2018

The Role of Genomic Versatility in Multi-Niche Preferences of *Escherichia Coli*

Gitanjali NandaKafle
South Dakota State University

Follow this and additional works at: https://openprairie.sdstate.edu/etd

Part of the Biology Commons, and the Microbiology Commons

THE ROLE OF GENOMIC VERSATILITY IN MULTI-NICHE PREFERENCES OF

ESCHERICHIA COLI

BY

GITANJALI NANDAKAFLE

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Microbiology

South Dakota State University

2018

THE ROLE OF GENOMIC VERSATILITY IN MULTI-NICHE PREFERENCES OF *ESCHERICHIA COLI*

GITANJALI NANDAKAFLE

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Biological Science degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Volker S. Brözel, PhD
Thesis Advisor   Date

Volker S. Brözel, PhD
Head, Department of Biology and Microbiology   Date

Kinchel Doerner, PhD
Dean, Graduate School   Date

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude and deep appreciation to my advisor Dr. Volker S. Brözel for his continuous support and guidance throughout my PhD study and research. I greatly appreciate the freedom he has given me to do my research and develop myself as researcher in the best possible way. I have been extremely lucky to have him as my supervisor, who cared so much about my research and, responded to my questions and queries promptly despite of his busy schedules. I could not have imagined having a better advisor and mentor for my PhD study. This dissertation would not have been possible without his guidance and persistent help.

I would like to thank my committee members Dr. Radhey S. Kaushik, Dr. Ruanbo Zhou and Dr. Madhav Nepal for their assistance, insightful comments and encouragements while pursuing my research. My Sincere thanks goes to all faculties and staffs of Biology and Microbiology department who directly and indirectly help me to achieve this goal. I wish to thank our department secretaries Jan Matson, Karen Barthel, Sharon Ellens and Carol Doyle for helping the department to run smoothly and for assisting me in many different ways.

I am thankful to my fellow lab mates Dakota York, Nabilah Alshibli, Amrit Koirala and Bikram Das for helping me various ways to accomplish my research. I will always remember those days we spent together in lab, all about exchange of views and about sharing joys, sorrows and lunches. Also, I would like to thank my undergrad interns Meghan Schenk, Taylor Haugen and Lane Blasius for their sincere hard work to generate

CONTENTS

LIST OF ABBREVIATIONS

A_{546}	Absorbance at 546 nanometers
ANOVA	Analysis of variance
AZM	Azithromycin
BF	Bovine feces
BC	Bovine cluster
°C	Degrees Centigrade
CFU	Colony forming units
CIP	Ciprofloxacin
CN	Gentamycin
CRO	Ceftriazone
CV	Crystal violet
DOM	Dissolved organic matter
EC	Environmental cluster
ECOR	E. coli Reference
HGG	Highly Grazed Group
g	Gram
K	Number of populations

LA   Lactose Agar

LB   Luria Bertani medium according to Miller

LGG   Least Grazed Group

MEM   Meropenem

MC   Mixed cluster

mL   Milliliter

MLST   Multilocus sequence type

μg   Microgram

μL   Microliter

OD   Optical density

PCA   Principal component analysis

RDAR   Rough, dry and red

SBG   Soil before grazing

SDS   Sodium dodecyl sulphate

SESOM   Soil extracted soluble organic matter

ST   Sequence Type

SWG   Sewage

TE   Tetracycline

UPGMA   Unweighted pair group method with arithmetic mean

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

THE ROLE OF GENOMIC VERSATILITY IN MULTI-NICHE PREFERENCES OF
ESCHERICHIA COLI

GITANJALI NANDAKAFLE

2018

Escherichia coli strains are naturally present as either commensals or pathogens in the gastrointestinal tract of mammals and some other vertebrates. Until recently, it was assumed that *E. coli* are solely associated with the gut and are unable to survive outside of a host for a long period of time, the basis of its use as an indicator organism. Recent reports suggest that *E. coli* can become naturalized to several tropical, subtropical or temperate soils and aquatic environments, where they have been isolated repeatedly. Several studies have shown that these strains are capable of surviving and proliferating in the environment under suitable conditions. Not only have these strains adapted to the environment but also, several studies have revealed that they are genetically distinct from their gut-associated counterparts. In this dissertation, I focused to understand the genomic versatility and adaptation strategies of *E. coli* in pasture and pond ecosystems. The objectives of my research were (I) to determine the *E. coli* diversity and niche partitioning in pasture and pond ecosystems, (II) to compare the growth and extended survival of environmental *E. coli* isolates and *E. coli O157:H7* in soil organic matter, (III) to determine *E. coli* fitness in soil by determining the antibiotic resistance, presence of virulence genes and susceptibility to grazing by *Dictyostelium discoideum,* and (IV) to compare the genotypic and phenotypic diversity of 20 representative isolates. These

objectives were achieved as follows. Sampling of the pasture and pond environments in this study involved various representative sample types of two ecosystems. Phylogrouping and phylogenetic analysis of *mutS* and *uidA* genes were used to determine the diversity within the *E. coli* populations obtained, and to find out if any possible unique environmental strains exist or any of these isolates belongs to one of the previously described *Escherichia* clades. Furthermore, to determine the survival ability of isolates in soil a long-term survival study was conducted in liquid soil organic matter (SESOM) at 25 °C as well as in sterile soil outside over winter. The ability of *E. coli* to survive in various environments depends on several factors. The fitness of these isolates to survive in soil and aquatic environments was determined by biofilm and RDAR (red, dry and rough) formation, antibiotic resistance, presence of virulence genes and protozoan grazing susceptibility. Comparative analyses of the whole genome sequences of 20 isolates were conducted using EDGAR computational platform and R programming. A Phenotypic microarray assay was used to obtain the nutrient utilization profile of 20 isolates.

The results of pasture isolate studies indicated the existence of environmental *E. coli* that are phylogenetically distinct from bovine fecal isolates, and which are able to better maintain populations in the soil environment. The pond isolates showed a distribution pattern of genotypic and phenotypic traits among isolates of various sample sources based on their niche preferences. Population genetic analysis of both the *uidA* and *mutS* genes supported the existence of three separate populations in the pond ecosystem. The bovine feces isolates belonged to one population and the snail isolates were of two, whereas the sediment, plant, and water isolates were an admixture of three different

populations. The antibiotic resistance pattern of snail and bovine feces isolates were very different from sediment, plant and water isolates. The environmental strains were found to be more resistant to protozoan grazing, suggesting these strains may have developed some mechanism to avoid grazing, thereby displaying enhanced survival in soil. *E. coli* isolates from pasture soil and bovine feces displayed a high genotypic and phenotypic diversity within phylogroups. However the genotype diversity did not mirror the phenotypic distribution. Further implementation of transcriptome, proteome and metabolomics data is necessary to understand the genotype and phenotypic relatedness of organisms.

These results suggest that *E. coli* strain*s* with the potential to be pathogenic are able to maintain populations in the environment more broadly than previously thought. The presence of naturalized or environmental populations of *E. coli* in soil and aquatic environments renders the use of this bacterium as an indicator organism ambiguous at best. The ability to distinguish between environmental and host associated strains could allow for more accurate use of *E. coli* as an indicator for recent fecal contamination.

Chapter 1: Literature Review

Diversity, Survival Potential and Virulence of Naturalized *Escherichia*

coli

1. Introduction

Escherichia coli is one of the most adaptable and extensively studied microorganisms. It is widely used as a host in many biotechnology laboratories to express several recombinant proteins, because of its rapid growth rate, easy genetic manipulation, and high level of recombinant protein synthesis rate. It is being exploited in the biotechnology industry for large-scale production of proteins for therapeutic use. Besides its use as model organism in research, it is an important member of the microbiome in the lower intestinal tract of humans and other vertebrates, its primary habitat. Initially it was believed that *E. coli* could only proliferate in the GI tract of warm-blooded animals. Recent research on the survival of *E. coli* in the natural environmental shows that *E. coli* can also reside outside the intestinal tract, the secondary habitat (Touchon et al., 2009). The occurrence of *E. coli* in environments outside the host has questioned the narrow view on its habitat.

Studies on the physiology, biochemistry and genetics of *E. coli* have focused mostly on the commensal and pathogenic strains, as it was considered that *E. coli* can only replicate and grow in the gastrointestinal tract.

The Genus *Escherichia* is a member of the class *Gammaproteobacteria* in the phylum *Proteobacteria* and belongs to the family *Enterobacteriaceae*. This family comprises a large number of Gram-negative bacteria that include pathogens and harmless symbionts.

Genera include *Citrobacter*, *Enterobacter*, *Klebsiella*, *Pantoea*, and *Salmonella, Shigella* (Farmer et al., 2010). So far the genus *Escherichia* is composed of six species including *E. coli* and less frequently occurring members; *E. albertii, E. blattae, E. fegussoni, E.hermanii* and *E. vulneris* (Farmer, 1999;Abbott et al., 2003). *E. adecarboxylata* has been renamed as *Leclercia adecarboxylata,* a new genus based on extensive biochemical study and DNA-DNA hybridization (Stock et al., 2004). *E. coli* and *Shigella* species share many common phenotypic characteristics, and have long been considered as a single species based on their DNA-DNA homology (Brenner 1984) cross ref (Fukushima et al., 2002). Biochemical and serological methods are used to differentiate these species; *Shigella* are non motile, unable to ferment lactose, non-gas producing isolates. However, this sometimes gives incorrect result due to inactive variants of *E. coli* (Khot and Fisher, 2013) or pathogenic *Shigella* strains that exhibited *E. coli* characteristics and display the ability to ferment sugars (Pupo et al., 2000).

Escherichia is most closely related to *Salmonella* species, based on the amino acid sequence data and 16SrRNA sequence data analysis, it has been estimated that the divergence of *E. coli* and *S.* Typhimurium from a common ancestor had occurred approximately 100-160 million years ago (Kumar and Hedges, 1998). These two species are considered to be phylogenetically related (Ochman and Wilson, 1987;Doolittle et al., 1996). They shared a lot of genetic materials; termed a core genome and 2500-3100 genes are part of the core genome that is about 50% of the total genome.

The distribution of *E. coli* population worldwide is estimated to be 10^{20}, and are widespread as both gut commensals in warm blooded animals and pathogens causing both intestinal and extra intestinal diseases in human (Whitman et al., 1998;Tenaillon et

al., 2010). Most of the *E. coli* strains pass through the GI tract in a short period of time without any effect on the host, but some are able to established in the intestine and become a part of the gut microbiome (Savageau, 1983). *E. coli* is identidfied by its Beta glucuronidase activity (Rice et al., 1990).

2. *E. coli* as an indicator organism

Indicators organisms are typically used to determine the presence or absence of any pathogenic group of bacteria in an environmental sample. Indicator organisms are categorized into three groups 1. General microbial indicators are a group of microorganisms that show the effectiveness of a process such as total coliform for chlorine disinfection. 2. Fecal indicators are a group of organisms that indicates the presence of fecal contamination, and 3. Index organisms or model organisms a group or species indicate the presence of pathogens and its behavior, for example *E. coli* is an index for *Salmonella,* and F-RNA coliphage is an index for enteric virus (Ashbolt et al., 2001).

The use of an indicator organism is cost effective and simple for analyzing environmental media compared to analyzing samples for individual pathogens. It is difficult to enumerate and detect the pathogenic organisms due to their low numbers and specific growth requirement (Stewart et al., 2007). Fecal indicator organisms are used to monitor the microbial safety of various water systems for public health (Anderson et al., 2005). As these organisms reside in the gut of warm blooded animals, their presence indicates fecal contamination in soil or water systems (Ishii and Sadowsky, 2008b). Several bacteria are currently being used as fecal indicator, however a fecal indicator is required

to be 1. A part of the gut microbiota of warm blooded animal, 2. Present in parallel with pathogens or absent otherwise. 3. Able to survive similarly in a condition as pathogens 4. Present in higher numbers than pathogens. 5. Incapable of proliferation in the environment, 6. Non pathogenic in nature, and 7. Identified and enumerated by simple, rapid and inexpensive procedures (Bitton, 2005). Traditionally total coliforms, fecal coliform enterococci, and *E. coli* have all been used as fecal indicator organisms (Ashbolt et al., 2001;Boehm and Sassoubre, 2014 ;Payment et al., 2003). *E.coli*, *Enterobacter*, *Klebsiella* and *Citrobacter* all fall under coliforms, lactose fermenting gram negative *Enterobacteriacea* (Leclerc et al., 2001). Fecal coliforms basically refer to thermotrophic coliforms that can grow at a temperature up to 44.5 °C, which was previously suggested to be more specific to fecal contamination. However some thermotrophic coliform members such as *Kleibsella* can be found from non-fecal sources.

To detect the best indicator of fecal contamination the U.S Environmental Protection Agency (USEPA) linked bacterial presence to swimming related gastrointestinal disease. The report suggested that enterococci disease incidents are related to marine water and *E. coli* with fresh water. According to the USEPA Ambient water quality; fresh water beaches should not be accessible if the *E. coli* count of a single sample exceeds 235 colonies per 100 mL of water or the geometric mean of *E. coli* counts of at least 5 samples equally spread over a 30 day period exceeds 126 colonies per 100 mL (Boehm and Sassoubre, 2014).
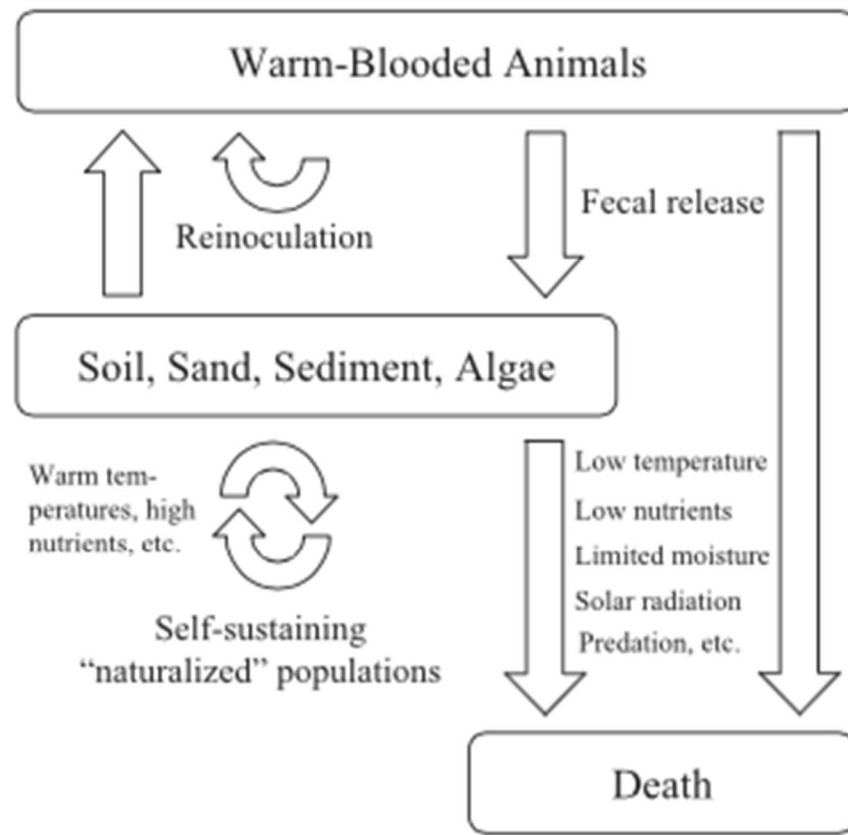
Fig.1.1 Schematic diagram of the lifecycle of *E. coli*. Once *E. coli* is released from its primary host (warm-blooded animals) through fecal droppings, the majority of the released bacteria die due to low nutrients and other environmental factors. Some of them however, become attached to soil, sand sediments or algae surfaces, and survive longer. In some conditions, these *E. coli* strains can grow and maintain their population long enough to become adapted or "naturalized" to the environment. The adapted or naturalized *E. coli* survive and replicate in the environment and can be reintroduced to animal hosts through contact with water and food. Adapted from (Ishii and Sadowsky, 2008b).

2.1. E. coli *in its Primary Habitat*

E. coli lifecycle occurs between two main habitats such as the intestine of warm blooded animals – the primary habitat, and outside of the host (water, sediments, soil, plants) the secondary habitat (Savageau, 1983). These two habitats have extremely different biotic and abiotic conditions and availability and types of nutrients. Savageau (1983) suggested that *E. coli* cells survive in this transition by having dual genetic regulatory mechanisms, where genes are positively controlled for certain function in one habitat and negatively in the other. According to Whittam (1989), selection pressure plays an important role while going through such transitions, which evolved strains that are primarily adapted to the primary habitat, while others are better adapted to the secondary habitat. Even though *E.coli* reside in the intestinal lumen of mammals as a commensal, certain group of *E. coli* can cause a widespread intestinal or extra-intestinal disease in humans and animals (Kaper et al., 2004;Croxen and Finlay, 2010). *E. coli* is classified in to three major groups based on its pathogenesis and genetic make up; commensal *E. coli*, intestinal pathogenic or diarrheagenic *E. coli*, and extra intestinal pathogenic *E. coli* (ExPEC) (Russo and Johnson, 2000).

2.1.1. Commensal *E. coli*

Commensal *E. coli* is one of the first bacteria to colonize shortly after birth, in the lower intestinal tract of human and other animal newborns including piglets, rats, mice and chickens (Benno et al., 1984;Penders et al., 2005;Palmer et al., 2007). There is also evidence of colonization of *E. coli* in fish and reptiles (cold blooded) at a suitable elevated temperature (Huggins and Rast, 1963;Gordon and Cowling, 2003a). *E. coli* is a

part of gut microbiome. The number of anaerobic bacteria in the gastrointestinal tract is much higher than *E. coli*, however, it is the predominant facultative anaerobic organism in the gut (Berg, 1996;Tenaillon et al., 2010). The density of *E. coli* varies among hosts; the occurrence in human is much higher compared to other animals, it is about 10^7-10^9 CFU per gram of feces (Slanetz and Bartley, 1957;Penders et al., 2006), the number in domestic animals is approximately between 10^4 and 10^6 CFU (Slanetz and Bartley, 1957). It has been reported that *E. coli* maintain a symbiotic relationship with other anaerobic gut microbiota. The anaerobes break down the complex polyscharides into simple mono or disaccharides so the *E. coli* can use them. On the other hand *E. coli* helps create an anaerobic environment by scavenging residual oxygen (Jones et al., 2007;Jones et al., 2011). Various strains of commensal *E. coli* together with other bacterial species form a stable gut microbiome, which prevents the colonization by invading pathogens, this barrier effect is called colonization resistance in host (Stecher and Hardt, 2011). Colonization of commensal *E. coli* Nissle 1917 and HS prevents colonization by pathogens by limiting nutrient availability (Maltby et al., 2013). The colonization of commensals also helps to stimulate the innate immune cells as well as the T-cell receptors (Williams et al., 2006;Petnicki-Ocwieja et al., 2009).

2.1.2. Diarrheagenic *E. coli*

Pathogenic *E. coli* are categorized into six pathotypes associated with diarrhea, they are collectively called diarrheagenic *E. coli*; 1. Shiga toxin producing *E. coli* or verocytotoxin producing *E. coli* or enterohemoragic *E. coli* (EHEC) 2. Enterotoxigenic *E.*

coli (ETEC), 3. Enteropathogenic *E. coli* (EPEC) 4. Enteroaggregative *E. coli* (EAEC) 5. Enteroinvasive *E. coli* (EIEC) 6. Diffusely Adherent *E. coli* (DAEC).

2.1.2.1. Shiga toxin producing *E.coli* (STEC)

STEC is commonly associated with most outbreaks of gastrointestinal diseases and poses a potential threat to public health (Heiman et al., 2015). STEC was first discovered in 1977 and its association with Hemolytic uremic syndrome was known in 1983 (Konowalchuk et al., 1977;Karmali et al., 1983). Enterohemorrhagic *E. coli* (EHEC), a subset of STEC is isolated from humans and responsible for several clinical symptoms such as hemorrhagic colitis and potential lethal hemolytic uremic syndrome (Karch et al., 2005;Karmali et al., 2010). Shiga toxin is the principal virulence factor in STEC infection and it consists of two main groups Stx1 and Stx2, and each group has several structurally and functionally similar toxin proteins. Besides Stx genes STEC strains often carry a gene coding for the adherence factor intimin (*eae*) which is an outer membrane protein (Nataro and Kaper, 1998). All STEC strains may not confer pathogenicity just by acquiring the *Stx* gene without having other virulence factors, however all EHEC strains are considered to be pathogenic as they expresses *Stx*, cause A/E lesions on epithelial cells, and possess a 60 MDa plasmid. A common example is *E. coli* O157-H7 (Levine, 1987;Nataro and Kaper, 1998). A single EHEC strain may express *Stx1* only, *Stx2* only or both or sometimes multiple forms of *Stx2*. Shiga toxin is a holotoxin composed of a single A subunit of approximately 33kDa in association with a B-subunit of 7.7kDa (Tesh and O'Brien, 1991). The B-pentamer of the toxin binds to a glycolipid receptor called Gb_3 present on the surface of eukaryotic cells. The A- subunit has N-glycosidase that removes

a single adenine from 28SrRNA of ribosomes, thereby inhibiting protein synthesis and leading to cell death (Nataro and Kaper, 1998).

2.1.2.2.   Enterotoxigenic *E. coli* (ETEC)

E. coli strains that possess genes for one of the two plasmid encoded enterotoxins; heat stable (ST) and heat labile are termed ETEC (Levine, 1987). The strain was first recognized in piglets as a cause of lethal infectious diarrhea (Alexander, 1994). ETEC strains cause diarrhea due the enterotoxins LT and ST, these strains may express only LT or only ST or both (Sears and Kaper, 1996;Hirayama and Wada, 2000). The LT toxins are oligomeric in nature and have two major serogroups LT-I and LT-II. LT-I is generally found in human *E. coli* isolates and is similar (almost 80%) to cholera enterotoxin (CT) expressed in *Vibrio cholera* (Sixma et al., 1993), and LT-II are mostly associated with non-human isolates (Nataro and Kaper, 1998;Qadri et al., 2005). LT toxins are AB_5 toxin (one A sub unit linked to a pentameric B subunit) and are transported through the bacterial membrane by a Type-II secretion system (Tauschek et al., 2002). LT-II shows 50-57% identity to LT-I and CT, but no substantial homology to the B-subunit (Guth et al., 1986;Pickett et al., 1989). Both LT-I and LT-II increase intracellular cAMP levels, resulting in osmotic diarrhea by a similar mechanism, however LT-I use GM-1 receptor and LT-II use GD1 receptor (Fukuta et al., 1988). LT-II is found in animal *E. coli* isolates and rarely in humans but there is no evidence of LT-II association in any human or animal diseases (Nataro and Kaper, 1998). STs are small monomeric toxins with cysteine residues and the disulfide bonds of cysteine account for the heat stability of this toxin. There are two classes of STs; ST-I or STa and ST-II/STb. STa toxin is commonly found

in humans and porcine isolates whereas STb toxins are primarily associated with porcine

isolates (Clements et al., 2012). Along with STa and STb ETEC strains have colonization

factors (surface fimbriae), which help them adhere and colonize in the intestinal mucosa

(Gaastra and Svennerholm, 1996), and the CF ETEC strains are mostly associated with

travellers' diarrhea and weanling diarrhea among children in developing countries. It is

also a major cause of weanling diarrhea in pigs (Amezcua et al., 2002;Qadri et al.,

2005;Daniels, 2006). The fimbriae or colonization factor shows species specificity, for

example; ETEC strains with K99 are pathogenic to calves, lambs and pigs whereas with

K88 are pathogenic to only pigs (Cassels and Wolf, 1995).

2.1.2.3.   Enteropathogenic *E. coli* (EPEC)

EPEC is another important group of diarrheagenic *E. coli* that has been a major cause of

infant diarrhea in the developing world. Attaching and effacing lesions in the ileum,

which appear like a pedestal, is a characteristic feature of EPEC infection. These types of

lesions are formed by the effacement of microvilli and intimate adherence of the

bacterium with the epithelium (Knutton et al., 1987;Jerse et al., 1990). There are several

virulence genes responsible for the pathogenicity of EPEC. An outer membrane protein

called intimin, encoded by *eae*, mediates intestinal cell attachment (Jerse et al., 1990). All

the genetic material required for A/E lesions is encoded by a large genomic pathogenicity

island called the locus of enterocyte effacement (LEE). It encodes a gene regulator,

structural components of a T3SS system, the bacterial surface protein intimin and a

number of translocator proteins (EspA, EspB, EspC and EspD) (Elliott et al., 1998). In

addition to LEE encoded effector genes there are also several non-LEE encoded (Nle)

effector genes classified in six pathogenicity island present throughout the genome, for

example Nle A-H, EspG2/Orf3, Cif, EspJ and EspL (Dean and Kenny, 2009). The EAF

plasmid encodes for BFP (bundle forming pilus) and a transcriptional regulator called Per

(plasmid encoded regulator) (Tobe et al., 1999;Khursigara et al., 2001). EPEC strain

E2348/69 (Serotype O127:H6) has been completely sequenced and widely used as a

prototype to study genetics, virulence and EPEC physiology (Iguchi et al., 2009).

2.1.2.4.   Enteroaggregative *E. coli* (EAEC)

EAEC are mainly associated with cases of acute and persistent diarrhea worldwide in

children and adults. It has a characteristic stacked brick aggregative adherence pattern

when attached to HEp-2 cells (Nataro and Kaper, 1998). Most EAEC strains harbor a 60-

65 MDa virulence plasmid (pAA), and a 1-Kb fragment of this plasmid that is known as

EAEC probe or CVD432 (Baudry et al., 1990) is commonly used for epidemiological

study (Elias et al., 2002;Scaletsky et al., 2002). pAA plasmid also encodes for AA

fimbriae (AAF) I, II and III(Nataro et al., 1992;Czeczulin et al., 1997;Bernier et al.,

2002), the transcriptional activator AggR(Nataro et al., 1994), enteroaggregative heat

stable enterotoxin 1(EAST-1) (Savarino et al., 1993), Pet; an autotransporter enterotoxin

(Eslava et al., 1998), and a novel antiaggregation protein dispersin encoded by the *aap*

gene (Sheikh et al., 2002). EAEC strains show a high degree of heterogeneous

pathogenicity. In a study by Nataro et al., (1995), EAEC strain 042 which produces

AAF/II- fimbriae, EASTI and the 108 kDa Pet toxin showed symptoms of infection in 4

out of 5 volunteers, whereas three other strains with AAF/I positive and one with EAST1

did not show any symptoms. This result suggests that Pet may play an important role in the pathogenicity of EAEC.

2.1.2.5.   Enteroinvasive *E. coli* (EIEC)

EIEC was first shown to cause diarrhea in 1971 in healthy volunteers as demonstrated by DuPont et al. (1971). The pathogenesis of EIEC strains is closely related to *Shigella,* the site of infection is colonic mucosa. The pathogenesis comprises of epithelial penetration, endocellular vacuole disruption, intracellular multiplication and infection to neighboring cells. It causes inflammation in the colon mucosa and in severe cases ulceration (Sansonetti, 1998). Both plasmid and chromosomal genes are involved in conferring pathogenesis. The *mxi* and *spa* loci in plasmid pInv (invasion related plasmid) encodes for the type III secretion system and IpaB, IpaC, IpaD are effector proteins, essential for invasion (Nataro and Kaper, 1998). VirG is a surface protein, that helps nucleation of actin filament and movement (Sansonetti, 1992). VirR, VirF and VirB chromosomally encoded proteins essential for regulatory cascade in *Shigella* virulence (Nataro and Kaper, 1998).

2.1.2.6.   Diffusely adherent *E. coli* (DAEC)

DAEC are characterized by their diffusely adherent pattern on epithelial HeLa or HEp-2 cells (Kaper et al., 2004). DAEC strains that harbor Afa/Dr family adhesins are able to cause enteric infection (Servin, 2005). Kyaw et al. (2003) demonstrated the presence of type three secretion system genes in DAEC. The secreted autotransporter toxin (SAT)

belonging to family serine protease autotransporter of *Enterobacteriaceae* (SPATE) is also considered to be an important virulence factor (Taddei et al., 2003).
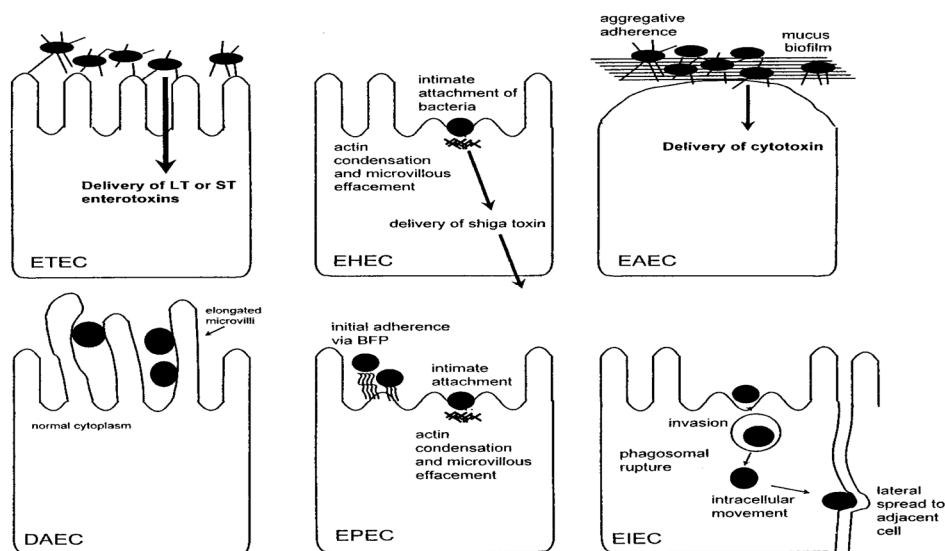
Fig. 1.2 Pathogenic schemes of diarrheagenic *E. coli*. The six recognized categories of diarrheagenic *E. coli* each have unique features in their interaction with eukaryotic cells. Here, the interaction of each category with a typical target cell is schematically represented. It should be noted that these descriptions are largely the result of *in vitro* studies and may not completely reflect the phenomena occurring in infected humans. Adapted from (Nataro and Kaper, 1998)

2.1.3.   Extra Intestinal Pathogenic *E. coli* (ExPEC)

The common extra intestinal diseases caused by ExPEC are urinary tract infection (UTI) (UPEC), neonatal meningitis (NMEC) and sepsis (SEPEC). In addition to this *E. coli* also cause intra-abdominal infections, nosocomial pneumonia, cellulitis, osteomyelitis and wound infection. (Johnson and Russo, 2002).

2.2. *Growth of E. coli in its primary habitat*

E. coli reside in the intestinal mucus layer as a part of the mixed biofilm and obtain nutrient to grow successfully. It has also been shown that each *E. coli* strain possesses a distinct nutritional practice in the intestine (Conway and Cohen, 2015). After ingestion *E.coli* survive the acid stress of the stomach by the stationary phase protective acid resistance system (Foster, 2004). Once they reach the colon, nutrients are necessary to grow from low to high numbers. The invading bacteria get eliminated if they fail to transition from lag to log phase. (Freter et al.,1983b). The ability of *E. coli* to colonize depends on several factors, but competition for nutrients is the main aspect for successful colonization in the intestine (Freter, 1992). The mammalian intestine is comparable to chemostat in which several hundred bacterial species live in equilibrium. To maintain this stable co-colonization each species must use one limiting nutrient better than all others (Freter et al., 1983b;Freter, 1988). The microbiota maintain a stable ecosystem in the healthy intestine, and resist the colonization of invading species. This ability of the resident microbiota to resist colonization is known as colonization resistance (Lawley and Walker, 2013).

E. coli is a facultative anaerobe with the ability to respire oxygen, use alternative electron acceptors or ferment depending on electron acceptor availability. The central metabolism system in *E. coli* consists of the EMP glycolytic pathway, the pentose phosphate pathway (PP), the Entner Doudoroff pathway (ED), the TCA cycle and diverse fermentation pathways. *E. coli* grows best on sugars, including a wide range of mono and disaccharides, but cannot grow on most of the complex polysaccharides as it does not possess the necessary hydrolytic enzymes (Fabich et al., 2008). It can also grow on amino

acids and carboxylates that are part of the TCA cycle. It depends on other intestinal anaerobes that hydrolyze complex polysaccharides to mono or disaccharides (Salyers and Pajeau, 1989;Goodman et al., 2009). The sources of nutrients that support intestinal colonization of *E. coli* are shed epithelial cells, dietary fibers and mucosal polysaccharides (Conway et al., 2004). Sugars released through hydrolysis of dietary fibers by other, mostly anaerobic species. Most of the required amino acids are available in the large intestine for the growth. *E. coli* strains have nearly identical catabolic potential, but they vary differentially in the sugars that support their colonization. For example pathogenic strains are predicted to grow on Sucrose while commensals are not. In contrast commensals are predicted to grow on galactonate while pathogens are not (Chang et al., 2004;Fabich et al., 2008;Maltby et al., 2013).

There is competition for limiting resources in the intestinal ecosystem (Tilman, 1982). The microbial community competes for carbon and energy sources, and terminal electron acceptors (Freter et al., 1983a). To maximize its population, *E. coli* uses various strategies. In a nutrient limiting condition *E. coli* is able to utilize up to 9 different sugars simultaneously (Fabich et al., 2008). It has the ability to induce a number of gene systems for carbon source transport and catabolism when growing slowly (Ihssen and Egli, 2005) or under nutrient deprived condition (Liu et al., 2005). The metabolic capacity of the cell expands in hunger states due to physiological and genetic changes (Ferenci, 2001). To maintain stable colonization in the intestine, *E. coli* must compete for limiting nutrient, so it uses several sugars at a time, and also has the capacity to metabolize glycogen carbon store in the intestine. *E. coli* compete for nutrients in the intestine in three ways; first, it can use nutrient that no other species in the community has used. Second, it can

outcompete other strains for the nutrients it prefers by superior uptake systems and faster growth (Fabich et al., 2011). Third, it can enter in to a symbiotic association with the anaerobes that release its preferred sugars (Leatham-Jensen et al., 2012).

2.3.   E. coli *in its secondary habitat (in the environment outside of hosts)*

It has been observed that most *E. coli* enter to the environment outside of their host at certain time periods, and may spend half their life there (Savageau, 1983;Gordon, 2001). The secondary habitat of *E. coli* differs from the primary habitat in large extent in both biotic and abiotic conditions. The gastrointestinal environment provides an optimal constant temperature, and availability of amino acids and sugars, which favor the growth of *E. coli* (Savageau, 1983). In the secondary environment *E. coli* struggles to survive the limited nutrient availability, osmotic stress, variable temperature and pH, UV radiation and predation (Rozen and Belkin, 2001;Brennan et al., 2010b). All these conditions lead to a decrease in density of specific strains in the secondary environment and often to undetectable levels (Ishii and Sadowsky, 2008b). As a result, it is believed that *E. coli* cannot grow in the external environment most of the time, the reason for its use as an indicator organism (Solo-Gabriele et al., 2000;Walk et al., 2007;Odonkor and Ampofo, 2013).

Recent reports suggest that *E. coli* are capable of surviving and even multiplying in the external environment in the absence of fecal contamination in various climatic conditions. Although most *E. coli* strains are commensals, many strains have diverged to take on a pathogenic life style. There are two schools of thought; *E. coli* originated from fecal contamination in the past, and over time, some strains have adapted to replicating

outside of their mammalian host and eventually form the part of the natural microbiota. A second group developed the hypothesis that *E. coli* was always part of the microbiota in the external environment and the same strains acquired the ability to cause disease in human and animal host. If either of these two scenarios is correct then the use of *E. coli* as an effective indicator organism is questionable (Power et al., 2005).

2.3.1. *E. coli* associated with soil, sands, sediment and aquatic plants

A substantial amount of research have been reported that *E. coli* is able to grow and persist in soil, beach, sand, sediments, water, and aquatic plants in various climatic condition (tropical or temperate) (Jang et al., 2017).

 Tropical soil environments have appropriate nutrients to support the growth of *E. coli* in the presence of other soil microbes (Byappanahalli and Fujioka, 1998). High growth rate of *E. coli* within river bank soil has been observed in a coastal subtropical region, where it serve as a constant source of *E. coli* while mixed with water in high tide or rain (Solo-Gabriele et al., 2000). Colonization and persistence of certain *E. coli* genotypes in the freshwater beaches of Lake Michigan was observed (Walk et al., 2007). Persistence and differential growth of *E. coli* have been reported in sediments and water columns of subtropical regions (Anderson et al., 2005). Naturalized *E. coli* strains had the ability to grow and maintain a population up to 10^5 CFU per gram of non-sterile non amended temperate soil of Lake Superior Watershed (Ishii et al., 2006). *E. coli* can occur and persist for an extended period of time in undisturbed temperate forest soil irrespective of any immediate contamination or seasonal changes (Byappanahalli et al., 2006). Several studies have suggested that beach sands and sediments may act as a reservoir for *E. coli*

(Wheeler Alm et al., 2003;Whitman and Nevers, 2003;Shibata et al., 2004;Ishii et al., 2007).

Marine/freshwater algae are reservoirs of many microbes (Egan et al., 2013), and these microbes play important role in the development, defense and metabolism of plants (Tujula et al., 2010). Studies have shown that at freshwater lakes, the macro alga *Cladophora* is an important reservoir for *E. coli,* and the algal mat provides a suitable environment for availability of nutrient from the exudates of alga, protection from UV rays and predation (Byappanahalli et al., 2003;Whitman et al., 2003;Englebert et al., 2008;Vanden Heuvel et al., 2010).

A study on coliform bloom of Australian lakes showed that *E. coli* isolated from the bloom are encapsulated, suggesting that some *E. coli* evolved a free living lifestyle and do not need any host in order to proliferate (Power et al., 2005).

From the above studies it could be implied that a high number of *E. coli* in environmental samples may not always be linked to fecal contamination.

2.3.2. Response of *E. coli* to Environmental challenges

E. coli survive and multiply in the secondary environment, in association with water, sand, sediments, and green algae. Soil and sediments in tropical and subtropical regions may allow the colonization of *E. coli* as it provides a suitable environment for growth temperature, nutrition availability, and protection from UV rays and predation (Wheeler Alm et al., 2003;Brennan et al., 2010a). It has been suggested that, to maintain the autochthonous population outside of its host, conditions should remain favorable. *E. coli*

survive in unfavorable temperature conditions like freezing cold of the winter months, and then when the temperature increases in the summer months the same strains of *E. coli* are able to multiply (Ishii et al., 2006).

The ability of *E. coli* to adapt and survive in the secondary environment may also be the result of its versatility in acquisition of energy sources or nutrients (Iuchi and Lin, 1993;van Elsas et al., 2011). It is able to survive on minimal carbon and nitrogen sources (van Elsas et al., 2011) and has the ability to utilize various aromatic compounds (Díaz et al., 2001). It is, therefore, inferred that *E. coli* with its ability to utilize various energy sources, grow at various temperatures, and both under aerobic and anaerobic conditions, and remain an integral part of the microbial communities in various environments (Ishii and Sadowsky, 2008b). It is, however, still not very clear what the fate of *E. coli* is under complex natural conditions. It is important to understand the environmental factors that affect the survival of *E. coli* in secondary environments, as in such habitats it has to face fluctuating or low nutrient condition, high or low oxygen levels, fluctuating temperatures, high or low pH, or high osmolarity.

2.3.3. Availability of resources

The availability of resources such as carbon substrate is an important factor which may affect the survival and growth of *E. coli* in the secondary environment. *E. coli* is a chemoheterotroph, and its survival depends on the acquisition of sufficient carbon compounds. To adapt to glucose-limited conditions, *E. coli* cells were shown to be prepared to efficiently take up various carbon sources by the upregulation of a large number of genes that encode periplasmic binding proteins (Franchini and Egli, 2006).

The level of RpoS increases upon glucose starvation, and upregulates the expression of a number of genes that help to combat various stresses (Mandel and Silhavy, 2005). *E. coli* can also assimilate a variety of nitrogen sources including ammonia. It responds to nitrogen limitation by activating nitrogen stress response (Ntr), which facilitates N scavenging from alternative sources by expressing about 100 of genes (Reitzer, 2003). Nitrogen starved *E. coli* cells synthesize a signal molecule, guanosine tetraphosphate (ppGpp), which serves as an effector molecules and initiates physiological changes known as the stringent response (Brown et al., 2014). *E. coli* was also found to show a high degree of catabolic flexibility, and often in nutrient limiting conditions different catabolic functions and binding protein become activated. This confers its fitness advantages to survive in the open environment (Ihssen and Egli, 2005).

2.3.4. Temperature

E. coli growth and survival is also influenced by temperature. In animal hosts the body temperature is usually stable, whereas it often fluctuates in the outside environment such as soil and water. Many survival studies of *E. coli* have been carried out under stable temperature conditions (Kudva et al., 1998;Franz et al., 2005). The effect of fluctuating temperature on survival and adaptation of *E. coli* in soil and water is not clearly understood. Studies on *E. coli* O157-H7 showed that survival in fluctuating temperature is generally lower compared to at stable temperature under manure (Semenov et al., 2007). It has been reported that *E. coli* Dh5 alpha can grow at an elevated temperature up to 49 °C, perhaps because of mutations that permit growth at this high temperature (Fotadar et al., 2005). *E. coli* showed a decreased level of gene expression involved in

glycolysis, PTS sugar transport, amino acid synthesis and transport at reduced temperatures (33 and 28C°) (Gadgil et al., 2005). Moreover gene expression patterns may keep altering with fluctuating temperature. The histone like nucleotide structuring protein in *E. coli* controls a majority of thermoregulatory genes such as multiple iron and other nutrient acquisition systems expressed at 37 °C and stress response, biofilm formation and cold shock genes expressed at 23C° (White-Ziegler and Davis, 2009). Hence in *E. coli* differential gene regulation may occur at a wide temperature range to adapt to the changing temperature in the environment.

In a long term study, *E. coli* survived for more than 260 d in autoclaved river water at a temperature range from 4°C to 25°C (Flint, 1987), which suggests that competition with other microbes in water was a cause for decline of *E. coli*.

2.3.5. Salinity / Osmolarity

The ability of *E. coli* to adapt to the fluctuation of ambient osmolarity in the secondary environment is of primary importance for their survival. A quick rise in the osmolarity of the environment causes a threat to the growth and survival of *E. coli*, due to water loss and decreased turgor of cells. At moderate to high osmolarities, potassium glutamate and other compatible solutes like polyols i.e. trehalose, aminoacid- proline and methyl amine-glycine betain are important osmoprotectants (Munro et al., 1989;Lucht and Bremer, 1994). The cells activated the osmoregulating system under increased osmotic pressure in the surrounding environment, which prevents cell shrinkage and plasmolysis. Osmotic pressure induces the Pex starvation protein, as well as HSPs in *E. coli*. In addition, high

osmolarity in *E. coli* cells is associated with reduced DNA replication, nutrient uptake and cell growth (Chung et al., 2006)

2.3.6. pH / acid stress

Acidic condition affects the cell function by interfering with nutrient acquisition, pH homeostasis in the cytoplasm, protein stability and integrity of DNA (Booth et al., 2002). *E. coli* encounter a number of potential acid stresses in nature, and they are capable of sensing and responding to this stress as a process of their protection mechanism. *E. coli* can survive many hours in extremely acidic (pH 2.h or lower) conditions by the use of effective acid resistance systems (ARs). There are three types of ARs, AR1 is glucose repressed and the stationary phase alternative sigma factor (RpoS) is required to develop acid tolerance, AR2 is based on glutamate decarboxylation, involves GadA and GAdB decarboxylase enzymes and GadC membrane transport protein to import glutamate, and AR3 requires external supply of arginine by the AdiC membrane transporter and AdiA arginine decarboxylase enzyme (Lin et al., 1995;Castanie-Cornet et al., 1999;Foster, 2004). It has been shown that OmpR regulates the components of the transcriptional program under acid stress conditions (Stincone et al., 2011). Acid resistance in *E. coli* has been found associated with an increased amount of cyclopropane fatty acid (CFA) in the membrane, which reduces the membrane permeability to protons (Brown et al., 1997;Chang and Cronan, 1999). It has been observed in several studies that *E. coli* O157:H7 in particular survives in low pH (Benjamin and Datta, 1995;Lin et al., 1996), and different strains showed different capacities to survive acid stress (Lin et al., 1996),

but in comparison to non O157 EHEC, they all do better in their survival (Bergholz and Whittam, 2007).

2.3.7. Oxidative stress

Oxidative stress occurs in cells due the production of reactive oxygen species (ROS), natural byproduct of aerobic metabolism. It can be detrimental to the cell by damaging several cellular sites like iron sulfur clusters, cysteine and methionine protein residues and DNA (Storz and Imlayt, 1999). *E. coli* response to superoxide and peroxide stress by inducing superoxide dismutases (SOD) which dismute superoxide to H_2O_2 (McCord and Fridovich, 1969) and catalases subsequently degrade H_2O_2 into H_2O and O_2.

2.3.8. Solar Radiation

Solar radiation is another abiotic factor that causes death of *E.coli* in environmental water and soil. Solar radiation, especially UV light can directly cause DNA damage or oxidation of cellular contents but these mechanisms is effective when cells are present on the soil surface or water surface where sunlight can reach. The effect of sunlight on *E. coli* survival may vary by exposure time or turbidity of the water environment. The impact of sunlight is less in soil and sediment than the water environment (Whitman et al., 2004).

2.3.9. Biotic Factors or other microbial communities

In natural habitats *E. coli* interact with other microbial communities. For survival it has to compete with other indigenous microorganisms for limited nutrients and also defend against antagonistic effects in the environment. It can also be susceptible to protozoan predation and lysed by phages. It has been reported that these two mechanisms can remove fecal indicator bacteria from river water very effectively (Korajkic et al., 2014). There is an increase in the *E. coli* population in sterile soil compared to non-sterile soil indicating that microbial communities play a crucial role in the survival of *E. coli* in the environment (Unc et al., 2006;Ishii et al., 2010a).

2.3.10. Ability to form Biofilm

Forming biofilm may be a survival strategy of *E.coli* that helps them to persist in the natural environment. Biofilms provide protection from hostile environmental condition such as desiccation, UV light, protozoan predation, antibiotics or disinfectants (McDougald et al., 2011). The bacteria may also use it as a source of nutrients (Jang et al., 2017). In addition to protection offered by the biofilm against different chemicals, a slow growing dormant sub-population called persister cells emerged more often from biofilm populations than planktonic populations (Lewis, 2010).

