SURVIVAL OF ESCHERICHIA COLI IN AGRICULTURAL SOILS

AND

STUDENT ENGAGEMENT IN ONLINE DISCUSSIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Allison Mary Truhlar

August 2017

SURVIVAL OF ESCHERICHIA COLI IN AGRICULTURAL SOILS

AND STUDENT ENGAGEMENT IN ONLINE DISCUSSIONS

Allison Mary Truhlar, PhD

Cornell University 2017

There is great interest in identifying manure management techniques to minimize the persistence of pathogenic *Escherichia coli* in agricultural fields and, thereby, decrease the risk of downstream contamination and human infection. The first two chapters of this dissertation seek to determine the how the agricultural environmental variable of manure application method shapes the genetic and phenotypic population structure of *E. coli*. First, we conducted a field and laboratory experiment that demonstrated the expression of extracellular fibers called curli, which are linked to environmental persistence of *E. coli*, was linked to surface-application of manure, as opposed to incorporation into the soil. Second, we applied whole genome sequencing technology to isolates collected from laboratory microcosms with differing manure application treatments. We found no systematic genomic differences (i.e. individual-level selection) that could be explained by week or manure application treatment.

As higher education institutions offer online courses to growing audiences, there is increasing desire to understand how best to engage students. The third chapter of this dissertation examines the effects of assigning chat roles and facilitating self and group reflection on student-content and student-student interaction outcomes in synchronous online chats. Group reflections were the only intervention that had a significant effect on both outcomes.

BIOGRAPHICAL SKETCH

Allison Truhlar attended college at Cornell University, receiving a Bachelor of Science degree in Biological Engineering in 2011. She then attended the University of Cambridge, receiving a Master of Philosophy degree in Biological Sciences (Zoology) in 2012. In 2013, Allison returned to Cornell University to pursue a doctorate in the Department of Biological and Environmental Engineering. During this most recent stint in Ithaca, Allison's other notable accomplishments include learning to swim freestyle, completing her first half-marathon and first sprint triathlon races, and making her first quilt.

DEDICATION

This dissertation is dedicated to my grandfather, John O'Dwyer, the first and so far only doctorate in my family. I'm proud to be joining him.

I also dedicate this dissertation to my mother, Mary Truhlar, who I look to as an example of how to be a strong, smart woman in the world, and to my father, Richard Truhlar, who always reminded me that I could do this.

ACKNOWLEDGEMENTS

During the course of this work, I was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144153, as well as by Cornell University Teaching Assistantships.

Work for Chapter 1 was additionally supported by the National Institute of Food and Agriculture (2008-35102-04462). Work for Chapter 3 was additionally supported by a grant from the Great Lakes Higher Education Corporation & Affiliates to the Cornell University Center for the Integration of Research, Teaching, and Learning.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH iii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

Chapter 1 : EFFECTS OF MANURE-APPLICATION PRACTICES ON CURLI PRODUCTION BY ESCHERICHIA COLI TRANSPORTED THROUGH SOIL 1

ABSTRACT 1

INTRODUCTION 2

MATERIALS AND METHODS 4

RESULTS 9

DISCUSSION 14

ACKNOWLEDGEMENTS 17

WORKS CITED 18

Chapter 2 : GENETIC SELECTION IN A PATHOGENIC *ESCHERICHIA COLI* STRAIN EXPOSED TO THE MANURE-AMENDED SOIL ENVIRONMENT 23

ABSTRACT 23

INTRODUCTION 25

METHODS 28

CONCLUSIONS 46

WORKS CITED 68

Chapter 3 : STUDENT ENGAGEMENT WITH COURSE CONTENT AND PEERS IN SYNCHRONOUS ONLINE DISCUSSIONS 74

ABSTRACT 74

INTRODUCTION 75

CONTEXT 79

PURPOSE 81

METHODS AND MATERIALS 82

RESULTS 86

DISCUSSION 94

IMPLICATIONS FOR PRACTICE 100

WORKS CITED 102

Appendices 106

APPENDIX A 106

APPENDIX B 107

APPENDIX C 109

APPENDIX D 110

LIST OF FIGURES

Figure 1-1 Escherichia coli morphotypes when plated on Congo Red media: (a) Red, dry and rough (R-DAR), indicating curli and cellulose production; (b) Red and smooth, indicating curli production; (c) White, indicating an absence of curli. 6

Figure 1 - 2 The proportion of E. coli isolates producing curli (± 1 SE) in manure from the manure storage area and in water from the subsurface drain pipe effluent on each of the three study farms. Different letters indicate significant between-farm differences of the proportion of curli-producing isolates sampled from the drain pipe effluent. An asterisk indicates a significant within-farm difference between the proportion of curli-producing isolates sampled from the manure storage and the drainage pipe effluent……10

Figure 1-3 The proportion of curli-producing E. coli isolates (± 1 SE) in effluent samples from soil columns with no manure (Control), manure incorporated into the top layer of soil (Incorp), or manure spread on the soil surface (surface). Different letters indicate significant differences between treatments. 12

Figure 1-4 Mean number of E. coli (± 1 SE) in the leachate of soil columns that had either no manure (Control), manure incorporated into the top layer of soil (Incorporated), or manure spread on the soil surface (Surface). Different letters indicate significant differences between treatments. 13

Figure 2-1 Schematic of the addition steps taken to ensure the mapping-to-reference method used in CLC Workbench identified all genomic differences. 34

Figure 2-2 Survival of E. coli O157:H7 EDL933 in soil microcosms with and without added manure ("Manure" and "Soil," respectively). "Incorp" indicates mixing of the soil after E. coli application; "SA" indicates surface application (i.e., no mixing). Data represents an average (+/- 1SD) of four replicates. 36

Figure 2-3 Specific decay constants for E. coli O157:H7 EDL933 in soil microcosms with and without added manure ("Manure" and "Soil," respectively). "Incorp" indicates mixing of the soil after E. coli application; "SA" indicates surface application (i.e., no mixing). Data represents an average (+/- 1SD) of four replicates. Different letters indicate significant differences in the specific decay constant between treatments. 37

Figure 2-4 Sample morphologies of isolates grown on Congo Red agar for seven days at 30C. Isolates (a) and (b) were collected from the surface-applied manure treatment after six weeks, and exemplify the red morphology overlain with raised white growth that was unique to isolates from this treatment. Isolate (c) was collected from the same soil microcosm after six weeks, and is representative of the smooth, uniform red morphology manifest by isolates collected from all treatments other than surface-applied manure. Pictures were taken under 16X magnification... 39

Figure 2-5 The number of discarded reads, obtained from each isolate after mapping reads to the reference genome, that were annotated with a predicted feature through the MG Rast server. The reads are plotted by both the week the isolate was collected and the soil-manure treatment to which the isolate was exposed. Four points with values greater than 25,000 were excluded to make the plot easier to read. These points came from week 3 soil-incorporated, week 5/6 manure-incorporated, week 5/6 manure surface-applied, and week 8 manure-incorporated strains. 41

Figure 3-1 Summary of the Teaching-as-Research process (Kwako et al., 2005; Williams, 2015). 79

Figure 3-2 Timeline of interventions used in the online chats. Once an intervention was introduced, it was continued for the rest of the semester, unless otherwise noted. 82

Figure 3-3 Proportion of total posts in the 2015 chat and Chat 1, 2016 that were coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions. Between 2015 and 2016, chat roles were introduced. The upper whisker extends to the largest observed value within the third quartile plus 1.5 times the interquartile range. The lower whisker extends to the smallest observed value within the first quartile minus 1.5 times the interquartile range. An asterisk indicates a significant difference at a 95% confidence level. 87

Figure 3-4 Proportion of total posts in Chat 1, 2016 and Chat 2, 2016 that were coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions. Between Chat 1 and Chat 2, self-reflections were introduced. The upper and lower whiskers are drawn as described in Figure 1. 88

Figure 3-5 The proportion of posts coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions, separated by whether the group went on to complete a group-reflection prior to chats 3 and 4. An asterisk indicates significance a 95% confidence level. A cross indicates significance at the same level after "high" outliers are removed. Lighter shading is used for

groups that went on to complete group-reflections before Chats 3 and 4, while darker shading is used for groups that were never asked to complete group-reflections. The upper and lower whiskers are drawn as described in Figure 1. 90

Figure 3-6 The number of individual posts from the two group reflections, before Chats 3 and 4, respectively, in which students either expressed or did not express agency. 92

Figure 3-7 The number of individual posts from the two group reflections, before Chats 3 and 4, respectively, in which students suggested a concrete or a vague change to the next chat..... 93

LIST OF TABLES

Table 1-1 Summary of daily precipitation during each of the three field sampling periods..... 10

Table 2-1 Soil and manure characteristics. 29

Table 2 -2 The read count for MGRAST-annotated predicted features were individually modeled as a linear function of week-treatment; the resulting p-values are shown here. The predicted features that were included in the analysis were the top 25 genuses (by frequency) and functional subsystems, except for those that were correlated with all other functional subsystems. Subsequently, all p values were then ordered and compared to threshold values calculated using the Benjamini-Hochberg (BH) procedure, to account for multiple comparisons. The false discovery rate (α) was set at 0.25 to ensure no possible significant factors were overlooked. A p value less than the corresponding threshold is considered significant. ……42

Table 2-3 Single nucleotide polymorphisms (SNPs) identified in all 94 resequenced isolates of E. coli O157:H7 EDL933 $\Delta stx_{1\text{-}2}$..... 44

Table 2-4 Summary of known curli regulators (Evans and Chapman, 2014). The location of each regulator in the E. coli O157:H7 EDL933 references genome used in this study is provided, along with an indication of whether a gap or SNP was found in the regulator through mapping to the reference genome on CLC Workbench..... 45

Table 3-1 Interventions applied in each of the four chats, with the corresponding hypotheses and t-test results. Bolded p-values indicate a significant result at the 95% confidence level. 84

Table 3-2 Modified version of Bloom's original taxonomy, adapted from Bradley et al. (2007). 85

Table 3-3 Chat aspect most frequency identified by each group as "ineffective" for the first and second group-reflection*s*..... 94

Chapter 1 : EFFECTS OF MANURE-APPLICATION PRACTICES ON CURLI PRODUCTION BY ESCHERICHIA COLI TRANSPORTED THROUGH SOIL[1]

ABSTRACT

The release of *Escherichia coli* into the environment from untreated manure can pose a threat to human health. Environmental survival of *E. coli* has been linked to extracellular fibers called curli. We investigated the effect of manure management (surface application followed by incorporation versus immediate incorporation) on the relative abundance of curli-producing *E. coli* in subsurface drainage effluent. Samples were collected from three dairy farms. The proportion of curli-producing *E. coli* in the manure storage facilities was uniform across the farms. However, the abundance of curli-producing *E. coli* was much greater ($P < 0.05$) in the tile drains of farms performing surface application of manure than in the tile drain of the farm that incorporated manure. This field observation was tested with controlled soil column experiments; the abundance of curli-producing *E. coli* in soil column effluents was greater ($P < 0.05$) when manure was surface-applied than when it was incorporated. Our findings suggest selection pressures resulting from the different manure application methods affected curli production by *E. coli* isolates transported through soil. Given the importance of curli production in pathogenesis, this work highlights the effect that manure management strategies may have on pathogenesis-associated phenotypes of bacteria in agricultural subsurface runoff.

---

[1] Reprinted with permission from Truhlar, A. M., Salvucci, A. E., Walter, M. T., Warnick, L. D., Hay, A. G., & Steenhuis, T. S. (2015). Effects of manure-application practices on curli production by Escherichia coli transported through soil. *Environmental Science & Technology*, *49*(4), 2099-2104. doi: 10.1021/es5053039. Copyright 2015, American Chemical Society.

INTRODUCTION

Escherichia coli is consistently found in the intestinal tract and feces of mammals; some strains, such as the verotoxigenic O157:H7, are important human pathogens. Animal feces are a major environmental source of *E. coli* due to livestock grazing and the deliberate spreading of manure on agricultural fields (Jamieson et al., 2002; Ogden et al., 2001). The protection of surface, drainage and ground waters from *E. coli* contamination is of particular interest as contaminated water poses a human health threat through both direct consumption and its use for the irrigation of produce (Jensen et al., 2013; Rangel et al., 2005; Solomon et al., 2002). Our failure to prevent water-associated infectious disease outbreaks is due in part to a lack of understanding of the factors that influence survival and transport of pathogenic bacteria in the environment (Stevik et al., 2004).

When *E. coli* are shed from the near constant conditions of the mammalian gut and enter the outside environment, they are subject to many new stresses including predation, limited nutrient availability, ultraviolet (UV) light, and osmostic stress, as well as variations in temperature, pH, and humidity (Savageau, 1983; Winfield and Groisman, 2003). In spite of these stressors, *E. coli* can survive in the environment for extended periods ranging from several days to more than a year (Entry et al., 2000; Fremaux et al., 2007a, 2007b; Kudva et al., 1998).

Escherichia coli are capable of expressing several surface structures that enhance their survival in the environment. Curli, curled polymeric protein structures expressed by most natural strains of *E. coli*, are capable of binding a variety of extracellular matrix proteins (e.g., fibronectin and plasminogen) and abiotic surfaces (Olsen et al., 1989; Römling et al., 1998; Sjöbring et al., 1994). They have been associated with such beneficial adaptations as biofilm formation (Römling et al., 1998; Vidal et al., 1998), reduced uptake of toxic metals (e.g. Hg(II)) (Hidalgo

et al., 2010), and improved survival in nutrient-limited conditions (Carter et al., 2011). Nearly identical structures known as thin aggregative fibers or TAFI have been shown to enhance long term survival and desiccation resistance in closely related *Salmonella* species (White et al., 2006). It is of great interest to understand how these bacterial phenotype traits are modulated by environmental factors.

There is evidence that environmental stressors can select for specific *E. coli* genotypes that are most able to persist in a given environment (Ishii et al., 2006; Walk et al., 2007). This may be due to the presence or absence of specific genes or changes in gene expression resulting in altered phenotypes; Moreira et al. (2012) found that *E. coli* populations isolated from freshwater periphyton had higher and more consistent biofilm formation capabilities than isolates taken from host environments. In the case of agriculture, environmental challenges experienced by *E. coli* vary with respect to method (e.g. surface applied or injected) and form (e.g. slurry or solid wastes) of manure application (Nicholson et al., 2005), but the effects of these differences on *E. coli* phenotype selection are not fully understood.

The objective of this study was to determine the relative abundance of curli-producing *E. coli* transported through soil after different manure application methods. Two sets of experiments were conducted. First, *E. coli* strains were isolated from the manure storage facilities and tile drains of three dairy farms that field-applied manure in Central New York and then examined for their ability to produce curli. Second, *E. coli* strains were isolated from the effluents of soil columns in which manure was applied via either immediate incorporation or surface application followed by incorporation, and then examined for curli formation. The results presented below clearly demonstrate that *E. coli* transported from soil amended with surface-applied manure were

more likely to produce curli than those transported from manure that had been incorporated into soil.

MATERIALS AND METHODS

Field Experiment

Sampling Locations

Samples were collected from three dairy farms within forty miles of Cornell University, Ithaca, New York. The study specifically targeted dairy farms with active subsurface drainage (tile) lines in crop fields with regular manure spreading. The predominant soils in the crop fields consist of fine-loamy, mixed active, mesic Glossaquic Hapludalfs, coarse-loamy, mixed, active, mesic Typic Fragiudepts, and fine-loamy, mixed, active, mesic Aeric Endoaqualfs which are common to the Central New York Finger Lakes region. The fields were kept in a corn and forage rotation. At the time of the study, the fields produced corn silage. Manure application occurred after fall harvest, and then again prior to planting in the spring.

At each farm, sampling was conducted once every seven days over a 3-4 week period at two distinct manure source locations: manure storage and drainage tile effluent. At farms 1 and 2, manure was surface applied to fields and incorporated several days after application. This method is hereafter referred to as "surface application." At farm 3, manure was incorporated at the time of application, a method hereafter referred to as "incorporation." All sampling was completed between October 2008 and October 2009. On each day during a sampling period, the daily maximum and minimum air temperature and daily precipitation were recorded.

Sampling Methods

Moore swabs were used to collect environmental samples from the manure storage and tile drains (Barrett et al., 1980). From manure storage areas, four independent samples were taken by swabbing a stainless steel rod that had been inserted into the manure outflow. These samples were pooled into a single flip-top container, resulting in one representative sample. For tile drain samples, the field most frequently treated with manure was identified by consultation with the farm manager, and then the tile drain with the most runoff from that field was subsequently selected for outflow sampling. Field tile drain outflows were sampled by hanging three individual Moore swabs from a six-foot fiberglass rod. The rod was clamped to the top of the discharge pipe, allowing the swabs to sit in the water flowing along the bottom of the tile drain. At each sampling point all three swabs were collected in a single flip-top container. Fresh Moore swabs were installed in the tile drain on days one, seven, and fourteen, and collected on days seven, fourteen, and twenty-one. All samples were refrigerated at 4°C until processed (within 36 hours).

Escherichia coli culturing

Moore swabs were enriched in 100 mL of Difco GN Hajna broth (BD, Sparks, MD, USA) and incubated at 37°C for 18–24 hours. Following incubation, the inoculated broth was streaked onto MacConkey agar (MAC) plates (Laboratory Services, Waterville, ME, USA), in order to isolate *E. coli*. The MAC plates were incubated for an additional 18-24 hours at 37^{o}C before approximately fifty *E. coli* isolates per sample were selected.

Curli producing morphotype identification

Each *E. coli* isolate was stabbed onto a YESCA-Congo Red agar plate as a qualitative means of assessing curli and cellulose production (Hammar et al., 1995). Plates were sealed with parafilm

and incubated for seven days at 30º C. *E. coli* isolates form red dry and rough colonies (R-DAR) on Congo Red media if they produce both curli and cellulose, with the latter being another major component of the extracellular matrix of some *E. coli* (Hammar et al., 1995). Isolates that produce curli but not cellulose are red and smooth, whereas isolates without either curli or cellulose form smooth white colonies.[25] The number of R-DAR and red-smooth colonies were added together to give the total number of curli-producing isolates. Figure 1 shows visual examples of each category used in classification. More than 10 proteins and six small RNAs control curli expression, thus there is no one gene to interrogate (Bordeau and Felden, 2014). Identifying curli production by ability to bind Congo Red dye is an efficient and effective alternative to molecular analysis.[30]

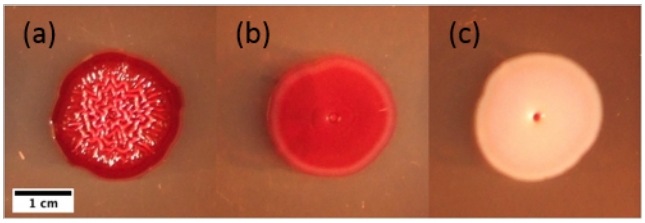


Figure 1-1 *Escherichia coli* morphotypes when plated on Congo Red media: (a) Red, dry and rough (R-DAR), indicating curli and cellulose production; (b) Red and smooth, indicating curli production; (c) White, indicating an absence of curli.

Laboratory Experiment

Soil column construction and treatment

To test the effects of manure application on the abundance of curli-producing E. coli in a more controlled manner, 5 cm soil columns were set up that contained one of three treatments: 1) soil only, 2) soil with manure applied to the surface, or 3) soil with manure uniformly incorporated. Each treatment was repeated in triplicate, resulting in nine columns total. The column apparati

consisted of PVC columns (7.7 cm inner diameter) attached to PVC funnels (7.7 cm outer diameter at opening). The outlet of the funnel was covered with mesh, and then each funnel was filled with 0.24 kg of 12/20 sand (Unimin Corp., New Canaan, CT, USA) to the level where the funnel met the column. For the soil only condition, 0.34 kg of soil was added to each column, to a depth of 5 cm. For the two soil and manure conditions, 0.32 kg of soil and 6.8 g of fresh manure were used. The manure was collected from one of the study farms. This ratio of soil to manure approximates the 0.02 manure to soil ratio found by Vidovic et al. (2007) in dairy farm soil samples. We assumed a mineral particle density of 2.65 g/cm^3 and a manure density of 1g/cm^3; thus, this packing protocol resulted in an overall column porosity of 0.45 and 0.46 for the soil and soil-manure mixtures, respectively, and the total pore volumes were 105 and 107 cm^3. Note, the soil was mixed to be homogenous before fabricating the soil columns.

The nine soil columns were exposed to conditions common in the field (UV, desiccation, temperature fluctuations) for a period of five days by leaving the columns near a large window. Both the maximum and minimum soil temperature at a depth of 3 to 4 cm was recorded for every night and day period. A leachate collection system was installed beneath each column consisting of a 250 mL wide-mouth bottle (Thermo Scientific Nalgene, Waltham, MA, USA) and a Moore swab. The columns were then arranged under a rain machine in a ten-space grid and rained on for five hours. Each column spent one half-hour in each square of the grid, for an average rain intensity of 0.52 cm/hr. The columns rested for 18 hours before the swabs were collected.

Escherichia coli culturing and enumeration

Escherichia coli to be selected for assessment of curli production were cultured following the methods described in Sections 2.1.4 and 2.1.5. This ensured that any influence of the media on

curli production would be consistent across experiments. To determine total E. coli, serial 10-fold dilutions of the soil column effluent were plated on MAC and incubated at 37^{o}C for 24 hours.

Statistical Analysis

All statistical analyses were performed in R version 3.0.2. The data describing the distribution of curli producing strains for both the field and soil column isolate populations were normal (Shapiro-Wilks test, $P > 0.05$) and homoscedastic (Barlett's test, $P > 0.05$). The soil column enumeration data were log transformed to meet the requirements of normality and homoscedasticity.

Field isolates

To address the questions of whether the proportion of curli producing isolates differed significantly between farms, and/or between sampling sites on the farms a one-way analysis of variance (ANOVA) was used. When the results of the ANOVA analysis were significant ($p < 0.05$), Tukey's HSD (Honestly Significant Difference) post hoc test was implemented to determine which farms differed significantly. To compare the proportion of curli producing isolates in the manure storage and drain tile effluent of a given farm, a t-test was used.

Soil column isolates

To determine whether the manure application method affected the proportion of isolates that produced curli in the soil column effluent, a one-way ANOVA was used, followed by Tukey's HSD post hoc test to determine which treatments differed significantly. Log-transformed

enumeration data were analyzed with a one-way ANOVA to determine if manure application method had an effect on number of isolates in the leachate.

RESULTS

Field experiment

The maximum air temperature recorded during any of the three sampling periods was 33.89°C; the minimum air temperature was -7.22°C. The mean daily maximum air temperature during all sampling periods was 13.34°C (SD = 6.48°C); the mean daily minimum air temperature was 1.33°C (SD = 4.83°C). A summary of the daily precipitation during each of the three sampling periods is shown in Table 1. The typical soil temperature at 1 m drain depth during the time of field sampling is 10 ± 3°C (L. Geohring, personal communication). Additionally, in order for there to be flow through the subsurface drains, the soil moisture must have held near field capacity water content (i.e., about 0.35 ± 0.05 on a volumetric water content basis) at all sampling periods in the experiment.

Prior to field application, approximately 45% of the E. coli isolated from manure produced curli. The proportion of E. coli isolates producing curli in manure from storage areas was not significantly different between farms (ANOVA, $F_{(2,6)} = 1.647$, $P = 0.269$; Figure 2). In samples taken from the drainage pipe effluent, however, the proportion of E. coli isolates producing curli was significantly affected by the farm of origin (ANOVA, $F_{(2,7)} = 165.7$, $P < 0.001$). Tukey's HSD test indicated that significant differences existed between Farm 3 and Farm 1 ($P < 0.001$), and between Farm 3 and Farm 2 ($P < 0.001$), with fewer curli-producing isolates being found in the drain effluent samples of Farm 3 in both cases (Figure 2). There was no significant difference

in the proportion of curli producing isolates in the drain effluent samples from Farm 1 and Farm 2 (P = 0.872; Figure 2).

Table 1-1 Summary of daily precipitation during each of the three field sampling periods.

Site	Treatment	Number of days with rain events	Average daily rain depth (in)	Maximum daily rain depth (in)	Minimum daily rain depth greater than zero (in)
1	Surface-applied	18 out of 29 days	0.17	1.18	0.02
2	Surface-applied	10 out of 33 days	0.04	0.34	0.04
3	Incorporated	15 out of 27 days	0.08	0.52	0.03

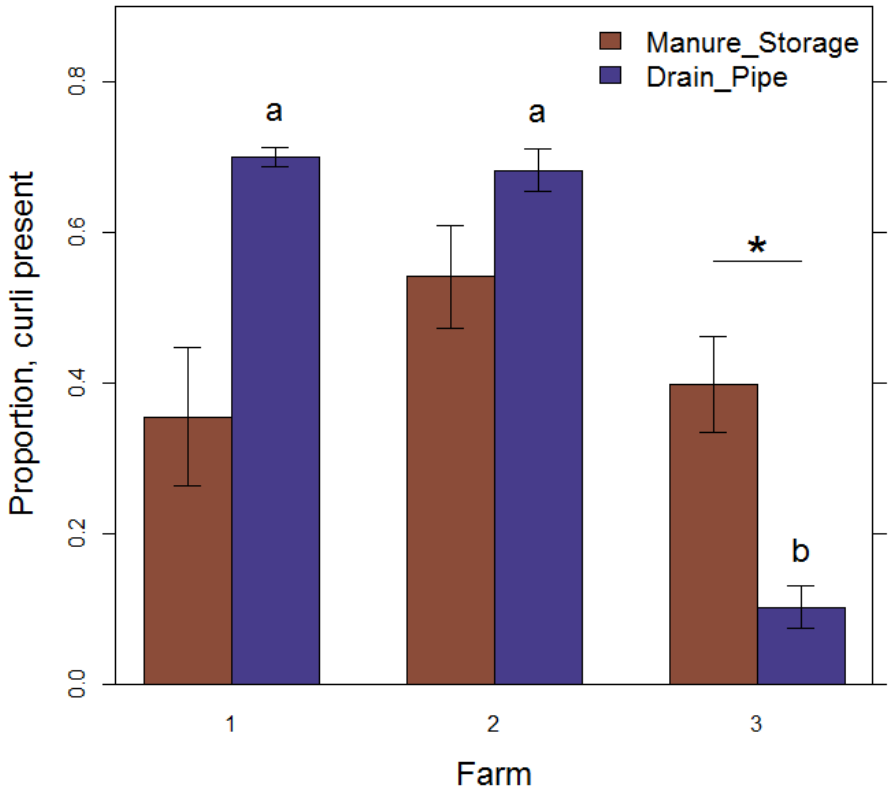


Figure 1-2 The proportion of *E. coli* isolates producing curli (± 1 SE) in manure from the manure storage area and in water from the subsurface drain pipe effluent on each of the three study farms. Different letters indicate significant *between-farm* differences of the proportion of curli-producing isolates sampled from the drain pipe effluent. An asterisk indicates a significant *within-farm* difference between the proportion of curli-producing isolates sampled from the manure storage and the drainage pipe effluent.

