.02 \times 10^{23} \text{ copy/mol}) \times (\text{DNA amount in grams})}{(\text{DNA length in bp}) \times (660 \text{ g/mol/bp})}$$

3. Set up 8 µL DNA mixtures containing competing and target DNA. For example:

Competing DNA in 1000-fold excess of target DNA (genome equivalents)

Type	Strain	Initial concentration	Volume added	Final concentration
Target	*E. coli* O157:H7 str. EDL933	2×10^4 genomes/µL	4 µL	1×10^4 genomes/µL
Competing	*E. coli* K-12 str. MG1655	2×10^7 genomes/µL	4 µL	1×10^7 genomes/µL

Competing DNA in 1000-fold excess of target DNA (amount)

Type	Strain	Initial concentration	Volume added	Final concentration
Target	*E. coli* O157:H7 str. EDL933	1.2×10^{-1} ng/µL	4 µL	0.6×10^{-2} ng/µL
Competing	Fish sperm DNA	1.2×10^2 ng/µL	4 µL	6.0×10^1 ng/µL

4. Set up and dispense the qPCR master mix as outlined in Appendix B.3.
5. Add 1 µL of the spiking mixtures to the appropriate wells. Touch the pipette tip to the side of the well when dispensing to avoid drawing up the reaction mix.
6. See Appendix B.3 for further instructions.

Reference: Whelan, J.A., Russell, N.B., and Whelan, M.A. (2003). A method for the absolute quantification of cDNA using real-time PCR. J Immunol Meth *278*, 261-269.

Appendix C: Defense presentation slides

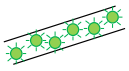
Insertion/deletion (indel) based approach for the detection of E. coli O157:H7 in freshwater environments

Shirley Wong
M.Sc. defence

May 29, 2014

Outline

- Background
- Part I: *E. coli* O157:H7-specific assay, EvaGreen dye format
- Part II: *E. coli* O157:H7-specific assay, TaqMan probe format



EvaGreen dye

TaqMan probe

Enterohaemorrhagic *E. coli* O157:H7

- *E. coli* – Gram-negative bacterium, facultative anaerobe, lives in the gastrointestinal tract of mammals
- O157:H7 – pathogenic serotype
- Pathogenesis:
  - Production of Shiga toxins (*stx* genes) [1]
  - Creation of attaching and effacing lesions (*eae* and *tir* genes) [2]
- Reservoirs – ruminants, swine
- Sources of infection – ground beef, vegetables, drinking water
- Association with outbreaks of gastrointestinal illnesses

[1] Lee, J.E., et al. 2002. Renal Failure **24**(5): 567-75.
[2] DeVinney, R., et al. 2001. Molecular Microbiology **41**(6): 1445-58.

DNA-based detection assays

- Simplex PCR – targets: *stx* [1], *eae* [2]
- Multiplex PCR – targets: *rfbE* (O-antigen), *fliC* (H-antigen), *stx*, *eae* [3]
- TaqMan probe-based qPCR targeting virulence genes [4]
- Identification of *E. coli* O157:H7 requires positive results from multiple gene targets
  - In the case of serotyping assays, requires positive results from O- and H-antigen-encoding genes

[1] Gannon, V.P., King, R.K., Kim, J.Y., and Thomas, E.J. (1992). Applied & Environmental Microbiology 58, 3809-3815.
[2] Gannon, V.P., Rashed, M., King, R.K., and Thomas, E.J. (1993). Journal of Clinical Microbiology 31, 1268-1274.
[3] Hu, Y., Zhang, Q., and Meitzler, J.C. (1999). Journal of Applied Microbiology 87, 867-876.
[4] Nagano, I., Kunishima, M., Itoh, Y., Wu, Z., and Takahashi, Y. (1998). Microbiology and Immunology 42, 371-376.

Conserved signature insertions/deletions (CSIs)

- Rare genetic changes in common ancestor of members in a taxonomic clade [1]
  - Used to elucidate evolutionary relationships among bacterial groups
- Has been used for development of a pyrosequencing assay (*Bacillus anthracis*) [2]
- Focus of qPCR assay design: *E. coli* O157:H7-specific insertion

[1] Gupta, R.S. (1998). Microbiology and molecular biology reviews : MMBR 62, 1435-1491.
[2] Ahmod, N.Z., Gupta, R.S., and Shah, H.N. (2011). Journal of microbiological methods 87, 278-285.

