8-2019

16S rRNA analysis and toxin gene presence in Escherichia coli isolated from beach water and sand at a public beach (Erie County, NY)

Jennifer D. Jackson
State University of New York College at Buffalo - Buffalo State College, jennifer6814@gmail.com

Advisor

Gary W. Pettibone, PhD., Professor of Biology

First Reader

Gary W. Pettibone, PhD., Professor of Biology and Thesis Advisor

Second Reader

Amy M. McMillan, PhD., Professor of Biology

Third Reader

Gregory J. Wadsworth, PhD., Associate Professor of Biology

Department Chair

Daniel L. Potts, Ph.D., Chair and Associate Professor of Biology

To learn more about the Biology Department and its educational programs, research, and resources, go to https://biology.buffalostate.edu/.

State University of New York
Buffalo State College
Department of Biology

16S rRNA analysis and toxin gene presence in *Escherichia coli* isolated from beach water and sand at a public beach (Erie County, NY)

An Abstract of a Thesis in
Biology

By

Jennifer D. Jackson

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts
August 2019

Abstract:

Every year, thousands of people utilize beaches for recreation, but most are unaware of *Escherichia coli* (*E. coli*) contamination and the possibility of acquiring an infection. In this study, 173 strains of *E. coli* were isolated from sand and adjacent waters from a public beach in Erie County, NY and analyzed for genetic relatedness based on sequence differences in the variable regions of the 16S rRNA gene. Some of the variable regions (V1 and V6) proved useful in constructing phylogenetic trees but the discriminatory power of these regions was inadequate to resolve intraspecies differences. Therefore, whether extant populations of *E. coli* differ between water and sand environments could not be determined. All environmental isolates also were analyzed for the presence of the toxin genes: *papC*, *sfa/foc*, stx_1 and stx_2. None of the isolates harbored the stx_1 or stx_2 genes, which code for a potentially lethal Shiga toxin. However, the *papC* gene and the *sfa/foc* gene were present in 5.2% (n=9) and 7.5% (n=13) of the isolates, respectively. These genes are known to be associated with the ability of *E. coli* to cause urinary tract infections, and their presence in beach sand and recreational water represents a health risk to user populations.

State University of New York
Buffalo State College
Department of Biology

16S rRNA analysis and toxin gene presence in *Escherichia coli* isolated from beach water and sand at a public beach (Erie County, NY)

A Thesis in
Biology

By

Jennifer D. Jackson

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts
August 2019

Approved by:

Gary W. Pettibone, Ph.D.
Professor of Biology
Chairperson of the Committee and Thesis Advisor

Daniel L. Potts, Ph.D.
Chair and Associate Professor
Department of Biology

Kevin J. Miller, Ed.D.
Dean
The Graduate School

Thesis Committee

Gary W. Pettibone, Ph.D.
Professor of Biology
Chairperson of the Committee and Thesis Advisor

Amy M. McMillan, Ph.D.
Professor of Biology

Gregory J. Wadsworth, Ph.D.
Associate Professor of Biology

Table of Contents

List of Figures and Tables

1.0 Introduction:

Microorganisms are ubiquitous in their distribution and inhabit areas from extreme environments, such as black smoker vents in the deep sea (Prieur 1997) and ice fields in Antarctica (Antony et al. 2016), to more temperate regions of the world. Since its discovery in 1885, the bacterium *Escherichia coli* (*E. coli*) has been one of the most studied and best characterized microorganisms in the world. *Escherichia coli* are members of a family of Gram-negative bacteria, Enterobacteriaceae, that mainly inhabit the intestinal tract of warm-blooded animals which include humans, livestock, birds and wild game, among others. Enterobacteriaceae is a large family that includes many pathogens to humans such as *Shigella*, *Salmonella*, *Klebsiella*, and *Yersinia pestis* as well as numerous beneficial bacteria.

Total coliform is a broad term that encompasses several genera of bacteria within the Enterobacteriaceae family, some of which can survive in the external environment under normal conditions. Fecal coliform, perhaps more accurately known as thermotolerant coliform, are a subset of the total coliform bacteria that, with a few exceptions, live only in the intestinal tract of warm-blooded animals. The few thermotolerant bacteria genera that can live in the external environment are *Klebsiella* (Knittel et al. 1977), *Enterobacter* (Patel et al. 2016), and *Citrobacter* (Rivera et al. 1988) (KEC coliforms) and to a lesser extent *E. coli* (Ishii et al. 2006). *Escherichia coli* is a fecal (thermotolerant) coliform and as such has historically been used as an indicator species for environmental

contamination by fecal matter since it is assumed that its only natural habitat is the intestinal tract of warm-blooded animals. In addition, it is not considered a pathogen to humans under normal circumstances and is believed to have a relatively fast die-off rate in the environment (Cote and Quessy 2005). Due to these characteristics, if *E. coli* is found in water or soil, it is assumed that there has been a recent fecal contamination event.

As natural microbiota, most strains of *E. coli* are considered harmless and even beneficial to the host. However, some strains are known to harbor toxin genes that, when in the wrong location such as the urinary tract, brain meninges, or respiratory tract, or in large quantities, can be harmful and in some cases cause the death of the human host. There are various exotoxins that can be produced by *E. coli*. Some of the most well-known and documented infections are caused by the enterohemorrhagic *E. coli* (EHEC) serotype O157:H7, a subcategory of the Shiga-toxin producing *E. coli* (STEC) pathotype. The STEC pathotype has historically been known as Verocytotoxin-producing *E. coli* (VTEC) (Karmali 1989). Invasion by these bacteria cause damage to blood vessels which can lead to hemorrhagic colitis, and a potentially fatal form of kidney failure known as hemolytic uremic syndrome (HUS) (https://www.cdc.gov/features/ecoliinfection/index.html). Outbreaks of STEC have been known to occur worldwide. Perhaps the most well-known outbreak of the EHEC O157:H7 bacterium in the United States happened in 1993 when over 500 (laboratory confirmed) persons were infected with the bacteria from several dozen Jack-in-the-Box restaurants in the northwestern United States. This

outbreak resulted in four deaths and almost 200 permanent life-altering injuries (https://www.cdc.gov/mmwr/preview/mmwrhtml/00020219.htm).

Despite being the most well-known foodborne illness outbreak, the 1993 case was not the first in the United States. Previous to 1993, there were over 20 outbreaks that had been documented since 1982 (Rangel et al. 2005). According to the Centers for Disease Control and Prevention (CDC), most of the outbreaks that have occurred since 2006 have been caused by consumption of ground beef and beef products that were contaminated with EHEC O157:H7. However, other outbreaks have arisen from a variety of foods such as spinach, lettuce, hazelnuts, and cookie dough. One outbreak in 2011 was even linked to travel outside of the United States (https://www.cdc.gov/ecoli/).

Shiga toxin (*stx*) is a type of AB toxin which causes dysentery and is among the most potent biological toxins known (Henkel et al. 2010). AB toxins are also associated with cholera, diphtheria, anthrax, tetanus, pertussis, and botulism, among others. *Stx* was discovered in the bacterium *Shigella dysenteriae* in the late 1800s by Dr. Kiyoshi Shiga but not characterized until the late-20th Century (Brown et al. 1982). Almost 80 years after the discovery of the *Shigella* bacterium, similar illnesses were identified in humans but were found to be caused by rare strains of *E. coli*, not *S. dysenteriae* (Riley et al. 1983). It was eventually discovered that there were two variations of *stx* in *E. coli*, *stx*$_1$ and *stx*$_2$. *Stx*$_1$ is nearly identical to *stx* of *S. dysenteriae* differing by only one amino acid (Calderwood et al. 1987, Pacheco and Sperandio 2012) while *stx*$_2$ has between 50%-60% similarity to *stx*$_1$ (Newland et al. 1987, Obrien and Holmes

1987, Fraser et al. 2004) and by extension *stx* itself. Konowalchuk et al. (1977) were the first to characterize these toxins identified from certain *E. coli* strains. It was shown that *stx*$_1$ and *stx*$_2$ had become incorporated into the genome of *E. coli* via bacteriophage (Huang et al. 1987, Makino et al. 1999, Plunkett et al. 1999, Yokoyama et al. 2000). The STEC pathotypes of *E. coli* got its name from the presence of shiga-like toxins in its genetic material so by definition contain at least one version of the shiga-like toxin, *stx*$_1$, *stx*$_2$, or both (Konowalchuk et al. 1977). On average, 5%-15% of STEC infections develop HUS, but this rate can be much higher depending on strain and geographic location (Mayer et al. 2012). It has been discovered that the stx2 gene is associated with more severe clinical diseases than those carrying only the stx1 gene (Boerlin et al. 1999, Ritchie et al. 2003).

In addition to STEC, other important toxigenic pathotypes, including enteroaggregative *E. coli* (EAggEC), enteroadherent *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), and diffusely adhering *E. coli* (DAEC), can lead to serious diarrheal diseases (Estrada-Garcia et al. 2009). Each of these virulent types has different mechanisms by which they infect the host (Clements et al. 2012) and can therefore cause diseases of varying severity ranging from mild intestinal discomfort up to and including death.

Another category of toxigenic *E. coli* is extraintestinal pathogenic *E. coli* (ExPEC). Extraintestinal pathogenic *E. coli* are the most common bacteria that cause urinary tract infections (UTIs), but they also cause newborn meningitis,

pneumonia and bacteremia (Escobar-Paramo et al. 2004, Iranpour et al. 2015, Poolman and Wacker 2016). Extraintestinal pathogenic *E. coli* have been shown to cause over 50% of UTIs, 25% of septic infections, 14% of skin and soft tissue infections, and 9% of lower respiratory infections from 19 United States hospitals between 2007 and 2010. The estimated number of deaths more than doubled during this time period from approximately 40,000 in 2001 to greater than 85,000 in 2014 (Poolman and Wacker 2016). Several genes have been determined to assist in the attachment of ExPEC strains to epithelial cells that line the urinary tract, but the two genes of interest in this study are *papC* and *sfa/foc*.

Pyelonephritis-associated pili (Pap) is an adhesion molecule that allows *E. coli* to tightly associate with epithelial cells in the kidney, ureter, and several other locations in the human body via a specific galactose-galactose moiety on eukaryotic extracellular glycolipids (Breimer and Karlsson 1983, Hultgren et al. 1991, Hacker 1992) and is the primary adherence factor of uropathogenic *E. coli* (Klemm 1994). Adherence factors allow for the bacteria to remain in contact with the eukaryotic cell despite the normal, regular flow of bodily fluids (Klemm 1994). The Pap adhesin is a cluster of 11 genes that all work in concert to form the pilus (Klemm 1994). The specific gene of interest in this cluster is *papC*, an outer membrane usher protein that is necessary for the assembly of the Pap pilus (Johnson 1991, Klemm 1994). Some studies have determined that the P-fimbrianated strains of *E. coli* constitute 60%-70% of all UTI cases (Hultgren et al. 1991, Hacker 1992), and other studies concluded that up to 50% of adult and

90% of children UTI *E. coli* strains expressed P-fimbriae (Kallenius et al. 1981, Johnson 1991).

Two other adhesion molecules are also implicated in the adherence of bacteria to eukaryotic cells: sialic acid-specific (S) fimbriae adhesin (*sfa*) and F1C fimbriae (*foc*). These adhesins recognize a sialic acid moiety on glycolipids. These adhesins can affect the glomeruli of the kidney (*sfa*) (Korhonen et al. 1986, Korhonen et al. 1990), the distal tubules and collecting ducts of the kidney (*foc*) (Korhonen et al. 1990) and the vascular epithelium of the kidney and bladder (*sfa* and *foc*) (Virkola et al. 1988). Kunin et al. (1993) determined that up to 13% of *E. coli* isolates from UTIs were positive for *sfa* and Pere et al. (1985) found that up to 30% of *E. coli* isolates from UTIs were positive for *foc*. Because of these alarming statistics for ExPEC (*papC* and *sfa/foc*) and the prevalence and severity of STEC illnesses (stx_1 and stx_2), it is important to determine if toxigenic *E. coli* can persist outside of its typical host habitat.

In 2004, Escobar-Paramo et al. tested 98 *E. coli* and *Shigella* strains for the presence of 17 virulence factors (VF). Of the 98 strains, there were 11 STEC (11%, including 9 [9%] EHEC) pathotypes and 9 (9%) ExPEC pathotypes. They developed a phylogenetic tree categorizing the 98 different strains into groups based on the genetic background of the individual strains using the sequences of six essential genes (*trpA*, *trpB*, *pabB*, *putP*, *icd*,and *polB*) that have been shown to exhibit low levels of horizontal gene transfer in *E. coli* (Lecointre et al. 1998, Denamur et al. 2000). *Escherichia coli* has been shown to consist of seven main phylogenetic groups (A, B1, B2, C, D, E and F) based on sequence analysis of

two genes (*chuA* and *yjaA*) and the small DNA fragment, TSPE4.C2 (Clermont et al. 2013). Escobar-Paramo, et al. (2004) were able to determine that certain VFs are contained within certain phylogroups. They found that *stx*$_1$ and *stx*$_2$ were present in 100% of both the STEC and EHEC pathotypes. The *stx*$_1$ and *stx*$_2$ genes were mainly detected within phylogroup E but also were found in phylogroups A and B1. The *papC* and the *sfa/foc* genes were present in 8 of 9 (~89%) of the ExPEC pathotypes. However, *papC* was present in 11 other *E. coli* strains and *sfa/foc* was present in four other *E. coli* strains as well. The ExPEC pathotypes were mostly placed in phylogroup B2 with a few placed in phylogroup D (Escobar-Paramo et al. 2004). Therefore, the *stx*$_1$ and *stx*$_2$ genes are rarely, if ever, found in the same genetic phylogroups as the *papC* and *sfa/foc* genes.

Over the last few years it has been shown that *E. coli* can persist in secondary habitats such as soil (Ishii et al. 2006, Brennan et al. 2010, Byappanahalli et al. 2012), wastewater treatment plant effluents (Zhi et al. 2016), *Cladophora* mats in the Great Lakes (Whitman et al. 2003, Ksoll et al. 2007), and beach sands and waters (both freshwater and marine) (Whitman and Nevers 2003, Shibata et al. 2004, Lee et al. 2006, Bonilla et al. 2007, Ishii et al. 2007, Walk et al. 2007, Yamahara et al. 2009, Abdelzaher et al. 2010, Solo-Gabriele et al. 2016) among other places (Burton et al. 1987, Lyautey et al. 2010, Chandrasekaran et al. 2015, Jang et al. 2015). Of particular interest is the bacterial load associated with beach water and sands since beach-related

activities are favorite pastimes for many in the summer. Stevenson (1953) showed that there have been a larger number of individuals who developed some kind of illness while swimming compared to non-swimmers. It also was shown in a multi-year epidemiological review of several locations that there was a strong correlation between bacterial loads on beaches and gastrointestinal illnesses (Cabelli et al. 1982). According to the Ambient Water Quality Criteria for Bacteria regulation released by the US Environmental Protection Agency (EPA) in 1986, when *E. coli* concentrations become greater than 235 colonies per 100mL of water from a single sample or if 5 samples within a 30 day period have a geometric mean of at least 126 colonies per 100mL of water, fresh water recreational beaches must close down (https://www.epa.gov/beaches).

However, the only samples considered during the monitoring of beaches are obtained from the water column. It has been shown that the levels of bacteria detected in beach sands may be several logs higher than those in water (Burton et al. 1987). Most beach sand studies to date have focused on sampling within one meter on either side (beachside or waterside) of the swash zone, or in the case of marine beaches, the area between low and high tides. Other studies have shown that the numbers of bacteria that inhabit the sand are on average one to two logs higher than the surrounding water (An et al. 2002, Alm et al. 2003, Whitman and Nevers 2003, Whitman et al. 2003, Yamahara et al. 2009, Abdelzaher et al. 2010). In 2010, Abdelzaher et al. showed that the bacterial levels in wet sands from an intertidal zone of a beach in Florida were two to four

orders of magnitude higher than in the adjacent water column, and that the bacterial levels in the dry sand above the intertidal zone were one to three orders of magnitude higher than levels observed in the wet sand. Whitman and Nevers (2003) noted that the *E. coli* counts were up to two orders of magnitude higher in foreshore beach sands than the adjacent waters at a Lake Michigan beach. Similarly, Bonilla et al. (2007) showed that bacterial levels were up to four orders of magnitude higher in dry sands than in adjacent water samples.

A previous study at Bennett Beach (Angola, NY) discovered an unexpectedly large *E. coli* population present in the beach sand six meters up beach from the swash zone (unpublished data) in what would seem to be conditions that are not conducive (e.g. relatively low water content in the sand) for persistence and growth. Since *E. coli* have been discovered to exist in sands/soils and waters, some researchers have postulated that certain strains may have become “naturalized” to these environments and are therefore questionable for use as a fecal indicator bacterium (FIB) for recent fecal contamination events (Ishii et al. 2006, Ishii and Sadowsky 2008, Brennan et al. 2010, Zhang and Yan 2012). Perchec-Merien and Lewis (2013) defined a naturalized population of *E. coli* as one that “does not derive directly from fecal sources and has evolved without host contact for a sufficiently long time to demonstrate reproducible characteristics”. Some researchers have established that the length of survivability in different types of soil for commensal *E. coli* and for EHEC O157:H7 is approximately the same, indicating that commensal *E. coli*

can be utilized as an indication of the presence of the more virulent strain (Mubiru et al. 2000). However, Smith et al. (2009) discovered that the incidence of FIB did not correlate to the presence of Shiga-toxin producing *E. coli*. In this case, the relative abundance of the Shiga-like toxin 1 and 2 (*stx*$_1$ and *stx*$_2$) genes were variable and did not show correlation to the abundance of non-toxigenic commensal *E. coli*.

Certain *E. coli* strains have been shown to have better adapted to growing outside of their primary habitat (Gordon et al. 2002). This information leads to the assumption that naturalized populations can occur and perhaps can begin to replicate in the secondary environment under certain conditions (Ishii et al. 2006, Ishii and Sadowsky 2008, Ishii et al. 2010). Over the past few years, several researchers have determined that among the different phylogroups within the *E. coli* species that the "naturalized" populations tend to fall into the B1 phylogroup (Escobar-Paramo et al. 2004, Walk et al. 2007). Walk et al. (2007) also hypothesized that the individuals in the B1 phylogroup may possess certain traits that allow them to persist and reproduce in the secondary habitat. Zhang and Yan (2012) found that the intracellular concentration of trehalose in sand-associated *E. coli* isolates was elevated and that assisted in the resistance to desiccation and the survival of the organism. Deng et al. (2014) also have postulated that there is a genetic marker, a putative glucosyltransferase gene (*ycj*M), that can differentiate between enteric strains of *E. coli* and naturalized strains because it appears to be enteric-specific.