2.4. **E. coli *Diversity and Population genetics***

Commensal and pathogenic *E. coli* strains display diverse phenotype and genotype variants. Multiple factors from both the host and environment shape the genetic structure of *E. coli.* To characterize how *E. coli* adapts to different niches it is necessary to unravel

how the species is genetically structured on a global scale. The balance between recombination and mutation largely defines a population structure. It keeps shifting from a clonal structure where mutation is low to a panmictic structure when recombination is high (Tenaillon et al., 2010). High genotypic diversity in *E. coli* has been identified by various DNA finger print technique i.e. repetitive extragenic palindromic (REP) PCR (Byappanahalli et al., 2012;Jang et al., 2014) and pulse field gel electrophoresis (Johnson et al., 2013). *E. coli* strains vary in their phenotypic characteristics as well, such as their ability to form biofilm, substrate utilization, antibiotic resistance and so on. This diversity of *E. coli* has been explained by the effect of the genomic makeup of the organism residing in the host intestine or in the natural environment (van Elsas et al., 2011). Horizontal gene transfer plays an important role in the acquisition of new genes, where gene mutations also contribute in *E. coli* phenotypic diversity, such as nutrient utilization. In multiple *E. coli* genomes, strain specific or group specific genes increase the pangenome size, suggesting the impact of horizontal gene transfer for genome plasticity of *E. coli* (Touchon et al., 2009).

The environmental conditions of *E. coli* habitat have been suggested to influence the genetic structure (Tenaillon et al., 2010). Unique genotypes representing environmental adapted strains have been reported in many studies (Anderson et al., 2005;Ishii and Sadowsky, 2008b;Byappanahalli et al., 2012). The comparative genome analysis between environmentally adapted strains and other enteric *E. coli* have broadened the understanding of the evolutionary lineage of this bacterium. Using an extended multi-locus sequence typing (MLST) approach, Walk et al. (2009) identified and characterized novel *Escherichia* clades (CI, CIII, CIV and CV). Whole genome phylogenetic analysis

of nine clade isolates and other commensal isolates showed that clade isolates were distinctly divergent from other enteric *E. coli* (Luo et al., 2011a). Therefore, genotypically the clade isolates may represent novel species even though they are indistinguishable by traditional phenotypic test and 16S rRNA gene based phylogenetic analysis (Luo et al., 2011a). Genetic exchange of core genes was detected, but only with the environmental clades or within the enteric strains and not between the two groups (Luo et al., 2011a). This result indicates the presence of a possible ecological barrier to gene flow between environmental and enteric strains. Environmental and commensal strains may behave differently as per their growth and survival mechanisms. Comparative transcriptome analysis suggests that environmental strains appeared to be better adapted in low nutrient conditions (Vital et al., 2015).

2.4.1. Tools for studying *E. coli* population genetics

Four main techniques have been used to study population genetics of *E. coli*

2.4.1.1. Serotyping

E.coli are serotyped based on a combination 173 O antigen (somatic), 80 K antigen (capsular) and 56 H antigens (flagellar). So far 700 *E. coli* isolates have been serotyped based on the O and H antigen combination (Nataro and Kaper, 1998). These methods have been used for many years for differentiating and characterizing *E. coli*. However, these methods are usually time consuming and not always accurate. Advanced in next generation sequencing technologies have made is possible to develop genetic based

subtyping and molecular serotyping method for *E. coli*, which is more discriminatory compared to phenotypic serotyping methods.

2.4.1.2. Multilocus Enzyme Electrophoresis-MLEE

Isolates are characterized by the relative electrophoretic mobility of several water - soluble housekeeping cellular enzymes. The variation in the mobility of an enzyme can directly relate to alleles at the corresponding locus, or in other words mutations in the gene locus that cause amino acid substitution in an enzyme coded by the gene. The allele at each locus is defined as electrophoretic type, and the relatedness of isolates can be visualized by a dendrogram produced from a matrix of pairwise differences between electrophoretic types (Selander et al., 1986).

2.4.1.3. Multi locus Sequence Typing –MLST

MLST is a powerful tool for bacterial population genetics. It refers to the systematic sequencing of six to ten well conserved housekeeping genes or loci within the bacterial genome. Allelic variation of each locus is listed, and sequence type (ST) or lineage is assigned by comparing to other isolate profiles in the database. The relatedness of isolates can be visualized by constructing a phylogenetic tree from the nucleotide sequences. Currently three MLST schemes are available for *E. coli*, each scheme uses a different combination of genes. The characteristics of the three schemes are summarized in Table1 below.

Table. 1 Main characteristics of the three *E. coli* MLST databases

Location	Genes	Website	No. STs	No. Strains
Michigan State University	*aspC, clpX, fadD, icd, lysP, mdh, uidA*	http://www.shigatox.net/ecmlst/cgi-bin/index	1081	3965
Warwick Medical School	*adk, fumC, gyrB, icd, mdh, purA, recA*	http://mlst.warwick.ac.uk/mlst/dbs/Ecoli	4499	7583
Pasteur Institute	*dinB, icd, pabB, polB, putP, trpA, trpB, uidA*	http://www.pasteur.fr/recherche/genopole/PF8/mlst/EColi.html	771	1311

2.4.1.4.   Phylogrouping Triplex PCR to multiplex PCR

This method allows strains to be assigned to one of the four main phylogroups A, B1, B2, and D. Since its introduction in 2000, it has been widely used because of its simplicity and fast results. This method is based on triplex PCR using the combination of three genes *yjaA, ChuA* and TSPE4.C2 (Clermont et al., 2000) Recently *E. coli* phylogenetic grouping has been revised based on multi locus sequence typing and genome sequence data and four new groups C, E, F and *Escherichia* Clade-I were added (Clermont et al., 2013). A multiplex PCR method was developed to rapidly classify *E. coli* strains into one of the seven phylogroups and *Escherichia* Clade-I without performing MLST or sequencing (Clermont et al., 2013). In general, strains belonging to different phylogroup display different phenotypic and genotypic traits (Gordon, 2004;Meric et al., 2013), so it is possible that phylogenetic group recognition of unknown *E. coli* isolates may provide some important information regarding their ecology and physiology.

This is the era of next generation sequencing, and it is now possible to study hundreds of strains to help understand the whole genome level and the evolutionary processes acting in a population; opening the era of population genomics (Liti et al., 2009;MacLean et al., 2009).

2.4.1.5. Pulse Field Gel Electrophoresis (PFGE)

PFGE allows the separation of DNA molecules over 1000 kb, following restriction digestion of genomic DNA with a soft agarose plug, DNA molecules are separated in agarose gel, periodically applying changes in the direction of the electrical field in which large DNA molecules are suspended. The unique restriction patterns of each isolate are then compared to one another to determine relatedness (Sander et al., 1998).

2.4.1.6. Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR

ERIC sequences are 127 base pair motifs that occurs in multiple copies across the genome of enteric bacteria or vibrios. ERIC PCR uses a combination of primers targeting to the conserved region in order to generate an electrophoretic banding pattern based on the frequency and orientation of ERIC sequences in the bacterial genome. The specific band pattern of amplified PCR products obtained using the sequences can be used to genotype the bacteria (Versalovic et al., 1991).

2.4.1.7.   Repetitive Extragenic Palindromic (REP) PCR

This method is used for fingerprinting bacterial genomes by examining strain specific pattern obtained from PCR amplification of repetitive DNA elements present within bacterial genomes (Versalovic et al., 1991). REP elements are 38 bp sequences consisting of six degenerate positions and a 5 bp variable loop between each side of a conserved palindromic stem (Stern et al., 1984). Both ERIC and REP PCR can be performed with a single primer, a single set of primers, or multiple sets of primers.

2.4.1.8.   Arbitrarily primed (AP) PCR or Random Amplified Polymorphic DNA Assay (RAPD)

It involves the use of a single arbitrary primer in the PCR reaction, resulting in the amplification of many discrete DNA products. This procedure detects nucleotide sequence polymorphisms in a DNA amplification based assay using only a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer binds to the genomic DNA at two different sites on opposite strands of the DNA template.

2.4.1.9.   Amplified Fragment Length Polymorphism (AFLP)

AFLP is based on the selective amplification of a subset of DNA fragments generated by restriction enzyme digestion (Vos et al., 1995). Two variation of AFLP have been described, one with two different restriction enzymes and two primers for amplification and a second with a single primer and restriction enzyme. The restriction fragments are then ligated to linkers containing each restriction site and a sequence homologous to a

PCR binding site. The PCR primers used for amplification contain DNA sequences homologous to the linker and contain one to two selective bases at their 3' ends. For example, a selective primer directed against an EcoRI site might have the sequence 5'-GAATTCAA-3', where the first six bases are complementary to the EcoRI site while the two A residues at the 3' end are selective and allow amplification of only those EcoRI sites with the sequence 3' -CTTAATT-5'. They would not amplify an EcoRI site with the sequence 3'-CTTAATC-5'. Thus, the selective nucleotides allow amplification of only a subset of the genomic restriction fragments. The banding patterns are used to compare the isolates, and it has a good reproducibility and ability to differentiate clonally derived strains (Olive and Bean, 1999).

3. Conclusions and Future Perspectives

The presence of environmental *E. coli* is now well established based on published reports over the last few decades. These *E. coli* may be of animal origin and have become naturalized in their surrounding environments; or may retain the genetic traits of their ancestral lineage which was environmental bacteria residing primarily in soils and sediments. Furthermore, these environmentally occurring *E. coli* may be genetically different from their commensal or pathogenic counterparts as a consequence of their adaptation to environment. If this is the situation then the suitability of *E. coli* as an indicator organism is highly questionable. To understand the evolutionary biology of *E. coli* it is important to elucidate the relationship between gene content and adaptation to the ecological niche. Strains within each phylogroup occupy various ecological niches, however there is no clear co-relation between phylogroups and their niches. There is

variation in the distribution of phylogroups in different niches. Occurrence of genes or gene cluster in the genomic island across *E. coli* strains of diverse phylogroups may enable them to thrive in multiple niches. The presence of *E. coli* in the secondary environment raises many questions: (i). can environmental *E. coli* still colonize the intestinal tract of warm-blooded animals, (ii). What mechanisms enable them to survive in multiple niches (iii). Are these bacteria potentially pathogenic to humans? Further research is needed to answer these questions.

4. References

Abbott, S.L., O'connor, J., Robin, T., Zimmer, B.L., and Janda, J.M. (2003). Biochemical Properties of a Newly Described *Escherichia* Species, *Escherichia albertii*. *Journal of Clinical Microbiology* 41**,** 4852-4854.

Alexander, T.J.L. (1994). *Neonatal diarrhoea in pigs.* In: *Escherichia coli* in domestic animals and humans, Wallingford, UK: Cab International.

Amezcua, R., Friendship, R.M., Dewey, C.E., Gyles, C., and Fairbrother, J.M. (2002). Presentation of postweaning *Escherichia coli* diarrhea in southern Ontario, prevalence of hemolytic *E. coli* serogroups involved, and their antimicrobial resistance patterns. *Canadian Journal of Veterinary Research* 66**,** 73-78.

Anderson, K.L., Whitlock, J.E., and Harwood, V.J. (2005). Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments. *Applied and Environmental Microbiology* 71**,** 3041-3048.

Ashbolt, N.J., Grabow, W., O.K. , and Snozzi, M. (2001). *Water Quality: Guidelines, Standards and Health.* London, UK: IWA Publishing.

Baudry, B., Savarino, S.J., Vial, P., Kaper, J.B., and Levine, M.M. (1990). A sensitive and specific DNA probe to identify enteroaggregative *Escherichia* coli, a recently discovered diarrheal pathogen. *Journal of Infectious Diseases* 161**,** 1249-1251.

Benjamin, M.M., and Datta, A.R. (1995). Acid tolerance of enterohemorrhagic *Escherichia coli. Applied and Environmental Microbiology* 61**,** 1669-1672.

Benno, Y., Sawada, K., and Mitsuoka, T. (1984). The intestinal microflora of infants: composition of fecal flora in breast-fed and bottle-fed infants. *Microbiology and immunology* 28**,** 975-986.

Berg, R.D. (1996). The indigenous gastrointestinal microflora. *Trends Microbiol* 4**,** 430-435.

Bergholz, T.M., and Whittam, T.S. (2007). Variation in acid resistance among enterohaemorrhagic *Escherichia* coli in a simulated gastric environment. *Journal of Applied Microbiology* 102**,** 352-362.

Bernier, C., Gounon, P., and Le Bouguenec, C. (2002). Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia* coli as a sensitive probe for detecting the AAF-encoding operon family. *Infection and Immunity* 70**,** 4302-4311.

Bitton, G. (2005). "Microbial Indicators of Fecal Contamination," in *Wastewater Microbiology*. John Wiley & Sons, Inc.), 153-171.

Boehm, A.B., and Sassoubre, L.M. (2014). *Enterococci* as Indicators of Environmental Fecal Contamination. Boston: Massachusetts Eye and Ear Infirmary.

Booth, I.R., Cash, P., and O'byrne, C. (2002). Sensing and adapting to acid stress. *Antonie van Leeuwenhoek* 81**,** 33-42.

Brennan, F.P., Abram, F., Chinalia, F.A., Richards, K.G., and O'flaherty, V. (2010a). Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil. *Applied and Environmental Microbiology* 76**,** 2175-2180.

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

Brown, D.R., Barton, G., Pan, Z., Buck, M., and Wigneshweraraj, S. (2014). Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nature Communications* 5.

Brown, J.L., Ross, T., Mcmeekin, T.A., and Nichols, P.D. (1997). Acid habituation of *Escherichia coli* and the potential role of cyclopropane fatty acids in low pH tolerance. *International Journal of Food Microbiology* 37, 163-173.

Byappanahalli, M.N., and Fujioka, R.S. (1998). Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water. Sci. Technol.* 38, 171-174.

Byappanahalli, M.N., Shively, D.A., Nevers, M.B., Sadowsky, M.J., and Whitman, R.L. (2003). Growth and survival of *Escherichia coli* and enterococci populations in the macro-alga Cladophora (Chlorophyta). *FEMS Microbiology Ecology* 46, 203-211.

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

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

Cassels, F., and Wolf, M. (1995). Colonization factors of diarrheagenic *E. coli* and their intestinal receptors. *Journal of Industrial Microbiology* 15, 214-226.

Castanie-Cornet, M.-P., Penfound, T.A., Smith, D., Elliott, J.F., and Foster, J.W. (1999). Control of Acid Resistance in *Escherichia coli*. *Journal of Bacteriology* 181, 3525-3535.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., and Conway, T. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proceeding of Natural Academy of Sciece U S A* 101, 7427-7432.

Chang, Y.-Y., and Cronan, J.E. (1999). Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. *Molecular Microbiology* 33, 249-259.

Chung, H.J., Bang, W., and Drake, M.A. (2006). Stress Response of *Escherichia coli*. *Comprehensive Reviews in Food Science and Food Safety* 5, 52-64.

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

Clermont, O., Bonacorsi, S., Bingen, E., and Bonacorsi, P. (2000). Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Applied Environmental Microbiology* 66.

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

Conway, T., and Cohen, P.S. (2015). Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbiology Spectrum* 3.

Conway, T., Krogfelt, K.A., and Cohen, P.S. (2004). The Life of Commensal *Escherichia coli* in the Mammalian Intestine. *EcoSal Plus* 1.

Croxen, M.A., and Finlay, B.B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Review Microbiology* 8**,** 26-38.

Czeczulin, J.R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M.H., Navarro-Garcia, F., and Nataro, J.P. (1997). Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. *Infection and Immunity* 65**,** 4135-4145.

Daniels, N.A. (2006). Enterotoxigenic *Escherichia coli*: Traveler's Diarrhea Comes Home. *Clinical Infectious Diseases* 42**,** 335-336.

Dean, P., and Kenny, B. (2009). The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. *Curr Opin Microbiol* 12**,** 101-109.

Díaz, E., Ferrández, A., Prieto, M.A., and García, J.L. (2001). Biodegradation of Aromatic Compounds by *Escherichia coli*. *Microbiology and Molecular Biology Reviews* 65**,** 523-569.

Doolittle, R.F., Feng, D.F., Tsang, S., Cho, G., and Little, E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 271**,** 470-477.

Dupont, H.L., Formal, S.B., Hornick, R.B., Snyder, M.J., Libonati, J.P., Sheahan, D.G., Labrec, E.H., and Kalas, J.P. (1971). Pathogenesis of *Escherichia coli* diarrhea. The *New England Journal of Med* 285**,** 1-9.

Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013). The seaweed holobiont: understanding seaweed-bacteria interactions. *FEMS Microbiology Reviews* 37**,** 462-476.

Elias, W.P., Uber, A.P., Tomita, S.K., Trabulsi, L.R., and Gomes, T.a.T. (2002). Combinations of putative virulence markers in typical and variant enteroaggregative *Escherichia coli* strains from children with and without diarrhoea. *Epidemiology Infection* 129**,** 49-55.

Elliott, S.J., Wainwright, L.A., Mcdaniel, T.K., Jarvis, K.G., Deng, Y.K., Lai, L.C., Mcnamara, B.P., Donnenberg, M.S., and Kaper, J.B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Molecular Microbiology* 28**,** 1-4.

Englebert, E.T., Mcdermott, C., and Kleinheinz, G.T. (2008). Effects of the nuisance algae, Cladophora, on *Escherichia coli* at recreational beaches in Wisconsin. *Science of Total Environment* 404**,** 10-17.

Eslava, C., Navarro-Garcia, F., Czeczulin, J.R., Henderson, I.R., Cravioto, A., and Nataro, J.P. (1998). Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. *Infection and Immunology* 66**,** 3155-3163.

Fabich, A.J., Jones, S.A., Chowdhury, F.Z., Cernosek, A., Anderson, A., Smalley, D., Mchargue, J.W., Hightower, G.A., Smith, J.T., Autieri, S.M., Leatham, M.P., Lins, J.J., Allen, R.L., Laux, D.C., Cohen, P.S., and Conway, T. (2008). Comparison of Carbon Nutrition for Pathogenic and Commensal *Escherichia coli* Strains in the Mouse Intestine. *Infection and Immunity* 76**,** 1143-1152.

Fabich, A.J., Leatham, M.P., Grissom, J.E., Wiley, G., Lai, H., Najar, F., Roe, B.A., Cohen, P.S., and Conway, T. (2011). Genotype and phenotypes of an intestine-adapted *Escherichia coli* K-12 mutant selected by animal passage for superior colonization. *Infect Immun* 79**,** 2430-2439.

Farmer, J.J., Farmer, M.K., and Holmes, B. (2010). "The *Enterobacteriaceae*: General Characteristics," in *Topley & Wilson's Microbiology and Microbial Infections*. John Wiley & Sons, Ltd).

Farmer, J.J., Iii. (1999). *Enterobacteriaceae:* introduction and identification. Washington, D.C.: American Society for Microbiology.

Ferenci, T. (2001). Hungry bacteria--definition and properties of a nutritional state. *Environmental Microbiology* 3**,** 605-611.

Flint, K.P. (1987). The long-term survival of *Escherichia coli* in river water. *Journal of Applied Bacteriology* 63**,** 261-270.

Foster, J.W. (2004). *Escherichia coli* acid resistance: Tales of an amateur acidophile. *Nature Reviews Microbiology* 2**,** 898-907.

Fotadar, U., Zaveloff, P., and Terracio, L. (2005). Growth of *Escherichia coli* at elevated temperatures. *Journal of Basic Microbiology* 45**,** 403-404.

Franchini, A.G., and Egli, T. (2006). Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiology* 152**,** 2111-2127.

Franz, E., Van Diepeningen, A.D., De Vos, O.J., and Van Bruggen, A.H. (2005). Effects of cattle feeding regimen and soil management type on the fate of *Escherichia coli* O157:H7 and *Salmonella enterica* serovar typhimurium in manure, manure-amended soil, and lettuce. *Applied Environmental Microbiology* 71**,** 6165-6174.

Freter, R. (1988). Mechanisms of bacterial colonization of the mucosal surfaces of the gut. Washington, D.C: American Society for Microbiology.

Freter, R. (1992). "Factors affecting the microecology of the gut," in *Probiotics: The scientific basis*. (Dordrecht: Springer Netherlands), 111-144.

Freter, R., Brickner, H., Botney, M., Cleven, D., and Aranki, A. (1983a). Mechanisms That Control Bacterial Populations in Continuous-Flow Culture Models of Mouse Large Intestinal Flora. *Infection and Immunity* 39**,** 676-685.

Freter, R., Brickner, H., Fekete, J., Vickerman, M.M., and Carey, K.E. (1983b). Survival and Implantation of *Escherichia coli* in the Intestinal Tract. *Infection and Immunity* 39**,** 686-703.

Fukushima, M., Kakinuma, K., and Kawaguchi, R. (2002). Phylogenetic Analysis of *Salmonella, Shigella*, and *Escherichia* coli Strains on the Basis of the gyrB Gene Sequence. *Journal of Clinical Microbiology* 40**,** 2779-2785.

Fukuta, S., Magnani, J.L., Twiddy, E.M., Holmes, R.K., and Ginsburg, V. (1988). Comparison of the carbohydrate-binding specificities of cholera toxin and *Escherichia coli* heat-labile enterotoxins LTh-I, LT-IIa, and LT-IIb. *Infection and Immunity* 56**,** 1748-1753.

Gaastra, W., and Svennerholm, A.M. (1996). Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends in Microbiology* 4**,** 444-452.

Gadgil, M., Kapur, V., and Hu, W.S. (2005). Transcriptional response of *Escherichia coli* to temperature shift. *Biotechnology Progress* 21**,** 689-699.

Goodman, A.L., Mcnulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host and Microbe* 6**,** 279-289.

Gordon, D.M. (2001). Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* 147**,** 1079-1085.

Gordon, D.M. (2004). The Influence of Ecological Factors on the Distribution and the Genetic Structure of *Escherichia coli*. *EcoSal Plus* 1.

Gordon, D.M., and Cowling, A. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149**,** 3575-3586.

Guth, B.E., Twiddy, E.M., Trabulsi, L.R., and Holmes, R.K. (1986). Variation in chemical properties and antigenic determinants among type II heat-labile enterotoxins of *Escherichia coli*. *Infection and Immunity* 54**,** 529-536.

Heiman, K.E., Mody, R.K., Johnson, S.D., Griffin, P.M., and Gould, L.H. (2015). *Escherichia coli* O157 Outbreaks in the United States, 2003-2012. *Emerging Infectious Diseases* 21**,** 1293-1301.

Hirayama, T., and Wada, A. (2000). "Heat-Stable Enterotoxin of *Escherichia coli*," in Bacterial Protein Toxins*,* eds. K. Aktories & I. Just. (Berlin, Heidelberg: Springer Berlin Heidelberg), 577-593.

Huggins, C., and Rast, H.V., Jr. (1963). Incidence of coliform bacteria in the intestinal tract of Gambusia affinis holbrooki (Girard) and in their habitat water. *Journal of bacteriology* 85**,** 489-490.

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., Kenny, B., Quail, M.A., Thurston, S., Dougan, G., Hayashi, T., Parkhill, J., and Frankel, G. (2009). Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *Journal Bacteriology* 191**,** 347-354.

Ihssen, J., and Egli, T. (2005). Global physiological analysis of carbon- and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environmental Microbiology* 7**,** 1568-1581.

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

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

Ishii, S., and Sadowsky, M.J. (2008). *Escherichia coli* in the Environment: Implications for Water Quality and Human Health. *Microbes and Environments* 23**,** 101-108.

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

Iuchi, S., and Lin, E.C. (1993). Adaptation of *Escherichia coli* to redox environments by gene expression. *Molecular Microbiology* 9**,** 9-15.

Jang, J., Di, D.Y.W., Lee, A., Unno, T., Sadowsky, M.J., and Hur, H.-G. (2014). Seasonal and Genotypic Changes in *Escherichia coli* Phylogenetic Groups in the Yeongsan River Basin of South Korea. *PLOS ONE* 9**,** e100585.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M.N., Yan, T., and Ishii, S. (2017). Environmental *Escherichia coli*: ecology and public health implications-a review. *Journal of Applied Microbiology* 123**,** 570-581.

Jerse, A.E., Yu, J., Tall, B.D., and Kaper, J.B. (1990). A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proceedings of the National Academy of Sciences of the United States of America* 87**,** 7839-7843.

Johnson, J.R., and Russo, T.A. (2002). Extraintestinal pathogenic *Escherichia coli*: "the other bad E coli". *Journal of Laboratory and Clinical Medicine* 139**,** 155-162.

Johnson, J.R., Tchesnokova, V., Johnston, B., Clabots, C., Roberts, P.L., Billig, M., Riddell, K., Rogers, P., Qin, X., Butler-Wu, S., Price, L.B., Aziz, M., Nicolas-Chanoine, M.H., Debroy, C., Robicsek, A., Hansen, G., Urban, C., Platell, J., Trott, D.J., Zhanel, G., Weissman, S.J., Cookson, B.T., Fang, F.C., Limaye, A.P., Scholes, D., Chattopadhyay, S., Hooper, D.C., and Sokurenko, E.V. (2013). Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *Journal of Infectious Disease* 207**,** 919-928.

Jones, S.A., Chowdhury, F.Z., Fabich, A.J., Anderson, A., Schreiner, D.M., House, A.L., Autieri, S.M., Leatham, M.P., Lins, J.J., Jorgensen, M., Cohen, P.S., and Conway, T. (2007). Respiration of *Escherichia coli* in the mouse intestine. *Infection and Immunity* 75**,** 4891-4899.

Jones, S.A., Gibson, T., Maltby, R.C., Chowdhury, F.Z., Stewart, V., Cohen, P.S., and Conway, T. (2011). Anaerobic Respiration of *Escherichia coli* in the Mouse Intestine. *Infection and Immunity* 79**,** 4218-4226.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic *Escherichia coli*. *Nature Review Microbiology* 2**,** 123-140.

Karch, H., Tarr, P.I., and Blelaszewska, M. (2005). Enterohaemorrhagic *Escherichia coli* in human medicine. *International Journal of Medical Microbiology* 295**,** 405-418.

Karmali, M.A., Gannon, V., and Sargeant, J.M. (2010). Verocytotoxin-producing *Escherichia coli* (VTEC). *Veterinary Microbiology* 140**,** 360-370.

Karmali, M.A., Steele, B.T., Petric, M., and Lim, C. (1983). Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet (London, England)* 1**,** 619-620.

Khot, P.D., and Fisher, M.A. (2013). Novel Approach for Differentiating *Shigella* Species and *Escherichia coli* by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. *Journal of Clinical Microbiology* 51**,** 3711-3716.

Khursigara, C., Abul-Milh, M., Lau, B., Girón, J.A., Lingwood, C.A., and Foster, D.E.B. (2001). Enteropathogenic *Escherichia coli* Virulence Factor Bundle-Forming Pilus Has a Binding Specificity for Phosphatidylethanolamine. *Infection and Immunity* 69**,** 6573-6579.

Knutton, S., Baldini, M.M., Kaper, J.B., and Mcneish, A.S. (1987). Role of plasmid-encoded adherence factors in adhesion of enteropathogenic *Escherichia coli* to HEp-2 cells. *Infection and Immunity* 55**,** 78-85.

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

Korajkic, A., Mcminn, B.R., Shanks, O.C., Sivaganesan, M., Fout, G.S., and Ashbolt, N.J. (2014). Biotic Interactions and Sunlight Affect Persistence of Fecal Indicator Bacteria and Microbial Source Tracking Genetic Markers in the Upper Mississippi River. *Applied and Environmental Microbiology* 80**,** 3952-3961.

Kudva, I.T., Blanch, K., and Hovde, C.J. (1998). Analysis of *Escherichia coli* O157:H7 Survival in Ovine or Bovine Manure and Manure Slurry. *Applied and Environmental Microbiology* 64**,** 3166-3174.

Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. *Nature* 392**,** 917-920.

Kyaw, C.M., De Araujo, C.R., Lima, M.R., Gondim, E.G.S., BríGido, M.M., and Giugliano, L.G. (2003). Evidence for the presence of a type III secretion system in diffusely adhering *Escherichia coli* (DAEC). *Infection, Genetics and Evolution* 3**,** 111-117.

Lawley, T.D., and Walker, A.W. (2013). Intestinal colonization resistance. *Immunology* 138**,** 1-11.

Leatham-Jensen, M.P., Frimodt-Moller, J., Adediran, J., Mokszycki, M.E., Banner, M.E., Caughron, J.E., Krogfelt, K.A., Conway, T., and Cohen, P.S. (2012). The streptomycin-treated mouse intestine selects *Escherichia coli* envZ missense mutants that interact with dense and diverse intestinal microbiota. *Infection and Immunity* 80**,** 1716-1727.

Leclerc, H., Mossel, D.A., Edberg, S.C., and Struijk, C.B. (2001). Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annual Review Microbiology* 55**,** 201-234.

Levine, M.M. (1987). *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *The Journal of infectious diseases* 155**,** 377-389.

Lewis, K. (2010). Persister cells. *Annual Review Microbiology* 64**,** 357-372.

Lin, J., Lee, I.S., Frey, J., Slonczewski, J.L., and Foster, J.W. (1995). Comparative analysis of extreme acid survival in *Salmonella* Typhimurium, *Shigella flexneri*, and *Escherichia coli*. *Journal of Bacteriology* 177**,** 4097-4104.

Lin, J., Smith, M.P., Chapin, K.C., Baik, H.S., Bennett, G.N., and Foster, J.W. (1996). Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. *Applied and Environmental Microbiology* 62**,** 3094-3100.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'kelly, M.J.T., Van Oudenaarden, A., Barton, D.B.H., Bailes, E., Nguyen, A.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., and Louis, E.J. (2009). Population genomics of domestic and wild yeasts. *Nature* 458**,** 337.

Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J., and Blattner, F.R. (2005). Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *Journal of Biological Chemistry* 280**,** 15921-15927.

Lucht, J.M., and Bremer, E. (1994). Adaptation of *Escherichia coli* to high osmolarity environments: osmoregulation of the high-affinity glycine betaine transport system proU. *FEMS Microbiology Reviews* 14**,** 3-20.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011). Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences USA* 108.

Maclean, D., Jones, J.D., and Studholme, D.J. (2009). Application of 'next-generation' sequencing technologies to microbial genetics. *Nature Reviews Microbiology* 7**,** 287-296.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and Conway, T. (2013). Nutritional Basis for Colonization Resistance by Human Commensal *Escherichia coli* Strains HS and Nissle 1917 against *E. coli* O157:H7 in the Mouse Intestine. *PLoS ONE* 8**,** e53957.

Mandel, M.J., and Silhavy, T.J. (2005). Starvation for Different Nutrients in *Escherichia coli* Results in Differential Modulation of RpoS Levels and Stability. *Journal of Bacteriology* 187**,** 434-442.

Mccord, J.M., and Fridovich, I. (1969). Superoxide Dismutase: An Enzymic Function for Erythrocuprein (hemocuprein). *Journal of Biological Chemistry* 244**,** 6049-6055.

Mcdougald, D., Rice, S.A., Barraud, N., Steinberg, P.D., and Kjelleberg, S. (2011). Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nature Reviews Microbiology* 10**,** 39.

Meric, G., Kemsley, E.K., Falush, D., Saggers, E.J., and Lucchini, S. (2013). Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environmental Microbiology* 15**,** 487-501.

Munro, P.M., Gauthier, M.J., Breittmayer, V.A., and Bongiovanni, J. (1989). Influence of osmoregulation processes on starvation survival of *Escherichia coli* in seawater. *Applied and Environmental Microbiology* 55**,** 2017-2024.

Nataro, J.P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S.J., Guers, L.D., Levine, M.M., and Tacket, C.O. (1995). Heterogeneity of enteroaggregative *Escherichia coli* virulence demonstrated in volunteers. *Journal of Infectious Diseases* 171**,** 465-468.

Nataro, J.P., Deng, Y., Maneval, D.R., German, A.L., Martin, W.C., and Levine, M.M. (1992). Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. *Infection and Immunity* 60**,** 2297-2304.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews* 11**,** 142-201.

Nataro, J.P., Yikang, D., Yingkang, D., and Walker, K. (1994). AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. *Journal of Bacteriology* 176**,** 4691-4699.

Ochman, H., and Wilson, A.C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26**,** 74-86.

Odonkor, S.T., and Ampofo, J.K. (2013). *Escherichia* coli as an indicator of bacteriological quality of water: an overview. *Microbiology research* 4**,** 2.

Olive, D.M., and Bean, P. (1999). Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. *Journal of Clinical Microbiology* 37**,** 1661-1669.

Palmer, C., Bik, E.M., Digiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. *Plos Biology* 5**,** 1556-1573.

Payment, P., Waite, M., and Dufour, A. (2003). Introducing parameters for the assessment of drinking water quality. *Assessing microbial safety of drinking water* 47.

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., Van Den Brandt, P.A., and Stobberingh, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118**,** 511-521.

Penders, J., Vink, C., Driessen, C., London, N., Thijs, C., and Stobberingh, E.E. (2005). Quantification of *Bifidobacterium* spp., *Escherichia coli* and *Clostridium difficile* in faecal samples of breast-fed and formula-fed infants by real-time PCR. *Fems Microbiology Letters* 243**,** 141-147.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y.-J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H., and Kobayashi, K.S. (2009). Nod2 is required for the regulation of commensal microbiota in the intestine. *Proceedings of the National Academy of Sciences of the United States of America* 106**,** 15813-15818.

Pickett, C.L., Twiddy, E.M., Coker, C., and Holmes, R.K. (1989). Cloning, nucleotide sequence, and hybridization studies of the type IIb heat-labile enterotoxin gene of *Escherichia coli*. *Journal of Bacteriology* 171**,** 4945-4952.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005). Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environmental Microbiology* 7**,** 631-640.

Pupo, G.M., Lan, R., and Reeves, P.R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proceedings of the National Academy of Sciences* 97**,** 10567-10572.

Qadri, F., Svennerholm, A.-M., Faruque, A.S.G., and Sack, R.B. (2005). Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clinical Microbiology Reviews* 18**,** 465-483.

Reitzer, L. (2003). Nitrogen Assimilation and Global Regulation in *Escherichia coli*. *Annual Reviews of Microbiology* 57**,** 155-176.

Rice, E.W., Allen, M.J., and Edberg, S.C. (1990). Efficacy of beta-glucuronidase assay for identification of *Escherichia coli* by the defined-substrate technology. *Applied and Environmental Microbiology* 56**,** 1203-1205.

Rozen, Y., and Belkin, S. (2001). Survival of enteric bacteria in seawater. *FEMS Microbiology Reviews* 25**,** 513-529.

Russo, T.A., and Johnson, J.R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *Journal of Infectious Diseases* 181**,** 1753-1754.

Salyers, A.A., and Pajeau, M. (1989). Competitiveness of different polysaccharide utilization mutants of *Bacteroides thetaiotaomicron* in the intestinal tracts of germfree mice. *Applied and Environmental Microbiology* 55**,** 2572-2578.

Sander, A., Ruess, M., Bereswill, S., Schuppler, M., and Steinbrueckner, B. (1998). Comparison of Different DNA Fingerprinting Techniques for Molecular Typing of *Bartonella* henselae Isolates. *Journal of Clinical Microbiology* 36**,** 2973-2981.

Sansonetti, P.J. (1992). "Molecular and Cellular Biology of Shigella flexneri Invasiveness: From Cell Assay Systems to Shigellosis," in *Pathogenesis of Shigellosis,* ed. P.J. Sansonetti. (Berlin, Heidelberg: Springer Berlin Heidelberg), 1-19.

Sansonetti, P.J. (1998). Pathogenesis of shigellosis: from molecular and cellular biology of epithelial cell invasion to tissue inflammation and vaccine development. *Japanese Journal of Medical Science and Biology* 51 Suppl**,** S69-80.

Savageau, M.A. (1983). *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *The American Naturalist* 122**,** 732-744.

Savarino, S.J., Fasano, A., Watson, J., Martin, B.M., Levine, M.M., Guandalini, S., and Guerry, P. (1993). Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1 represents another subfamily of *E. coli* heat-stable toxin. *Proceedings of the National Academy of Sciences U S A* 90**,** 3093-3097.

Scaletsky, I.C., Fabbricotti, S.H., Aranda, K.R., Morais, M.B., and Fagundes-Neto, U. (2002). Comparison of DNA hybridization and PCR assays for detection of putative pathogenic enteroadherent *Escherichia coli*. *Journal of Clinical Microbiology* 40**,** 1254-1258.

Sears, C.L., and Kaper, J.B. (1996). Enteric bacterial toxins: Mechanisms of action and linkage to intestinal secretion. *Microbiological Reviews* 60**,** 167-+.

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., and Whittam, T.S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology* 51**,** 873-884.

Semenov, A.V., Van Bruggen, A.H.C., Van Overbeek, L., Termorshuizen, A.J., and Semenov, A.M. (2007). Influence of temperature fluctuations on *Escherichia coli* O157:H7 and *Salmonella enterica* serovar Typhimurium in cow manure. *FEMS Microbiology Ecology* 60**,** 419-428.

Servin, A.L. (2005). Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*. *Clinical Microbiology Reviews* 18**,** 264-292.

Sheikh, J., Czeczulin, J.R., Harrington, S., Hicks, S., Henderson, I.R., Le Bouguenec, C., Gounon, P., Phillips, A., and Nataro, J.P. (2002). A novel dispersin protein in enteroaggregative *Escherichia coli*. *Journal of Clinical Investigation* 110**,** 1329-1337.

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

Sixma, T.K., Kalk, K.H., Vanzanten, B.a.M., Dauter, Z., Kingma, J., Witholt, B., and Hol, W.G.J. (1993). Refined Structure of *Escherichia coli* Heat-Labile Enterotoxin, A Close Relative of Cholera-Toxin. *Journal of Molecular Biology* 230**,** 890-918.

Slanetz, L.W., and Bartley, C.H. (1957). Numbers of *Enterococci* in Water, Sewage, and Feces Determined by the Membrane Filter Technique with an Improved Medium. *Journal of Bacteriology* 74**,** 591-595.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000). Sources of *Escherichia coli* in a coastal subtropical environment. *Applied and Environmental Microbiology* 66**,** 230-237.

Stecher, B., and Hardt, W.-D. (2011). Mechanisms controlling pathogen colonization of the gut. *Current Opinion in Microbiology* 14**,** 82-91.

Stern, M.J., Ames, G.F., Smith, N.H., Robinson, E.C., and Higgins, C.F. (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37**,** 1015-1026.

Stewart, J.R., Santo Domingo, J.W., and Wade, T.J. (2007). "Fecal Pollution, Public Health, and Microbial Source Tracking," in *Microbial Source Tracking*. American Society of Microbiology).

Stincone, A., Daudi, N., Rahman, A.S., Antczak, P., Henderson, I., Cole, J., Johnson, M.D., Lund, P., and Falciani, F. (2011). A systems biology approach sheds new light on *Escherichia coli* acid resistance. *Nucleic Acids Research*.

Stock, I., Burak, S., and Wiedemann, B. (2004). Natural antimicrobial susceptibility patterns and biochemical profiles of Leclercia adecarboxylata strains. *Clinical Microbiology and Infection* 10**,** 724-733.

Storz, G., and Imlayt, J.A. (1999). Oxidative stress. *Current opinion in microbiology* 2**,** 188-194.

Taddei, C.R., Moreno, A.C.R., Fernandes Filho, A., Montemor, L.P.G., and Martinez, M.B. (2003). Prevalence of secreted autotransporter toxin gene among diffusely adhering *Escherichia coli* isolated from stools of children. *FEMS Microbiology Letters* 227**,** 249-253.

Tauschek, M., Gorrell, R.J., Strugnell, R.A., and Robins-Browne, R.M. (2002). Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proceedings of the National Academy of Sciences U S A* 99**,** 7066-7071.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology* 8**,** 207-217.

Tesh, V.L., and O'brien, A.D. (1991). The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins. *Molecular Microbiology* 5**,** 1817-1822.

Tilman, D. (1982). Resource competition and community structure. *Monographs in Population Biology* 17**,** 1-296.

Tobe, T., Hayashi, T., Han, C.G., Schoolnik, G.K., Ohtsubo, E., and Sasakawa, C. (1999). Complete DNA sequence and structural analysis of the enteropathogenic Escherichia coli adherence factor plasmid. *Infection and Immunity* 67**,** 5455-5462.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., and Bouvet, O. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5.

Tujula, N.A., Crocetti, G.R., Burke, C., Thomas, T., Holmstrom, C., and Kjelleberg, S. (2010). Variability and abundance of the epiphytic bacterial community associated with a green marine Ulvacean alga. *ISME J* 4**,** 301-311.

Unc, A., Gardner, J., and Springthorpe, S. (2006). Recovery of *Escherichia coli* from Soil after Addition of Sterile Organic Wastes. *Applied and Environmental Microbiology* 72**,** 2287-2289.

Van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5**,** 173-183.

Vanden Heuvel, A., Mcdermott, C., Pillsbury, R., Sandrin, T., Kinzelman, J., Ferguson, J., Sadowsky, M., Byappanahalli, M., Whitman, R., and Kleinheinz, G.T. (2010). The green alga, Cladophora, promotes *Escherichia coli* growth and contamination of recreational waters in Lake Michigan. *Journal of Environmental Quality* 39**,** 333-344.

Versalovic, J., Koeuth, T., and Lupski, J.R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Research* 19**,** 6823-6831.

Vital, M., Chai, B., Ostman, B., Cole, J., Konstantinidis, K.T., and Tiedje, J.M. (2015). Gene expression analysis of E. coli strains provides insights into the role of gene regulation in diversification. *Isme journal* 9**,** 1130-1140.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Et Al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23**,** 4407-4414.

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

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., and Whittam, T.S. (2009). Cryptic lineages of the genus *Escherichia*. *Applied and Environmental Microbiology* 75**,** 6534-6544.

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

White-Ziegler, C.A., and Davis, T.R. (2009). Genome-Wide Identification of H-NS-Controlled, Temperature-Regulated Genes in *Escherichia coli* K-12. *Journal of Bacteriology* 191**,** 1106-1110.

Whitman, R.L., and Nevers, M.B. (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., Nevers, M.B., Korinek, G.C., and Byappanahalli, M.N. (2004). Solar and temporal effects on *Escherichia coli* concentration at a Lake Michigan swimming beach. *Applied and Environmental Microbiology* 70**,** 4276-4285.

Whitman, R.L., Shively, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N. (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.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* 95**,** 6578-6583.

Whittam, T.S. (1989). Clonal dynamics of *Escherichia coli* in its natural habitat. *Antonie van Leeuwenhoek Journal of Microbiology* 55**,** 23-32.

Williams, A.M., Probert, C.S.J., Stepankova, R., Tlaskalova-Hogenova, H., Phillips, A., and Bland, P.W. (2006). Effects of microflora on the neonatal development of gut mucosal T cells and myeloid cells in the mouse. *Immunology* 119**,** 470-478.

Chapter 2: Distribution of Diverse *Escherichia coli* between Cattle and Pasture

The contents of this chapter has appeared as Nandakafle, G., Seale, T., Flint, T., Nepal, M., Venter, S.N., and Brozel, V.S. (2017). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. *Microbes Environ*. 32, 226-233.

1. Introduction

Escherichia coli occur as part of the intestinal microbiota of many warm blooded animals, their primary habitat, and were long thought to survive for only short periods outside the host (Savageau, 1983) The use of *E. coli* as an indicator organism is based on the assumption that it does not persist and grow in secondary environments such as soil, water and sediments, thus indicating the presence of recent fecal contamination (Lang et al., 2003;Osborn and Trussell, 2004). Recent studies have demonstrated that *E. coli* is able to maintain populations in aquatic and soil environments (Winfield and Groisman, 2003;Ishii and Sadowsky, 2008a). The occurrence of *E. coli* in soil, sediments, and water in tropical and sub-tropical regions has been widely documented, and the species is now considered to be autochthonous to soil within such warm regions (Byappanahalli and Fujioka, 1998;Solo-Gabriele et al., 2000;Desmarais et al., 2002;Byappanahalli and Fujioka, 2004;Hartz et al., 2008;Byappanahalli et al., 2012). *E. coli* can also survive for long periods and potentially replicate in temperate environments. Strains have been repeatedly isolated from undisturbed riparian soils of Southern Lake Michigan, Indiana (Byappanahalli et al., 2006). Similarly Ishii *et al*. (Ishii et al., 2006) also reported the

isolation of naturalized *E. coli* strains from temperate soil of Lake Superior Watershed, Minnesota. *E. coli* were also found attached to the macro-alga *Cladophora* in Lake Michigan (Byappanahalli et al., 2003), to periphyton in Lake Superior (Ksoll et al., 2007), and in beach sand and sediments (Beversdorf et al., 2007;Ishii et al., 2007). Persistent strains have also been reported from alpine pasture soils, whether sampled from under or away from cowpats (Texier et al., 2008).

The genus *Escherichia* comprises *E. coli, E. hermanii, E. vulneris, E. fergusonii* and *E. albertii*, with *E. marmotae* recently being added (Liu et al., 2015). Only *E. coli* contains a functional ß-glucuronidase encoded by *uidA* (Hayes et al., 1995), allowing distinction on differential media such as Membrane Lactose Glucuronide Agar (MLGA). Although enterohemorrhagic *E. coli* O157:H7 lack functional ß-glucuronidase, positive variants have been reported (Hayes et al., 1995;Sanchez et al., 2010). A collection of environmental *Escherichia* isolates initially considered to be *E. coli* have been assigned to four distinct genetic clusters based on their unique Multi-Locus Sequence Type profiles, and named clades I, III, IV, and V(Walk et al., 2009). According to the extent of recombination between isolates of *E. coli*, Clade I is viewed as part of *E. coli sensu stricto* (Luo et al., 2011b), whereas clades III, IV and V are phylogenetically distant and not part of the species. Environmental *E. coli* may be defined as resident in a primary habitat in an environment outside of a host. Clade I is most closely related to *E. coli* with evidence for genetic exchange between members of the two. The clade strains are phenotypically similar to *E. coli*, and generally positive for ß-glucuronidase; however clade III cannot ferment sorbitol and sucrose or utilize lysine (Walk et al., 2009). They

may be distinguished from *E. coli sensu stricto* by multi locus sequence typing (MLST) and phylogrouping.