CSI: 2-oxoglutarate and Fe(II)-dependent oxygenase

- Encoded by *ybiX*
- Red box: O157:H7-specific insertion
- "–" denotes identity to first amino acid

Host Type	Strain	Region of Interest (170 … 220)
Human pathogen	Escherichia coli O157:H7 EDL933	AIFINIQGMIKDKKKKAMLFELDK WIQ WIQSLKSYGEMEEIALLNVYNH
Human pathogen	Escherichia coli O157:H7 Sakai	---------- --- ----------
Human pathogen	Escherichia coli O157:H7 EC4013	---------- --- ----------
Human pathogen	Escherichia coli O157:H7 1125	---------- --- ----------
Human pathogen	Escherichia coli O55:H7 CB9615	---------- --- ----------
Human pathogen	Escherichia coli WPO21017.9	---M---------- ----------S----------
Human pathogen	Escherichia coli WPO21017.11	---M---------- ----------S----------
Human pathogen	Escherichia coli WPO21552.7	---M---------- ----------S----------
Human pathogen	Escherichia coli WP305240.5	---M---------- ----------S----------
Human pathogen	Escherichia coli WP305293.10	---M---------- ----------S----------
Human pathogen	Escherichia coli WP302308.14	---M---------- ----------S----------
Human pathogen	Escherichia coli O45:K1:H7 S88	---M---------- ----------S----------
Human pathogen	Escherichia coli EC93	---M---------- ----------S----------
Human pathogen	Escherichia coli O128:H21 DEC14A	---M---------- ----------S----------
Human pathogen	Escherichia coli P12b	---M---------- ----------S----------
Human pathogen	Escherichia coli B494	---M---------- ----------S----------
Human commensal	Escherichia coli 3431	---M---------- ----------S----------
Human commensal	Escherichia coli NC99945-2	---M---------- ----------S----------
Human commensal	Escherichia coli NC119FL	---M---------- ----------S----------
Human commensal	Escherichia coli ETEC102	---M---------- ----------S----------
Human commensal	Escherichia coli K-12 MG1655	---M---------- ----------S----------
Human commensal	Escherichia coli HCD2	---M---------- ----------S----------
Human commensal	Escherichia coli W6 182-1	---M---------- ----------S----------
Avian pathogen	Escherichia coli O08	---M---------- ----------S----------
Porcine pathogen	Escherichia coli O149 UMN888	---M---------- ----------S----------
Human pathogen	Escherichia coli O44:H18 042	---M---------T---- ----------S----------
Human pathogen	Escherichia coli UMEA 3805-1	---M---------- ----------
Human pathogen	Escherichia coli O8 IAI1	---M---------- ----------
Human pathogen	Escherichia coli O111:H- 11671	---M---------- ----------
Human pathogen	Escherichia coli O139:H2 E24377A	---M---------- ----------
Human pathogen	Escherichia coli WP304816.8	---M---------- ----------
Human pathogen	Escherichia coli 2735000	---M---------- ----------
Human commensal	Escherichia coli DEC12B	---M---------- ----------
Human commensal	Escherichia coli HS11	---M---------- ----------
Bovine commensal	Escherichia coli O91:H- 99.0741	---M---------- ----------
Bovine commensal	Escherichia coli O5:H- 97.0246	---M---------- ----------
Human pathogen	Escherichia coli O7:K1 IA139	---M----------T ----------R--S----------
Human pathogen	Escherichia coli O2:K25 ETEC_7v	---M----------T ----------R--S----------
Human commensal	Escherichia coli ETEC196	---M---------- ----------
Human commensal	Escherichia coli 113290	---M---------- ----------S----------

Part I:
E. coli O157:H7-specific assay, EvaGreen dye format

7

Objectives

- To develop EvaGreen-based qPCR assay to detect *E. coli* O157:H7 strains
- To establish a range of linearity and specificity of the assay

8

Strains used in this study

Seropathotype[a]	Serotype	Strain	Pathotype[b]	Reference[c]
A	O157:H7	EDL933	EHEC	[1]
B	O26:H11	CL1	EHEC	[2]
	O111:NM	R82F2	EHEC	[2]
	O121:H19	CL106	EHEC	[2]
	O145:NM	N00-6496	EHEC	[2]
C	O5:NM	N00-4067	EHEC	[2]
	O113:H21	CL3	EHEC	[2]
	O121:NM	N99-4390	EHEC	[2]
D	O103:H25	N00-4859	EHEC	[2]
	O172:NM	EC6-484	EHEC	[2]
E	O84:NM	EC2-044	EHEC	[2]
	O98:H25	EC3-377	EHEC	[2]
----	O6:K2:H1	CFT073	UPEC	[3]
----	OR:H48:K-	MG1655	K-12	CGSC