Escherichia coli is still used as a FIB to indicate recent fecal contamination but its effectiveness as a FIB would be questionable if naturalized *E. coli* have a distinct genetic background. In 1994 Schmidt and Relman laid out a fairly new method of determining evolutionary relatedness among species. It focused on the small subunit (SSU) of the ribosomal RNA (16S rRNA) gene. At the time it was best done by manually aligning an RNA hairpin structure against known organisms and then forming a "best-fit" tree that was compatible with the sequence data. Over the years the techniques for analyzing evolutionary relatedness have improved greatly. In the last several decades, many researchers have used 16S rRNA gene sequencing to identify and classify microorganisms from multiple environments, proving its usefulness (Loong et al. 2016, Makapela et al. 2016, Pu and Hou 2016, Andini et al. 2017). Because 16S rRNA gene sequencing has been shown to accurately determine genetic variation among organisms, the use of this technique should be considered for determining possible genetic variabilities in microorganisms (e.g. *E. coli*) found in beach water and sands.

Since its first use in 1977 for distinguishing different organisms from each other (Woese and Fox), the SSU of ribosomes (16S for prokaryotes and 18S for eukaryotes) has been established as a new, more accurate classification system (Three Domains: Bacteria, Archaea, and Eukarya) for all organisms based on their molecular components (Woese et al. 1990) than the five kingdom classification system. Because ribosomes are a central part of cellular

machinery, the molecular makeup of ribosomes of all organisms within their respective domains is highly conserved. This allows the SSU to be an effective chronometer for determining how closely species are related to one another.

The 16S species of the SSU in prokaryotic ribosomes has been used successfully in the comparison of genomes for profiling bacterial communities (Lane et al. 1985, Bernhard and Field 2000a, b, Chakravorty et al. 2007, Guo et al. 2013, Loong et al. 2016). The gene that encodes the 16S ribosomal RNA (rRNA) protein is approximately 1500bp in length and contains highly conserved regions as well as 9 interspersed hypervariable regions (named V1 through V9) which have been shown to be useful in distinguishing among bacteria (Chakravorty et al. 2007, Yang et al. 2016). Coenye and Vandamme (2003) determined that there were specific variable regions that were more likely to distinguish different levels of taxa. In their study, it was determined that variable regions one and six had the most heterogeneity and were therefore more likely to distinguish organisms at the species level. According to Chakravorty et al. (2007), variable regions two and three can determine organisms at the genus level and V3 also contained the most single-nucleotide polymorphisms and therefore was able to distinguish between closely related species from the Enterobacteriaceae family. Variable region four was shown by Yang et al. (2016) to most closely resemble the same relatedness as if the entire 16S rRNA gene were used in the classification and was therefore not the best region to determine species identity. Because of the characteristics mentioned above and the fact

that there are highly conserved regions as well as nine interspaced hypervariable regions (Tringe and Hugenholtz 2008), this study focused on the use of four hypervariable regions (V1, V2, V3, and V6) to determine genetic relatedness of *E. coli* isolates from beach sands and adjacent beach waters.

The aim of this project was two-fold. The first aim was to determine if there was a genetic difference between the *E. coli* populations in water and beach sands six meters up from the swash zone. To do this, the 16S rRNA genes of 173 different isolates of biochemically identified *E. coli* cultures were sequenced to determine the relative genetic relatedness of the isolates from the beach sand to the isolates from the water. The second goal was to determine if those *E. coli* in beach sands and in the adjacent waters were carrying the *stx*$_1$, *stx*$_2$, *sfa/foc* or *papC* genes. The presence of these four toxin genes would indicate that there are potential pathogens in these external environments that could cause public health problems.

2.0 Materials and Methods

2.1 Beach sand and water collection:

Sand and water samples were collected from Bennett Beach, a public beach in Erie County, NY, on three different occasions during the summer of 2016 (from a previous project): May 25, June 29, and July 26; and on two different occasions during the summer of 2017: June 21 and August 2. There were four different collection sites, two water sites and two sand sites. Sites 2

(water) and 13 (sand) were located on the South (left) side and sites 3 (water) and 9 (sand) were located on the North (right) side of the beach looking lake-ward (Figure 1). During sampling select environmental parameters were recorded: water and air temperatures, rainfall within the previous three and/or seven days, number of waterfowl on the beach, and number of people on the beach [Appendix D].

In 2016, sterile garden spades were used to collect beach sand approximately six-m up beach from the swash zone. The upper two cm of the sand were removed with a sterile spatula and sand was then collected down to the ten-cm mark based on the markings of the spades and placed inside sterile Whirl-Pak® bags (Nasco).

In 2017, sand samples were collected using a core method. Using sterile razor blades, the end of a 60cc syringe was cut off aseptically in a Biosafety cabinet and placed back into the original packaging until on site at the beach. As in 2016, the top two cm of sand were removed, the plunger was removed from the syringe, and a core sample was taken to fill up the tube of the syringe. The filled syringe was placed inside sterile Whirl-Pak® bags (Nasco) to contain anything that would potentially fall out. All samples were stored on ice until processing.

Water samples from both years were collected in sterile one-L Corning bottles containing 0.83 mL of a 10% sodium thiosulfate solution to neutralize chlorine. Wading out to one-m depth in the water column and facing lake-ward, the bottle was submerged approximately 45 cm. The cap was removed

underwater and the bottle filled completely and then recapped while still submerged. The water was stored on ice until processing. All samples were processed within four hours of collection.

2.2 Beach sand processing:

In 2016, 50g of sand were aseptically measured out in a Biosafety cabinet and placed into a sterile 500-mL Gibco bottle containing a stir bar. In 2017, the sample was sifted through a clean and ethanol sterilized 4mm sieve to remove all large objects from the sample. Fifty-g of sieved sand were then placed into a sterile 500-mL Corning bottle containing a stir bar. Four-hundred and fifty mL of sterile 18MΩ (Millipore) water was added to each bottle to create a one to ten dilution. The soil and water were stirred at full speed for two min on a magnetic stir plate and then allowed to settle for three min.

2.3 IDEXX Most Probable Number (MPN) processing:

Using the standardized method 9223B of the National Environmental Methods Index (NEMI) (https://www.nemi.gov/methods/method_summary/5583/), 100-mL of water from the lake or the agitated sand solution was added to an IDEXX (Westbrook, ME) sample bottle containing sodium thiosulfate (cat #WV120PET-200) for chlorine inactivation. A pouch of IDEXX Colilert 18 (cat #WP100I-18) was added to the water and agitated until completely dissolved. Two drops of IDEXX AntiFoaming agent (cat #WAFDB) were added to the bottles to eliminate bubbles. The entire volume was poured into a 97-well IDEXX tray (cat #WQT2K) and then the tray was resealed using the IDEXX Quantitray Plus sealer (cat #WQTSPLUS). Samples were incubated at 44.5°C in a circulating

water bath (ThermoScientific Lindberg Blue M) for 18-22 hr. All wells were scored as either positive for Fecal Coliforms (FC) if the wells were yellow or negative for FC if there was no color change. The yellow wells also were scored as positive for *E. coli* if they demonstrated blue fluorescence under long wavelength (366nm) UV light. The numbers of positive wells were compared to a published MPN chart (IDEXX, Westbrook, Maine) to determine the number of FC and *E. coli* per 100-mL of water or 100-g of sand.

2.4 Sample processing:

Selected IDEXX tray wells that scored as positive for *E. coli* (yellow with blue fluorescence) were subsampled to isolate *E. coli* for subsequent 16S rRNA analysis. A sterile one-mL tuberculin syringe was used to aseptically aspirate 0.8-mL of liquid from wells scored as positive for FC and *E. coli*. This was added to sterile glycerol (20% final concentration) and frozen at -80°C until further processing. Since there was a possibility of mixed cultures in the IDEXX positive wells, the frozen samples were thawed on ice and resuscitated for *E. coli* isolation and identification.

All samples collected in 2016 were for a different study and as a consequence not many *E. coli* positive IDEXX wells were selected from dates I (May 25, 2016) and II (June 29, 2016). On sample date III (July 26, 2016), all *E. coli* positive IDEXX wells from each site were subsampled to increase the number of *E. coli* isolates from that previous study. On sample date IV (June 21, 2017), 20 IDEXX wells that were positive for *E. coli* were randomly subsampled for each site. However, for sample date V (August 2, 2017), there were not 20

positive wells in the IDEXX trays for every site. Therefore, all positive *E. coli* samples were saved and used from each site to ensure the number of isolates would be large enough to provide meaningful comparisons (Table 1).

The thawed samples were then streaked onto MacConkey agar (Difco) for isolation and putative *E. coli* identification. Colonies that exhibited a pink color, indicating the fermentation of lactose, on MacConkey agar were picked and put onto nutrient agar slants (Difco) to be used as stocks for further biochemical testing. The pure cultures were then added to 1% tryptone broth to test for the presence of tryptophanase (the indole test) and to urea broth to test for the production of urease. Those isolates that were indole positive and urease negative were then added to EC broth with MUG (4-methylumbelliferyl-β-D-glucuronide) (Difco) for 18-22h in a 44.5°C circulating waterbath. *Escherichia coli* contains the β-galactosidase enzyme and therefore breaks down lactose to glucose and galactose, which is then metabolized and creates gas as an end product that will accumulate in an inverted Durham tube. Most *E. coli* also contain β-glucuronidase, an enzyme that will hydrolyze MUG to 4-methylumbelliferone, which fluoresces as a blue-white light under long wavelength (366nm) ultraviolet light. All selected colonies that were positive for gas production and fluorescence as well as lactose fermentation, urease negative, and indole positive were considered to be *E. coli*. All samples biochemically determined to be *E. coli* [Table 2] were grown in nutrient broth

(Gibco) and then frozen with 20% glycerol as a pure culture for further processing.

2.5 Cell lysis and PCR:

Pure cultures of biochemically identified *E. coli* were lysed in order to collect the genomic DNA (gDNA) of the cells. In brief, one mL of an overnight *E. coli* LB broth culture (at 37°C with 200rpm agitation) was added to a sterile 1.5-mL Eppendorf tube for processing. The sample was centrifuged for five min at 13,300xg at room temperature (RT) after which the supernatant was discarded. Four-hundred μL of Fast Lysis buffer (*mericon*® DNA Bacteria Plus kit, cat no. 69534) were added to the cell pellet and the pellet was resuspended. This entire cell mixture was then added to a pathogen lysis tube supplied with the kit (containing ~0.5mm glass beads) and vortexed for 10 min at maximum speed. The vortexed tubes were then centrifuged at RT for five min at 13,300xg. At least 100 μL were taken from the lysis tubes and placed into a new set of microcentrifuge tubes. The gDNA was then quantitated using the Implen NanoPhotometer™ (Implen P-360, 7122 V2.3.1), diluted with nuclease-free water to a concentration of 10 ng per μL, and immediately used for polymerase chain reaction (PCR) for the 16S rRNA gene as well as all four toxin genes (see below).

Using the diluted gDNA, the 16S rDNA was amplified by PCR using published primers: 8F and 1492R(s) (Eden et al. 1991) [Appendix A]. By using

the 8F and 1492R(s) primers, the entire ~1500bp 16S rRNA gene was amplified with a thermocycler (Bio-Rad T100 ThermoCycler) [Appendix B].

2.6 Sequencing:

After amplification of the 16S rRNA gene, the PCR product was quantitated using the Qubit fluorometer (Invitrogen, Qubit 2.0 Fluorometer, V3.10) with the broad range protocol and reagents. The samples were then diluted with nuclease-free water to the appropriate concentration of four ng per μL in a total volume of 10 μL for all beach *E. coli* samples as well as two control (i.e. known) *E. coli* cultures (ECOR60 and *E. coli* K-12), and sent to GenScript (Piscataway, NJ, www.genscript.com) for sequencing using the 8F primer [Appendix A].

2.7 Sequence preparation:

All sequences were retrieved from the GenScript website in .ab1 file format. In order to compare the sequences, they were individually converted into FASTA file format using the online converter at sequenceconversion.bugaco.com/converter/biology/sequences. The FASTA files were opened and saved using the Notepad program in order to keep the file as text only. As a control for sequence quality, the free program Chromas (v. 2.6.5) was used to visualize the electropherograms and trim any low quality reads. As another level of identification, all files were BLAST searched on the NCBI website (http://www.ncbi.nlm.nih.gov) directly from the Chromas program to verify that they were *E. coli*.

2.8 Multiple sequence alignment:

Using the statistical program, R (https://www.r-project.org, V. 3.5.2 "Eggshell Igloo"), the sequences of all 173 beach samples, two control *E. coli* strains and three *E. coli* 16S rDNA sequences selected from the NCBI website (www.ncbi.nlm.nih.gov, accession numbers: MH782105.1, NR 024570.1, and AR657803.1) were aligned and compared using the add-on msa package (Bodenhofer et al. 2015). All parameters used were the default parameters in the ClustalOmega format to obtain relatedness for all sequences. Knowing that specific variable regions are more applicable for certain levels of classification (Chakravorty et al. 2007, Yang et al. 2016) variable regions one, two, three, and six were analyzed individually to determine if they could be used to differentiate between the sand isolates and the water isolates.

In order to determine the similarity of the samples for variable regions one and two, the default ClustalOmega format was used and specific base pair ranges were parsed out of the aligned sequences and compared [Appendix C]. Variable regions three and six were compared using the more stringent MUSCLE default parameters (Edgar 2004). Therefore, in order to compare the differences of those variable regions, any sequence that was less than 960bp long (n=65) was removed and the remaining sequences were aligned and compared using the default Muscle alignment parameters in R [Appendix C]. The command line program TeXstudio (v. 2.12.14) was also utilized to create a graphical representation of the level of conservation of aligned bases of the V6 region of the 16S rRNA gene [Appendix G].

2.9 Toxin gene amplification:

The toxin genes of interest in this study were *stx*$_1$, *stx*$_2$, *sfa/foc*, and *papC*. Using published primers (Escobar-Paramo et al. 2004, Zhi et al. 2016) [Appendix A] PCR was performed on all beach *E. coli* strains isolated in this study to look for the presence of those toxin genes [Appendix B]. Analysis of the presence of the toxin genes was recorded as percentages of all *E. coli* isolates.

All PCR reactions included positive and negative controls. Genomic DNA for the *stx*$_1$ and *stx*$_2$ genes was purchased through ATCC (cat# BAA-2196D-5) and for the *sfa/foc* and *papC* genes *E. coli* reference strain, ECOR60, was purchased from the STEC Center at Michigan State University (www.shigatox.net). Two negative controls (*E. coli* strain K-12 [ATCC # 25404] and nuclease-free water) were utilized.

3.0 Results

3.1 Sample collection:

Since the study samples were from the external environment, many factors could potentially play a role in the number of positive samples that were collected. On average, the air temperature was 21.8°C and the water temperature was 21.6°C. There was an average of 27.4 waterfowl (mostly gulls) and 8 people on the beach for all sample dates [Appendix D].

The uneven distribution in samples from 2016 versus those from 2017, with 2016 making up only 15.6% of the total number of samples and 2017

making up the remaining 84.4% (Table 2), did not allow for an accurate comparison across two summers therefore the remainder of this study focused on determining if there was a difference of *E. coli* populations among four collection sites.

3.2 Enumerations:

Using the IDEXX identification system (standard method 9223B), the MPN of *E. coli* positive wells ranged from one organism/100mL of sample at site 2 (water) on May 25, 2016 to over 28,000 organisms/100g of sample at site 9 (sand) on July 26, 2016 (Figure 2a). The ratio of *E. coli* to fecal coliforms in most water and sand samples ranged between 60% and 80%. However, several samples had 100% *E. coli* to fecal coliforms ratio, while the lowest percentage was 4% from site 13 (sand) on June 29, 2016 (Figure 2b).

In all, 173 *E. coli* isolates from the beach environment: two from May 25, 2016; six from June 29, 2016; 19 from July 26, 2016; 75 from June 21, 2017; and 71 from August 2, 2017, were used in the analysis of the 16S rRNA gene as well as the presence of the toxin genes (Table 2). The water sites had 32 and 35 (18.5% and 20.2%) *E. coli*, respectively and the sand sites had 61 and 45 (35.3% and 26%) *E. coli*, respectively (Figure 3). Water isolates made up approximately one-third of the tested samples so the discrepancy in numbers should be considered during analysis.

3.3 16S rRNA gene analysis:

The full 16S rRNA gene phylogenetic alignment of all 178 sequences is found in Appendix E which indicates that there are no genetic differences

between the water sites (2 and 3) and the sand sites (9 and 13). Phylogenetic trees were also constructed for variable regions one and two using the ClustalOmega multiple sequence alignment parameters as well [Appendices E2 and E3]. However, phylogenetic trees for variable regions three and six were formed using the Muscle alignment program parameters in R-studio [Appendices E4 and E5]. The variable region 6 showed the most variation, but as is shown in Appendix G the level of conservation of bases is still rather high. None of the phylogenetic trees constructed indicated a genetic difference between water (sites 2 & 3) and sand (sites 9 & 13) isolates.

Since the shorter sequences affected the distance matrix calculated in the R program code, a more accurate representation of the relationship of the samples was developed by removing all sequences less that 960bp (n=65). Therefore, all sequences over 960bp (n=110) as well as the three sequences retrieved from NCBI were analyzed using the Muscle default parameters encompassing V1 to V6 regions (Figure 4). As a further breakdown, all samples from the North side of the beach (sites 3 and 9) [Appendix E6], those from the South side of the beach (sites 2 and 13) [Appendix E7], the sand isolates only (sites 9 and 13) [Appendix E8], and the water isolates only (sites 2 and 3) [Appendix E9] which were over 960bp were aligned using the Muscle multiple sequence alignment parameters encompassing variable regions one through six.