Laboratory incorporation experiment

The maximum soil temperature reached in any of the soil columns over the course of the experiment was 21.1^{o}C; the minimum soil temperature was 11.2^{o}C. The mean daily maximum temperature of all the soil columns was 18.6^{o}C (SD = 1.1^{o}C); the mean daily minimum temperature was 14.9^{o}C (SD = 2.1^{o}C).

The proportion of E. coli isolates producing curli in soil column leachate samples was significantly affected by the manure application method (ANOVA, $F_{(2,6)} = 8.696$, P = 0.017; Figure 3). Significant differences at the 95% confidence level existed between the surface applied manure and control (no manure) columns (Tukey's HSD, P = 0.017) and between the surface applied manure and incorporated manure columns (Tukey's HSD, P = 0.049). In both cases, a greater proportion of the isolates produced curli in the effluent samples from the surface applied manure treatment (Figure 3).

In order to determine if genes associated with curli production were present in curli negative isolates (isolates that appeared white when plated on Congo Red agar), we completed a follow-up assessment using strains isolated from the column effluent. The strains had been frozen at -80^{o}C since the time of the original experiment. Seven strains, two that had been identified as curli-positive and five that had been identified as curli-negative, were selected. We targeted the strains for regions related to curli production, specifically, the functional gene csgA. We observed no correlation between the amplification of either of these genes and whether the strain was curli positive or negative; csgA was present in all of the seven strains.

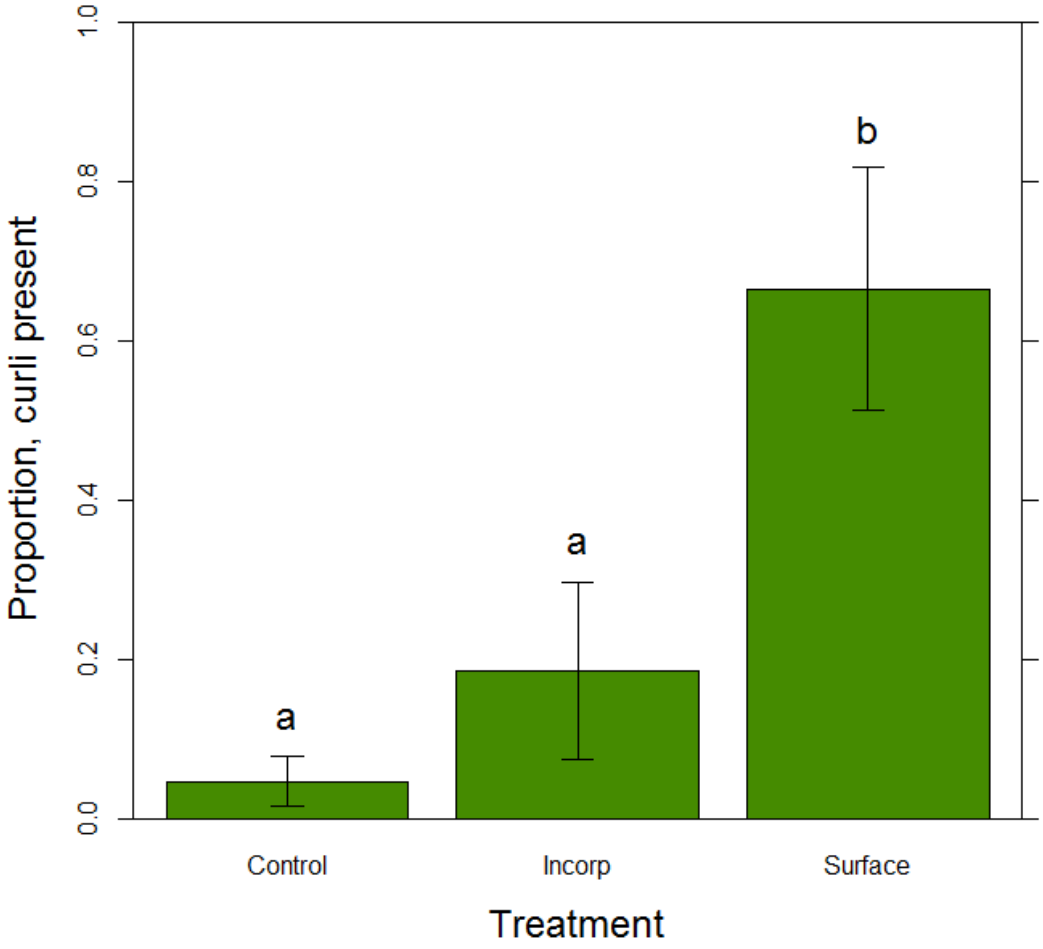


Figure 1-3 The proportion of curli-producing *E. coli* isolates (± 1 SE) in effluent samples from soil columns with no manure (Control), manure incorporated into the top layer of soil (Incorp), or manure spread on the soil surface (surface). Different letters indicate significant differences between treatments.

As expected based on our farm observations, there was a significant effect of manure application on the total number of bacteria isolated from the soil column effluent (ANOVA, $F_{(2,6)} = 8.895$, $P = 0.016$; Figure 4). Specifically, the control effluent had significantly fewer E. coli isolates than both the surface-applied manure treatment ($P = 0.032$; Figure 4) and the incorporated manure treatment ($P = 0.020$; Figure 4). However, there was no significant difference between the total

number of E. coli in the effluent of the surface-applied manure treatment and that of the incorporated manure treatment (P = 0.908; Figure 4). The initial E. coli count in the manure was 5 x 10^7 CFU/mL, which is over 3500 times the maximum amount of E. coli recovered from any of the column treatments.

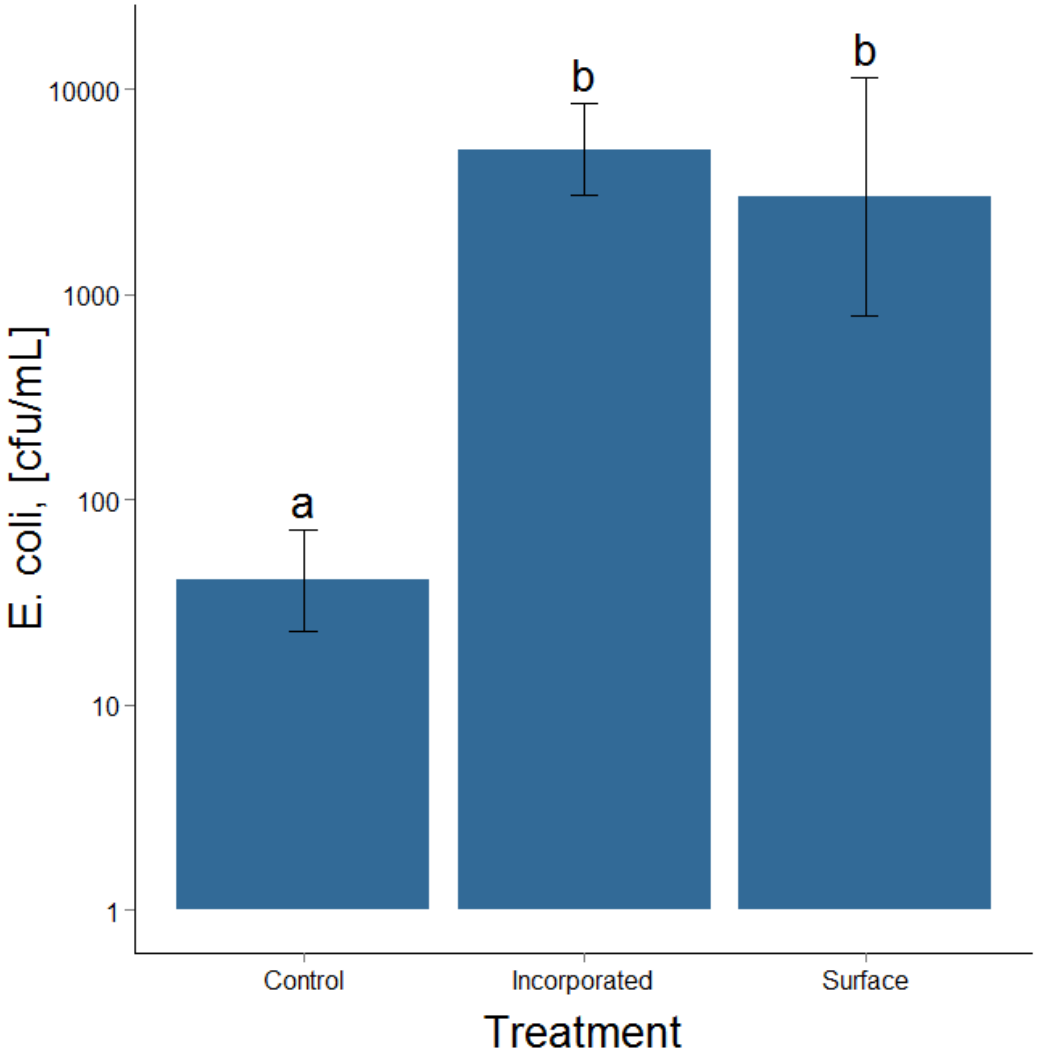


Figure 1-4 Mean number of E. coli (± 1 SE) in the leachate of soil columns that had either no manure (Control), manure incorporated into the top layer of soil (Incorporated), or manure spread on the soil surface (Surface). Different letters indicate significant differences between treatments.

DISCUSSION

Once removed from their mammalian hosts, *E. coli* adapt to the environmental challenges they face by altering their expression of surface associated biomolecules (Ishii et al., 2006; Walk et al., 2007). In this study, we sought to understand what effect different methods of manure application would have on the number of curli-producing *E. coli* transported through soil. Curli are polymeric surface structures known to convey numerous survival advantages and to be associated with pathogenicity (Carter et al., 2011; Hidalgo et al., 2010; Römling et al., 1998; Vidal et al., 1998); TAFI are the homologue structure in Salmonella. Previous work has investigated the survival advantages conferred to bacteria by curli and TAFI (Carter et al., 2011; Hidalgo et al., 2010; Römling et al., 1998; Vidal et al., 1998; White et al., 2006), and the effect of these structures on bacterial transport in soil (Brombacher, 2003; Salvucci et al., 2009), but we are unaware of any work specifically linking manure management strategy to expression of these cell surface features and subsequent downstream transport.

Given their role in biofilm formation and reported tendency to retard transport through sand columns (Brombacher, 2003), we initially hypothesized that curli-producing *E. coli* would not be transported efficiently through the soil column and would therefore be underrepresented in tile drains when compared to the manure of origin. Somewhat surprisingly, this was not the case for the first two farms we sampled. Rather, the opposite occurred, with curli-producing *E. coli* accounting for approximately 70% of the tile drain isolates despite representing only 45% of the isolates from the original manure. One potential interpretation of these results was that only curli-producing *E. coli* survived because the manure, which was surface applied at these farms, was subject to environmental stress, such as UV radiation, temperature fluctuations, and desiccation. Similar environmental conditions have been shown to select for the production of

thin aggregative fimbriae (TAFI) by *Salmonella* species (White et al., 2006). The genes encoding TAFI in *Salmonella* are 95% similar to those encoding curli in *E. coli* (Barnhart and Chapman, 2006). It is therefore possible curli-producing *E. coli* survived the initial application at a higher rate in these conditions than those that did not produce curli, and were therefore over-represented in the tile drain effluent.

These results from the two farms that surface-applied manure contrasted dramatically with the results from the third farm we sampled, where the manure was incorporated. At the third farm less than 10% of the *E. coli* in the tile drain effluent produced curli. This latter result was more in keeping with our original hypothesis, but the fact that all three farms had similar levels of curli-producing organisms in the original manure and such markedly different levels in the effluent suggested that the presence or absence of curli was not the main factor governing *E. coli* transport.

A major difference between farms in the field study was the manure management strategy, with curli-producing *E. coli* strains being more abundant in the tile drains of farms that surface applied their manure. We therefore undertook a controlled study to determine if manure application method impacted the proportion of curli-producing *E. coli* transported through soil columns. Consistent with our farm-based observations, we found that the relative abundance of curli-producing *E. coli* was significantly higher in the effluent of columns where the manure was surface applied than in the effluent of columns where the manure was incorporated (Figure 3). Furthermore, because the manure for the column experiments originated from one source, these results rule out the possibility that observed differences in curli expression between treatments were caused solely by the presence of different strains of *E. coli* in the manure used at each farm.

The total number of *E. coli* transported through each column, however, did not significantly vary by treatment (Figure 4). This result appears to contradict previous work by Brombacher et al. (2003) that correlated biofilm production by *E. coli* to cell adhesion in sand columns. A possible explanation for this divergence is the relatively long time scale of leachate collection in our experiment, which could have allowed for the recovery of even those isolates that would have been transported more slowly during the shorter experiments of Brombacher et al. (2003) This, however, is not the first report that found no correlation between the production of curli-like structures and cell retention in soil columns; Salvucci et al. (2009) demonstrated that the transport of *Salmonella* spp. was not retarded by the expression of TAFI (*Salmonella's* curli homologue). Finally, we note that the *E. coli* counts may be limited by our choice of method. It is known that culturing on solid media, especially from environmental samples, can result in under-representation of the actual microbial population (Epstein, 2013; Hoefel et al., 2003).

Our results clearly demonstrate that the method of manure application impacts the relative abundance of curli-producing *E. coli* that are transported to tile drains. The strength of this impact is emphasized by the consistent results seen for both surface-applied sampling periods, despite different rainfall patterns (Table 1). Given that similar numbers of all *E. coli* were transported in our column study regardless of the manure application method (Figure 4), it seems unlikely that the increased abundance of curli-producers from surface applied manure was simply due to the die-off of *E. coli* that did not produce curli. However, further work is needed to determine if similar results are seen in the field. An alternative explanation for our column results is curli-specific phenotypic shifting, which has been observed in TAFI-producing *Salmonella* in the laboratory (White and Surette, 2006). Phenotypic shifting requires some metabolic activity in order for *E. coli* to respond to environmental stimuli. At the temperatures

reported for both the field and column experiments, *E. coli* population growth would likely range from stasis to moderate growth (Kudva et al., 1998).

From a practical standpoint, the results of this study make it clear that manure management strategy has an important impact on the type of *E. coli* that get transported in both field and laboratory settings. Our results suggest that the *E. coli* transported from surface-applied manure are likely to have greater advantages for environmental survival and increased pathogenicity compared to those transported from incorporated manure, due to greater rates of curli expression. More work is needed to clearly characterize the downstream risks associated with the differences in *E. coli* that produce curli and those that do not, especially with regard to the spread of foodborne illness. For example, while shiga-like toxin-producing *E. coli* strains such as O157:H7 are known to produce curli (Carter et al., 2011; Moreira et al., 2012), future work could ask specifically whether curli expression in these strains responds to the method of manure application as observed here. Such information could help inform regulations on untreated animal-based soil amendments, such as those imposed by Food and Drug Administration in the Food Safety and Modernization Act (Harris et al., 2013).

ACKNOWLEDGEMENTS

Support from the National Institute of Food and Agriculture (2008-35102-04462) is gratefully acknowledged. Additional funding for the laboratory experiments was provided by the Dextra Undergraduate Research Endowment Fund. We thank M. Elton, J. D. Siler, W. Zhang, B. K. Richards, and L. D. Geohring for their help with performing field and laboratory work, and with editing the manuscript.

WORKS CITED

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

Barrett, T.J., Blake, P.A., Morris, G.K., Puhr, N.D., Bradford, H.B., and Wells, J.G. (1980). Use of Moore swabs for isolating Vibrio cholerae from sewage. J. Clin. Microbiol. *11*, 385–388.

Bordeau, V., and Felden, B. (2014). Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone. Nucleic Acids Res. *42*, 4682–4696.

Brombacher, E. (2003). The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in Escherichia coli. Microbiology *149*, 2847–2857.

Carter, M.Q., Brandl, M.T., Louie, J.W., Kyle, J.L., Carychao, D.K., Cooley, M.B., Parker, C.T., Bates, A.H., and Mandrell, R.E. (2011). Distinct Acid Resistance and Survival Fitness Displayed by Curli Variants of Enterohemorrhagic Escherichia coli O157:H7. Appl. Environ. Microbiol. *77*, 3685–3695.

Entry, J.A., Hubbard, R.K., Theis, J.E., and Fuhrmann, J.J. (2000). The Influence of Vegetation in Riparian Filterstrips on Coliform Bacteria: II. Survival in Soils. J. Environ. Qual. *29*, 1215–1224.

Epstein, S. (2013). The phenomenon of microbial uncultivability. Curr. Opin. Microbiol. *16*, 636–642.

Fremaux, B., Prigent-Combaret, C., Delignette-Muller, M.L., Dothal, M., and Vernozy-Rozand, C. (2007a). Persistence of Shiga toxin-producing *Escherichia coli* O26 in cow slurry. Lett. Appl. Microbiol. *45*, 55–61.

Fremaux, B., Delignette-Muller, M.L., Prigent-Combaret, C., Gleizal, A., and Vernozy-Rozand, C. (2007b). Growth and survival of non-O157:H7 Shiga-toxin-producing *Escherichia coli* in cow manure. J. Appl. Microbiol. *102*, 89–99.

Hammar, M., Arnqvist, A., Bian, Z., Olsen, A., Normark, S., and others (1995). Expression of two csg operons is required for production of fibronectin-and Congo red-binding curli polymers in *Escherichia coli* K-12. Mol. Microbiol. *18*, 661–670.

Harris, L.J., Berry, E.D., Blessington, T., Erickson, M., Jay-Russell, M., Jiang, X., Killinger, K., Michel, F.C., Millner, P., Schneider, K., et al. (2013). A framework for developing research protocols for evaluation of microbial hazards and controls during production that pertain to the application of untreated soil amendments of animal origin on land used to grow produce that may be consumed raw. J. Food Prot. *76*, 1062–1084.

Hidalgo, G., Chen, X., Hay, A.G., and Lion, L.W. (2010). Curli Produced by *Escherichia coli* PHL628 Provide Protection from Hg(II). Appl. Environ. Microbiol. *76*, 6939.

Hoefel, D., Grooby, W.L., Monis, P.T., Andrews, S., and Saint, C.P. (2003). Enumeration of water-borne bacteria using viability assays and flow cytometry: a comparison to culture-based techniques. J. Microbiol. Methods *55*, 585–597.

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.

Jamieson, R.C., Gordon, R.J., Sharples, K.E., Stratton, G.W., and Madani, A. (2002). Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review. Can. Biosyst. Eng. *44*, 1–9.

Jensen, A.N., Storm, C., Forslund, A., Baggesen, D.L., and Dalsgaard, A. (2013). *Escherichia coli* Contamination of Lettuce Grown in Soils Amended with Animal Slurry. J. Food Prot. *76*, 1137–1144.

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. Appl. Environ. Microbiol. *64*, 3166–3174.

Moreira, S., Brown, A., Ha, R., Iserhoff, K., Yim, M., Yang, J., Liao, B., Pszczolko, E., Qin, W., and Leung, K.T. (2012). Persistence of *Escherichia coli* in freshwater periphyton: biofilm-forming capacity as a selective advantage. FEMS Microbiol. Ecol. *79*, 608–618.

Nicholson, F.A., Groves, S.J., and Chambers, B.J. (2005). Pathogen survival during livestock manure storage and following land application. Bioresour. Technol. *96*, 135–143.

Ogden, I.D., Fenlon, D.R., Vinten, A.J., and Lewis, D. (2001). The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. Int. J. Food Microbiol. *66*, 111–117.

Olsen, A., Jonsson, A., and Normark, S. (1989). Fibronectin Binding Mediated by a Novel Class of Surface Organelles on *Escherichia coli*.. Nature *338*, 652–655.

R Core Team. (2013) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria.

Rangel, J.M., Sparling, P.H., Crowe, C., Griffin, P.M., and Swerdlow, D.L. (2005). Epidemiology of *Escherichia coli* O157: H7 outbreaks, United States, 1982–2002. Emerg. Infect. Dis. *11*.

Römling, U., Sierralta, W.D., Eriksson, K., and Normark, S. (1998). Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the agfD promoter. Mol. Microbiol. *28*, 249–264.

Salvucci, A.E., Zhang, W., Morales, V.L., Cakmak, M.E., Hay, A.G., and Steenhuis, T.S. (2009). The impact of biofilm-forming potential and tafi production on transport of environmental *Salmonella* through unsaturated porous media. Biologia (Bratisl.) *64*, 460–464.

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

Sjöbring, U., Pohl, G., and Olsén, A. (1994). Plasminogen, absorbed by *Escherichia coli* expressing curli or by Salmonella enteritidis expressing thin aggregative fimbriae, can be activated by simultaneously captured tissue-type plasminogen activator (t-PA). Mol. Microbiol. *14*, 443–452.

Solomon, E.B., Yaron, S., and Matthews, K.R. (2002). Transmission of *Escherichia coli* O157:H7 from Contaminated Manure and Irrigation Water to Lettuce Plant Tissue and Its Subsequent Internalization. Appl. Environ. Microbiol. *68*, 397–400.

Stevik, T.K., Kari Aa, Ausland, G., and Fredrik Hanssen, J. (2004). Retention and removal of pathogenic bacteria in wastewater percolating through porous media: a review. Water Res. *38*, 1355–1367.

Vidal, O., Longin, R., Prigent-Combaret, C., Dorel, C., Hooreman, M., and Lejeune, P. (1998). Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression. J. Bacteriol. *180*, 2442–2449.

Vidovic, S., Block, H.C., and Korber, D.R. (2007). Effect of soil composition, temperature, indeigenous microflora, and environmental conditions on the survival of *Escherichia coli* O157:H7. Can. J. Microbiol. *53*, 822–829.

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.

White, A.P., and Surette, M.G. (2006). Comparative Genetics of the rdar Morphotype in Salmonella. J. Bacteriol. *188*, 8395–8406.

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.

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.

Chapter 2 : GENETIC SELECTION IN A PATHOGENIC *ESCHERICHIA COLI* STRAIN EXPOSED TO THE MANURE-AMENDED SOIL ENVIRONMENT

ABSTRACT

There is great interest in identifying manure management techniques to minimize the persistence of pathogenic *Escherichia coli* in agricultural fields and, thereby, decrease the risk of downstream contamination and human infection. Increased expression of extracellular fibers called curli, which are linked to environmental persistence of *E. coli*, has previously been linked to surface-application of manure, as opposed to incorporation into the soil. However, the selection mechanism behind the differing curli expression remains unknown. We hypothesized that initial and surviving isolates of a single pathogenic *E. coli* strain collected from agricultural soil microcosms would have distinct genomic differences that vary both by manure treatment method and time, indicative of individual-level genetic selection driven by exposure to the soil-manure environment. We applied whole genome sequencing technology to isolates collected from the microcosms over the course of eight weeks to address this question. In the 94 isolate sample, a total of two single-nucleotide polymorphisms (SNPs) were found, in unique strains from different weeks and manure treatments. Furthermore, a Poisson regression model revealed there was no significant difference ($\alpha = 0.05$) between the count of gaps in any gene that could be explained by week or treatment, and a linear regression model revealed no significant difference ($\alpha = 0.25$) in the number of discarded reads. The SNP, gap, and discarded read results demonstrate that there were no systematic genomic differences (i.e. individual-level selection) that can be explained by week or manure application treatment. Therefore, we recommend future research focus on the population-level selection of *E. coli* strains in the manure-amended soil environment.

INTRODUCTION

Water contamination with pathogenic *Escherichia coli* (*E. coli*) has been designated with "high" health significance by The World Health Organization, based on both the incidence and severity of outbreaks (World Health Organization, 2011). *Escherichia coli* is a minor though important component of the gut microbiota of many mammals, and exists in both commensal and pathogenic forms (Blount, 2015). Verotoxigenic (VTEC) *E. coli* strains produce Shiga toxin, which is responsible for hemolytic uremic syndrome, characterized by the destruction of red blood cells and acute renal failure (van Elsas et al., 2011). *Escherichia coli* O157:H7 is a VTEC serotype that significantly contributes to disease outbreaks in North America and worldwide (Carter et al., 2011; van Elsas et al., 2011).

Human exposure to pathogenic *E. coli* can occur through routes including consumption of contaminated food or surface water. In agricultural landscapes, *E. coli* sourced from manure application can contaminate surface waters through two major hydrologic routes: surface runoff and subsurface drainage (Jamieson et al., 2002). Controlling for initial concentration, the primary factor influencing the availability of *E. coli* for transport is the die-off rate of the bacteria (Reddy et al., 1981). However, the pathogenic *E. coli* O157:H7 has been found to persist in manure-amended soil for periods over 200 days (Franz et al., 2011). Anthropogenic alterations of the natural topography in central New York may increase *E. coli* exposure risk due to decreased length and duration of surface and groundwater flow pathways from agricultural fields to surface waters (Buchanan et al., 2013; Falbo et al., 2013). Therefore, manure management techniques to minimize the persistence of *E. coli*, and thereby decrease the risk of downstream contamination, are of great interest.