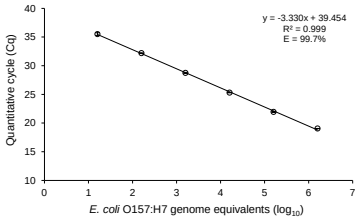
[a] seropathotypes classified according to Karmali et al.[2]
[b] EHEC = enterohaemorrhagic *E. coli*; UPEC = uropathogenic *E. coli*
[c] CGSC = Coli Genetic Stock Center

- One *E. coli* O157:H7 strain
- 12 non-O157:H7 pathogenic *E. coli* strains
- One *E. coli* K-12 strain

[1] Riley, L.W., et al. (1983). New England Journal of Medicine *308*, 681-685.
[2] Karmali, M.A., et al. (2003). Journal of Clinical Microbiology *41*, 4930-4940.
[3] Culham, D.E., et al. (2001). Microbiology *147*, 1657-1670.

9

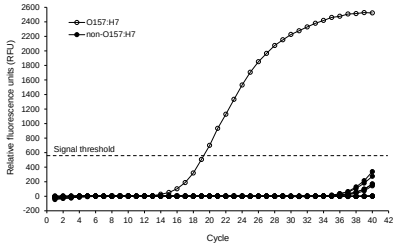
Range of linearity



- *E. coli* O157:H7 DNA extract
- Range of linearity from 16 to 1.6 x 10^6 genome equivalents
- Amplification efficiency = 99.7%

10

Specificity



- DNA extracts of *E. coli* strains
- Only *E. coli* O157:H7 produced amplification signal above the signal threshold (570 RFU)

11

Conclusions

- Initial study demonstrates applicability of this CSI to qPCR assays
- Improvement: using TaqMan probes
  - Probes are sequence-specific, while EvaGreen intercalates into double-stranded DNA
    - EvaGreen: high background fluorescence in samples with background flora

12

Part II:
E. coli O157:H7-specific assay, TaqMan probe format

13

Objectives

- To design a TaqMan probe-based qPCR that identifies *E. coli* O157:H7 strains using the CSI found in the *ybiX* gene
- To assess the performance of the *ybiX* CSI assay in the presence of competing non-target DNA
  - Spiking experiments

14

Strains used in this study

Seropathotype[a]	Serotype	Strain	Pathotype[b]	Reference[c]
A	O157:H7	EDL933	EHEC	[1]
B	O26:H11	CL1	EHEC	[2]
	O26:H11	CL9	EHEC	[2]
	O111:NM	R82F2	EHEC	[2]
	O121:H19	CL106	EHEC	[2]
	O145:NM	N00-6496	EHEC	[2]
	O145:NM	N02-5149	EHEC	[2]
C	O5:NM	N00-4067	EHEC	[2]
	O113:H21	CL3	EHEC	[2]
	O121:NM	N99-4390	EHEC	[2]
D	O103:H25	N00-4859	EHEC	[2]
	O172:NM	EC6-484	EHEC	[2]
E	O84:NM	EC2-044	EHEC	[2]
	O98:H25	EC3-377	EHEC	[2]
----	O6:K2:H1	CFT073	UPEC	[3]
----	OR:H48:K-	MG1655	K-12	CGSC

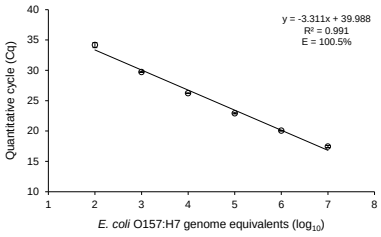
[a] seropathotypes classified according to Karmali et al.[2]
[b] EHEC = enterohaemorrhagic *E. coli*; UPEC = uropathogenic *E. coli*
[c] CGSC = Coli Genetic Stock Center

- One *E. coli* O157:H7 strain
- 14 non-O157:H7 pathogenic *E. coli* strains
- One *E. coli* K-12 strain

[1] Riley, L.W., et al. (1983). New England Journal of Medicine *308*, 681-685.
[2] Karmali, M.A., et al. (2003). Journal of Clinical Microbiology *41*, 4930-4940.
[3] Culham, D.E., et al. (2001). Microbiology *147*, 1657-1670.