There is increasing evidence to show that *E. coli* strains occurring in secondary environments are genetically and phenotypically distinct from *E. coli* inhabiting the gut (Gordon et al., 2002). The allocation of *E. coli* to four Phylogroups by multiplex PCR (Clermont et al., 2000) has indicated a degree of niche partitioning. Phylogroups A and B1 appear dominated by environmentally occurring strains (Walk et al., 2007), whereas B2 and D are predominated by mammalian isolates (Le Gall et al., 2007;Diard et al., 2010). These findings suggest niche partitioning of diverse *E. coli* strains across various environments, and thus, some strains may be autochthonous to soils. This phylogrouping protocol was recently refined to yield seven groups (Clermont et al., 2013). Isolates from temperate soil of Lake Superior Watershed displayed DNA fingerprints distinct from animal-derived isolates (Ishii et al., 2006). *E. coli* from fresh water beaches along Lake Huron and the St. Clair River in Michigan revealed extensive genetic diversity of MLST (Walk et al., 2007). The *uidA* sequences of *E. coli* from alpine pasture soil were distinct from fecal *E. coli*, indicating a naturalized population that was part of the indigenous soil community (Texier et al., 2008).

Previous studies have established occurrence of *E. coli* in soil, water and sediments under various climatic conditions. The objective of the present study was to investigate whether niche partitioning of *E. coli* occurs between cattle and their pasture. We attempted to clarify whether *E. coli* from bovine feces differed phenotypically and genotypically from isolates maintaining a population in pasture soil over winter. *E. coli* strains that survived in pasture soil through the extreme South Dakota winter displayed a different genotype

compared to bovine fecal isolates and need to be considered as environmental or naturalized.

2. Materials and methods

2.1. Sample collection

Samples were collected from a cattle pasture (12.14 Ha) divided into four separate encampments (GPS co-ordinates 44°22'17.70"N 96°58'1.54"W), at Volga, SD, USA in May, June and July, 2013. This pasture had been cleared of cattle at the end of July 2012. Before the reintroduction of five or more cows per encampment, soil cores (4 cm in depth) were collected over three weeks, between May and June, and designated Soil Before Grazing (SBG). Following the introduction of cattle at the start of July, soil cores, run-off and cattle feces samples were collected once per week for four weeks and from each of the four encampments. Every week a transect was drawn at random across each encampment, and five soil samples, five run-off samples and two fresh fecal samples were collected per transect. Soil samples were taken to a depth of 4 cm using a soil borer (2 cm in diameter), and soil cores transferred to sterile 50-mL conical screw-cap tubes. A simulation of run-off was performed using a Cornell infiltrometer (Ogden et al., 1997) fed with sterile dH$_2$O. Runoff was collected into sterile 100-mL screw-cap flasks. Fresh fecal samples were taken directly from the pasture by scooping into sterile 50-mL conical screw-cap tubes. Samples were transported to the laboratory in a cooled container and processed on the same day. Run-off and soil samples collected at the time of grazing were designated as pasture samples.

2.2. Isolation of E. coli from soil, run-off and feces sample

Twenty grams of soil were mixed vigorously with 100 mL sterile water and allowed to settle for 1 h. One- and ten-ml aliquots were then filtered through a sterile 0.45 μm mixed cellulose ester filter (Millipore), which was placed on Membrane Lactose Glucuronide Agar (MLGA, Fluka Analytical). Run-off samples (1 and 10 ml) were filtered directly through the 0.45 μm nitrocellulose filters which were then placed on MLGA. Aliquots of ten-fold dilutions of bovine fecal samples were plated on MLGA. Green colonies were positive for β-Galactosidase and β-Glucuronidase activities, and were assumed to be *E. coli*. This protocol excluded ß-glucuronidase negative O157:H7 strains. An average of 2 colonies were picked from the highest dilutions showing growth, streaked onto MLGA to confirm purity, sub-cultured onto LB agar and stored at -80 °C in 50% glycerol. All 15 cryptic species isolates obtained from Dr. Seth Walk grew on MLGA and 8 formed green colonies; therefore, it is unlikely that our isolation method excluded potential members of clades I, III, IV or V (data not shown).

2.3. Analysis of the uidA and mutS gene sequences

Genomic DNA was extracted from overnight LB agar cultures suspended in 10 mM phosphate buffer (pH 7.0) using the genomic DNA Quick Prep Kit (Zymo Research), and stored at -20°C. The *uidA* and *mutS* genes were amplified by PCR using primers described previously (Walk et al., 2009) (Table S1), with *E. coli* MG1655 as positive control. PCR reactions (25 μl) were set up as follows: 2.5 μl reaction buffer (10X) (New England Biolab), 0.5 μl MgCl$_2$ (25mM), 0.5 μl dNTPs (40mM), 0.1 μl forward primer and 0.1 μl reverse primer (100 μmol), 0.125 μl of Taq polymerase (NE Biolab), 0.5 μl of

a DNA template and 20.7 µl sterile nano pure water. The amplification cycle was initiated with 95°C for 2 min, followed by 30 cycles of denaturing at 95°C for 30 s, annealing at 56°C for 30 s and extension at 72°C for 1 min, with a final extension at 72°C for 5 min. DNA sequences were elucidated by the dideoxy chain termination method (Beckman Coulter Genomic Center at Denver, MA). *uidA* sequences were submitted to Genbank (http://www.ncbi.nlm.nih.gov/genbank/) under the BankIt number 1841773 (accession numbers KT311394 – KT311756), and *mutS* sequences as the BankIt number 1841687 (accession numbers KT311004 – KT311366). The DNA sequences were aligned using ClustalW (Tamura et al., 2013), and overhangs were trimmed using SeAl (Rambaut, 2002). The *uidA* and *mutS* sequences for all isolates and reference strains (Leimbach et al., 2013) were concatenated using SeAl. A maximum likelihood analysis using model GTR+G+I with 1,000 bootstrap replicates was performed in the program MEGA6.06 (Tamura et al., 2013). The tree was then annotated and visualized using the ITOL online tool (Letunic and Bork, 2011).

2.4. Identification of Phylogroups

Isolates were assigned to phylogroups using the protocol described by Clermont et al. (2013). In order to avoid ambiguity, PCR was performed separately for each primer set (Table S1). In order to clarify whether the distribution of phylogroups differed by source or cluster, we used multinomial log-linear regression models. Multiple logistic regression is used when the dependent variable is nominal and there is more than one independent variable. It is a classification method that generalizes logistic regression to multiclass issues having more than two possible discrete outcomes (McDonald, 2009). The models

were fit using the nnet package in R (v.3.2.2)(R Development Core Team, 2015). The response variable in this analysis was the phylogroup of each isolate (A, B1, B2, C, D, E, and Unknown), and the explanatory variables were the sample source and clusters associated with origin of the isolates. In order to visualize the effect of significant explanatory variables, we used regression trees fit using Package Party (Hothorn, 2014) in R.

2.5. *Identification of the RDAR (red, dry and rough) morphotype*

The RDAR morphotype was determined as described by White et al. (2010). Briefly *E. coli* isolates were grown at 37°C overnight on LBns agar (LB without salt), followed by culturing overnight at 37°C in LBns broth with shaking. Spot colonies were prepared by inoculation of 1μl of the overnight broth culture onto LBns agar supplemented with 100 μg.mL^{-1} Congo red. Colonies were observed under a stereo microscope (Olympus SZX16), after an incubation at 28°C for 72 h. Colony morphologies were assigned to four groups where white smooth colonies were "ws", red smooth colonies were "rs", slight rough colonies were "sc" and highly wrinkled (curli) colonies were "c".

2.6. *Degree of biofilm formation in LB and SESOM media*

Regarding the biofilm quantification in LB media, 5 μL of overnight broth culture was mixed with 195 μL of LB broth in a 96-well plate and incubated for 16 h at 37°C. *Pseudomonas aeruginosa* (PAO) (obtained from Dr Sang-Jin Suh, Auburn University, Alabama, USA) was used as a positive control. Soluble Extractable Soil Organic Matter

(SESOM) was prepared for culturing in liquid soil extract using air-dried soil from the cattle pasture as described previously (Vilain et al., 2006). Five microliters of overnight broth culture was mixed with 195 µL of SESOM, harvested by centrifugation and the pellet re-suspended in 200 µL of SESOM. The staining and quantification of biofilm were performed using the Crystal Violet (CV) assay as described by O'Toole (2011), but with 95% ethanol instead of 30% acetic acid for solubilization of CV. Each treatment was repeated four times. Isolates were assigned to four groups based on the quantity of biofilm formed (absorbance at 560 nm). In SESOM group" 0" was < 0.01, group "1" was <0.025, group "2" was < 0.05 and group "3" was > 0.05. In LB group "0" was <0.025, group "1" was <0.05, group "2" was <0.1 and group "3" was >0.1.

2.7. Long term survival

In order to study the long-term survival of *E. coli* in soil under environmental conditions, 45 isolates were selected to represent the three sample sources and three cluster types (see results). A loop full of culture was then washed two times with sterile water and the pellet suspended in 500 µl SESOM. Two hundred microliters of this cell suspension was inoculated into 20 mL of SESOM and incubated at 25ºC overnight. The overnight liquid culture was diluted to an optical density of 0.05 at 546 nm (10^6 CFU.mL^{-1}) and 1mL of this diluted cell suspension was used to inoculate 5 g of double autoclaved pasture soil placed in a sterile 50-mL conical tube. A further 2 mL of sterile dH$_2$O was added to moisten the soil, which was shaken vigorously for 1 min, and the culturable count measured for time zero. Each soil microcosm was set up three times. Tubes were placed on the soil surface outdoors from November 2014 till May 2015. Following an incubation

under environmental conditions for six months, each sample was supplemented with 10 mL of sterile dH_2O, and serial dilutions were prepared to measure the culturable count. The $\log_{10}$ decline was calculated for each isolate. An Analysis of Variance (ANOVA) was performed to examine the significance of differences in log decline among groups and sample types using R (v.3.2.2) (R Development Core Team, 2015).

3. Results

3.1. Isolation of E. coli

We isolated *E. coli* from pasture soil (120 isolates), run-off (163 isolates), and fresh bovine feces (35 isolates) while cattle were grazing. *E. coli* was also isolated from the same pasture soil before cattle were introduced for the summer, with 45 SBG (soil before grazing) isolates. SBG samples contained between 0 and 25 $CFU.g^{-1}$ of soil, while most pasture samples (soil after grazing) tested positive for *E. coli*, with up to 100 $CFU.g^{-1}$ of soil (Fig.1). While the average culturable count in SBG samples was lower than that during grazing, it was not significant according to the Welch two sample t-test ($p = 0.054$). Since cattle were removed during the previous summer and re-introduced only after the SBG samples had been collected, these results indicated that at least some strains maintained populations in soil through the previous fall, winter and spring. It was not possible to control the access of birds or small animals such as rodents to the pasture. Therefore, some *E. coli may have been* deposited by small mammals or birds entering the pasture before or during the sampling period, adding to the diversity obtained.
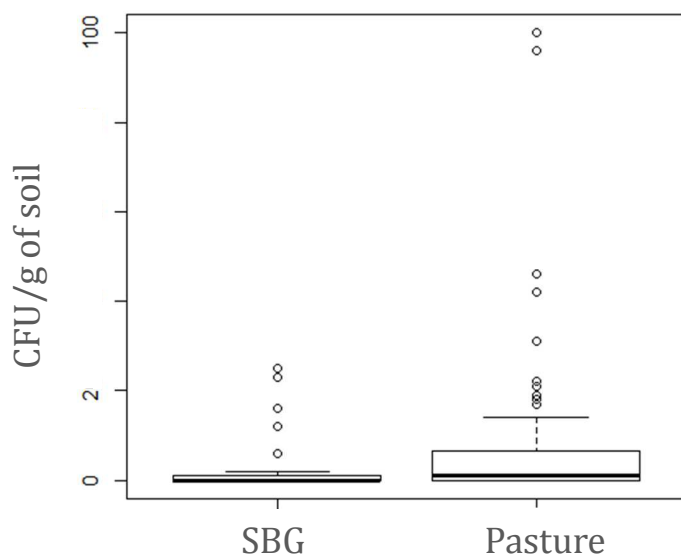
Fig. 2. 1 Box and whisker plot showing the culturable population density of E. coli in soil before grazing (SBG) and in pasture at the time of grazing (p = 0.084).

3.2. Phylogenetic data analysis

A phylogenetic analysis of the concatenated *mutS* and *uidA* sequences of all isolates (363 isolates), 25 human pathogen and commensal reference strains (Leimbach et al., 2013), and representatives of the cryptic species Clades I, III, IV and V (Walk et al., 2009) exhibited multiple distinct clusters (Fig. 2). These clusters were classified into three groups based on the origin of isolates; (i) isolates from all sample types, except bovine feces, i.e. SBG and pasture; designated Environmental, (ii) isolates from all sample types except SBG i.e. feces and pasture; designated Bovine, and (iii) groups containing all three sample types; designated Mixed. Only two such mixed clusters were obtained, and these

mostly contained SBG and pasture isolates, and only three bovine isolates. These three

bovine isolates may be adapted to the soil and gut environments. Five clusters of isolates

fell into the environmental class. Six well-separated bovine clusters were observed, two

clusters with strong bootstrap support (Bov-3 and, Bov-4 with 100). Most of the

reference strains clustered separately from our isolates, indicating *E. coli* different from

current reference strains (Leimbach et al., 2013). None of our isolates clustered with any

of the cryptic species.

3.3. Phylogroup distribution

Isolates were assigned to six out of seven possible phylogroups: A, B1, B2, C, D, and E.

Phylogroup distribution varied across sample types (Fig. 3a). Some isolates were not

allocated to any of these phylogroups, and were termed unknown, whereas none of the

isolates were assigned to group-F or Clade I. Phylogroup distribution across all isolates

showed an overall predominance of B1 and E. SBG samples contained a higher

percentage of B1 (66%) than bovine feces (32%) or pasture samples (40%), indicating

that some B1 strains maintained populations in soil slightly more effectively than their

counterparts. Bovine feceal isolates displayed a higher percentage of E (40%) than SBG

(7%) or pasture sample (28%), suggesting that phylogroup E was primarily bovine

associated and less able to maintain populations in pasture soil. Bovine clusters Bov-3

and Bov-4 were comprised almost exclusively of phylogroup E (Fig. 2). There were

very few phylogroup A isolates in feces and soil samples but the distribution pattern was

similar to B1, indicating that A is somewhat more environmental than bovine associated,

but not particularly competitive in either environment. Bov-5 cluster was comprised of

Sample source
- Pasture
- Bovine Feces
- SBG
- References incl.

Clades
Cluster type
- Environment
- Bovine
- Mixed
- No cluster

Phylogroup
- A
- B1
- B2
- C
- D
- E
- Unknown
- Clades
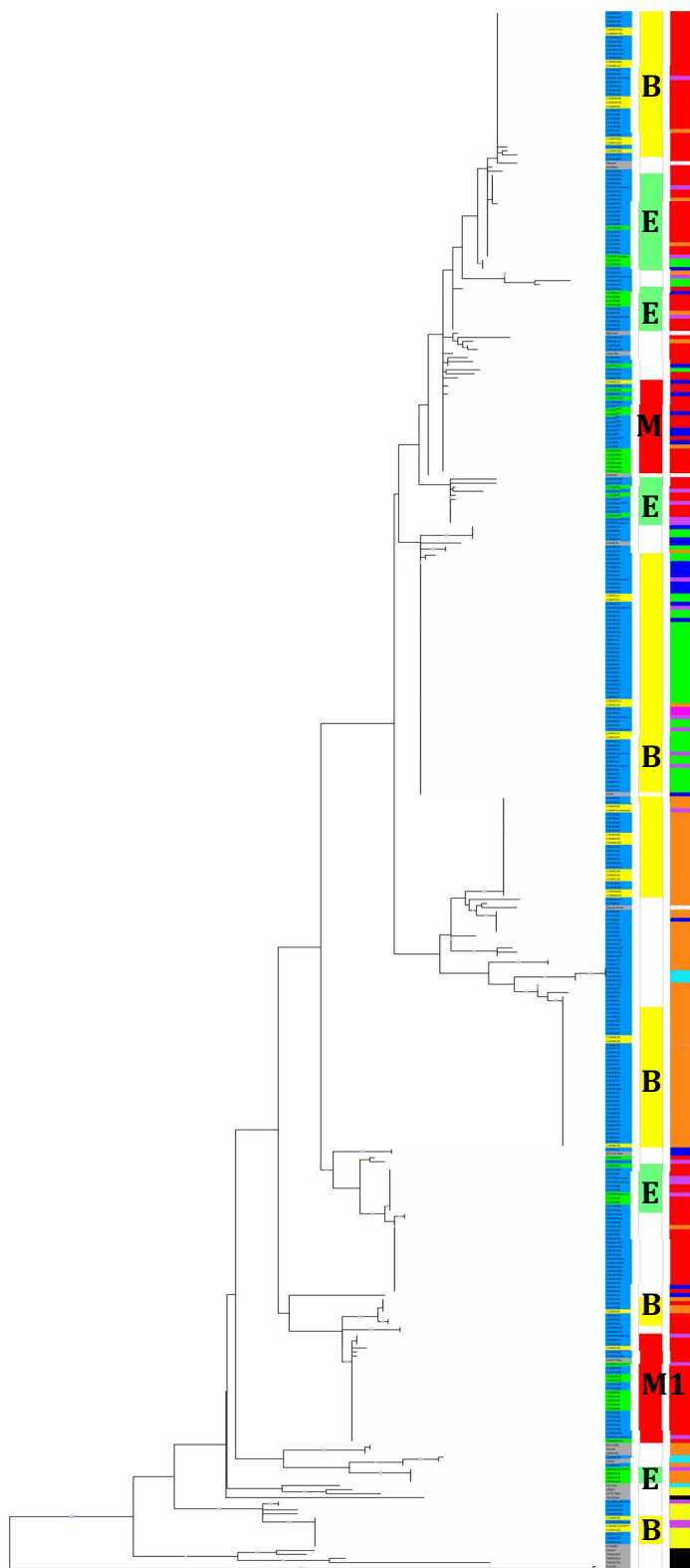
Fig. 2. 2 Phylogenetic analysis of the concatenated *uidA and mutS* gene sequences of *E. coli* isolates, reference strains and cryptic species of E. coli (Walk et al., 2009). Sequences were aligned using ClustalW and manually trimmed using Se-Al. The best Model: Maximum Likelihood analysis with GTR and G+I was performed in the program MEGA 6. Numbers represent branch support of 1000 bootstrap replicates. The phylogenetic tree was color coded and visualized using the Interactive Tree of Life. Isolates are color coded based on their sources (left panel), cluster type (center panel), and phylogroups (right panel). Grey circles on branches indicate a bootstrap value of > 80% (1000 bootstraps).

mostly phylogroup C, while cluster Bov-6 was primarily phylogroup B1. The distribution of the C- phylogroup was similar to E, suggesting that the source of C in soil is mostly from cattle feces. The distribution of unknown isolates was similar to phylogroup B1.

A multinomial log linear regression analysis of phylogroup distribution based on source types (Fig. 3b) suggested similarity between bovine fecal and pasture communities, while fecal and pasture communities were different from SBG ($p = 0.01$), indicating that SBG populations were not from cattle. Cluster-wise comparison of phylogroup distribution (Fig. 3c) revealed that environmental and mixed clusters were similar, indicating that the three bovine isolates in mixed clusters were unique, displaying fitness in the bovine colon and pasture. Environmental and mixed cluster composition differed from bovine cluster populations ($p < 0.001$).

a)

b)

c)
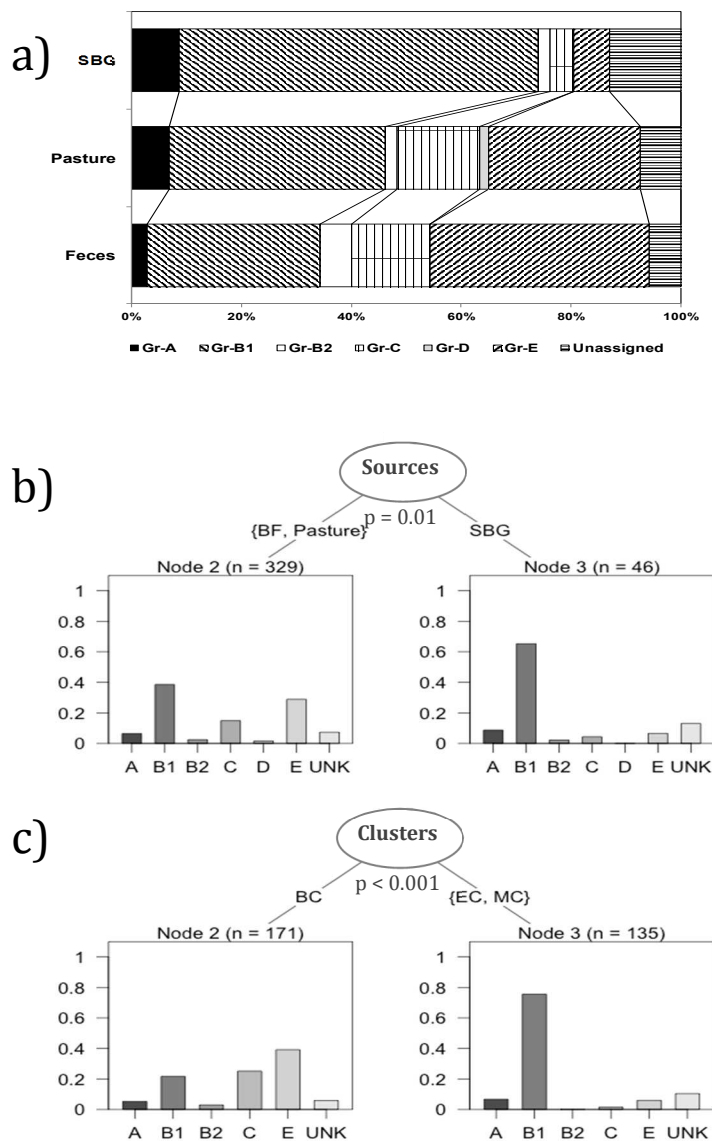
Fig. 2. 3 Phylogroup distribution of isolates across sample and cluster types.

Distribution of phylogroups of isolates across soil before grazing (SBG), pasture soil

while grazing, bovine feces and raw sewage samples (a). Phylogrouping was

performed according to the scheme of Clermont et al., 2013. Regression tree

showing the difference in distribution of phylogenetic groups among sources and

clusters (b). The X axis denotes phylogroups and the Y-axis represents proportion of

isolates. BF-bovine feces, Past-pasture, SBG-soil before grazing, SWG-sewage, BC-bovine cluster, EC-environmental cluster, MC-mixed cluster

3.4. *Curli and biofilm formation*

In order to characterize phenotypes potentially associated with long-term survival in soil, isolates were compared for curli formation. RDAR phenotyping showed that 38% of the isolates were intense curli formers. SBG isolates had the greatest proportion of curli formers, whereas pasture isolates (38%) fell between bovine feces (26%) and SBG (61%) (Fig. 4). A Chi square test of distribution of RDAR by isolate source yielded $p = 1.801$ x 10^{-11}.
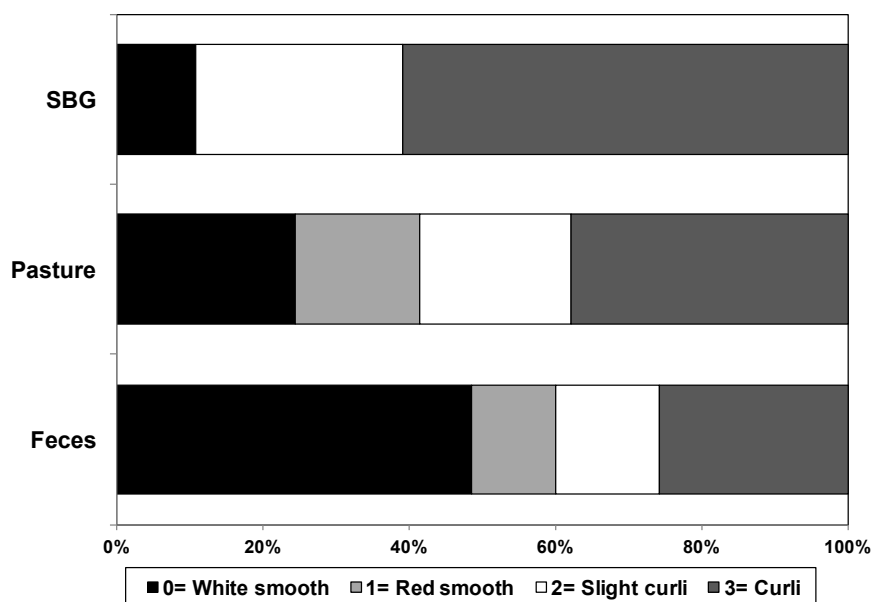
Fig. 2. 4 Distribution of RDAR groups among sample sources.

This indicated that curli formation is associated with soil fitness of some but not all *E. coli*. No clear association was found between curli formation and phylogroup (data not shown), whereas most of phylogroup B1 formed curli.

The degree of biofilm formation varied widely among isolates (Fig. S1), although all isolates formed more sparse biofilm than *P. aeruginosa* PAO (OD= 1.36). Many isolates formed more biofilm in LB media than SESOM; however, the reverse was observed for a small number, indicating variations in biofilm formation across the strains. Most SBG isolates formed sparse biofilm, indicating that persistence in soil was not associated with biofilm formation. There was no correlation between curli and biofilm formation in either medium (data not shown), in contrast to previous findings on isolates from human origin (Barnhart and Chapman, 2006). There was also no correlation between phylogroup and biofilm formation (data not shown).

3.5. *Winter survival of selected isolates*

In order to study the long-term fate of various isolates in pasture soil, strains selected from all three sample types and clusters were incubated in soil and kept outdoors from November until May. All 45 isolates were recovered after the winter, with the highest population decline being $\log_{10} 2.5$ (Fig.5a, b). SBG isolates had the lowest average decline, but due to the few isolates with higher decline the SBG group was not significantly better than the pasture and fecal isolates (Fig. 5a). Mixed cluster strains demonstrated a significantly greater degree of survival than the bovine cluster (Fig. 5b). Although not significant, environmental cluster strains showed a lower mean decline than

bovine strains. This result suggested that mixed and environmental clusters strains had a

higher
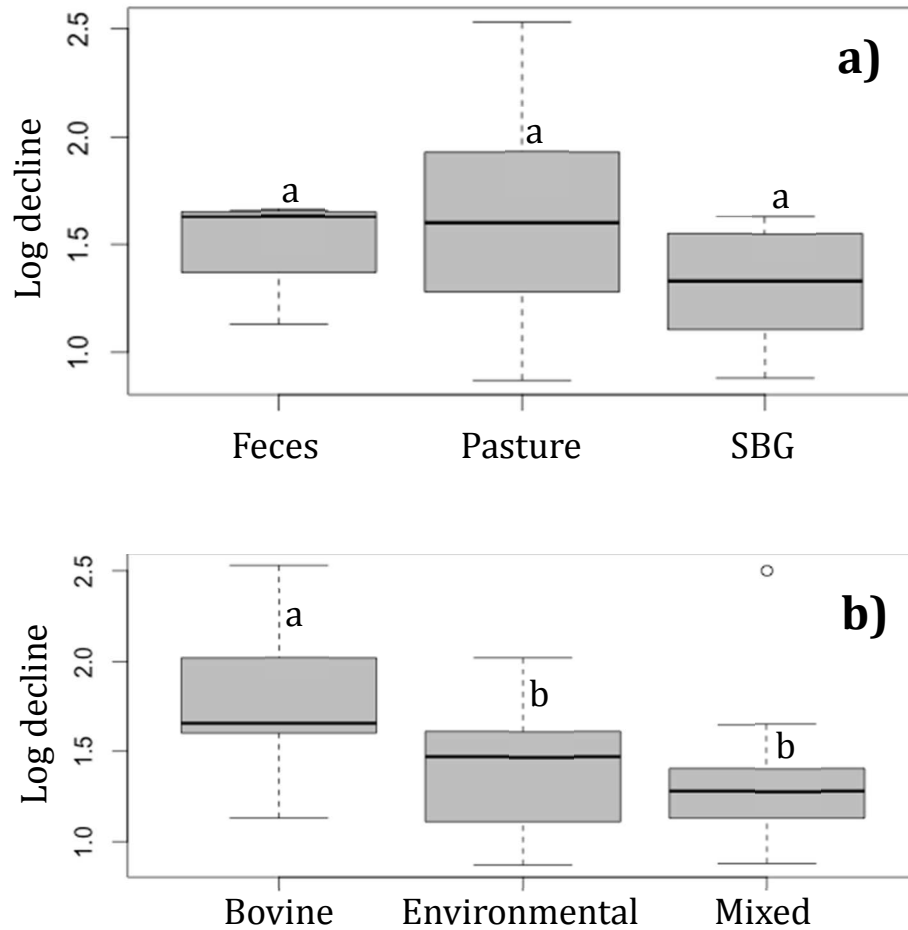
Fig. 2. 5 Log decline of *E. coli* isolates representing bovine cluster (B),

environmental cluster (E), mixed cluster (M) and sewage (SWG)

propensity to survive the winter in soil than bovine cluster isolates (p<0.05). Although

the difference in log decline was small, it did suggest a difference in fitness that may lead

to shift in the predominance of strains over time. Since sterile soil was used, this experiment was not performed in presence of competitors, therefore, a greater decline may occur under natural conditions.

4. Discussion

We herein attempted to clarify whether *E. coli* from bovine feces differed phenotypically and genotypically from isolates maintaining a population in pasture soil over winter, indicating niche-partitioning. *E. coli* strains that survived in pasture soil through the extreme South Dakota winter displayed a different genotype compared to bovine fecal isolates, while only few of the bovine derived genotypes were isolated from pasture after the winter. These results indicated that the niche partitioning of *E. coli* occurs between cattle and their pasture.

The phylogenetic analysis of concatenated *uidA* and *mutS* sequences showed diverse groups of isolates separated into different, well-supported clusters (Fig. 2). It is unlikely that all isolates in one cluster were clonal because multiple samples were taken from across a 12.14 Ha surface area populated with multiple cattle. Most of the reference strains used in this study clustered separately from our isolates, but were of human origin, and most of them were pathogenic or related to a clinical condition (Leimbach et al., 2013). None of these reference strains clustered in environmental clusters (Fig. 2). Some strains collected before grazing clustered separately from any bovine associated strains. As shown in previous studies, strains isolated from different aqueous and soil habitats showed a high genetic diversity (Buchan et al., 2001;Gordon, 2001;Lasalde et al., 2005;Higgins et al., 2007). A large number of soil isolates clustered with bovine isolates,

but with no SBG isolates in these clusters, suggesting that they were introduced recently to the pasture through feces. Mixed clusters contained only one and two bovine isolates respectively, indicating that select strains thrive in multiple niches, i.e. the bovine gastro-intestinal tract as well as pasture soil under varying weather conditions. Five small clusters contained only SBG and pasture isolates, suggesting these *E. coli* strains had the ability to survive the winter months through freeze- thaw cycles and subsequently grow in the summer months. Since these isolates may be environmentally adapted or naturalized (Ishii et al., 2006), we designated them environmental. *E. coli* populations detected in a number of environments such as soil and water in tropical, temperate, or alpine climates have been designated as naturalized or environmental (Byappanahalli et al., 2006;Ishii et al., 2006;Beversdorf et al., 2007;Texier et al., 2008;Brennan et al., 2010b). About 12% of our isolates fell under the environmental category and, thus, may represent multiple naturalized populations adapted to niches in pasture.

E. coli is considered to be a highly versatile and diverse species equipped with the ability to survive in many different habitats potentially stressful to other strains of the species. The genome flexibility of *E. coli* plays a key role in its metabolic and phenotypic diversity, increasing the competitiveness and fitness of individual variants in specific niches (Leimbach et al., 2013). Recently several groups of phylogenetically distinct *E. coli* found only in the environment were grouped into four "Clades", viz. I, III, IV and V (Luo et al., 2011b). These were designated by Luo et al. (2011b) as the "Environmental *E. coli*", implying only cryptic clades are environmental in nature. Walk, *et al* (Walk et al., 2009) reported that the members of Clades I, III, IV and V were found in higher abundance in environmental samples relative to human isolates. We did not obtain any of

these Clade isolates as per the *uidA* and *mutS* sequence analysis or by phylogrouping. Since all 15 clade strains evaluated showed the ability to grow on MLGA, were β-galactosidase positive, and more than half were also β-glucuronidase positive, our isolation protocol may have yielded Clade isolates if dominant in pasture samples. Members of Clusters Env 1-5 are possible environmental *sensu stricto E. coli*. In contrast to our results, Clades I, III, IV and V occur in association with backyard poultry (Blyton et al., 2015).

Most studies have used the initial phylogrouping protocol described by Clermont *et al* (Clermont et al., 2000). Phylogroups A and B1 are widely viewed as primarily environmental, whereas B2 and D are of mammalian origin (Gordon and Cowling, 2003b;Escobar-Paramo et al., 2006). The protocol was recently refined to yield eight phylogroups viz. A, B1, B2, C, D, E, F and Clade I (Clermont et al., 2013). Phylogroup C previously fell within A, and groups E and F fell within D. Phylogroup D was previously viewed as a pathogen group, and the newer group E emanating from it also contains pathogenic strains including enterohaemorrhagic *E. coli* such as O157:H7(Leimbach et al., 2013). We obtained many group E isolates, predominantly from bovine feces and recently grazed pasture, consistent with the mammalian association of the old phylogroup D. Phylogroup C isolates were primarily from bovine feces and pasture, with only a small proportion from SBG. A small number of Phylogroup B2 isolates were obtained from bovine feces, pasture or SBG, although association of B2 is mostly with humans and pigs (Carlos et al., 2010). B1 was the most numerous group in pasture soil, both before and after grazing. These have been reported to survive in the environment, and possibly become naturalized in fresh water beaches (Walk et al., 2007), and estuarine microcosms

(Berthe et al., 2013). The difference in the distribution of *E.coli* subpopulations (phylogroups) among various ecological sources is also influenced by fecal deposition of small animals (Bergholz et al., 2011).

The RDAR (red, dry and rough) morphotype is associated with multicellular growth and biofilm formation of *E. coli* for survival under harsh conditions (White et al., 2006). *E. coli* biofilm formation is associated with the expression of curli and extracellular polymeric substances such as adhesin, amyloid-forming protein, and exopolysaccharide. (Jonas et al., 2007;Romero et al., 2010;Vlamakis et al., 2013)We did not observe a correlation between phylogroup and RDAR morphotype, or between RDAR morphotype and biofilm formation, indicating that neither phenotype is conserved phylogenetically across bovine and pasture –associated *E. coli*. Phylogroup B1 was the notable exception with a high proportion of curli forming isolates. B1 phylogroup isolates from various animals, humans, and water had a markedly higher RDAR positive rate than the other phylogroups (White et al., 2011).

All *E. coli* isolates evaluated showed population survival in soil through the winter. However, the log decline was significantly different between members of the bovine versus mixed clusters. Since the mixed clusters contained only one and two bovine isolates, it is possible that these bovine isolates displayed fitness for the soil and GI tract niches. The overall log decline was not very high, this may have been because there was no competition with other microorganisms or predators. A previous study reported that *E. coli* K12 viability declined at a higher rate in non-sterile water, soil and sea-water than under sterile conditions (Bogosian et al., 1996). *E. coli* O157–H7 in pure culture maintained the population until day 30, and there was only one $\log_{10}$ decline after 179

days. Naturalized *E. coli* isolated from soil survived and grew better in sterile than non-sterile soil, indicating the presence of indigenous microbes negatively affected the growth of *E. coli* (Ishii et al., 2010b).

5. Conclusion

We demonstrate the occurrence of diverse groups of *sensu stricto E. coli* in a cattle pasture, some clustering with SBG, but not cattle isolates, indicating niche partitioning. The strains of these clusters were better able to survive winter freezing in soil. Amarked variation in the distribution of phylogroups also described the genetic diversity among isolates. Our results add further support to the existence of environmental *sensu stricto* (non-cryptic species) *E. coli* that has become naturalized in soil and form a reservoir of populations in the environment. Further studies are required in order to characterize the biology of these environmental *E. coli* isolates, including their adaptive abilities and pathogenicity.

6. References

Barnhart, M.M., and Chapman, M.R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.* 60**,** 131-147.

Bergholz, P.W., Noar, J.D., and Buckley, D.H. (2011). Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. *Appl. Environ. Microbiol.* 77**,** 211-219.

Berthe, T., Ratajczak, M., Clermont, O., Denamur, E., and Petit, F. (2013). Evidence for coexistence of distinct *Escherichia coli* populations in various aquatic Environments and their Survival in estuary water. *Appl. Environ. Microbiol.* 79**,** 4684-4693.

Beversdorf, L.J., Bornstein-Forst, S.M., and Mclellan, S.L. (2007). The potential for beach sand to serve as a reservoir for *Escherichia coli* and the physical influences on cell die-off. *J. Appl. Microbiol.* 102**,** 1372-1381.

Blyton, M.D.J., Pi, H., Vangchhia, B., Abraham, S., Trott, D.J., Johnson, J.R., and Gordon, D.M. (2015). Genetic structure and antimicrobial resistance of *Escherichia coli* and cryptic clades in birds with diverse human associations. *Appl. Environ. Microbiol.* 81**,** 5123-5133.

Bogosian, G., Sammons, L.E., Morris, P.J., O'neil, J.P., Heitkamp, M.A., and Weber, D.B. (1996). Death of the *Escherichia coli* K-12 strain W3110 in soil and water. *Appl. Environ. Microbiol.* 62**,** 4114-4120.

Brennan, F.P., O'flaherty, V., Kramers, G., Grant, J., and Richards, K.G. (2010). Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. *Appl. Environ. Microbiol.* 76**,** 1449-1455.

Buchan, A., Alber, M., and Hodson, R.E. (2001). Strain-specific differentiation of environmental *Escherichia coli* isolates via denaturing gradient gel electrophoresis (DGGE) analysis of the 16S-23S intergenic spacer region. *FEMS Microbiol. Ecol.* 35**,** 313-321.

Byappanahalli, M., and Fujioka, R. (2004). Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. *Water. Sci. Technol.* 50**,** 27-32.

Byappanahalli, M.N., and Fujioka, R.S. (1998). Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water. Sci. Technol.* 38**,** 171-174.

Byappanahalli, M.N., Shively, D.A., Nevers, M.B., Sadowsky, M.J., and Whitman, R.L. (2003). Growth and survival of *Escherichia coli* and enterococci populations in the macro-alga Cladophora (Chlorophyta). *FEMS Microbiol. Ecol.* 46**,** 203-211.

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

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

Carlos, C., Pires, M., Stoppe, N., Hachich, E., Sato, M., Gomes, T., Amaral, L., and Ottoboni, L. (2010). *Escherichia coli* phylogenetic group determination and its

application in the identification of the major animal source of fecal contamination. *BMC Microbiol.* 10**,** 161.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* 66**,** 4555-4558.

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

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. (2002). Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl. Environ. Microbiol.* 68**,** 1165-1172.

Diard, M., Garry, L., Selva, M., Mosser, T., Denamur, E., and Matic, I. (2010). Pathogenicity-associated islands in extraintestinal pathogenic *Escherichia coli* are fitness elements involved in intestinal colonization. *J. Bacteriol.* 192**,** 4885-4893.

Escobar-Paramo, P., Le Menac'h, A., Le Gall, T., Amorin, C., Gouriou, S., Picard, B., Skurnik, D., and Denamur, E. (2006). Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ. Microbiol.* 8**,** 1975-1984.

Gordon, D.M. (2001). Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiol.* 147**,** 1079-1085.

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

Gordon, D.M., and Cowling, A. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiol.* 149**,** 3575-3586.

Hartz, A., Cuvelier, M., Nowosielski, K., Bonilla, T.D., Green, M., Esiobu, N., Mccorquodale, D.S., and Rogerson, A. (2008). Survival potential of *Escherichia coli* and Enterococci in subtropical beach sand: implications for water quality managers. *J. Environ. Qual.* 37**,** 898-905.

Hayes, P.S., Blom, K., Feng, P., Lewis, J., Strockbine, N.A., and Swaminathan, B. (1995). Isolation and characterization of a beta-D-glucuronidase-producing strain of *Escherichia coli* serotype O157:H7 in the United States. *J. Clin. Microbiol.* 33**,** 3347-3348.

Higgins, J., Hohn, C., Hornor, S., Frana, M., Denver, M., and Joerger, R. (2007). Genotyping of *Escherichia coli* from environmental and animal samples. *J. Microbiol. Meth.* 70**,** 227-235.

Hothorn, T. (2014). "Party, v.1.0-17. Comprehensive R Archive Network". (Vienna, Austria: R Foundation for Statistical Computing).

Ishii, S., Hansen, D.L., Hicks, R.E., and Sadowsky, M.J. (2007). Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in lake Superior. *Environ. Sci. Technol.* 41**,** 2203-2209.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006). Presence and growth of naturalized *Escherichia coli* in temperate soils from lake Superior watersheds. *Appl. Environ. Microbiol.* 72**,** 612-621.

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

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

Jonas, K., Tomenius, H., Kader, A., Normark, S., Romling, U., Belova, L., and Melefors, O. (2007). Roles of curli, cellulose and BapA in *Salmonella* biofilm morphology studied by atomic force microscopy. *BMC Microbiol.* 7**,** 70.

Ksoll, W.B., Ishii, S., Sadowsky, M.J., and Hicks, R.E. (2007). Presence and sources of fecal coliform bacteria in epilithic periphyton communities of lake Superior. *Appl. Environ. Microbiol.* 73**,** 3771-3778.

Lang, N.L., Smith, S.R., Bellett-Travers, D.M., Pike, E.B., and Rowlands, C.L. (2003). Decay of *Escherichia coli* in soil following the application of biosolids to agricultural land. *Water. Environ. J.* 17**,** 23-28.

Lasalde, C., Rodriguez, R., and Toranzos, G.A. (2005). Statistical analyses: Possible reasons for unreliability of source tracking efforts. *Appl. Environ. Microbiol.* 71**,** 4690-4695.

Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E., and Tenaillon, O. (2007). Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol. Biol. Evol.* 24**,** 2373-2384.

Leimbach, A., Hacker, J., and Dobrindt, U. (2013). *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* 358**,** 3-32.

Letunic, I., and Bork, P. (2011). Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39**,** W475-W478.

Liu, S., Jin, D., Lan, R., Wang, Y., Meng, Q., Dai, H., Lu, S., Hu, S., and Xu, J. (2015). *Escherichia marmotae* sp. nov., isolated from faeces of *Marmota himalayana*. *Int. J. Syst. Evol. Microbiol.* 65**,** 2130-2134.

Luo, C.W., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011). Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc. Natl. Acad. Sci.* 108**,** 7200-7205.

Mcdonald, J.H. (2009). *Handbook of Biological Statistics.* Baltimore, Maryland, U.S.A.: Sparky House Publishing.

O'toole, G.A. (2011). Microtiter dish biofilm formation assay. *J. Vis. Exp.: JoVE***,** 2437.

Ogden, C.B., Vanes, H.M., and Schindelbeck, R.R. (1997). Miniature rain simulator for field measurement of soil infiltration. *Soil Sci. Soc. Am. J.* 61**,** 1041-1043.

Osborn, M.J., and Trussell, R.R. (2004). *Indicators for waterborne pathogens.* Washington, DC: The National Academies Press.

R Development Core Team (2015). "R: A language and environment for statistical computing". (Vienna, Austria: R Foundation for Statistical Computing).

Rambaut, A. (2002). SE-AL v. 2.0a11: sequence alignment program. (http://tree.bio.ed.ac.uk/software/seal/).

Romero, D., Aguilar, C., Losick, R., and Kolter, R. (2010). Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc. Natl. Acad. Sci.* 107**,** 2230-2234.

Sanchez, S., Martinez, R., Rey, J., Garcia, A., Blanco, J., Blanco, M., Blanco, J.E., Mora, A., Herrera-Leon, S., Echeita, A., and Alonso, J.M. (2010). Pheno-genotypic characterisation of *Escherichia coli* O157:H7 isolates from domestic and wild ruminants. *Vet. Microbiol.* 142**,** 445-449.

Savageau, M.A. (1983). *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Amer. Nat.* 122**,** 732-744.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000). Sources of *Escherichia coli* in a coastal subtropical environment. *Appl. Environ. Microbiol.* 66**,** 230-237.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30**,** 2725-2729.

Texier, S., Prigent-Combaret, C., Gourdon, M.H., Poirier, M.A., Faivre, P., Dorioz, J.M., Poulenard, J., Jocteur-Monrozier, L., Moenne-Loccoz, Y., and Trevisan, D. (2008). Persistence of culturable *Escherichia coli* fecal contaminants in dairy alpine gassland soils. *J. Environ. Qual.* 37**,** 2299-2310.

Vilain, S., Luo, Y., Hildreth, M.B., and Brozel, V.S. (2006). Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl. Environ. Microbiol.* 72**,** 4970-4977.

Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013). Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat. Rev. Microbiol.* 11**,** 157-168.

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

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., and Whittam, T.S. (2009). Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* 75**,** 6534-6544.

White, A.P., Gibson, D.L., Kim, W., Kay, W.W., and Surette, M.G. (2006). Thin aggregative fimbriae and cellulose enhance long-term survival and persistence of *Salmonella*. *J. Bacteriol.* 188**,** 3219-3227.

White, A.P., Sibley, K.A., Sibley, C.D., Wasmuth, J.D., Schaefer, R., Surette, M.G., Edge, T.A., and Neumann, N.F. (2011). Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. *Appl. Environ. Microbiol.* 77**,** 7620-7632.

White, A.P., Weljie, A.M., Apel, D., Zhang, P., Shaykhutdinov, R., Vogel, H.J., and Surette, M.G. (2010). A global metabolic shift is linked to *Salmonella* multicellular development. *PLoS ONE* 5**,** e11814.

Winfield, M.D., and Groisman, E.A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl. Environ. Microbiol.* 69**,** 3687-3694.