Upon leaving the hospitable mammalian intestinal environment, *E. coli* are exposed to a variety of different stressors that impact their survival, including nutrient limitation, UV radiation, and fluctuations in temperature, pH, and humidity (van Elsas et al., 2011). The specific environmental conditions *E. coli* encounter in the soil impact the duration of their survival. Multiple studies have suggested that water availability is the principal environmental factor affecting *E. coli* survival in soil. In general, *E. coli* has reduced mortality in soils with greater water content (Berry and Miller, 2005; Cools et al., 2001; Mubiru et al., 2000). For example, Berry and Miller (2005) found that *E. coli* O157:H7 populations kept under conditions of 0.11 g H2O g^{-1} dry soil decreased from an initial concnentation of 10^5 CFU g^{-1} dry soil to below the detection limit (10^1 CFU g^{-1} dry soil) in two days. In contrast, populations kept under conditions of 0.25 g H2O g^{-1} dry soil maintained populations of 104 CFU g^{-1} dry soil or greater for the 14 days of the experiment (Berry and Miller, 2005). Furthermore, cycling of the soil water content increases the mortality rate of *E. coli* (Berry and Miller, 2005). Prolonged *E. coli* survival has also been positively correlated with the amount of bioavailable carbon in the soil (Franz et al., 2008; Ma et al., 2011), finer textured soils (Brennan et al., 2014; Ma et al., 2011), lower soil temperatures (Cools et al., 2001; Kudva et al., 1998; Sjogren, 1994), and neutral to alkaline pH (Sjogren, 1994). One unique aspect of the agricultural soil environment is the added variable of manure application method. Soils with incorporated or injected manure support prolonged *E. coli* survival compared to soils with surface applied manure, due to less microbial exposure to atmospheric drying and UV irradiation from sunlight (Hutchison et al., 2004).

In a previous study, we demonstrated that the method of manure application affected the *E. coli* phenotypes that persisted in the soil, to be ultimately transported through subsurface drainage (Truhlar et al., 2015). Specifically, the phenotype affected was the presence or absence of curli.

Curli is the main proteinaceous component of *E. coli* biofilm, the extra-cellular matrix that encases aggregated cells, promoting survival (Costerton et al., 1995). Furthermore, curli expression has been associated with a number of stimuli including temperatures above 37^{o}C, low osmolarity, and aerobic conditions (Evans and Chapman, 2014). Of the surviving *E. coli* population collected from subsurface drainage, a significantly greater proportion of isolates from soil with incorporated manure did not express curli, compared to soil with surface-applied manure (Truhlar et al., 2015).

This phenotypic observation leaves open the question of the mechanism behind the differing curli expression. Specifically, is there an individual- or population-level shift in the gene pool? To date, genetic analyses aimed at understanding how genomic differences might influence *E. coli* survival in soil target a handful of genes associated with expression of the surface protein 'curli' and stress adaptations. For instance, both inter- and intra-strain comparisons of *E. coli* O157:H7 indicate that an intact *rpoS* gene, which is a global regulator of the general stress response, is correlated with increased survival in soil without manure (van Hoek et al., 2013; Ravva et al., 2014). However, it is unknown whether this variant is actively selected for in soil, or manure-amended soil, and at what level this potential selection occurs. Furthermore, we suggest that limiting genetic analyses to specific genes associated with one phenotype might result in missing other important changes occurring in populations. Whole genome sequencing data can provide a complete picture of the genetic differences that exist in surviving cells.

The present study aimed to determine how environmental challenges posed by the agricultural environment, and specifically different manure management techniques, impact the genomic population structure of *E. coli*. This work chose to focus on the variable of desiccation in order to isolate impacts of this important stressor in the soil environment. We hypothesized that initial

and surviving isolates of a single *E. coli* O157:H7 strain collected from agricultural soil microcosms would have distinct genotypes that vary both by manure treatment method and time, indicative of individual-level genetic selection driven by exposure to the soil-manure environment. We applied whole genome sequencing technology to discover genetic evolution in the released population, including single nucleotide polymorphisms (SNPs), insertions, and deletions.

METHODS

Seven treatments were investigated: (1) Soil only, (2) soil with un-inoculated, surface-applied manure, (3) soil with un-inoculated, incorporated manure, (4) soil with surface applied inoculant, (5) soil with incorporated inoculant, (6) soil with inoculated, surface-applied manure, and (7) soil with inoculated, incorporated manure. Four replicates were created for each treatment.

Microcosm setup

Soil (Langford channery silt loam) was collected from the upper 30 cm of a marginal agricultural field in upstate New York. All soil was air dried, sieved at 2 mm and homogenized. Soil properties characterized included clay, silt and sand content, water content, water holding capacity (WHC), percent carbon and percent nitrogen. Bovine manure was collected from the Cornell University teaching dairy facility. Both soil and manure were stored at 4^{o}C. Manure properties characterized included water content, percent carbon and percent nitrogen. Clay, silt and sand content were determined according to the hydrometer method (Kroetsch and Wang, 2008). Water content and WHC were determined according to the methods described by Franz et al. (2011). Percent carbon and nitrogen analyses were conducted at the Cornell University

Stable Isotope Laboratory (COIL; Ithaca, NY). The soil and manure characteristics are described in Table 1.

Table 2-1 Soil and manure characteristics.

	Sand (%)	**Silt (%)**	**Clay (%)**	**Bulk density (g/cm^3)**	**WHC (g water/g dry material)**	**Initial water content (g water/g dry material)**	**Nitrogen (%)**	**Carbon (%)**
Soil	36	62	2	0.79	0.7	0.17	0.3	3.2
Manure						4.4	2.3	40

Microcosms were established in plastic containers covered with perforated aluminum foil lids to permit drying. Twenty-four hours before the start of the experiment, the air-dried soil was adjusted to 60% WHC using sterile deionized water (Franz et al., 2008; Ma et al., 2011) and returned to storage at 4°C. Each soil microcosm received 100 g of the wetted soil.

Escherichia coli O157:H7 EDL933 Δ$stx_{1\text{-}2}$ was used as the inoculant. The genome of *E. coli* O157:H7 EDL933 has been fully sequenced and annotated (Latif et al., 2014). Its isogenic mutant derivative *E. coli* O157:H7 EDL933 Δ$stx_{1\text{-}2}$, which lacks the Shiga toxins Stx_1 and Stx_2 and has added cassettes for resistance to kanamycin (Km) and chloramphenicol (Cm), has been demonstrated to have an indistinguishable survival curve from the wild type strain in soils (Ma et al., 2011).

The overnight cultures of *E. coli* O157:H7 EDL933 Δ$stx_{1\text{-}2}$ (EcO157) in LB with Km (50 μg/ml) and Cm (25 μg/ml) were harvested by centrifugation at 4°C, washed three times with phosphate buffer, and re-suspended in sterile deionized water (Ma et al., 2011). The number of cells per ml of suspension was determined with a spectrophotometer, and used to calculate the volume of

inoculant required to reach a final concentration of 1 x 10^7 colony forming unit (cfu) per gram dry weight of soil.

To create the inoculated manure, a 25 g manure subsample was combined with an appropriate volume of inoculant based on the above calculations. Another 25 g manure subsample was combined with an equal volume of sterile deionized water to create the un-inoculated manure. From these slurries, two grams plus the weight of the inoculant volume required to achieve the desired final cell concentration were added to the inoculated and un-inoculated manure-treated microcosms. This is equivalent to a manure application rate of 2% w/w, which was used in related experiments by (Truhlar et al., 2015), and follows the manure application rates suggested by (Vidovic et al., 2007). The manure slurry was then either fully incorporated by mixing into the microcosm soil or left on the surface. For the inoculated soil treatments, inoculant plus a volume of sterile deionized water equal to the water content of the manure was added and then either incorporated or left on the surface. For the soil only (control) treatment, a volume of sterile deionized water equal to the total inoculant and manure moisture volume was added.

The microcosms were maintained at 15^{o}C for the course of the experiment using a temperature-controlled growth chamber. The microcosms were weighed weekly, and sterile deionized water was added to return the microcosms to their initial water content.

Microcosm sampling and E. coli enumeration

Sampling occurred at time 0 and weekly thereafter. Each week, samples were obtained both for later DNA extraction and for immediate enumeration of EcO157 colonies. At each sampling event, two 0.5 g soil samples were taken from each microcosm. Samples designated for DNA extraction were immediately frozen at -80^{o}C.

One mL of 0.1% peptone buffer was added to each soil sample designated for enumeration (Ma et al., 2011). The samples were then vortexed for 2 x 20 s (Ma et al., 2011). The resulting soil slurry was subjected to 10-fold serial dilutions to concentrations determined in preliminary experiments. Ten µL of the two to three highest dilutions were plated in triplicate. For inoculated treatments, LB agar supplemented with Km and Cm was used. For un-inoculated treatments, MacConkey agar was used. The plates were incubated overnight at 30^{o}C, and the results expressed as log colony forming units of *E. coli* per gram dry weight of soil.

From each of the two soil subsamples per microcosm, six EcO157 isolates (where possible) were selected from the enumeration plates. These were inoculated into 100 µL LB broth and grown at 37^{o}C for 16 h. One hundred µL of 1:1 glycerol and sterile deionized water solution was added to the overnight cultures to achieve a final glycerol concentration of 25%. The cell suspensions were frozen at -80^{o}C for later use.

Selection of isolates for sequencing

To identify potential isolates of interest for sequencing, two phenotypic assays were completed. Fifteen EcO157 isolates were randomly selected from each of the four treatments collected during week zero and week six, for a total of 120 isolates. The first assay consisted of plating the isolates on Congo Red (CR) agar and growing them at 30^{o}C for seven days. The plates were inspected every 24 hr for morphology differences, indicative of differences in curli and cellulose production. Specifically, red and rough colonies indicate the production of curli and cellulose, respectively, while white and smooth colonies indicate the absence of these components of the extracellular matrix (Hammar et al., 1995). *E. coli* PHL628 WT and *E. coli* PHL628 Δ*csgA*, which has a known curli deficiency, were used as visual controls. The second assay consisted of

growing a random subset of the Congo Red assay isolates (six per treatment for both weeks zero and six) in LB broth at 37°C and 15°C. Growth was measured over a 15 hr period using a spectrophotometer. A specific growth constant was calculated for each treatment under these conditions using the specific growth rate equation for bacteria (Equation 1) (Neidhardt et al., 1990). The non-parametric Kruskal-Wallis test was used to determine whether the specific growth constant differed significantly between treatments.

$$ln\frac{N_2}{N_1} = k(t_2 - t_1) \qquad \textbf{Eqn. 1}$$

The following distribution of EcO157 isolates from inoculated samples were selected for whole genome sequencing: six isolates representing the initial population, 24 isolates (six per treatment) collected after three weeks, 48 isolates (12 per treatment) collected after six weeks, and 18 isolates (six per treatment, except soil surface-applied). All isolates were randomly selected except for those collected during week six; these were selected to include isolates with observed morphology differences in the CR agar test described above.

DNA extraction, sequencing, and bioinformatics

DNA was extracted using a QIAamp DNA minikit (Qiagen, Hilden, Germany) with addition of an RNase A (100 mg/ml; Qiagen) treatment as described in Denes et al. (2015).

A Nextera XT DNA sample preparation kit (Illumina, Inc., San Diego, CA) was used to prepare the library, and 2x75 bp paired-end reads were obtained by sequencing the library on the Illumina NextSeq platform. Single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs) were called using both a reference-based and *de novo* detection method. For the reference-based method, EcO157 isolate reads were mapped against the chromosome and

plasmid of *E. coli* O157:H7 str. EDL933 (GenBank accession numbers CP008957.1 and CP008958.1) using CLC Genomics Workbench (CLC Bio, Qiagen), following the pipeline described by (Van den Hoecke et al., 2015). The Cortex variation assembler, cortex_var, version 1.0.5.14 (Iqbal et al., 2012, 2013), was used for the *de novo* variant detection (both SNP and insertion/deletion events) as outlined by (den Bakker et al., 2014). For both methods, SNPs with a minimal coverage of 50% of the genome-wide average coverage (GAC), a minimal variant coverage of 50% of the GAC, and minimum alternative variant frequency of 95% were considered for analysis (Denes et al., 2015).

We performed several more tests to further explore possible genomic differences overlooked by the mapping-to-reference method used in CLC Workbench (Figure 1). First, we extracted the reference genome coverage for each sequenced strain to identify all regions with zero coverage. Within each strain, these regions were grouped based on gene annotation. The rates of occurrence of zero-coverage regions in a given gene were compared using a Poisson model for count data, with the combination of week-treatment as the sole explanatory variable. This was done to accommodate the fact that the control occurred only in week 0 and therefore could not have week as a separate explanatory variable. The same was true for the soil, surface-applied treatment, which could not be isolated in week 8.

Finally, since mapping to the reference resulted in a fairly consistent number of reads that did not match to the reference, we investigated the content of these discarded reads, which might represent additional genetic material that the strains received through horizontal gene transfer from other bacteria present in the soil (Davison, 1999). The metagenomics RAST server (MG-RAST) pipeline was used to annotate the discarded reads (Figure 1; Meyer et al., 2008). We then compared the total number of discarded reads that could be annotated with a predicted

feature using a linear model with week-treatment as the sole explanatory variable. Again, this was done to accommodate the fact that the control occurred only in week 0, and that the soil, surface-applied treatment could not be isolated in week 8, and therefore both these instances could not have week as a separate explanatory variable. The predicted features that were used in this comparison were the top 25 genuses (by frequency). Functional subsystems were also included, except for those that were correlated with all other subsystems (all $p < 0.01$; data not shown). The number of predicted features that met these criteria were rank-transformed to meet normality requirements. Additionally, specific annotations, normalized to the number of predicted features for a given strain, were individually fit to a linear regression with week-treatment as the explanatory variable (Table S1). The *p*-values from these regressions were adjusted to account for multiple comparisons, using the Benjamini-Hochberg method (McDonald, 2014).

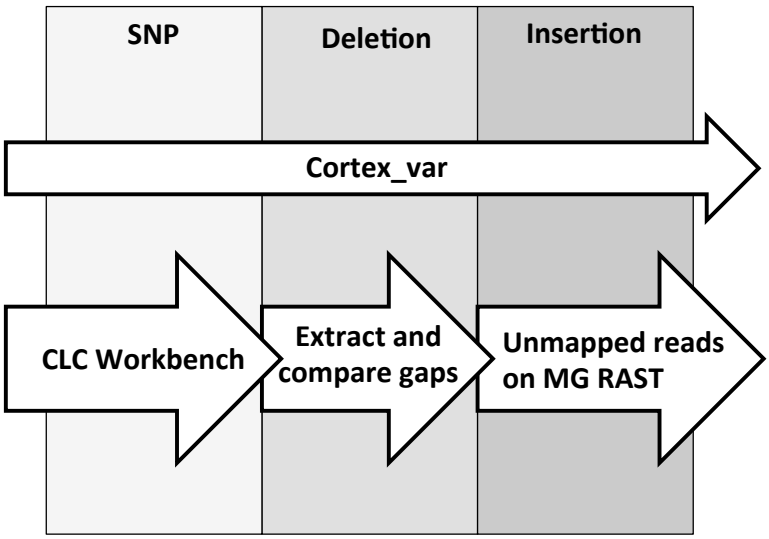


Figure 2-1 Schematic of the addition steps taken to ensure the mapping-to-reference method used in CLC Workbench identified all genomic differences.

RESULTS AND DISCUSSION

E. coli survival in soil

In all inoculated treatments, *E. coli* O157:H7 EDL933 Δ*stx*$_{1\text{-}2}$ (EcO157) survived at least six weeks after the start of the experiment, but experienced a steady decay over this time period (Figure 2). After seven weeks, EcO157 could no longer be detected in the surface applied inoculant treatment (Figure 2). For all other treatments, EcO157 was present at the end of the experiment, eight weeks after the initial inoculation (Figure 2). This falls within the range of *E. coli* soil survival durations reported in previous studies using a similar initial inoculation density (ca. 1 x 10^7 cfu per gram dry weight of soil) in unautoclaved, manure-amended soil. For instance, Ma et al. (2011) reported the same strain survived between 32 and 113 days depending on the soil type used.

A decay constant was calculated for each survival curve using the specific growth rate equation for bacteria (Eqn. 1) (Neidhardt et al., 1990). Overall, the surface-applied treatments decayed faster (had more negative decay constants) than their incorporated treatment counterparts. The decay constant for the soil surface-applied treatment was significantly more negative than the decay constants for both manure treatments ($p < 0.05$; Figure 3), indicating faster die-off. Hutchison et al. (2004) found a similar trend, with *E. coli* surviving longer when injected into the soil than when spread on the soil surface. This is thought be driven by greater exposure to drying and UV light exposure (Hutchison et al., 2004). ,However, in the current experiment, drying would be the only relevant stressor.

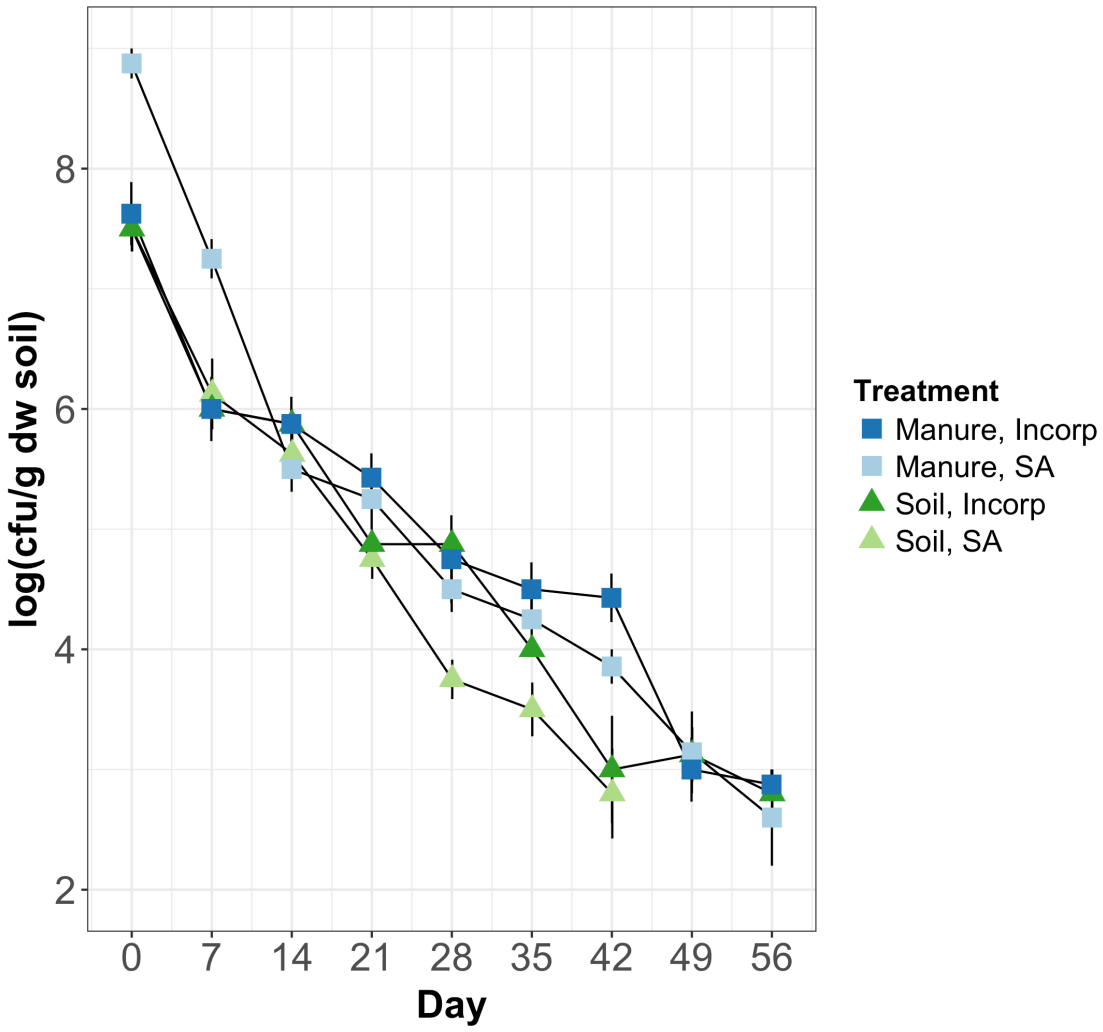


Figure 2-2 Survival of *E. coli* O157:H7 EDL933 in soil microcosms with and without added manure ("Manure" and "Soil," respectively). "Incorp" indicates mixing of the soil after *E. coli* application; "SA" indicates surface application (i.e., no mixing). Data represents an average (+/- 1SD) of four replicates.

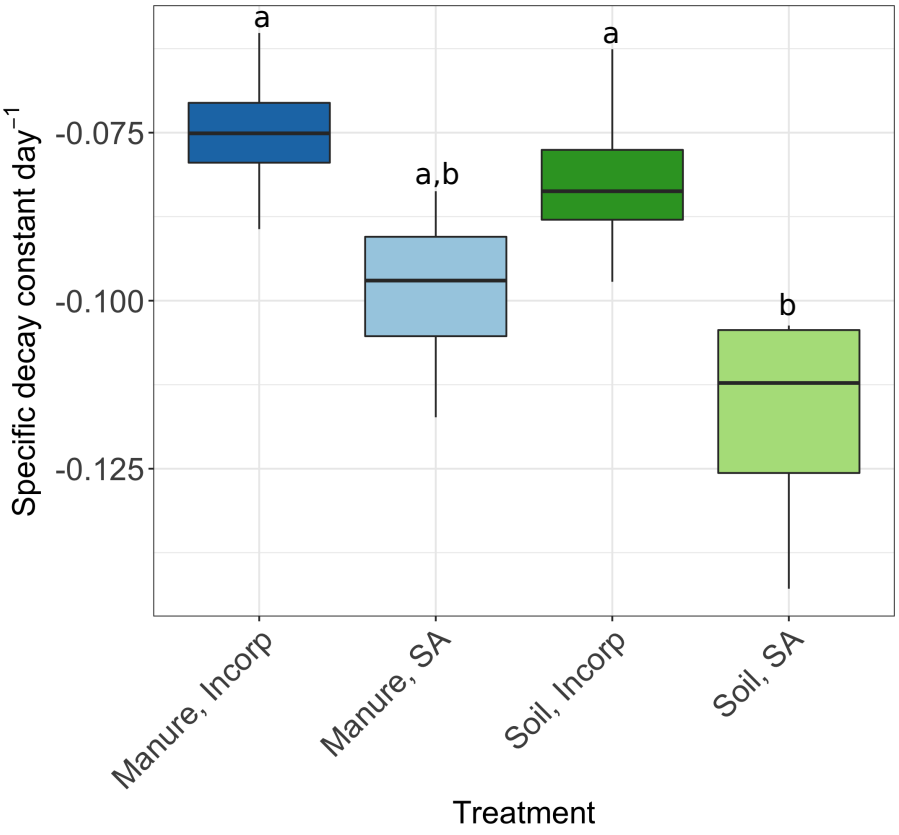


Figure 2-3 Specific decay constants for *E. coli* O157:H7 EDL933 in soil microcosms with and without added manure ("Manure" and "Soil," respectively). "Incorp" indicates mixing of the soil after *E. coli* application; "SA" indicates surface application (i.e., no mixing). Data represents an average (+/- 1SD) of four replicates. Different letters indicate significant differences in the specific decay constant between treatments.

Phenotypic assays

The specific growth rate for EcO157 isolates from the different treatments grown over a 14 hr period at 37^{o}C in LB broth did not differ significantly ($p > 0.50$; Figures S1 and S2). However, the specific growth rate calculated only for the time period where the isolates were leaving the lag phase (6 to 10 hrs), did differ significantly between treatments, with all the treatments except soil surface-applied leaving the lag phase at a significantly greater rate than the control (week 0)

treatment (Figure S3). Growth could not be detected for isolates kept at 15°C (data not shown). Ishii et al. (2006) observed similar growth patterns for naturalized *E. coli* soil populations. While the *E. coli* grew in soils kept at or above 30°C, there was no cell density increase when soil was kept at 15°C. However, upon increasing the temperature of the soil to 37°C, cell growth resumed. This suggests that *E. coli* can persist at low temperatures and then return to growth once the conditions become suitable (Ishii et al., 2006). Furthermore, the water content of our experimental soil was 0.39 g H_2O per g dry soil (60% WHC), which likely places our microcosms within a range of conditions suitable for aerobic activity (Miller and Berry, 2005). Miller and Berry (2005) identified a water content cutoff for the soil used in their experiment, below which cellular inactivity was the dominant form of microbial metabolism. A direct comparison is not possible because our manure application rate of 2% w/w was less than the lowest reported rate of 5% w/w, and water content translates to different water availability in different soils (Miller and Berry, 2005). However, speculate that our EcO157 populations existed in an inactive state, characterized by primarily cell maintenance with minimal to no cell division and obvious cell die-off, while exposed to cyclic drying periods at 15°C. This is supported by the lack of observed growth during any portion of the experiment, and suggested by the combination of cooler temperature (15°C), less available nutrients from manure, and cyclic drying that characterized the conditions of our experiment.

The isolates from distinct treatments showed phenotypic differences when grown on Congo Red agar for seven days at 30°C. Specifically, seven out of 15 randomly selected week 6 isolates from the surface applied manure treatment developed a white growth on top of an otherwise red colony morphology (Figure 4). All other isolates, including the control "week 0" isolates, exhibited red colony morphologies. Red colonies indicate the production of curli, while white

colonies indicate the absence of curli (Hammar et al., 1995). While the results of a binomial regression for two disjoint outcomes (i.e., red or mixed morphotype) by treatment was not significant ($p > 0.05$; data not shown), the Congo Red morphology results still suggest that prolonged exposure to different manure application methods, specifically to surface-applied manure in this case, resulted in distinct phenotypic populations in the soil microcosms. In unpublished data from an earlier study (Figure S4), we found a similar trend. Namely, the ratio of isolates with a mixed red and white morphotype to isolates with a red morphotype was highest in the *E. coli* population sampled from the top 2 cm of soil in a column treated with surface applied manure. For all treatments, the red morphotype was the predominant morphotype in the sampled population.