To compare all of the sequences to one another, ClustalOmega was utilized due to its accuracy and speed over other multiple sequence alignment programs such as Kalign, MAFFT, and T-Coffee since the number of sequences

was small (t<200) (Sievers et al. 2011). In the ClustalOmega alignment and successive phylogenetic trees, there was no apparent pattern of genetic differences from water samples (sites 2 and 3) and sand samples (sites 9 and 13) [Appendix E1]. As an additional level of comparison, certain variable regions were analyzed individually, and it was shown that variable regions one and six were able to establish different genetic signatures and create phylogenetic trees with some level of variation [Appendices E2 and E5]. In analyzing variable regions two and three, there was no difference between any of the 178 samples that were aligned for variable region two [Appendix E3] and the 113 samples aligned for variable region three [Appendix E4]. By using the more stringent parameters of the MUSCLE multiple sequence alignment program (Edgar 2004), the analysis of all of the combined variable regions available (one through six) showed the most promising phylogenetic tree (Figure 4).

3.4 Toxin gene analysis:

All 173 beach environment *E. coli* gDNA samples were subjected to PCR to determine if they contained the toxin genes: *papC*, *sfa/foc*, stx_1, and stx_2 using the primers and conditions listed in Appendices A and B. The stx_1 and stx_2 genes were not found in any of the 173 isolates (data not shown). It was also determined that the *papC* gene was present in 5.2% (n=9) of the isolates, and that the *sfa/foc* gene was present in 7.5% (n=13) of the isolates (Figure 5a). It is important to note that some bacteria contained both genes and others contained only one or the other. There were five isolates (2.9%) that contained only the

papC gene, nine (5.2%) that contained only the *sfa/foc* gene, and four (2.3%) that contained both genes (Figure 5b).

Initially, the annealing temperature for PCR of the *papC* and *sfa/foc* genes was 63°C. It was determined that there was non-specific binding of the primers therefore the annealing temperature was increased by 2°C to 65°C for the final data presentation. Since increasing the annealing temperature increases the stringency of the PCR and reduces the number of false-positive results (Ishii and Fukui 2001), it was not necessary to repeat all of the samples that originally tested negative for those toxin genes.

4.0 Discussion

In general, every strain of *E. coli* can be categorized into one of seven different phylogenetic groups based upon a multiplex PCR system that determines the genetic background of the strain (Clermont et al. 2013). Genetic testing has categorized the highly virulent EHEC O157:H7 strains, which are known to originate from ruminants (Lim et al. 2010), mostly in phylogroup E with a few being placed in A and B1 (Clermont et al. 2013). In addition, Tymensen et al. (2015) concluded that naturalized *E. coli* populations fell within the B1 phylogroup as well. This suggests that the genetic background between certain extremely virulent strains of *E. coli* and the naturalized populations are similar. In 2000, Clermont et al. determined that virulent extraintestinal pathogenic *E. coli* were placed in the phylogroup B2 and occasionally in phylogroup D. This was

further corroborated by Johnson et al. (2017), who also found in the same study that ExPEC strains were found in many surface waters as determined by the presence of the *papC* and *sfa/foc* genes among others. Based upon these findings, ExPEC and EHEC strains do not have the same genetic background.

Since variable regions two and three are utilized to differentiate organisms to the genus level (Chakravorty et al. 2007), it is not surprising that the phylogenetic trees constructed using these variable regions [Appendices E3 and E4] did not demonstrate strain differences because all of the organisms analyzed belong to the genus *Escherichia*. Also because V3 region is highest in single nucleotide polymorphisms (SNPs) (Chakravorty et al. 2007), the one outlying sample (V13I) [Appendix E4] could be explained by a single nucleotide difference.

Given the transient nature of water, it was expected that there would be fewer *E. coli* found in the water versus the sand as has been previously shown (Burton et al. 1987, Whitman et al. 2003). The pattern observed in this study agreed with other investigators who have reported differences of one to two orders of magnitude higher levels of *E. coli* in sands versus adjacent water (An et al. 2002, Alm et al. 2003, Whitman and Nevers 2003, Whitman et al. 2003, Abdelzaher et al. 2010) [Figure 2a and Appendix F].

Considering that all *E. coli* originates from an intestinal environment, there is potential for strains of different genetic backgrounds to exchange genetic information via horizontal transfer. Enterohemorrhagic *E. coli* (EHEC) O157:H7 is the most common intestinal infection that results in hemolytic uremic syndrome

(HUS) (Bitzan et al. 1993). However in Germany in 2011 there was an outbreak of *E. coli* infections resulting in HUS from strain O104:H4 that is usually characterized as a different pathotype: enteroaggregative *E. coli* (EAEC) (Navarro-Garcia 2014). This novel hybrid was determined to have a prophage encoding Shiga-like toxin that is characteristic of EHEC, which causes diarrheal diseases that can produce HUS (Bielaszewska et al. 2011, Mellmann et al. 2011, Rasko et al. 2011). This hybrid, which is considerably more virulent than its EHEC cousin, acquired the prophage via horizontal gene transfer (Rasko et al. 2011). Recently it has been shown that EHEC has formed hybrid pathotypes with enteropathogenic *E. coli* (EPEC) as well (van Hoek et al. 2019). While hybrid pathotypes of EHEC and EPEC or EAEC have been shown to exist, it is seemingly rare for hybrids of extraintestinal pathogenic *E. coli* (ExPEC) and EHEC to exist. However, in January of 2019, De Rauw et al. established the emergence of one such hybrid in which the source of infection is unknown. The development of HUS from urinary tract infections (UTI), while uncommon, has been recorded as well. In a case report in 2001, Hogan et al. reported of several cases of HUS resulting from UTI in which death of a few patients had resulted, as well as many long-term effects on the health of the individuals. While this study did not apprise on the presence of the *papC* or *sfa/foc* genes, it is probable that the *E. coli* that originally caused those UTI were of the ExPEC pathotype. Hogan et al. (2001) suggested that any future cases of UTI should be screened for Shiga toxin production especially if the patients exhibit HUS. These examples indicate that the sharing of virulent characteristics is a common phenomenon and

is of great concern to public health. Indeed, Wright (2010) reported that a clinical isolate of *A. baumannii* exhibited resistance characteristics from an historical horizontal gene transfer event that initiated in a soil environment. However, that transfer appears to be plasmid-based and the location of the *papC* and *sfa/foc* genes are on pathogenicity islands on the bacterial chromosome (Najafi et al. 2018) which would decrease the possibility of horizontal gene transfer.

Because of the nature of the *E. coli* numbers in sand and the relative stability of the sand environment, it was expected that there would be a higher percentage of *E. coli* strains from the sand sites, 9 and 13, that would show positive results for the toxin genes versus the water sites, 2 and 3. However, unexpectedly, site 3 (up-beach water) had higher numbers of toxin gene positive results relative to the sand. Thirteen of the 18 samples that tested positive came from site 3 while four came from site 9 (up-beach sand adjacent to site 3) and only one came from site 13 (down-beach sand) (Figure 6). Site 2 showed no positive samples for toxin genes. One possible explanation for this is that the water site 3 is closer to the Big Sister creek output into Lake Erie than the water site 2 (Figure 1). In a previous study conducted on Bennett Beach, the MPN of *E. coli* water isolates from Big Sister Creek averaged 1049.5 CFU/100mL over three sample dates (unpublished data). This average suggests the strong possibility that the E*. coli* that were found in the water site (site 3) in this study could, in part, originate from Big Sister Creek. However, it cannot be definitively determined that Big Sister creek is the source of the influx of *papC* and *sfa/foc*

positive *E. coli* isolates since there were no samples taken from that creek in this study.

This study found the presence of markers that are associated with ExPEC in beach sand and water but did not demonstrate the presence of toxin genes associated with EHEC. There is no indication that the *E. coli* collected from the beach sands and waters in this study were of a hybrid pathotype. This agrees with Bauer and Alm (2012) who found that the presence of stx_1 and stx_2 toxin genes were uncommon in sand-isolated *E. coli* while attachment genes such as *papC* and *sfa/foc* were frequently detected. However, the phenomenon of hybrid pathotypes should be considered during future studies.

The presence of *papC* in *E. coli*-caused UTI samples is between 60-90% (Kallenius et al. 1981, Hultgren et al. 1991, Johnson 1991, Hacker 1992), and *sfa/foc* presence ranges from 13-30% (Pere et al. 1985, Kunin et al. 1993) as stated previously. Rahdar et al. (2015) also found that the clinical isolates of *E. coli* UTI samples contained high percentages of *papC* and *sfa/foc* genes. This study showed the percentage of *papC* was only 5.2% and *sfa/foc* was only 7.5%. However, the samples collected in this study are from an external environmental source which could explain the discrepancy in percent positive samples. Preliminary data from an ongoing project involving 100 known *E. coli* UTI samples suggests that *papC* is present in approximately 75% and *sfa/foc* is present in approximately 30% of the samples. These data are supportive of the average percent positive in those previous studies.

The current study was designed to evaluate the possibility of discerning a genetic difference in *E. coli* isolates from beach sands and adjacent waters over two years and four sites based on 16S rRNA gene sequences, and also to determine if the presence of certain toxin genes could be detected in those strains. Toxin genes that are associated with extraintestinal infections were found to be present in both sand and water isolates. However, there is no distinguishing genetic pattern based on the 16S rRNA gene sequences obtained in this study of samples from sites 2 and 3 (water) to sites 9 and 13 (sand) based upon the 16S rRNA gene. This does not necessarily mean that there is no difference in the genetic background of samples from water versus samples from sand, however. It should be noted that none of the samples had the full 16S rRNA gene sequenced. Having the entire gene sequenced with high accuracy or using alternative methods could potentially lead to the ability to notice differences between the water and sand populations of *E. coli*. For example, using horizontal, fluorophore-enhanced repetitive extragenic palindromic PCR (HFERP) analysis Ishii et al. (2007) identified putative naturalized populations of *E. coli* isolated from freshwater sediments and sands.

Also, there are other portions of the genome that should be considered in order to determine whether the samples are genetically distinct. The addition of genes used by Escobar-Paramo et al. (2004) (*trpA*, *trpB*, *pabB*, *putP*, *icd*,and *polB*) could aid in the ability to distinguish distinct genetic backgrounds if they are present. Another possibility is to look at the intracellular concentration of the sugar trehalose which has been determined to help *E. coli* resist desiccation

(Zhang and Yan 2012). Using any of the above genetic markers or a combination of them in addition to the 16S rRNA gene will increase the likelihood that any differences in genetic background could be discovered.

While the 16S rRNA gene is traditionally used as a chronometer for evolutionary relationships among species (Woese et al. 1990), it should be combined with other genes and tests to determine if there is a genetic distinction among *E. coli*. Also, it was shown in this study that toxin genes can exist in *E. coli* collected from an external environment; however, it is not known at this time how long the cultures that harbor these toxin genes will persist outside of the normal animal host. The presence of the genes found in *E. coli*-caused urinary tract infections in beach waters and sands could lead to an increased incidence of UTIs seen in recreational beach users and potentially alter the current practice of using *E. coli* as a FIB for recent contamination events for the purposes of beach closures.

5.0 Figures and Tables:

Figure 1: Bennett Beach aerial photograph with sample sites identified. Sample sites two and three were taken approximately six-m offshore from a depth of 45-cm in the water column. Samples nine and 13 were collected six-m inland from the swash zone at a depth of 10-cm from the top layer of the sand.

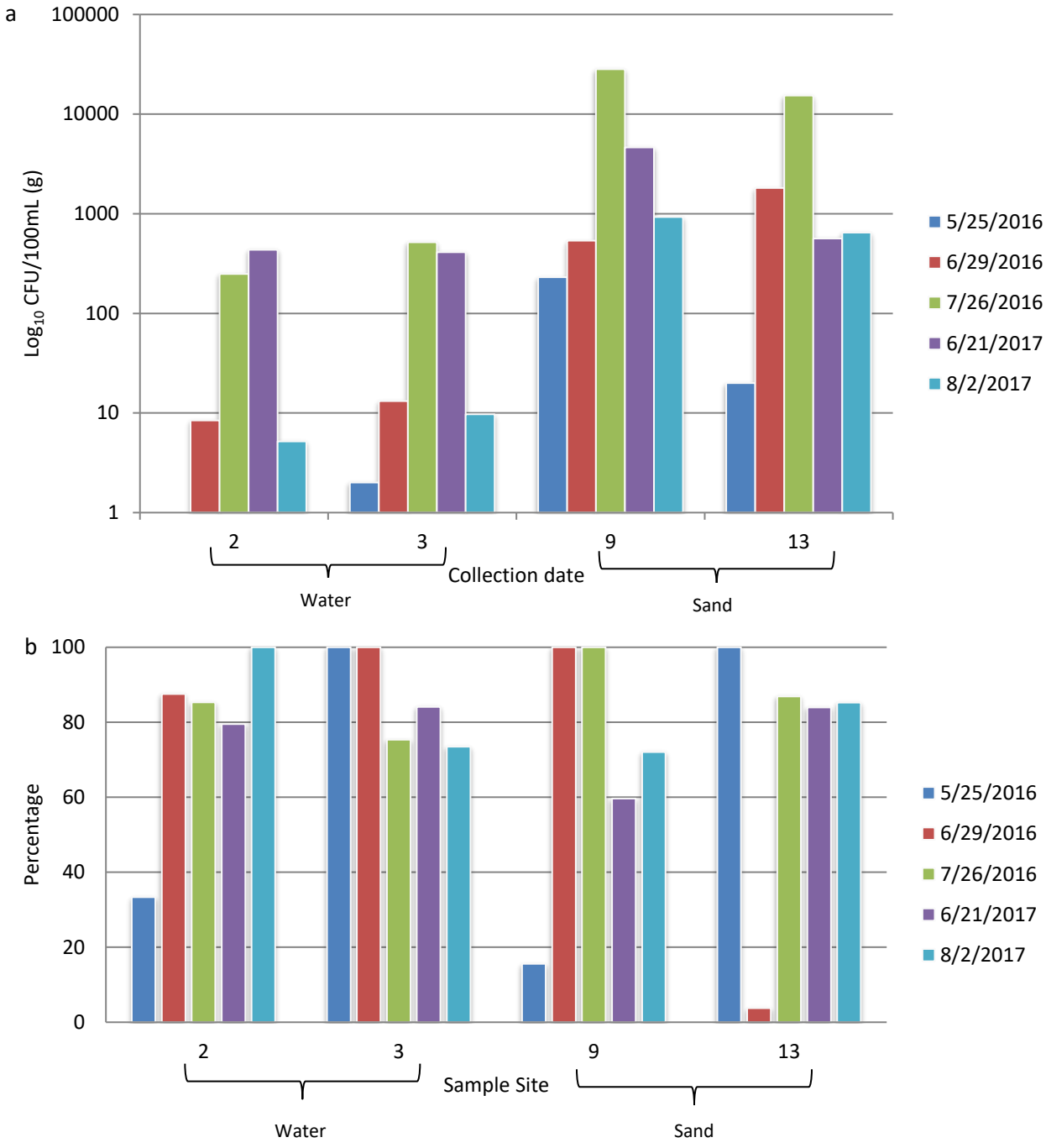


Figure 2: a) Log_{10} of *E. coli* CFU/100mL by site. Wells were scored as positive for *E. coli* if the contents turned yellow (indicating the activity of β-galactosidase) and also had a blue florescence when exposed to 366nm UV light (indicating the activity of β-glucuronidase). b) Percent *E. coli* of Fecal Coliforms by site. The number of *E. coli* positive wells (yellow and fluorescent wells) was divided by the number of fecal coliforms positive wells (yellow only wells) to determine overall percentages.

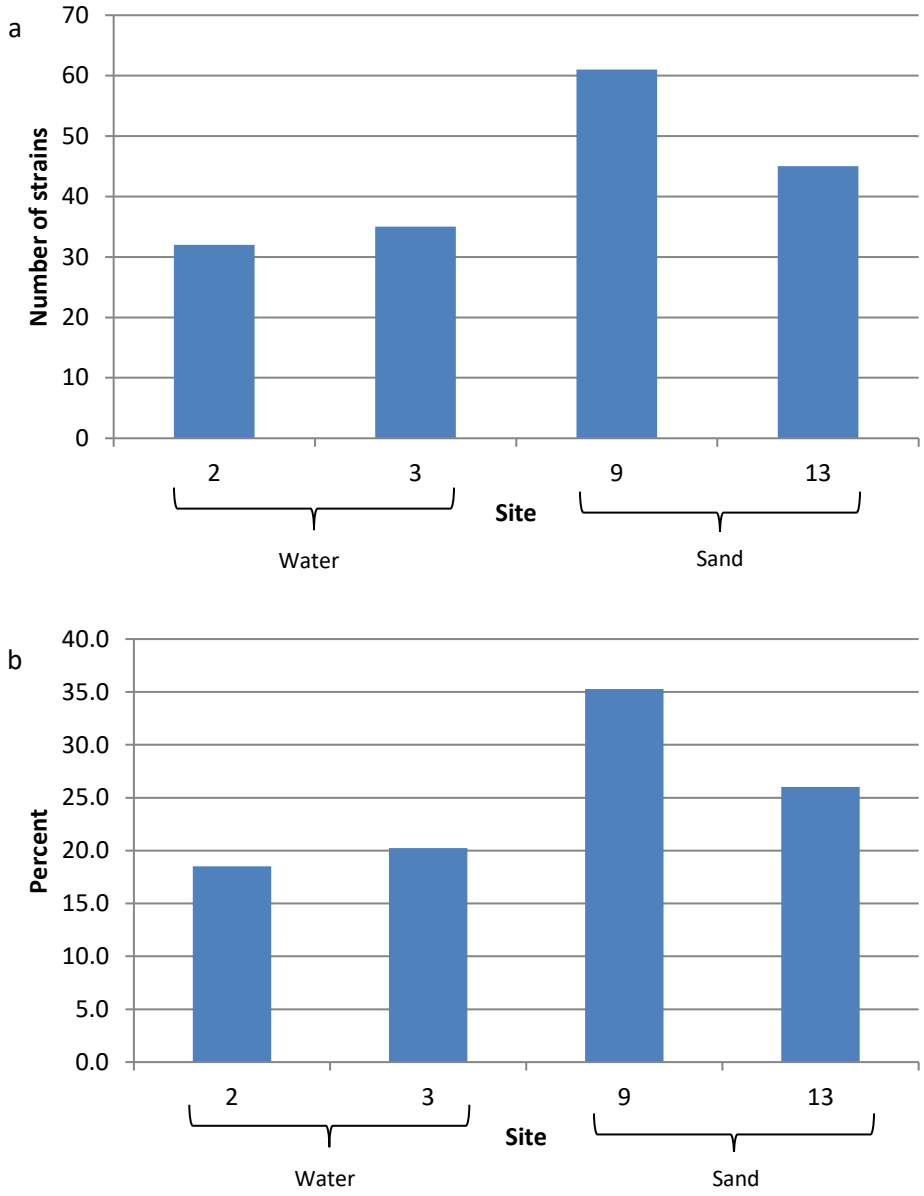


Figure 3: a) Actual number of strains analyzed by site. b) Percent of overall strains by site. The number of strains selected from each site was divided by the total number of strains analyzed (n=173) to determine the percentage of strains from each site.