Table S2. 1 List of Primers

Primer Name	Sequence (5' – 3')	Size of product	References
uidA	CATTACGGCAAAGTGTGGGTCAAT (F) TCAGCGTAAGGGTAATGCGAGGTA (R)	658 bp	Walk *et al.*, 2009
mutS	GGCCTATACCCTGAACTACA (F) GCATAAAGGCAATGGTGTC (R)	596 bp	Walk *et al.*, 2009
chuA	ATGGTACCGGACGAACCAAC (F) TGCCGCCAGTACCAAAGACA (R)	288 bp	Clermont et al. , 2013
yjaA	CAAACGTGAAGTGTCAGGAG (F) AATGCGTTCCTCAACCTGTG (R)	211 bp	Clermont et al. , 2013
TspE4	CACTATTCGTAAGGTCATCC (F) AGTTTATCGCTGCGGGTCGC (R)	152 bp	Clermont et al. , 2013
arpA	AACGCTATTCGCCAGCTTGC (F) TCTCCCCATACCGTACGCTA (R)	400 bp	Clermont et al. , 2013
Group E Specific ArpAgpE	GATTCCATCTTGTCAAAATATGCC (F) GAAAAGAAAAAGAATTCCCAAGAG (R)	301 bp	Clermont et al. , 2013
Group C specific trpAgpC	AGTTTTATGCCCAGTGCGAG (F) TCTGCGCCGGTCACGCCC (R)	219 bp	Clermont et al. , 2013

bp : Base pair

Chapter 3: Niche Preferences of *Escherichia coli* in a Peri-Urban Pond

The work reported in this chapter was conducted by Gitanjali NandaKafle in collaboration with Taylor Huegen, Sarah C. Potgieter, Emma Steenkamp and Stephanus N. Venter

1. Introduction

Escherichia coli is a very diverse species with a large pan genome, and viewed primarily as a resident of the mammalian colon (Touchon et al. 2009). The sources of *E. coli* found in water bodies are believed to be humans, farm animals or wild animals, so-called fecal contamination (Ishii et al. 2006; Ishii et al. 2007). This includes commensals and diverse strains pathogenic to humans and animals. *E. coli* was thought unable to grow in secondary habitats such as water, soil, and sediments (Winfield and Groisman 2003). Recent studies have shown that *E. coli* can survive and even grow in water, sediments, soil, and water-plants in various climatic regions where no evidence for fecal contamination exists (Byappanahalli and Fujioka 1998; Byappanahalli et al. 2006; Ishii et al. 2006; Beversdorf et al. 2007). Some *E. coli* appear to become naturalized in the environment, with distinct genotypes from strains in animal hosts (Ishii et al. 2006; NandaKafle et al. 2017). *E. coli* populations from different aquatic environments also showed extensive genetic diversity (McLellan 2004; Byappanahalli et al. 2007; Casarez et al. 2007). Previous studies have suggested that there is a relationship between

genotypes of *E. coli* found among specific animal hosts and the geographic location from which they were isolated (Byappanahalli et al. 2006; Ishii et al. 2006; Petit et al. 2017).

DNA fingerprinting methods have revealed the diversity of *E. coli* populations obtained from different sources (McLellan et al. 2003; Anderson et al. 2006; Chandran and Mazumder 2013; 2014). *E. coli* strains are typically classified into four major phylogroups (A, B1, B2 and D)(Clermont et al. 2000), and later eight phylogroups based on their genomic information. Of these, seven (A, B1, B2, C, D, E, F) belong to *E. coli sensu stricto* whereas the eighth one is represented by cryptic clade-I (Clermont et al. 2013). There are variations in genotypic and phenotypic traits among strains of different phylogroups (Bergthorsson and Ochman 1998; Gordon 2004). It is widely held that the four major phylogroups differ in their ecological habitats, with phylogroups A and B1 occurring more frequently in the environment than B2 and D (Walk et al. 2007). Moreover, phylogroup B2 and D strains were frequently isolated from extra-intestinal sites within host bodies (Gordon 2004). Some strains belonging to phylogroup B1, were reported to persist in water (Walk et al. 2007; Ratajczak et al. 2010) and soil (NandaKafle et al. 2017). Many studies have also reported that phylogroup B2 and, to a lesser extent D strains are likely to be more virulent than other phylogroups (Picard et al. 1999; Johnson and Stell 2000; Le Gall et al. 2007). Interestingly, virulence genes are more frequently present in phylogroup B1 isolates from environments where phylogroup B2 strains are absent (Unno et al. 2009). Thus, identification of the phylogroup of unknown isolates may provide information on their physiological and ecological characteristics.

Although most *E. coli* are commensal, eight pathovars have been well characterized. Each of these pathovars uses a large collection of virulence factors to disrupt host cellular

functions to augment its virulence activities (Croxen and Finlay 2010). Among the eight pathovars, Enterohemorrhagic *E. coli* (EHEC) and Enterotoxigenic *E. coli* (ETEC) are well recognized for causing infection in humans and domestic animals. EHEC is a human pathogen responsible for bloody diarrhea and hemolytic uremic syndrome (HUS). The morbidity and mortality associated with several outbreaks of EHEC disease poses a serious public health concern (Nguyen and Sperandio 2012). EHECs are often characterized by the presence of specific virulence factors; most strains produce at least one Shiga like toxin (encoded by *Stx1* or *Stx2*), hemolysin (EHEC specific plasmid encoded hemolysin, encoded by *hlyA*), and intimin (attachment and effacement protein, encoded by *eaeA*) (Fagan et al. 1999). ETEC strains that cause diarrhea produce heat labile (LT) and heat stable (ST) enterotoxins (Erume et al. 2008). These are the common cause of diarrhea in children and adults living in the developing world with inadequate clean water and poor sanitation, also known for causing traveler's diarrhea (Qadri et al. 2005). Neonatal and post weaning diarrhea due to ETEC is an important cause of economic loss to the pig industry (Francis 2002).

Bacterial infections are becoming increasingly difficult to treat due to wide spread antibiotic resistance among pathogens. The current emergence of antibiotic resistant pathogenic bacteria in worldwide has become a matter of concern for public and animal health. Antibiotics are often used for the treatment of *E. coli* infections and may be incorporated into commercial livestock and poultry feed at a sub-therapeutic dose to promote growth. Over time, selection pressure selects resistant strains with specific resistance genes, and *E. coli* populations may show a specific antibiotic resistance pattern

depending on their habitats (Smith and Coast 2002; Collignon et al. 2009; Amaya et al. 2012).

We hypothesized that environmentally adapted *E. coli* associate preferentially with specific niches. A secluded peri-urban pond adjacent to cattle pasture was selected as sampling site. We isolated *E. coli* from water, sediment, submerged water plants and water snails, as well as from bovine feces in adjacent pasture. To determine evidence of niche partitioning, isolates were characterized genotypically by phylogrouping and analysis of their *uidA* and *mutS* sequences, and phenotypically for antibiotic resistance and virulence gene distribution.

2. Materials and Methods:

2.1. *Sample source*

Samples were collected from a secluded pond (GPS co-ordinate 44.2719° N, 96.7736° W) at the edge of Brookings, SD, USA during June and July 2013. This pond is located between the edge of town and a nature park, and surrounded by dense scrub and trees, rarely frequented by humans. Water (31), sediment (27), water plant (35), and snail samples (20) were collected from the pond, and bovine feces (7) was collected from an adjoining cattle pasture. Samples were placed into sterile 50 mL conical screw cap tubes, brought to laboratory on ice and processed on the same day.

2.2. *Isolation of* **E. coli**

Water samples were filtered (10 mL and 1 mL) through a sterile 0.45 μm mixed cellulose ester filter (Milipore) and the filters placed on Membrane Lactose Glucuronide agar (MLGA, Fluka analytical). Sediment samples were mixed with 15 mL of sterile dH_2O and 1 and 10 mL aliquots filtered before placing filters onto MLGA. Water plants, snails were rinsed with sterile dH_2O, then crushed in 10 mL sterile dH_2O, and 100 μL plated directly on to MLGA plates. An aliquot of tenfold dilutions of suspended feces samples and all other samples were also plated onto MLGA. Green colonies indicated positive for β-Galactosidase and β-Glucuronidase and were assumed to be *E. coli*. This protocol excluded ß-glucuronidase negative O157:H7 strains. An average of two colonies were selected at random from the highest dilutions showing growth, streaked onto MLGA to confirm purity, sub-cultured on LB agar, and stored at -80 °C in 50% glycerol.

2.3. *Analysis of the* **uidA** *and* **mutS** *gene sequences*

Genomic DNA was extracted from overnight LB agar cultures suspended in 10 mM phosphate buffer (pH 7.0), using the genomic DNA Quick Prep Kit (Zymo Research), and stored at -20°C. The *uidA* and *mutS* genes were amplified by PCR using primers described previously (Walk et al. 2009) (Table S1). PCR reactions (25 μl) were set up as follows: 2.5 μl reaction buffer (10X) (New England Biolab), 1.5 μl $MgCl_2$ (25mM), 0.5 μl dNTPs (40mM), 0.1 μl forward primer and 0.1 μl reverse primer (100 μmol), 0.125 μl

of Taq polymerase (NE Biolabs), 0.5 µl of DNA template and 20.7 µl sterile nano pure

water. The amplification cycle was initiated with 95°C for 2 min, followed by 30 cycles

of denaturing at 95°C for 30 s, annealing at 56°C for 30 s and extension at 72°C for 1

min, with a final extension at 72°C for 5 min. DNA sequences were determined by the

dideoxy chain termination method (Beckman Coulter Genomic Center at Denver, MA).

The *uidA* and *mutS* sequences were submitted to Genbank

(http://www.ncbi.nlm.nih.gov/genbank/) under BankIt2031081: MF459726 - MF459846

and BankIt2031086: MF459847 - MF459967 respectively.

The sequences generated were analyzed using two approaches. To infer the relationships

among isolates, DNA sequences were aligned using ClustalW (Tamura et al. 2011), and

overhangs were trimmed using SeAl (Rambaut 2002). The *uidA* and *mutS* sequences for

all isolates and reference strains (Leimbach et al. 2013) were concatenated using SeAl. A

maximum likelihood analysis using model GTR+G+I with 1,000 bootstrap replicates was

performed in the program MEGA6.06 (Tamura et al. 2011). The tree was then annotated

and visualized using the ITOL online tool (Letunic and Bork 2011).

2.4. *Population genetic analysis*

To infer population structure and assign isolates to distinct populations, we employed a

model-based clustering method using the STRUCTURE software (Falush et al. 2003).

More specifically the admixture model was applied using sample locations as prior

(LOCPRIOR). By assuming mixed ancestry, individuals within a population are thought

to have inherited a fraction of their genome from an ancestor in the population (Pritchard

et al. 2000). Ln probability values and the variance of Ln likelihood scores were

estimated for the concatenated *uidA-mutS* sequences, assuming the presence of 2 populations (K = 2, with an adjusted alpha = 0.5) and performing twenty iterations for each K from K = 1 to K = 6. Estimated burn-in period of 10 000 and a run length of 500 000 (Hubisz et al. 2009). All other parameters in STRUCTURE were left as default. The resulting data from STRUCTURE were collated and visualized using the web-based program Structure Harvester (Pritchard et al. 2000) to assess which likelihood values across the multiple estimates of K best that explained the data (in this case K=3 was the best) using the Evanno method (Evanno et al. 2005; Earl and Vonholdt 2012). Furthermore, optimal alignments for the number of replicate cluster analyses were generated using the FullSearch algorithm in CLUMPP (Jakobsson and Rosenberg 2007) and resulting output files were used directly for cluster visualization as plots in Excel and the program Distruct 1.1 (Rosenberg 2004).

2.5.   *Phylogroup analysis*

Isolates were assigned to phylogroups using the protocol of Clermont et al. (2013). To avoid ambiguity, PCR was performed separately for each primer set (Table S1). Phylogroup similarity among the sample courses was determined by carrying out UPGMA analysis using the constrained Jaccard coefficient in PAST version 3.14 (http://folk.uio.no/ohammer/past) (Hammer et al. 2001). To determine whether the distribution of phylogroups differed by source or cluster we used multinomial log-linear regression models. The models were fitted using the nnet package in R (v.3.2.2)(R Development Core Team 2015). The response variable in this analysis was the phylogroup of each isolate (A, B1, B2, C, D, E, and Unknown), and the explanatory

variables were the sample source and clusters associated with origin of the isolates. To

visualize the effect of significant explanatory variables, we used regression trees fitted

using Package Party (Hothorn 2014) in R.

2.6. *Virulence gene assays*

PCR for detection of *stx1*, *stx2*, *eaeA*, *hlyA* genes, was performed using primers as

described by Fagan et al. (1999) (Table S-2) and for ST and LT virulence genes as

described by Osek (2001) (Table-S2). DNA samples for PCR were prepared by the

boiling method. Stock cultures were recovered on LBA, two colonies suspended in 500

μL dH_2O, washed by centrifugation and suspension in fresh sterile dH_2O, lysed by

incubating at 100°C for 10 min, and immediately chilled on ice for 5 min. Debris was

removed by centrifugation for 1 min at 12,000 X g and the supernatant was carefully

transferred to a new sterile tube and stored at -20°C for further use as PCR template. PCR

reactions were carried out in 25 μl volume containing 1 μl of DNA template, 2.5 μl

reaction buffer (10X) (New England Biolab), 1.5 μl MgCl_2 (25mM), 0.5 μl dNTPs

(40mM), 0.1 μl forward primer and 0.1 μl reverse primer (100 μmol), 0.1 μl of Taq

polymerase (NE Biolab), and 19.2 μl sterile nano pure water. PCR amplification for *stx1,*

stx2, eaeA, and *hlyA* were performed under the following conditions: initial 95°C

denaturation step for 3 min followed by 35 cycles of 20 s denaturation at 95°C, 40 s

primer annealing at 58°C, and 90 s extension at 72°C. The final cycle was followed by a

72°C incubation for 5 min (Fagan et al. 1999). LT and ST were amplified under the

following conditions: an initial DNA denaturation step at 94 C for 5 min followed by 30

cycles of 1 min of denaturation at 94°C, 1 min of primer annealing at 55°°C, and 2 min of

extension at 72°C. The final extension step was performed at 72°C for 5 min (Osek

2001).

2.7. *Antibiotic resistance assays*

Antibiotic susceptibility of 120 *E. coli* isolates was determined using disk diffusion assay

following the CLSI standard (CLSI 2015). Stock cultures were recovered in 5 mL

Mueller Hinton (MH, Oxoid) broth at 37°C for 16h. Cells were harvested by

centrifugation (10,000 X g, 2 min), re-suspended in sterile tap water and the cell density

adjusted to 0.5 on the McFarland turbidity standard. Cell suspensions were spread onto

the MH agar (Oxoid), and antibiotic disks Ciprofloxacin (5 µg), Meropenem (10 µg),

Ceftrixome (30 µg), Gentamicin (10 µg), Azythromycin (15 µg), Tetracycline (30 µg),

with Penicillin (10 µg) as control) were placed on the surface. After 18h incubation at

37°C, zone diameters were measured and isolates scored as intermediately or fully

resistant according to the (CLSI 2015). Isolates resistance to two or more antimicrobials

were defined as multidrug resistant. *E. coli* ATCC 25922 was included for each assay as

a negative control.

3. Results

E. coli were obtained from water, sediment, water plants and water snails in the pond, as

well as from bovine feces in the adjoining pasture. The population composition as

determined by both genotypic and phenotypic traits was similar for water, sediment, and

water plant niches, while the population composition of snail and bovine fecal

populations differed.

3.1. *Phylogroup distribution*

Populations obtained from the water, sediment and submerged water plants showed

similar phylogroup distribution (Fig. 1), predominated by phylogroups B1, E and some

B2 isolates. Sediment isolate distribution differed.
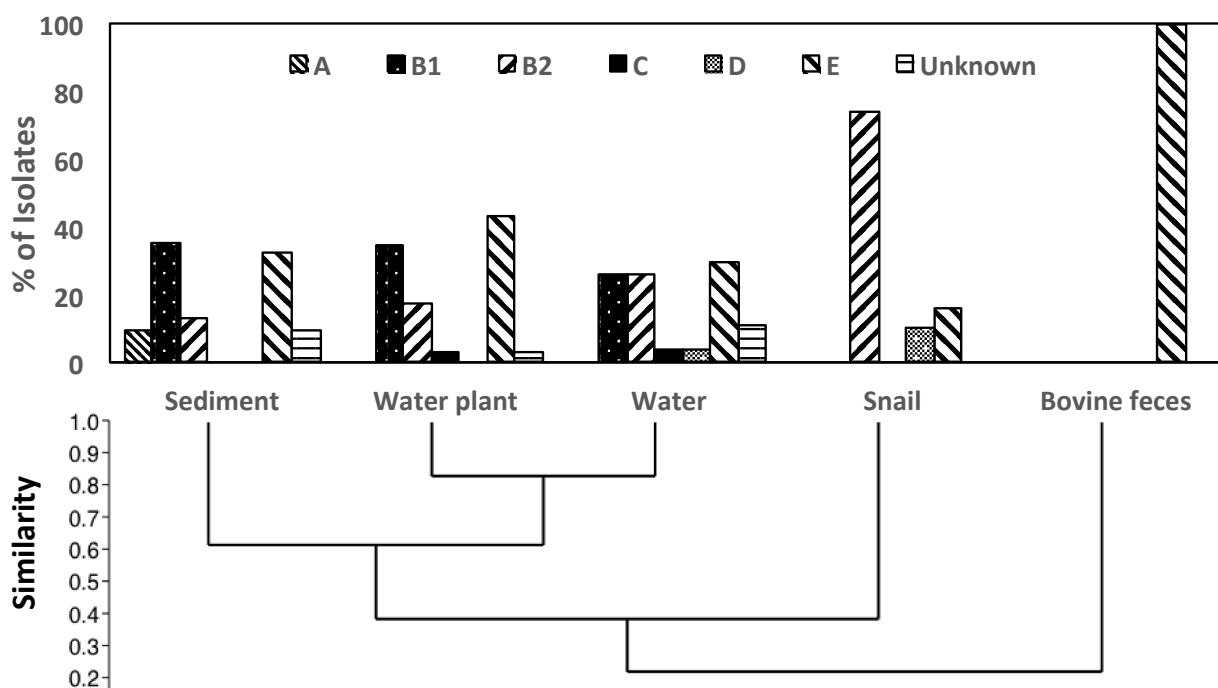
Fig. 3. 1 Phylogroup distribution across isolates from the five sample types.

Phylogrouping was performed according to the scheme of Clermont et al., 2013. The

relatedness between virulence gene distribution profiles was determined by

UPGMA using the constrained Jaccard coefficient.

In contrast, snail populations were predominated by phylogroup B2. Multinomial log

linear regression supported a significant difference (p < 0.001) between snail populations

and the water, sediment and plant populations (Fig. 1S). Fecal isolates were all

phylogroup E. As our isolation method was based on MLGA (β-Glucuronidase and β-Galactosidase), phylogroup E strains lacking the *uidA* gene for β-Glucuronidase would have been excluded (Chang et al. 1989). Yet, we obtained several isolates from feces.
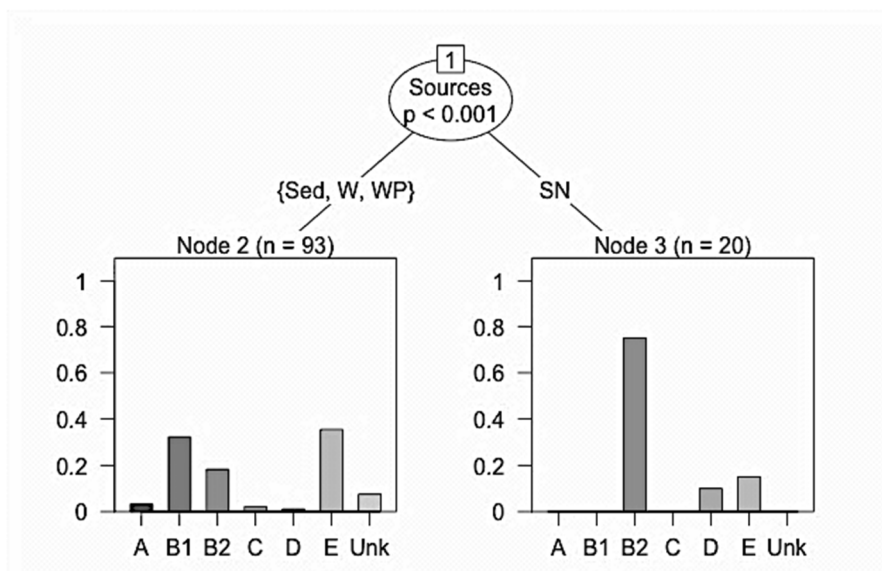
Fig. S3. 1 Multinomial log-linear regression analysis of phylogroup distribution of isolates across sample types. Phylogrouping was performed according to the scheme of Clermont et al., 2013. The X axis denotes phylogroups and the Y-axis represents proportion of isolates. Sed – sediment, W – water, WP – water plant, SN – snail

3.2. Phylogenetic analysis

The concatenated *mutS* and *uidA* sequence phylogeny formed many well-separated clusters with strong bootstrap support (Fig. 2). Grey circles denote greater that 80% bootstrap support. Most of the water, water plant, and sediment isolates fell into mixed clusters, some with reference strains. This indicated the co-occurrence of these strains across the three niches. None of our isolates aligned with any of the Clade I, III, IV or V

Fig. 3. 2 Phylogenetic analysis of the concatenated *uidA* and *mutS* gene sequences of *E. coli* isolates, reference strains and cryptic species of *E. coli*. Sequences were aligned using ClustalW and manually trimmed using Se-Al. The best Model: Maximum Likelihood analysis with GTR and G+I was performed in the program MEGA 6. Numbers represent branch support of 1000 bootstrap replicates. The phylogenetic tree was color-coded and visualized using the Interactive Tree of Life.

Isolates are color-coded based on their sources. Grey circles on branches indicate a bootstrap value of > 80% (1000 bootstraps).

strains. The majority of snail isolates grouped into three unique clusters that contained no water, sediment or water plant isolates, and also no reference strains, indicating unique strains, and supporting a preference of these strains for snails over surrounding water, sediment or water-plant isolates. Three of the snail isolates did cluster with water plant, sediment, water and reference strains. All bovine fecal isolates formed a separate cluster with no reference strains (Fig. 2), indicating hitherto poorly studied diversity of species.

3.3. *Population genetic analysis*

Population genetic analysis of concatenated uidA and mutS genes was performed assuming one aquatic and one fecal population (i.e. K = 2, alpha = 0.5). The result obtained from the Evanno table was K = 3, supporting the existence of three separate genetic backgrounds (Fig. 3). The bovine fecal population was homogenous, indicating one genetic background. The snail population comprised of two homogenous backgrounds, one of these being identical to fecal background. In contrast, water, sediment and water plant population isolates comprised of a mixture of three backgrounds. Some were as bovine feces, some as the second snail background, but a distinct third admixture was observed, unique to all aquatic populations but snails. Thus the pond ecosystem comprised of an admixture of strains representing three populations, one likely due to introduction of bovine-derived strains, a second associated with snail
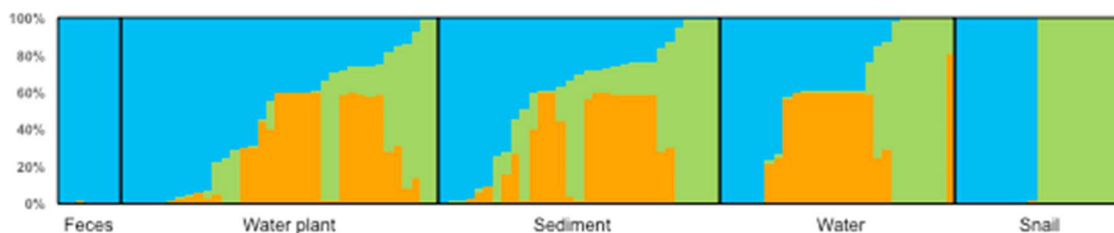
Fig. 3. 3 Population structure analysis of isolates based on their *uidA* and *mutS* sequences. The admixture model was applied, assuming one aquatic and one fecal population, for K=2, alpha = 0.5. The result obtained from the Evanno table was K = 3, supporting the existence of three separate genetic backgrounds represented by the three colors.

populations, and a third unique to the aquatic environment. This indicates that the water, sediment and water plant populations have shared gene flow, but with admixture from the fecal and snail populations. The water, water plant, sediment, and snail populations did not contribute any admixture to the fecal population. The pond ecosystem isolates represented three genetic backgrounds that are not specifically linked to phylotypes.

3.4. *Virulence gene distribution*

All isolates were screened for the presence of major virulence genes associated with diarrhaeagenic *E. coli* in order to determine their pathogenic potential. Out of six genes four (*Stx2, eaeA, hlyA* and *STb*) were detected in these isolates. Among these *eaeA* was the most frequently detected (36.13%), then *Stx2* (12. 61%), LTa (10.9%) and *hlyA* (3.36%). Distribution of the virulence genes in *E. coli* populations of water, sediment

and water plants was similar, supporting exchange of isolates among these niches (Fig. 4).
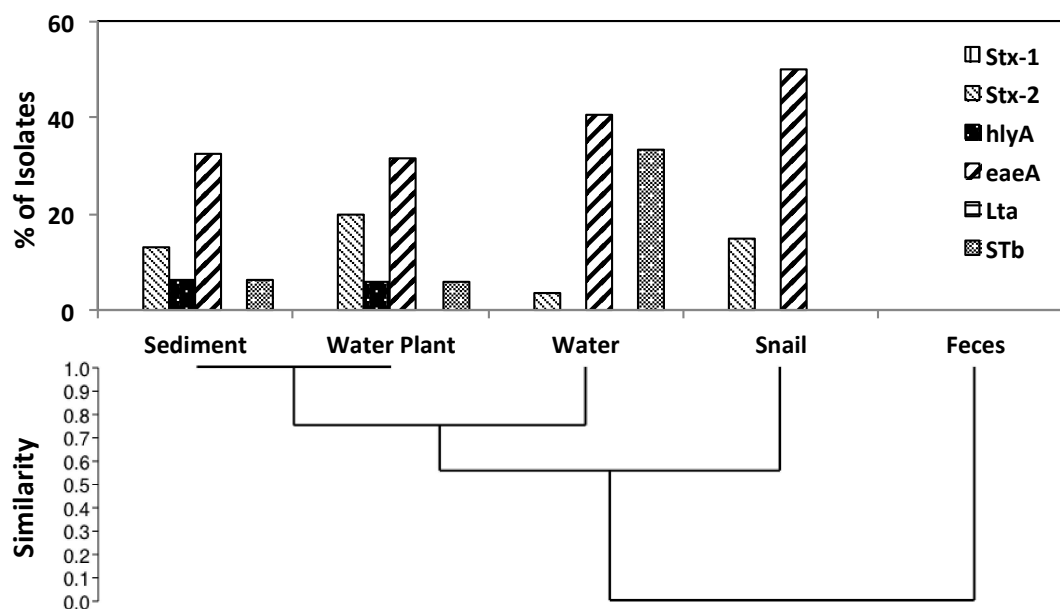
Fig. 3. 4 Virulence gene distribution across isolates from the five sample types.

The relatedness between virulence gene distribution profiles was determined by UPGMA using the constrained Jaccard coefficient.

Yet the water population was much richer in prevalence of the *STb* gene, and had no isolates with the *hlyA* gene. Virulence gene distribution of snail populations was different, with more than half the isolates containing the *eaeA* gene. We did not detect any isolates with the *Stx-1* and LT genes, although the positive control EDL933D and O157:K87:K88 (Francis and Willgohs 1991) yielded positive results confirming reliability of the assay. While all bovine feces isolates belonged to phylogroup E, none contained any of the six virulence genes (Fig. 4). The MLGA used for primary isolation would not obtain β-glucuronidase negative strains, so some phylogroup E strains that may have contained some virulence genes may have been excluded.

3.5. *Antibiotic resistance profiling*
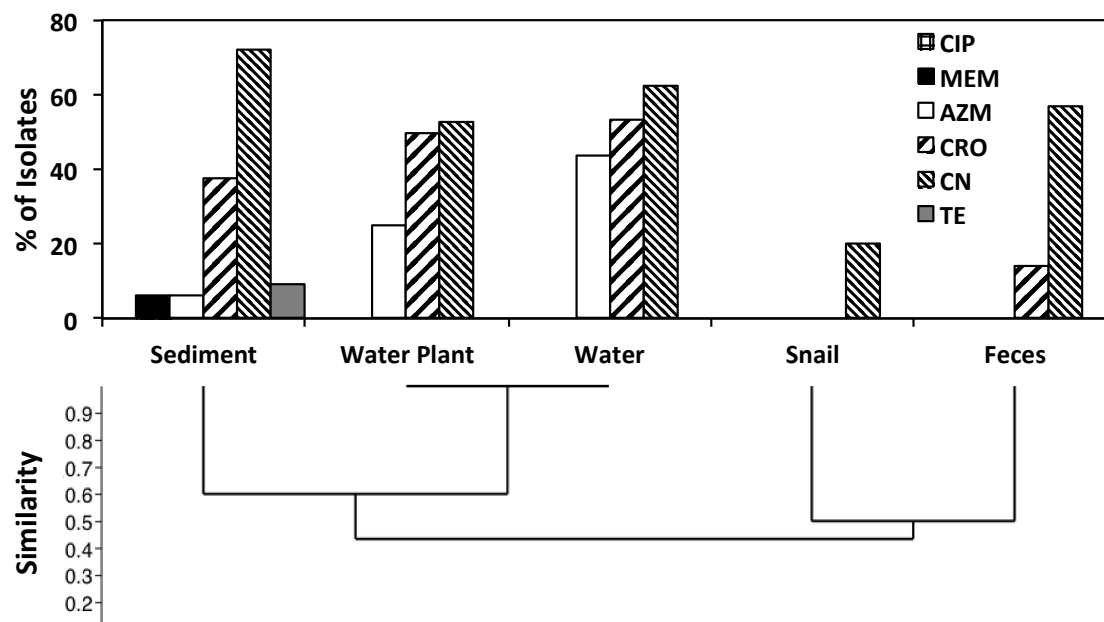
Fig. 3. 5 Antibiotic resistance across isolates from the five sample types. The relatedness between resistance profiles was determined by UPGMA using the constrained Jaccard coefficient.

One antibiotic was chosen from each of seven target classes to test the susceptibility of isolates: ceftriazone (class cephalosporins), ciprofloxacin (class- fluoroquinolones), gentamicin (class-aminoglycosides), azithromycin (class-macrolides), meropenem (class-carbapenems), and tetracycline. Water and water plant populations showed a similar distribution, with 60% of isolates resistant to gentamicin (Fig. 5). Sediment antibiotic resistance was different somewhat from water and water plant populations. Water, water plant, and sediment samples contained isolates resistant to three antibiotics, many of which also contained the *eaeA* gene as well as either *STb* or *hlyA* (Fig. 6). Snail

populations had a unique antibiotic resistance profile, with 80% of isolates sensitive to all antibiotics (Fig. 6), whereas only 20% of water, sediment and water plant isolates were not resistant to any of the antibiotics. However, most of the snail isolates displayed intermediate resistance to three or four antibiotics (Fig. S2). Bovine feces populations displayed their own antibiotic resistance profile (Fig. 5), with 80% of isolates displaying intermediate resistance to either two or three antibiotics.

Virulence gene	Sediment				Water Plant				Water				Snail				Feces			
Stx1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stx2	0	11.8	28.6	0	14.2	21.4	27.3	0	0	0	12.5	0	12.5	25	0	0	0	0	0	0
hlyA	0	5.9	0	50	0	14.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
eaeE	60	29.4	28.6	50	14.3	21.4	54.5	66.7	14.3	42.9	62.5	40	43.8	50	0	0	0	0	0	0
Lta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stb	0	11.8	0	0	28.6	0	0	0	0	28.6	62.5	40	0	0	0	0	0	0	0	0
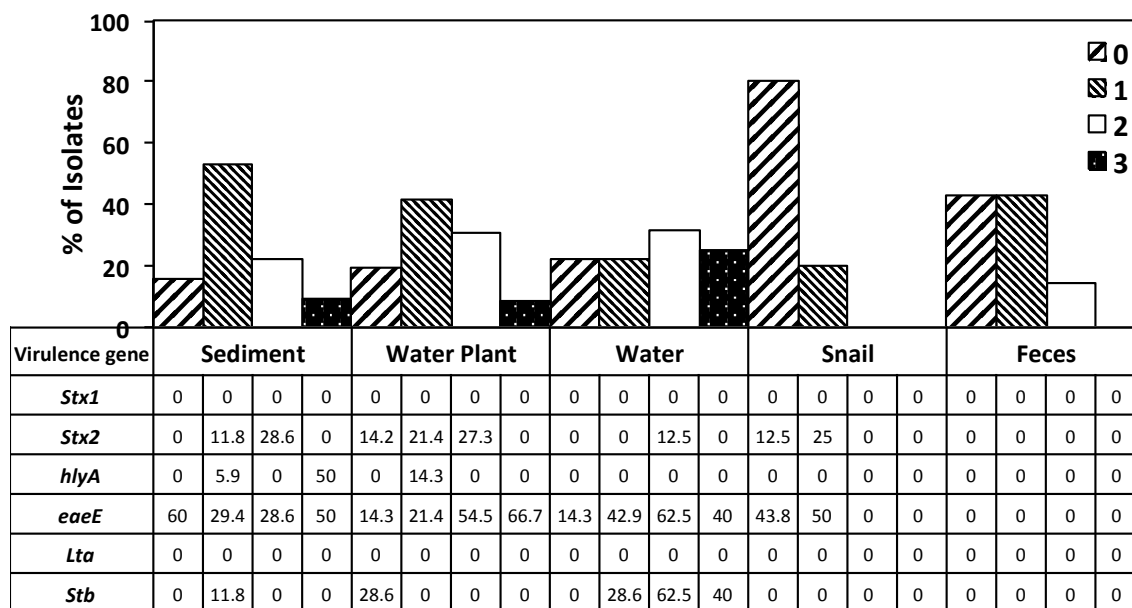
Fig. 3. 6 Sensitivity and multidrug resistance (resistance to 0, 1, 2 or 3 antibiotics) across sample types, compared to occurrence of virulence genes (percentage)

4. Discussion

The goal of this study was to determine whether environmental *E. coli* associate with specific niches. To obtain the evidence in support of niche partitioning, isolates from five niches were characterized genotypically and phenotypically.

Phylogroup distribution in water was similar to water plants while the sediment population differed, containing phylogroup A but not phylogroup C. Yet no statistical significance difference could be shown among the three. Snail population were significantly different to all other populations, predominantly phylogroup B2. Isolates from water, water plant and sediment were predominantly B1 and E, with fewer B2. This result was consistent with previous studies where B1 were generalist and harbor traits linked to plant association whereas B2 strains are more host associated (White et al. 2011; Meric et al. 2013). Phylogroup distribution within the *E. coli* population in both water and superficial sediments showed spatial variation (Petit et al. 2017). It has also been reported that phylogenetic groups are adaptable and genotypically influenced by changes in environmental conditions, however phylogroup B1 isolates seem to persist in water (Ratajczak et al. 2010; Jang et al. 2014). Our data indicated that B2 populations occurring in fresh water pond persist in snails. Likewise, phylogroup E strains predominant in feces deposited nearby did not thrive in the pond. The phylogroup E strains in the pond differ from those in feces, indicating niche preference among them. The composition differences of phylogroups among populations in different environments may be caused by differences in adaptability and plasticity of *E. coli* strains (Jang et al. 2014). Such variation in phylogroup distribution suggests that *E. coli* phylogroups are affected by niche specific selective pressures (Meric et al. 2013).

The *mutS* and *uidA* phylogenetic tree analysis showed that some clusters are devoid of reference strains and some are with reference strains. B1 phylogroups clustered with reference strains, indicating some of these strains may come from human sources. All isolates from the bovine feces fell under phylogroup E but formed a completely different cluster in the phylogenetic tree of *mutS* and *uidA* sequence, indicating a separate group of isolates. Phylogroup E did not cluster with any reference strains suggesting these isolates are different to those from humans. In our previous study we also found a higher percentage of phylogroup E in bovine fecal isolates compared to soil isolates (NandaKafle et al. 2017). Population genetic analysis of *mutS* and *uidA* supported the existence of two separate populations. The bovine fecal population has no admixture whereas pond ecosystem has an admixture of two separate populations.

High prevalence of *eaeA* was observed in all four pond niches, but not in feces, indicating presence of *eaeA* may play a role in aquatic fitness that is distinct from virulence. Byappanahalli et al. (2015) detected a high level of *eaeA* in algae and to a lesser extent in water and sand samples from lake Michigan. *eaeA* is one of the most frequently detected pathogen genes in the environment (Hamilton et al. 2010; Chandran and Mazumder 2015; Zhang et al. 2016). It is not certain that these isolates with virulence genes are pathogenic and survive in the environment, or whether they acquire these genes from these environments. The presence of *eaeA*, *Stx2*, *hlyA*, and *STb* indicates the presence of potential pathogens, though it has been suggested that the occurrence of single or multiple virulence genes in *E. coli* does not confirm its pathogenicity, unless it has the appropriate combination of VGs to cause disease to the host. It has been reported that enteric pathogens exposed to vegetables express similar genes those required to the

colonize host intestine, indicating that enteric bacteria may have the ability to colonize vegetables by using similar mechanism required for animal cells (Goudeau et al. 2013). The antibiotic resistance showed different pattern for sample types. Isolates all five sample types showed some resistance to gentamicin, and all samples types but snail showed resistance to ceftriazone. The antibiotic resistance pattern varied among sample types, except water and water plants showed more than 90% similarity. Sediment samples showed about 60% similarity with water and water plant, some isolates in sediment samples showed resistance to meropenem, azithromycin and tetracycline. The similarities of patterns of resistance in different sample types suggested there might be a common source of resistant strains. The resistance pattern in snail and bovine fecal samples were very different from water, water plant and sediment isolates. This pond is not being used for any human or domestic animal activities and also there is no direct in put of any wastewater or farm run-off. It is interesting to find strains with multiple antibiotic resistance in all sample types but snail. Previous studies have shown that *E. coli* isolated from various sampling sources showed variation in the antibiotic resistance patterns depending on the use of antibiotics and their exposure to environments (Sayah et al. 2005; Ibekwe et al. 2011; Amaya et al. 2012). We are interested to understand these strains with multiple antibiotic resistance, whether they have acquired antibiotic resistance from various antibiotic exposure or they have these genes naturally (Vaz-Moreira et al. 2014).

In conclusion, our study showed a distribution pattern of genotypic and phenotypic traits among isolates of various sample sources based on their niche preferences. Here, sediment, water and water plants isolates showed similarities in phylogroup distribution,

occurrence of virulence genes and antibiotic resistance pattern, whereas snail and bovine

fecal isolates were different.

5. References

Amaya, E., Reyes, D., Paniagua, M., Calderon, S., Rashid, M.U., Colque, P., Kuhn, I., Mollby, R., Weintraub, A., and Nord, C.E. (2012). Antibiotic resistance patterns of *Escherichia coli* isolates from different aquatic environmental sources in Leon, Nicaragua. *Clin Microbiol Infect* 18**,** E347-E354.

Anderson, K.L., Whitlock, J.E., and Harwood, V.J. (2005). Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments. *Appl Environ Microbiol* 71**,** 3041-3048.

Anderson, M.A., Whitlock, J.E., and Harwood, V.J. (2006). Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. *Appl Environ Microbiol* 72**,** 6914-6922.

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol Biol Evol* 15**,** 6-16.

Beversdorf, L.J., Bornstein-Forst, S.M., and Mclellan, S.L. (2007). The potential for beach sand to serve as a reservoir for *Escherichia coli* and the physical influences on cell die-off. *J Appl Microbiol* 102**,** 1372-1381.

Byappanahalli, M.N., and Fujioka, R.S. (1998). Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water Sci Technol* 38**,** 171-174.

Byappanahalli, M.N., Nevers, M.B., Whitman, R.L., and Ishii, S. (2015). Application of a Microfluidic Quantitative Polymerase Chain Reaction Technique To Monitor Bacterial Pathogens in Beach Water and Complex Environmental Matrices. *Environ Sci Technol Lett* 2**,** 347-351.

Byappanahalli, M.N., Richard, L.W., Shivelya, D.A., John, F., Ishii, S., and Sadowsky, M.J. (2007). Population structure of cladophora-borne *Escherichia coli* in nearshore water of lake Michigan. *Water Res* 41**,** 3649-3654.

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

Casarez, E.A., Pillai, S.D., and Di Giouanni, G.D. (2007). Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. *Water Res* 41**,** 3643-3648.

Chandran, A., and Mazumder, A. (2013). Prevalence of Diarrhea-Associated Virulence Genes and Genetic Diversity in *Escherichia coli* Isolates from Fecal Material of Various Animal Hosts. *Appl Environ Microbiol* 79**,** 7371-7380.

Chandran, A., and Mazumder, A. (2014). Occurrence of Diarrheagenic Virulence Genes and Genetic Diversity in *Escherichia coli* Isolates from Fecal Material of Various Avian Hosts in British Columbia, Canada. *Applied and Environmental Microbiol* 80**,** 1933-1940.

Chandran, A., and Mazumder, A. (2015). Pathogenic Potential, Genetic Diversity, and Population Structure of *Escherichia* coli Strains Isolated from a Forest-Dominated Watershed (Comox Lake) in British Columbia, Canada. *Appl Environ Microbiol* 81**,** 1779-1789.

Chang, G.W., Brill, J., and Lum, R. (1989). Proportion of beta-D-glucuronidase-negative *Escherichia* coli in human fecal samples. *Appl Environ Microbiol* 55**,** 335-339.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 66**,** 4555-4558.

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

Clsi (2015). "Twenty-Fifth Informational Performance Standards for Antimicrobial Susceptibility Testing Supplement.", in: *CLSI document M100-S25*. (Wayne, PA: Clinical and Laboratory Standards Institute).

Collignon, P., Powers, J.H., Chiller, T.M., Aidara-Kane, A., and Aarestrup, F.M. (2009). World Health Organization ranking of antimicrobials according to their importance in human medicine: A critical step for developing risk management strategies for the use of antimicrobials in food production animals. *Clin Infect Dis* 49**,** 132-141.

Croxen, M.A., and Finlay, B.B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 8**,** 26-38.

Earl, D.A., and Vonholdt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4**,** 359-361.

Erume, J., Berberov, E.A., Kachman, S.D., Scott, M.A., Zhou, Y., Francis, D.H., and Moxley, R.A. (2008). Comparison of the contributions of heat-labile enterotoxin and heat-stable enterotoxin b to the virulence of enterotoxigenic *Escherichia coli* in F4ac receptor-positive young pigs. *Infect Immun* 76**,** 3141-3149.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14**,** 2611-2620.

Fagan, P.K., Hornitzky, M.A., Bettelheim, K.A., and Djordjevic, S.P. (1999). Detection of Shiga-Like Toxin (stx(1) and stx(2)), Intimin (eaeA), and Enterohemorrhagic *Escherichia coli* (EHEC) Hemolysin (EHEC *hlyA*) Genes in Animal Feces by Multiplex PCR. *Appl Environ Microbiol* 65**,** 868-872.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164**,** 1567-1587.

Francis, D.H. (2002). Enterotoxigenic *Escherichia coli* infection in pigs and its diagnosis. *J Swine Health Prod* 10**,** 171-175.

Francis, D.H., and Willgohs, J.A. (1991). Evaluation of a Live Avirulent *Escherichia coli* Vaccine For K88 Positive Lt Positive Enterotoxigenic Colibacillosis In Weaned Pigs. *Am J Vet Res* 52**,** 1051-1055.

Gordon, D.M. (2004). The Influence of Ecological Factors on the Distribution and the Genetic Structure of *Escherichia coli*. *EcoSal Plus* 1.

Goudeau, D.M., Parker, C.T., Zhou, Y.G., Sela, S., Kroupitski, Y., and Brandl, M.T. (2013). The Salmonella Transcriptome in Lettuce and Cilantro Soft Rot Reveals a Niche Overlap with the Animal Host Intestine. *Appl Enviro Microbiol* 79**,** 250-262.

Hamilton, M.J., Hadi, A.Z., Griffith, J.F., Ishii, S., and Sadowsky, M.J. (2010). Large scale analysis of virulence genes in *Escherichia coli* strains isolated from Avalon Bay, CA. *Water Res* 44**,** 5463-5473.

Hammer, Ø., Harper, D.a.T., and Ryan, P.D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Paleontol Electronica* 4**,** 1-9.

Hothorn, T. (2014). "Party, v.1.0-17. Comprehensive R Archive Network". (Vienna, Austria: R Foundation for Statistical Computing).

Hubisz, M.J., Falush, D., Stephens, M., and Pritchard, J.K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9**,** 1322-1332.

Ibekwe, A.M., Murinda, S.E., and Graves, A.K. (2011). Genetic Diversity and Antimicrobial Resistance of *Escherichia coli* from Human and Animal Sources Uncovers Multiple Resistances from Human Sources. *Plos One* 6.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006). Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl Environ Microbiol* 72**,** 612-621.

Ishii, S., Meyer, K.P., and Sadowsky, M.J. (2007). Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of *Escherichia coli* strains from diverse human and animal sources. *Appl Environ Microbiol* 73**,** 5703-5710.

Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23**,** 1801-1806.

Jang, J., Di, D.Y.W., Lee, A., Unno, T., Sadowsky, M.J., and Hur, H.G. (2014). Seasonal and Genotypic Changes in *Escherichia coli* Phylogenetic Groups in the Yeongsan River Basin of South Korea. *Plos One* 9.

Johnson, J.R., and Stell, A.L. (2000). Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* 181**,** 261-272.

Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E., and Tenaillon, O. (2007). Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 24**,** 2373-2384.

Leimbach, A., Hacker, J., and Dobrindt, U. (2013). *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. *Curr Top Microbiol Immunol* 358**,** 3-32.

Letunic, I., and Bork, P. (2011). Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39**,** W475-W478.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 60**,** 708-720.

Mclellan, S.L. (2004). Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. *Appl Environ Microbiol.* 70**,** 4658-4665.

Mclellan, S.L., Daniels, A.D., and Salmore, A.K. (2003). Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. *Appl Environ Microbiol* 69**,** 2587-2594.

Meric, G., Kemsley, E.K., Falush, D., Saggers, E.J., and Lucchini, S. (2013). Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environ Microbiol* 15**,** 487-501.

Nandakafle, G., Seale, T., Flint, T., , Nepal, M., Venter, S.N., and Brözel, V.S. (2017). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. . *Microbes environ* ME17030.