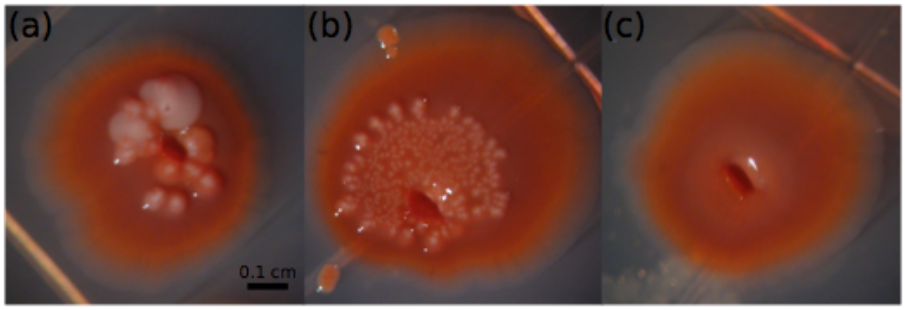


Figure 2-4 Sample morphologies of isolates grown on Congo Red agar for seven days at 30C. Isolates (a) and (b) were collected from the surface-applied manure treatment after six weeks, and exemplify the red morphology overlain with raised white growth that was unique to isolates from this treatment. Isolate (c) was collected from the same soil microcosm after six weeks, and is representative of the smooth, uniform red morphology manifest by isolates collected from all treatments other than surface-applied manure. Pictures were taken under 16X magnification.

Genetic sequencing

There was no significant difference between the number of gaps in any gene that could be explained by week or treatment ($p > 0.05$; Table S2). Gaps did tend to occur in the same genes

across isolates, including Week 0 isolates, suggesting that our initial population differed from the reference sequence (Table S2). Furthermore, there was no significant difference found in the number of unmapped reads that could be annotated as predicted features or assigned to any specific taxa, or subsystem via MGRAST that could be explained by week or treatment (Figure 5; Table 2). Combined, these results demonstrate that there was no significant difference in the amount or content of genetic material either deleted from the genome or discarded through the reference mapping process, for strains from different weeks or treatments.

Our analyses in CLC Workbench and Cortex_Var identified two mutations in the 94 strains we sampled. Both mutations were missense, present only once in these strains, and occurred in all strains regardless of sample week or treatment (Table 3). The SNP in SWL0050 was confirmed with Sanger sequencing. The full metrics for mapping reads to each isolate in CLC Workbench are provided in Table S3.

Of the two mutations, only one occurred in a protein with an identified function, the AIDA-I adhesion-like protein. AIDA-I was first identified as conferring the diffuse adherence phenotype to enteropathogenic *E. coli* strains (Benz and Schmidt, 1989; Benz and Schmidt, 1992), characterized by uniform adherence to the whole host cell surface (Scaletsky et al., 1984). It also enhances bacterial aggregation, biofilm formation, and invasion of host cells (Klemm et al., 2006). It has been shown that insertions in different regions of the protein can result in reduced ability of AIDA-I to mediate one or more of these functions, depending on the location of the insertion (Charbonneau and Mourez, 2007). However, the same study identified multiple insertion locations through the gene that did not alter function (Charbonneau and Mourez, 2007). It is therefore uncertain whether the present mutation, a non-synonymous single nucleotide polymorphism (SNP) in the N-terminal third of the protein, would have any effect on AIDA-I

function. The N-terminal third of AIDA-I contains the cell-binding domain (Charbonneau and Mourez, 2007). A first attempt to understand the effect of this mutation would be to compare the mutant strain's ability to bind to epithelial cells with a control strain.

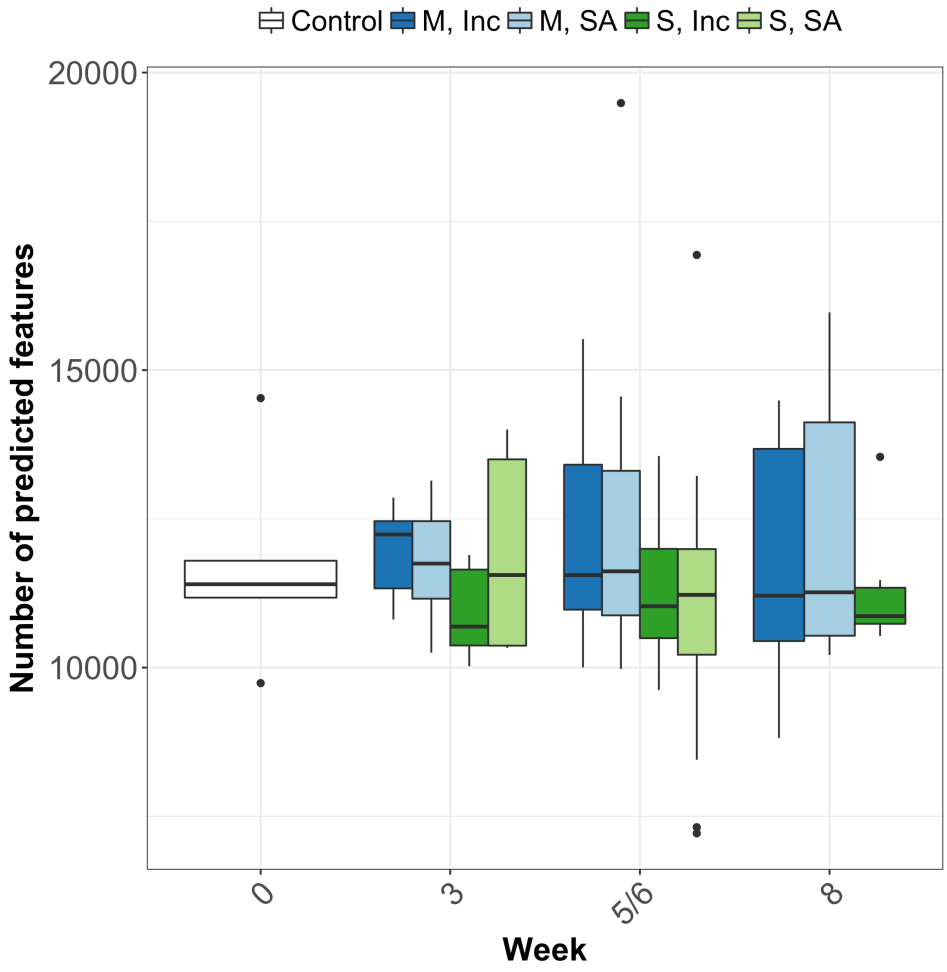


Figure 2-5 The number of discarded reads, obtained from each isolate after mapping reads to the reference genome, that were annotated with a predicted feature through the MG Rast server. The reads are plotted by both the week the isolate was collected and the soil-manure treatment to which the isolate was exposed. Four points with values greater than 25,000 were excluded to make the plot easier to read. These points came from week 3 soil-incorporated, week 5/6 manure-incorporated, week 5/6 manure surface-applied, and week 8 manure-incorporated strains.

Table 2-2 The read count for MGRAST-annotated predicted features were individually modeled as a linear function of week-treatment; the resulting *p*-values are shown here. The predicted features that were included in the analysis were the top 25 genuses (by frequency) and functional subsystems, except for those that were correlated with all other functional subsystems. Subsequently, all *p* values were then ordered and compared to threshold values calculated using the Benjamini-Hochberg (BH) procedure, to account for multiple comparisons. The false discovery rate (α) was set at 0.25 to ensure no possible significant factors were overlooked. A *p* value less than the corresponding threshold is considered significant.

Response variable	*P* value	BH Threshold
Yersinia	0.11	0.01
Phages, Prophages, Transposable elements, Plasmids	0.17	0.01
Salmonella	0.25	0.02
Nitrogen Metabolism	0.31	0.03
Shigella	0.31	0.03
Enterobacter	0.33	0.04
Escherichia	0.38	0.05
Respiration	0.40	0.05
Phosphorus Metabolism	0.41	0.06
Bacteria	0.41	0.07
Iron acquisition and metabolism[1]	0.31	0.07
Stress Response	0.48	0.08
Pectobacterium	0.51	0.09
Dormancy and Sporulation[2]	0.54	0.09
Serratia	0.52	0.10
Lambda like viruses	0.55	0.11
Viruses	0.56	0.11
Ruminococcus	0.59	0.12
Cronobacter	0.60	0.13
Bacteroides	0.61	0.14
Populus	0.61	0.14
Erwinia	0.62	0.15
Citrobacter	0.66	0.16
Klebsiella	0.68	0.16
Gallus	0.72	0.17
Pseudomonas	0.73	0.18
Streptococcus	0.74	0.18
Metabolism of Aromatic Compounds[1]	0.84	0.19
Potassium metabolism	0.79	0.20
Motility and Chemotaxis[1]	0.89	0.20
Vibrio[2]	0.76	0.21
Haemophilus	0.83	0.22
Acinetobacter	0.84	0.22
Neisseria	0.85	0.23
Drosophila	0.87	0.24
Propionibacterium	0.90	0.24
Mycobacterium[2]	0.90	0.25

[1]Data were log transformed
[2]Data were square-root transformed

Table 2-3 Single nucleotide polymorphisms (SNPs) identified in all 94 resequenced isolates of *E. coli* O157:H7 EDL933 Δ*stx*$_{1\text{-}2}$.

Nucleotide position*	Feature	Mutation Type	WT	Mutant	Strain	Strain Description
380467	AidA-I adhesion-like protein	Missense	A	G	SWL0019	Week 3, surface-applied, with manure
1380405	Hypothetical protein	Missense	C	T	SWL0050	Week 5, incorporated, soil only

*Positions are relative to Modified_ref_1_v2

The SNP, gap, and discarded read results demonstrate that there were no systematic genomic differences that can be explained by week or manure application treatment. This suggests that survival was not driven by genetic selection at the individual level. However, the clear trend seen in the Congo Red phenotypes between treatments leads us to conclude that individual survival was not completely stochastic. No genetic changes were detected in any known curli regulators in the isolates we sequenced (Table 4), therefore, the differential expression resulting in te altered phenotypes might be driven by epigenetic regulation of the curli genes. For instance, it is known that *Salmonella* strains lacking the DNA adenine methyltransferase enzyme, which methylates N-6 of adenine in GATC sequences, show diminished production of curli (Aya Castañeda et al., 2015). Therefore, methylation is a potentially important epigenetic control on curli production in *E. coli* as well.

Finally, our results lead us to hypothesize that the phenotypic differences observed by Truhlar et al. (2015), between populations of E. coli that had been transported from soil, were driven by population-level selection dependent on some combination of the environment (incorporated manure versus surface-applied manure) and the transport process. It has been shown that the variables of pH, water content, and percent organic matter can partially account for genotype

Table 2-4 Summary of known curli regulators (Evans and Chapman, 2014). The location of each regulator in the E. coli O157:H7 EDL933 references genome used in this study is provided, along with an indication of whether a gap or SNP was found in the regulator through mapping to the reference genome on CLC Workbench.

Regulator	**Location in Modified_ref_1_v2**	**CLC gap?**	**CLC SNP?**
RpoS	3674778 – 3674548 and 3675540 – 3674818	No	No
Curlin genes transcriptional activator (Crl)	295717 – 296118	No	No
CsgD	1549726 – 1549076	No	No
MlrA	3031981 – 3032628	No	No
Cra (FruR)	92634 – 93638	No	No
Cyclic AMP receptor protein (Crp)	4293493 – 4294125	No	No
Transcriptional regulator TetR family (RcdA)	1012481 – 1013017	No	No
Integration host factor (IHF)	1180803 – 1181087 and 2489382 – 2489083	No	No
DNA binding protein H-NS	1822808 – 1822395	No	No
CpxA/R	4991028 – 4992401; 4992398 – 4993030	No	No
EnvZ/OmpR	4334177 – 4335529; 4335526 – 4336245	No	No
Sensory histidine kinase in two-component regulatory system (RstB)/RstA	2376799 – 2378100; 2376076 – 2376795	No	No
ArcA/B	545101 – 5545817; 4180509 – 4182845	No	No
BasS/R	5235336 – 5236436; 5236437 – 5237105	No	No
OmrA/B	*3786055 – 3785968; *3786252 – 3786171	No	No
McaS	*2180827 – 2180922	No	No
GcvB	*3752562 – 3752767	No	No
RprA	*2464200 – 2464305	No	No
ArcZ	*4180397 – 4180517	No	No
SdsR	*2619232 – 2619112	No	No

*Indicates that the gene annotation did not exist in EDL933, but a match was found after BLASTing the E. coli K12 segment against EDL933.

Variation across a landscape for microbial species including *E. coli, Fusarium,* and *Burkholderia cepacia* (Bergholz et al., 2011; Ramette and Tiedje, 2007; Yergeau et al., 2010). Further work could seek to test (1) if differing manure management practices influence the population-level genetic structure of *E. coli* and (2) whether the transport process itself is another variable that determines the genetic structure of *E. coli* populations sourced from subsurface drainage. We plan to use environmental *E. coli* isolates collected during this experiment to test the first question. This knowledge would help guide manure management practices to avoid creation of reservoirs of stress-tolerance genes, such as those promoting curli formation, in both the soil and surface-water landscapes.

CONCLUSIONS

Through whole-genome sequencing, we demonstrated that there was no selective pressure on individual genomes for pathogenic *E. coli* O157:H7 EDL933 Δ$stx_{1\text{-}2}$ (EcO157) driven by exposure to cyclic drying in the soil and different manure application methods. This result brings us closer to a mechanistic understanding of how manure application method relates to population-wide phenotypic patterns previously observed in *E. coli* isolated from the soil. EcO157 survival in this experiment was not completely stochastic, as evinced by differing curli phenotypes by manure application method, suggesting that epigenetic regulation might also play a role in phenotypic patterns. However, we suggest that the greatest need for future research is investigating whether and how population-level genetic selection results from different manure application methods. This could provide a much-needed mechanistic basis for choosing manure management techniques that do not favor the selection of stress-resistance genes, such as those promoting curli formation, thereby minimizing the persistence of pathogenic *E. coli* in the agricultural environment.

Supplemental Table 2-1 Annotations downloaded from the MG-RAST server for unmapped reads from each isolate.

Variable group	Variable
Overall sequence breakdown	Predicted feature
	*Unknown
	*Failed QC
Domain	*Eukaryota
	Bacteria
	Viruses
	*Other sequences
	*None
Genus	Escherichia
	Shigella
	Salmonella
	Yersinia
	Citrobacter
	Erwinia
	Klebsiella
	Enterobacter
	Bacteroides
	Lambda like viruses
	Gallus
	Ruminococcus
	Drosophila
	Haemophilus
	Acinetobacter
	Pseudomonas
	Neisseria
	Vibrio
	Serratia
	Mycobacterium
	Populus
	Cronobacter
	Propionibacterium
	Streptococcus
	Pectobacterium
	*Unclassifed derived from Podovirae
	*Unclassified derived from viruses
	*Unclassifed derived from other sequences
	*None
Subsystems	Respiration
	Iron acquisition and metabolism
	Motility and Chemotaxis
	Nitrogen Metabolism
	+RNA Metabolism
	+Protein Metabolism
	Phosphorus Metabolism
	+Fatty Acids Lipids and Isoprenoids
	Stress Response
	+Sulfur Metabolism
	+Cell Division and Cell Cycle
	+DNA Metabolism
	+Virulence Disease and Defense

Supplemental Table 2.1 (Continued)

Variable group	Variable
Subsystems	[+]Cofactors Vitamins Prosthetic Groups Pigments
	[+]Carbohydrates
	[+]Cell Wall and Capsule
	[+]Membrane Transport
	[+]Dormancy and Sporulation
	[+]Clustering based subsystems
	[+]Amino Acids and Derivatives
	Potassium metabolism
	Phages Prophages Transposable elements Plasmids
	Metabolism of Aromatic Compounds
	[+]Regulation and Cell signaling
	[+]Nucleosides and Nucleotides

*Variable was excluded from further analysis

[+]Subsystem was positively correlated with all other subsystems, and was therefore excluded from the analysis for parsimony

Supplemental Table 2-2 Comparison of the number of gaps found in genes in weeks 0, 3, 5/6, and 8 isolates using a Poisson model for count data, and week-treatment as the explanatory variable. Protein IDs are the Genbank accession numbers for the proteins from the chromosome and plasmid of *E. coli* O157:H7 str. EDL933 (GenBank accession numbers CP008957.1 and CP008958.1; Latif et al., 2014).

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Hypothetical protein CDS	AIG66563.1	(Intercept)	-2.2E+01	-1.2E-03	1.00
		WeekTreatWeek 3, M, Inc	8.0E-10	3.1E-14	1.00
		WeekTreatWeek 3, M, SA	8.0E-10	3.1E-14	1.00
		WeekTreatWeek 3, S, Inc	8.0E-10	3.1E-14	1.00
		WeekTreatWeek 3, S, SA	8.0E-10	3.0E-14	1.00
		WeekTreatWeek 5/6, M, Inc	8.0E-10	3.5E-14	1.00
		WeekTreatWeek 5/6, M, SA	8.0E-10	3.5E-14	1.00
		WeekTreatWeek 5/6, S, Inc	8.0E-10	3.5E-14	1.00
		WeekTreatWeek 5/6, S, SA	2.1E+01	1.1E-03	1.00
		WeekTreatWeek 8, M, Inc	8.0E-10	3.1E-14	1.00
		WeekTreatWeek 8, M, SA	8.0E-10	3.1E-14	1.00
		WeekTreatWeek 8, S, Inc	8.0E-10	3.1E-14	1.00
Hypothetical protein CDS	AIG66691.1	(Intercept)	-2.2E+01	-1.2E-03	1.00
		WeekTreatWeek 3, M, Inc	9.9E-10	3.9E-14	1.00
		WeekTreatWeek 3, M, SA	9.9E-10	3.9E-14	1.00
		WeekTreatWeek 3, S, Inc	9.9E-10	3.9E-14	1.00
		WeekTreatWeek 3, S, SA	9.9E-10	3.7E-14	1.00
		WeekTreatWeek 5/6, M, Inc	9.9E-10	4.4E-14	1.00
		WeekTreatWeek 5/6, M, SA	9.9E-10	4.4E-14	1.00
		WeekTreatWeek 5/6, S, Inc	2.0E+01	1.0E-03	1.00
		WeekTreatWeek 5/6, S, SA	2.0E+01	1.0E-03	1.00
		WeekTreatWeek 8, M, Inc	9.9E-10	3.9E-14	1.00
		WeekTreatWeek 8, M, SA	9.9E-10	3.9E-14	1.00
		WeekTreatWeek 8, S, Inc	9.9E-10	3.9E-14	1.00
Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N- terminal CDS	AIG67225.1	(Intercept)	-7.4E-17	-1.7E-16	1.00
		WeekTreatWeek 3, M, Inc	5.8E-17	9.5E-17	1.00
		WeekTreatWeek 3, M, SA	7.2E-17	1.2E-16	1.00
		WeekTreatWeek 3, S, Inc	9.6E-17	1.6E-16	1.00
		WeekTreatWeek 3, S, SA	3.4E-17	5.3E-17	1.00
		WeekTreatWeek 5/6, M, Inc	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 5/6, M, SA	4.5E-17	8.4E-17	1.00
		WeekTreatWeek 5/6, S, Inc	8.0E-02	1.5E-01	0.88
		WeekTreatWeek 5/6, S, SA	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 8, M, Inc	4.2E-17	7.0E-17	1.00
		WeekTreatWeek 8, M, SA	4.0E-17	6.5E-17	1.00
		WeekTreatWeek 8, S, Inc	4.9E-17	8.1E-17	1.00
Transposase C CDS	AIG67229.1	(Intercept)	-1.6E+00	-1.6E+00	0.11
		WeekTreatWeek 3, M, Inc	5.1E-01	4.2E-01	0.68

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Transposase C CDS	AIG67229.1	WeekTreatWeek 3, M, SA	1.2E+00	1.1E+00	0.28
		WeekTreatWeek 3, S, Inc	5.1E-01	4.2E-01	0.68
		WeekTreatWeek 3, S, SA	1.5E-14	1.1E-14	1.00
		WeekTreatWeek 5/6, M, Inc	2.2E-01	1.9E-01	0.85
		WeekTreatWeek 5/6, M, SA	1.2E+00	1.1E+00	0.26
		WeekTreatWeek 5/6, S, Inc	2.2E-01	1.9E-01	0.85
		WeekTreatWeek 5/6, S, SA	1.1E+00	1.0E+00	0.32
		WeekTreatWeek 8, M, Inc	1.4E+00	1.3E+00	0.19
		WeekTreatWeek 8, M, SA	5.1E-01	4.2E-01	0.68
		WeekTreatWeek 8, S, Inc	-1.8E-01	-1.3E-01	0.90
Transposase CDS	AIG67230.1	(Intercept)	-9.2E-01	-1.3E+00	0.20
		WeekTreatWeek 3, M, Inc	9.2E-01	1.1E+00	0.26
		WeekTreatWeek 3, M, SA	7.3E-01	8.8E-01	0.38
		WeekTreatWeek 3, S, Inc	9.2E-01	1.1E+00	0.26
		WeekTreatWeek 3, S, SA	6.9E-01	8.0E-01	0.42
		WeekTreatWeek 5/6, M, Inc	2.2E-01	2.7E-01	0.78
		WeekTreatWeek 5/6, M, SA	8.3E-01	1.1E+00	0.28
		WeekTreatWeek 5/6, S, Inc	8.3E-01	1.1E+00	0.28
		WeekTreatWeek 5/6, S, SA	7.3E-01	9.5E-01	0.34
		WeekTreatWeek 8, M, Inc	9.2E-01	1.1E+00	0.26
		WeekTreatWeek 8, M, SA	9.2E-01	1.1E+00	0.26
		WeekTreatWeek 8, S, Inc	5.1E-01	5.9E-01	0.56
Hypothetical protein CDS	AIG67231.1	(Intercept)	-2.2E-01	-4.5E-01	0.66
		WeekTreatWeek 3, M, Inc	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 3, M, SA	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 3, S, Inc	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 3, S, SA	-1.0E-15	-1.5E-15	1.00
		WeekTreatWeek 5/6, M, Inc	-1.8E-01	-3.0E-01	0.77
		WeekTreatWeek 5/6, M, SA	1.4E-01	2.3E-01	0.82
		WeekTreatWeek 5/6, S, Inc	1.4E-01	2.3E-01	0.82
		WeekTreatWeek 5/6, S, SA	-6.5E-02	-1.1E-01	0.91
		WeekTreatWeek 8, M, Inc	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 8, M, SA	4.1E-02	6.1E-02	0.95
		WeekTreatWeek 8, S, Inc	2.2E-01	3.5E-01	0.73
Antigen 43 precursor CDS	AIG67322.1	(Intercept)	-2.1E+01	-1.9E-03	1.00
		WeekTreatWeek 3, M, Inc	2.0E+01	1.7E-03	1.00
		WeekTreatWeek 3, M, SA	-3.5E-10	-2.3E-14	1.00
		WeekTreatWeek 3, S, Inc	-3.5E-10	-2.3E-14	1.00

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Antigen 43 precursor CDS	AIG67322.1	WeekTreatWeek 3, S, SA	-3.5E-10	-2.2E-14	1.00
		WeekTreatWeek 5/6, M, Inc	1.9E+01	1.6E-03	1.00
		WeekTreatWeek 5/6, M, SA	-3.5E-10	-2.6E-14	1.00
		WeekTreatWeek 5/6, S, Inc	1.9E+01	1.6E-03	1.00
		WeekTreatWeek 5/6, S, SA	1.9E+01	1.6E-03	1.00
		WeekTreatWeek 8, M, Inc	-3.5E-10	-2.3E-14	1.00
		WeekTreatWeek 8, M, SA	-3.5E-10	-2.3E-14	1.00
		WeekTreatWeek 8, S, Inc	-3.5E-10	-2.3E-14	1.00
Antigen 43 precursor CDS	AIG67323.1	(Intercept)	9.4E-18	2.1E-17	1.00
		WeekTreatWeek 3, M, Inc	8.7E-17	1.4E-16	1.00
		WeekTreatWeek 3, M, SA	-2.0E-18	-3.3E-18	1.00
		WeekTreatWeek 3, S, Inc	9.0E-17	1.5E-16	1.00
		WeekTreatWeek 3, S, SA	-9.1E-17	-1.4E-16	1.00
		WeekTreatWeek 5/6, M, Inc	8.0E-02	1.5E-01	0.88
		WeekTreatWeek 5/6, M, SA	1.5E-01	3.0E-01	0.77
		WeekTreatWeek 5/6, S, Inc	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 5/6, S, SA	8.0E-02	1.5E-01	0.88
		WeekTreatWeek 8, M, Inc	6.1E-17	1.0E-16	1.00
		WeekTreatWeek 8, M, SA	4.4E-17	7.3E-17	1.00
		WeekTreatWeek 8, S, Inc	6.5E-17	1.1E-16	1.00
Ferredoxin reductase CDS	AIG67324.1	(Intercept)	-2.2E-01	-4.5E-01	0.66
		WeekTreatWeek 3, M, Inc	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 3, M, SA	-4.7E-01	-6.2E-01	0.54
		WeekTreatWeek 3, S, Inc	-8.8E-01	-1.0E+00	0.31
		WeekTreatWeek 3, S, SA	2.2E-01	3.3E-01	0.74
		WeekTreatWeek 5/6, M, Inc	-6.5E-02	-1.1E-01	0.91
		WeekTreatWeek 5/6, M, SA	-1.8E-01	-3.0E-01	0.77
		WeekTreatWeek 5/6, S, Inc	-4.7E-01	-7.3E-01	0.47
		WeekTreatWeek 5/6, S, SA	-1.8E-01	-3.0E-01	0.77
		WeekTreatWeek 8, M, Inc	-1.8E-01	-2.6E-01	0.80
		WeekTreatWeek 8, M, SA	-8.8E-01	-1.0E+00	0.31
		WeekTreatWeek 8, S, Inc	-8.8E-01	-1.0E+00	0.31
Transposase CDS	AIG67682.1	(Intercept)	-5.1E-01	-8.8E-01	0.38
		WeekTreatWeek 3, M, Inc	-1.8E-01	-2.2E-01	0.82
		WeekTreatWeek 3, M, SA	5.1E-01	7.2E-01	0.47
		WeekTreatWeek 3, S, Inc	-1.8E-01	-2.2E-01	0.82
		WeekTreatWeek 3, S, SA	-4.1E-01	-4.4E-01	0.66
		WeekTreatWeek 5/6, M, Inc	-1.8E-01	-2.6E-01	0.80
		WeekTreatWeek 5/6, M, SA	1.1E-01	1.6E-01	0.88