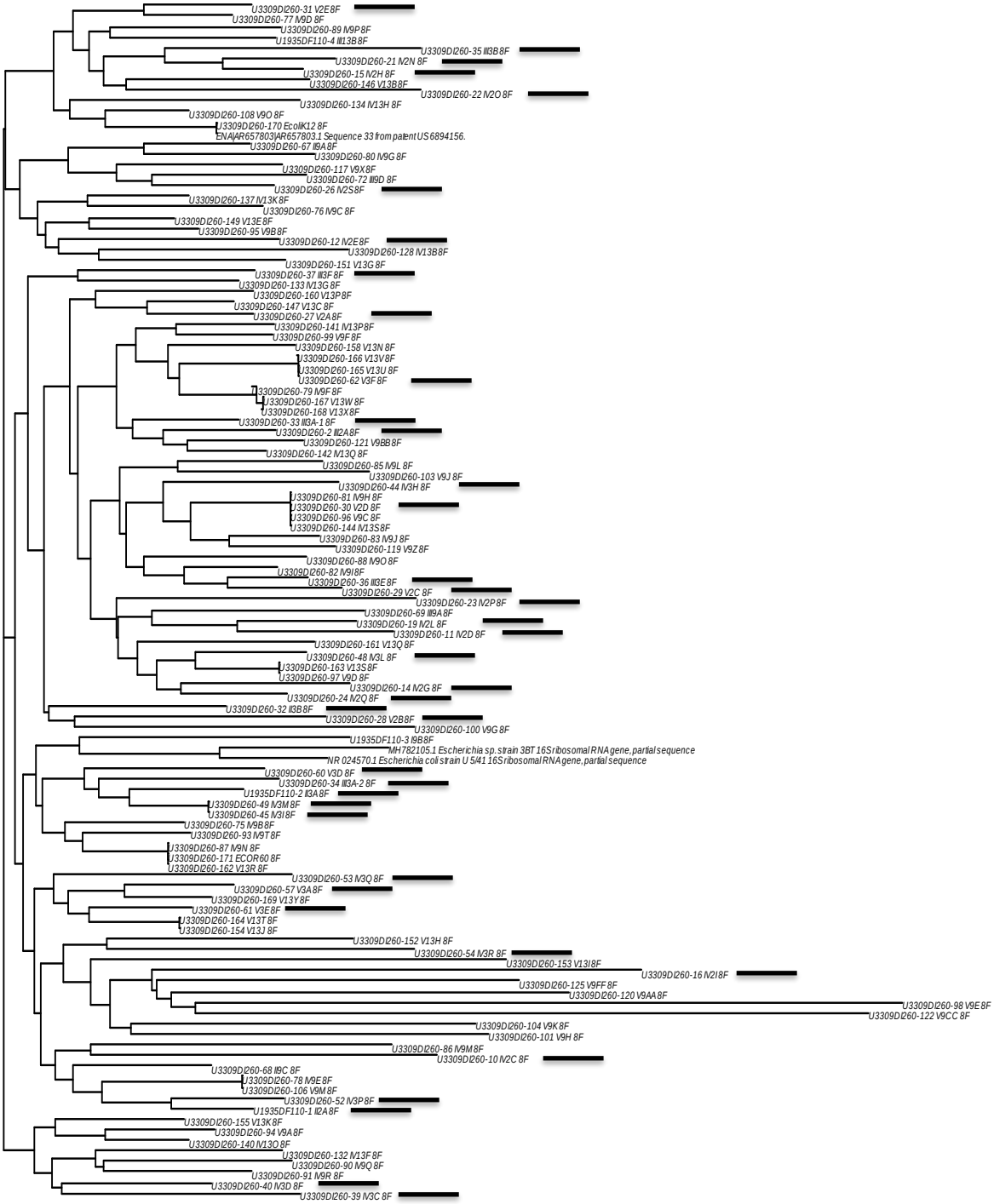


Figure 4: Phylogenetic tree of variable regions 1 through 6 using all samples over 960bp (n=113) analyzed using the MUSCLE multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 1043 were parsed out and this tree was generated. The thick black line indicates the phylogenetic location of all water samples showing no pattern of relatedness.

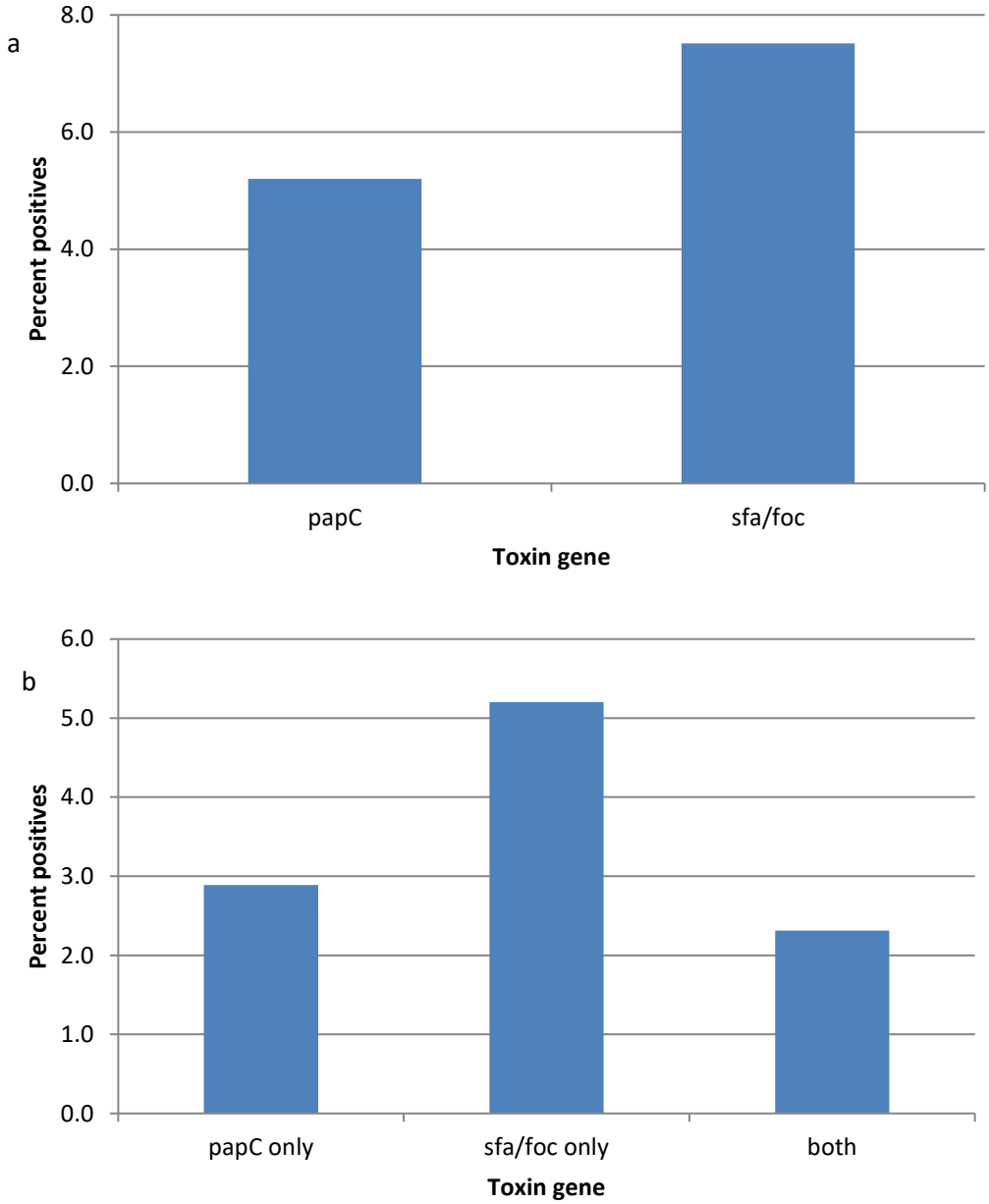


Figure 5: a) Percent positive toxin genes overall. b) Percent of toxin gene combinations. Positive samples of all combinations (total *papC*, total *sfa/foc*, *papC* only, *sfa/foc* only, and both *papC* and *sfa/foc* positive) were divided by the total number of samples (n=173) analyzed for the toxin genes.

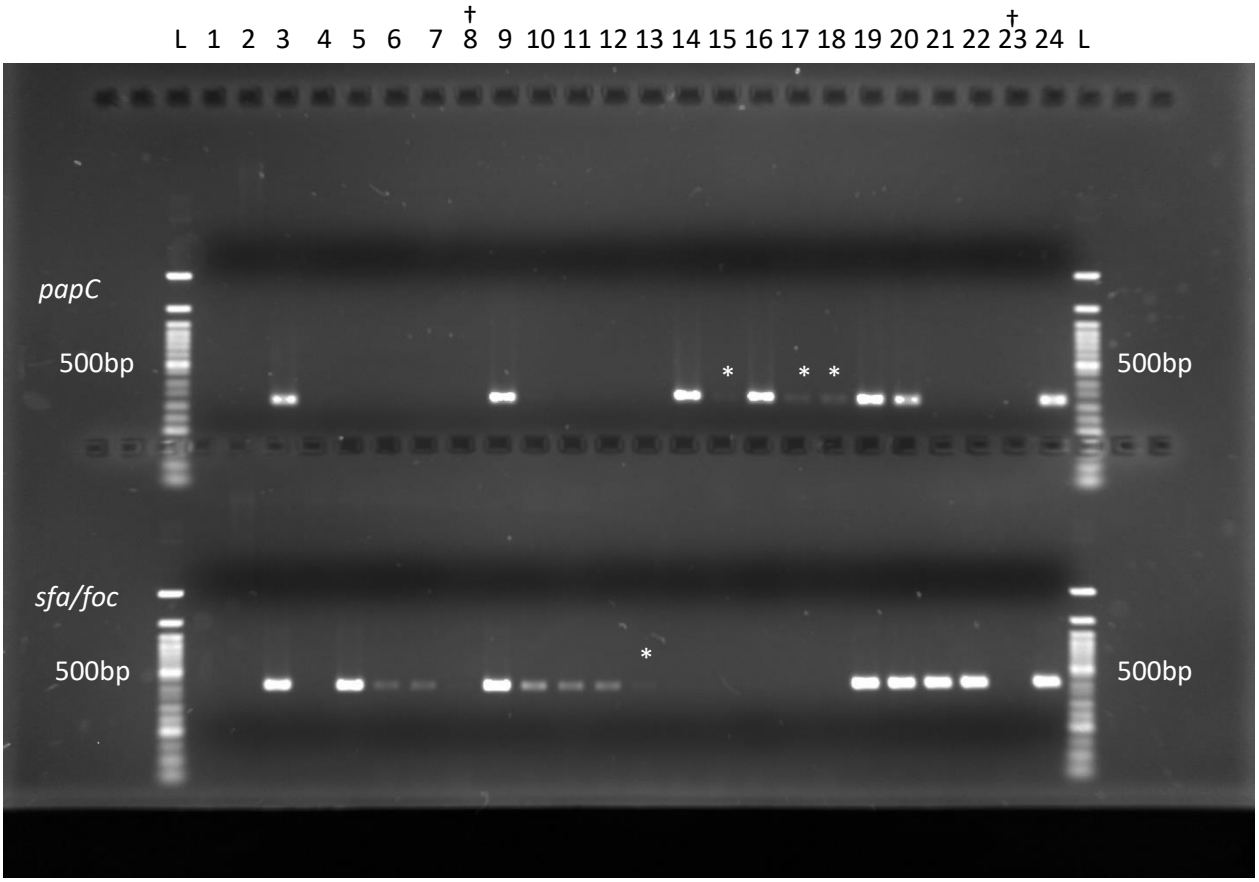


Figure 6: 2% TAE agarose gel of *papC* and *sfa/foc* positive samples. The *papC* gene is 328bp and the *sfa/foc* gene is 407bp. Ladder is the Quick-Load Purple 50bp DNA ladder from New England Biolabs (cat# N0556S) with the 500bp marker identified. Samples: L=DNA ladder, 1=dH_2O (-) control, 2=*stx* gDNA (-) control, 3=ECOR60 (+) control, 4=*E. coli* K-12 (-) control, 5=II3A, 6=III3E, 7=IV3B, 8=IV3E†, 9=IV3I, 10=IV3J, 11=IV3K, 12=IV3L, 13=IV3M, 14=V3B, 15=V3C, 16=V3G, 17=V3H, 18=V3I, 19=I9B-1, 20=I9C, 21=IV9C, 22=V9J, 23=III13D†, 24=IV13D, and L=DNA ladder. The * indicates faint bands. The † indicates samples that originally had faint bands which disappeared when the annealing temperature was raised.

Year	Collection Date	Collection Sites				
		2 (w)	3 (w)	9 (s)	13 (s)	Totals
2016	I	0	0	4	0	31
	II	2	2	2	1	
	III	6	5	5	4	
2017	IV	20	20	20	20	151
	V	5	9	32	25	
	Totals:	33	36	63	50	182

Table 1: Number of isolates collected from IDEXX trays as putative *E. coli*. Collection dates: I = May 25, 2016; II = June 29, 2016; III = July 26, 2016; IV = June 21, 2017; V = August 2, 2017;
w = water; s = sand

Year	Collection Date	Collection Sites				
		2 (w)	3 (w)	9 (s)	13 (s)	Totals
2016	I	0	0	2	0	27
	II	2	2	2	0	
	III	6	5	5	3	
2017	IV	19	19	20	17	146
	V	5	9	32	25	
	Totals:	32	35	61	45	173

Table 2: Number of isolates collected from IDEXX trays biochemically identified as *E. coli* and selected to proceed in this study. Collection dates: I = May 25, 2016; II = June 29, 2016; III = July 26, 2016; IV = June 21, 2017; V = August 2, 2017; w = water; s = sand

6.0 Bibliography:

Abdelzaher, A. M., M. E. Wright, C. Ortega, H. M. Solo-Gabriele, G. Miller, S. Elmir, X. Newman, P. Shih, J. A. Bonilla, T. D. Bonilla, C. J. Palmer, T. Scott, J. Lukasik, V. J. Harwood, S. McQuaig, C. Sinigalliano, M. Gidley, L. R. W. Plano, X. F. Zhu, J. D. Wang, and L. E. Fleming. 2010. Presence of pathogens and indicator microbes at a non-point source subtropical recreational marine beach. Applied and Environmental Microbiology **76**:724-732.

Alm, E. W., J. Burke, and A. Spain. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Research **37**:3978-3982.

An, Y. J., D. H. Kampbell, and G. P. Breidenbach. 2002. *Escherichia coli* and total coliforms in water and sediments at lake marinas. Environmental Pollution **120**:771-778.

Andini, N., B. Wang, P. Athamanolap, J. Hardick, B. J. Masek, S. Thair, A. Hu, G. Avornu, S. Peterson, S. Cogill, R. E. Rothman, K. C. Carroll, C. A. Gaydos, J. T. H. Wang, S. Batzoglou, and S. Yang. 2017. Microbial typing by machine learned DNA melt signatures. Scientific Reports **7**.

Antony, R., A. Sanyal, N. Kapse, P. K. Dhakephalkar, M. Thamban, and S. Nair. 2016. Microbial communities associated with Antarctic snow pack and their biogeochemical implications. Microbiological Research **192**:192-202.

Bauer, L., and E. Alm. 2012. *Escherichia coli* toxin and attachment genes in sand at Great Lakes recreational beaches. Journal of Great Lakes Research **38**:129-133.

Bernhard, A. E., and K. G. Field. 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Applied and Environmental Microbiology **66**:1587-1594.

Bernhard, A. E., and K. G. Field. 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. Applied and Environmental Microbiology **66**:4571-4574.

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

Bitzan, M., K. Ludwig, M. Klemt, H. Konig, J. Buren, and D. E. Mullerwiefel. 1993. The role of *Escherichia coli* O157 infections in the classical (enteropathic) hemolytic uremic syndrome - results of a central European, multicenter study. Epidemiology and Infection **110**:183-196.

Bodenhofer, U., E. Bonatesta, C. Horejs-Kainrath, and S. Hochreiter. 2015. msa: an R package for multiple sequence alignment. Bioinformatics **31**:3997-3999.

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

Bonilla, T. D., K. Nowosielski, M. Cuvelier, A. Hartz, M. Green, N. Esiobu, D. S. McCorquodale, J. M. Fleisher, and A. Rogerson. 2007. Prevalence and distribution of fecal indicator organisms in South Florida beach sand and preliminary assessment of health effects associated with beach sand exposure. Marine Pollution Bulletin **54**:1472-1482.

Breimer, M. E., and K. A. Karlsson. 1983. Chemical and immunological identification of glycolipid-based blood group ABH and Lewis antigens in human kidney. Biochimica Et Biophysica Acta **755**:170-177.

Brennan, F. P., V. O'Flaherty, G. Kramers, J. Grant, and K. G. Richards. 2010. Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. Applied and Environmental Microbiology **76**:1449-1455.

Brown, J. E., D. E. Griffin, S. W. Rothman, and B. P. Doctor. 1982. Purification and biological characterization of Shiga toxin from *Shigella dysenteriae* 1. Infection and Immunity **36**:996-1005.

Burton, G. A., D. Gunnison, and G. R. Lanza. 1987. Survival of pathogenic bacteria in various freshwater sediments. Applied and Environmental Microbiology **53**:633-638.

Byappanahalli, M. N., T. Yan, M. J. Hamilton, S. Ishii, R. S. Fujioka, R. L. Whitman, and M. J. Sadowsky. 2012. The population structure of *Escherichia coli* isolated from subtropical and temperate soils. Science of the Total Environment **417**:273-279.

Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin. 1982. Swimming-associated gastroenteritis and water quality. Am J Epidemiol **115**:606-616.

Calderwood, S. B., F. Auclair, A. Donohuerolfe, G. T. Keusch, and J. J. Mekalanos. 1987. Nucleotide sequence of the Shiga-like toxin genes of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **84**:4364-4368.

Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods **69**:330-339.

Chandrasekaran, R., M. J. Hamilton, P. Wang, C. Staley, S. Matteson, A. Birr, and M. J. Sadowsky. 2015. Geographic isolation of *Escherichia coli* genotypes in sediments and water of the Seven Mile Creek - A constructed riverine watershed. Science of the Total Environment **538**:78-85.

Clements, A., J. C. Young, N. Constantinou, and G. Frankel. 2012. Infection strategies of enteric pathogenic *Escherichia coli*. Gut Microbes **3**:71-87.

Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Applied and Environmental Microbiology **66**:4555-4558.

Clermont, O., J. K. Christenson, E. Denamur, and D. M. Gordon. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environmental Microbiology Reports **5**:58-65.

Coenye, T., and P. Vandamme. 2003. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. Fems Microbiology Letters **228**:45-49.

Cote, C., and S. Quessy. 2005. Persistence of *Escherichia coli* and *Salmonella* in surface soil following application of liquid hog manure for production of pickling cucumbers. Journal of Food Protection **68**:900-905.

De Rauw, K., D. Thiry, B. Caljon, M. Saulmont, J. Mainil, and D. Piérard. 2019. Characteristics of Shiga toxin producing- and enteropathogenic Escherichia coli of the emerging serotype O80:H2 isolated from humans and diarrhoeic calves in Belgium. Clinical Microbiology and Infection **25**:111.e115-111.e118.

Denamur, E., G. Lecointre, P. Darlu, O. Tenaillon, C. Acquaviva, C. Sayada, I. Sunjevaric, R. Rothstein, J. Elion, F. Taddei, M. Radman, and I. Matic. 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103**:711-721.

Deng, D. Y., N. Zhang, A. Mustapha, D. Xu, T. Wuliji, M. Farley, J. Yang, B. Hua, F. J. Liu, and G. L. Zheng. 2014. Differentiating enteric *Escherichia coli* from environmental bacteria through the putative glucosyltransferase gene (*ycjM*). Water Research **61**:224-231.

Eden, P. A., T. M. Schmidt, R. P. Blakemore, and N. R. Pace. 1991. Phylogenetic analysis of *Aquaspirillum magnetotacticum* using polymerase chain reaction-amplified 16S ribosomal-RNA-specific DNA. International Journal of Systematic Bacteriology **41**:324-325.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Molecular Biology and Evolution **21**:1085-1094.

Estrada-Garcia, T., C. Lopez-Saucedo, R. Thompson-Bonilla, M. Abonce, D. Lopez-Hernandez, J. I. Santos, J. L. Rosado, H. L. DuPont, and K. Z. Long. 2009. Association of diarrheagenic *Escherichia coli* pathotypes with infection and diarrhea among Mexican children and association of atypical enteropathogenic *E-coli* with acute diarrhea. Journal of Clinical Microbiology **47**:93-98.

Fraser, M. E., M. Fujinaga, M. M. Cherney, A. R. Melton-Celsa, E. M. Twiddy, A. D. O'Brien, and M. N. G. James. 2004. Structure of Shiga toxin type 2 (stx_2) from *Escherichia coli* O157 : H7. Journal of Biological Chemistry **279**:27511-27517.

Gordon, D. M., S. Bauer, and J. R. Johnson. 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology-Sgm **148**:1513-1522.

Guo, F., F. Ju, L. Cai, and T. Zhang. 2013. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. Plos One **8**.

Hacker, J. 1992. Role of fimbrial adhesins in the pathogenesis of *Escherichia coli* infections. Canadian Journal of Microbiology **38**:720-727.

Henkel, J. S., M. R. Baldwin, and J. T. Barbieri. 2010. Toxins from bacteria. Exs **100**:1-29.

Hogan, M. C., J. M. Gloor, J. R. Uhl, F. R. Cockerill, and D. S. Milliner. 2001. Two cases of non-O157 : H7 *Escherichia coli* hemolytic uremic syndrome caused by urinary tract infection. American Journal of Kidney Diseases **38**:art. no.-E22.

Huang, A., J. Friesen, and J. L. Brunton. 1987. Characterization of a bacteriophage that carries the genes for production of Shiga-Like Toxin-1 in *Escherichia coli*. Journal of Bacteriology **169**:4308-4312.

Hultgren, S. J., S. Normark, and S. N. Abraham. 1991. Chaperone assisted assembly and molecular architecture of adhesive pili. Annual Review of Microbiology **45**:383-415.

Iranpour, D., M. Hassanpour, H. Ansari, S. Tajbakhsh, G. Khamisipour, and A. Najafi. 2015. Phylogenetic groups of *Escherichia coli s*trains from patients with urinary tract infection in Iran based on the new Clermont phylotyping method. BioMed Research International **2015**:7.

Ishii, K., and M. Fukui. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Applied and Environmental Microbiology **67**:3753-3755.

Ishii, S., D. L. Hansen, R. E. Hicks, and M. J. Sadowsky. 2007. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. Environmental Science & Technology **41**:2203-2209.

Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky. 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Applied and Environmental Microbiology **72**:612-621.

Ishii, S., and M. J. Sadowsky. 2008. *Escherichia coli* in the environment: implications for water quality and human health. Microbes and Environments **23**:101-108.

Ishii, S., T. Yan, H. Vu, D. L. Hansen, R. E. Hicks, and M. J. Sadowsky. 2010. Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. Microbes and Environments **25**:8-14.

Jang, J., D. Y. W. Di, D. Han, T. Unno, J. H. Lee, M. J. Sadowsky, and H. G. Hur. 2015. Dynamic changes in the population structure of *Escherichia coli* in the Yeongsan River basin of South Korea. Fems Microbiology Ecology **91**.

Johnson, J. R. 1991. Virulence factors in *Escherichia coli* urinary tract infection. Clinical Microbiology Reviews **4**:80-128.

Johnson, J. R., B. D. Johnston, P. Delavari, P. Thuras, C. Clabots, and M. J. Sadowsky. 2017. Phylogenetic backgrounds and virulence-associated traits of *Escherichia coli* isolates from surface waters and diverse animals in Minnesota and Wisconsin. Applied and Environmental Microbiology **83**:e01329-01317.

Kallenius, G., R. Mollby, S. B. Svenson, I. Helin, H. Hultberg, B. Cedergren, and J. Winberg. 1981. Occurence of p-fimbriated *Escherichia coli* in urinary tract infections. Lancet **2**:1369-1372.

Karmali, M. A. 1989. Infection by verocytotoxin-producing *Escherichia coli*. Clinical Microbiology Reviews **2**:15-38.

Klemm, P. 1994. Fimbriae: adhesion, genetics, biogenesis, and vaccines. CRC Press.

Knittel, M. D., R. J. Seidler, C. Eby, and L. M. Cabe. 1977. Colonization of the botanical environment by *Klebsiella* isolates of pathogenic origin. Applied and Environmental Microbiology **34**:557-563.

Konowalchuk, J., J. I. Speirs, and S. Stavric. 1977. Vero response to a cytotoxin of *Escherichia coli*. Infection and Immunity **18**:775-779.

Korhonen, T. K., J. Parkkinen, J. Hacker, J. Finne, A. Pere, M. Rhen, and H. Holthofer. 1986. Binding of *Escherichia coli* s-fimbriae to human kidney epithelium. Infection and Immunity **54**:322-327.

Korhonen, T. K., R. Virkola, B. Westurlund, H. Holthofer, and J. Parkkinen. 1990. Tissue tropism of *Escherichia coli* adhesins in human extraintestinal infections. Current Topics in Microbiology and Immunology **151**:115-127.

Ksoll, W. B., S. Ishii, M. J. Sadowsky, and R. E. Hicks. 2007. Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. Applied and Environmental Microbiology **73**:3771-3778.

Kunin, C. M., T. H. Hua, C. Krishnan, L. V. White, and J. Hacker. 1993. Isolation of a nicotinamide-requiring clone of *Escherichia coli* O18-K1-H7 from women with acute cystitis - resemblance to strains found in neonatal meningitis. Clinical Infectious Diseases **16**:412-416.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences of the United States of America **82**:6955-6959.

Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur. 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Molecular Biology and Evolution **15**:1685-1695.

Lee, C. M., T. Y. Lin, C. C. Lin, G. A. Kohbodi, A. Bhattl, R. Lee, and J. A. Jay. 2006. Persistence of fecal indicator bacteria in Santa Monica Bay beach sediments. Water Research **40**:2593-2602.

Lim, J. Y., J. Yoon, and C. J. Hovde. 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. Journal of microbiology and biotechnology **20**:5-14.

Loong, S. K., C. S. Khor, F. L. Jafar, and S. AbuBakar. 2016. Utility of 16S rDNA sequencing for identification of rare pathogenic bacteria. Journal of Clinical Laboratory Analysis **30**:1056-1060.

Lyautey, E., Z. Lu, D. R. Lapen, G. Wilkes, A. Scott, T. Berkers, T. A. Edge, and E. Topp. 2010. Distribution and diversity of *Escherichia coli* populations in the South Nation River Drainage Basin, Eastern Ontario, Canada. Applied and Environmental Microbiology **76**:1486-1496.

Makapela, B., K. Okaiyeto, N. Ntozonke, U. U. Nwodo, E. Green, L. V. Mabinya, and A. I. Okoh. 2016. Assessment of *Bacillus pumilus* isolated from fresh water milieu for bioflocculant production. Applied Sciences-Basel **6**.

Makino, K., K. Yokoyama, Y. Kubota, C. H. Yutsudo, S. Kimura, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, T. Iida, K. Yamamoto, M. Onishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 1999. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. Genes Genet Syst **74**:227-239.

Mayer, C. L., C. S. Leibowitz, S. Kurosawa, and D. J. Stearns-Kurosawa. 2012. Shiga toxins and the pathophysiology of hemolytic uremic syndrome in humans and animals. Toxins **4**:1261-1287.

Mellmann, A., D. Harmsen, C. A. Cummings, E. B. Zentz, S. R. Leopold, A. Rico, K. Prior, R. Szczepanowski, Y. M. Ji, W. L. Zhang, S. F. McLaughlin, J. K. Henkhaus, B. Leopold, M. Bielaszewska, R. Prager, P. M. Brzoska, R. L. Moore, S. Guenther, J. M. Rothberg, and H. Karch. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. Plos One **6**.

Mubiru, D. N., M. S. Coyne, and J. H. Grove. 2000. Mortality of *Escherichia coli* O157 : H7 in two soils with different physical and chemical properties. Journal of Environmental Quality **29**:1821-1825.

Najafi, A., M. Hasanpour, A. Askary, M. Aziemzadeh, and N. Hashemi. 2018. Distribution of pathogenicity island markers and virulence factors in new phylogenetic groups of uropathogenic *Escherichia coli* isolates. Folia Microbiologica **63**:335-343.

Navarro-Garcia, F. 2014. *Escherichia coli* O104:H4 pathogenesis: an enteroaggregative *E. coli*/Shiga toxin-producing *E. coli* explosive cocktail of high virulence. Microbiology Spectrum **2**.

Newland, J. W., N. A. Strockbine, and R. J. Neill. 1987. Cloning of genes for producion of *Escherichia coli* Shiga-like toxin type II. Infection and Immunity **55**:2675-2680.

Obrien, A. D., and R. K. Holmes. 1987. Shiga and Shiga-like toxins. Microbiological Reviews **51**:206-220.

Pacheco, A. R., and V. Sperandio. 2012. Shiga toxin in enterohemorrhagic *E.coli*: regulation and novel anti-virulence strategies. Frontiers in Cellular and Infection Microbiology **2**.

Patel, C. B., R. Shanker, V. K. Gupta, and R. S. Upadhyay. 2016. Q-PCR based culture-independent enumeration and detection of *Enterobacter*: an emerging environmental human pathogen in riverine systems and potable water. Frontiers in Microbiology **7**.

Perchec-Merien, A. M., and G. D. Lewis. 2013. Naturalized *Escherichia coli* from New Zealand wetland and stream environments. Fems Microbiology Ecology **83**:494-503.

Pere, A., M. Leinonen, V. Vaisanenrhen, M. Rhen, and T. K. Korhonen. 1985. Occurence of type-1C fimbriae on *Escherichia coli* strains isolated from human extraintestinal infections. Journal of General Microbiology **131**:1705-1711.

Plunkett, G., D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157 : H7: Shiga toxin as a phage late-gene product. Journal of Bacteriology **181**:1767-1778.

Poolman, J. T., and M. Wacker. 2016. Extraintestinal pathogenic *Escherichia coli,* a common human pathogen: challenges for vaccine development and progress in the field. Journal of Infectious Diseases **213**:6-13.

Prieur, D. 1997. Microbiology of deep-sea hydrothermal vents. Trends in Biotechnology **15**:242-244.

Pu, Y. C., and Y. M. Hou. 2016. Isolation and identification of bacterial strains with insecticidal activities from *Rhynchophorus ferrugineus* Oliver (Coleoptera: Curculionidae). Journal of Applied Entomology **140**:617-626.

Rahdar, M., A. Rashki, H. R. Miri, and M. R. Ghalehnoo. 2015. Detection of *pap, sfa, afa, foc*, and *fim* adhesin-encoding operons in uropathogenic *Escherichia coli* isolates collected from patients with urinary tract infection. Jundishapur Journal of Microbiology **8**.

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

Rasko, D. A., D. R. Webster, J. W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E. E. Paxinos, R. Sebra, C. S. Chin, D. Iliopoulos, A. Klammer, P. Peluso, L. Lee, A. O. Kislyuk, J. Bullard, A. Kasarskis, S. Wang, J. Eid, D. Rank, J. C. Redman, S. R. Steyert, J. Frimodt-Moller, C. Struve, A. M. Petersen, K. A. Krogfelt, J. P. Nataro, E. E. Schadt, and M. K. Waldor. 2011. Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic-Uremic Syndrome in Germany. New England Journal of Medicine **365**:709-717.

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

Ritchie, J. M., C. M. Thorpe, A. B. Rogers, and M. K. Waldor. 2003. Critical roles for stx_2, *eae*, and *tir* in enterohemorrhagic *Escherichia coli*-induced diarrhea and intestinal inflammation in infant rabbits. Infection and Immunity **71**:7129-7139.

Rivera, S. C., T. C. Hazen, and G. A. Toranzos. 1988. Isolation of fecal coliforms from pristine sites in a tropical rain forest. Applied and Environmental Microbiology **54**:513-517.

Schmidt, T. M., and D. A. Relman. 1994. Phylogenetic identification of uncultured pathogens using ribosomal-RNA sequences. Bacterial Pathogenesis, Pt A **235**:205-222.

Shibata, T., H. M. Solo-Gabriele, L. E. Fleming, and S. Elmir. 2004. Monitoring marine recreational water quality using multiple microbial indicators in an urban tropical environment. Water Research **38**:3119-3131.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology **7**.

Smith, C. J., A. M. Olszewski, and S. A. Mauro. 2009. Correlation of Shiga toxin gene frequency with commonly used microbial indicators of recreational water quality. Applied and Environmental Microbiology **75**:316-321.

Solo-Gabriele, H. M., V. J. Harwood, D. Kay, R. S. Fujioka, M. J. Sadowsky, R. L. Whitman, A. Wither, M. Canica, R. C. Da Fonseca, A. Duarte, T. A. Edge, M. J. Gargate, N. Gunde-Cimerman, F. Hagen, S. L. McLellan, A. N. Da Silva, M. N. Babic, S. Prada, R. Rodrigues, D. Romao, R. Sabino, R. A. Samson, E. Segal, C. Staley, H. D. Taylor, C. Verissimo, C. Viegas, H. Barroso, and J. C. Brandao. 2016. Beach sand and the potential for infectious disease transmission: observations and recommendations. Journal of the Marine Biological Association of the United Kingdom **96**:101-120.

Stevenson, A. H. 1953. Studies of bathing water quality and health. American Journal of Public Health and the Nations Health **43**:529-538.

Tringe, S. G., and P. Hugenholtz. 2008. A renaissance for the pioneering 16S rRNA gene. Current Opinion in Microbiology **11**:442-446.

Tymensen, L. D., F. Pyrdok, D. Coles, W. Koning, T. A. McAllister, C. C. Jokinen, S. E. Dowd, and N. F. Neumann. 2015. Comparative accessory gene fingerprinting of surface water *Escherichia coli* reveals genetically diverse naturalized population. Journal of Applied Microbiology **119**:263-277.

van Hoek, A., J. N. J. van Veldhuizen, I. Friesema, C. Coipan, J. Rossen, I. L. Bergval, and E. Franz. 2019. Comparative genomics reveals a lack of evidence for pigeons as a main source of stx_{2f}-carrying *Escherichia coli* causing disease in humans and the common existence of hybrid Shiga toxin-producing and enteropathogenic *E. coli* pathotypes.

Virkola, R., B. Westerlund, H. Holthofer, J. Parkkinen, M. Kekomaki, and T. K. Korhonen. 1988. Binding characteristics of *Escherichia coli* adhesins in human urinary bladder. Infection and Immunity **56**:2615-2622.

Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky, and T. S. Whittam. 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. Environmental Microbiology **9**:2274-2288.

Whitman, R. L., and M. B. Nevers. 2003. Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. Applied and Environmental Microbiology **69**:5555-5562.

Whitman, R. L., D. A. Shively, H. Pawlik, M. B. Nevers, and M. N. Byappanahalli. 2003. Occurrence of *Escherichia coli* and enterococci in *Cladophora (Chlorophyta)* in nearshore water and beach sand of Lake Michigan. Applied and Environmental Microbiology **69**:4714-4719.

Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of prokaryotic domain - primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America **74**:5088-5090.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms - proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences of the United States of America **87**:4576-4579.

Wright, G. D. 2010. Antibiotic resistance in the environment: a link to the clinic? Current Opinion in Microbiology **13**:589-594.

Yamahara, K. M., S. P. Walters, and A. B. Boehm. 2009. Growth of enterococci in unaltered, unseeded beach sands subjected to tidal wetting. Applied and Environmental Microbiology **75**:1517-1524.

Yang, B., Y. Wang, and P. Y. Qian. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. Bmc Bioinformatics **17**.

Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, S. Kimura, C. H. Yutsudo, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, M. Yoh, T. Iida, M. Ohnishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 2000. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157 : H7 strain derived from the Sakai outbreak. Gene **258**:127-139.