Nguyen, Y., and Sperandio, V. (2012). Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front Cell Infect Microbiol* 2.

Osek, J. (2001). Multiplex polymerase chain reaction assay for identification of enterotoxigenic *Escherichia coli* strains. *J Vet Diagn Invest* 13**,** 308-311.

Petit, F., Clermont, O., Delannoy, S., Servais, P., Gourmelon, M., Fach, P., Oberle, K., Fournier, M., Denamur, E., and Berthe, T. (2017). Change in the Structure of *Escherichia coli* Population and the Pattern of Virulence Genes along a Rural Aquatic Continuum. *Front Microbiol* 8.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67**,** 546-553.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155**,** 945-959.

Qadri, F., Svennerholm, A.M., Faruque, A.S.G., and Sack, R.B. (2005). Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clin Microbiol Rev* 18**,** 465-483.

R Development Core Team (2015). "R: A language and environment for statistical computing". (Vienna, Austria: R Foundation for Statistical Computing).

Rambaut, A. (2002). SE-AL v. 2.0a11: sequence alignment program. (http://tree.bio.ed.ac.uk/software/seal/).

Ratajczak, M., Laroche, E., Berthe, T., Clermont, O., Pawlak, B., Denamur, E., and Petit, F. (2010). Influence of hydrological conditions on the *Escherichia coli* population structure in the water of a creek on a rural watershed. *BMC Microbiol* 10**,** 222.

Rauch, E.M., and Bar-Yam, Y. (2004). Theory predicts the uneven distribution of genetic diversity within species. *Nature* 431**,** 449-452.

Rosenberg, N.A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4**,** 137-138.

Sayah, R.S., Kaneene, J.B., Johnson, Y., and Miller, R. (2005). Patterns of antimicrobial resistance observed in *Escherichia coli* isolates obtained from domestic- and wild-animal fecal samples, human septage, and surface water. *Appl Environ Microbiol* 71**,** 1394-1404.

Smith, R.D., and Coast, J. (2002). Antimicrobial resistance: a global response. *Bull World Health Organ* 80**,** 126-133.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., and Bouvet, O. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5.

Unno, T., Han, D., Jang, J., Lee, S.N., Ko, G., Choi, H.Y., Kim, J.H., Sadowsky, M.J., and Hur, H.G. (2009). Absence of *Escherichia coli* phylogenetic group B2 strains in humans and domesticated animals from Jeonnam Province, Republic of Korea. *Appl Environ Microbiol* 75**,** 5659-5666.

Vaz-Moreira, I., Nunes, O.C., and Manaia, C.M. (2014). Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *Fems Microbiol Rev* 38**,** 761-778.

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

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., and Whittam., T.S. (2009). Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* 75.

White, A.P., Sibley, K.A., Sibley, C.D., Wasmuth, J.D., Schaefer, R., Surette, M.G., Edge, T.A., and Neumann, N.F. (2011). Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. *Appl. Environ. Microbiol.* 77**,** 7620-7632.

Winfield, M.D., and Groisman, E.A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl Environ Microbiol* 69**,** 3687-3694.

Zhang, Q., Eichmiller, J.J., Staley, C., Sadowsky, M.J., and Ishii, S. (2016). Correlations between pathogen concentration and fecal indicator marker genes in beach environments. *Sci Total Environ* 573**,** 826-830.

Table S3. 1: Primers used for determining the *uidA* and *mutS* genes, and for phylogrouping.

Primer	Sequence (5' – 3')	Size	Reference
uidA	CATTACGGCAAAGTGTGGGTCAAT (F) TCAGCGTAAGGGTAATGCGAGGTA (R)	658 bp*	(Walk et al. 2009)
mutS	GGCCTATACCCTGAACTACA (F) GCATAAAGGCAATGGTGTC (R)	596 bp	(Walk et al. 2009)
chuA	ATGGTACCGGACGAACCAAC (F) TGCCGCCAGTACCAAAGACA (R)	288 bp	(Clermont et al. 2013)
yjaA	CAAACGTGAAGTGTCAGGAG (F) AATGCGTTCCTCAACCTGTG (R)	211 bp	(Clermont et al. 2013)
TspE4	CACTATTCGTAAGGTCATCC (F) AGTTTATCGCTGCGGGTCGC (R)	152 bp	(Clermont et al. 2013)
arpA	AACGCTATTCGCCAGCTTGC (F) TCTCCCCATACCGTACGCTA (R)	400 bp	(Clermont et al. 2013)
Group E ArpAgpE	GATTCCATCTTGTCAAAATATGCC (F) GAAAAGAAAAAGAATTCCCAAGAG (R)	301 bp	(Clermont et al. 2013)
Group C rpAgpC	AGTTTTATGCCCAGTGCGAG (F) TCTGCGCCGGTCACGCCC (R)	219 bp	(Clermont et al. 2013)

* bp = base pair

Table S3. 2: Primers used for amplification of virulence genes.

Primers	Sequence (5' – 3')	Size (bp)	Reference
Stx1	Forward ACACTGGATGATCTCAGTGG Reverse CTGAATCCCCCTCCATTATG	614	Fagan et al.(1999)
Stx2	Forward CCATGACAACGGACAGCAGTT Reverse CCTGTCAACTGAGCAGCACTTTG	779	Fagan et al. (1999)
eaeA	Forward GTGGCGAATACTGGCGAGACT Reverse CCCCATTCTTTTTCACCGTCG	890	Fagan et al. (1999)
hlyA	Forward ACGATGTGGTTTATTCTGGA Reverse CTTCACGTGACCATACATAT	165	Fagan et al. (1999)
STa	Forward GCCTATGCATCTACACAATC Reverse TGAGAAATGGACAATGTCCG	278	Osek (2001)
LTb	Forward TATCCTCTCTATATGCACAG Reverse CTGTAGTGGAAGCTGTTATA	480	Osek (2001)

* bp = base pair

Chapter 4: Growth and extended survival of *Escherichia coli* O157:H7 in soil organic matter

1. Introduction

Escherichia coli O157:H7 and related enterohaemorrhagic strains have been associated with many serious food-associated outbreaks, (Hilborn et al. 1999; Currie et al. 2007; Grant et al. 2008; King et al. 2009). The infectious dose is low so that food products are required to be free from enterohaemorrhagic *Escherichia coli* O157:H7, but despite various measures taken during processing, consumers can still be exposed to this pathogen (LeBlanc 2003; Yang et al. 2017). Cattle are widely believed to be the primary host and several outbreaks have been associated with beef–based products (Currie et al. 2007; King et al. 2009). *E. coli* O157:H7 is known to be associated with the bovine gastrointestinal tract, specifically the cecum (Yoon and Hovde 2008; Wang et al. 2017), currently believed to be the primary source of entry into the food chain. More recently various plant foods such as spinach, tomato, lettuce and fresh fruits have been identified as sources (Grant et al. 2008; Herman et al. 2015; Denis et al. 2016). Initially these foods were thought to be fecally contaminated, but recent reports suggest growth of *E. coli* O157:H7 (Brandl 2008; Wright et al. 2013; Wright et al. 2017) in tissues of salad leaves and tomatoes. Upon inoculation from an unknown source, the enteric bacteria multiply

inside the growing plant, and cannot be removed through surface treatment such as washing. The annual nature of these crop plants excludes them as an environmental reservoir of these enteric bacteria. Rather, these crop plants would need to be infected during growth.

E. coli are found in both gastrointestinal systems, and in the environment (Adamowicz et al. 1991; Ishii et al. 2006; Ksoll et al. 2007). Once shed from a mammalian host, *E. coli* populations are widely believed to enter a dead end, relying for extended survival on stress responses (Winfield and Groisman 2003). The paradigm assumes slow decline following fecal contamination, the basis of the fecal coliform test (Tallon et al. 2005). This is supported by decline of *E. coli* O157:H7 in manure (Williams et al. 2008; Looper et al. 2009) and in soil (Berry and Miller 2005) over time. Yet enterohaemorrhagic *E. coli* maintain culturable populations in various soils for many months, even moisture is limited, and with slower decline at lower temperatures (Berry and Miller 2005; Fremaux et al. 2008). Some *E. coli* appear to grow in sub-tropical environments such as riverbank soil and river sediment (Desmarais et al. 2002). More recently persistent *E. coli* populations have been reported from temperate forest, watershed soils (Byappanahalli et al. 2006; Ishii et al. 2006), and pasture (NandaKafle et al. 2017). Naturalized *E. coli* strains believed to be autochthonous to soil were able to maintain populations in soils from Lake Superior shore.

Persistence of bacterial populations in soil would require a suitable nutrient pool. Soil is a complex assemblage of particulate components with varying concentrations of organic and inorganic matter. The dissolved organic matter (DOM) in soils is a cocktail of sugars, aromatic compounds, amino acids, and organic and fatty acids between C_{14} and C_{54}

(Huang et al. 1998; Kalbitz et al. 2000). The concentrations of solutes like amino acids range from 0.1-5 µM. Monoprotic acids (e.g. formate, acetate and lactate) range from 1 µM to 1000 µM, and di- and trivalent low molecular organic acids (e.g. oxalate, malate and citrate) from 0.1-50 µM (Strobel 2001; Pizzeghello et al. 2006). Monomeric intermediates such as carboxylic acids and amino acids have residence times in the order of hours in soils (Jones et al. 2005; Van Hees et al. 2005). Carbohydrates like mono-, di- and oligosaccharides vary in presence and concentration (Lynch 1982; Guggenberger and Zech 1993b; a; Kaiser et al. 2001; Kalbitz et al. 2003). Surprisingly, glucose is present in soils up to 100 µM concentrations (Schneckenberger et al. 2008). The variety of sugars, organic and amino acids in these soils suggest that enteric bacteria should, generally, be able to grow here. We have reported a detailed analysis of liquid extract of deciduous forest soil, able to support growth of *Salmonella* Typhimurium (Liebeke et al. 2009).

The source of contamination of annual food crops by enterrohaemorrhagic *E. coli* is unresolved. Soil has been underestimated as a potential reservoir. As persistence of bacterial populations in open systems is the product of growth, death, predation, and competition, measurement of numbers over time shows the overall net effect, and cannot inform autecology of the species. Whether population maintenance of *E. coli* O157:H7 in soils is due to a combination of cell division and death, predation and competition, or simply to extended survival alone, is unresolved. It has been shown that *E. coli* O157:H7 is able to grow in sterile fresh water (Vital et al. 2008). In order to understand how soil is a potential reservoir for enterrohaemorrhagic *E. coli*, autecological studies are required. Here we report that *E. coli* O157:H7 is able to grow in liquid extract of soil. Furthermore, soil extract-grown populations demonstrate extended culturability over cultures grown in

laboratory media, and display a unique stationary phase proteome.

2. Materials and Methods

2.1. Culture and culture media:

The partially attenuated *E. coli* O157:H7 933D (*stx*-II) (Strockbine et al. 1986) was

maintained in 50% (v/v) glycerol at -80ºC. Soils used were corn field soil (Brandt silty

clay loam, Aurora, South Dakota, USA), a commercially available garden top soil, and

deciduous forest soil (Oak Lake Field Station, Brookings, South Dakota, USA). Cow

manure from a herd fed an antibiotic-free diet was obtained from the South Dakota State

University Beef Unit. Soil-extracted solubilized organic matter (SESOM) was prepared

as described previously (Vilain et al. 2006). Briefly, 100g of air-dried soil, or 90g soil

and 10g manure, was suspended in 500 mL MOPS buffer (10 mM, pH 7, 50ºC) and kept

shaking at 200 rpm for 1h. The extracts were filtered sequentially through filter paper,

hydrophilic PVDF membranes with 5, 1.2, and 0.45, pore sizes to remove particulates,

and sterilized using a polyethersulfone membrane with a 0.22 µm pore size. The sterility

of each batch was determined by placing 5µl SESOM onto LB agar plates and incubating

at 30˚C for 24h.

2.2. Culturing conditions:

Growth and survival in the various liquid extracts was determined by measuring

the optical density periodically. Overnight cultures of *E. coli* O157:H7 were prepared in

LB broth, diluted 1:1,000 into 50mL fresh LB broth and incubated to mid-exponential

phase at 28°C (3h, A_{546} = 0.41). Cells were harvested by centrifugation (10,000 x g, 10 min, 30 °C), washed twice, and re-suspended in 2mL sterile tap water. Triplicate 250mL flasks with 50mL pre-warmed liquid medium (LB, 1/40th strength LB and SESOM from deciduous forest soil) were inoculated to an initial A_{546} of 0.005, and incubated at 28°C while shaking (120 rpm). The culturable count was determined every hour till 8h, at 24h and daily till 24d by the droplet plate technique (Lindsay and Von Holy 1999). Briefly, 20 µL volumes of serial dilutions were plated onto LB agar and incubated for 18h at 30°C. Culturable counts reflected the average of nine droplet counts, with three droplet counts from each of three replicate cultures. The homogeneity of variance was checked at 1d, and then every four days till 24d. There was no reason to reject the null hypothesis, meaning that homogeneity of variance of CFU of three media was equal. Thus, the data were subjected to an ANOVA test with multiple LSD comparison using Statistix 9.0.5 (Informer Technologies, Inc.).

2.3. *Effect of cell density on culturability:*

The effect of cell density on extended culturability of populations was investigated by concentrating or diluting populations, and re-suspending in cell-free supernatants of the same culture type. LB-grown populations were harvested at 24h of incubation (10,000 x g, 10 min, 30 °C) and re-suspended to one tenth their density in cell-free supernatant. Conversely, 1/40th strength LB and SESOM-grown populations were re-suspended to ten-fold density in their respective cell-free supernatants. All cultures were then incubated at 28 °C while shaking, and the culturable count determined every 24h to day 24. Culturable counts reflected the average of nine droplet counts, with three droplet

counts from each of three replicate cultures. Statistical analysis was performed as described above.

2.4. Protein sample preparation:

SESOM and LB − grown populations of *E. coli* O157:H7 were harvested in mid-exponential phase (180min and 140min) at A_{546} 0.055 and 0.183, respectively, and SESOM, LB, and 1/40th strength LB populations were harvested in late stationary phase (3d). Cells were harvested by centrifugation (10,000 x g, 10 min, 4ºC), washed in 5mL potassium phosphate buffer (100 mM, pH 7.0), and re-suspended in 2ml IEF buffer (7 M urea, 2 M thiourea, 2% (w/v) 3-[3-chloamidopropyl] dimethylammonio-1-propanesulfonate (CHAPS), 2% (w/v) Amidosulfobetaine-14 (ASB14), 10mM dithiothreitol (DTT) and 2% (v/v) carrier ampholytes (pH 3.5 - 10; Amersham)). Cells were disrupted by two cycles of freeze thaw (from -80°C to 20°C) followed by ultrasonication at 4°C (15W, 12 pulses of 3min). Cell debris was removed by centrifugation (10,000 x g, 10min), and the protein concentration was determined using the Bradford protein assay (BioRAD), with bovine serum albumin as the standard (Vilain et al. 2001).

2.5. Two-dimensional gel electrophoresis (2DE):

IPG strips (pH 4 − 7, 18 cm, GE Healthcare) were re-hydrated for 16h with 400 μL IEF buffer containing 50 μg protein for 2D gel map construction, and 200 μg protein per IPG strip for protein identification. Proteins were separated by IEF on an Amersham

Pharmacia horizontal electrophoresis system for a total of 44 kVh (150 V for 1 h, 350 V for 1 h, 500 V for 4 h, 750 V for 1h, 1 kV for 1 h, 1.5 kV for 1 h and 3.5 kV for 11 h). After IEF, the IPG gel strips were frozen at -80 °C, thawed and equilibrated for 10 min in equilibration buffer (6 M urea and 30% glycerol, 1% SDS) with 20mg/mL DTT, and for 10 min in equilibration buffer with 260 mM iodoacetamide. The second dimension consisted of SDS-PAGE using a 12.5% (w/v) running polyacrylamide gel and a 4.65% stacking gel (width, 18 cm; length, 20 cm; thickness, 1 mm). Gels were stained with silver (Rabilloud 1992) for spot detection and protein map construction, and with colloidal Coomassie Blue G250 for protein identification (Vilain et al. 2001). Uninoculated SESOM was run on a one-dimensional SDS PAGE to check for proteins present, but following staining, none were found.

2.6. *Gel analysis, spot detection and protein map construction:*

Gels were scanned using a transmission scanner (ScanMaker 9800XL, Microtek) in transmission mode. Gel images were analyzed using PDQuest software (version 7.3.1; Bio-Rad) which allows detection, quantification and matching of protein spots. Spots were quantified on a Gaussian image and pooled on a reference image. The following formula was used to calculate the quantity of Gaussian spot: Spot height $\times \sigma_x \times \sigma_y \times \pi$; where: Spot height is the peak of the Gaussian representation of the spot, σ_x is the standard deviation of the Gaussian distribution of the spot in the direction of the x axis, and σ_y is the standard deviation in the direction of the y axis. SESOM-derived spots either higher than two-fold or less than half the intensity in LB broth were excised from stationary phase LB, LB 1/40, and SESOM derived gels, and identified by MALDI-TOF

mass spectrometry of tryptic digests as described previously (Voigt et al. 2006), but using the *E. coli* O157:H7 EDL933 sequence database (ftp://www.expasy.org/databases/complete_proteomes/fasta).

Principal component analysis (PCA) of the 2D electrophoretograms was performed as described previously (Vilain and Brozel 2006), using Statgraphics Plus 4.0 (Manugistics). Briefly, calculations of the Eigen value were comprised by taking the data set and subtracting the mean value from each dimension (ie. effect of culture medium) until all means were zero. A covariance matrix was then calculated since the data set has more than one dimension. By calculating the covariance matrix on means of zero a line develops that characterizes the data. The lines, or Eigen values, determine the statistical significance of each of the components.

3. Results

The enterohaemorrhagic pathogen *E. coli* O157:H7 was able to grow using water-soluble organic matter from various soils, as indicated by increases in optical density during incubation (Fig. 1). The yield in SESOM was 1 Log lower than in LB broth, and varied among extracts of various soils. To determine whether cell density during entry into stationary phase affects future culturability, we sought to culture in LB to the same population density achieved in SESOM. Various dilutions of LB (1/30, 1/40, 1/50, 1/70, and 1/100) were evaluated to determine which supported a final optical density similar to SESOM (results not shown). LB diluted 40 times yielded the desired density, and the resulting populations remained culturable longer than in LB, with a stable population of 10^8 CFU/mL at d9 (Fig. 2a), after which culturability declined.
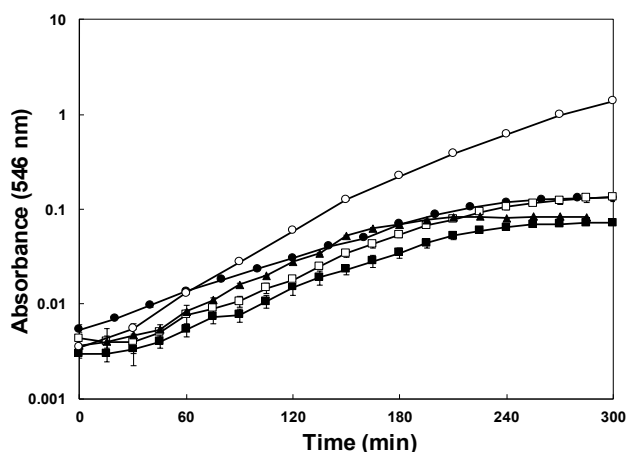
Fig. 4. 1 Growth of *E. coli* 0157:H7 933D stxII- in SESOM from deciduous forest soil (●), corn field soil (■), corn field soil supplemented with 10% (m/v) cow manure (□), garden soil (▲), and LB broth (○) while shaking at 30°C.
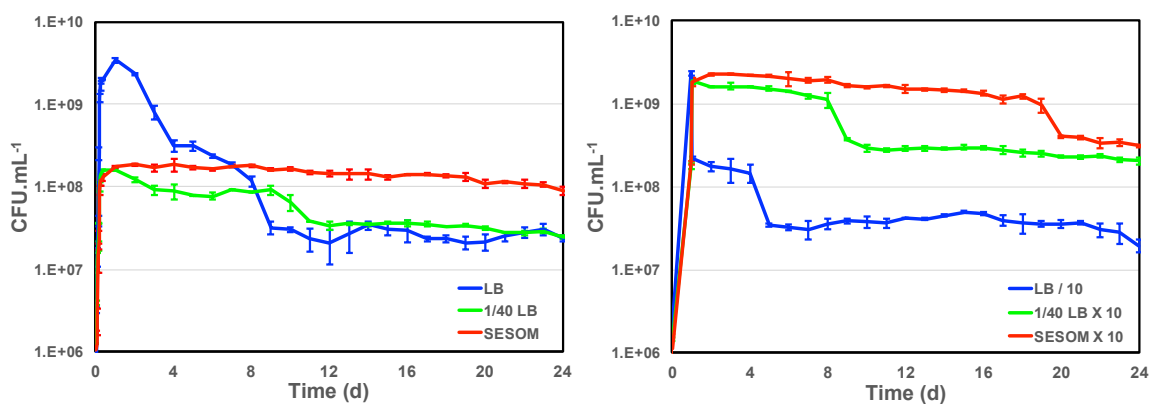
Fig. 4. 2 Growth and survival of *E. coli* O157:H7 933D in LB, dilute LB (1/40) and SESOM (a), and when cultures were either concentrated ten-fold in own supernatant (SESOM and 1/40th LB - grown), or diluted ten-fold (LB - grown) at 24h (b). Error bars indicate one standard error of the mean.

This indicated that population density in stationary phase may play a role in maintenance of culturability of cells, with higher density associated with decreased survival. The pH on d12 was 6.6. Cells that entered stationary phase due to nutrient limitation in $1/40^{th}$ strength LB were more resilient than populations grown to higher density in LB. SESOM grown cells were, however, more resilient than $1/40^{th}$ LB grown cells, although both entered stationary phase due to nutrient limitation (Fig. 2a). This indicated that soil grown *E. coli* populations would persist longer in soil than predicted by laboratory experiments. Cultures in M9 minimal medium with 10 $g.L^{-1}$ glucose displayed loss in culturability over time, similar to in LB (data not shown), indicating that increased longevity could not be attributed to growth requiring a greater degree of anabolic reactions.

Cell density appeared to play a role in stationary phase survival of LB-grown populations (Fig. 2a). To further investigate the role of cell density in survival, we modified cell density 10-fold upon entry into stationary phase. LB-grown stationary phase cells (24h) were harvested and resuspended to one tenth their original density in their own spent broth, and the culturable count determined for 24d. The population lost one log_{10} of culturability after d4 (Fig. 2b), as opposed to 2 log_{10} in undiluted culture (Fig. 2a). This could indicate that LB-grown *E. coli* are able to maintain only a certain cell density into stationary phase. To determine whether the resilience of populations grown in 1/40 strength LB was due to lower final density, stationary phase populations were concentrated ten-fold and resuspended in their own supernatant. The increased cell density did not initially lead to much loss of culturability, similar to the un-concentrated culture (Fig. 2 a & b). After decline at d8, late stationary phase population density

remained at ten-fold that of the original 1/40 th LB grown culture. Importantly, the concentrated 1/40th LB population was at the same density as LB-grown population entering stationary phase, but did not undergo the 2 $\log_{10}$ decline. Thus 1/40th LB – grown cells were more likely to survive than LB-grown ones, irrespective of cell density post-stationary phase. These results indicated that conditions upon entry into stationary phase affect the condition of the cells, thereby determining their potential for survival over long term incubation.

SESOM-grown populations maintained at ten-fold concentration in their own spent medium declined slowly, only showing a five-fold decline at d19 (Fig. 2b). Thus SESOM – grown cells were more likely to survive than 1/40th and LB-grown ones, irrespective of cell density post-stationary phase. Collectively the results indicated that extended longevity of SESOM-grown populations was due to both a lower cell density, but also to a SESOM-associated factor. These results suggested that SESOM-grown populations had an altered physiological state when entering stationary phase.

To gain insight into possible physiological reasons underlying the extended longevity of SESOM-grown populations, the proteomes of LB and SESOM-grown cultures in exponential and stationary phase (d3), were determined by 2DE, and compared to the 1/40th LB stationary phase proteome. For exponential phase, care was taken that populations had not yet begun transition to stationary phase. The five proteomic datasets were then subjected to principle component analysis (PCA). Four components were revealed at an Eigen value greater than 1, *viz*. 2.69, 2.51, 1.53 and 1.20. These components were sequentially compared in pair-wise fashion using biplots (Fig. 3). The results showed that exponential phase LB- and SESOM grown populations differed

significantly, as did stationary phase populations in the two media. Intriguingly, the LB 1/40th stationary proteome was very similar to the SESOM- proteome in the first three of four coordinates, and quite different to the LB stationary phase proteome.
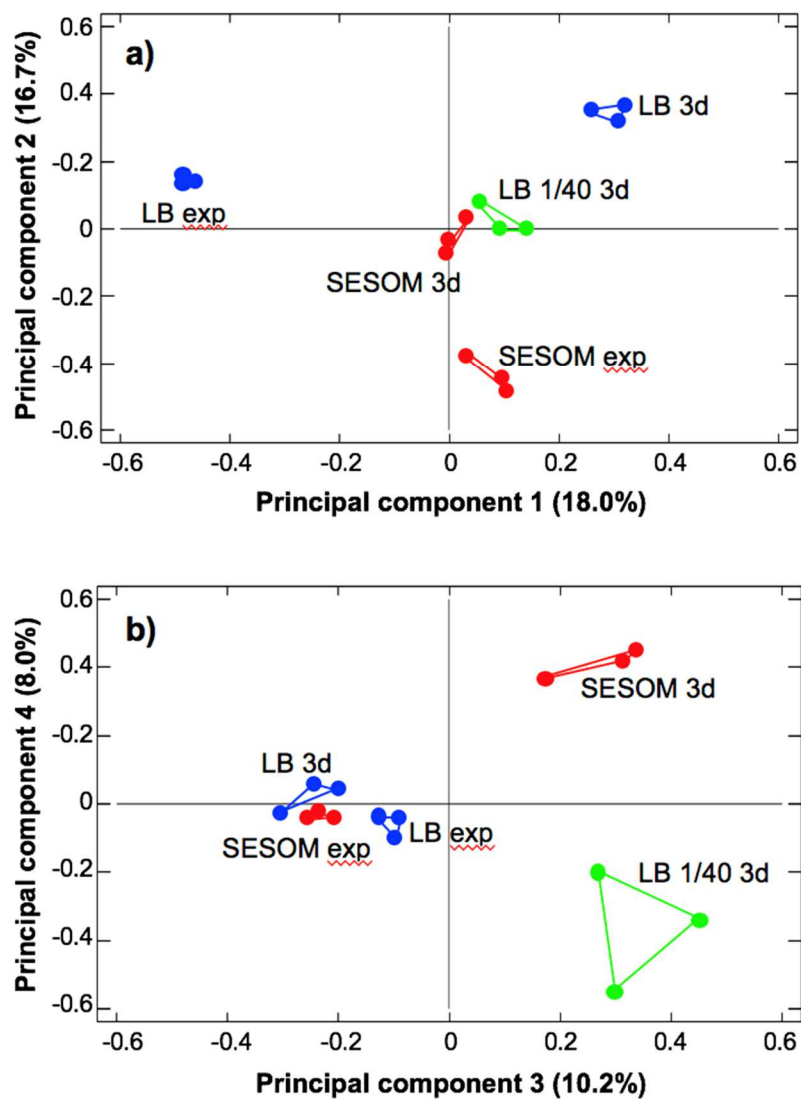
Fig. 4. 3 Principle component analysis of exponential (exp) and stationary phase (3d) proteomes of *E. coli* O157:H7 933D cultured in LB, 1/40strength LB and SESOM at 30°C. Four components with Eigen values >1 were revealed, shown as principle components 1 and 2 (A) and 3 and 4 (B).

Collectively the PCA analysis showed that stationary phase populations had culture medium-specific proteomes that could explain the different physiological states and propensity to survive.

A large number of protein spots had significantly different abundance as determined using the criteria outlined above in materials and methods. All these spots were identified by MALDI-TOF MS, and collectively paint a unique physiological state of *E. coli* O157:H7 persisting in soil organic matter (Table 1). Stationary phase LB populations appeared to experience several stresses as indicated by elevated levels of the universal stress protein UspA and the carbon starvation protein Slp. They also had elevated levels of the alkyl hydroperoxide reductase AhpC. By contrast SESOM-grown cells appeared less stressed and more active, indicated by increased levels in transcriptional (DksA and RpoA) and translational proteins (GroEL, TufA and YeiP). This suggested sustained transcriptional and translational activity during stationary phase in SESOM versus LB. Many uptake systems were either over or under-expressed in SESOM-grown populations, including outer membrane and periplasmic uptake systems. This indicates that cells growing on SESOM have the ability to sense what the surrounding environment has to offer. These proteins reinforce the notion that *E. coli* O157:H7 is very adaptable to non-host environments such as soil. In addition to various uptake systems, several systems involving substrate metabolism were found to be over and under-expressed in SESOM-grown cells, indicating different approaches to catabolic activity in stationary phase.

Structurally, cells grown in SESOM appear to be different based on the expression of several membrane and cell structure proteins, primarily those involved in membrane lipid

biosynthesis. YmcD and Adk were both up-regulated in SESOM. This suggests that the

cellular envelope is formed differently in SESOM-grown populations as opposed to LB-

or LB 1/40-grown populations. Perhaps the cellular envelope is thicker to provide

protection from adverse conditions. Both structural and regulatory flagellar proteins were

present in increased abundance in SESOM-grown cells suggesting that the cells are

potentially motile and responsive to chemotactic behavior in soil organic matter. Overall,

**Table 4. 1. Proteins of different abundance in stationary phase (3d) populations of
E. coli O157:H7 grown and maintained in LB, 1/40-strength LB and SESOM at 30
°C. Blue color-less expressed, white color- intermediate and red color- highly
expressed**

Protein name	Function	Relative amount[1]		
		LB	**LB 1/40**	**SESOM**
Stress responses				
AhpC	Alkyl hydroperoxide reductase	12154	6044	3483
OsmY	Hyperosmotically inducible periplasmic protein, RpoS-inducible	1668	11249	3075
Slp	C starvation and stationary phase inducible outer membrane lipoprotein	4807	1755	505
UspA	Universal stress protein	14202	4034	1136
Motility				
FliC	Flagellin filament structural protein	21614	4395	122538
FliY	Cystine-binding protein; not required for motility; may regulate FliA (sigma F)	3680	9158	14844
Outer membrane proteins				
OmpA	Outer membrane protein 3a	12108	16971	30333
OmpC	Outer membrane protein 1b	22919	29601	662
OmpW	OmpW functions as an ion channel in planar lipid bilayers, global iron-dependent gene regulation in Escherichia coli	58638	6649	1190
Membrane and wall functions				
Adk	Pleiotropic effects on glycerol-3-phosphate acyltransferase activity - plays a role in phospholipid biosynthesis	267	311	1810
ClsC	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases	1376	2791	15315
MipA	Mediates assembly of MltA to PBP1B into a complex. MltA is Lipoprotein lytic transglycosylase; membrane-bound murein hydrolase, affecting sacculus maturation	3590	219	552
NemA	N-Ethylmaleimide reductase - induced by menadione, dimethyl maleate and linoleic acid, possibly due to lipid peroxidation	7294	8193	576
Pbl	Lytic Transglycosylase family, catalyze the cleavage of the beta-1,4-glycosidic bond between N-acetylmuramic acid (MurNAc) and N-acetyl-D-glucosamine (GlcNAc).	8967	9143	1884
Periplasmic uptake systems				
AnsB	Periplasmic L-asparaginase II	13285	2099	1726
ArgT	Lysine-, arginine-, ornithine-binding periplasmic protein	4800	14535	30498

ArtJ	Arginine 3rd transport system periplasmic binding protein	2830	403	334
GlnH	Periplasmic glutamine-binding protein; permease	3131	419	678
HisJ	histidine-binding periplasmic protein of high-affinity histidine transport system	1055	2117	6397
MalE	Periplasmic maltose-binding protein	7356	8303	33319
ManX	PTS enzyme IIAB, mannose-specific	9091	2549	10576
ModA	Molybdate-binding periplasmic protein; permease	3109	684	5682
PtsI	PEP-protein phosphotransferase system enzyme I	3194	635	386
TbpA	Thiamine-binding periplasmic protein	6422	4762	1045
Transcription				
DksA	Involved in control of transcription initiation	480	2241	3971
RpoA	RNA polymerase, alpha subunit	69	605	495
Translation and protein processing				
ClpP	ATP-dependent proteolytic subunit of clpA-clpP serine protease	3499	181	667
GroL	GroEL, chaperone Hsp60, peptide-dependent ATPase, heat shock protein	1009	321	9468
HtpG	Chaperone Hsp90, heat shock protein	5681	1028	1316
PheS	Phenylalanine tRNA synthetase, alpha-subunit	1748	1144	125
RpsA	30S ribosomal subunit protein S1	7350	412	565
Tsf	Protein chain elongation factor EF	43187	50227	17501
TufA	TufA - duplicate gene for EF-Tu subunit; elongation factor, unstable	1872	2890	11999
YbdQ	Universal stress protein G enhances cell survival during prolonged stress	8891	1620	1024
YedU	Chaperone protein HchA - Type 1 glutamine amidotransferase	4852	4958	982
YeiP	Putative translation elongation factor	1104	1251	13845
Central metabolism				
AckA	Acetate kinase, acetate to acetyl phosphate: in acetate utilization	2402	2247	260
AldA	Aldehyde dehydrogenase, NAD-linked	6097	1557	1903
Eno	Enolase	1810	419	332
FrdA	Fumarate reductase, anaerobic, flavoprotein subunit	1309	756	446
Lcd	Isocitrate dehydrogenase, specific for NADP+	9996	287	433
LpcA	Phosphoheptose isomerase	1144	126	595
Lpd	Subunit of various ezymes: dihydrolipoate dehydrogenase, 2 oxoglutare dehydrogenase and pyruvate dehydrogenase	4928	5364	1148
MaeB	Putative NADP+-linked malic enzyme	1878	2983	438
Mdh	Malate dehydrogenase	863	1346	5757
PckA	Phosphoenolpyruvate carboxykinase	1994	2875	749
Pgk	Phosphoglycerate kinase	272	424	9162
PpsA	Phosphoenolpyruvate synthase	2972	1017	620
PrpR	Propionate catabolism operon	2330	440	734
SdhA	Succinate dehydrogenase, flavoprotein subunit	5316	7050	1387
TtdA	L-tartrate dehydratase, subunit A	2071	147	584
YbhE	Putative, 6-phosphogluconolactonase, or also 3-carboxymuconate cyclase	537	910	4429
YfiD	Putative formate acetyltransferase	3946	590	748
ATPase function				
AtpA	Membrane-bound ATP synthase, F1 sector, alpha-subunit	31963	31551	3123
AtpH	Membrane-bound ATP synthase, F1 sector, delta-subunit	5737	207	255
Ppa	Inorganic pyrophosphatase - hydrolyzes diphosphate to 2 Pi	11687	677	2019
Amino acid biosynthesis				
AspA	Aspartate ammonia-lyase (aspartase)	17030	3517	2553
CysK	Cysteine synthase A, O-acetylserine sulfhydrolase A	1812	5513	4567
FklB	Peptidyl-prolyl cis-trans isomerase	9539	635	1158
GcvT	Aminomethyltransferase (tetrahydrofolate-dependent) of glycine cleavage system	4291	828	773
GdhA	NADP-specific glutamate dehydrogenase	2287	1747	314
GlyA	Serine hydroxymethyltransferase – glycine synthesis	2051	5348	3469
SerC	3-phosphoserine aminotransferase – serine biosynthesis	11626	2421	4729

TnaA	Tryptophanase	18108	5384	1077
WrbA	Affects association between Trp repressor and operator in stationary phase	3109	345	226
Virulence factors, toxins and resistance				
AcrA	Acridine efflux pump, related to MAR system	7322	214	1496
Hha	Modulates expression of haemolysin genes *hly*	1328	3285	4941
PmbA	Antibiotic peptide MccB17	10512	29643	25983
TerZ	Putative phage inhibition, colicin resistance and tellurite resistance protein	4970	119	962
TolB	Periplasmic protein involved in the tonB-independent uptake of group A colicins	4304	927	363
FolA	Dihydrofolate reductase type I; trimethoprim resistance	187	482	1461
Miscellaneous				
CcmH	Required for synthesis of c-type cytochromes	71	479	756
NrdH	Glutaredoxin-like protein involved in electron transport system for ribonucleotide reductase system NrdEF	5360	492	325
RibB	3,4 dihydroxy-2-butanone-4-phosphate synthase – riboflavin biosynthesis	544	2255	4423
CchA	Putative acetyl/butyryl P transferase	124	82	1069
NohB	Putative DNA packaging protein of prophage CP-933R	2689	528	708
Unknown function				
YbiM	Unknown Hypothetical protein	1131	228	2423
YdcL	Predicted lipoprotein	857	3147	171
YidQ	Putative periplasmic lipoprotein	5199	769	398

1. The relative amount is the average normalized amount of protein per spot across three separate gels.

SESOM-grown stationary phase cells appeared less stressed, more motile, metabolically different, and with suggestions of less altered membrane composition when compared to LB-grown populations.

4. Discussion

E. coli is not thought to survive for long periods outside the host intestine, so produce-associated outbreaks have widely been ascribed to recent fecal contamination. The suspected sources of produce contamination include soil amendments (manure or compost), irrigation water contaminated with cattle feces, or contaminated surface runoff (Ongeng et al. 2015). Our results showed that *E. coli* O157 can grow using nutrients available in soils (Fig. 1). There have been countless studies reporting numbers of *E. coli* O157 in soils over time, and some have suggested growth in soil. Survival of *E. coli* in

soil has been reported by many researchers; more than 200 d under natural environmental conditions and 500 d in frozen soil and on plant roots (Gagliardi and Karns 2002; Islam et al. 2004). This is the first report showing definitively that *E. coli* O157 is able to grow using water soluble nutrients in soil.

Soil-grown *E. coli* O157 appeared more resilient than laboratory-grown cultures, with almost 100% culturability maintained over 28 d (Fig. 2). This finding pointed to an altered physiological state of SESOM-grown cells entering stationary phase. This suggests that *E. coli* responds differently to nutrient limitation in SESOM, preparing cells for stationary phase differently. To gain insight into possible physiological reasons underlying the extended longevity of SESOM-grown populations, the proteomes of LB and SESOM-grown stationary phase cultures were compared (Table 1). The stationary phase proteome of SESOM-grown *E. coli* differed significantly from LB-grown and dilute LB-grown populations (Fig. 3), indicating cells with substantially altered composition, and therefore catalytic and structural properties.

SESOM grown cells had lower levels of several proteins associated with cellular responses to stress, including Alkyl hydroperoxide reductase (AhpC), the carbon starvation response lipoprotein Slp, and the universal stress protein UspA (Table 1). AhpC is the primary degrader of hydrogen peroxide and reactive nitrogen intermediates in *E. coli*, protecting the cell against oxidative stresses (Chen et al. 1998). The substantially lower concentration of AhpC in SESOM-grown cells indicated decreased oxidative stress, or possible alternative mechanisms to cope with reactive oxygen and nitrogen species. Slp accumulates in response to carbon starvation (Alexander and St John 1994), but our data showed that LB grown cells expressed the most Slp, although

SESOM-grown cells were clearly nutrient starved following entry in stationary phase (Liebeke et al. 2009). Clearly, SESOM-grown cells responded differently to nutrient starvation and entry in stationary phase. UspA is induced as soon as the growth rate falls below the maximum rate supported by the medium (Nystrom and Neidhardt 1994). Despite the abrupt transition from exponential to stationary phase in SESOM, cells expressed less UspA than in LB. SESOM populations contained a much greater amount of the flagellar components FliC and FliY, indicating increased motility.

OmpA, the major outer membrane protein in *E. coli*, was more prevalent in the SESOM population. Loss of OmpC in *E. coli* contributes to antibiotic resistance (Liu et al. 2012), but this is only significant in the exponential-phase, while such difference in stress-resistance becomes trivial after bacteria reach the stationary phase (Wang 2002). The elevated OmpA in LB populations is likely due to the high NaCl concentration. The ratio of OmpC to OmpF increases at higher temperature and pH, as well as under oxidative stress (Snyder L et al 2013 4th edition), consistent with increased level of AhpC in LB cells. Elevated levels of Adk and ClsC indicated differences in membrane lipid composition of SESOM versus LB-grown populations due to their role in synthesis of phospholipids. An addition, Adk has been linked to mutational fitness effects. It was also observed that the length of the lag phase is more sensitive to variation in Adk catalytic capacity than is the exponential growth rate, so that the lag phase appears to be optimal with respect to variation in Adk catalytic capacity (Adkar et al. 2017). NemA, abundant in LB cultures, is involved in reductive degradation of toxic nitrous compounds (Umezawa et al. 2008), again consistent with elevated AhpC in LB populations. Enhanced levels of the peptidoglycan-modulating factors MipA and Pbl in LB cultures

indicates differences in cell wall structure between the stationary phase cultures. High levels of MipA have been reported in sessile compared to planktonic cultures (Rivas et al. 2008).

Periplasmic nutrient uptake systems varied in quantity across the three culture media, but would be remnants from exponential phase where amino acid and sugar uptake were required. The high levels of AnsB, ArtJ and GlnH in LB cultures is puzzling as LB supplies ample amino acids derived from tryptone and yeast extract. Our forest SESOM did not contain detectable levels of lysine, arginine, ornithine or histidine (Liebeke et al. 2009), explaining the enhanced level of ArgT and HisJ. ArgT expression is increased in response to nitrogen starvation and during early response glucose limitation (Kabir et al. 2004; Franchini and Egli 2006), and our SESOM contained very little glucose. Molybdenum (molybdate) is essential as cofactor for the assembly and function of several enzymes including nitrate reductase, formate dehydrogenase, dimethyl-sulfoxide reductase, trimethylamine-N-oxide reductase, and biotin-sulfoxide reductase (Rajagopalan and Johnson 1992). PtsI, more prevalent in LB, is a component of the glucose uptake system that is inhibited by α-ketoglutarate during nitrogen limitation, when it was over expressed the metabolic rate was increased fourfold (Chubukov et al. 2017).

DksA was highly expressed in SESOM grown cells, suggesting a stringent response with induction of ppGpp synthesis due to nutrient limitation. DksA activated by ppGpp binds to the β-subunit of RNA polymerase, directly affecting the affinity to different promoters and thus altering the expression level of more than 80 genes, most importantly suppression of all components of the protein biosynthesis system: rRNA, ribosomal

proteins, and translation factors (Pletnev et al. 2015). This indicates that transcription is shut down tightly in SESOM-grown stationary phase cells.

LB populations appeared more stressed as indicated by elevated levels of various stress proteins and chaperones. LB populations contained more ClpP, part of the proteosomal protein degradation system. Controlled degradation of cytoplasmic proteins has long been considered essential for survival of bacteria under stress conditions, due to the requirement for efficient removal of misfolded or otherwise damaged proteins by ClpP (Weichart et al. 2003). The corresponding low abundance of ClpP in cultures with no decline phase indicated either a reduced need for protein turnover, or a lower degree of damaged proteins. A different profile of damaged proteins was supported by the differences in chaperones GroL (SESOM and 1/40 LB) and HtpG (LB). HtpG expression is increased in cells grown in a complex medium with ample amino acid availability (LB) following heat shock, but low in glucose minimal medium. HtpG expression unaffected or even repressed by imposition of a nutrient stress condition in minimal medium (Mason et al. 1999). The stressed nature of LB stationary cells was supported by elevated levels of Tsf (Elongation factor EF), which plays a role in sequestering surfaces of heterologous proteins to prevent protein–protein interactions leading to formation of inclusion bodies (Han et al. 2007). Elevated levels of the putative stress proteins YbdQ and YedU further indicated greater degree of stress in LB populations, as also indicated by alkyl hyodroperoxide reductase. The elevated levels of TufA (Elongation factor Tu) in SESOM indicates minor starvation due to nutrient limitation. TufA plays an important role in a minor starvation defense mechanism where it helps in rescuing stuck ribosomes (Pletnev et al. 2015).

A total of 17 central carbon metabolism enzymes were detected, and 14 of these were more abundant in LB, indicating that SESOM populations had prepared for reduced metabolic activity going into stationary phase. *E. coli* grown in rich medium undergo a reconstruction of their proteome in stationary phase, with increases in proteins required for scavenging and metabolizing rare nutrients and general cell protection (Li et al. 2014). An example was AckA (acetate kinase), part of the acetate switch that occurs as cells deplete their environment of acetate-producing carbon sources and scavenge for acetate. The accumulation of extracellular acetate during stationary phase occurs as cells co-metabolize acetogenic amino acids, e.g. l-threonine and l-alanine, with those that require the TCA cycle, e.g., L-glutamate (Wolfe 2005). A second example was the elevated levels of ATP synthase components in LB populations, indicating continued need for ATP synthesis driven by periplasmic proton motive force. A third example was PckA (phosphoenolpyruvate carboxykinase), elevated in LB and 1/40th LB. PckA increases 100-fold in the stationary phase independent of cyclic AMP, probably to provide carbohydrates required for energy reserves after cessation of growth, since protease activity, Krebs cycle enzyme activities, and glycogen synthesis all increase in the stationary phase (Goldie and Sanwal 1980).

LB-grown populations had higher overall levels of amino acid biosynthetic enzymes, than both SESOM and dilute LB cultures. This contrasts with the abundance of metabolizable oligopeptides available in LB. However, bioassay of LB medium after growth of *E. coli* showed that it no longer contains significant amounts of recoverable L-serine, L-threonine, L-proline, glycine, L-arginine, L-glutamine, L-asparagine, L-

cysteine, and L-lysine (Sezonov et al. 2007), indicating a need for synthesis in stationary phase.

E. coli O157 933D appears well adapted to grow using soluble nutrients available in soil (SESOM). Moreover, SESOM grown populations did not display a detectable death phase, but remained culturable for at least 24 d. This was supported by the substantially altered proteome of SESOM-grown stationary phase populations. Our results suggest that *E. coli* may well be a soil commensal that maintains stable populations in soil, as growth supported by soil nutrients combined with enhanced longevity of cells would help counter the effects of competition and predation. Soil itself should, therefore, be included as potential source of contamination of fresh produce. Future work should investigate the roles of competition and predation affecting *E. coli* populations in soil.