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Transposase CDS	AIG67682.1	WeekTreatWeek 5/6, S, Inc	2.2E-01	3.3E-01	0.74
		WeekTreatWeek 5/6, S, SA	-2.8E-02	-4.1E-02	0.97
		WeekTreatWeek 8, M, Inc	-1.8E-01	-2.2E-01	0.82
		WeekTreatWeek 8, M, SA	-1.8E-01	-2.2E-01	0.82
		WeekTreatWeek 8, S, Inc	-1.8E-01	-2.2E-01	0.82
Antigen 43 precursor CDS	AIG67774.1	(Intercept)	-1.6E+00	-1.6E+00	0.11
		WeekTreatWeek 3, M, Inc	-2.0E+01	-1.9E-03	1.00
		WeekTreatWeek 3, M, SA	-2.0E+01	-1.9E-03	1.00
		WeekTreatWeek 3, S, Inc	5.1E-01	4.2E-01	0.68
		WeekTreatWeek 3, S, SA	-1.3E-14	-8.9E-15	1.00
		WeekTreatWeek 5/6, M, Inc	-2.0E+01	-2.7E-03	1.00
		WeekTreatWeek 5/6, M, SA	-2.0E+01	-2.7E-03	1.00
		WeekTreatWeek 5/6, S, Inc	-2.0E+01	-2.7E-03	1.00
		WeekTreatWeek 5/6, S, SA	-1.8E-01	-1.5E-01	0.88
		WeekTreatWeek 8, M, Inc	-2.0E+01	-1.9E-03	1.00
		WeekTreatWeek 8, M, SA	-2.0E+01	-1.9E-03	1.00
		WeekTreatWeek 8, S, Inc	-2.0E+01	-1.9E-03	1.00
Putative vimentin CDS	AIG67776.1	(Intercept)	2.2E-17	4.8E-17	1.00
		WeekTreatWeek 3, M, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, M, SA	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, S, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, S, SA	-1.3E-17	-2.1E-17	1.00
		WeekTreatWeek 5/6, M, Inc	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, M, SA	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, S, Inc	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, S, SA	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 8, M, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 8, M, SA	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 8, S, Inc	-1.3E-17	-2.2E-17	1.00
Hypothetical protein CDS	AIG68237.1	(Intercept)	-2.3E+01	-7.5E-04	1.00
		WeekTreatWeek 3, M, Inc	-2.5E-09	-6.0E-14	1.00
		WeekTreatWeek 3, M, SA	-2.5E-09	-6.0E-14	1.00
		WeekTreatWeek 3, S, Inc	2.2E+01	6.9E-04	1.00
		WeekTreatWeek 3, S, SA	-2.5E-09	-5.8E-14	1.00
		WeekTreatWeek 5/6, M, Inc	-2.5E-09	-6.9E-14	1.00
		WeekTreatWeek 5/6, M, SA	-2.5E-09	-6.8E-14	1.00
		WeekTreatWeek 5/6, S, Inc	-2.5E-09	-6.8E-14	1.00
		WeekTreatWeek 5/6, S, SA	-2.5E-09	-6.9E-14	1.00
		WeekTreatWeek 8, M, Inc	-2.5E-09	-6.0E-14	1.00

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
		WeekTreatWeek 8, M, SA	-2.5E-09	-6.0E-14	1.00
		WeekTreatWeek 8, S, Inc	-2.6E-09	-6.1E-14	1.00
Hypothetical protein CDS	AIG68238.1	(Intercept)	-2.2E+01	-1.2E-03	1.00
		WeekTreatWeek 3, M, Inc	-3.0E-09	-1.2E-13	1.00
		WeekTreatWeek 3, M, SA	-3.0E-09	-1.2E-13	1.00
		WeekTreatWeek 3, S, Inc	2.2E+01	1.1E-03	1.00
		WeekTreatWeek 3, S, SA	-3.0E-09	-1.1E-13	1.00
		WeekTreatWeek 5/6, M, Inc	-3.0E-09	-1.4E-13	1.00
		WeekTreatWeek 5/6, M, SA	-3.0E-09	-1.4E-13	1.00
		WeekTreatWeek 5/6, S, Inc	-3.0E-09	-1.4E-13	1.00
		WeekTreatWeek 5/6, S, SA	-3.0E-09	-1.4E-13	1.00
		WeekTreatWeek 8, M, Inc	-3.0E-09	-1.2E-13	1.00
		WeekTreatWeek 8, M, SA	-3.0E-09	-1.2E-13	1.00
		WeekTreatWeek 8, S, Inc	-3.0E-09	-1.2E-13	1.00
Phage capsid and scaffold protein CDS	AIG68254.1	(Intercept)	-5.1E-01	-8.8E-01	0.38
		WeekTreatWeek 3, M, Inc	3.3E-01	4.5E-01	0.65
		WeekTreatWeek 3, M, SA	5.1E-01	7.2E-01	0.47
		WeekTreatWeek 3, S, Inc	3.3E-01	4.5E-01	0.65
		WeekTreatWeek 3, S, SA	2.9E-01	3.8E-01	0.71
		WeekTreatWeek 5/6, M, Inc	2.2E-01	3.3E-01	0.74
		WeekTreatWeek 5/6, M, SA	3.3E-01	5.0E-01	0.62
		WeekTreatWeek 5/6, S, Inc	3.3E-01	5.0E-01	0.62
		WeekTreatWeek 5/6, S, SA	4.2E-01	6.5E-01	0.52
		WeekTreatWeek 8, M, Inc	3.3E-01	4.5E-01	0.65
		WeekTreatWeek 8, M, SA	1.1E-01	1.4E-01	0.89
		WeekTreatWeek 8, S, Inc	3.3E-01	4.5E-01	0.65
Phage capsid and scaffold protein CDS	AIG68258.1	(Intercept)	-9.2E-01	-1.3E+00	0.20
		WeekTreatWeek 3, M, Inc	-8.8E-01	-7.1E-01	0.47
		WeekTreatWeek 3, M, SA	-1.7E+01	-7.4E-03	0.99
		WeekTreatWeek 3, S, Inc	-8.8E-01	-7.1E-01	0.47
		WeekTreatWeek 3, S, SA	-6.9E-01	-5.7E-01	0.57
		WeekTreatWeek 5/6, M, Inc	-8.8E-01	-8.8E-01	0.38
		WeekTreatWeek 5/6, M, SA	-8.8E-01	-8.8E-01	0.38
		WeekTreatWeek 5/6, S, Inc	2.2E-01	2.7E-01	0.78
		WeekTreatWeek 5/6, S, SA	-1.8E-01	-2.1E-01	0.83
		WeekTreatWeek 8, M, Inc	-1.8E-01	-1.8E-01	0.86
		WeekTreatWeek 8, M, SA	-1.7E+01	-7.4E-03	0.99
		WeekTreatWeek 8, S, Inc	-1.8E-01	-1.8E-01	0.86

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Phage capsid and scaffold protein CDS	AIG68656.1	(Intercept)	-6.2E-18	-1.4E-17	1.00
		WeekTreatWeek 3, M, Inc	-1.8E-01	-2.9E-01	0.77
		WeekTreatWeek 3, M, SA	-4.1E-01	-6.0E-01	0.55
		WeekTreatWeek 3, S, Inc	-1.8E-01	-2.9E-01	0.77
		WeekTreatWeek 3, S, SA	-2.2E-01	-3.3E-01	0.74
		WeekTreatWeek 5/6, M, Inc	-1.8E-01	-3.3E-01	0.74
		WeekTreatWeek 5/6, M, SA	-1.8E-01	-3.3E-01	0.74
		WeekTreatWeek 5/6, S, Inc	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 5/6, S, SA	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 8, M, Inc	-2.7E-17	-4.4E-17	1.00
		WeekTreatWeek 8, M, SA	-1.8E-01	-2.9E-01	0.77
		WeekTreatWeek 8, S, Inc	-1.8E-01	-2.9E-01	0.77
Hypothetical protein CDS	AIG68658.1	(Intercept)	1.4E-15	3.1E-15	1.00
		WeekTreatWeek 3, M, Inc	-1.5E-15	-2.5E-15	1.00
		WeekTreatWeek 3, M, SA	-4.1E-01	-6.0E-01	0.55
		WeekTreatWeek 3, S, Inc	-6.9E-01	-9.5E-01	0.34
		WeekTreatWeek 3, S, SA	-9.2E-01	-1.1E+00	0.27
		WeekTreatWeek 5/6, M, Inc	-6.9E-01	-1.1E+00	0.25
		WeekTreatWeek 5/6, M, SA	-4.1E-01	-7.1E-01	0.48
		WeekTreatWeek 5/6, S, Inc	-1.8E-01	-3.3E-01	0.74
		WeekTreatWeek 5/6, S, SA	-4.1E-01	-7.1E-01	0.48
		WeekTreatWeek 8, M, Inc	2.9E-01	5.0E-01	0.61
		WeekTreatWeek 8, M, SA	-4.1E-01	-6.0E-01	0.55
		WeekTreatWeek 8, S, Inc	-6.9E-01	-9.5E-01	0.34
Phage terminase small subunit CDS	AIG68659.1	(Intercept)	4.7E-01	1.3E+00	0.18
		WeekTreatWeek 3, M, Inc	-4.7E-01	-8.7E-01	0.38
		WeekTreatWeek 3, M, SA	-1.8E-01	-3.6E-01	0.72
		WeekTreatWeek 3, S, Inc	-1.8E-01	-3.6E-01	0.72
		WeekTreatWeek 3, S, SA	-4.7E-01	-8.2E-01	0.41
		WeekTreatWeek 5/6, M, Inc	-4.7E-01	-1.0E+00	0.30
		WeekTreatWeek 5/6, M, SA	-3.9E-01	-8.7E-01	0.39
		WeekTreatWeek 5/6, S, Inc	-3.2E-01	-7.1E-01	0.48
		WeekTreatWeek 5/6, S, SA	-3.2E-01	-7.1E-01	0.48
		WeekTreatWeek 8, M, Inc	-4.7E-01	-8.7E-01	0.38
		WeekTreatWeek 8, M, SA	-3.2E-01	-6.1E-01	0.54
		WeekTreatWeek 8, S, Inc	-1.8E-01	-3.6E-01	0.72
Putative Dnase CDS	AIG68660.1	(Intercept)	-5.1E-01	-8.8E-01	0.38

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Putative Dnase CDS	AIG68660.1	WeekTreatWeek 3, M, Inc	1.1E-01	1.4E-01	0.89
		WeekTreatWeek 3, M, SA	5.1E-01	7.2E-01	0.47
		WeekTreatWeek 3, S, Inc	-1.8E-01	-2.2E-01	0.82
		WeekTreatWeek 3, S, SA	-4.1E-01	-4.4E-01	0.66
		WeekTreatWeek 5/6, M, Inc	2.2E-01	3.3E-01	0.74
		WeekTreatWeek 5/6, M, SA	2.2E-01	3.3E-01	0.74
		WeekTreatWeek 5/6, S, Inc	-2.8E-02	-4.1E-02	0.97
		WeekTreatWeek 5/6, S, SA	-1.8E-01	-2.6E-01	0.80
		WeekTreatWeek 8, M, Inc	5.1E-01	7.2E-01	0.47
		WeekTreatWeek 8, M, SA	5.1E-01	7.2E-01	0.47
		WeekTreatWeek 8, S, Inc	-1.8E-01	-2.2E-01	0.82
Phage tail assembly chaperone CDS	AIG68661.1	(Intercept)	-2.2E+01	-1.2E-03	1.00
		WeekTreatWeek 3, M, Inc	3.2E-10	1.3E-14	1.00
		WeekTreatWeek 3, M, SA	3.2E-10	1.3E-14	1.00
		WeekTreatWeek 3, S, Inc	3.2E-10	1.3E-14	1.00
		WeekTreatWeek 3, S, SA	3.2E-10	1.2E-14	1.00
		WeekTreatWeek 5/6, M, Inc	2.0E+01	1.0E-03	1.00
		WeekTreatWeek 5/6, M, SA	3.2E-10	1.4E-14	1.00
		WeekTreatWeek 5/6, S, Inc	3.2E-10	1.4E-14	1.00
		WeekTreatWeek 5/6, S, SA	3.2E-10	1.4E-14	1.00
		WeekTreatWeek 8, M, Inc	3.2E-10	1.3E-14	1.00
		WeekTreatWeek 8, M, SA	3.2E-10	1.3E-14	1.00
		WeekTreatWeek 8, S, Inc	2.1E+01	1.1E-03	1.00
Phage protein CDS	AIG68664.1	(Intercept)	-1.2E-16	-2.8E-16	1.00
		WeekTreatWeek 3, M, Inc	3.5E-16	5.8E-16	1.00
		WeekTreatWeek 3, M, SA	9.8E-17	1.6E-16	1.00
		WeekTreatWeek 3, S, Inc	2.6E-16	4.3E-16	1.00
		WeekTreatWeek 3, S, SA	1.4E-16	2.2E-16	1.00
		WeekTreatWeek 5/6, M, Inc	2.7E-16	5.1E-16	1.00
		WeekTreatWeek 5/6, M, SA	2.0E-16	3.7E-16	1.00
		WeekTreatWeek 5/6, S, Inc	-1.8E-01	-3.3E-01	0.74
		WeekTreatWeek 5/6, S, SA	-8.7E-02	-1.6E-01	0.87
		WeekTreatWeek 8, M, Inc	-6.6E-18	-1.1E-17	1.00
		WeekTreatWeek 8, M, SA	-1.8E-01	-2.9E-01	0.77
		WeekTreatWeek 8, S, Inc	1.4E-17	2.3E-17	1.00
Phage capsid and scaffold protein CDS	AIG68665.1	(Intercept)	-2.2E-01	-4.5E-01	0.66
		WeekTreatWeek 3, M, Inc	4.1E-02	6.1E-02	0.95
		WeekTreatWeek 3, M, SA	4.1E-02	6.1E-02	0.95

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Phage capsid and scaffold protein CDS	AIG68665.1	WeekTreatWeek 3, S, Inc	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 3, S, SA	5.2E-16	7.3E-16	1.00
		WeekTreatWeek 5/6, M, Inc	-6.5E-02	-1.1E-01	0.91
		WeekTreatWeek 5/6, M, SA	4.1E-02	6.9E-02	0.94
		WeekTreatWeek 5/6, S, Inc	1.4E-01	2.3E-01	0.82
		WeekTreatWeek 5/6, S, SA	4.1E-02	6.9E-02	0.94
		WeekTreatWeek 8, M, Inc	4.1E-02	6.1E-02	0.95
		WeekTreatWeek 8, M, SA	2.2E-01	3.5E-01	0.73
		WeekTreatWeek 8, S, Inc	2.2E-01	3.5E-01	0.73
Putative transport system permease protein CDS	AIG69104.1	(Intercept)	2.2E-17	4.8E-17	1.00
		WeekTreatWeek 3, M, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, M, SA	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, S, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 3, S, SA	-1.3E-17	-2.1E-17	1.00
		WeekTreatWeek 5/6, M, Inc	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, M, SA	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, S, Inc	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 5/6, S, SA	-1.3E-17	-2.5E-17	1.00
		WeekTreatWeek 8, M, Inc	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 8, M, SA	-1.3E-17	-2.2E-17	1.00
		WeekTreatWeek 8, S, Inc	-1.3E-17	-2.2E-17	1.00
Hypothetical protein CDS	AIG69401.1	(Intercept)	-1.6E+00	-1.6E+00	0.11
		WeekTreatWeek 3, M, Inc	-1.8E+01	-4.6E-03	1.00
		WeekTreatWeek 3, M, SA	-1.8E+01	-4.6E-03	1.00
		WeekTreatWeek 3, S, Inc	5.1E-01	4.2E-01	0.68
		WeekTreatWeek 3, S, SA	-1.8E+01	-4.2E-03	1.00
		WeekTreatWeek 5/6, M, Inc	5.1E-01	4.6E-01	0.65
		WeekTreatWeek 5/6, M, SA	-8.8E-01	-6.2E-01	0.54
		WeekTreatWeek 5/6, S, Inc	-1.8E-01	-1.5E-01	0.88
		WeekTreatWeek 5/6, S, SA	-1.8E-01	-1.5E-01	0.88
		WeekTreatWeek 8, M, Inc	5.1E-01	4.2E-01	0.68
		WeekTreatWeek 8, M, SA	-1.8E-01	-1.3E-01	0.90
		WeekTreatWeek 8, S, Inc	-1.8E-01	-1.3E-01	0.90
Hypothetical protein CDS	AIG69967.1	(Intercept)	-2.0E+01	-2.9E-03	1.00
		WeekTreatWeek 3, M, Inc	-1.0E-10	-1.1E-14	1.00
		WeekTreatWeek 3, M, SA	1.9E+01	2.7E-03	1.00
		WeekTreatWeek 3, S, Inc	-1.0E-10	-1.1E-14	1.00
		WeekTreatWeek 3, S, SA	-1.0E-10	-1.1E-14	1.00

Supplemental Table 2-2
(Continued)

Annotation	Protein ID	Week-Treatment	Estimate	z value	Pr(>|z|)
Hypothetical protein CDS	AIG69967.1	WeekTreatWeek 5/6, M, Inc	-1.0E-10	-1.3E-14	1.00
		WeekTreatWeek 5/6, M, SA	1.9E+01	2.7E-03	1.00
		WeekTreatWeek 5/6, S, Inc	1.9E+01	2.7E-03	1.00
		WeekTreatWeek 5/6, S, SA	1.9E+01	2.7E-03	1.00
		WeekTreatWeek 8, M, Inc	1.9E+01	2.7E-03	1.00
		WeekTreatWeek 8, M, SA	-1.0E-10	-1.1E-14	1.00
		WeekTreatWeek 8, S, Inc	1.9E+01	2.7E-03	1.00
GTP-binding protein TypA/BipA CDS	AIG71329.1	(Intercept)	-2.1E+01	-1.9E-03	1.00
		WeekTreatWeek 3, M, Inc	2.0E+01	1.7E-03	1.00
		WeekTreatWeek 3, M, SA	2.0E+01	1.8E-03	1.00
		WeekTreatWeek 3, S, Inc	5.1E-10	3.3E-14	1.00
		WeekTreatWeek 3, S, SA	5.1E-10	3.1E-14	1.00
		WeekTreatWeek 5/6, M, Inc	1.9E+01	1.6E-03	1.00
		WeekTreatWeek 5/6, M, SA	5.1E-10	3.7E-14	1.00
		WeekTreatWeek 5/6, S, Inc	5.1E-10	3.7E-14	1.00
		WeekTreatWeek 5/6, S, SA	1.9E+01	1.6E-03	1.00
		WeekTreatWeek 8, M, Inc	5.1E-10	3.3E-14	1.00
		WeekTreatWeek 8, M, SA	5.1E-10	3.3E-14	1.00
		WeekTreatWeek 8, S, Inc	5.1E-10	3.3E-14	1.00

Supplemental Table 2-3 Metrics from mapping isolate reads to the reference genome in CLC Workbench. The reference genome (main chromosome plus plasmid) is 5,639,239 bp.

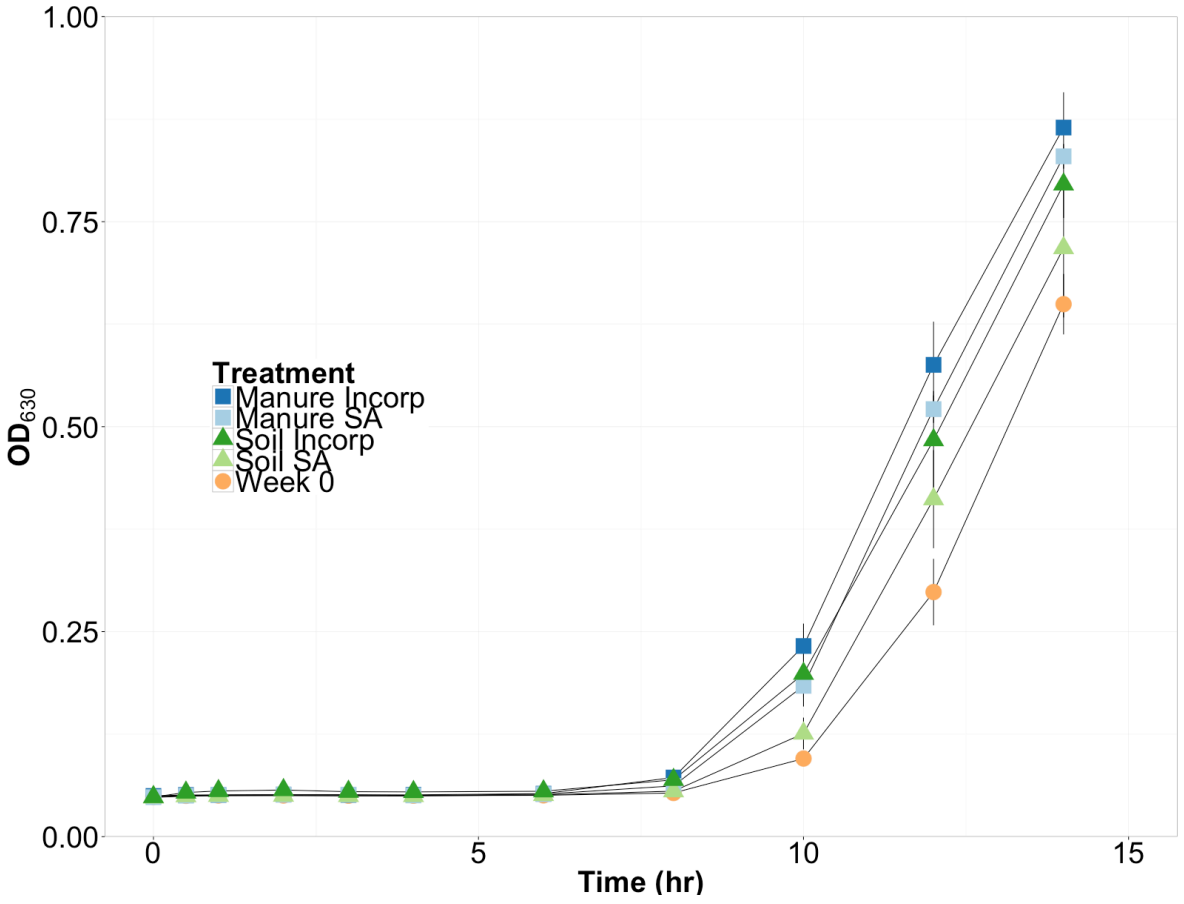
Isolate	Total reads	Mapped reads	Unmapped reads	Percentage unmapped reads	Genome-wide average coverage	Number of bp with zero coverage	Percentage of genome with zero coverage
SWL0001	5115512	5008033	107479	2.1%	110	832	0.015%
SWL0002	3463972	3379949	84023	2.5%	76	987	0.018%
SWL0003	2404902	2335309	69593	3.0%	53	1376	0.024%
SWL0004	2441250	2304023	137227	6.0%	52	1066	0.019%
SWL0005	3251463	3159144	92319	2.9%	70	1127	0.020%
SWL0006	3276316	3209806	66510	2.1%	73	1250	0.022%
SWL0007	4451822	4334152	117670	2.7%	97	926	0.016%
SWL0008	3713149	3621714	91435	2.5%	82	986	0.017%
SWL0009	4142715	4044635	98080	2.4%	89	971	0.017%
SWL0010	2427728	2333294	94434	4.0%	52	1111	0.020%
SWL0012	3536595	3458660	77935	2.3%	77	1075	0.019%
SWL0013	3614221	3497951	116270	3.3%	79	1051	0.019%
SWL0014	2758139	2667328	90811	3.4%	59	1122	0.020%
SWL0015	2204881	2091796	113085	5.4%	48	1286	0.023%
SWL0016	3628416	3549309	79107	2.2%	79	1645	0.029%
SWL0017	3452624	3400978	51646	1.5%	77	1052	0.019%
SWL0018	2919246	2837237	82009	2.9%	64	1075	0.019%
SWL0019	2384764	2284697	100067	4.4%	51	1268	0.022%
SWL0020	2275215	2066457	208758	10.1%	48	1485	0.026%
SWL0021	3548739	3471336	77403	2.2%	78	909	0.016%
SWL0022	2455883	2344974	110909	4.7%	54	1159	0.021%
SWL0023	2767158	2607100	160058	6.1%	57	1206	0.021%
SWL0024	1994310	1765660	228650	12.9%	40	1376	0.024%
SWL0025	2202961	2058912	144049	7.0%	47	1346	0.024%
SWL0026	3112547	2987157	125390	4.2%	68	1011	0.018%
SWL0027	4331064	4245170	85894	2.0%	95	947	0.017%
SWL0028	2860889	2697001	163888	6.1%	61	1265	0.022%
SWL0029	3670006	3612451	57555	1.6%	80	1112	0.020%
SWL0030	1814541	1769883	44658	2.5%	41	1681	0.030%
SWL0031	2730522	2606682	123840	4.8%	59	1276	0.023%
SWL0032	2324018	2268683	55335	2.4%	52	1221	0.022%
SWL0033	2663637	2525443	138194	5.5%	58	1224	0.022%
SWL0034	2533253	2450801	82452	3.4%	56	1338	0.024%
SWL0035	5473772	5317958	155814	2.9%	116	840	0.015%
SWL0036	1266599	1207174	59425	4.9%	28	1460	0.026%
SWL0037	3468920	3397718	71202	2.1%	75	970	0.017%
SWL0038	2894467	2746831	147636	5.4%	62	1163	0.021%
SWL0039	3638577	3551097	87480	2.5%	80	1020	0.018%
SWL0040	3778062	3709085	68977	1.9%	84	1116	0.020%
SWL0041	3562279	3484006	78273	2.2%	79	1020	0.018%

Supplemental Table 2-3 (Continued)