Zhang, Q., and T. Yan. 2012. Correlation of intracellular trehalose concentration with desiccation resistance of soil *Escherichia coli* populations. Applied and Environmental Microbiology **78**:7407-7413.

Zhi, S., G. Banting, Q. Z. Li, T. A. Edge, E. Topp, M. Sokurenko, C. Scott, S. Braithwaite, N. J. Ruecker, Y. Yasui, T. McAllister, L. Chui, and N. F. Neumann. 2016. Evidence of naturalized stress-tolerant strains of *Escherichia coli* in municipal wastewater treatment plants. Applied and Environmental Microbiology **82**:5505-5518.

Appendices

Appendix A: PCR primers used

Gene of Interest	Name	Primer Sequence	Citation	Size of Amplicon (bp)
16S rRNA	8F	5'-AGAGTTTGATCCTGGCTCAG-3'	(Eden et al. 1991)	~1500
	1492R(s)	5'-GGTTACCTTGTTACGACTT-3'		
papC	papC F-A	5'-GACGGCTGTACTGCAGGGTGTGGCG-3'	(Escobar-Paramo et al. 2004)	~328
	papC R	5'-ATATCCTTTCTGCAGGGATGCAATA-3'		
sfa/foc	sfa/foc F	5'-CTCCGGAGAACTGGGTGCATCTTAC-3'	(Escobar-Paramo et al. 2004, Zhi et al. 2016)	~407
	sfa/foc R	5'-CGGAGGAGTAATTACAAACCTGGCA-3'		
stx_1	stx1 F-A	5'-CATCGCGAGTTGCCAGAAT-3'	(Escobar-Paramo et al. 2004)	~130
	stx1 R-A	5'-GCGTAATCCCACGGACTCTTC-3'		
stx_2	stx2 F-B	5'-CCGGAATGCAAATCAGTC-3'	(Zhi et al. 2016)	~113
	stx2 R-B	5'-CAGTGACAAAACGCAGAACT-3'		

Appendix B: PCR conditions

		16S rRNA	*papC* and *sfa/foc*	stx_1 and stx_2
Step	Time	Temperature (°C)		
Initial denaturation	3 min	95	94	94
Denaturation	45 sec	95	94	94
Annealing	60 sec	50	63/65*	58
Elongation	60sec	72	72	72
Number of cycles	29 repeats to Denaturation step (30 total cycles)			
Final elongation	10 min	72	72	72
Hold	∞	4	4	4

Appendix C: R code used for alignment and comparisons

All code used was modified from that in the msa alignment add-on package for R (Bodenhofer et al. 2015)

For entire 16S rDNA alignments of all 172 sequenced organisms:

```
## Analyzing entire 16S gDNA Alignments for all 172 samples ##

## Read Packages ##

library("msa")

library("ape")

library("seqinr")

library("ggtree")

## Point to data ##

SequenceFile <- "<file path>"  ## File path of ALL fasta file sequences ##

## Read data as DNA ##

Sequences_FromFile <- readDNAStringSet(SequenceFile)

## Create Alignment using ClustalOmega ##

Alignment_ClustalOmega <- msa(Sequences_FromFile, "ClustalOmega")

## Optional: To see the whole alignment ##

print(Alignment_ClustalOmega, show="complete")

## Making neighbor joining tree ##

Aln_Omega <- msaConvert(Alignment_ClustalOmega, type="seqinr::alignment")

## Make Distance Matrix (using library(seqinr))##

dist_matrix <- dist.alignment(Aln_Omega, "identity")

## Make Neighbor-joining Tree (using library(ape)) ##

NJ_Tree <- nj(dist_matrix)

## Basic Plot of NJ Tree ##
```

```
plot(NJ_Tree, main="16S rDNA alignment of all E. coli samples collected")

##To read plot better##

plot(unroot(NJ_Tree),lab4ut="axial",cex=0.35, main="16S rRNA Alignment of E. coli")

##To expand output of tree##

plot(unroot(NJ_Tree),no.margin=TRUE,lab4ut="axial",cex=0.3)

##To number the nodes to expand a certain area of the tree##

ggtree(NJ_Tree,branch.length="none")+geom_text2(aes(label=node))+geom_tiplab()

##To zoom in on specific nodes/branches##

p=ggtree(NJ_Tree)

viewClade(p+geom_tiplab(),node=**) ##Where ** is specific node chosen to zoom in on##
```

For analyzing V1 and V2 of all 172 sequenced organisms:

```
## Analyzing V1 and V2 regions of 16S gDNA Alignments for all 172 samples ##

## Read Packages ##

library("msa")

library("ape")

library("seqinr")

library("ggtree")

## Point to data ##

SequenceFile <- "<file path>"  ##File path of ALL fasta file sequences ##

## Read data as DNA ##

Sequences_FromFile <- readDNAStringSet(SequenceFile)

## Create Alignment using ClustalOmega ##

Alignment_ClustalOmega <- msa(Sequences_FromFile, "ClustalOmega")

## Optional: To see the whole alignment ##

print(Alignment_ClustalOmega, show="complete")

## Making neighbor joining tree ##
```

```
Aln_Omega <- msaConvert(Alignment_ClustalOmega, type="seqinr::alignment")

##Specific Variable region (here is specific for v1) alignment##

##V2 region is from 137 to 242##

final=list()

for(i in 1:length(Aln_Omega$seq)){

 sequences=Aln_Omega$seq[i]

 sequences=unlist(strsplit(sequences,''))

 sequences=sequences[69:99]

 final[i]=print(paste(sequences,collapse=''))}

finalAlign_V1=as.alignment(nb=Aln_Omega$nb,nam=Aln_Omega$nam,seq=final,com=Aln_Omega$com)

## Make Distance Matrix (using library(seqinr))##

dist_matrix_v1 <- dist.alignment(finalAlign_V1, "identity")

## Make Neighbor-joining Tree (using library(ape)) ##

NJ_Tree_v1 <- nj(dist_matrix_v1)

## Basic Plot of NJ Tree ##

plot(NJ_Tree_v1, main="V1 alignment of all E. coli samples collected")

##To read plot better##

plot(unroot(NJ_Tree_v1),lab4ut="axial",cex=0.35, main="V1 Alignment of E. coli")

##To expand output of tree##

plot(unroot(NJ_Tree_v1),no.margin=TRUE,lab4ut="axial",cex=0.3)

##To number the nodes to expand a certain area of the tree##

ggtree(NJ_Tree_v1,branch.length="none")+geom_text2(aes(label=node))+geom_tiplab()

##To zoom in on specific nodes/branches##

p=ggtree(NJ_Tree_v1)

viewClade(p+geom_tiplab(),node=**) ##Where ** is specific node chosen to zoom in on##
```

The above code is for V1 region only. For V2 region, the base pairs to compare after alignment are 137 to 242. Change the "V1" portion of the commands as needed in order to differentiate the different variable regions being compared.

For analyzing V3 and V6 regions of 16S rDNA (113 samples):

Because some sequences did not reach over 960bp, and because of the strict parameters for the Muscle alignment protocol, I removed those sequences (n=65) that were less than 960bp from this alignment protocol.

```
## Analyzing V3 and V6 regions of 16S gDNA Alignments for 110 samples (ones over 960bp in length) ##

## Read Packages ##

library("msa")

library("ape")

library("seqinr")

library("ggtree")

## Point to data ##

SequenceFile_960 <- "<file path>"## File path of fasta file sequences over 960bp ##

## Read data as DNA ##

Sequences_FromFile_960 <- readDNAStringSet(SequenceFile_960)

## Create Alignment using Muscle for more stringent alignment ##

Alignment_Muscle <- msa(Sequences_FromFile_960, "Muscle")

## Optional: To see the whole alignment ##

print(Alignment_Muscle, show="complete")

## Making neighbor joining tree ##

Aln_Muscle <- msaConvert(Alignment_Muscle, type="seqinr::alignment")

##Specific Variable region (here is specific for v6) alignment##

##V3 region is from 433 to 497##

final=list()
```

```
for(i in 1:length(Aln_Muscle$seq)){
 sequences=Aln_Muscle$seq[i]
 sequences=unlist(strsplit(sequences,''))
 sequences=sequences[986:1043]
 final[i]=print(paste(sequences,collapse=''))}
finalAlign_v6=as.alignment(nb=Aln_Muscle$nb,nam=Aln_Muscle$nam,seq=final,com=Aln_Muscle$com)
## Make Distance Matrix (using library(seqinr))##
dist_matrix_v6 <- dist.alignment(finalAlign_v6, "identity")
## Make Neighbor-joining Tree (using library(ape)) ##
NJ_Tree_v6 <- nj(dist_matrix_v6)
## Basic Plot of NJ Tree ##
plot(NJ_Tree_v6, main="V6 alignment of E. coli samples collected over 960bp")
##To read plot better##
plot(unroot(NJ_Tree_v6),lab4ut="axial",cex=0.35, main="V6 Alignment of E. coli")
##To expand output of tree##
plot(unroot(NJ_Tree_v6),no.margin=TRUE,lab4ut="axial",cex=0.3)
##To number the nodes to expand a certain area of the tree##
ggtree(NJ_Tree_v6,branch.length="none")+geom_text2(aes(label=node))+geom_tiplab()
##To zoom in on specific nodes/branches##
p=ggtree(NJ_Tree_v6)
viewClade(p+geom_tiplab(),node=**) ##Where ** is specific node chosen to zoom in on##
##msa Pretty Print##
msaPrettyPrint(Alignment_Muscle, output=c("asis"), subset=c(1:20), y=c(986, 1043), showNames =
"none", showLogo = "top", logoColors = "rasmol", shadingMode = "identical", showLegend = FALSE,
askForOverwrite = FALSE)  ## Creates command line code for LaTeX in TeXStudio ##
msaPrettyPrint(Alignment_Muscle, output=c("pdf"), subset=c(1:20), y=c(986, 1043), showNames =
"none", showLogo = "top", logoColors = "rasmol", shadingMode = "identical", showLegend = FALSE,
askForOverwrite = FALSE)  ##automatically generates pdf file of alignment##
```

The above code is for V6 region only. For V3 region, the base pairs to compare after alignment are 433 to 497. Change the "V6" portion of the commands as needed in order to differentiate the different variable regions being compared. In the "msaPrettyPrint" code line, the subset variables may be changed to include as many as you would like to compare. The "y" variables may also be changed to select a different region to compare.

For analyzing V1 to V6 regions of 16S rDNA (113 samples):

```
## Analyzing V1 to V6 regions of 16S gDNA Alignments for 110 samples (ones over 960bp in length) ##

## Read Packages ##

library("msa")

library("ape")

library("seqinr")

library("ggtree")

## Point to data ##

SequenceFile_960 <- "<file path>" ##File path of fasta file sequences over 960bp ##

## Read data as DNA ##

Sequences_FromFile_960 <- readDNAStringSet(SequenceFile_960)

## Create Alignment using Muscle for more stringent alignment ##

Alignment_Muscle <- msa(Sequences_FromFile_960, "Muscle")

## Optional: To see the whole alignment ##

print(Alignment_Muscle, show="complete")

## Making neighbor joining tree ##

Aln_Muscle <- msaConvert(Alignment_Muscle, type="seqinr::alignment")

##Specific Variable region (here is specific for V1 to V6) alignment##

final=list()

for(i in 1:length(Aln_Muscle$seq)){

  sequences=Aln_Muscle$seq[i]
```

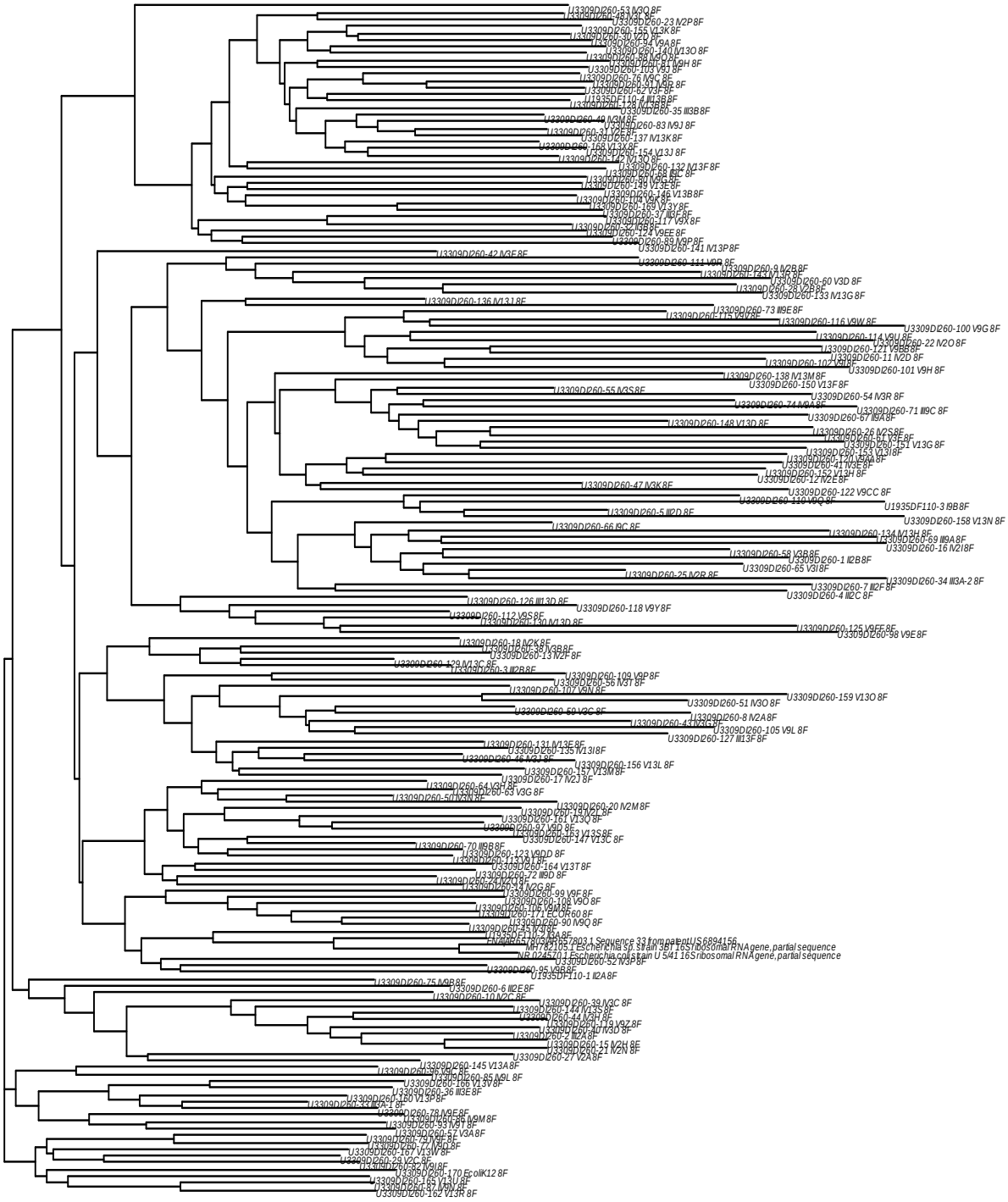
```
 sequences=unlist(strsplit(sequences,''))

 sequences=sequences[69:1043]

 final[i]=print(paste(sequences,collapse=''))}

finalAlign_v1to6=as.alignment(nb=Aln_Muscle$nb,nam=Aln_Muscle$nam,seq=final,com=Aln_Muscle$com)

## Make Distance Matrix (using library(seqinr))##

dist_matrix_v1to6 <- dist.alignment(finalAlign_v1to6, "identity")

## Make Neighbor-joining Tree (using library(ape)) ##

NJ_Tree_v1to6 <- nj(dist_matrix_v1to6)

## Basic Plot of NJ Tree ##

plot(NJ_Tree_v1to6, main="V1 to V6 alignment of E. coli samples collected over 960bp")

##To read plot better##

plot(unroot(NJ_Tree_v1to6),lab4ut="axial",cex=0.35, main="V1 to V6 Alignment of E. coli")

##To expand output of tree##

plot(unroot(NJ_Tree_v1to6),no.margin=TRUE,lab4ut="axial",cex=0.3)

##To number the nodes to expand a certain area of the tree##

ggtree(NJ_Tree_v1to6,branch.length="none")+geom_text2(aes(label=node))+geom_tiplab()

##To zoom in on specific nodes/branches##

p=ggtree(NJ_Tree_v1to6)

viewClade(p+geom_tiplab(),node=**)  ##Where ** is specific node chosen to zoom in on##

##msa Pretty Print##

msaPrettyPrint(Alignment_Muscle, output=c("asis"), subset=c(1:20), y=c(900, 1043), showNames = "none", showLogo = "top", logoColors = "rasmol", shadingMode = "identical", showLegend = FALSE, askForOverwrite = FALSE) )  ## Creates command line code for LaTeX in TeXStudio ##

msaPrettyPrint(Alignment_Muscle, output=c("pdf"), subset=c(1:20), y=c(900, 1043), showNames = "none", showLogo = "top", logoColors = "rasmol", shadingMode = "identical", showLegend = FALSE, askForOverwrite = FALSE) ##automatically generates pdf file of alignment##
```

In the “msaPrettyPrint” code line, the subset variables may be changed to include as many as you would like to compare. The “y” variables may also be changed to select a different region to compare.