5. References

Adamowicz, M., Conway, T., and Nickerson, K.W. (1991). Nutritional complementation of oxidative glucose metabolism in *Escherichia coli* via pyrroloquinoline quinone-dependent glucose dehydrogenase and the Entner-Doudoroff pathway. *Appl. Environ. Microbiol.* 57**,** 2012-2015.

Adkar, B.V., Manhart, M., Bhattacharyya, S., Tian, J., Musharbash, M., and Shakhnovich, E.I. (2017). Optimization of lag phase shapes the evolution of a bacterial enzyme. *Nat. Ecol. Evol.* 1**,** 149.

Alexander, D.M., and St John, A.C. (1994). Characterization of the carbon starvation-inducible and stationary phase-inducible gene slp encoding an outer membrane lipoprotein in *Escherichia coli*. *Mol. Microbiol.* 11**,** 1059-1071.

Berry, E.D., and Miller, D.N. (2005). Cattle feedlot soil moisture and manure content: II. Impact on *Escherichia coli* 0157. *Journal of Environmental Quality* 34**,** 656-663.

Brandl, M.T. (2008). Plant lesions promote the rapid multiplication of *Escherichia coli* O157:H7 on postharvest lettuce. *Appl. Environ. Microbiol.* 74**,** 5285-5289.

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

Chen, L., Xie, Q.W., and Nathan, C. (1998). Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol. Cell* 1**,** 795-805.

Chubukov, V., Desmarais, J.J., Wang, G., Chan, L.J.G., Baidoo, E.E.K., Petzold, C.J., Keasling, J.D., and Mukhopadhyay, A. (2017). Engineering glucose metabolism of *Escherichia coli* under nitrogen starvation. *NPJ Syst. Biol. Appl.* 3**,** 16035-.

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

Denis, N., Zhang, H., Leroux, A., Trudel, R., and Bietlot, H. (2016). Prevalence and trends of bacterial contamination in fresh fruits and vegetables sold at retail in Canada. *Food Control* 67**,** 225-234.

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. (2002). Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl. Environ. Microbiol.* 68**,** 1165-1172.

Franchini, A.G., and Egli, T. (2006). Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiol* 152**,** 2111-2127.

Fremaux, B., Prigent-Combaret, C., Delignette-Muller, M.L., Mallen, B., Dothal, M., Gleizal, A., and Vernozy-Rozand, C. (2008). Persistence of Shiga toxin-producing *Escherichia coli* O26 in various manure-amended soil types. *J. Appl. Microbiol.* 104**,** 296-304.

Gagliardi, J.V., and Karns, J.S. (2002). Persistence of *Escherichia coli* O157 : H7 in soil and on plant roots. *Environ. Microbiol.* 4**,** 89-96.

Goldie, A.H., and Sanwal, B.D. (1980). Genetic and physiological characterization of *Escherichia coli* mutants deficient in phosphoenolpyruvate carboxykinase activity. *J. Bacteriol.* 141**,** 1115-1121.

Grant, J., Wendelboe, A.M., Wendel, A., Jepson, B., Torres, P., Smelser, C., and Rolfs, R.T. (2008). Spinach-associated *Escherichia coli* O157:H7 outbreak, Utah and New Mexico, 2006. *Emerg. Infect. Dis.* 14**,** 1633-1636.

Guggenberger, G., and Zech, W. (1993a). Dissolved organic carbon control in acid forest soils of the Fichtelgebirge (Germany) as revealed by distribution patterns and structural composition analyses. *Geoderma* 59**,** 109-129.

Guggenberger, G., and Zech, W. (1993b). Dissolved Organic-Matter (Dom) Dynamics in Spruce Forested Sites - Examinations by Analytical Dom Fractionation. *Zeitschrift Fur Pflanzenernahrung Und Bodenkunde* 156**,** 341-347.

Han, K.Y., Song, J.A., Ahn, K.Y., Park, J.S., Seo, H.S., and Lee, J. (2007). Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. *FEMS Microbiol. Lett.* 274**,** 132-138.

Herman, K.M., Hall, A.J., and Gould, L.H. (2015). Outbreaks attributed to fresh leafy vegetables, United States, 1973–2012. *Epidemiol. Infect.* 143**,** 3011-3021.

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

Huang, Y., Eglinton, G., Van Der Hage, E.R.E., Boon, J.J., Bol, R., and Ineson, P. (1998). Dissolved organic matter and its parent organic matter in grass upland soil horizons studied by analytical pyrolysis techniques. *Eur. J. Soil. Sci.* 49**,** 1-15.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006). Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl. Environ. Microbiol.* 72**,** 612-621.

Islam, M., Doyle, M.P., Phatak, S.C., Millner, P., and Jiang, X.P. (2004). Persistence of enterohemorrhagic *Escherichia coli* O157 : H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. *J. Food Prot.* 67**,** 1365-1370.

Jones, D.L., Kemmitt, S.J., Wright, D., Cuttle, S.P., Bol, R., and Edwards, A.C. (2005). Rapid intrinsic rates of amino acid biodegradation in soils are unaffected by agricultural management strategy. *Soil Biology and Biochemistry* 37**,** 1267-1275.

Kabir, M.S., Sagara, T., Oshima, T., Kawagoe, Y., Mori, H., Tsunedomi, R., and Yamada, M. (2004). Effects of mutations in the rpoS gene on cell viability and global gene expression under nitrogen starvation in *Escherichia coli. Microbiol.* 150**,** 2543-2553.

Kaiser, K., Guggenberger, G., Haumaier, L., and Zech, W. (2001). Seasonal variations in the chemical composition of dissolved organic matter in organic forest floor layer leachates of old-growth Scots pine (*Pinus sylvestris* L.) and European beech (*Fagus sylvatica* L.) stands in northeastern Bavaria, Germany. *Biogeochemistry* 55**,** 103-143.

Kalbitz, K., Schwesig, D., Schmerwitz, J., Kaiser, K., Haumaier, L., Glaser, B., Ellerbrock, R., and Leinweber, P. (2003). Changes in properties of soil-derived

dissolved organic matter induced by biodegradation. *Soil Biol. Biochem.* 35**,** 1129-1142.

Kalbitz, K., Solinger, S., Park, J.H., Michalzik, B., and Matzner, E. (2000). Controls on the dynamics dissolved organic matter in soils: A review. *Soil Science* 165**,** 277-304.

King, L.A., Mailles, A., Mariani-Kurkdjian, P., Vernozy-Rozand, C., Montet, M.P., Grimont, F., Pihier, N., Devalk, H., Perret, F., Bingen, E., Espie, E., and Vaillant, V. (2009). Community-wide outbreak of *Escherichia coli* O157:H7 associated with consumption of frozen beef burgers. *Epidemiol. Infect.* 137**,** 889-896.

Ksoll, W.B., Ishii, S., Sadowsky, M.J., and Hicks, R.E. (2007). Presence and sources of fecal coliform bacteria in epilithic periphyton communities of Lake Superior. *Appl Environ Microbiol* 73**,** 3771-3778.

Leblanc, J.J. (2003). Implication of virulence factors in *Escherichia coli* O157 : H7 pathogenesis. *Crit. Rev. Microbiol.* 29**,** 277-296.

Li, Z.P., Nimtz, M., and Rinas, U. (2014). The metabolic potential of *Escherichia coli* BL21 in defined and rich medium. *Microb. Cell. Fact.* 13.

Liebeke, M., Brozel, V.S., Hecker, M., and Lalk, M. (2009). Chemical characterization of soil extract as growth media for the ecophysiological study of bacteria. *Appl. Microbiol. Biotechnol.* 83**,** 161-173.

Lindsay, D., and Von Holy, A. (1999). Different responses of planktonic and attached *Bacillus subtilis* and *Pseudomonas fluorescens* to sanitizer treatment. *J Food Prot* 62**,** 368-379.

Liu, Y.F., Yan, J.J., Lei, H.Y., Teng, C.H., Wang, M.C., Tseng, C.C., and Wu, J.J. (2012). Loss of Outer Membrane Protein C in *Escherichia coli* Contributes to Both Antibiotic Resistance and Escaping Antibody-Dependent Bactericidal Activity. *Infect. Immun.* 80**,** 1815-1822.

Looper, M.L., Edrington, T.S., Callaway, T.R., and Rosenkrans, C.F., Jr. (2009). Fate of *Escherichia coli* O157:H7 and *Salmonella* from contaminated manure slurry applied to soil surrounding tall fescue. *Lett Appl Microbiol* 48**,** 513-516.

Lynch, J.M. (1982). Limits to Microbial-Growth in Soil. *Journal of General Microbiology* 128**,** 405-410.

Mason, C.A., Dünner, J., Indra, P., and Colangelo, T. (1999). Heat-Induced Expression and Chemically Induced Expression of the *Escherichia coli* Stress Protein HtpG Are Affected by the Growth Environment. *Appl. Environ. Microbiol.* 65**,** 3433-3440.

Nandakafle, G., Seale, T., Flint, T., Nepal, M., Venter, S.N., and Brozel, V.S. (2017). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. *Microbes Environ.* 32**,** 226-233.

Nystrom, T., and Neidhardt, F.C. (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol. Microbiol.* 11**,** 537-544.

Ongeng, D., Geeraerd, A.H., Springael, D., Ryckeboer, J., Muyanja, C., and Mauriello, G. (2015). Fate of *Escherichia coli* O157:H7 and Salmonella enterica in the manure-amended soil-plant ecosystem of fresh vegetable crops: A review. *Crit. Rev. Microbiol.* 41**,** 273-294.

Pizzeghello, D., Zanella, A., Carletti, P., and Nardi, S. (2006). Chemical and biological characterization of dissolved organic matter from silver fir and beech forest soils. *Chemosphere* 65**,** 190-200.

Pletnev, P., Osterman, I., Sergiev, P., Bogdanov, A., and Dontsova, O. (2015). Survival guide: *Escherichia coli* in the stationary phase. *Acta Naturae* 7**,** 22-33.

Rabilloud, T. (1992). A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* 13**,** 429-439.

Rajagopalan, K.V., and Johnson, J.L. (1992). The pterin molybdenum cofactors. *J. Biol. Chem.* 267**,** 10199-10202.

Rivas, L., Fegan, N., and Dykes, G.A. (2008). Expression and putative roles in attachment of outer membrane proteins of *Escherichia coli* O157 from planktonic and sessile culture. *Foodborne Pathog. Dis.* 5**,** 155-164.

Schneckenberger, K., Demin, D., Stahr, K., and Kuzyakov, Y. (2008). Microbial utilization and mineralization of [^{14}C]glucose added in six orders of concentration to soil. *Soil Biology and Biochemistry* 40**,** 1981-1988.

Sezonov, G., Joseleau-Petit, D., and D'ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* 189**,** 8746-8749.

Strobel, B.W. (2001). Influence of vegetation on low-molecular-weight carboxylic acids in soil solution - a review. *Geoderma* 99**,** 169-198.

Strockbine, N.A., Marques, L.R., Newland, J.W., Smith, H.W., Holmes, R.K., and O'brien, A.D. (1986). Two toxin-converting phages from *Escherichia coli* O157:H7 strain 933 encode antigenically distinct toxins with similar biologic activities. *Infect. Immun.* 53**,** 135-140.

Tallon, P., Magajna, B., Lofranco, C., and Leung, K.T. (2005). Microbial indicators of faecal contamination in water: A current perspective. *Water Air Soil Pollut.* 166**,** 139-166.

Umezawa, Y., Shimada, T., Kori, A., Yamada, K., and Ishihama, A. (2008). The uncharacterized transcription factor YdhM is the regulator of the nemA gene, encoding N-ethylmaleimide reductase. *J Bacteriol* 190**,** 5890-5897.

Van Hees, P.a.W., Jones, D.L., Finlay, R., Godbold, D.L., and Lundstromd, U.S. (2005). The carbon we do not see - The impact of low molecular weight compounds on carbon dynamics and respiration in forest soils: A review. *Soil Biol Biochem* 37**,** 1-13.

Vilain, S., and Brozel, V.S. (2006). Multivariate approach to comparing whole-cell proteomes of *Bacillus cereus* indicates a biofilm-specific proteome. *J. Proteome. Res.* 5**,** 1924-1930.

Vilain, S., Cosette, P., Charlionet, R., Hubert, M., Lange, C., Junter, G.A., and Jouenne, T. (2001). Substituting Coomassie Brilliant Blue for bromophenol blue in two-dimensional electrophoresis buffers improves the resolution of focusing patterns. *Electrophoresis* 22**,** 4368-4374.

Vilain, S., Luo, Y., Hildreth, M.B., and Brözel, V.S. (2006). Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl. Environ. Microbiol.* 72**,** 4970-4977.

Vital, M., Hammes, F., and Egli, T. (2008). *Escherichia coli* O157 can grow in natural freshwater at low carbon concentrations. *Environ. Microbiol.* 10**,** 2387-2396.

Voigt, B., Schweder, T., Sibbald, M.J.J.B., Albrecht, D., Ehrenreich, A., Bernhardt, J., Feesche, J., Maurer, K.H., Gottschalk, G., Van Dijl, J.M., and Hecker, M. (2006). The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. *Proteomics* 6**,** 268-281.

Wang, O., Mcallister, T.A., Plastow, G., Stanford, K., Selinger, B., and Guan, L.L. (2017). Host mechanisms involved in cattle *Escherichia coli* O157 shedding: a fundamental understanding for reducing foodborne pathogen in food animal production. *Sci. Rep.* 7.

Wang, Y. (2002). The function of OmpA in *Escherichia coli. Biochem. Biophys. Res. Commun.* 292**,** 396-401.

Weichart, D., Querfurth, N., Dreger, M., and Hengge-Aronis, R. (2003). Global Role for ClpP-Containing Proteases in Stationary-Phase Adaptation of *Escherichia coli*. *J Bacteriol* 185**,** 115-125.

Williams, A.P., Mcgregor, K.A., Killham, K., and Jones, D.L. (2008). Persistence and metabolic activity of *Escherichia coli* O157:H7 in farm animal faeces. *FEMS Microbiol Lett* 287**,** 168-173.

Winfield, M.D., and Groisman, E.A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl. Environ. Microbiol.* 69**,** 3687-3694.

Wolfe, A.J. (2005). The Acetate Switch. *Microbiol. Mol. Biol. Rev.* 69**,** 12-50.

Wright, K.M., Chapman, S., Mcgeachy, K., Humphris, S., Campbell, E., Toth, I.K., and Holden, N.J. (2013). The endophytic lifestyle of *Escherichia coli* O157:H7: quantification and internal localization in roots. *Phytopathology* 103**,** 333-340.

Wright, K.M., Crozier, L., Marshall, J., Merget, B., Holmes, A., and Holden, N.J. (2017). Differences in internalization and growth of *Escherichia coli* O157:H7 within the apoplast of edible plants, spinach and lettuce, compared with the model species Nicotiana benthamiana. *Microb. Biotechnol.* 10**,** 555-569.

Yang, S.C., Lin, C.H., Aljuffali, I.A., and Fang, J.Y. (2017). Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. *Arch. Microbiol.* 199**,** 811-825.

Yoon, J.W., and Hovde, C.J. (2008). All blood, No stool: enterohemorrhagic *Escherichia coli* O157 : H7 infection. *J. Vet. Sci.* 9**,** 219-231.

Chapter 5: Growth and Population Maintenance of *Escherichia coli* in Soil Organic Matter

The work reported in this chapter was conducted by Gitanjali NandaKafle in collaboration with Alexander W. Kena

1. Introduction

Escherichia coli is the foundational model organism for molecular biology (Russo, 2003), used widely as a tool to produce various biomolecules including plasmids and proteins (Rosano and Ceccarelli, 2014), and is the primary indicator of fecal contamination (Edberg et al., 2000). When exposed to amenable organic energy and carbon sources it grows by binary fission, displaying exponential growth until resources fall below the required threshold, leading to growth arrest or stationary phase (Nystrom, 2004). Transition to stationary phase entails preparation for enhanced tolerance to stress, and stationary phase cells synthesize proteins at about 20% or the exponential rate (Reeve et al., 1984). Stationary phase endures for 2 to 5 days, after which cells begin to die, the decline phase (Ericsson et al., 2000;Finkel, 2006). Intriguingly about 1% of cells survive, and enter Long Term Stationary Phase (LTSP) (Finkel, 2006), or Constant Activity in Stationary Phase (CASP) (Gefen et al., 2014).

In nature bacteria are largely nutrient limited, dividing rarely, and spending most of their time in stationary phase (Gefen et al., 2014). The ability of *E. coli* to enter LTSP and remain culturable for long periods in absence of added nutrients suggests an ability to survive in the environment. *E. coli* is widely viewed as a gastrointestinal bacterium, but

can be isolated from a range of habitats, several of which are subject to fluctuating

conditions including changes in pH, oxygen availability, nutrient availability,

temperature, and osmolarity (van Elsas et al., 2011). Sources include surface waters,

sediments, fresh produce and soil (Winfield and Groisman, 2003;NandaKafle et al.,

2017). *E. coli* has been isolated from various soils with no evidence for recent fecal

contamination, including cattle pasture devoid of grazing cattle for 10 months

(NandaKafle et al., 2017). Population maintenance of bacteria in soil is the product of

growth, predation, competition and death, so growth through cell division should be a

prerequisite for population maintenance. It has been shown that *E. coli* can survive in

different soil environments and possibly grow when conditions are suitable (Morris et al.,

1998;Ishii et al., 2006;Brennan et al., 2010;van Elsas et al., 2011). Most studies on

survival of *E. coli* in soil have used enterohemorrhagic strains such as O157:H7, which

can survive in soil for more than 90 days, even after fumigation (Ibekwe et al., 2007).

Several studies have shown progressive decline to below the detection limit in soil and

soil-like environments, but there was very little decline when occurring in sterile soil,

indicating that other soil organisms may play a role in *E. coli* population density

(Bogosian et al., 1996;Duffitt et al., 2011). Soil provides a wide variety of nutrients that

can be utilized by a diverse group of microorganisms (Killham, 1998). Our laboratory

established a protocol to prepare Soil Extracted Soluble Organic Matter (SESOM) as

culture medium (Vilain et al., 2006). SESOM contains a range of amino acids,

carbohydrates, low molecular weight organic acids and a range of inorganic compounds,

and supports the growth of *Bacillus cereus* (Vilain et al., 2006) and various other Gram

negative and positive bacteria (Liebeke et al., 2009). We have recently reported that *E.*

coli O157:H7 is able to grow in liquid extract of various soils (NandaKafle et al., 2018). Populations of the O157:H7 strain 933D grown in deciduous forest soil extract did not display a decline phase but retained 100% culturability for at least 24d (NandaKafle et al., 2018).

Factors leading to presence of *E. coli* in soils without recent fecal contamination are still poorly understood. While many studies have reported survival, there is scant direct evidence for growth using the nutrients available in soils. We hypothesized that diverse *E. coli* other than the enterohaemorrhagic O157 can grow using soluble components in soil and display the same soil-associated extended survival phenotype with no decline phase. *E. coli* K12, Clade I, 25 isolates obtained from cattle pasture and 933D were inoculated into sterile pasture soil SESOM and incubated for 24d, using LB broth as control. All strains could grow in the pasture SESOM, with no decline phase observed.

2. Materials and methods

2.1. *Source of strains used:*

We selected 25 *E. coli* isolates from a recent collection of 390 isolates obtained from cattle pasture (NandaKafle et al., 2017). The experimental pasture was enclosed as four separate units, and cattle were introduced for one month during July every year. Isolates were obtained during June from soil before cattle were introduced (Soil Before Grazing, SBG), soil during grazing, and from fresh cattle feces. Phylogeny was determined using concatenated *mutS* and *uidA* genes, and yielded multiple clusters, classified as environmental, bovine and mixed clusters based on the sample types in each cluster (NandaKafle et al., 2017). Isolates were allocated to phylogroups using the scheme of

(Clermont et al., 2013). Isolates were selected to represent a cross section of sample types and phylogroups. *E. coli* MG1655 (K12), 933D and TW 10509 (Clade I) were included as controls.

2.2. *Preparation of culture media:*

SESOM was prepared as described by Vilain et al. (2006). Briefly 100 g of the dried pasture soil was mixed with 500 mL warm (60°C) 2mM MOPS buffer (pH 7.0) in a 2 L flask and shaken for 1 h at 200 rpm. The suspension was filtered using filter paper, and then through 5 µm and 1.2 µm membrane filters (Millipore), and filter sterilized using 0.2 µm pore size filters (Thermo Scientific Nalgene) membrane filter under vacuum. Sterility of SESOM medium was confirmed by inoculating a few drops onto LB agar plates and incubating at 28°C for 3 d and observing for any growth. LB broth (Miller) containing 10 g NaCl per liter was used as control.

2.3. *Incubation conditions:*

All *E. coli* isolates were retrieved from -80°C glycerol stocks onto LB agar and incubated at 37°C overnight. Colonies were inoculated into 50 mL SESOM and LB broth in 250mL flasks and incubated overnight while shaking at 25° C. Cells were harvested by centrifugation (10,000 x g, 5 Min), washed two times with sterile tap water and re-suspended in SESOM or LB broth respectively to $A_{546} = 2.5$. SESOM and LB were inoculated with 100 µL of this suspension, to yield an initial density of $A_{546} = 0.005$. Samples for culturable counting were taken immediately (0 h), at 1, 2, 3, 4, 5, 6, 7 h after

inoculation, and on days 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 18, 20, 22 and 24. The culturable count was determined using the droplet plate technique (Lindsay and Von Holy, 1999). Briefly, 20 µL volumes of serial dilutions were plated onto LB agar in triplicate and incubated for 18h at 30°C.

2.4. *Statistical Analysis:*

Data were analyzed using the R statistical language. A broken line linear regression model fitted in the R package Segmented (Muggeo, 2008) was used to identify the transition points of lag phase, log phase and decline phase as three separate slopes. The results of 95% confidence intervals were visualized using the ggplot2 package.

3. Results

E. coli O157:H7 933D is able to grow in soluble organic matter available in deciduous forest soil, and survive for extended periods (NandaKafle et al., 2018). To determine whether a range of *E. coli* displayed exponential, stationary and decline phase when growing in pasture SESOM, pasture and bovine fecal isolates were incubated in SESOM for 24d. The culturable count data were broken into three phases, growth (slope1), stationary (slope2) and decline phase (slope3) and analyzed using broken line linear regression. The long-term culturable density of *E. coli* strains in batch culture of SESOM and LB media showed four different patterns (Fig.1a). Few strains followed the classical pattern of exponential, stationary and decline phase, especially in LB broth. For some there was no detectable stationary phase, while in others, stationary was not followed by

a decline phase. A significant increase or decrease in culturable count over time was indicated by confidence intervals (95% confidence) either completely above (increase) or below (decrease) the zero line (Fig. 2).

All 25 pasture and bovine fecal isolates with K12, 933D and Clade-I grew in the pasture SESOM, evidenced by a significant increase in their culturable count (Fig 1), and a confidence interval in slope_1 greater than 0 (Fig. 2a). This indicated presence of growth-supporting nutrients in the pasture soil. It further indicated that growth using soluble organic matter in soil is not unique to *E. coli* O157:H7 strain 933D (NandaKafle et al., 2018), but displayed by a range of *E. coli*.

In SESOM, 22 of the 25 isolates and strains 933D and K-12 did not show any decline directly following the end of exponential phase (Slope_2, Fig. 2b), displaying a classical stationary phase for 4 – 5d. Three strains and Clade I went directly from exponential phase into decline. Only 6 isolates and 933D displayed a stationary phase in LB, while all others transitioned from exponential directly to decline phase. The Clade 1 strain, purportedly an environmental *E. coli* (Luo et al., 2011), went directly from exponential into decline phase, both in SESOM and in LB.

The classical stationary phase ends after a few days, followed by decline or death phase (Finkel, 2006). All isolates and the three control strains displayed unchanged culturability in SESOM for the remainder of the experiment. The three pasture isolates and Clade-I that declined during d1 – 5 (stationary phase) in SESOM, did not display further decline from d5, showing a growth curve pattern of 'b' (Fig. 1a). The remaining 22 isolates showed a growth curve pattern of 'd' (Fig. 1a) in SESOM. In LB the majority of isolates declined, or kept on declining after d 5, following classical decline phase or growth curve

pattern a. Out of 28 strains, 5 followed growth curve pattern 'a', 7 pattern 'b', 15 pattern

'c' and only one of the isolate followed pattern 'd' (Fig. 1a). This indicated that a wide

array of *E. coli* are somehow more resilient when grown in SESOM, displaying

significantly greater longevity than in laboratory media such as LB.
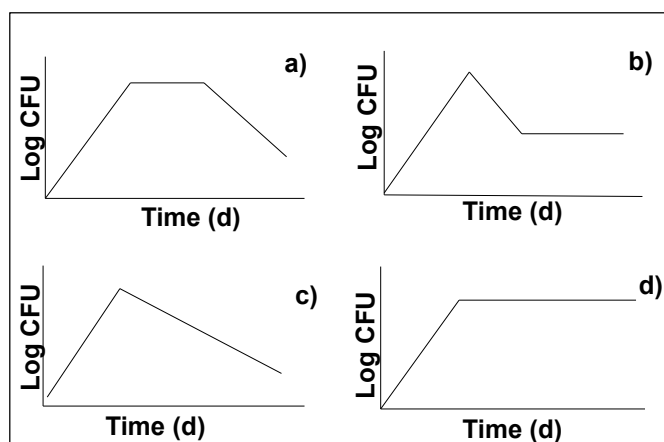
Fig. 5. 1 Four possible patterns of *E. coli* growth curve
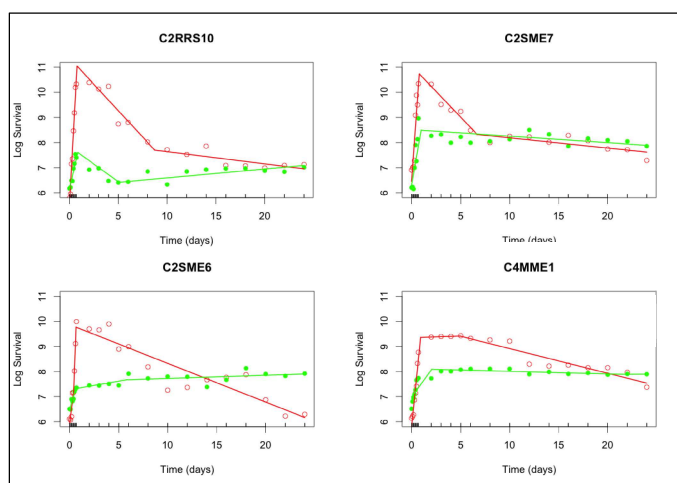
Fig. 5. 2 Culturable counts of four *E. coli* isolates in LB and SESOM
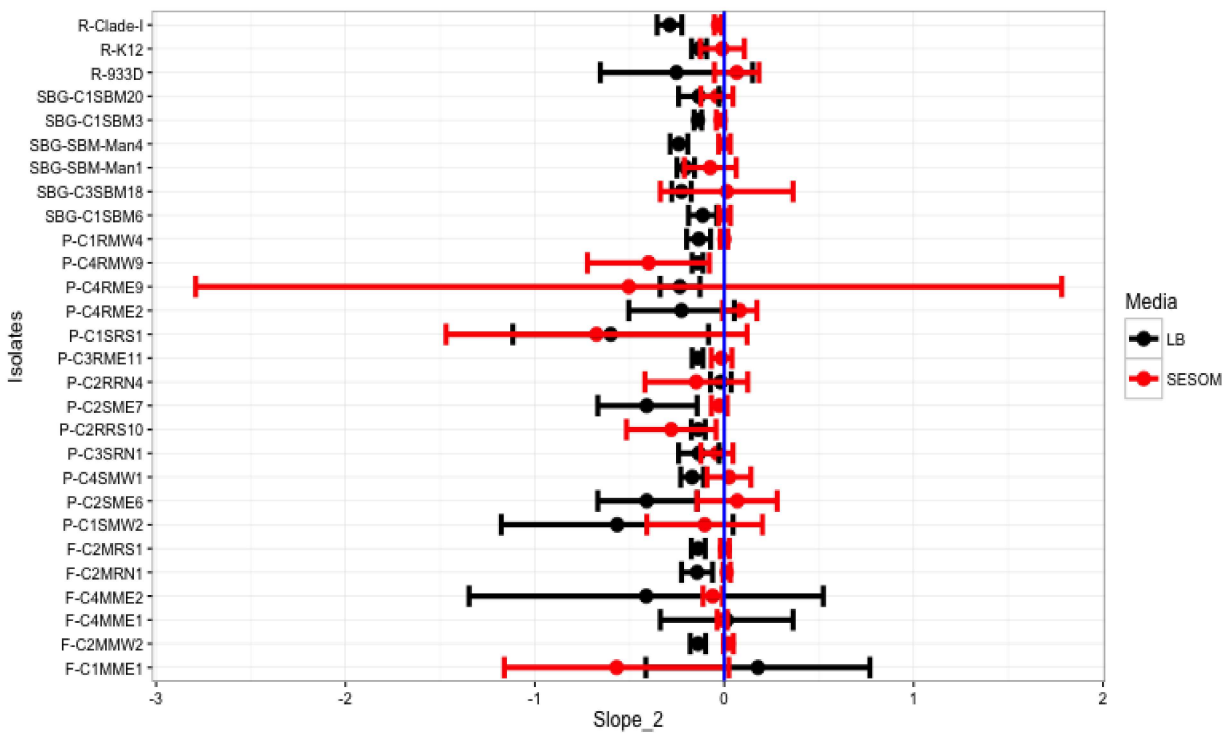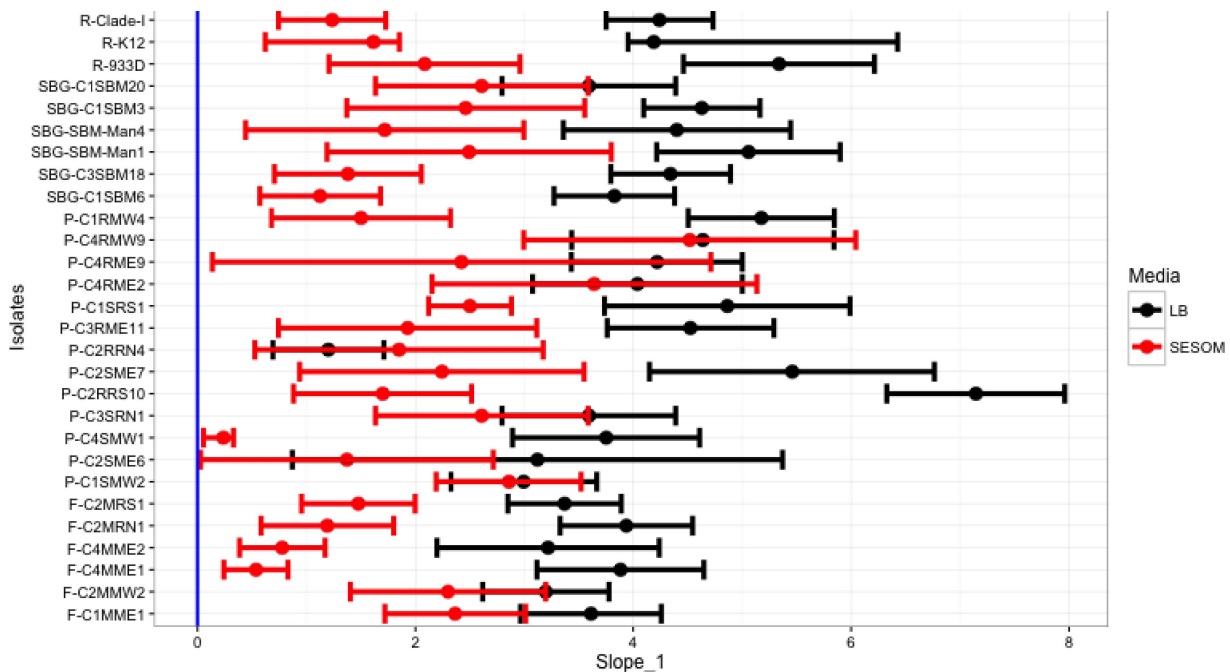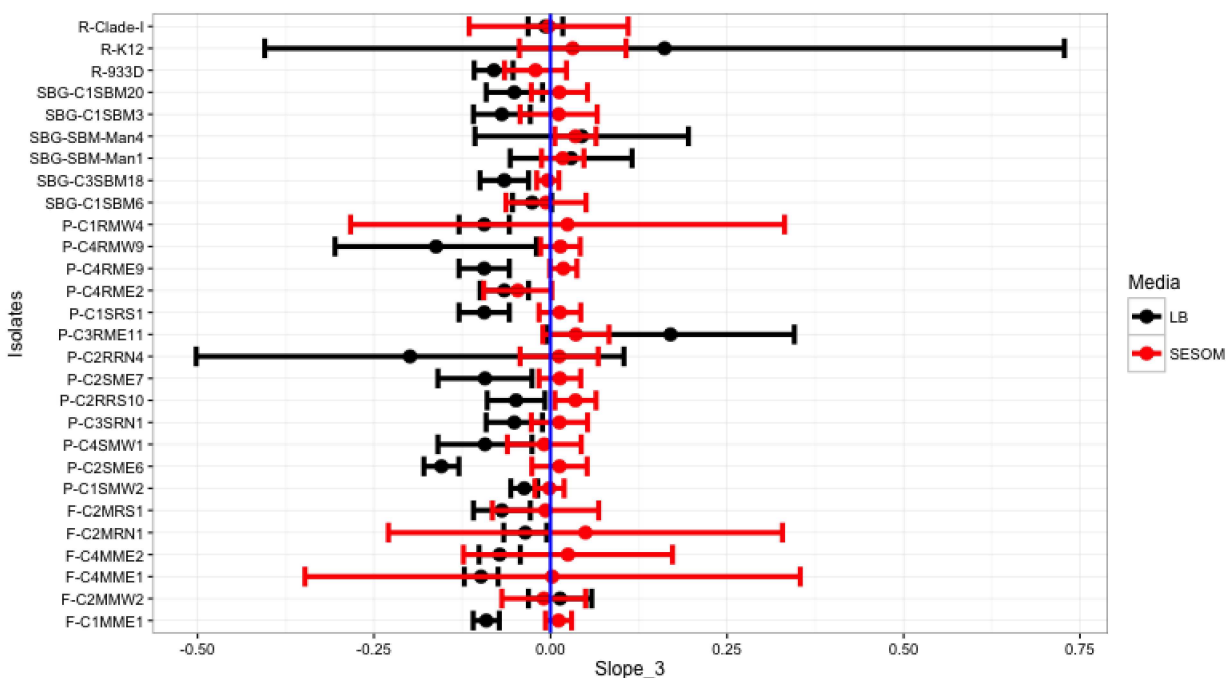
Fig. 5. 3 Plots of 95% confidence intervals using ggplot2 in R. Slope_1 represents lag and log phase; Slope_2 represents "stationary" (or decline) phase; and Slope_3 represents "death" (stationary, decline or incline phase). Range >0 indicates significant increase over time, range<0 indicates significant decrease over time, and if zero falls between maximum and minimum range there was no significant change.

4. Discussion

Factors leading to the presence of *E. coli* in soils without recent fecal contamination are still poorly understood. The aim of this work was to determine whether diverse *E. coli* strains were able to grow using nutrients available in pasture soil, and how this affected their long-term survival. We followed the culturable count of 25 environmental *E. coli* isolates in pasture soil SESOM to determine whether isolates a) were able to grow using soluble nutrients in pasture soil, and b) they survive or decline following classical

stationary phase. All 25 isolates obtained from cattle pasture grew in nutrient extract of pasture soil, indicating that population maintenance is possible as *E. coli* can grow using available nutrients.

Cells cultured in various laboratory media show a population density pattern of 5 phases i.e. lag, exponential, stationary, decline and long term stationary phase. The isolates did not transition from stationary to decline phase during the 24d incubation period, indicating that none of the cells senesced. This implies that cells grown using soil organic matter display superior longevity, remaining culturable for long periods. Growth and survival of *E. coli* has been studied using a plethora of complex and minimal culture media. The nature and concentration of nutrients affect growth rate and cell size (Schaechter et al., 1958). Balanced exponential growth transitions to stationary phase. Cessation of proliferation in rich media such as LB is triggered by cell density, but in nutrient limited media, stationary phase is reached due to the exhaustion of specific nutrients required for biomass production, commonly carbon, nitrogen or phosphate (Gonidakis and Longo, 2013). The nature of the limiting nutrient that causes transition to stationary phase affects the composition and gene regulation of cells (Matin, 1991;Ericsson et al., 2000). After stationary phase for 3-5 days (Ericsson et al., 2000), cells enter the decline phase and about 99% of the cell die (Finkel, 2006;Todar, 2006), but those patients enough to follow the fate of cultures discovered that a subset of the population does not die, but enters Long-Term Stationary Phase (LTSP) (Finkel and Kolter, 1999;Lewis, 2010). Cells that survive during LTSP are subject to selection of mutants with beneficial alleles, or adaptive evolution, expressing the Growth Advantage in Stationary Phase (GASP) phenotype (Finkel and Kolter, 1999). LTSP cultures have

been reported to remain viable for 5 years with addition of just water (Zambrano and Kolter, 1996;Zinser and Kolter, 2004). *E. coli* grown in pasture SESOM did not exit stationary into decline phase, with all cells surviving for 24d. Select cultures followed for 120 d still maintained 100% culturability. Cells grown in very rich media experience higher oxidative stress, which leads to higher glycation levels and mutation frequency, and sometimes do not enter LTSP (Kram and Finkel, 2015). It is possible that the opposite holds for very low nutrient environments such as SESOM, where no decline was observed. While the mechanism underlying zero decline is unknown at present, the phenomenon appears conserved across many *E. coli*. It could be linked to the nutrient composition of SESOM, which contains a diverse array of sugars and other carbohydrates, organic and amino acids in very low concentrations (Liebeke et al., 2009).

Soil persistent strains have been reported to have a unique growth and metabolic characteristics compared to control strains such as K-12 and its derivatives (Brennan et al., 2013). Soil persistent cells were shown to retain fully functional RpoS regulated general stress response, which indicates that the isolates have gone through continuous selective pressure, resulting in well maintained stress resistance despite nutrient limitations (Somorin et al., 2016). In our hands, K-12 and 933D strains displayed the same absence of decline phase in SESOM as the 25 environmental and bovine isolates, suggesting that cells grown in SESOM showed a similar adaptation by all isolates in response to nutrient limitations. This suggests that the development of longer-living cells in SESOM is conserved across the species. Isolation of *E. coli* from diverse soils with no recent evidence of fecal contamination points to maintenance of populations in soils. This

populations maintenance, while under pressure from decimating effects such as predation

and competition, is possible through growth and extended longevity of *E. coli* in soil.

5. References

Bogosian, G., Sammons, L.E., Morris, P.J.L., Oneil, J.P., Heitkamp, M.A., and Weber, D.B. (1996). Death of the Escherichia coli K-12 strain W3110 in soil and water. *Appl. Environ. Microbiol.* 62**,** 4114-4120.

Brennan, F.P., Abram, F., Chinalia, F.A., Richards, K.G., and O'flaherty, V. (2010). Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil. *Appl. Environ. Microbiol.* 76**,** 2175-2180.

Brennan, F.P., Grant, J., Botting, C.H., O'flaherty, V., Richards, K.G., and Abram, F. (2013). Insights into the low-temperature adaptation and nutritional flexibility of a soil-persistent Escherichia coli. *FEMS Microbiol Ecol* 84**,** 75-85.

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

Duffitt, A.D., Reber, R.T., Whipple, A., and Chauret, C. (2011). Gene Expression during Survival of Escherichia coli O157:H7 in Soil and Water. *Int. J. Microbiol.* 2011**,** 12.

Edberg, S.C., Rice, E.W., Karlin, R.J., and Allen, M.J. (2000). Escherichia coli: the best biological drinking water indicator for public health protection. *J. Appl. Microbiol.* 88**,** 106S-116S.

Ericsson, M., Hanstorp, D., Hagberg, P., Enger, J., and Nyström, T. (2000). Sorting Out Bacterial Viability with Optical Tweezers. *J. Bacteriol.* 182**,** 5551-5555.

Finkel, S.E. (2006). Long-term survival during stationary phase: evolution and the GASP phenotype. *Nat. Rev. Micro.* 4**,** 113-120.

Finkel, S.E., and Kolter, R. (1999). Evolution of microbial diversity during prolonged starvation. *Proc. Natl. Acad. Sci.* 96**,** 4023-4027.

Gefen, O., Fridman, O., Ronin, I., and Balaban, N.Q. (2014). Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity *Proc. Natl. Acad. Sci USA* 111**,** 556.

Gonidakis, S., and Longo, V.D. (2013). "Assessing Chronological Aging in Bacteria," in *Cell Senescence: Methods and Protocols,* eds. L. Galluzzi, I. Vitale, O. Kepp & G. Kroemer. (Totowa, NJ: Humana Press), 421-437.

Ibekwe, A.M., Grieve, C.M., and Yang, C.H. (2007). Survival of Escherichia coli O157 : H7 in soil and on lettuce after soil fumigation. *Can. J. Microbiol.* 53**,** 623-635.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006). Presence and growth of naturalized *Escherichia coli* in temperate soils from lake Superior watersheds. *Appl. Environ. Microbiol.* 72**,** 612-621.

Killham, K. (1998). Soil Microbiology and Biochemistry. By E. A. PAUL and F. E. CLARK. 23×15 cm. Pp. xiii+340 with 108 text-figures. San Diego, CA, USA:

Academic Press: 2nd Edition, 1996. Price h/b: £29.95, ISBN 0 12 546806 7. *New Phytol.*138**,** 563-566.

Kram, K.E., and Finkel, S.E. (2015). Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Appl. Environ. Microbiol.* 81**,** 4442-4450.

Lewis, K. (2010). Persister Cells. *Annual Review of Microbiology* 64**,** 357-372.

Liebeke, M., Brozel, V.S., Hecker, M., and Lalk, M. (2009). Chemical characterization of soil extract as growth media for the ecophysiological study of bacteria. *Appl. Microbiol. Biotechnol.* 83**,** 161-173.

Lindsay, D., and Von Holy, A. (1999). Different responses of planktonic and attached *Bacillus subtilis* and *Pseudomonas fluorescens* to sanitizer treatment. *J. Food Prot.* 62**,** 368-379.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011). Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. *Proc. Natl. Acad. Sci. USA* 108.

Matin, A. (1991). The molecular basis of carbon-starvation-induced general resistance in Escherichia coli. *Mol. Microbiol.* 5**,** 3-10.

Morris, R., Byappanahalli, M.N., and Fujioka, R.S. (1998). Health-related Water Microbiology 1998Evidence that tropical soil environment can support the growth of Escherichia coli. *Water Sci. Technol.* 38**,** 171-174.

Muggeo, V. (2008). segmented: An {R} package to Fit Regression Models with Broken-Line Relationships. *R News* 8**,** 20-25.

Nandakafle, G., Christie, A.A., Vilain, S., and Brozel, V.S. (2018). Growth and extended survival of *Escherichia coli* O157:H7 in soil organic matter. *Front Microbiol* 9:762. doi: 10.3389/fmicb.2018.00762.

Nandakafle, G., Seale, T., Flint, T., Nepal, M., Venter, S.N., and Brozel, V.S. (2017b). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. *Microbes Environ.* 32**,** 226-233.

Nystrom, T. (2004). Stationary-phase physiology. *Annu. Rev. Microbiol.* 58**,** 161-181.

Reeve, C., Amy, P., and Matin, A. (1984). Role of protein synthesis in the survival of carbon-starved Escherichia coli K-12. *J. bacteriol.* 160**,** 1041-1046.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* 5**,** 172.

Russo, E. (2003). Special Report: The birth of biotechnology. *Nature* 421**,** 456.

Schaechter, M., Maaløe, O., and Kjeldgaard, N.O. (1958). Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *Microbiol.* 19**,** 592-606.

Somorin, Y., Abram, F., Brennan, F., and O'byrne, C. (2016). The General Stress Response Is Conserved in Long-Term Soil-Persistent Strains of Escherichia coli. *Appl. Environ. Microbiol.* 82**,** 4628-4640.

Todar, K. (2006). *Todar's online textbook of bacteriology.* University of Wisconsin-Madison Department of Bacteriology Madison, Wis, USA.

Van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5**,** 173-183.

Vilain, S., Luo, Y., Hildreth, M.B., and Brozel, V.S. (2006). Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl. Environ. Microbiol.* 72**,** 4970-4977.

Winfield, M.D., and Groisman, E.A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl. Environ. Microbiol.* 69**,** 3687-3694.

Zambrano, M.M., and Kolter, R. (1996). GASPing for life in stationary phase. *Cell* 86**,** 181-184.

Zinser, E.R., and Kolter, R. (2004). Escherichia coli evolution during stationary phase. *Res. Microbiol.* 155**,** 328-336.

Chapter 6: Differentiating *Escherichia coli* Fitness in Soil by

Susceptibility to Grazing by *Dictyostelium discoideum*

The work reported in this chapter was conducted by Gitanjali NandaKafle in

collaboration with Lane A. Blasius

1. Introduction

In the last decade, research efforts have found substantial populations of *E. coli* harboring

in the soil, freshwater sediments, water plants, and even in beach sand (Ishii et al., 2006;

Byappanahalli et al., 2007;Byappanahalli et al., 2012;Jang et al., 2017). When dealing

with particular pathogenic strains of *E. coli* in the open environment, the degree of its

survival and factors that affect this survival rate are crucial from the fundamental point of

view (van Elsas et al., 2011). However, survival of *E. coli* in an open environment

requires the ability to overcome environmental stresses, such as nutrient deprivation,

fluctuating temperature, salinity, exposure to solar radiation, competition with

autochthonous microbial communities, and protozoan grazing. Although both biotic and

abiotic factors play roles in the growth and decline of *E. coli* populations in the

environment, studies have shown that *E. coli* populations decline in natural soils while

there is a minimal or no decline in sterile soils (Jiang et al., 2002;Semenov et al., 2007).

Korajkic et al. (2013) reported that *E. coli* decay was minimal in outdoor microcosms that

were exposed to natural UV radiation when the natural microbiota (pedation and

competition) was removed by disinfection. This suggests that the natural microbiota plays

a relatively important role in controlling the growth of *E. coli* in the natural habitat.