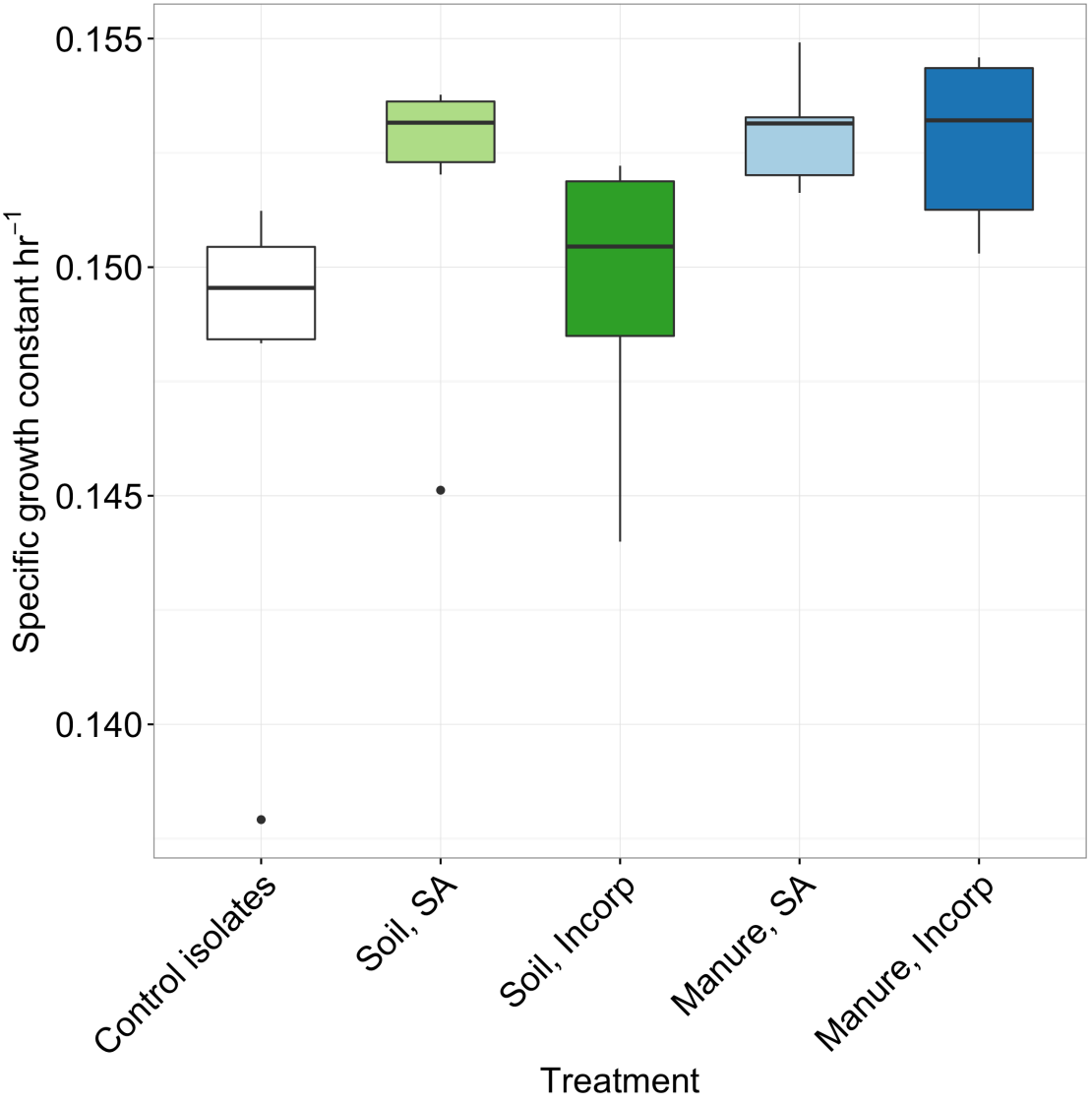
Isolate	Total reads	Mapped reads	Unmapped reads	Percentage unmapped reads	Genome-wide average coverage	Number of bp with zero coverage	Percentage of genome with zero coverage
SWL0042	3459679	3329811	129868	3.9%	75	953	0.017%
SWL0043	3202458	3083829	118629	3.8%	70	1126	0.020%
SWL0044	4013766	3944224	69542	1.8%	88	892	0.016%
SWL0045	4373289	4283928	89361	2.1%	94	912	0.016%
SWL0046	3458706	3397807	60899	1.8%	77	979	0.017%
SWL0047	2946169	2873500	72669	2.5%	63	1260	0.022%
SWL0048	2613328	2552320	61008	2.4%	57	1247	0.022%
SWL0049	1938180	1842619	95561	5.2%	43	1346	0.024%
SWL0050	3510426	3418640	91786	2.7%	77	1062	0.019%
SWL0051	3110401	2955209	155192	5.3%	67	1061	0.019%
SWL0052	2599003	2517452	81551	3.2%	57	1011	0.018%
SWL0053	3033119	2972523	60596	2.0%	66	1209	0.021%
SWL0054	1875493	1748583	126910	7.3%	39	1350	0.024%
SWL0055	2240281	2136059	104222	4.9%	48	1324	0.023%
SWL0056	2620681	2507293	113388	4.5%	57	1080	0.019%
SWL0057	3088347	3008942	79405	2.6%	68	1120	0.020%
SWL0058	3029425	2838507	190918	6.7%	63	1115	0.020%
SWL0059	2148067	2053371	94696	4.6%	47	1330	0.024%
SWL0060	3211254	3063443	147811	4.8%	69	1088	0.019%
SWL0061	3278275	3007683	270592	9.0%	66	1077	0.019%
SWL0062	3710794	3607623	103171	2.9%	80	886	0.016%
SWL0063	4163842	4061398	102444	2.5%	90	940	0.017%
SWL0064	3361621	3275006	86615	2.6%	73	1164	0.021%
SWL0065	2489341	2363361	125980	5.3%	54	1271	0.023%
SWL0066	3947989	3883124	64865	1.7%	87	1045	0.019%
SWL0067	3611196	3496159	115037	3.3%	78	1026	0.018%
SWL0068	3056001	2852790	203211	7.1%	66	1212	0.021%
SWL0069	3232967	3148094	84873	2.7%	71	1056	0.019%
SWL0070	3135195	3016236	118959	3.9%	68	1102	0.020%
SWL0071	1872896	1738028	134868	7.8%	40	1452	0.026%
SWL0072	2746907	2671809	75098	2.8%	61	1263	0.022%
SWL0073	3039790	2779227	260563	9.4%	64	1066	0.019%
SWL0074	2205149	2068738	136411	6.6%	47	1320	0.023%
SWL0075	3297976	3202865	95111	3.0%	72	1134	0.020%
SWL0076	5278034	5171966	106068	2.1%	112	819	0.015%
SWL0077	3887568	3800661	86907	2.3%	86	1178	0.021%
SWL0078	1940277	1825813	114464	6.3%	41	1384	0.025%
SWL0079	2938780	2845342	93438	3.3%	63	1078	0.019%

Supplemental Table 2-3 (Continued)

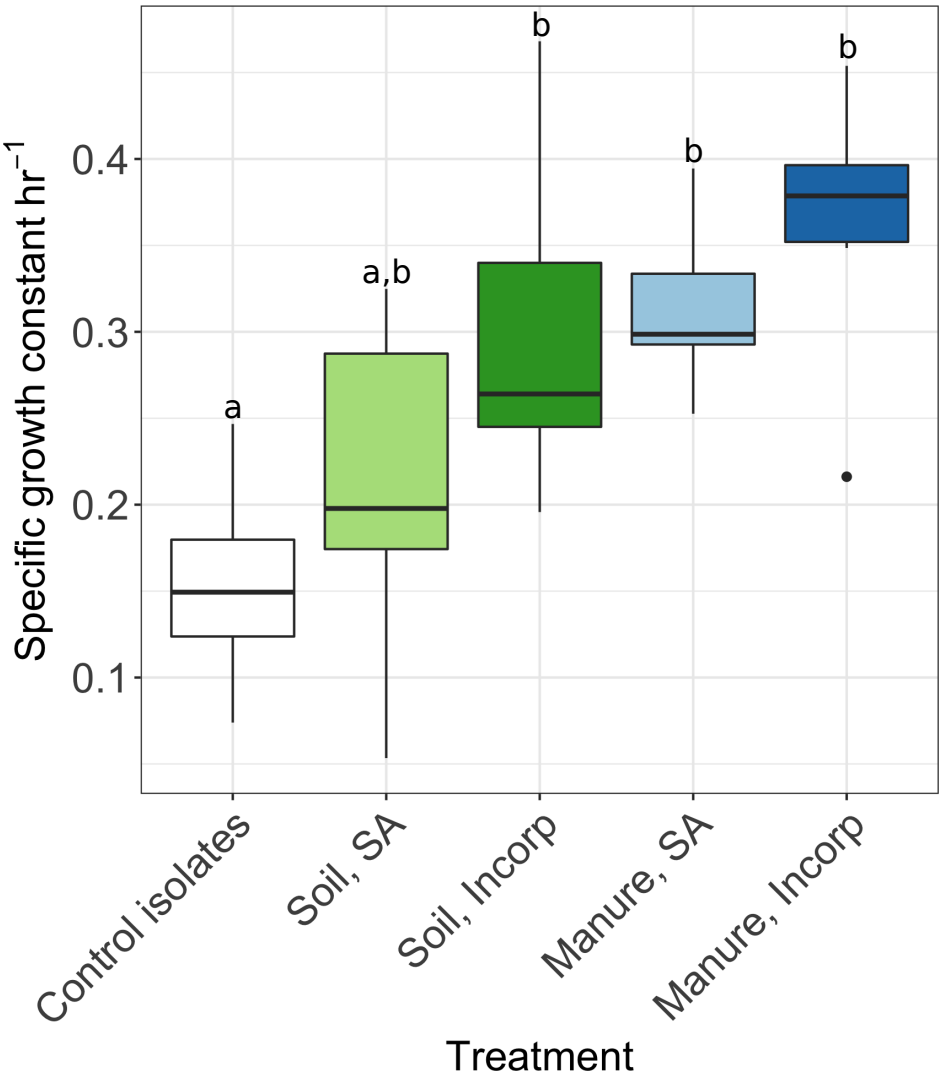
Isolate	Total reads	Mapped reads	Unmapped reads	Percentage unmapped reads	Genome-wide average coverage	Number of bp with zero coverage	Percentage of genome with zero coverage
SWL0080	3386280	3321620	64660	1.9%	75	1309	0.023%
SWL0081	2437111	2338500	98611	4.2%	53	1310	0.023%
SWL0082	3822114	3760757	61357	1.6%	83	1048	0.019%
SWL0083	3753978	3613259	140719	3.9%	81	1052	0.019%
SWL0084	3402067	3336196	65871	2.0%	74	992	0.018%
SWL0085	3173600	3104967	68633	2.2%	71	1123	0.020%
SWL0086	3110182	2833383	276799	9.8%	65	1308	0.023%
SWL0087	3710315	3510998	199317	5.7%	79	1030	0.018%
SWL0088	4427458	4076768	350690	8.6%	91	1015	0.018%
SWL0089	2711828	2621168	90660	3.5%	60	1048	0.019%
SWL0090	2998881	2919782	79099	2.7%	67	1309	0.023%
SWL0091	3026168	2937974	88194	3.0%	65	1138	0.020%
SWL0092	3204218	3113797	90421	2.9%	71	1094	0.019%
SWL0093	2709041	2661366	47675	1.8%	60	1248	0.022%
SWL0094	3278896	3167775	111121	3.5%	72	1088	0.019%
SWL0095	4369886	4254489	115397	2.7%	94	895	0.016%



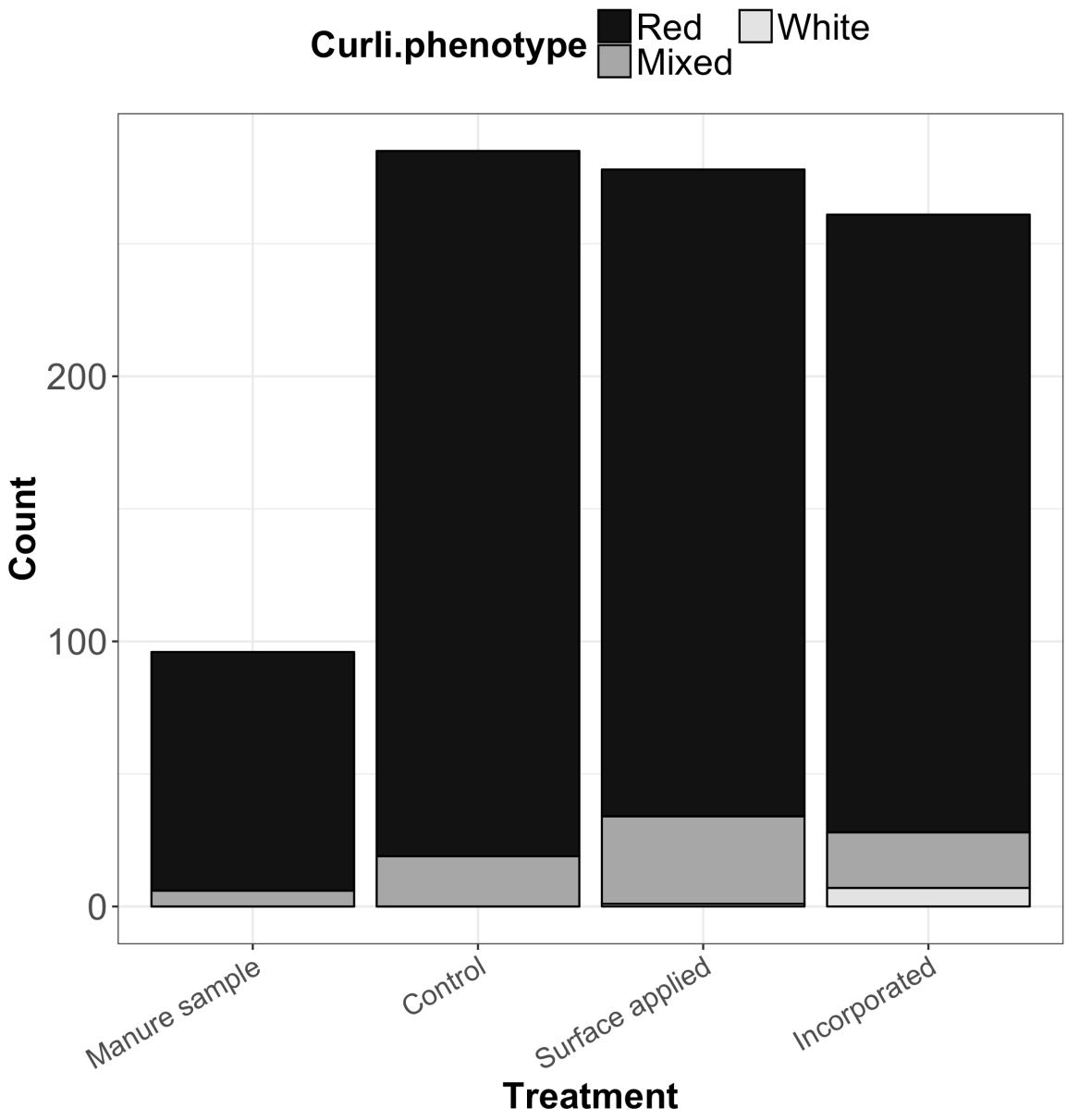
Supplemental Figure 2-1 Growth curves of *E. coli* O157:H7 EDL933 isolates in LB at 37°C, collected from each treatment type at week 6 and initial population isolates (Week 0). Data represents an average (+/- 1SD) of six replicates.



Supplemental Figure 2-2 Specific growth constants for *E. coli* O157:H7 EDL933 isolates collected from each treatment type at week 6, and control isolates (Week 0), and grown over a 14-hr period in LB at 37°C. Data represents six replicates for each treatment.



Supplemental Figure 2-3 Specific growth constants for *E. coli* O157:H7 EDL933 isolates leaving the lag phase when grown in LB at 37°C (6 to 10-hr). Each boxplot includes data from six isolates collected from each treatment type at week 6, and control isolates (Week 0). Different letters represent significant differences ($p < 0.05$) in the specific growth constant between treatments.



Supplemental Figure 2-4 Previously unpublished data collected by Truhlar et al. (2015) that shows the ratio of isolates with a mixed red and white morphotype to isolates with a red morphotype was highest in the population sampled from the top 2 cm of soil in the surface applied treatment column. A red colony morphotype on Congo Red agar indicates curli production, while a white morphotype indicates no curli production. For all treatments, the red morphotype was the predominant morphotype in the population. Control treatments had no added manure, incorporated treatments mixed the manure into the top 5 cm of soil, and surface-applied treatments left the manure on the soil surface. The manure sample was taken from the manure used in all treatments.

WORKS CITED

Aya Castañeda, M. del R., Sarnacki, S.H., Noto Llana, M., López Guerra, A.G., Giacomodonato, M.N., and Cerquetti, M.C. (2015). Dam methylation is required for efficient biofilm production in *Salmonella enterica* serovar Enteritidis. Int. J. Food Microbiol. *193*, 15–22.

Den Bakker, H.C., Allard, M.W., Bopp, D., Brown, E.W., Fontana, J., Iqbal, Z., Kinney, A., Limberger, R., Musser, K.A., Shudt, M., et al. (2014). Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. Emerg. Infect. Dis. *20*, 1306–1314.

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.

Berry, E.D., and Miller, D.N. (2005). Cattle feedlot soil moisture and manure content. J. Environ. Qual. *34*, 656–663.

Blount, Z.D. (2015). The unexhausted potential of *E. coli*. Elife *4*, e05826.

Brennan, F.P., Moynihan, E., Griffiths, B.S., Hillier, S., Owen, J., Pendlowski, H., and Avery, L.M. (2014). Clay mineral type effect on bacterial enteropathogen survival in soil. Sci. Total Environ. *468-469*, 302–305.

Buchanan, B.P., Falbo, K., Schneider, R.L., Easton, Z.M., and Walter, M.T. (2013). Hydrological impact of roadside ditches in an agricultural watershed in Central New York: implications for non-point source pollutant transport. Hydrol. Process. *27*, 2422–2437.

Carter, M.Q., Brandl, M.T., Louie, J.W., Kyle, J.L., Carychao, D.K., Cooley, M.B., Parker, C.T., Bates, A.H., and Mandrell, R.E. (2011). Distinct acid resistance and survival fitness displayed by curli variants of enterohemorrhagic *Escherichia coli* O157:H7. Appl. Environ. Microbiol. *77*, 3685–3695.

Charbonneau, M.-E., and Mourez, M. (2007). Functional organization of the autotransporter 69dhesion involved in diffuse adherence. J. Bacteriol. *189*, 9020–9029.

Cools, D., Merckx, R., Vlassak, K., and Verhaegen, J. (2001). Survival of *E. coli* and *Enterococcus* spp. Derived from pig slurry in soils of different texture. Appl. Soil Ecol. *17*, 53–62.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995). Microbial biofilms. Annu. Rev. Microbiol. *49*, 711–745.

Davison, J. (1999). Genetic exchange between bacteria in the environment. Plasmid *42*, 73–91.

Denes, T., den Bakker, H.C., Tokman, J.I., Guldimann, C., and Wiedmann, M. (2015). Selection and characterization of phage-resistant mutant strains of *Listeria monocytogenes* reveal host genes linked to phage adsorption. Appl. Environ. Microbiol. *81*, 4295–4305.

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.

Evans, M.L., and Chapman, M.R. (2014). Curli biogenesis: Order out of disorder. Biochim. Biophys. Acta BBA – Mol. Cell Res. *1843*, 1551–1558.

Falbo, K., Schneider, R.L., Buckley, D.H., Walter, M.T., Bergholz, P.W., and Buchanan, B.P. (2013). Roadside ditches as conduits of fecal indicator organisms and sediment: Implications for water quality management. J. Environ. Manage. *128*, 1050–1059.

Franz, E., Semenov, A.V., Termorshuizen, A.J., de Vos, O.J., Bokhorst, J.G., and van Bruggen, A.H.C. (2008). Manure-amended soil characteristics affecting the survival of *E. coli* O157:H7 in 36 Dutch soils. Environ. Microbiol. *10*, 313–327.

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.

Hammar, M., Arnqvist, A., Bian, Z., Olsen, A., Normark, S., and others (1995). Expression of two *csg* operons is required for production of fibronectin-and Congo red-binding curli polymers in *Escherichia coli* K-12. Mol. Microbiol. *18*, 661–670.

Van Hoek, A.H.A.M., Aarts, H.J.M., Bouw, E., van Overbeek, W.M., and Franz, E. (2013). The role of *rpoS* in *Escherichia coli* O157 manure-amended soil survival and distribution of allelic variations among bovine, food and clinical isolates. FEMS Microbiol. Lett. *338*, 18–23.

Hutchison, M.L., Walters, L.D., Moore, A., Crookes, K.M., and Avery, S.M. (2004). Effect of length of time before incorporation on survival of pathogenic bacteria present in livestock wastes applied to agricultural soil. Appl. Environ. Microbiol. *70*, 5111–5118.

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., and McVean, G. (2012). *De novo* assembly and genotyping of variants using colored de Bruijn graphs. Nat. Genet. *44*, 226–232.

Iqbal, Z., Turner, I., and McVean, G. (2013). High-throughput microbial population genomics using the Cortex variation assembler. Bioinformatics *29*, 275–276.

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.

Jamieson, R.C., Gordon, R.J., Sharples, K.E., Stratton, G.W., and Madani, A. (2002). Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review. Can. Biosyst. Eng. *44*, 1–9.

Klemm, P., Vejborg, R.M., and Sherlock, O. (2006). Self-associating autotransporters, SAATs: Functional and structural similarities. Int. J. Med. Microbiol. *296*, 187–195.

Kroetsch, D., and Wang, C. (2008). Particle size distribution, in: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis. Canadian Society of Soil Science.

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. Appl. Environ. Microbiol. *64*, 3166–3174.

Latif, H., Li, H.J., Charusanti, P., Palsson, B.O., and Aziz, R.K. (2014). A gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. Genome Announc. *2*, e00821–14 – e00821–14.

Ma, J., Ibekwe, A.M., Yi, X., Wang, H., Yamazaki, A., Crowley, D.E., and Yang, C.-H. (2011). Persistence of *Escherichia coli* O157:H7 and its mutants in soils. PloS ONE *6*, e23191.

McDonald, J.H. (2014). Multiple comparisons. In Handbook of Biological Statistics, (Baltimore, Maryland: Sparky House Publishing), pp. 254–260.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics *9*, 386.

Miller, D.N., and Berry, E.D. (2005). Cattle feedlot soil moisture and manure content. J. Environ. Qual. *34*, 644–655.

Mubiru, D.N., Coyne, M.S., and Grove, J.H. (2000). Mortality of *Escherichia coli* O157:H7 in two soils with different physical and chemical properties. J. Environ. Qual. *29*, 1821–1825.

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990). Physiology of the bacterial cell: a molecular approach (Sunderland, Mass.: Sinauer).

Ramette, A., and Tiedje, J.M. (2007). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proc. Natl. Acad. Sci. *104*, 2761–2766.

Ravva, S., Cooley, M., Sarreal, C., and Mandrell, R. (2014). Fitness of outbreak and environmental strains of *Escherichia coli* O157:H7 in aerosolizable soil and association of clonal variation in stress gene regulation. Pathogens *3*, 528–548.

Reddy, K.R., Khaleel, R., and Overcash, M.R. (1981). Behavior and transport of microbial pathogens and indicator organisms in soils treated with organic wastes. J. Environ. Qual. *10*, 255–266.

Scaletsky, I.C., Silva, M.L.M., and Trabulsi, L.R. (1984). Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. Infect. Immun. *45*, 534–536.

Sjogren, R.E. (1994). Prolonged survival of an environmental *Escherchia coli* in laboratory soil microcosms. Water. Air. Soil Pollut. *75*, 389–403.

Truhlar, A.M., Salvucci, A.E., Walter, M.T., Warnick, L.D., Hay, A.G., and Steenhuis, T.S. (2015). Effects of manure-application practices on curli production by *Escherichia coli* transported through soil. Environ. Sci. Technol. *49*, 2099–2104.

Van den Hoecke, S., Verhelst, J., Vuylsteke, M., and Saelens, X. (2015). Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. BMC Genomics *16*, 79.

Vidovic, S., Block, H.C., and Korber, D.R. (2007). Effect of soil composition, temperature, indeigenous microflora, and environmental conditions on the survival of *Escherichia coli* O157:H7. Can. J. Microbiol. *53*, 822–829.

World Health Organization (2011). Guidelines for drinking-water quality (Geneva: World Health Organization).

Yergeau, E., Labour, K., Hamel, C., Vujanovic, V., Nakano-Hylander, A., Jeannotte, R., and St-Arnaud, M. (2010). Patterns of *Fusarium* community structure and abundance in relation to spatial, abiotic and biotic factors in soil: *Fusarium* and space, biotic and abiotic factors. FEMS Microbiol. Ecol. *71*, 34–42.

Chapter 3 : STUDENT ENGAGEMENT WITH COURSE CONTENT AND PEERS IN SYNCHRONOUS ONLINE DISCUSSIONS

ABSTRACT

As higher education institutions in United States introduce online courses to their curricula and offer these courses to growing audiences, there is increasing desire to understand how best to engage students with both course content and their peers. In one online introductory-level sustainability course, synchronous chats have been the primary means by which students interact with one another. This study examines the effects of assigning chat roles and facilitating self and group reflection on student-content and student-student interaction outcomes in the synchronous chats. We also considered what occurred within group reflections to inform how they are structured in the future. We found that assigning roles increased the proportion of critical student-content interactions and significantly increased critical student-student interactions; $t(9.42) = 0.94$, $p = 0.19$ and $t(12.85) = 3.10$, $p = 0.004$, respectively. Self-reflections had no effect on either interaction type. Groups completing group reflections had a significantly greater proportion of critical student-content interactions in the third chat and critical student-student interactions in the fourth chat than the groups that did not complete the group reflections; $t(7.99) = 2.07$, $p = 0.04$ and $t(6.73) = 1.94$, $p = 0.05$, respectively. This suggests group reflections help students to maintain critical thinking throughout the course of the semester. Based on these results combined with the qualitative analysis of the group reflections, we plan to keep roles going forward and eliminate the self-reflections. Furthermore, to increase the effectiveness of the group reflections, we recommend changes to increase student agency and ability to convert their ideas into change during the next chat.