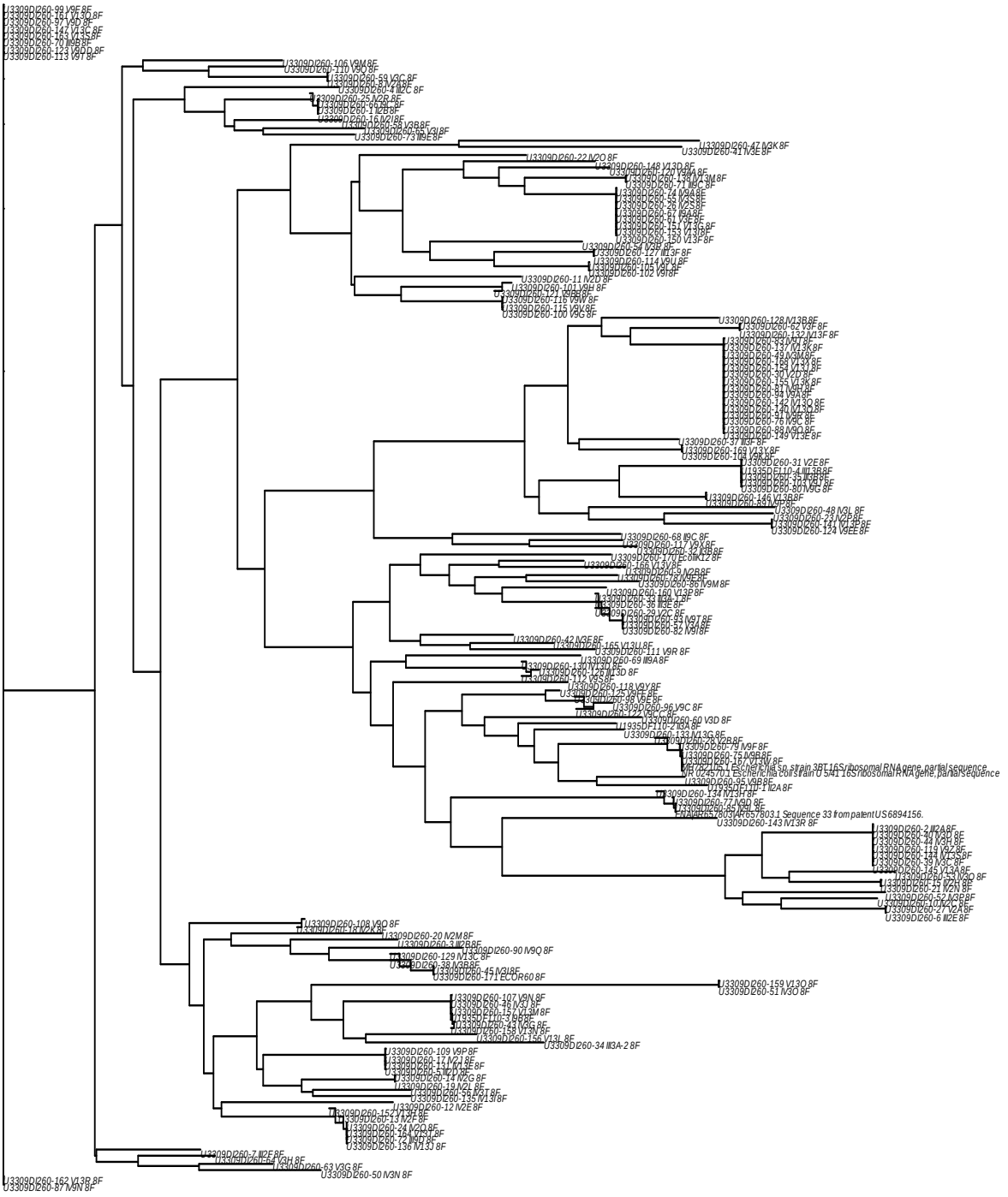
Appendix D: Data collected per sample date

Sample Date	Air Temperature (°C)	Water Temperature (°C)	No. Waterfowl	No. People	Precipitation within previous 3 days	Precipitation within previous 7 days
5/25/2016	18	17	21	10	No	No
6/29/2016	21	21	45	9	No	Yes
7/26/2016	24	25	20	11	Yes	
6/21/2017	18	21	1	0	Yes	
8/2/2017	28	24	50	10	Yes	
Average:	21.8	21.6	27.4	8		

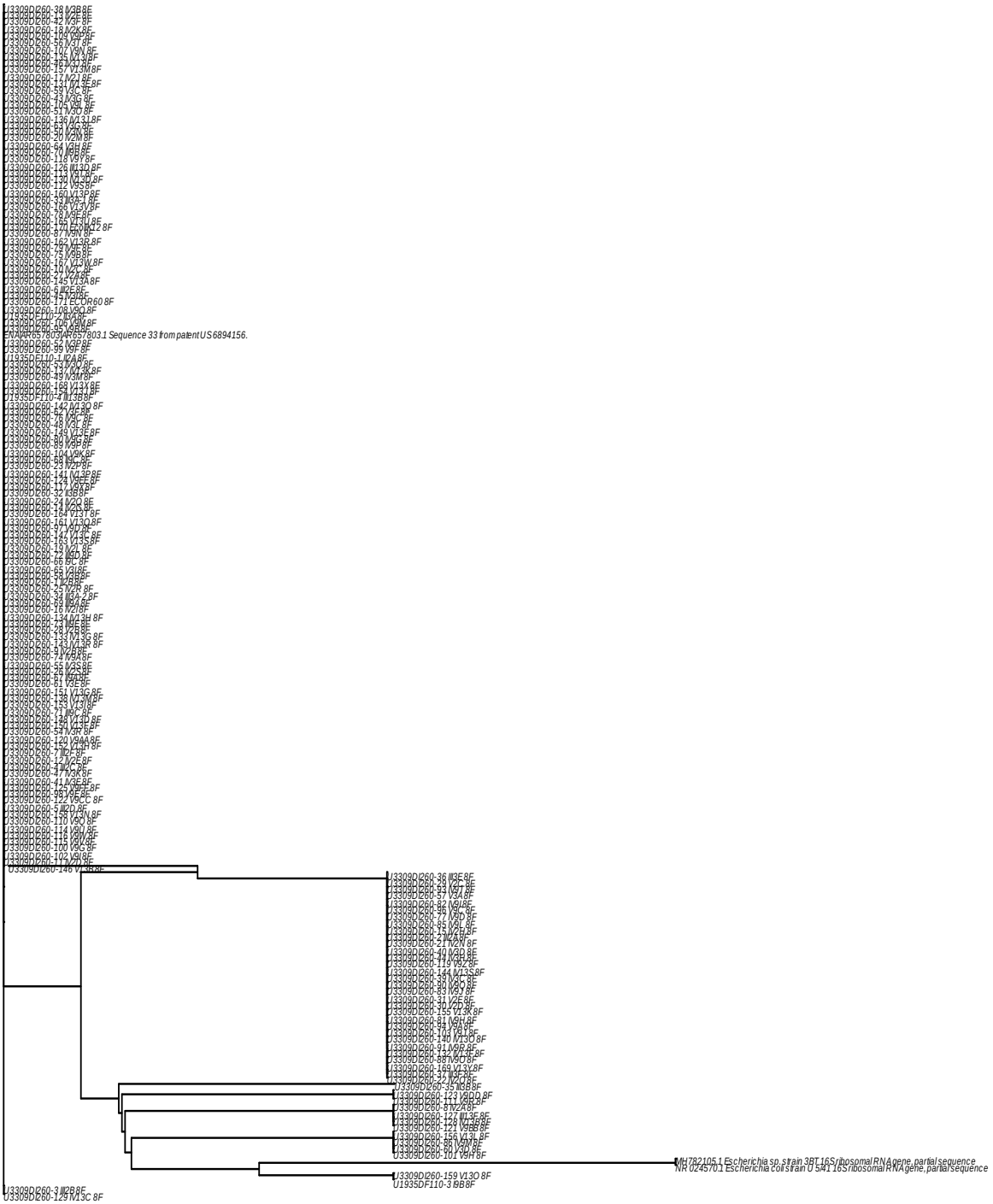
Appendix E: Phylogenetic trees created using R Studio



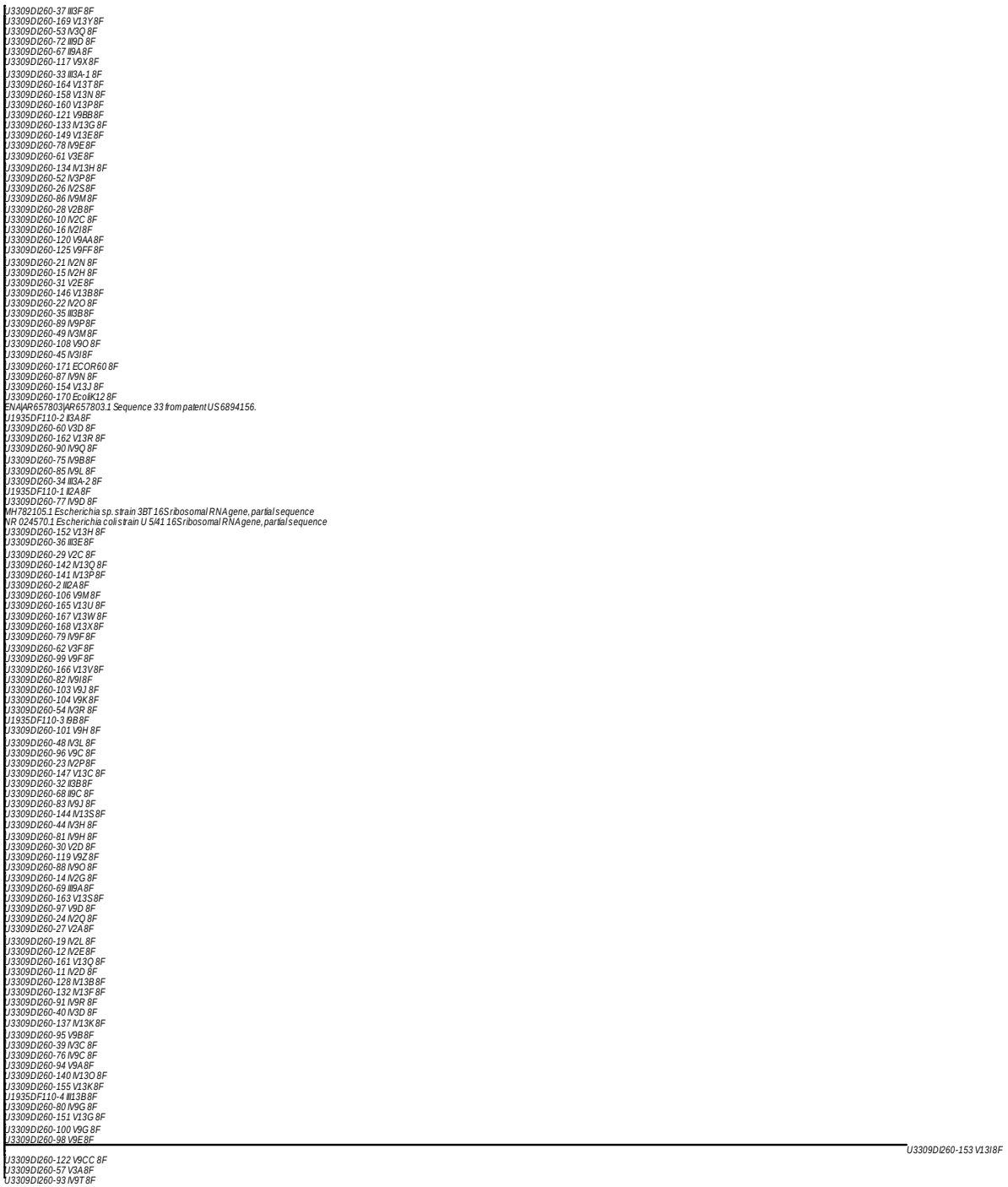
1. Complete 16S rDNA gene sequence analysis of all 178 samples using the ClustalOmega multiple sequence alignment parameters.



2. Variable region 1 sequence analysis of all 178 samples using the ClustalOmega multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 99 were parsed out and this tree was generated.

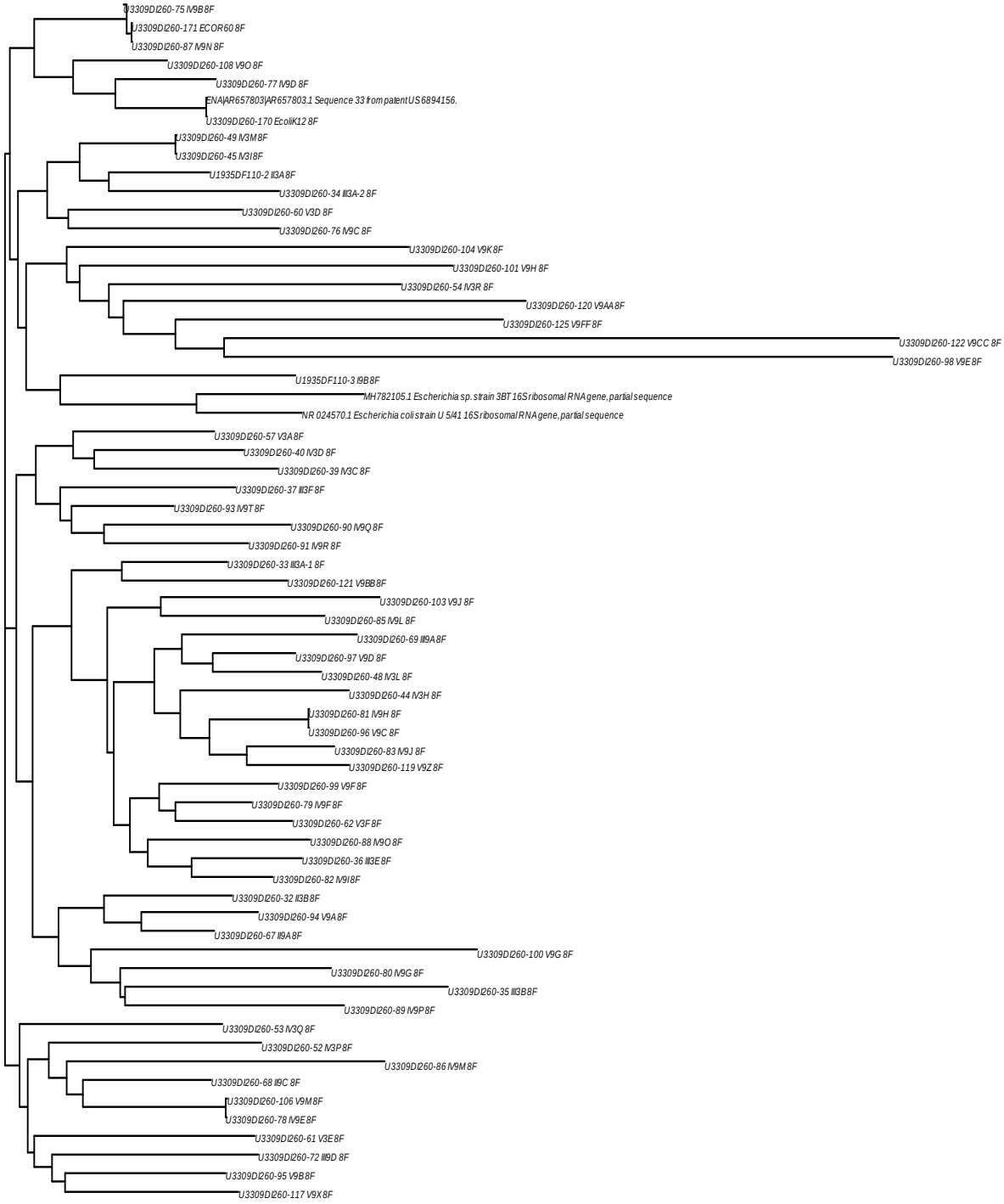


3. Variable region 2 sequence analysis of all 178 samples using the ClustalOmega multiple sequence alignment parameters. After sequences were aligned, base pairs 137 to 242 were parsed out and this tree was generated.

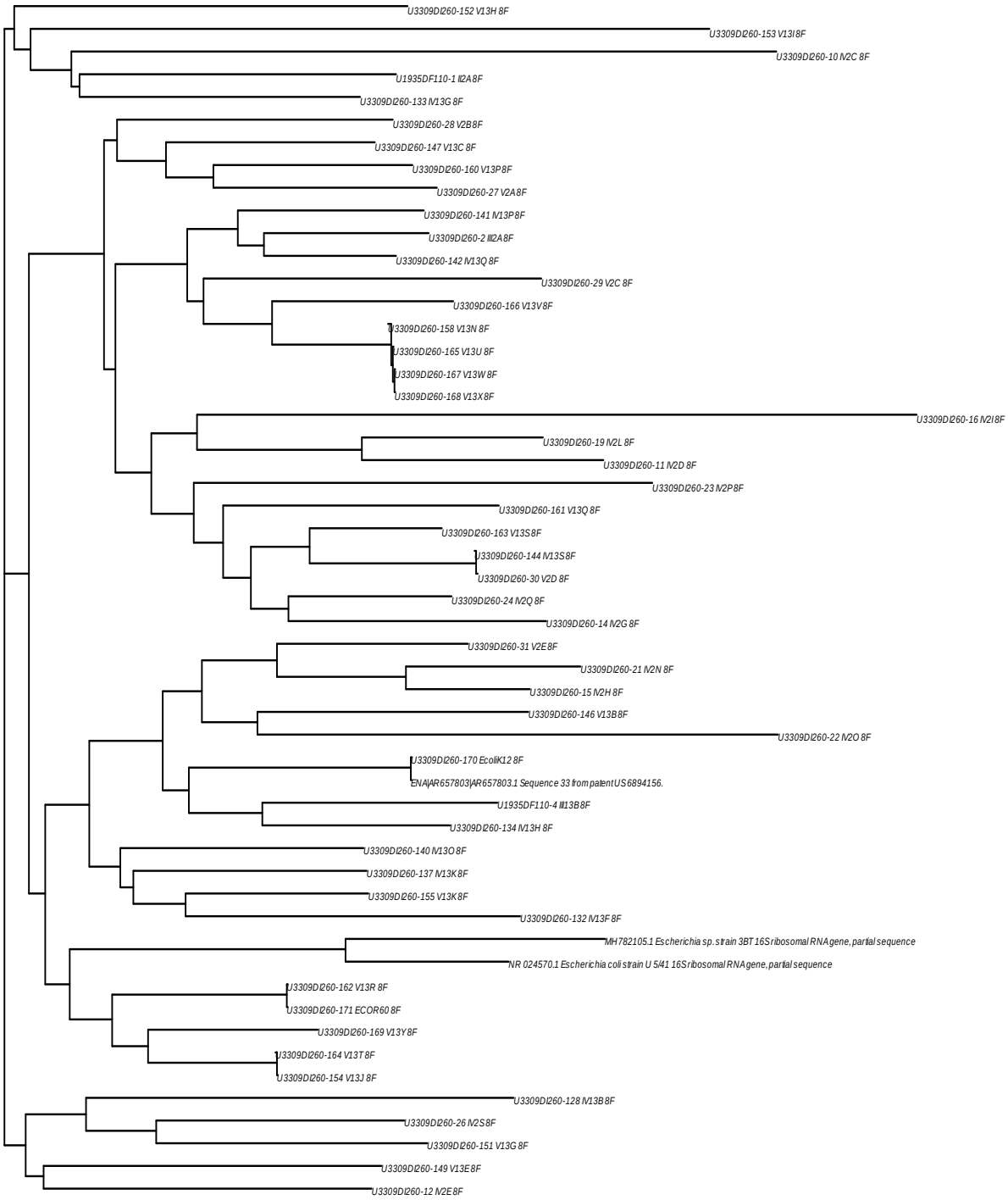


4. Variable region 3 sequence analysis of 113 samples using the Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 433 to 497 were parsed out and this tree was generated.