Some strains of *E. coli* survive by acquiring selective growth advantages to maintain its population for long periods of time and become a persistent or naturalized community. To date, few studies have assessed the survival of different *E. coli* isolates in the presence of natural microbiota, specifically protozoa. This is important because the bacterial elimination rate by natural protozoa varies depending on different bacterial characteristics such as cell size, cell wall composition, presence of virulence factors, and location (González et al., 1990;Iriberri et al., 1994;Steinberg and Levin, 2007;Adiba et al., 2010). Protozoa are the most prevalent predators of bacteria in soil, and play a major role in controlling the bacterial population, although not all bacteria seem to be an equally suitable food source for protozoa. *E. coli* has been found to be an excellent food source for three types of amoeba: *Acanthamoeba polyphaga*, *A. castellanii*, and *H. vermiformis* (Weekers et al., 1993). Various protozoa isolated from dairy wastewater has been reported to have different grazing effects of E. coli (Ravva et al., 2010).

The fitness of *E.coli* O157 to resist predation was tested for the curli positive and curli negative phenotypes and it was concluded that the curli negative trait is selective for survival against predation (Ravva et al., 2014).

Virulence factors of *E. coli* have also been shown to be responsible for enhancing their survival by providing protection against predation from bactivorous protozoa, nematodes and other predators in the soil, water or gastrointestinal tract of bovine hosts (Steinberg and Levin, 2007). Protozoan predation is considered to be an important factor in shaping the genotypic and phenotypic structure of planktonic and terrestrial bacterial communities (Hahn and Hofle, 2001;Jurgens and Matz, 2002). Since protozoan predation contributes to the decline of bacterial populations, it is possible that bacteria can become resistant to

this predation through evolutionary processes. There are various defense mechanisms that bacteria can use to either avoid or endure predation (Matz and Kjelleberg, 2005). It has been reported that the susceptibility of bacteria to predation varies within and between species. Some serotypes of *Salmonella enterica* are more resistant to predation by amoeba than others (Tezcan-Merdol et al., 2004; Wildschutte et al., 2004) *Salmonella enterica* serovar Thompson survives better than *Listeria monocytogenes* within the food vacuoles of *Tetrahymena pyriformis* (Brandl et al., 2005). Smith et al. (2016) have shown that isolates *Pantoea ananatis* exhibit differential grazing susceptibility with some being resistant to amoeboid grazing; they identified *rhlA* and *rhlB* genes involved in the biosynthesis of surfactant glycolipid that enables swarming motility of *P. ananatis* BRT175 and is cytotoxic to amoeba. Some studies, comparing the survival of commensal *E. coli* and *E. coli* O157:H7 in situ and in vitro showed that *E. coli* O157 was more resistant to predation (Jenkins et al., 2011), and other virulence strains (ExPec carrying virulence genes iroN, irp2, fyuA involved in iron uptake) were resistant to the grazing of *Dictyostelium discoideum* (Adiba et al., 2010). It is likely that *E. coli* O157 has the ability to survive in the food vacuoles of protozoa, which enhance their survival advantages in the environment and thereby their chances of transmission to humans (Brandl, 2006). There are also reports with contradictory findings showing that *E. coli* is equally susceptible to predation compared with commensal *E. coli* and other fecal indicator bacteria (Artz and Killham, 2002;Avery et al., 2008). Steinberg and Levin (2007) have reported that the *Stx* encoding prophage of *E.coli* O157:H7 provides protection against predation by grazing protozoa, while (Schmidt et al., 2016) did not find any evidence of protective effects of neither Stx nor the products of other bacteriophage genes on

protozoan predation (*Paramecium caudatum* and *Tetrahymena pyriformis*) on *E. coli* (*E. coli* O157:H7 (EDL933D and its isogenic mutant).

The soil dwelling amoeba *Dictystellium discoedium* is a good model organism to study the interaction of bacteria and protozoan predation. It lives as an independent haploid cell under favorable conditions and feeds on bacteria. When food is scarce it co-aggregates into a multicellular motile slug and then forms fruiting bodies. About 20% of cells die to form a long thin stalk that the rest of the cells ascend. A globular structure formed at the tip of the stalk by the remaining cells is known as a sorus, which contains the spores. This is a strategy for spore dispersal by passing animals (smith et al., 2014). Once a favorable new environment is found the spores hatch into vegetative cells and the cycle continues.

In addition to eating bacteria, *D. discoedium* can also form symbiotic association with different bacterial species. In amoeba this trait appears to be binary, some clones pick up and carry their edible bacterial source throughout their life cycles (DiSalvo et al., 2014).

In this experiment we were sought to find out if *E. coli* isolated from pasture soil and bovine feces   exhibited different grazing susceptibilities to the amoeba *Dictyostelium discoideum,* and if there is any correlation between the presence of virulence factors and protozoan resistance of the strains.

2. Materials and Methods:

E. coli culture condition: *E. coli* isolates were collected from pasture soil (126), run-off (160) SBG (46) and bovine feces (35) as described in our previous work (NandaKafle et

al., 2017). For grazing assay *E.coli* isolates were recovered from -80° C glycerol stock on a LB plate and for grazing assay *E. coli* was grown in HL5 medium (10 g L^{-1} protease peptone, 10 g L^{-1} glucose, 5 g L^{-1} yeast extract, 0.35 g L^{-1} Na_2HPO_4 $7H_2O$, 0.35 g L^{-1} KH_2PO_4, pH 6.5) in a shaker incubator at 28° C (Adiba et al., 2010). Cells were washed once and then resuspended in HL5 media and the optical density was adjusted to 0.5 A_{546}. The bacterial culture (4 µL) then applied on lactose agar plates (1 g L^{-1} lactose, 1 g L^{-1} proteose peptone and 20 g L^{-1} agar).

2.1. *Amoeba strain and culture condition*

Amoeba *D. discoideum* axenic strain was obtained from Carolina Biological Supply. Amoeba was grown in 50 mL HL5 medium at 24°C in a shaker incubator for overnight, cells were washed once and the optical density was adjusted to 0.5 A_{546} to use for grazing assay. For HL5 agar media and broth we replaced thiotone E peptone as mentioned in (Adiba et al., 2010) with protease peptone as it was not available to purchase. Initially we used HL5 agar media to grow *D. discoideum*, but it did not see any grazing effect (no fruiting body was formed) so, we tested various media such that LB, R2A and LA (lactose agar). In LA media *D. discoideum* cells were able to make fruiting bodies, which is a condition when there is no available of immediate food source for amoeba.

2.2. *Grazing resistance assay*

All E. coli isolates were evaluated for grazing resistance by using a quantitative assay as described by (Moore et al., 2016) with some modifications. For this assay D. discoideum

was co-cultured with *E. coli* on LA plates. To test a particular isolate 4 μL (0.5 OD A_{546}) bacterial suspension applied on plates in triplicate lines. Four microliter of *D. discoideum* broth culture was inoculated on the middle of each line as shown in (Fig.1a). All plates were incubated at room temperature for five days. As *D. discoideum* consumed the bacteria and proliferated or grazed. The proliferating (grazing) front advanced along the bacterial lines (Fig.1b). The distance of amoeba grazing was measured in mm. To determine the difference in grazing susceptibility among sample types an ANOVA test was performed using R program.

2.3. Grazing preferences by amoeba

For grazing preferences determination we chose two highly susceptible (strains with highest grazing distance) and six resistant (strains with least grazing distance) strains. Each susceptible strain was inoculated on LA plates in combination with resistant strains. Four μL cultures were streaked on the plate as straight lines touching each other at one end, four μL of amoeba broth culture was then placed on the touching point of the *E. coli* cultures as shown in Fig. 2. Plates were incubated at room temperature for 5 days and grazing distances were measured.

2.4. Virulence gene determination

To determine the presence of six virulence genes (stx1, stx2, eaeA, hlyA, ST and LT) in *E. coli* isolates we used PCR based method as described previously in Chapter # 3.

2.5. *Whole genome of E. coli isolates*

Out of 20 whole genome sequences available (previously sequenced for genome analysis study in our laboratory Chapter # 6) 5 least grazed isolates and 5 most grazed isolates were chosen to study their genetic relatedness. The two groups were denoted as least grazed group (LGG) and highly grazed group (HGG). The unique genes related to each group and the core genes were identified using EDGAR bioinformatics platform (Blom et al., 2016) and R program (R Development Core Team, 2015).

3. Results

3.1. *Grazing resistance*

Grazing susceptibility of *E. coli* isolates from soil, run-off, SBG and bovine feces were measured by the grazing distances of *D. discoideum* on *E. coli* as shown in Fig. 1. The grazing susceptibility varied from strains to strain suggesting some strains are capable of resisting
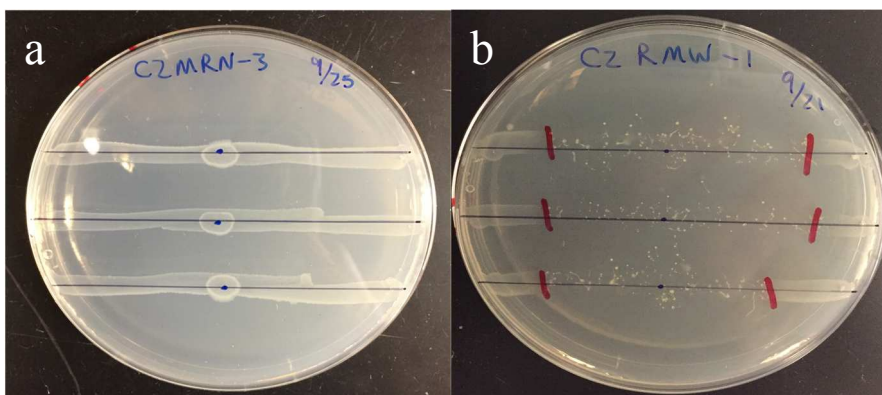
Fig 6. 1. Growth of *E. coli* on LA agar after 24h at 25 °C, and with *D. discoideum* applied at center (a), and after a further 96h incubation at 25 °C in the dark (b).
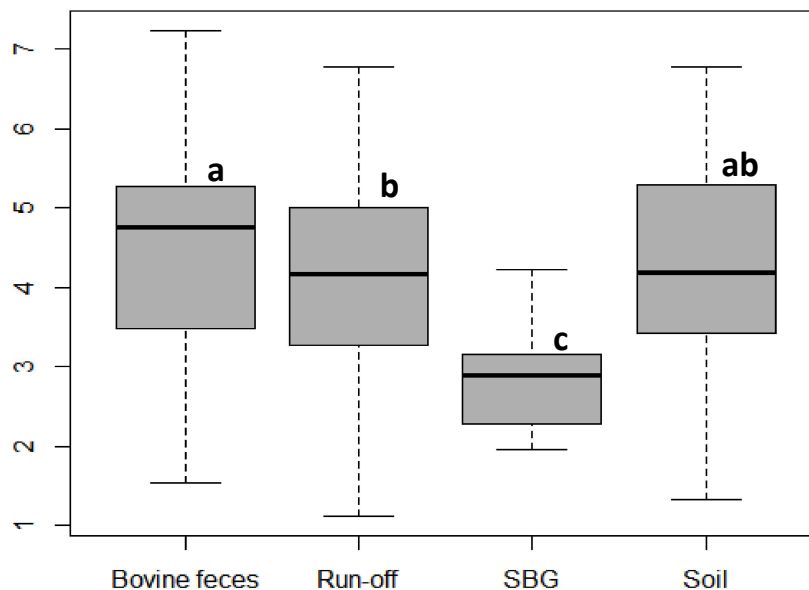
Fig 6. 2. Box and Whisker Plot depicting the grazing distances of *D. discoideum* on E. coli isolates from different sources. Sample groups with the same letter were not significantly difference as determined by ANOVA.

predation may be due to difference in genotype and so phenotypic behavior. The distribution of grazing susceptibility significantly varied among sample types; isolates previously (NandaKafle et al., 2017) found to persist in soil (SBG) showed the least susceptibility to grazing by *D. discoideum* (Fig. 2). This result indicates that *E. coli* population persistence in soil is at least in part due to decreased grazing susceptibility.

3.2. Grazing preferences by D. discoideum

In this assay, when *D. discoideum* was grown in presence of two *E. coli* isolates (one least susceptible and one highly susceptible) it preferred one over the other. The grazing was initiated first on the highly susceptible isolates and grazed a longer distance and later

it grazed on the least susceptible isolates (Fig. 3). Our results indicate that even though *D. discoideum* grazes both isolates but it has a preference for bacterial cells.
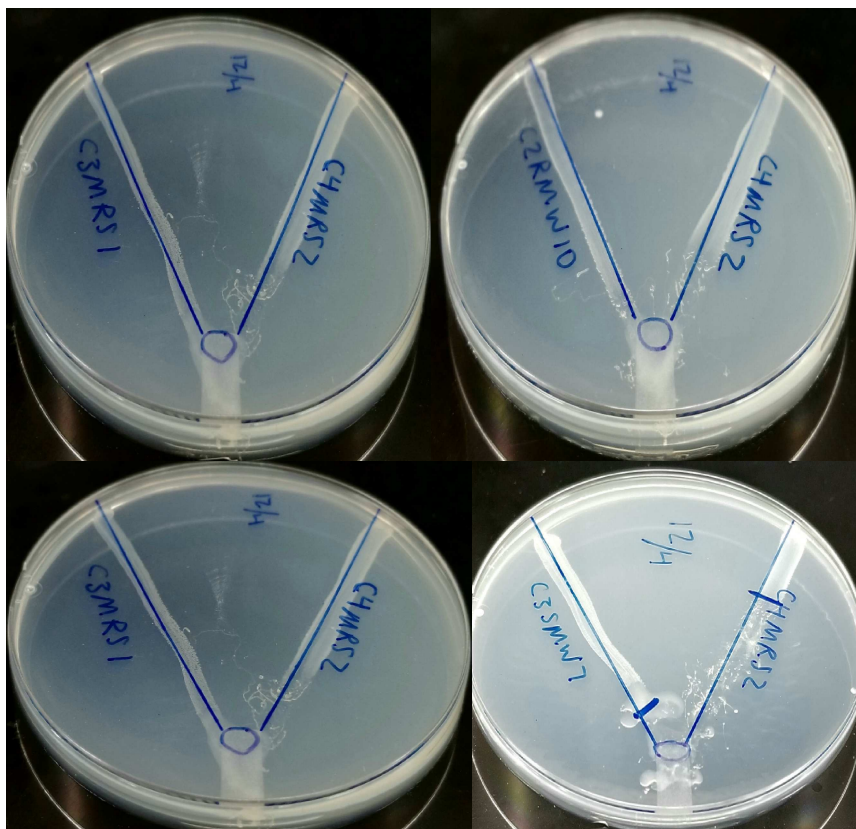
Fig. 6. 3 Grazing preference of *D. discoideum* between two different *E. coli* isolates

3.3. *Presence of virulence genes and grazing susceptibility*

To determine if there is any relationship of the presence of virulence genes and grazing susceptibility of *E. coli*, six virulence genes (*stx1, stx2, eaeA, hlyA*, ST and LT) were checked for all isolates by PCR method. Our result did not show any co-relation between the presence of virulence genes and grazing susceptibility of *E. coli* Fig. S1.

3.4. *Comparison of whole genomes of least grazed and most grazed groups*

The average genome size for the least grazed group (LGG) was 4852 genes and the highly grazed group (HGG) had 5100 genes and there were 3414 core genes between these two groups. It was interesting to find that there were more than double unique genes specific to the highly grazed groups than there were unique genes specific to the least grazed groups, LGG had 130 genes unique to the group and the HGG had 389. These unique genes were grouped based on their function. The numbers of membrane related genes in HGG was 33 and in LGG it was only 8, suggesting there is a huge difference in their membrane structure and the substrates they secrete. The HGG group contained 23 transporter genes compared to LGG having only 10 transporter genes. The HGG group also possessed many fimbial and flagellar genes with small toxic proteins and hemolysin genes.

The LGG group had only three fimbrial and invasion related genes, suggesting that the HGG group could possibly contribute more virulence genes compared to LGG. There are also a high number of toxin-antitoxin system genes in HGG compared to LGG. The quorum sensing molecules such as autoinducer-2 related genes were more abundant in HGG compared to LGG, where only one autoinducer 2-binding protein gene, LsrB, was present. Surprisingly, we found that the secretory proteins related genes in HGG were only four and were related to secretory system-III, whereas in LGG there were 10 secretory system-II proteins related genes present.

Our results suggested that the LGG and HGG strains are phenotypically and genotypically different from each other.

Table 6. 1 List of unique genes in two groups (HGG and LGG) of *E. coli* isolates

Highly Grazed HGG (Susceptible)	Least grazed-LGG (Resistant)
Membrane protein	**Membrane protein**
00706,D-methionine-binding lipoprotein MetQ precursor	02333,LPS-assembly protein LptD
01754,Outer membrane protein IcsA autotransporter precursor	04132,Periplasmic trehalase
00603,Outer membrane usher protein FimD precursor	04745,Inner membrane protein YijD
02959,putative outer membrane usher protein ElfC precursor	02531,Cryptic outer membrane porin BglH
03321,Threonine-rich inner membrane protein GfcA precursor	02786,Inner membrane protein YmfA
03320,putative lipoprotein GfcB precursor	02787,Inner membrane protein YcfZ
04701,Bestrophin, RFP-TM, chloride channel	03086,Outer membrane usher protein PapC
04713,Outer membrane usher protein FimD precursor	02865,Major curlin subunit
05195,Outer membrane porin protein OmpD precursor	
05048,Inner membrane protein YnbA	
03566,Outer membrane protein G precursor	
03559,Inner membrane ABC transporter permease protein YcjP	
03558,Inner membrane ABC transporter permease protein YcjO	
00078,Inner membrane protein YnjI	
00071,Inner membrane metabolite transport protein YdjE	
00065,Inner membrane metabolite transport protein YdjE	
04633,Inner membrane protein YedR	
04857,putative inner membrane protein	
05409,Lipoprotein YlpA precursor	
05429,Type IV conjugative transfer system lipoprotein (TraV)	
03405,Outer membrane protein IcsA autotransporter precursor	
03404,Inner membrane protein YmgF	
02742,InvH outer membrane lipoprotein	
02737,Lipoprotein PrgK precursor	
05051,Phospholipase YtpA	
00929,Putative penicillin-binding protein PbpX	
01085,Inner membrane protein YiaV precursor	
01084,Inner membrane protein YiaW	
03190,Inner membrane protein YidI	
01369,putative outer membrane usher protein ElfC precursor	
01208,Inner membrane protein YihN	
02823,Energy-coupling factor transporter transmembrane protein EcfT	
02855,Inner membrane protein YhaI	
Transporter protein	**Transporter protein**

03297,putative transporter YycB	00866,Ribose import permease protein RbsC
01955,4-hydroxybenzoate transporter PcaK	00865,Ribose import ATP-binding protein RbsA
02919,Inner membrane transporter YcaM	03286,Putrescine transporter PotE
03317,Polysialic acid transport protein KpsD precursor	00867,Autoinducer 2 import system permease protein LsrD
03444,putative autotransporter precursor	01281,Electron transport complex subunit RsxC
04676,Proline/betaine transporter	02631,Fe(2+) transporter FeoB
04692,Sugar efflux transporter	01383,L-fucose-proton symporter
00208,Multidrug transporter EmrE	01554,C4-dicarboxylate TRAP transporter large permease protein DctM
05180,putative D,D-dipeptide transport system permease protein DdpB	02501,D-galactonate transporter
05182,putative D,D-dipeptide transport ATP-binding protein DdpD	01553,Sialic acid TRAP transporter small permease protein SiaQ
05183,putative D,D-dipeptide transport ATP-binding protein DdpF	
00040,L-carnitine/gamma-butyrobetaine antiporter	
04851,putative autotransporter precursor	
03444,putative autotransporter precursor	
02739,Yop proteins translocation protein F	
00997,Sugar efflux transporter C	
03543,Putrescine importer PuuP	
00040,L-carnitine/gamma-butyrobetaine antiporter	
00871,putative formate transporter 1	
03372,Inner membrane ABC transporter permease protein YtfT	
02824,Energy-coupling factor transporter ATP-binding protein EcfA1	
02825,Energy-coupling factor transporter ATP-binding protein EcfA1	
02206,High-affinity glucote transporter	
Virulence related	**Virulence related**
00605,Type-1 fimbrial protein, A chain precursor	04227,putative fimbrial-like protein YadM
00604,Chaperone protein FimC precursor	00840,Invasin
00602,putative fimbrial-like protein ElfG precursor	03085,putative fimbrial-like protein YbgD
00601,Type-1 fimbrial protein, A chain precursor	04227,putative fimbrial-like protein YadM
01601,Fimbria A protein precursor	
00600,Virulence factors putative positive transcription regulator BvgA	
02958,putative fimbrial chaperone protein ElfD precursor	

02957,Fimbrial subunit ElfA precursor	
02960,putative fimbrial-like protein ElfG precursor	
02961,putative fimbrial-like protein YcbU precursor	
02962,putative fimbrial-like protein YcbV precursor	
02963,putative fimbrial chaperone YcbF precursor	
03417,Hemolysin E, chromosomal	
04573,Small toxic polypeptide LdrD,IDENTICAL PARALOGS:,04574,Small toxic polypeptide LdrD	
04714,S-fimbrial protein subunit SfaG precursor	
04715,S-fimbrial adhesin protein SfaS precursor	
04716,S-fimbrial protein subunit SfaH	
02361,Flagellin	
03185,Small toxic protein TisB	
02744,flagellar biosynthesis protein FliR	
02745,flagellar biosynthesis protein FliQ	
02746,Flagellar biosynthetic protein FliP precursor	
01370,putative fimbrial-like protein ElfG precursor	
01367,Type-1 fimbrial protein, A chain precursor	
Toxin antitoxin system	**Toxin antitoxin system**
00418,Antitoxin DinJ	**02097,Toxin YoeB**
03953,Antitoxin ParD1	**02096,Antitoxin YefM**
01856,Antitoxin HicB	**04560,Toxin HigB-2**
03953,Antitoxin ParD1	**04561,Antitoxin HigA-2**
01875,Antitoxin VapB	
04711,Antitoxin HipB	
03952,Toxin ParE1	
02588,Antitoxin MazE	
04802,Antitoxin PrlF	
04801,Toxin YhaV	
02876,Antitoxin HigA	
03367,Antitoxin ChpS	
Quorum sensing	**Quorum sensing**
04705,Autoinducer 2-binding protein LsrB precursor	**00864,Autoinducer 2-binding protein LsrB**
04706,Autoinducer 2 import system permease protein LsrD	
04707,Autoinducer 2 import system permease protein LsrC	
04708,Autoinducer 2 import ATP-binding protein LsrA	
04709,Transcriptiol regulator LsrR	
04710,Autoinducer 2 kise LsrK	
04703,Autoinducer 2-degrading protein LsrG	
stress protein	**stress protein**
00310,Stress-induced bacterial acidophilic repeat motif	**02418,General stress protein A**

00069,General stress protein 69	02415,General stress protein A
Secretory protein	**Secretory protein**
04251,Secreted effector protein pipB2	04794,Putative type II secretion system protein D
02748,type III secretion system protein SpaO	03006,Type II secretion system protein E
02740,Type III secretion system protein PrgH-EprH (PrgH)	03005,Type II secretion system protein F
02748,type III secretion system protein SpaO	04791,Putative type II secretion system protein G
	02998,Type II secretion system protein M
	02999,Type II secretion system protein L
	03000,Putative type II secretion system protein K
	03001,Type II secretion system protein J
	03002,Type II secretion system protein I
	03003,Type II secretion system protein H
	03004,Type II secretion system protein G
	03007,Type II secretion system protein D
	03008,Type II secretion system protein C

4. Discussion

This study showed that *E. coli* isolates showed different susceptibility to protozoan predation and the grazing distance of *E. coli* isolates were different even though they share the same habitat. However, we reported that the majority of isolates from SBG samples that are considered as environmental (NandaKafle et al., 2017) showed significantly higher resistance to grazing compared to soil, run-off and the bovine feces sample. This indicates that *E. coli* may have harbored a different trait to escape predation and survive in the soil. We also confirmed our result by growing least susceptible and highest susceptible strains together with *D. discoeideum* (Fig. 3) to see if there is a preference for grazing on *E. coli* isolates, it was clearly shown that *D. discoeideum*

initiated grazing on the least susceptible isolates. Adiba et al. (2010) have shown that *D. discoideum* was able to survive and phagocytize *E. coli* strains not harboring virulence genes involved in iron capture (*ironN, fyuA, irp*), are not resistance to bile, serum, lactoferin or that do not belong to phylogroup B2.

In our study interestingly, we also found that isolates belong to B2 phylogroup showed resistance to protozoan grazing although there were very few B2 isolates in our collection (total 368 isolates and only 9 B2 isolates). The highest grazing distance was 7.2 cm and the range of grazing distance for isolates belong to phylogroup B2 was 0-3.1 cm. It has been shown *E.coli* strains that harbor virulence genes are able to survive and replicate in common environmental protozoa such as *E. coli* O157, (Barker et al., 1999;Steinberg and Levin, 2007) or extra intestinal pathogenic *E. coli* (Adiba et al., 2010). To determine the correlation between the presence of virulence genes and grazing resistance, we detected the presence and absence of six virulence genes in all isolates. We did not find any correlation between the presence of virulence genes and grazing resistance of *E. coli*. We also measured the grazing distance of *E. coli* O157:H7 strains and did not find any significant resistance by the strain. Our result is consistent with Schmidt et al. (2016) as reported previously that *Paramecium caudatum* consistently reduced both *E. coli* O157:H7 (EDL933D) and non Shiga toxin cattle commensal *E.coli* population by 1-3 log CFU when grown together in broth culture with over three days in an ambient laboratory temperature.

If virulence genes are not the major factor for *E. coli* to be resistant to predation then what are the traits responsible for their ability to evade grazing instead. To find out the difference between the least grazed isolates (resistant isolates) and the highly grazed

group (susceptible isolates) we chose 5 isolates from least grazed group (denoted as LGG) and 5 isolates from highest grazed group (HGG) and made two groups to compare their genome data. We found that the two groups shared a core genome consisting of 3414 genes, while each group also has some unique genes they do not share. For the HGG groups there were 389 genes, and for the LGG there were 130 unique genes. It was interesting to know that the HGG group has a higher abundance of membrane protein, transporter protein, fimbrial and flagellar protein, toxin-antitoxin system related protein and autoinducer-2. However, the LGG has a high number of secretory system-II proteins compared to HGG, which has fewer Secretory system-III proteins.

 A recent study by Snyder et al. (2017) found that mutant strains of *E. coli* that are resistant to *D. discoideum* phagocytosis possess several genes related to flagella, oxidoreductase and acid resistance. These genes may have the potential to develop a mechanism to resist *D. discoideum* predation, which contributes to selection and maintenance of bacterial virulence factors against mammalian host. *Salmonella enterica* subsp. Typhimurium inhibits the *D. discoideum* starvation response through the type III secretion system there by preventing sporulation (Sillo et al., 2011). The type-III secretion system in HGG may also play a role in secreting substrate that may allow the starvation response of *D. discoideum*. Type –II secretion system occur in both pathogenic and non-pathogenic *E. coli,* and the output T2S secretory proteins can be a diverse group of toxins, degradative enzymes and other effector proteins. This system is clearly used by bacteria for environmental survival and virulence (Cianciotto and White, 2017). This report suggests that the T2S system may play an important role for LGG to resist predation. We also found that Autoinducer-2 related genes in HGG which are part of

quorum sensing system that allows communication with many different bacterial species (Federle, 2009). It has been also reported that functional quorum sensing is important for interaction of *Vibrio cholera* and the amoeba *Acanthamoeba castellanii.,* Upon phagocytized by the amoeba *Vibrio cholera* can resist intracellular killing (Van der Henst et al., 2015). The presence of autoinducer II in HGG indicates that the cells interact with *D. discoideum* to phagocytose. It may be possible that the cells are not completely killed, but form a symbiotic association with amoeba of farmer clones that carry bacteria through their social stages or dispersal stages and can be identified by the presence of bacteria in their sorus (Brock et al., 2011). It will be interesting to investigate the presence of *E. coli* cells in the sorus of *D. discoideum* that has grazed on HGG isolates.

Our study did not give any detailed information about the association of specific genes to *E. coli* survival from protozoan predation. The presence of genes unique to HGG and LGG may play a role in grazing susceptibility or grazing resistance of bacteria. To determine the role of these genes of protozoan predation more investigation is needed.

Our results of characterizing amoeba grazing on distinct *E. coli* isolates and a correlation between the presence of virulence genes and grazing resistance, deviate from previous reports (Adiba et al., 2010;Jenkins et al., 2011). These inconsistencies could easily be attributed to differences in amoeba clones, plating methods, nutrient conditions and the laboratory atmosphere. In our study we distinctly found that plating medium clearly affects the growth of amoeba clones on distinct *E. coli*.

In conclusion, our study clearly depicts that there is a difference in grazing susceptibility of *E. coli* isolates and in that environmental *E. coli* are found to be more resistance to

grazing. The highly grazed group and least grazed group possess unique genes that may play roles in their ability to be grazed or resist grazing.

5. References:

Adiba, S., Nizak, C., Van Baalen, M., Denamur, E., and Depaulis, F. (2010). From Grazing Resistance to Pathogenesis: The Coincidental Evolution of Virulence Factors. *PLOS ONE* 5, e11882.

Artz, R.R., and Killham, K. (2002). Survival of *Escherichia coli* O157:H7 in private drinking water wells: influences of protozoan grazing and elevated copper concentrations. *FEMS Microbiol Lett* 216, 117-122.

Avery, L.M., Williams, A.P., Killham, K., and Jones, D.L. (2008). Survival of *Escherichia coli* O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Sci Total Environ* 389, 378-385.

Barker, J., Humphrey, T.J., and Brown, M.W. (1999). Survival of *Escherichia coli* O157 in a soil protozoan: implications for disease. *FEMS Microbiol Lett* 173, 291-295.

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., and Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Research* 44, W22-W28.

Brandl, M.T. (2006). Fitness of human enteric pathogens on plants and implications for food safety. *Annu Rev Phytopathol* 44, 367-392.

Brandl, M.T., Rosenthal, B.M., Haxo, A.F., and Berk, S.G. (2005). Enhanced survival of *Salmonella enterica* in vesicles released by a soilborne *Tetrahymena* species. *Appl Environ Microbiol* 71, 1562-1569.

Brock, D.A., Douglas, T.E., Queller, D.C., and Strassmann, J.E. (2011). Primitive agriculture in a social amoeba. *Nature* 469, 393.

Byappanahalli, M.N., Richard, L.W., Shivelya, D.A., John, F., Ishii, S., and Sadowsky, M.J. (2007). Population structure of cladophora-borne *Escherichia coli* in nearshore water of lake Michigan. *Water Research* 41, 3649-3654.

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

Cianciotto, N.P., and White, R.C. (2017). The Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond. *Infection and Immunity*.

Disalvo, S., Brock, D.A., Smith, J., Queller, D.C., and Strassmann, J.E. (2014). In the social amoeba Dictyostelium discoideum, density, not farming status, determines predatory success on unpalatable *Escherichia coli*. *BMC Microbiology* 14, 328.

Federle, M.J. (2009). Autoinducer-2-Based Chemical Communication in Bacteria: Complexities of Interspecies Signaling. *Contributions to microbiology* 16, 18-32.

González, J.M., Iriberri, J., Egea, L., and Barcina, I. (1990). Differential Rates of Digestion of Bacteria by Freshwater and Marine Phagotrophic Protozoa. *Appl Environ Microbiol* 56, 1851-1857.

Hahn, M.W., and Hofle, M.G. (2001). Grazing of protozoa and its effect on populations of aquatic bacteria. *FEMS Microbiol Ecol* 35**,** 113-121.

Iriberri, J., Azúa, I., Labirua-Iturburu, A., Artolozaga, I., and Barcina, I. (1994). Differential elimination of enteric bacteria by protists in a freshwater system. *Journal of Applied Bacteriology* 77**,** 476-483.

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

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M.N., Yan, T., and Ishii, S. (2017). Environmental *Escherichia coli*: ecology and public health implications-a review. *J Appl Microbiol* 123**,** 570-581.

Jenkins, M.B., Fisher, D.S., Endale, D.M., and Adams, P. (2011). Comparative Die-off of *Escherichia coli* 0157:H7 and Fecal Indicator Bacteria in Pond Water. *Environmental Science & Technology* 45**,** 1853-1858.

Jiang, X., Morgan, J., and Doyle, M.P. (2002). Fate of *Escherichia coli* O157:H7 in Manure-Amended Soil. *Applied and Environmental Microbiology* 68**,** 2605-2609.

Jurgens, K., and Matz, C. (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie Van Leeuwenhoek* 81**,** 413-434.

Korajkic, A., Wanjugi, P., and Harwood, V.J. (2013). Indigenous Microbiota and Habitat Influence *Escherichia coli* Survival More than Sunlight in Simulated Aquatic Environments. *Applied and Environmental Microbiology* 79**,** 5329-5337.

Matz, C., and Kjelleberg, S. (2005). Off the hook--how bacteria survive protozoan grazing. *Trends Microbiol* 13**,** 302-307.

Moore, M.D., Withers, T.R., Yu, H., and Trzyna, W. (2016). A simple plating assay using Acanthamoeba to evaluate the effect of environmental stressors on growing and dividing cells. *Proceedings of the West Virginia Academy of Science* 88.

Nandakafle, G., Seale, T., Flint, T., Nepal, M., Venter, S.N., and Brozel, V.S. (2017). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. *Microbes and Environments* 32**,** 226-233.

R Development Core Team (2015). "R: A language and environment for statistical computing". (Vienna, Austria: R Foundation for Statistical Computing).

Ravva, S.V., Sarreal, C.Z., and Mandrell, R.E. (2010). Identification of Protozoa in Dairy Lagoon Wastewater that Consume *Escherichia coli* O157:H7 Preferentially. *PLOS ONE* 5**,** e15671.

Ravva, S.V., Sarreal, C.Z., and Mandrell, R.E. (2014). Strain Differences in Fitness of *Escherichia coli* O157:H7 to Resist Protozoan Predation and Survival in Soil. *PLoS ONE* 9**,** e102412.

Schmidt, C.E., Shringi, S., and Besser, T.E. (2016). Protozoan Predation of *Escherichia coli* O157:H7 Is Unaffected by the Carriage of Shiga Toxin-Encoding Bacteriophages. *PLOS ONE* 11**,** e0147270.

Semenov, A.V., Van Bruggen, A.H., Van Overbeek, L., Termorshuizen, A.J., and Semenov, A.M. (2007). Influence of temperature fluctuations on *Escherichia coli* O157:H7 and Salmonella enterica serovar Typhimurium in cow manure. *FEMS Microbiol Ecol* 60**,** 419-428.

Sillo, A., Matthias, J., Konertz, R., Bozzaro, S., and Eichinger, L. (2011). *Salmonella* Typhimurium is pathogenic for Dictyostelium cells and subverts the starvation response. *Cell Microbiol* 13**,** 1793-1811.

Smith, D.D.N., Nickzad, A., Déziel, E., and Stavrinides, J. (2016). A Novel Glycolipid Biosurfactant Confers Grazing Resistance upon *Pantoea ananatis* BRT175 against the Social Amoeba *Dictyostelium discoideum*. *mSphere* 1.

Smith, J., Queller, D.C., and Strassmann, J.E. (2014). Fruiting bodies of the social amoeba *Dictyostelium discoideum* increase spore transport by *Drosophila*. *BMC Evolutionary Biology* 14**,** 105.

Snyder, M., Weichseldorfer, M., Lane, J., Kazmierski, R., Bains, H., White, T., Parker, D., Aklkharouf, N., and Hemm, M. (2017). Identification of Virulence and Virulence Regulatory Factors in *Escherichia coli* using *Dictyostelium discoideum* as a Model Phagocyte. *The Journal of Immunology* 198**,** 131.119.

Steinberg, K.M., and Levin, B.R. (2007). Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. *Proc Biol Sci* 274**,** 1921-1929.

Tezcan-Merdol, D., Ljungström, M., Winiecka-Krusnell, J., Linder, E., Engstrand, L., and Rhen, M. (2004). Uptake and Replication of *Salmonella enterica* in *Acanthamoeba rhysodes*. *Applied and Environmental Microbiology* 70**,** 3706-3714.

Van Der Henst, C., Scrignari, T., Maclachlan, C., and Blokesch, M. (2015). An intracellular replication niche for *Vibrio cholerae* in the amoeba *Acanthamoeba castellanii*. *The Isme Journal* 10**,** 897.

Van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5**,** 173-183.

Weekers, P.H.H., Bodelier, P.L.E., Wijen, J.P.H., and Vogels, G.D. (1993). Effects of Grazing by the Free-Living Soil Amoebae *Acanthamoeba castellanii*, *Acanthamoeba polyphaga*, and *Hartmannella vermiformis* on Various Bacteria. *Applied and Environmental Microbiology* 59**,** 2317-2319.

Wildschutte, H., Wolfe, D.M., Tamewitz, A., and Lawrence, J.G. (2004). Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proceedings of the National Academy of Sciences of the United States of America* 101**,** 10644-10649.

y = -0.0568x + 1.1398
R² = 0.00726
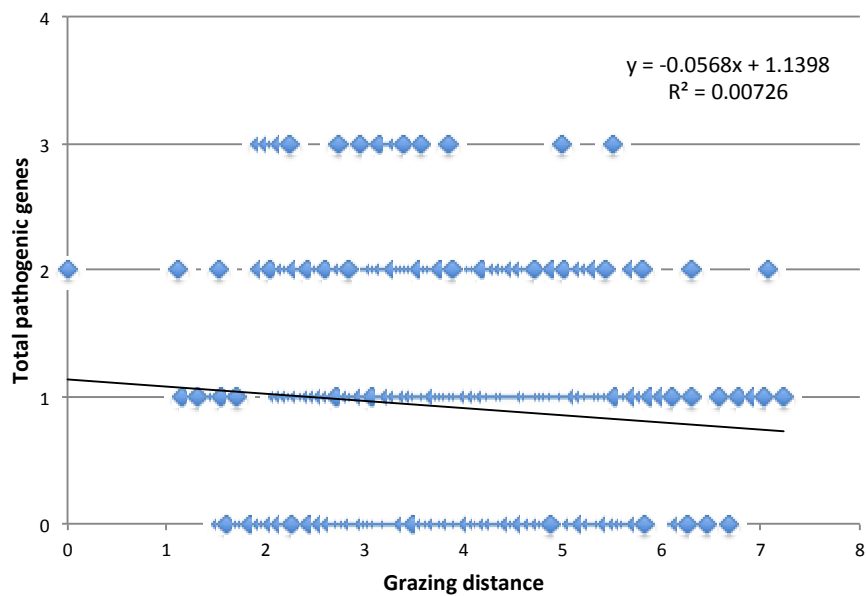
Fig S6. 1 Correlation between grazing distance and presence of pathogenic genes

Chapter 7: Comparative Analysis of Whole Genome Sequences of

Escherichia coli Isolated from Pasture Soil

The work reported in this chapter was conducted by Gitanjali NandaKafle in collaboration with Dakota York and Joy Scaria

1. Introduction:

Escherichia coli is a commensal bacterium found in the lower intestine of warm-blooded animals, but it can also be pathogenic in nature causing serious illness to humans and domestic animals (Kaper et al., 2004;Croxen and Finlay, 2010;Tenaillon et al., 2010). The primary habitat of *E. coli* is assumed to be within the animal host and to thrive in such a diverse nutrient rich habitat the bacterium needs to be genetically adapted (Winfield and Groisman, 2003). Despite the intense competition within the densely populated gut ecosystem**,** *E. coli* is the most abundant facultative anaerobe species in human intestine (Dubreuil, 2012). *E. coli* also often enters the environment (considered as secondary habitat) through fecal deposition, where until recently it assumed to survive for a short period of time (Ishii and Sadowsky, 2008). In recent years wide spread reports on *E. coli* survival and growth in soil, water, sediments, water plants in tropical, sub-tropical and temperate regions have been reported (Byappanahalli and Fujioka, 1998;Solo-Gabriele et al., 2000;Desmarais et al., 2002;Whitman et al., 2006;Ishii et al., 2007;Texier et al., 2008;Brennan et al., 2010b). The considerable disparity in conditions out in the environment**,** and in the primary host habitat**,** raises the question of how the

organism survives and competes for niche spaces in the outside environment. Some reports indicate the naturalized *E. coli* form genetically distinct populations with DNA finger prints different from that of the primary host isolates (Gordon et al., 2002;Walk et al., 2007;Texier et al., 2008). *E. coli* is highly versatile in its ability to adapt in multiple environments, which justifies its immense diversity within species (Bergthorsson and Ochman, 1998). Recently, strains phenotypically indistinguishable but genetically distinct from *E. coli* have been reported and named *Escherichia* clades I to V (Walk et al., 2009;Luo et al., 2011). A multiplex PCR based method that enables strains of *E. coli* to be assigned to a phylogroup based on the presence and absence of genes (*yjaA*, *arpA*, *chuA*, TSPE4.C2 and *trpA*) and *E. coli* strains can be classified into one of the seven phylogroups denoted as A, B1, B2, C,D, E, F and recently, Clade-I is considered as the eighth phylogroups (Clermont et al., 2013). Although, the distribution of phylogroups depends on the diet or geographic conditions, strains belonging to phylogroup A and B1 are highly adapted to humans and vertebrate animals, the A phylogroups being predominant in humans and B1 strains in animals (Duriez et al., 2001;Gordon and Cowling, 2003;Skurnik et al., 2008). Some strains from phylogroup B1 were found to persist in water and soil environment (Walk et al., 2007;Ratajczak et al., 2010;NandaKafle et al., 2017). The new phylogroup assignment method has not been used extensively, however human fecal isolates screened using this method demonstrated that about 13% of *E. coli* isolates belong to newly described phylogroups C, E, F and Clade-I (Clermont et al., 2013). In a recent study phylogroup B1 and E were found to be the major groups in *E. coli* isolated from cattle (Morcatti Coura et al., 2015). Our previous work also showed similar results with the majority of cattle isolates belonging to

phylogroup E and environmental isolates belonging to phylogroup B1(NandaKafle et al., 2017). Although *E. coli* O157:H7 belongs to phylogroup E, very few strains of this group has been recorded so far.

"The ability to identify clonal /clonal complex /phylogroups is crucial as a strains ecological niche, life style and propensity to cause disease vary with its phylogenetic origin" (Clermont et al., 2015). There are currently three multi locus sequence type (MLST) schemes; Michigan State University (http://www.shigatox.net/ecmlst/cgi-bin/index), Warwick Medical School (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli), and Pasteur Institute (http://www.pasteur.fr/recherche/genopole/PF8/mlst/EColi.html) in use to assign a sequence type (ST) to phylogroups. The Warwick scheme is largely being used for ST nomenclature.

Analysis of sequenced *E. coli* genomes shows a broad variability of their size and genes content, more importantly, sequencing has revealed the extraordinary flexibility and dynamics of the *E. coli* genome that contribute to its phenotypic and genotypic diversity. Comparative genome analysis of bacterial genomes of same species suggested that bacterial species can be characterized by pan-genome which consist of the core genome (genes common among all strains), a dispensible genome or accessory genome (genes that are present in subset of strains) and the unique genes (specific to strain)(Tettelin et al., 2005). The popular lab strains K-12, O157:H7, and the uropathogenic strain CFT073 share only 39% of their genes (Welch et al., 2002). Subsequent sequencing of more genomes reduced the core genome to less than 20% of the pangenome of 16000 genes or more (Lukjancenko et al., 2010;Kaas et al., 2012). Sixty-one publicly available genome sequences of *E. coli* strains were compared and it was found that 20% of the genes

belong to the core and the remaining 80% were not found in all other *E. coli* genomes (Lukjancenko et al., 2010). The size of the *E. coli* genome can vary greatly among strains. The standard laboratory strains have genomes of ~4.5 million base pairs with about 4000 genes, while pathogenic strains can have genomes of over 5.9 million base pairs with 5500 genes (Lukjancenko et al., 2010;de Muinck et al., 2013). Extensive gene acquisition and loss has led to lineages that differ in their abilities to use diverse metabolites and survive in harsh environments, and their potential to be pathogenic, and resistant to various antibiotics. Such mobile genetic modules along with point mutations facilitate the rapid adaptation of *E. coli* to changing environments (Brzuszkiewicz et al., 2009;Dobrindt et al., 2010).

E. coli displayed phenotypic characteristics which may increase their ability to survive and compete for resources in low temperature soil (Brennan et al., 2010a). The versatile nature of nutrient utilization of this facultatively anaerobic bacterium has also been suggested to contribute to survival and growth in the environment (Ishii and Sadowsky, 2008;O'Reilly et al., 2010). Indeed, the ability to metabolize a wide array of carbon, nitrogen, phosphorus and sulphur sources would represent a significant advantage under nutrient-limiting conditions, typical of *E. coli* secondary habitat (Durso et al., 2004).

Our goal in this study was to understand the genetic diversity of -20 *E. coli* isolates (of phylogroup A (2 isolate), B1(8 isolate), E (9 isolate) and B2 (1 isolate)) collected from environmental samples and their comparison with other reference strains (disease associated E. coli O157:H7, environmental clade-I -TW10509 and laboratory strain K12 MG1655), and to characterize their ability to metabolize various C, N, P and S sources.

2. Materials ad Methods

2.1. Selection of Escherichia coli isolates:

Twenty isolates were selected from three clusters of *E. coli* environmental, bovine and mixed clusters (NandaKafle et al., 2017) of Phylogroup B1, E, A and B2 for whole genome sequencing and phenotypic microarray assay, BIOLOG.