INTRODUCTION

The introduction of the World Wide Web in 1991 allowed for the growth of online education in the United States (Sun and Chen, 2016). Twenty years after this milestone, 89% and 60% of public and private four-year colleges in the US, respectively, offered either fully online or blended online courses (Parker et al., 2011). More recently, "massive open online courses" (MOOCs) were introduced to online education and heralded as a way to make higher education accessible to anyone with a computer. The percentage of higher education institutions offering MOOCs has steadily increased from 2.6% in 2012 to 11.3% in 2015 (Allen and Seaman, 2016). This paper explores strategies for promoting student engagement and learning in synchronous online discussions, particularly in the case where the class is too large for the primary instructor to facilitate every discussion group. At a time when the higher education landscape is increasingly shifting towards online learning, and in many cases, online learning for the "masses," this investigation is a timely addition to the conversation surrounding student engagement and achievement of learning outcomes in online education.

Interaction in successful online learning

The US Distance Learning Association indicates that interaction is an integral component of quality distance education, including online education (Holden and Westfall, 2007). In any educational context, interaction falls into three categories: student-instructor, student-student, or student-content (Moore, 1989). A meta-analysis of 74 studies, found that implementing at least one type of interaction intervention in distance-education courses increased student understanding of course content (Bernard et al., 2009). Among the three interaction types, student-student and student-content interactions had significantly greater effect sizes on student learning.

In online courses, the majority of student-student interactions take place in asynchronous (e.g. discussion forums students can access on their own time) or synchronous (e.g. live chat rooms) discussions carried out over a technological platform. Aside from individual assignments, discussions are also where many student-content interactions occur. Therefore, there is great interest in understanding how to best encourage and support interactions in these discussion formats. Over the past two decades many interventions have been studied for their ability to improve outcomes of student interactions in online discussions. Examples include assigning roles to students (Wise et al., 2012) and asking students to reflect on their contributions to the discussion (Kayler and Weller, 2007).

Assigning discussion roles to students has been suggested to have many benefits including increasing student participation (Tagg, 1994), interaction (Hara et al., 2000), and knowledge integration and construction (Schellens, 2005; Strijbos et al., 2007). However, the research on whether these outcomes are actually achieved is inconclusive. Wise and colleagues (2012) suggest this is because research has focused on roles, but not on the actual functions the roles perform in conversation. Therefore, depending on how the instructor or researcher specifies the role to students, roles with the same name could be asked to fulfill different functions, or vice-versa. Wise and colleagues (2012) identified six functions common to the roles assigned in 12 studies. Of these six, three were suggested to be most helpful in discussions, namely: (1) give direction, (2) summarize, and (3) provide critical response. The "give direction" function is typically assigned to the student expected to start the conversation and can be expanded to include keeping the conversation on topic for its duration (Hara et al., 2000; Persell, 2004; Tagg, 1994; Wise et al., 2012; Zhu, 1998). The "summarize" function asks students to synthesize what has been stated in the discussion up to a certain point, usually set at the end of the discussion

period. Posts made by students assigned this function are consistently associated with higher levels of thinking due to the synthesis required by the function description itself (De Wever et al., 2007; Schellens, 2005). Summary posts have also been shown to stimulate higher-order responses from other students (Wise and Chiu, 2011). The "critical response" function is defined as "the right to be critical of ideas in the discussion" (Wise et al., 2012). A lack of critical comments on other students' thinking is common to online discussions (De Wever et al., 2008; Schellens, 2005; Schellens, et al., 2007). Assigning the critical response function gives students permission to challenge their peers, which they might otherwise avoid (Wise et al., 2012).

Both individual and group reflections are commonly used pedagogical tools to promote better group work (Gibbs, 1994). However, there are few instances of these techniques being applied to online discussions, even though this specific type of group work has the unique quality of producing a written transcript. Therefore, instructors have the ability to hold students accountable to what actually transpired in the discussion. Multiple researchers have suggested that the reflective reading of transcripts as an intervention to improve students' online discussion strategies is an area that needs more investigation (De Wever et al., 2008; Murphy and Jerome, 2005; Walker, 2004). Qualitative work has found that structured, written self-reflection on contributions to an online discussion can allow students to identify their strengths and areas for improvement (Kayler and Weller, 2007; Murphy and Jerome, 2005). Furthermore, Kayler and Weller (2007) found that students who favored written reflection in their learning style were particularly empowered by this activity, and used it to determine how they could continue to improve their own contributions even when they were not happy with the group as a whole. Students who did not favor written reflection as a learning tool tended to blame bad group discussion experiences on perceived deficiencies in the participation of their group-mates.

Therefore, one of the outcomes of this study was the modification of the self-reflection activity to accommodate diverse learning styles with the inclusion of an in-person group-reflection. This work seeks to build on the work by Kayler and Weller (2007), and determine the improvements in chat quality, if any, gained from the addition of group reflections.

Critical interactions

As noted by Ertmer and colleagues (2011), measuring interactions in online discussions comes with challenges. Specifically, not every post is meaningful and relevant to course content (Ertmer et al., 2011). Likewise, many student-to-student posts might be social interactions and not active discussions of course content. Therefore, purely quantitative measures of interaction (e.g. post counts) have not been as useful as those mixed with qualitative measures (e.g. post quality), often defined in terms of critical thinking (Ertmer et al., 2011; Walker, 2004; Wise and Chiu, 2011). For the remainder of this manuscript, "critical interaction" will be considered, as opposed to all student-content or student-student interaction.

Bloom's original taxonomy (Bloom et al., 1956) is often applied to measuring critical thinking about course content. The higher levels of Bloom's original taxonomy (i.e., analysis, synthesis, and evaluation) correspond to the cognitive processes associated with critical thinking (Pear et al., 2001). For student-student interactions, critical thinking occurs when participants hold ideas "publicly accountable" and work to reason out their understanding of those ideas (Mercer et al., 1999). This is opposed to a discussion where participants reflexively agree or disagree, and thereby only superficially engage with the ideas put forward. Indicators of critical thinking in a conversation are questions that elicit further explanation of a provided fact or opinion, questions that require a defense of a line of argument, and statements containing a counter or alternative

argument (Walker, 2004). These question types are labeled by Walker (2004) as "probe," "challenge," and "counter" questions, respectively.

CONTEXT

Our research is situated in the framework of "Teaching-as-Research," a formalized process by which an instructor applies research methods to evaluate their own teaching practice in order to improve learning outcomes for their students (Kwako et al., 2005). The steps of this process are outlined in Figure 1.

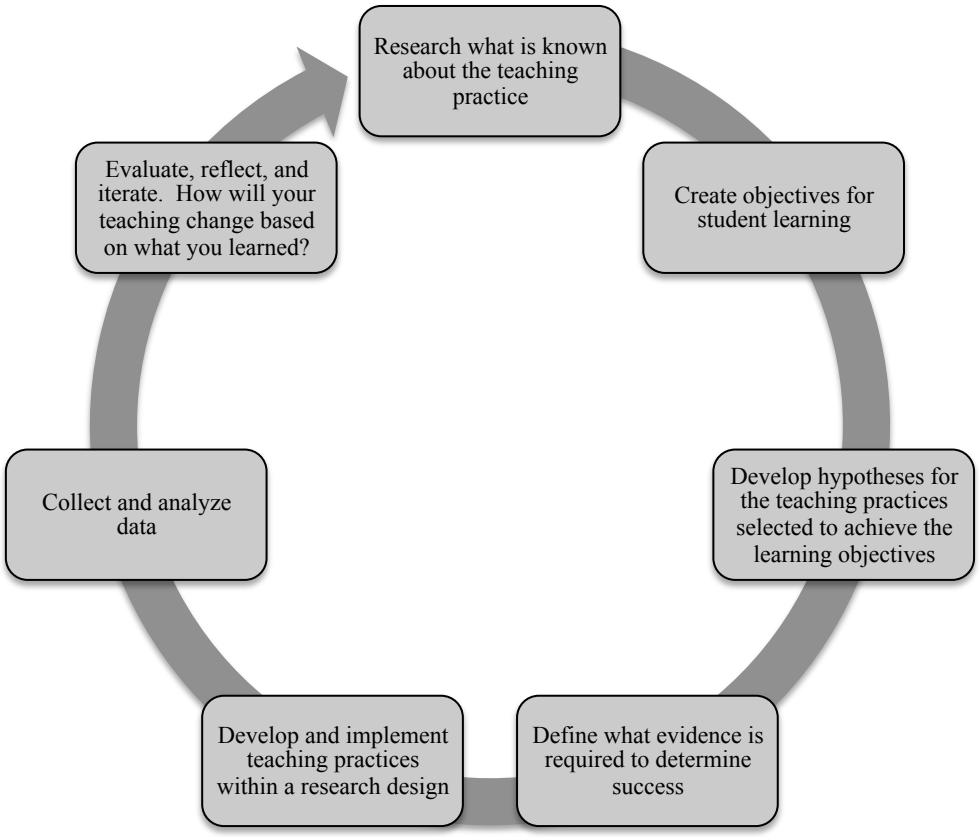


Figure 3-1 Summary of the Teaching-as-Research process (Kwako et al., 2005; Williams, 2015).

Our study pertains to an introductory-level sustainable development course that has been offered in an entirely online format since its inception in 1998. During this time period, synchronous chats, generally held four times over the semester, have been the primary means by which students engage with one another and develop a learning community. In spring 2015, enrollment in the course exceeded three hundred students, requiring the use of 40 undergraduate teaching assistants (TAs) to facilitate chats (one group per TA) and complete other tasks such as grading. In fall 2016, enrollment was capped at 200 students and 16 undergraduate TAs were recruited. In both semesters, the first author served as the head graduate TA and primary coordinator for the undergraduate TAs.

In spring 2015, the first author closely read and qualitatively analyzed a selection of chat transcripts to understand how students were engaging in the chats. Three main themes emerged from this analysis, two of which are directly relevant to this study: (1) students were not critically engaging with the course information, as evidenced by a majority of postings being made at low levels of thinking on Bloom's taxonomy, and (2) students were not critically engaging with information provided by their peers, as evidenced by a lack of challenges to new information. The first author had an intuitive sense from reading through chat transcripts throughout the semester that students were not fully engaged. For example, as soon as new questions were posted in a chat, it was common for a rapid series of long responses to follow. The questions were provided ahead of time to allow students to prepare, and the first author suspected that students were simply copying and pasting pre-written responses to earn participation points for being present during the chat. This was anecdotally confirmed through conversations with past students.

As a result of these findings, for the last chat in the spring 2015 semester we changed the second half (about 30 minutes) of the chat format to center around a debate, as opposed to the previous series of open-ended questions. The idea behind this change was that students would be forced to respond spontaneously to each other, as they could not completely anticipate the arguments of the opposing team. This format change still did not resolve the problem of low student engagement and led us to search the literature for other intervention options. As the literature reviewed here indicates, the majority of online interaction research relates to asynchronous, as opposed to synchronous, discussions. However, we decided to implement one intervention that has been routinely shown to increase both student-content and student-student interactions in asynchronous discussions: assigning discussion roles. In addition, we decided to test the effect of self- and group-reflections on student-content and student-student interaction, as this is an area that has been identified in the literature for further research.

PURPOSE

This study examines the effects of assigning chat roles and facilitating self and group reflection on student-content and student-student interaction outcomes for an online synchronous chat. The main hypotheses addressed in this study are: (1) assigning chat roles will increase the proportion of chat posts that contain critical student-content and student-student interactions, (2) self-reflections will further increase the proportion of chat posts containing critical student-content and student-student interactions compared to roles alone, (3) when roles are assigned, groups completing group-reflections will have a greater proportion of chat posts containing critical student-content and student-student interactions compared to groups completing only self-reflections, and (4) when no roles are assigned, groups completing group-reflections will have a greater proportion of critical student-content and student-student interactions than groups without

group-reflections. In addition to these four hypotheses and in an effort to inform the development of future group-reflection assignments, we ask the question: what happens during an online, asynchronous, group-reflection process?

METHODS AND MATERIALS

Interventions

The changes made to the course for the fall 2016 semester were guided by the emergent themes from the earlier qualitative analysis and the existing literature. A timeline of the interventions is shown in Figure 2. Students were assigned to a discussion group from the start of the course and conducted all their chats within this group. All chats were based around a one-hour version of the informal debate structure introduced in spring 2015. As the first intervention, before the first chat each student was assigned one of three specific roles to focus on during the debate (see Appendix A for full language of the role assignments). In brief, the three roles were the Traffic-Director, which was responsible for keeping the discussion on topic, the Questioner, which was responsible for asking critical questions of their peers, and the Synthesizer which was responsible for identifying themes amongst posts and highlighting topics that still need to be addressed. Students were assigned roles for Chats 1, 2, and 3, and changed roles each time.

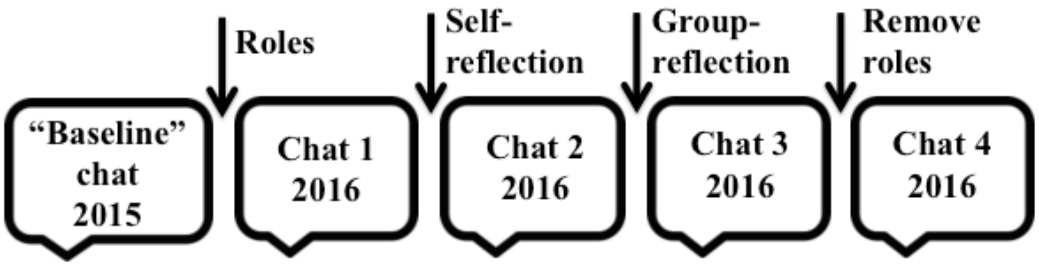


Figure 3-2 Timeline of interventions used in the online chats. Once an intervention was introduced, it was continued for the rest of the semester, unless otherwise noted.

Following the first chat, students were asked to complete a self-reflection on how they fulfilled their assigned role, drawing on examples of both things they did well and things they could improve on from the chat transcript. To ensure their familiarity with all three roles, they were also asked to reflect on examples from the two roles they were not assigned (see Appendix B for full assignment).

The second chat followed the same procedure as the first, but after the chat eight of the sixteen groups were assigned to complete both a written self-reflection (see Appendix C for full assignment) and an asynchronous group reflection on what went well in the chat and what could be improved (see Appendix D for full assignment). The other eight groups were assigned to complete a written self-reflection only, and to complete an asynchronous group discussion activity that built off the topic discussed in the synchronous chats.

The third chat followed the same procedure as the second chat. The fourth chat followed the same procedure as the second and third chats, except students were not assigned chat roles. Table 1 summarizes the interventions used in each chat and how they relate to our four hypotheses.

Table 3-1 Interventions applied in each of the four chats, with the corresponding hypotheses and *t*-test results. Bolded *p*-values indicate a significant result at the 95% confidence level.

Intervention	Hypothesis	Dependent variable	*t*-value	df	*p*-value
Chat 1 – assign roles	Assigning chat roles increases critical student-content and student-student interactions. Ho: Chat 1, 2016 = Chat 4, 2015 Ha: Chat 1, 2016 > Chat 4, 2015	Proportion of higher level thinking posts	0.94	9.42	0.19
		Proportion of critical questions	3.10	12.85	**0.004**
Chat 2 – self-reflection	Self-reflections further increase critical student-content and student-student interactions. (Paired t-test) Ho: Chat 2, 2016 = Chat 1, 2016 Ha: Chat 2, 2016 > Chat 1, 2016	Proportion of higher level thinking posts	0.32	9	0.38
		Proportion of critical questions	-1.14	9	0.85
Chat 3 – group-reflection	Group-reflections further increase critical student-content and student-student interactions, as compared to self-reflections alone. Ho: Chat 3, 2016, group reflections = Chat 3, 2016, no group reflections Ha: Chat 3, 2016, group reflections > Chat 3, 2016, no group reflections	Proportion of higher level thinking posts	2.07	7.99	**0.04**
		Proportion of critical questions	-1.03	7.98	0.83
Chat 4 – remove roles	Groups completing group-reflections have greater proportion of critical student-content and student-student interactions than groups without group-reflections, when no roles are assigned. Ho: Chat 4, 2016, group reflections = Chat 4, 2016, no group reflections Ha: Chat 4, 2016, group reflections > Chat 4, 2016, no group reflections	Proportion of higher level thinking posts	0.65	5.96	0.27*
		Proportion of critical questions	1.94	6.73	**0.05**

*Note: If both the yes group and no group "high" outliers are removed, the *p* value becomes **0.04**; $t(3.53) = 2.40$.

Measuring student-content and student-student interactions

The first forty posts in five chat transcripts from the last chat in 2015 and ten chat transcripts from each 2016 chat were coded for critical student-content interactions, using the higher levels of thinking from Bloom's original taxonomy (e.g., analysis, synthesis, evaluation) as a proxy. The level of thinking for each post was coded based on a modified version of Bloom's original taxonomy (Table 2; Bradley et al., 2008). If a post included multiple statements, it was coded according to the highest level of thought it contained. The transcripts were also coded for critical student-student interactions using the three types of questions (probe, challenge, or counter) suggested by Walker (2004) as a proxy. One-sided *t*-tests were used to test each hypothesis; see Table 1 for a full description of the hypotheses and corresponding *t*-tests used. All statistical analyses were performed with R version 3.2.3 (R Core Team, 2015).

Table 3-2 Modified version of Bloom's original taxonomy, adapted from Bradley et al. (2007).

Level	**Example**
Opinion	"I think…" with no "because'
Knowledge	Cite reading
Understanding	Paraphrase reading
Application	Relate to own experience
Analysis	Compare information
Synthesis	Draw connections
Evaluation	Justify position

Qualitative analysis

In order to understand what was happening in the group-reflections, we closely read transcripts from three of the five groups participating in group-reflections. These groups, arbitrarily named A, B, and C, were selected because they had the same students participating in each chat. The other two of the five groups completing group-reflections had minor changes in student chat attendance over the course of the semester due to scheduling conflicts, although all the students in these groups did complete a group reflection. General themes were identified following the initial reading. We then re-read the transcripts, looking for evidence of each theme in individual posts and labeling them accordingly. The number of times a theme was mentioned was tabulated.

RESULTS

Quantitative results

Hypothesis 1

Compared to a chat with no roles assigned, chats with assigned roles had a greater proportion of critical student-content and student-student interactions; $t(9.42) = 0.94$, $p = 0.19$ and $t(12.85) = 3.10$, $p = 0.004$, respectively (Table 1, Figure 3). However, only the increase in the proportion of critical student-student interactions was significant.

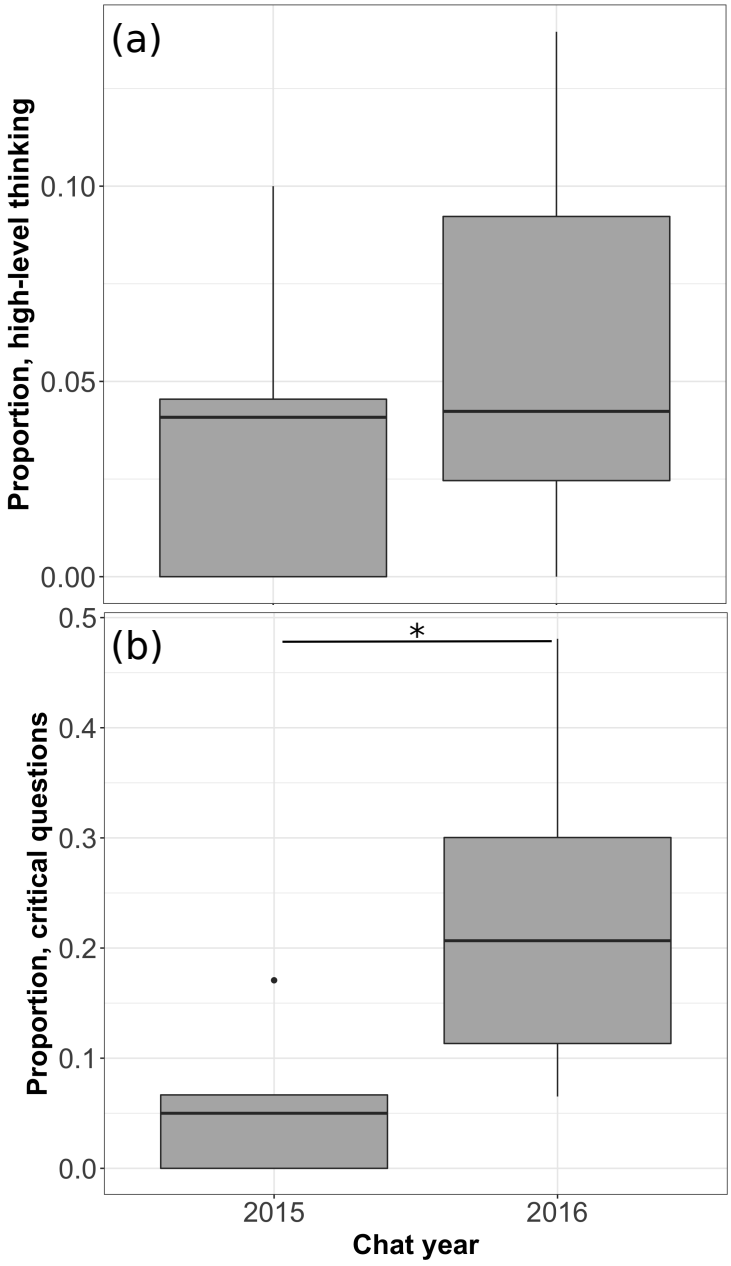


Figure 3-3 Proportion of total posts in the 2015 chat and Chat 1, 2016 that were coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions. Between 2015 and 2016, chat roles were introduced. The upper whisker extends to the largest observed value within the third quartile plus 1.5 times the interquartile range. The lower whisker extends to the smallest observed value within the first quartile minus 1.5 times the interquartile range. An asterisk indicates a significant difference at a 95% confidence level.

Hypothesis 2

The addition of pre-chat self-reflections to the chat format did not increase either the proportion of critical student-content interactions, or the proportion of critical student-student interactions; $t(9) = 0.32$, $p = 0.38$ and $t(9) = -1.14$, $p = 0.85$, respectively (Table 1, Figure 4).

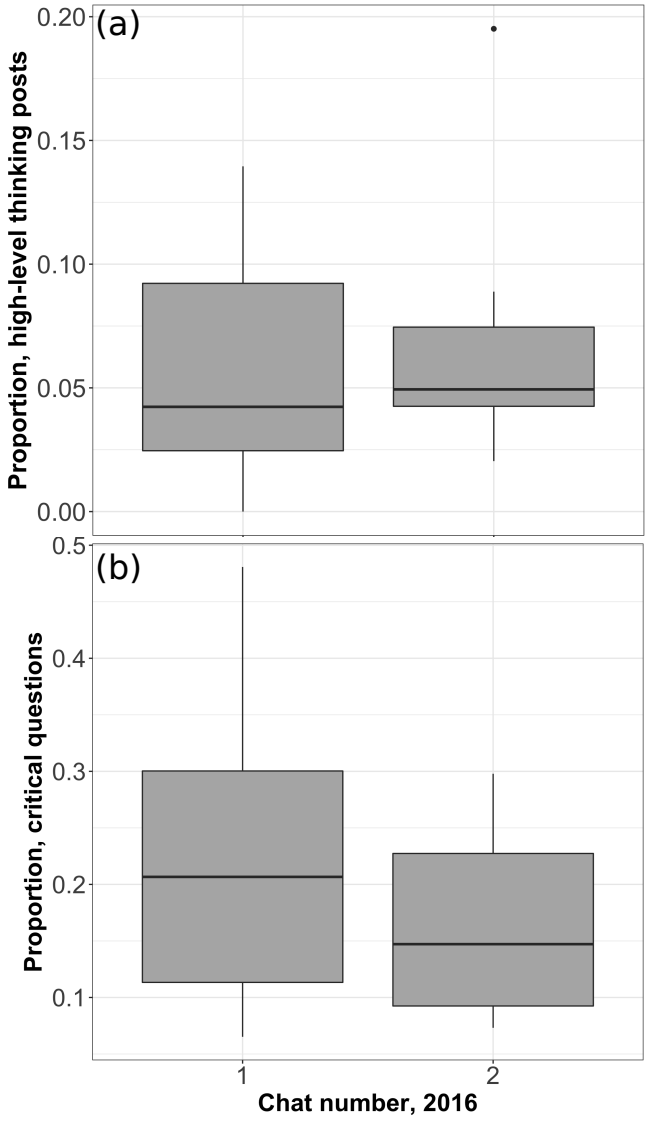


Figure 3-4 Proportion of total posts in Chat 1, 2016 and Chat 2, 2016 that were coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions. Between Chat 1 and Chat 2, self-reflections were introduced. The upper and lower whiskers are drawn as described in Figure 1.

Hypothesis 3

The proportion of critical student-content interactions was significantly greater for groups completing group-reflections, compared to groups that did not complete group-reflections; $t(7.99) = 2.07$, $p = 0.04$ (Table 1, Figure 5). There was no difference in the proportion of critical student-student interactions between groups in the two treatments; $t(7.98) = -1.03$, $p = 0.83$ (Table 1, Figure 5).

Hypothesis 4

When roles were removed for the fourth chat, the groups that completed a group-reflection had a greater proportion of critical student-content interactions than groups that did not complete a group-reflection; $t(5.96) = 0.65$, $p = 0.27$ (Table 1, Figure 5). If "high" outliers from both the group with reflections and the group without reflections are removed, this difference becomes significant; $t(3.53) = 2.40$, $p = 0.04$. High outliers were defined as values greater than the 75^{th} percentile plus 1.5 times the interquartile range, or the range between the 25^{th} and 75^{th} percentile (Tukey, 1977). In general, the groups completing group reflections maintained higher-level thinking in the chats over the course of the semester (Figure 5). Furthermore, groups completing group-reflections had a greater proportion of critical student-student interactions in the fourth chat than groups that did not complete the group-reflection; $t(6.73) = 1.94$, $p = 0.05$ (Table 1).