15

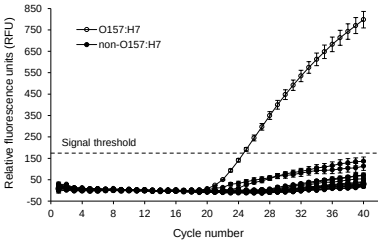
Range of linearity



- *E. coli* O157:H7 DNA extract
- Range of linearity from 1 x 10^2 to 1 x 10^7 genome equivalents
- Amplification efficiency = 100.5%

16

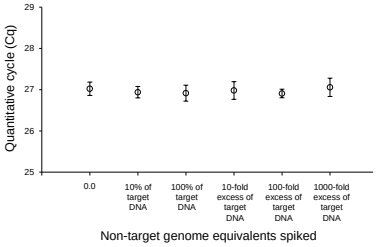
Specificity



- DNA extracts of *E. coli* strains
- Only *E. coli* O157:H7 produced amplification signal above the signal threshold (170 RFU)

17

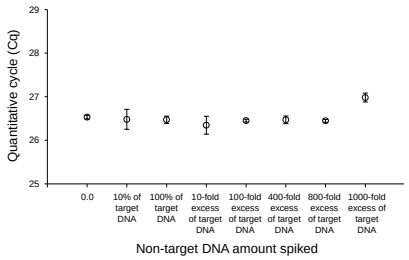
DNA competition: *E. coli* K-12 DNA + O157:H7 DNA



- *E. coli* K-12 (**non-target**) – genetically similar DNA
  - Possibility of TaqMan probes adhering to non-target DNA
- No lowered sensitivity to *E. coli* O157:H7 DNA up to non-target DNA spiked in at 1000-fold excess of target

18

DNA competition: fish sperm DNA + O157:H7 DNA



- Fish sperm DNA (**non-target**) – complex heterologous DNA
  - Proxy for environmental DNA and potentially obscures target DNA template
- No lowered sensitivity to *E. coli* O157:H7 DNA up to 800-fold excess

19

Conclusions and future directions

- Potential use of TaqMan probe-based assay on environmental samples
  - Sample types with high levels of background flora, including storm water or agricultural run-off
- Further application of this assay on environmental samples
  - Impact of different factors on assay's sensitivity
- Strategy of CSI assay development can be applied to other indels

20

Acknowledgements

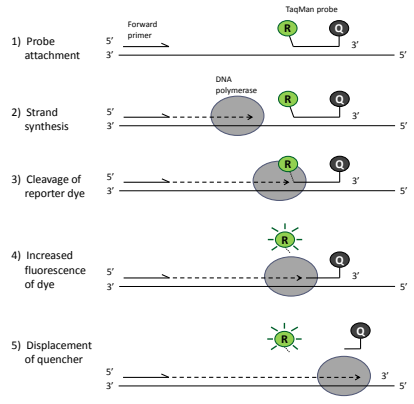
- Dr. Schellhorn
- Dr. Gupta
- Dr. Kolasa
- Schellhorn lab members, past and present
- The audience

21

Thank you!

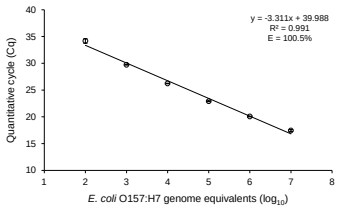
Supplementary Information

TaqMan probes



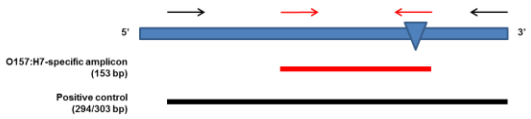
Calculating amplification efficiency

- Efficiency = $(10^{(-1/\text{slope})}) - 1$

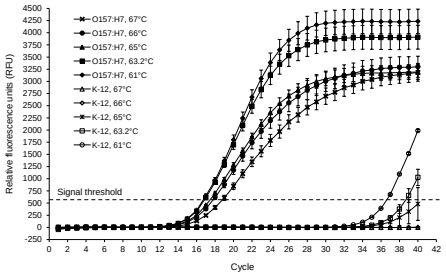


EvaGreen-based assay – primer locations

- O157:H7-specific primers (red)
  - *ybiX* (O157:H7 str. EDL933)
- Control primers (black)
  - Conserved regions from multiple nucleotide sequence alignment

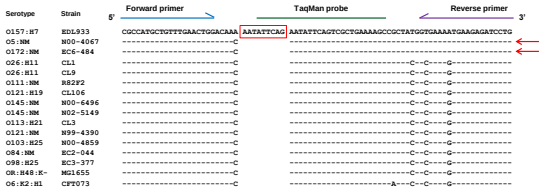


EvaGreen-based assay: determining annealing temperature and signal threshold



- O157:H7 – target strain K-12 – non-target strain

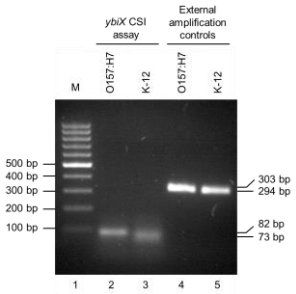
TaqMan probe-based assay – primers and probe



- "–" denotes identity to first row

PCR products from TaqMan probe-based assay

- O157:H7 – target strain
- K-12 – non-target strain



TaqMan probe-based assay – primers and probe