2.2. Whole genome sequencing of Escherichia coli isolates:

Genomic DNA was extracted from overnight LB agar cultures suspended in 10 mM phosphate buffer (pH 7.0) using the genomic DNA Quick Prep Kit (Zymo Research), and all extracted DNA samples were quantified using Nanodrop Spectrophotometer as well as Quibit Fluorometer. The DNA samples were sent to Microbes NG, UK for sequencing (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). The protocol used for sequencing is briefly explained; the Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: two nanograms of DNA instead of one were used as input, and PCR elongation time was increased to 1 min from 30 seconds. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cutoff of Q15 (Bolger et al., 2014). De novo assembly was performed on samples using SPAdes version 3.7 (Bankevich et al., 2012), and contigs were annotated using Prokka 1.11(Seemann, 2014)

2.3. *BIOLOG Phenotypic Microarray (PM):*

We performed the BIOLOG phenotypic microarray (BIOLOG, Hayward, CA, USA) according to the manufacturer's recommendations. The Biolog microplate assay is based on measurement of bacterial respiration which produces NADH (Bochner et al., 2001). If *E. coli* metabolizes a specific substrate, electrons from NADH reduce a tetrazolium dye in an irreversible reaction generating a purple color in the PM plates well. The isolates were tested with the 96 well plates PM1 to PM4, containing 190 carbon (PM1 and PM2), 95 nitrogen (PM3), 59 phosphorous (PM4) and 35 sulfur (PM4) substrates. Bacterial colonies obtained after 18 h at 37°C on BUG-B agar plates were suspended in the inoculating fluid from BIOLOG, each well of the plate was inoculated with 100 µL of cell suspension. The plates were incubated at 37°C for 24 h. Following incubation the metabolism of various substrate was measured spectrophotometrically at 590 nm, using a microtiter plate reader. The color intensities of the wells were normalized against that of the negative control well, combined with the visual increase in turbidity. Plates were scored in binary fashion as either positive or negative (Classen et al., 2003).

2.4. *Data Analysis:*

The BIOLOG data were normalized and arranged based on their positive and negative substrate utilization (positive was denoted as 1 and negative utilization was denoted as 0). Genome data were also arranged based on their presence and absence of genes in the pangenomes. For both pangenome and BIOLOG results heatmaps were created using the heatmap.2 function from the package: "gplots" in R (R Development Core Team, 2015) to observe the similarity among isolates.

A comparative analysis of 20 pasture isolates with reference strains *E. coli* O157 EDL933D, K-12 MG1655 and Clade-I (TW10509) were conducted using EDGAR 2.0 platform (Blom et al., 2016). The pan and core genome development plot, neighbor-joining tree using the core genes of 23 isolates were constructed.

2.5. Multi locus Sequence Type (MLST):

Multi locus sequence typing of 20 isolates was performed by uploading the whole genome assembled contigs in the database https://cge.cbs.dtu.dk/services/. Two different MLST schemes (Achtman or Warwick scheme as MLST #1 and Pasteur scheme as MLST # 2) were used to identify the sequence types of isolates by using MLST 1.8 (Larsen et al., 2012). For Achtman (Warwick) scheme internal fragments of the seven house keeping genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA* (Wirth et al., 2006) and for Pasteur scheme eight house keeping genes *dinB*, *icdA*, *padB*, *polB*, *putP*, *trpA*, *trpB*, *uidA* (Jaureguy et al., 2008)were used.

2.6. Virulence Factors:

Virulence genes of *E. coli* isolates were determined by uploading the assembled genome sequences in the CGE tool for virulence finder database (Joensen et al., 2014). It is possible to select configurations for the organism of interest, and in addition, it is possible to select percent identity (%ID) threshold between the input and the best matching database gene. The output consists of best-matching genes from BLAST analysis of the selected database, against the submitted genome, with genes set to cover a minimum of

three-fifths of the length of the database genes. The output contains information on the virulence gene, the %identity, the length of query and database gene, the position of the hit in the contig, and the accession number of the hit.

3. Results

3.1. Whole genome sequence analysis

A total of 23 isolates (2 isolates of phylogroup A, 8 of phylogroup B1, 1 isolate of phylogroup B2, 9 isolates of phylogroup E and reference strains K12, 933D and TW10509) whole genome sequences were analyzed to understand the similarity and diversity of various isolates of diiferent phylogroups. The total numbers of core genes for all 20 isolates were comprised of only 33.8%, the accessory or indispensible genes were of 46.6% and the unique genes were 19.6% (Fig.1). Each genome had more than 20% genes of unidentified function named as hypothetical genes (Table. 2). The average numbers of genes of 20 isolates were of 4,948, where as the E phylogroup isolates were consist of 5,267 genes and both phylogroup A and B1 together the average genes were 4,638 (Table. 2). The higher genome size of phylogroup E isolates suggesting a great diversity among these isolates.
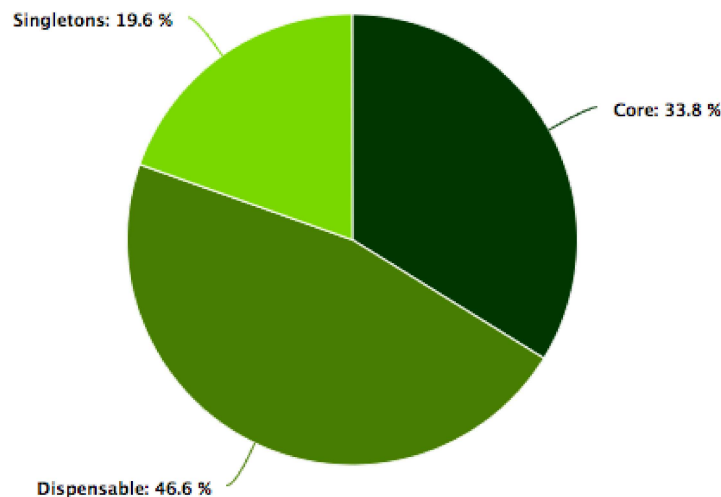
Fig. 7. 1 Genomic subset distribution of 20 isolates. Core gene consist of 33.8%, dispensable or accessory genes 46.6% and singletons or unique genes are of 19.6%

3.2. Pan-genome and core genome analysis

The neighbor joining tree constructed by 3,238 core genes sequence alignment showed that the Clade-I and B2 phylogroup isolates are very different from other isolates in the group (Fig. 2). Two environmental isolates (NandaKafle et al., 2017) of E phylogroup formed very distinct cluster whereas two E isolates clustered together with B1 and A suggesting a very diverse group of E phylogroups (Fig. 2). The reference strain 933D and K12 did not cluster with any E or A phylogroup isolates. But, 933D was closer to E and K12 was closer to A phylogroup isolates (Fig.2). The heat map constructed based on the presence and absence of pangenes showing the similarity among isolates (Fig. 3). Based on presence and absence of pangenes the three isolates clade-I, 933D and the B2 isolates made a separate cluster, however all three were very different from each other. Phylogroup A and B1 isolates were very similar genotype compared to E isolates.
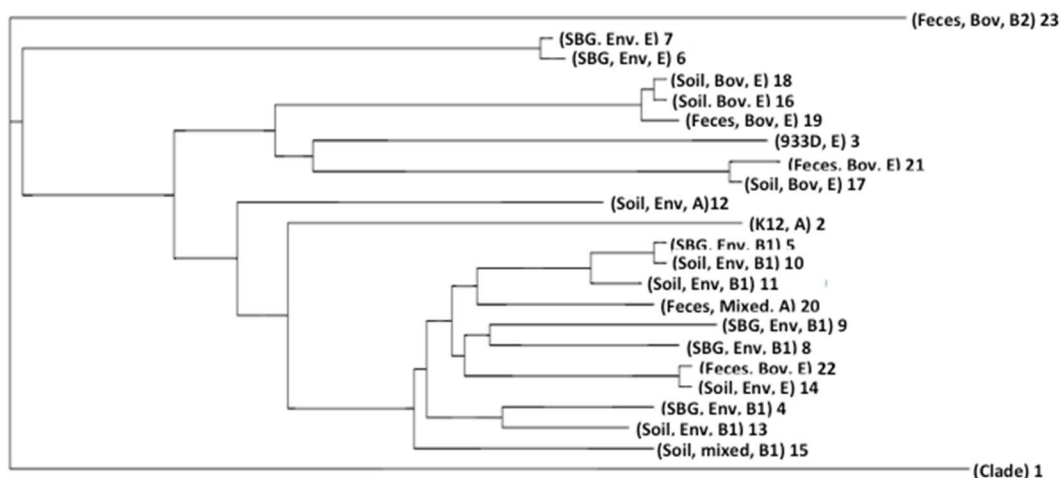
Fig. 7. 2 Neighbor joining tree constructed by EDGAR using sequences of core genome of 20 pasture isolates and 3 reference genomes; Clade-I, *E. coli* O157:H7 933D and *E. coli* K-12 MG1655

However, in the E phylogroups, 3 isolates clustered together at a closer distance from reference strains EDL933D and TW10509. Other six isolates make three clusters each of 2 isolates (7-8, 9-10, 11-12), those two isolates of each cluster appeared to be very similar (Fig. 3). These E isolate clusters are closer to A and B1 clusters. The core genome tree and pan-genome tree showed similarity with respect to the clustering of isolates, but the reference strains 933D in the core genome tree clustered with E isolates but in pan-genome tree it was distinctly clustered from E isolates, indicating that the accessory genes of reference strains were very different. The reference strain K-12 was clustered with phylogroup A isolates in pan-tree whereas in core –tree it was not clustered with any A isolate.
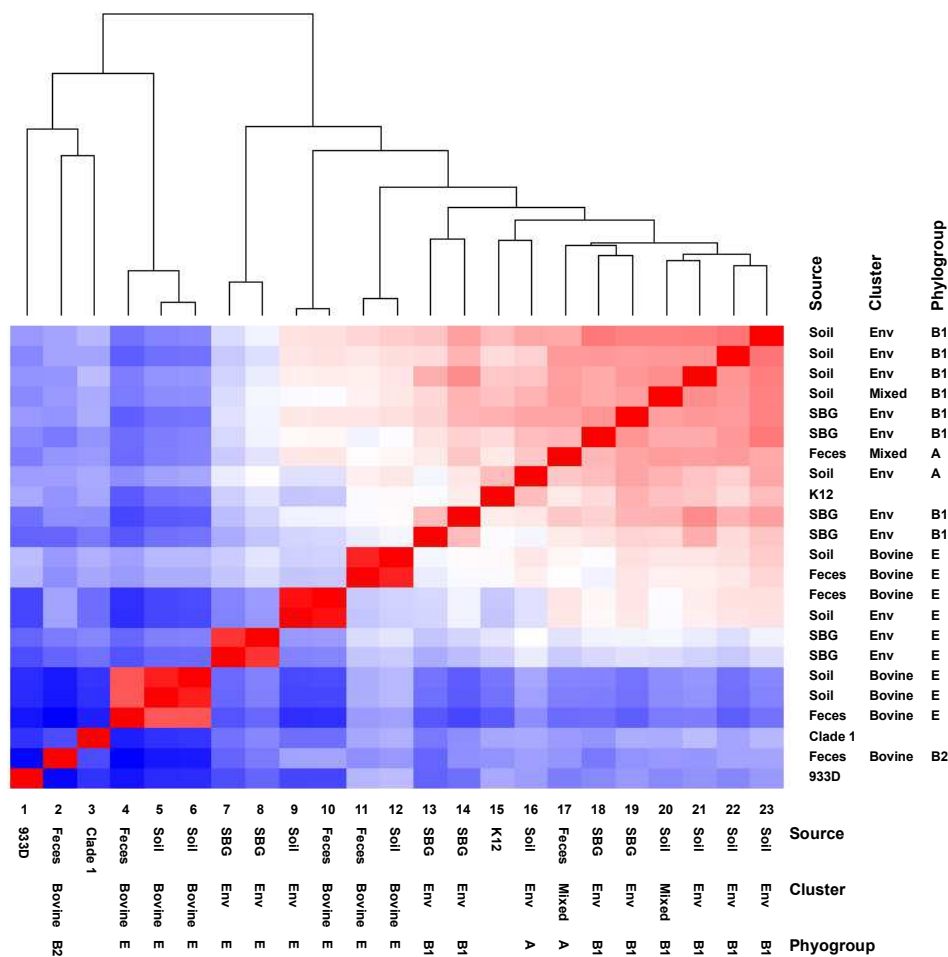
Fig. 7. 3 Heat map of total genome composition of 23 isolates, based on the presence and absence of genes. Red color represents high similarity and blue color less similarity

3.3. *Core-pan genome development plot*

The contribution of each genome to complete pan-genome of *E. coli* is demonstrated in Fig. 4 where the pan-genome and core-genome was plotted for 20 isolates including three reference strains (K-12, EDL 933D and TW10509). The size of pan- and core genome for different groups of isolates was summarized in Table-1. For all 20 isolates and 3

reference strains the pangenome was 11,550 and the core genome was 3,238. When the Clade-I genome was removed the pangenome reduced to 10,939 and core genome did not change a lot, it was only increased to 3,251. This suggested that the Clade-I has contributed 611 genes in the pan-genome curve, which are unique to this strain. The pan-core plot for only 20 isolates showed pangenome size of 9,946 and the core genome size was 3,362. So, the three reference strains contributed about 1600 genes suggesting our isolates are very different from the reference strains based on their pan genome. To find out the contribution of phylogroup A, B1 and E we plotted two pan-core development curves one with phylgroup E isolates and 933D and other with phylogroup A, B1 and K12 together. The pangenome for phylogorup E isolates together with 933D was 8,746, which was 1,754 genes more than the pan-genome of A, B1 phylogroup together. This result also suggests that E phylogroup isolates are very diverse in nature, with larger genome size than B1 and A (Table.2).
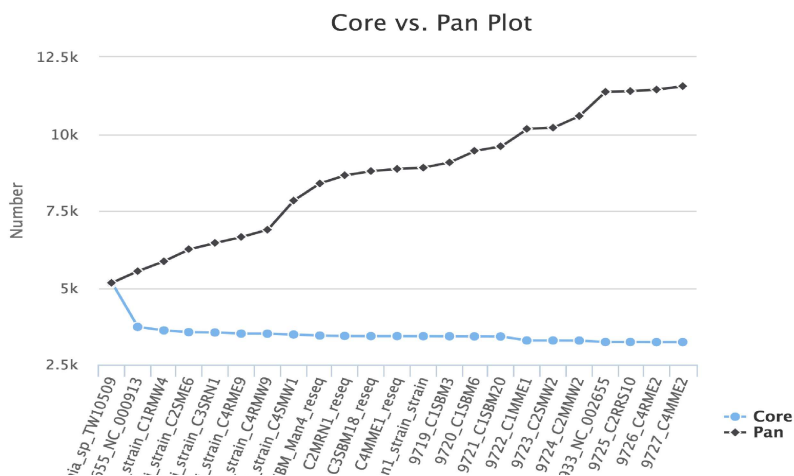

Fig. 7. 4 Pan-and core genome plot of 23 isolates. The black pan-genome curve represents the cumulative number of gene families present in the total genomes and the blue core genome curve represent the conserved number of gene families

Table 7. 1 list of pan-genes and core genes while comparing the whole genome of different groups of isolates

Group of isolates	Core gene	Pan genes
20 isolates with K12, 933D and TW10509	3,238	11,550
20 isolates with 933D and K12	3,251	10,939
20 isolates only	3,362	9,946
Phylogroup E isolates with 933D	3,575	8,746
Phylogroup A and B with K12	3,578	6,992

3.4. Sequence typing

All *E. coli* isolates were sequence typed by two MLST scheme. According to the MLST # 1 scheme, out of 20 isolates analyzed 18 were assigned to different sequence types (STs) and two isolates could not be assigned to any available STs. Two isolates from phylogroup B1 was belonged to ST294, whereas from phylogroup E, ST392 was assigned to 2, ST6645 was assigned 3 and ST392 was assigned to 2 isolates indicating the presence of clonal isolates. Other nine isolates were assigned to different STs. Phylogroup B2 isolate was assigned to ST95, which is same as a reference strain *E. coli* S88 (Clermont et al., 2015). According to MLST # 2 scheme only seven isolates were assigned to sequence type and 13 isolates were remained unassigned. Two isolates were assigned with ST295 other five sequence types were ST654, ST339, ST303, ST363 and ST1. The isolate with phylogroup B2 was assigned with ST1 (by MLST#2) and ST95 (by MLST # 1) it is consistent with the result previously reported (Clermont et al., 2015).

3.5. *Virulence genes*

The presence of virulence related genes were identified in 20 isolates. A total of 10

virulence related genes were detected through out 20 isolates. The list of virulence related

genes were given in Table. 2.

Table 7. 2 Detail information on genomes of 20 *E. coli* isolates

Isolates	MLST #1	MLST #2	Phylo	Virulence related genes in genome	Total genes	Unique genes	ABR	RDAR	Biofilm LB-SE	Hypoth genes
C4RME9	ST -1850	ST-Unk	A	gad, lpfA	4408	77	0	3	1, 3	1207
C4MME2	ST-2713	ST-654	A	lpfA, gad	4774	60	1	3	2, 1	1264
C1RMW4	ST-201	ST-294	B1	iss, gad, lpfA	4530	26	0	1	1, 0	1266
C1SBM3	ST-201	ST-294	B1	lpfA, gad, astA	4677	78	0	3	0, 0	1156
C1SBM6	ST-388	ST-339	B1	gad, lpfA	4953	110	0	2	3, 1	1407
C1SBM20	ST-164	ST-303	B1	astA, iss, gad, lpfA	4774	67	0	3	1, 1	1277
C3SBM18	ST-937	ST-363	B1	lpfA, gad	4424	68	0	3	1, 2	697
C3SRN1	ST-111	ST-Unk	B1	lpfA,gad,iss	4529	109	2	0	0, 1	1234
C4RME2	ST-906	ST- unk	B1	lpfA, gad, iss	4601	26	0	1	1, 0	1095
C4RMW9	ST-162	ST-Unk	B1	gad, lpfA, iss	4710	56	3	3	0, 1	1317
C2MRN1	ST-6645	ST- Unk	E	lpfA, gad, air, eilA	5764	45	1	0	1, 2	1515
C2RRS10	ST-392	ST-Unk	E	iss, gad, lpfA, ironN	5298	10	0	3	2, 1	1637
C2MMW2	ST-392	ST-Unk	E	iss, gad, lpfA, ironN	5274	3	0	3	0, 0	1612
C2SME6	ST-753	ST-Unk	E	gad, iss, air, eilA	4780	8	3	0	2, 2	1387
C4MME1	ST-753	ST-Unk	E	iss, air, eilA	4821	22	1	1	2, 1	855
C2SMW2	ST-6645	ST-Unk	E	gad, eilA, air	5644	7	1	0	0, 0	1932
C4SMW1	ST-6645	ST-Unk	E	gad, eilA, lpfA, celB	5679	13	1	1	0, 1	1944
SBM_Man1	ST-Unk*	ST- Unk	E	air, gad, astA, eilA	4999	78	1	3	0, 1	1409
SBM_Man4	ST-Unk*	ST- Unk	E	air, gad, astA, eilA	5143	31	1	3	1, 1	1065
C1MME1	ST-95	ST-1	B2	iss, gad, vat, cdtB	5171	306	1	0	0, 0	1578
*The sequence type is identical										

The distribution of these genes were differ from strain to strain, *iss* (increased serum

survival) was present in 12 isolates, *gad* (glutamate decarboxylase) was present in all 20

isolates, vat (vacuolating autotransporter toxin) and *cdtB* (cytolethal destending toxin)

was present in only in B2 isolate, *ipfA* (long polar fimbriae) was present in 14 isolates (all

B1 and A isolates and 4 E isolates), *astA* (EAST-1 heat stable toxin) was present four

isolates 2 B1 and 2 E phlogroup, *eilA* (*Salmonella hilA* homolog) was present in 7

isolates (all 7 were of phylogroup E), *air* (enteroaggregative immunoglobulin repeat

protein) was present in 6 isolates (all 6 are phylogroup E), *ironN* (enterobactin

siderophore receptor protein) was present in only 2 isolates (both are of E phylogroups),

celB (endonuclease colicin E2) was present in one phylogroup E isolate. The presence

more virulence related genes in phylogroup E isolates suggested that phylogroup E

isolates may have the potential to be more pathogenic in comparison with A and B1.

3.6. *Phenotypic microarray or BIOLOG Assay*

The results of phenotypic microarray assay for nutrient utilization of 20 isolates and 3

reference strains were very different from their genotypes. The utilization C, N, P and S

were different from strain to strain even though they are genotypically similar. The

nutrient utilization pattern of Clade I and 933D strains was very similar to other isolates,

however the K-12 strain was very different (Fig. 5). Phylogroup E, B1 and A did not

show any distinct pattern in their nutrient utilization. Some of them are similar with

Clade-I and 933D whereas some strains of phylogroup B1 and E did not show any

similarity in their nutrient utilization pattern. This result suggested that isolates with

similar genotypes might have different gene functionality. The utilization of sulfur and

phosphorous was maximum compared to carbon and nitrogen sources (Fig. 6). However

one isolate showed about 40% P and S utilization but more than 60% carbon utilization,
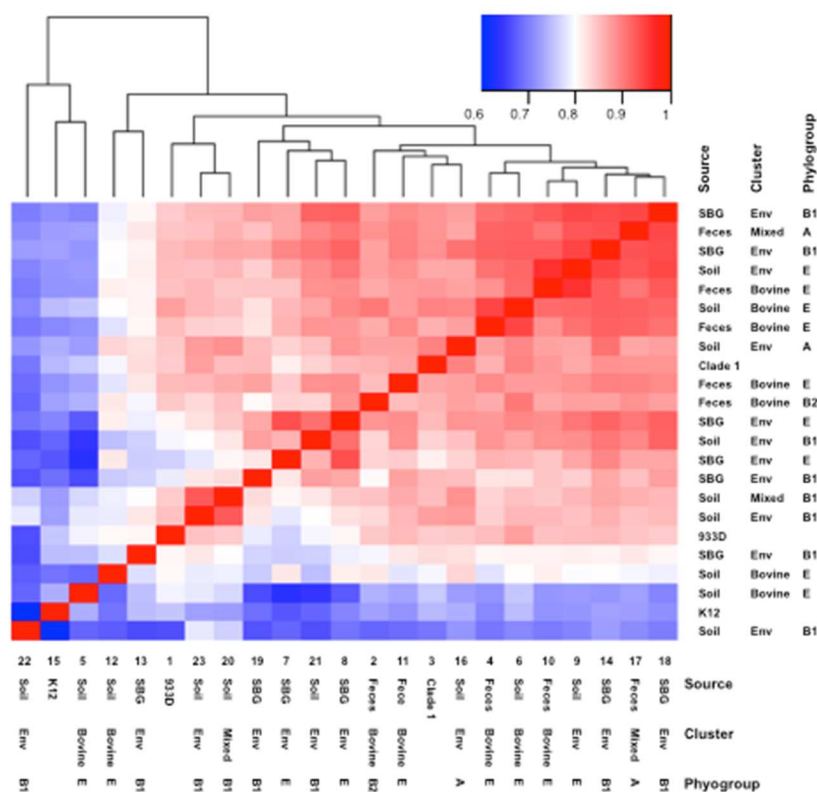
Fig. 7. 5 Heat map of substrate utilization profiles evaluated using BIOLOG Pheneplates 1 – 4. Red represents high similarity and blue is less similarity.

which is very different from other isolates. Our results also showed that many of the C, N, P and S sources were utilized by more than 20 isolates where as some of the sources were not being utilized by any of the isolates. Out of 190 C sources 59 sources were utilized by more than 20 isolates and 21 sources were not utilized by any of the isolates, out of 95 nitrogen sources 15 sources were utilized by more than 20 isolates and 7 sources were not utilized by any of the isolates. For phosphorus out of 59 sources 47 nutrient sources were utilized by more than 20 isolates and only one source was not in use by any of the isolates. In case of sulfur as nutrient source only 19 sources were

utilized by more than 20 isolates and all other nutrient sources were utilized my at least one of the isolates.
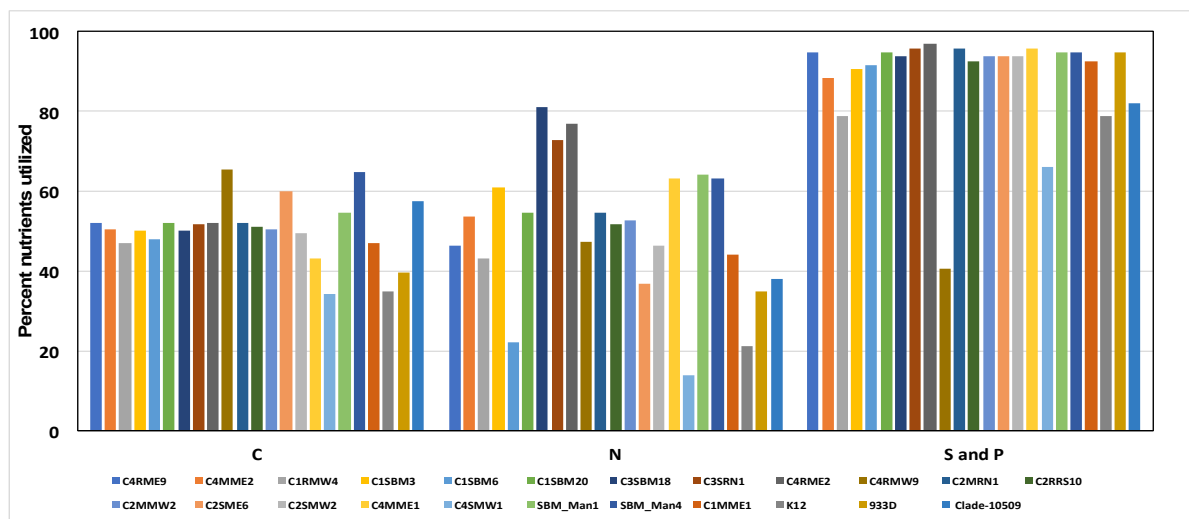
Fig. 7. 6 Nutrient utilization pattern of 23 isolates as determined using BIOLOG Pheneplates 1 – 4.

4. Discussion

In this study we sequenced and analyzed 20 soil isolates from a pasture together with 3 reference genomes. The phylogenetic analysis of using the core genes of 20 genomes with three reference strains revealed an interesting pattern of clustering of isolates. Comparison of 20 *E. coli* from the pasture soil including bovine feces and environmental isolates (NandaKafle et al., 2017) revealed that 3,362 genes were conserved in all isolates, and the pan-genome contained a reservoir of about 10,000 gene pool (Table. 1) However, when reference strains were included the gene pool went up to 11,500. In a previous pan-genome study of 17 pathogenic and commensal *E. coli* isolates the core

genome was found to be 2,200 whereas the pan-genome was consist of more than 13,000 genes (Rasko et al., 2008). In a genome comparison study of 61 *E. coli* strains showed that the predicted pangenome was comprising of about 16,000 genes and the core genome was about 1000 genes, which is one fifth of the typical *E. coli* genome. Many of the accessory genes making more than 90% of the pangemone are four fifth of the typical genome and are often found to be co localized in the genomic islands (Lukjancenko et al., 2010).

This result suggests that the pathogenic and commensal *E. coli* genomic diversity represent an open pan-genome model. The size of the pan-genome is greatly depends on the existing balance between gain and loss activities. When the bacteria are in very diverse environment together in a group gaining of genes are common and the genome size is strongly related to the selfish genes that is parasitic and constitute the mobilome (den Bakker et al., 2010;Rouli et al., 2015). The E phylogroup isolates had a larger pangenome pool compared to phylogroup B1 and A together suggesting that E isolates may harbor more virulent genes. The virulence gene analysis of all isolates showed that E isolates have more virulence related genes compared A and B1 (Table. 2). Phylogenetic analysis based on core genome sequence alignment, the isolates clustered differently from the pangenome tree. The pangenome tree portrayed two separate clusters of one with E phylogroups and one with A and B1. Phylogroup B2 and 933D and CladeI isolates were outlier (Fig. 3). This result indicates that pan-genome phylogeny delineate the phylogroup distribution more clearly than core. We also noticed that the B1 and A phylogroups are less diverse compare to E phylogroups as there was less than 75% similarity among the strains even though they clustered together. The diversity within the

species of *E. coli* due to the accessory gene content may allow them to adapt in various environments.

The survival of *E. coli* in low nutrient and highly competitive conditions depends on its ability to efficiently utilize diverse nutrients (van Elsas et al., 2011). The carbon utilization pattern of *E. coli* K12 demonstrated that this organism has a physiological adaptation to use mixed carbon and energy substrate rather than single substrates as used in BIOLOG plates (Ihssen and Egli, 2005). This could be a strategy of *E. coli* to maintain population under nutrient limiting conditions by scavenging trace amount of nutrients (Ihssen and Egli, 2005;Franchini and Egli, 2006). Most of the nutrient profiling studies to investigate metabolic flexibility of *E. coli* have conducted with pathogenic and commensal strains and mainly focused on carbon utilization (Durso et al., 2004;Ihssen and Egli, 2005;Franz et al., 2011;Xavier et al., 2014). The physiology of *E. coli* that survives in soil remains poorly understood; there is a need to investigate ecological characteristics of *E. coli* in soil. Assessment of nutrient utilization profile using the BIOLOG system can provide some insight into the nutrient flexibility of this organism. The nutrient utilization profile of environmental persistence *E. coli* was found to be highly versatile with the ability to use at 15 °C, 34%, 32%, 78% and 46% of C, N, P, and S respectively and it also maintained the same fitness with a better utilization of these substrates at 37°C, suggesting no trade-off in thermal tolerance of cold adapted environmental isolates (Brennan et al., 2013). Overall in this study, all isolates from different sources of diverse phylogroups were found to be highly versatile, with the ability of 20 strains (except 933D, K12 and one soil isolate) to use more than 40% of carbon substrates. The nitrogen utilization pattern was very diverse among the isolates

whereas the P and S utilization was more than 80% by all the isolates except two soil isolates (Fig. 6). Based on the nutrient utilization profile *E. coli* isolates clustered very differently from their genotype, Clade I and 933D reference strains cluster together with other isolates whereas K12 showed a very different pattern of nutrient utilization (Fig.5). Our results was supported by the previous studies showing environmental *E. coli* strains that were different from enteric strains genotypically, behaved very similar phenotypically (Luo et al., 2011). A study on pan and core metabolism was performed on 29 *E. coli* strains species (Vieira et al., 2011) which resulted a 1545 pan-metabolism reaction including 885 core. It was detected that the proportion of core gene and the nature of pan-genome did not reflect the pan-metabolism distribution. The carbon utilization of K-12, 933D and a pasture isolate were less than 40% but all other isolates oxidized more substrate. Indicating that 933D and K-12 were metabolically less diverse based on their carbon utilization. However 933D utilized 5% more carbon substrate compare to K-12 which contradict the previous report explaining commensal strains oxidize more substrate compare to *E. coli* O157:H7 strain (Durso et al., 2004). However, to depict the correlation between genomic and phenotypic microarray based data based on the presence and absence of genes is not reasonable as the phenotypic diversity of cell is also affected by mechanisms as for example regulatory and signal transduction systems and membrane transporter whose functionality could not be directly detected by Phenotypic microarray (PM) system or genome annotation. Further implementation of transcriptome, proteome and metabolomics data is necessary to understand the genotype and phenotypic relatedness of organisms.

The sequence types (ST) for all isolates were determined based on two MLST schemes. As per MLST #1 all isolates were assigned as some STs but two were unknown, this suggests that these two isolates from E phylogroups may be new STs not listed in the collection of strains of MLST#1 scheme. Only 7 isolates were assigned with STs based on MLST # 2 scheme (Table. 2) indicating that this scheme has not been listed enough isolates (Clermont et al., 2015). Interestingly, the phylogroup B2 strain was assigned as ST-95 based on MLST # 1, and ST1 as per MLST#2. These STs matches with the *E. coli* reference (ECOR) strain S88 belongs to phylogroupB2 (Clermont et al., 2015). This strain also harbor virulence related genes *iss*, *gad*, *vat* and *cdt.* It has also been reported that *E. coli* O18:K1:H7 with sequence type ST95 and phylogroup B2 was found to be the cause of neonatal sepsis in Barcelona (Sáez-López et al., 2017). These results indicate that potential pathogenic strains can be found in the soil environment without any fecal contamination. Our results also showed that phylogroup E had higher numbers of virulence related genes compared to phylogroup B1 and A, indicating E isolates could be potential pathogens. Because of many of the initial phylogenetic studies are based on initial rapid determination scheme proposed by Clermont et al. (2000), which only allow to distinguish only among the four mail phylogroups (A, B1, B2, and D). So fewer number of phylogroup E strains may have been reported previously. Phylogroup E isolates were part of   phylogroup D before the Clermont modified multiplex PCR method (Clermont et al., 2013). It has been reported that pathogenic strains causing extra-intestinal diseases mainly belong to phylogroup B2 and D (Picard et al., 1999;Chakraborty et al., 2015), so it is possible that phylogroup E isolates have the potential to be pathogens.

In conclusion, *E. coli* isolates from pasture soil and bovine feces belong to various phylogroups were very diverse in their genotype and phenotype. The pangenome analyses of isolates distribute the phylogroups more precisely than the core genomes. We found isolates of phylogroup E that may represent some new sequence types. Some of the E phylogroup isolates may be potential pathogens present in the environment. The isolate belong to phylogroup B2 we isolated from the bovine feces in pasture soil was a pathogenic strains with *cdtB* and *vat* virulene factors. These findings indicate a regular need of monitoring *E. coli* from the environmental sample without any fecal contamination to understand the changing pattern of their phylogroup distribution, nutrient utilization or and change in their virulence related genes. This may help us to understand the evolution of environmental *E. coli* and their population genetics.

5. References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., and Pevzner, P.A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19**,** 455-477.

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol Biol Evol* 15**,** 6-16.

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., and Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 44**,** W22-W28.

Bochner, B.R., Gadzinski, P., and Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 11**,** 1246-1255.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-2120.

Brennan, F.P., Abram, F., Chinalia, F.A., Richards, K.G., and O'flaherty, V. (2010a). Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil. *Appl Environ Microbiol* 76**,** 2175-2180.

Brennan, F.P., Grant, J., Botting, C.H., O'flaherty, V., Richards, K.G., and Abram, F. (2013). Insights into the low-temperature adaptation and nutritional flexibility of a soil-persistent *Escherichia coli*. *FEMS Microbiol Ecol* 84**,** 75-85.

Brennan, F.P., O'flaherty, V., Kramers, G., Grant, J., and Richards, K.G. (2010b). Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. *Appl Environ Microbiol* 76**,** 1449-1455.

Brzuszkiewicz, E., Gottschalk, G., Ron, E., Hacker, J., and Dobrindt, U. (2009). Adaptation of Pathogenic *E. coli* to Various Niches: Genome Flexibility is the Key. *Genome Dyn* 6**,** 110-125.

Byappanahalli, M.N., and Fujioka, R.S. (1998). Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water. Sci. Technol.* 38**,** 171-174.

Chakraborty, A., Saralaya, V., Adhikari, P., Shenoy, S., Baliga, S., and Hegde, A. (2015). Characterization of *Escherichia coli* Phylogenetic Groups Associated with Extraintestinal Infections in South Indian Population. *Ann Med Health Sci Res* 5**,** 241-246.

Classen, A.T., Boyle, S.I., Haskins, K.E., Overby, S.T., and Hart, S.C. (2003). Community-level physiological profiles of bacteria and fungi: plate type and incubation temperature influences on contrasting soils. *Fems Microbiol Ecol* 44**,** 319-328.

Clermont, O., Bonacorsi, S., Bingen, E., and Bonacorsi, P. (2000). Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Appl Environ Microbiol* 66.

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

Clermont, O., Gordon, D., and Denamur, E. (2015). Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiol* 161**,** 980-988.

Croxen, M.A., and Finlay, B.B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 8**,** 26-38.

De Muinck, E.J., Lagesen, K., Afset, J.E., Didelot, X., Rønningen, K.S., Rudi, K., Stenseth, N.C., and Trosvik, P. (2013). Comparisons of infant *Escherichia coli* isolates link genomic profiles with adaptation to the ecological niche. *BMC Genomics* 14**,** 1-21.

Den Bakker, H.C., Cummings, C.A., Ferreira, V., Vatta, P., Orsi, R.H., Degoricija, L., Barker, M., Petrauskene, O., Furtado, M.R., and Wiedmann, M. (2010). Comparative genomics of the bacterial genus Listeria: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* 11**,** 688.

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. (2002). Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl Environ Microbiol* 68**,** 1165-1172.

Dobrindt, U., Chowdary, M.G., Krumbholz, G., and Hacker, J. (2010). Genome dynamics and its impact on evolution of *Escherichia coli*. *Med Microbio Immunol* 199**,** 145-154.

Dubreuil, J.D. (2012). The whole Shebang: the gastrointestinal tract, *Escherichia coli* enterotoxins and secretion. *Curr Issues Mol Biol* 14**,** 71-82.

Duriez, P., Clermont, O., Bonacorsi, S., Bingen, E., Chaventre, A., Elion, J., Picard, B., and Denamur, E. (2001). Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiol* 147**,** 1671-1676.

Durso, L.M., Smith, D., and Hutkins, R.W. (2004). Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Appl Environ Microbiol* 70**,** 6466-6472.

Franchini, A.G., and Egli, T. (2006). Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiol* 152**,** 2111-2127.

Franz, E., Van Hoek, A.H.a.M., Bouw, E., and Aarts, H.J.M. (2011). Variability of *Escherichia coli* O157 Strain Survival in Manure-Amended Soil in Relation to Strain Origin, Virulence Profile, and Carbon Nutrition Profile. *Appl Environ Microbiol* 77**,** 8088-8096.

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

Gordon, D.M., and Cowling, A. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiol* 149**,** 3575-3586.

Ihssen, J., and Egli, T. (2005). Global physiological analysis of carbon- and energy-limited growing Escherichia coli confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environ Microbiol* 7**,** 1568-1581.

Ishii, S., Hansen, D.L., Hicks, R.E., and Sadowsky, M.J. (2007). Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in lake Superior. *Environ Sci Technol* 41**,** 2203-2209.

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

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X., and Brisse, S. (2008). Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9**,** 560.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., and Aarestrup, F.M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 52**,** 1501-1510.

Kaas, R.S., Friis, C., Ussery, D.W., and Aarestrup, F.M. (2012). Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 13**,** 1-13.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2**,** 123-140.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., and Lund, O. (2012). Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J Clin Microbiol* 50**,** 1355-1361.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 60**,** 708-720.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011). Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* 108.

Morcatti Coura, F., Diniz, S.D.A., Silva, M.X., Mussi, J.M.S., Barbosa, S.M., Lage, A.P., and Heinemann, M.B. (2015). Phylogenetic Group Determination of *Escherichia coli* Isolated from Animals Samples. *Sci World J* 2015**,** 258424.

Nandakafle, G., Seale, T., Flint, T., Nepal, M., Venter, S.N., and Brozel, V.S. (2017). Distribution of Diverse *Escherichia coli* between Cattle and Pasture. *Microbes Environ* 32**,** 226-233.

O'reilly, K.M., Low, J.C., Denwood, M.J., Gally, D.L., Evans, J., Gunn, G.J., Mellor, D.J., Reid, S.W., and Matthews, L. (2010). Associations between the presence of virulence determinants and the epidemiology and ecology of zoonotic *Escherichia coli*. *Appl Environ Microbiol* 76**,** 8110-8116.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67**,** 546-553.

R Development Core Team (2015). "R: A language and environment for statistical computing". (Vienna, Austria: R Foundation for Statistical Computing).

Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., and Chaudhuri, R. (2008). The

pangenome structure of *Escherichia coli*: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J Bacteriol* 190.

Ratajczak, M., Laroche, E., Berthe, T., Clermont, O., Pawlak, B., Denamur, E., and Petit, F. (2010). Influence of hydrological conditions on the *Escherichia coli* population structure in the water of a creek on a rural watershed. *BMC Microbiol* 10**,** 222.

Rouli, L., Merhej, V., Fournier, P.E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infect* 7**,** 72-85.

Sáez-López, E., Bosch, J., Salvia, M.D., Fernández-Orth, D., Cepas, V., Ferrer-Navarro, M., Figueras-Aloy, J., Vila, J.P., and Soto, S.M. (2017). Outbreak Caused by *Escherichia coli* O18: K1H7 Sequence Type 95 in a Neonatal Intensive Care Unit in Barcelona, Spain. *The Pediatric Infectious Disease Journal* 36**,** 1079-1086.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30**,** 2068-2069.

Skurnik, D., Bonnet, D., Bernede-Bauduin, C., Michel, R., Guette, C., Becker, J.M., Balaire, C., Chau, F., Mohler, J., Jarlier, V., Boutin, J.P., Moreau, B., Guillemot, D., Denamur, E., Andremont, A., and Ruimy, R. (2008). Characteristics of human intestinal *Escherichia coli* with changing environments. *Environ Microbiol* 10**,** 2132-2137.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000). Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol* 66**,** 230-237.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., and Fraser, C.M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102**,** 13950-13955.

Texier, S., Prigent-Combaret, C., Gourdon, M.H., Poirier, M.A., Faivre, P., Dorioz, J.M., Poulenard, J., Jocteur-Monrozier, L., Moenne-Loccoz, Y., and Trevisan, D. (2008). Persistence of culturable *Escherichia coli* fecal contaminants in dairy alpine gassland soils. *J. Environ. Qual.* 37**,** 2299-2310.

Van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011). Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* 5**,** 173-183.

Vieira, G., Sabarly, V., Bourguignon, P.-Y., Durot, M., Le Fèvre, F., Mornico, D., Vallenet, D., Bouvet, O., Denamur, E., Schachter, V., and Médigue, C. (2011). Core and Panmetabolism in *Escherichia coli*. *J Bacteriol* 193**,** 1461-1472.

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

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., and Whittam., T.S. (2009). Cryptic lineages of the genus Escherichia. *Appl Environ Microbiol* 75.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L.T., Donnenberg, M.S., and Blattner, F.R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 99**,** 17020-17024.

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

Winfield, M.D., and Groisman, E.A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl Environ Microbiol* 69**,** 3687-3694.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C.J., Ochman, H., and Achtman, M. (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 60.

Xavier, R.N., Morgan, H.W., Mcdonald, I.R., and Withers, H. (2014). Effect of Long-Term Starvation on the Survival, Recovery, and Carbon Utilization Profiles of a Bovine *Escherichia coli* O157:H7 Isolate from New Zealand. *Appl Environ Microbiol* 80**,** 4383-4390.

Chapter 8: Conclusions

This study was based on increasing evidence that *E. coli* not only occur but also grow and proliferate in soil and aquatic environments, outside of mammalian hosts. Furthermore, several studies have suggested that strains that persist in the environment are genetically distinct from gut commensal. The presence of *E. coli* in these secondary environments in the absence of any fecal contamination will likely confound the use of *E. coli* as a reliable fecal indicator. The overall goal of this project was to understand the genomic and phenotypic versatility of *E. coli* isolated from pasture soil and pond ecosystem.

Several *E. coli* isolates were found from pasture soils that survived in the field throughout the year. These isolates were diverse from bovine isolates based on *mutS* and *uidA* phylogeny. The phylogroup distribution was also very different with predominantly B1 phylogroup compared to the bovine isolates with phylogroup E and B1. Phenotypically these isolates were distinct from bovine isolates, with better winter survival and higher RDAR formation. These results indicate the presence of unique environmental strains in the pasture. Some isolates seem to be with mixed character and may have the ability to survive well in pasture and bovine gut.

E. coli isolates were collected from various sources in a pond environment. The distribution pattern of genotypic and phenotypic traits among isolates of various sample sources based on their niche preferences. Here, sediment, water and water plants isolates showed similarities in phylogroup distribution, occurrence of virulence genes and antibiotic resistance pattern, whereas snail and bovine fecal isolates were different.

What factors helped these environmental isolates to survive in the secondary environment was the next question; it was clearly found that the environmental strains formed high percentage of RDAR, resistant to protozoan grazing and also survive better in the long-term winter survival experiment. The results of antibiotic resistance and virulence genes study did not show any correlation to their survival in the secondary environment. The long-term survival of 25 representative strains with three reference strains; *E. coli* O157:H7, *E. coli* K-12 and Clade-I (TW-10509) was tested in soluble soil organic matter (SESOM) to find out if it supports the growth of *E. coli*. The results of this experiment showed a very different pattern of survival curve compared to LB (general lab media). Interestingly, the SESOM grown cells did not decline even after long-term stationary phase. Proteome study of SESOM and LB grown *E. coli* O157:H7 showed that very different stress related, central metabolism and membrane transporter proteins were expressed in two different media grown cells. This result suggested that cells maintained its population better in SESOM, could be well adapted in soil environment.

E. coli is genotypically and phenotypically very diverse. A comparative genome analysis of 20 representative isolates of phylogroup A, B1, B2 and E from pasture together with *E. coli* K12, *E. coli* O157:H7 and Clade-I (TW-10509) was performed. The result based on the core-genome phylogeny showed that phylogroup E isolates were very diverse within the group compared to phylogroup B1, Clade-I, isolates and B2 isolates were remained as outlier in the phylogenetic tree. However, the pan-genome heatmap similarity based on the presence and absence of genes showed a very clear separation of phylogroup E and phylogroup B1 isolates indicating accessory genes play an important role in determining the distribution of phylogroup and niche partitioning. It was also interesting to find that

phylogroup E isolates had a larger pangenome compare to A and B1 together. The phenotypic diversity of these 20 isolates with three reference strains was determined by nutrient utilization profiling. The pattern of nutrient utilization was very different for each strain. However, the results of phenotypic microarray assay may not justify the physiology of E. coli nutrient utilization as it is based on the utilization of single substrate, whereas in nature it is always a mixed substrate condition. The genotypic diversity did not show any correlation with the phenotypic diversity. However, to depict the correlation between genomic and phenotypic microarray based data based on the presence and absence of genes is not reasonable as the phenotypic diversity of cell is also affected by mechanisms as for example regulatory and signal transduction systems and membrane transporter whose functionality could not be directly detected by Phenotypic microarray (PM) system or genome annotation. Further implementation of transcriptome, proteome and metabolomics data is necessary to understand the genotype and phenotypic relatedness of organisms.

This study supports the idea that genetically distinct population of naturalized *E. coli* may exist which is different from the *Escherichia* clade. Environmental *E. coli* stains have developed various strategies to survive in the environment under low nutrient conditions. However, there are many more interesting questions to further investigate the ecology of naturalized or environmental *E. coli*; what mechanisms enable these bacteria to grow and survive in soil relatively better than non-naturalized *E. coli*? What are the functions of the unique genes present in those strains? How these environmental strains avoid protozoan predation? Whether they have the ability to recolonize in gut environment? What is the mutation rate of naturalized *E. coli* in soil and in laboratory media?