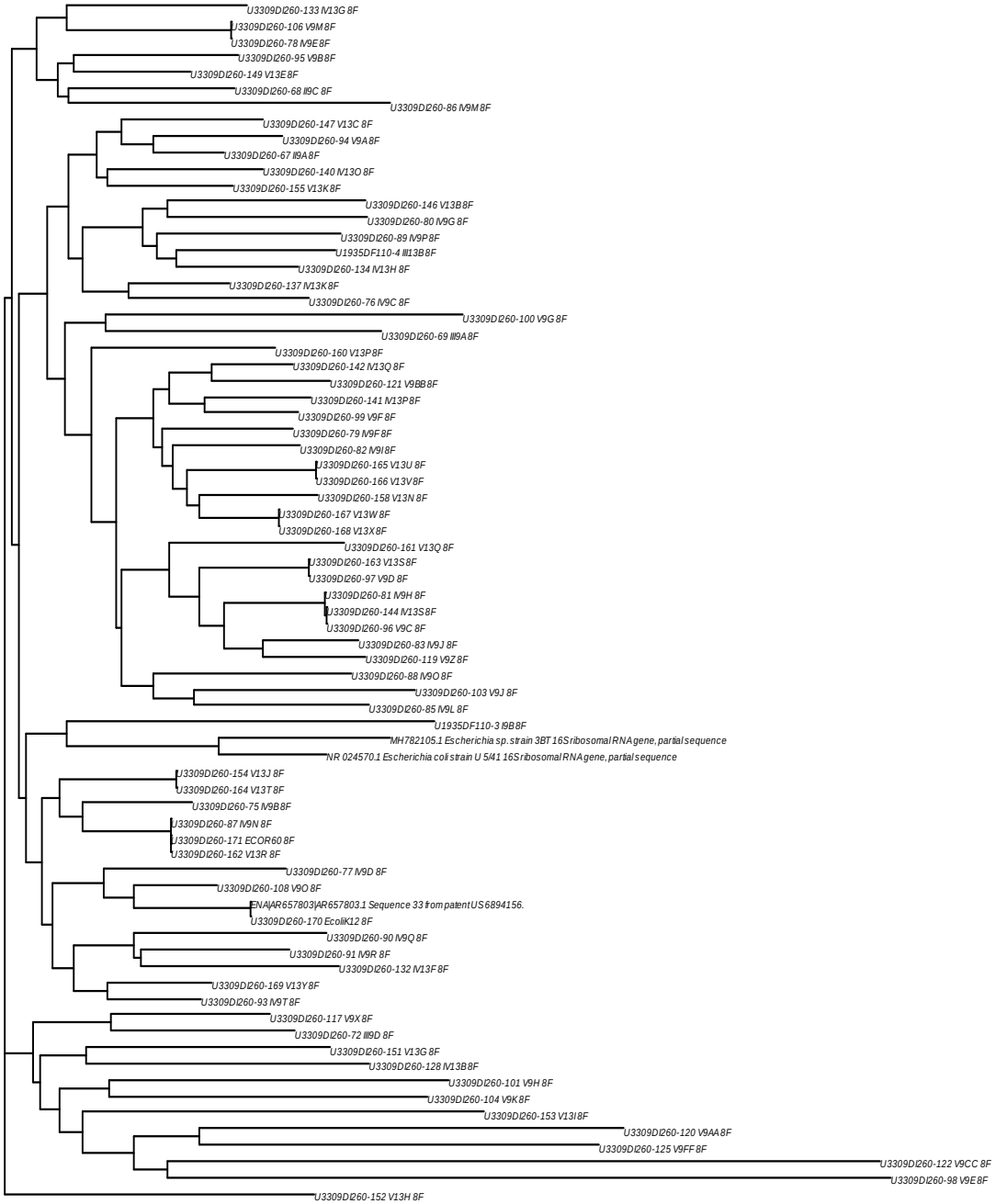
5. Variable region 6 sequence analysis of 113 samples using the Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 986 to 1043 were parsed out and this tree was generated.



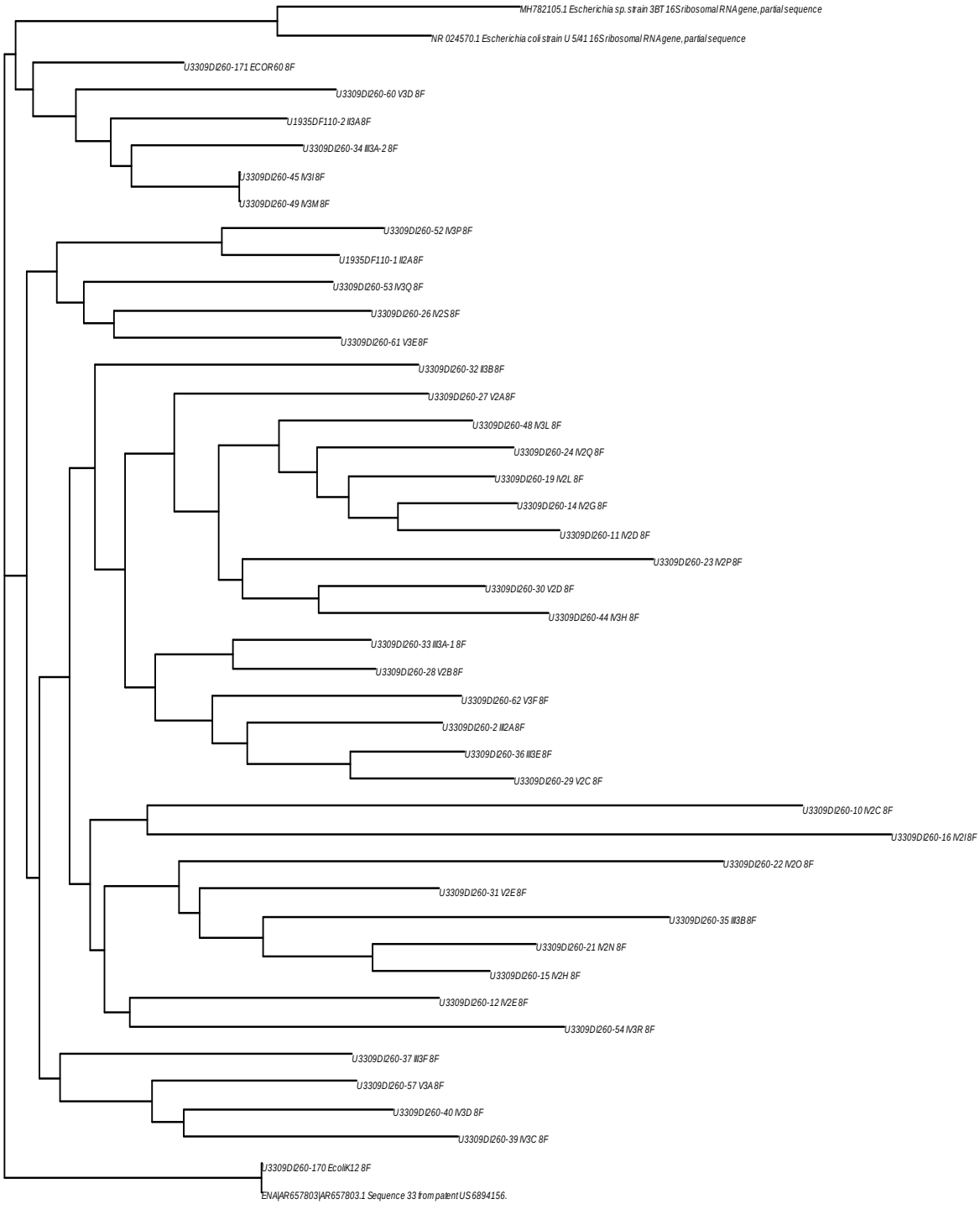
6. Sites 3 and 9 only (North side of beach, n_t=60: n_3=20, n_9=40) analysis covering variable regions 1 through 6 using Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 1043 were parsed out and this tree was generated.



7. Sites 2 and 13 only (South side of beach, n_t=48: n_2=19, n_{13}=29) analysis covering variable regions 1 through 6 using Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 1043 were parsed out and this tree was generated.

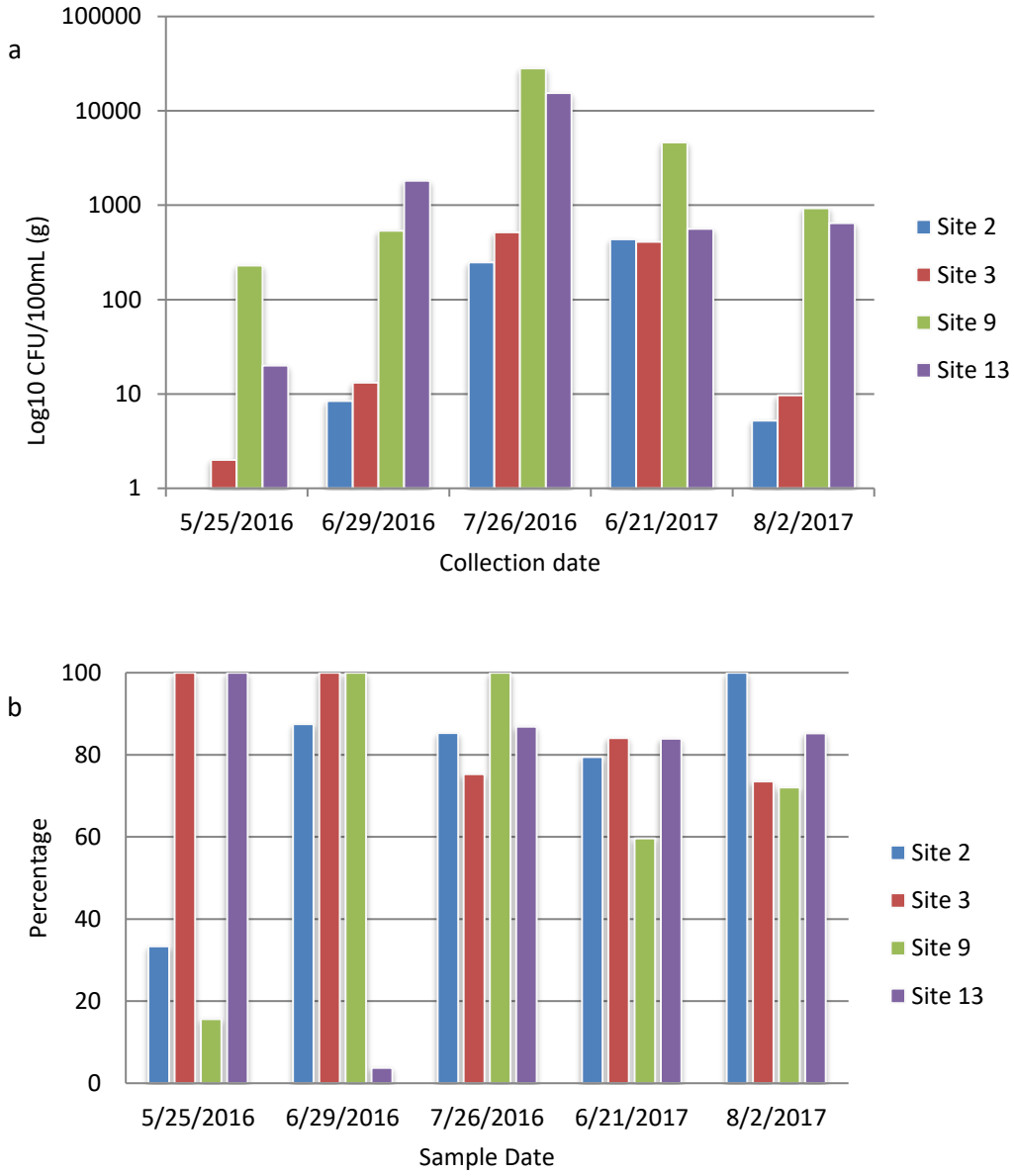


8. Sites 9 and 13 only (Sand isolates only, n_t=69: n_9=40, n_{13}=29) analysis covering variable regions 1 through 6 using Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 1043 were parsed out and this tree was generated.



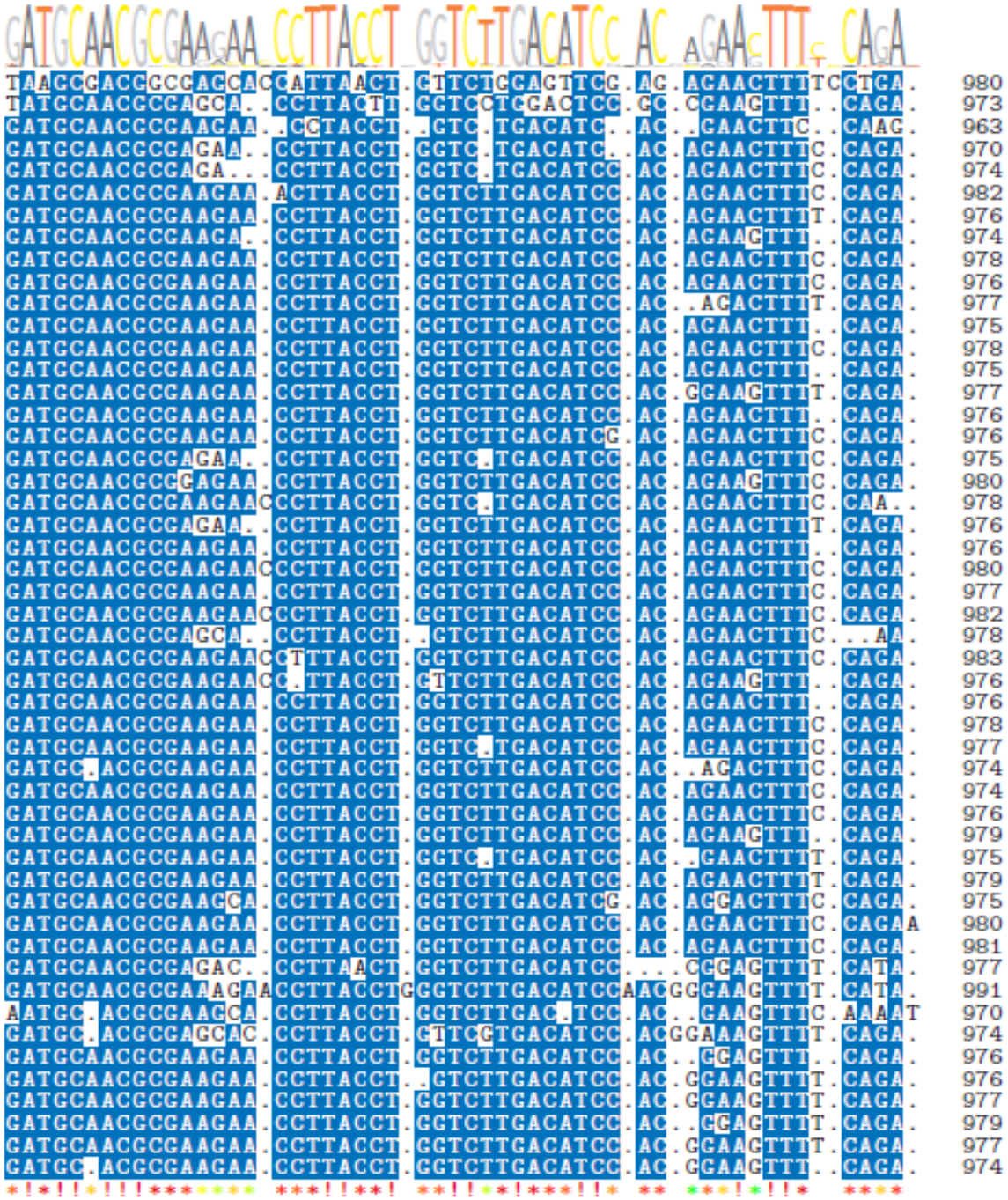
9. Sites 2 and 3 only (Water isolates only, n_t=39: n_2=19, n_3=20) analysis covering variable regions 1 through 6 using Muscle multiple sequence alignment parameters. After sequences were aligned, base pairs 69 to 1043 were parsed out and this tree was generated.

Appendix F: *E. coli* IDEXX data by collection date



a) Log_{10} of *E. coli* CFU/100mL by sample date. Wells were scored as positive for *E. coli* if the contents turned yellow (indicating the activity of β-galactosidase) and also had a blue florescence when exposed to 366nm UV light (indicating the activity of β-glucuronidase). b) Percent *E. coli* of Fecal Coliforms by sample date. The number of *E. coli* positive wells (yellow and fluorescent wells) was divided by the number of fecal coliforms positive wells (yellow only wells) to determine overall percentages. Sites 2 & 3 = water sites; Sites 9 & 13 = sand sites

Appendix G: LaTeX alignment and conservation of bases



Representation of the overall conservation of bases in the V6 region of the 16S rRNA gene. Bases highlighted in blue indicate conserved bases. The size of the letters on top indicate the level of conservation. The exclamation point under the alignment indicates complete conservation of base while the asterisks indicate the level where green is less conserved than red.

Bibliography:

Bodenhofer, U., E. Bonatesta, C. Horejs-Kainrath, and S. Hochreiter. 2015. msa: an R package for multiple sequence alignment. Bioinformatics **31**:3997-3999.

Eden, P. A., T. M. Schmidt, R. P. Blakemore, and N. R. Pace. 1991. Phylogenetic analysis of *Aquaspirillum magnetotacticum* using polymerase chain reaction-amplified 16S ribosomal-RNA-specific DNA. International Journal of Systematic Bacteriology **41**:324-325.

Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur. 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Molecular Biology and Evolution **21**:1085-1094.

Zhi, S., G. Banting, Q. Z. Li, T. A. Edge, E. Topp, M. Sokurenko, C. Scott, S. Braithwaite, N. J. Ruecker, Y. Yasui, T. McAllister, L. Chui, and N. F. Neumann. 2016. Evidence of naturalized stress-tolerant strains of *Escherichia coli* in municipal wastewater treatment plants. Applied and Environmental Microbiology **82**:5505-5518.