Qualitative results

Three major themes arose around the group-reflection process. First, among the groups that completed group reflections, each group differed in the amount of perceived agency (as defined and discussed below) they expressed about their chat experience. Second, each group varied in their ability to suggest concrete changes to the chat. Third, as a whole, groups tended to favor

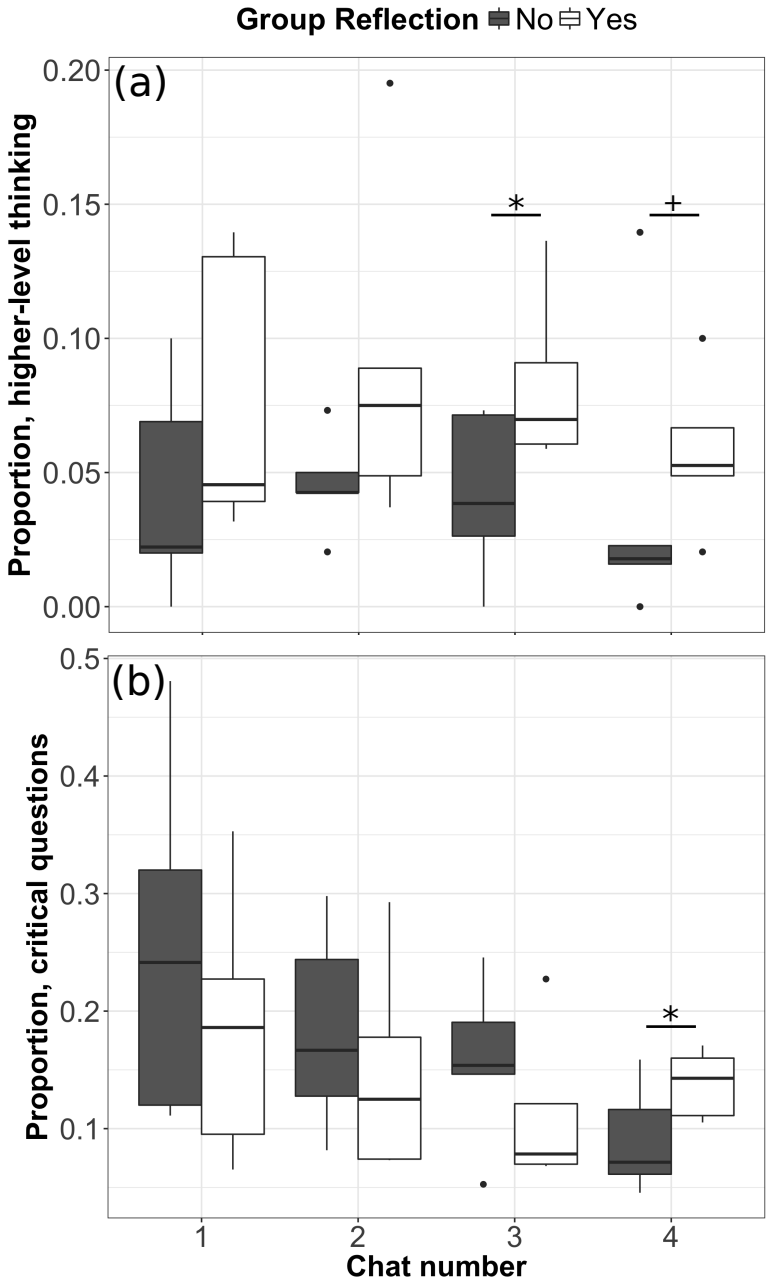


Figure 3-5 The proportion of posts coded as being (a) higher-level thinking, a proxy for critical student-content interactions, and (b) critical questions, a proxy for critical student-student interactions, separated by whether the group went on to complete a group-reflection prior to chats 3 and 4. An asterisk indicates significance a 95% confidence level. A cross indicates significance at the same level after "high" outliers are removed. Lighter shading is used for groups that went on to complete group-reflections before Chats 3 and 4, while darker shading is used for groups that were never asked to complete group-reflections. The upper and lower whiskers are drawn as described in Figure 1.

suggesting logistical changes to the chats over learning-outcome related changes. All the above themes held true both with and without role assignment.

Agency

One group, "Group A," demonstrated more agency in both the first and second group-reflection compared to the other groups (Figure 6). Here, we define agency as "feeling enabled and empowered to act" and solve problems (Kayler and Weller, 2007). When members of Group A discussed aspects of the chat they thought were least effective, they tended to emphasize what they could do, or had already done, to affect change (Example 1). Many also put forth a narrative that they were learning and improving with each chat (Example 1). Contrary to these patterns, members from the other two groups expressed feeling that the ability to change what was happening in the chat was out of their hands, placing responsibility on inadequacies in the chat platform, other students, or the teaching assistant facilitating the discussion (Example 2).

Example 1: "The most effective thing about our group's chat was the organization and smooth manner it was carried out [in] because we had discussed it beforehand. After having two chats, we knew how out of hand it can get. So we went subject by subject by first tackling one point, then another, and then the last one."

Example 2: "I think that the digital format of the chat is responsible for the constraint on timely responses…"

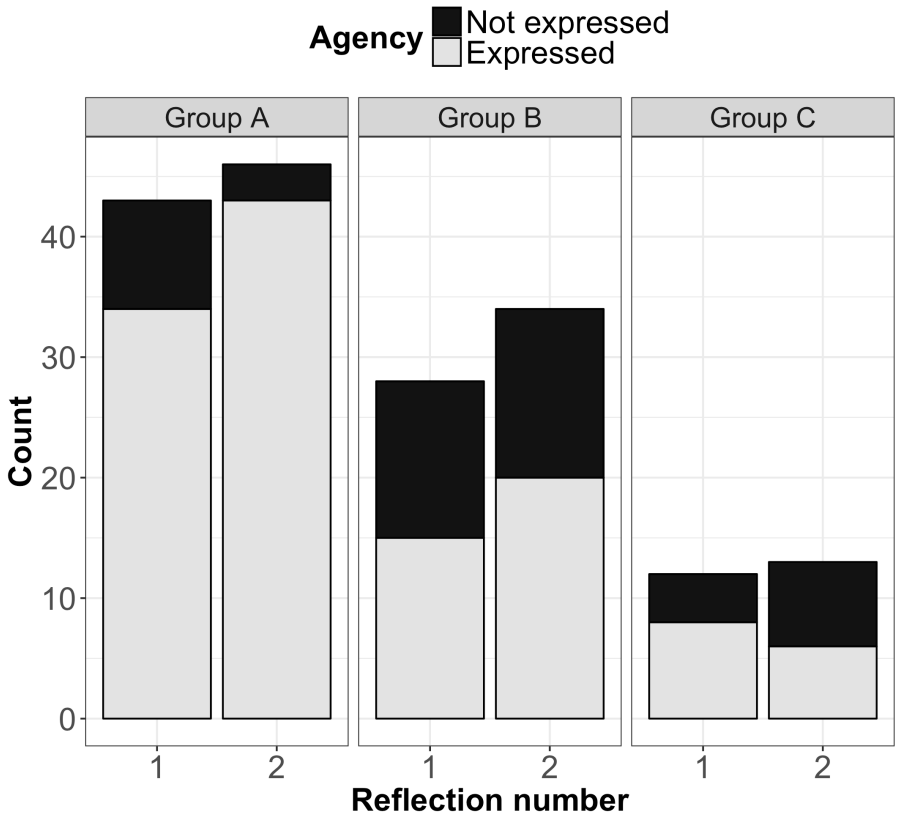


Figure 3-6 The number of individual posts from the two group reflections, before Chats 3 and 4, respectively, in which students either expressed or did not express agency.

Concrete versus vague suggestions for improvement

Members of Group A consistently proposed concrete actions they could take to improve ineffective aspects of the chat (Figure 7; Example 3). In the other two groups, members often did not propose actions to address ineffective aspects of the chat. When they did, these suggestions were vague, as indicated by the frequent use of words like "hopefully" and "maybe" (Example 4).

Example 3: "I agree with you that the fast-paced nature of the debate is probably the main issue. We should definitely utilize the Raise Hand function."

Example 4: "Hopefully next time, all questions will be answered or at least expanded on."

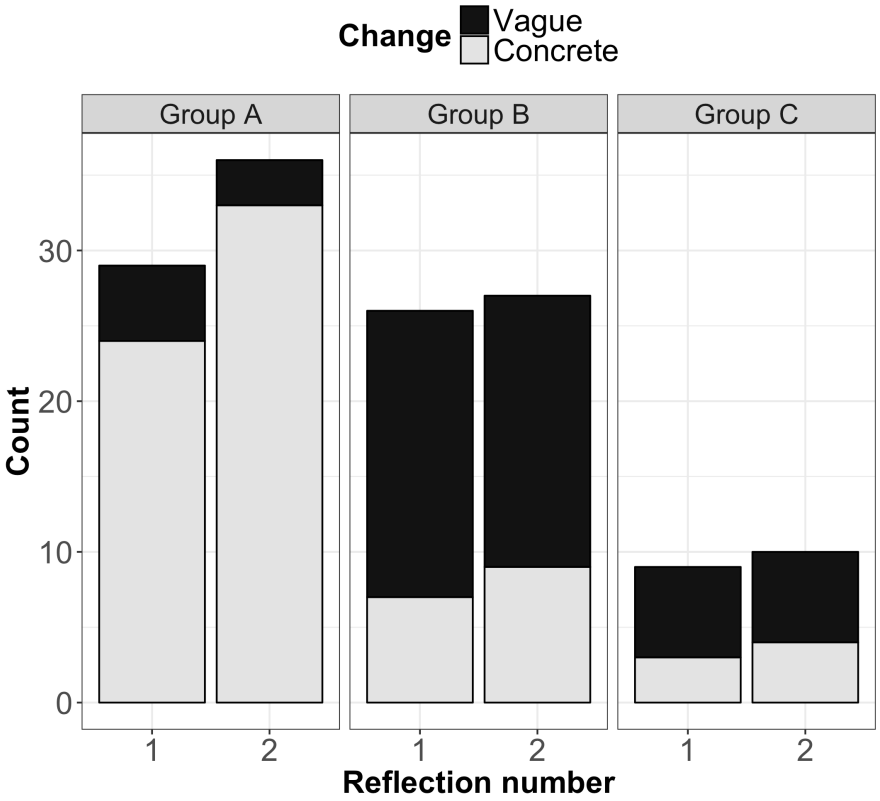


Figure 3-7 The number of individual posts from the two group reflections, before Chats 3 and 4, respectively, in which students suggested a concrete or a vague change to the next chat.

Mixed focus on logistics and learning outcomes

The chat aspect that was identified most frequently by each group as needing improvement varied, but overall tended to favor logistical aspects over actions directly related to achieving the target pedagogical outcomes (Table 3).

Table 3-3 Chat aspect most frequency identified by each group as "ineffective" for the first and second group-reflections.

Group	Reflection Number	General category	Specific description	Number of distinct posts
A	1	Logistics	Multiple conversations occurring simultaneously	46
	2	Logistics	Schedule of the debate, specifically the order and timing of when main points should be presented and discussed	25
B	1	Logistics	Schedule of the debate, specifically, the "for" team received more time to present their arguments	19
	2	Outcomes/ Logistics	Questions were left unanswered, but it is unclear whether concern about this was driven more by lack of information from peers (outcomes) or by multiple conversations happening at once (logistics)	19
C	1	Outcomes	The debate lost focus due to students getting too bogged down in "specifics" – for example, specific definitions	19
	2	Logistics	Multiple conversations occurring simultaneously	9

DISCUSSION

Assigned roles and group-reflections emerged as the most useful interventions for achieving our pedagogical goals of critical student-content and student-student interactions in the synchronous online chats. Specifically, roles significantly increased student-student interaction, while the group-reflections served to maintain higher-level thought throughout the semester, even when roles were removed for the final chat. Group-reflections also resulted in significantly more student-student interactions when roles were removed. The third intervention we tested, self-reflections, did not result in gains in either higher-level thought or student interaction.

Roles

Our results are in line with previous work that demonstrates roles increase student-student interaction in online chats (Hara et al., 2000). Of the posts directed to another student in Chat 1, 2016, and the 2015 chat, an average of 59% and 24%, respectively, were content-based questions. This increase in questioning suggests much of the student interaction in the 2016 chat was driven by the Questioner role. There is qualitative evidence that other roles, specifically the Traffic Director role, increased student-to-student posts as well. In Examples 5.1 and 5.2, Teammates A and C are the Traffic Directors. In these exchanges, the Traffic Director posts did not include information relevant to the chat content but were rather intended to organize group members to respond to questions. No organization by students in preparation to answer questions was seen in the 2015 chat, which lacked roles.

Example 5.1: Teammate A (Traffic Director): Any thoughts on this Teammate B?

Example 5.2: Teammate C (Traffic Director): I can address Student X's question.

> Teammate C (Traffic Director): And then Teammate D idk [I don't know] if you want to address Student Y's?

Roles did not increase the proportion of posts made at higher levels of thinking. Two roles, the Synthesizer and the Questioner, were specifically implemented due to previous work demonstrating that the functions contained within these roles cause an increase in both critical student-content and student-student interactions (De Wever et al., 2007; Schellens, 2005; Wise and Chiu, 2011; Wise et al., 2012). A couple of observations can explain the failure of these

roles to increase higher-level thought in our case. First, we found students in the Synthesizer role emphasized the "summarize" function found in the role description and would provide short restatements of the main points made throughout the debate, rather than connecting points and identifying areas yet to be explored (Table 1). These summaries were explicitly mentioned as being useful in the group-reflections, due to the fast nature of the synchronous chat (Example 6). Second, of the questions asked across all groups in Chat 1, 83 were posed at lower levels of thinking ("know" or "understand" on Bloom's original taxonomy), while 45 were posed at higher levels of thinking. Notably, no questions were coded as "counter" questions during Chat 1.

Example 6: "…it is nice to have the synthesizers to summarize what is going on so that if someone does have a really long response that it can be summarized and easily understood."

Combined, our observations regarding the Synthesizer and Questioner role implementation suggest that students will pick and choose the functions they find most advantageous or expedient within a role definition. In the case of the Synthesizer role, it seems that students found summaries to better serve their immediate needs due to the fast-paced nature of the synchronous chat, where keeping up with information is the primary challenge and the synthesis of information becomes a secondary concern. For the Questioner role, although students were given three categories of questions to ask, they neglected to ask any counter questions, which were intended to fulfill the critique function described by Wise and colleagues (2012).

Therefore, in both of these roles students effectively selected the functions of the role that favor lower-level thinking.

Self-reflections

After the addition of self-reflections before the second chat, neither the proportion of critical student-content nor student-student interactions increased. In reading the reflections we did find that students, as instructed by the self-reflection prompt, considered whether each role was properly executed in Chat 1 and how the roles could be better fulfilled in Chat 2. This is similar to the qualitative results from previous work that found self-reflections caused students to "identify patterns of participation which could enhance the larger learning community" (Kayler and Weller, 2007; Murphy and Jerome, 2005). However, neither of the studies report whether the ability to identify these patterns led to subsequent changes in chat quality (Kayler and Weller, 2007; Murphy and Jerome, 2005). In contrast to these two qualitative studies, one quantitative study found that students assigned to complete self-reflections for online asynchronous discussions did not demonstrate an increase in knowledge construction compared to students who did not complete self-reflections (De Wever et al., 2008). Our results support these quantitative findings and suggest that self-reflections alone, even if thoughtfully completed by students as indicated by our qualitative results, are not sufficient to cause students to increase their critical thinking or interaction with other students in online chats.

An important similarity between our work and that of De Wever and colleagues (2008) is that undergraduates were enrolled in the course under consideration. Undergraduates may not be able to accurately self-asses the extent of their knowledge construction (De Wever et al., 2008), and in this case, the extent of their critical thinking. The qualitative studies supporting the use of

self-reflections both considered graduate-level courses (Kayler and Weller, 2007; Murphy & Jerome, 2005). We therefore echo earlier calls for further research on the effects of providing training for undergraduates prior to completing self-assessments (De Wever et al., 2008). We also suggest that future work could look into whether changing the focus of the self-reflection to emphasize the desired pedagogical outcomes, or adding a step to make the self-reflection salient immediately before the subsequent chat, could result in a stronger quantifiable effect from the self-reflection intervention.

Group-reflections

For both Chat 3 and Chat 4, groups that completed group-reflections wrote a greater proportion of higher-level thinking posts than groups that did not complete group-reflections. This difference was significant for Chat 3, and significant for Chat 4 when all "high" outliers were removed. This suggests that providing groups an opportunity to discuss what happened in a chat and collectively decide on steps to improve the chat quality can help maintain higher-level thinking throughout the semester, and can allow students to sustain thoughtful discussion outside the framework of assigned roles. Furthermore, these results offer evidence that the self-reflection intervention alone, which was completed by all groups for Chats 2 through 4, were not enough to sustain higher-level thinking in chats throughout the semester.

In our qualitative analysis, one group, "Group A," possessed unique qualities amongst the three groups analyzed. This group demonstrated more agency than the other two groups and more frequently suggested concrete changes to address inadequacies they saw in the chats. We suggest these two results are logically related; if Group A felt empowered to solve their own problems, it follows that they would also be more likely to suggest concrete actions to take in

solving these problems. Previous work has shown that students who feel agency in online chats generally have a more positive chat experience (Kayler and Weller, 2007). This work suggests that feeling agency over the chat may also translate into students being able to suggest practical actions for improving ineffective aspects of the chat. Therefore, increasing the feeling of agency among group members seems to be a desirable outcome for chat groups. Kayler and Weller (2007) observed a correlation between the preference of face-to-face discussion as a learning style and expressing less agency in self-reflections. Thus, they added a group reflection component that required periodic in-person group meetings as a means of accommodating diverse learning styles, and hopefully promoting a feeling of personal agency in all their students. However, this does not address the question of how to promote personal and group agency in an all-online setting, with the further complication of discussion groups that are too large to be practical for video conferencing. Especially with the increased prevalence of MOOCs, we see this as an important question for future research.

Even though Group A was successful in identifying aspects of the chat, under their control, that were ineffective, and subsequently proposed concrete actions to affect change, the group had mixed success in implementing these actions in the following chat. The group was able to implement the suggestion that they focus on one topic at a time and discussed this in the subsequent reflection as a successful achievement and improvement in chat quality. However, there was a notable instance where some individuals from the group attempted to try the hand-raise feature included in the chat software, which was agreed upon by many during the group-reflection, but the idea was rejected when proposed during the chat. While students were instructed to come to a consensus over what they would change in the subsequent chat, the discussion forum was not set up in a way to facilitate reaching this consensus in a way that

included everyone. In future group-reflection activities, it seems essential that the group have a way of determining buy-in from all group-members. Recapping the group-reflection immediately prior to the subsequent chat may also help to make the reasoning for the change more salient.

The final theme that emerged was that the deficiencies identified in the chats were a mix of logistical and learning-outcome related aspects, favoring logistics over outcomes. In all three groups, a logistical aspect of the chat was identified as the primary ineffective aspect in at least one of the reflections. Specifically, multiple conversations occurring at once and an unequal distribution of time to different portions of the chat were mentioned. There are a couple possible explanations for the focus on these logistical issues. First, it seems evident that the chat format, specifically, the "rush" created by the synchronous posting, was not satisfactory and was seen by students as inhibiting their ability to have an effective chat. Second, the group-reflection prompt did not explicitly instruct students to focus on the pedagogical outcomes of higher-order thinking and questioning, therefore, it seems reasonable that students would not choose to focus on these aspects. Going forward, changes need to be made both to the chat format and the group-reflection prompt to ensure students are able to focus on the desired learning outcomes over logistics.

IMPLICATIONS FOR PRACTICE

Based on the results of this study, we plan to make the following changes in the next semester of this course:

- Focus roles on one specific function rather than wrapping multiple functions into a role, to enhance fulfillment of target functions such as "synthesize" and "critique."

- Remove self-reflections from the course assignments.
- Change chat format from synchronous to asynchronous to reduce rushing and multiple conversations.
- Change the group-reflection activity to ask students to (1) focus on the desired learning outcomes, (2) reach a consensus and get buy-in from whole group, and (3) require students to revisit the group-reflection prior to starting the next chat, in order to make the motivation behind the agreed-upon change more salient.

While our study has the limitation of being a case-study of a single course, we feel some aspects of our results can be applied more generally. We add to the body of literature demonstrating that roles help engage students in chats, especially with their peers, however, our research highlights concerns that are specific to synchronous chats when designing role descriptions. Specifically, the fast-paced nature of the chat might cause students to need the logistical functions in the roles simply to have a manageable discussion. Our research also explores the relatively undocumented process of group-reflections in online settings and provides several guidelines for improving the group-reflection process. We feel group-reflection assignments could allow for increased student autonomy in running online discussion and agency in achieving the pedagogical goals for the course.

WORKS CITED

Allen, I.E., and Seaman, J. (2016). Online report card: Tracking online education in the United States.

Bernard, R.M., Abrami, P.C., Borokhovski, E., Wade, C.A., Tamim, R.M., Surkes, M.A., and Bethel, E.C. (2009). A Meta-Analysis of Three Types of Interaction Treatments in Distance Education. Rev. Educ. Res. *79*, 1243–1289.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. (1956). Taxonomy of educational objectives, handbook I: The cognitive domain (New York: David McKay Co Inc).

Bradley, M.E., Thom, L.R., Hayes, J., and Hay, C. (2008). Ask and you will receive: how question type influences quantity and quality of online discussions. Br. J. Educ. Technol. *39*, 888–900.

De Wever, B., Van Keer, H., Schellens, T., and Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. Learn. Instr. *17*, 436–447.

De Wever, B., Schellens, T., Van Keer, H., and Valcke, M. (2008). Structuring Asynchronous Discussion Groups by Introducing Roles: Do Students Act in Line With Assigned Roles? Small Group Res. *39*, 770–794.

Ertmer, P.A., Sadaf, A., and Ertmer, D.J. (2011). Student-content interactions in online courses: the role of question prompts in facilitating higher-level engagement with course content. J. Comput. High. Educ. *23*, 157–186.

Gibbs, G. (1994). Learning in teams (Oxford: Oxford Centre for Staff Development).

Hara, N., Bonk, C.J., and Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. Instr. Sci. *28*, 115–152.

Holden, J.T., and Westfall, P.J.-L. (2007). An Instructional Media Selection Guide for Distance Learning.

Kayler, M., and Weller, K. (2007). Pedagogy, self-assessment, and online discussion groups. Educ. Technol. Soc. *10*, 136–147.

Mercer, N., Wegerif, R., and Dawes, L. (1999). Children's Talk and the Development of Reasoning in the Classroom. Br. Educ. Res. J. *25*, 95–111.

Moore, M.G. (1989). Editorial: Three types of interaction. Am. J. Distance Educ. *3*, 1–7.

Murphy, E., and Jerome, T. (2005). Assessing students' contributions to online asynchronous discussions in university-level courses. Electron. J. Instr. Sci. Technol. *8*.

Parker, K., Lenhart, A., and Moore, K. (2011). The digital revolution and higher education: College presidents, public differ on value of online learning. Pew Internet Am. Life Proj.

Pear, J.J., Crone-Todd, D.E., Wirth, K.M., and Simister, H.D. (2001). Assessment of thinking levels in students' answers. Acad. Exch. Q. *5*, 94–99.

Persell, C.H. (2004). Using focused web-based discussions to enhance student engagement and deep understanding. Teach. Sociol. *32*, 61–78.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org.

Schellens, T. (2005). The Impact of Role Assignment on Knowledge Construction in Asynchronous Discussion Groups: A Multilevel Analysis. Small Group Res. *36*, 704–745.

Schellens, T., Van Keer, H., De Wever, B., and Valcke, M. (2007). Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? Int. J. Comput.-Support. Collab. Learn. *2*, 225–246.

Strijbos, J.-W., Martens, R.L., Jochems, W.M.G., and Broers, N.J. (2007). The effect of functional roles on perceived group efficiency during computer-supported collaborative learning: a matter of triangulation. Comput. Hum. Behav. *23*, 353–380.

Sun, A., and Chen, X. (2016). Online Education and Its Effective Practice: A Research Review. J. Inf. Technol. Educ. Res. *15*, 157–190.

Tagg, A.C. (1994). Leadership from within: Student moderation of computer conferences. Am. J. Distance Educ. *8*, 40–50.

Walker, S.A. (2004). Socratic strategies and devil's advocacy in synchronous CMC debate. J. Comput. Assist. Learn. *20*, 172–182.

Wise, A.F., and Chiu, M.M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. Int. J. Comput.-Support. Collab. Learn. *6*, 445–470.

Wise, A.F., Saghafian, M., and Padmanabhan, P. (2012). Towards more precise design guidance: specifying and testing the functions of assigned student roles in online discussions. Educ. Technol. Res. Dev. *60*, 55–82.

Zhu, E. (1998). Learning and Mentoring: Electronic Discussion in a Distance-Learning Course. In Electronic Collaborators: Learner-Centered Technologies for Literacy, Apprenticeship, and Discourse, (New York, NY: Routledge), pp. 233–257.

APPENDICES

Appendix A

Each team member will be assigned a specific role that they should focus on playing during the debate. These roles represent different functions that are critical for a good discussion. Please note that you are asked to focus on your assigned role during the debate, but we expect everyone to actively provide information during the debate, justify their opinions, and challenge the views of others. The specific roles are described below:

***The Traffic-Director*:** This person is responsible for **giving direction to the debate**. The Traffic-Director **starts the debate** off, introducing the main arguments for his/her team's side. Throughout the debate, if the conversation gets stuck or off-topic, the Traffic-Director **redirects conversation** back to the main purpose of the debate.

The Questioner: This person is responsible for **prodding, challenging, and countering** the arguments made by the other team. Let's take these one by one:

- Prodding - asking questions that require the other team to provide clarifying information about a fact or an opinion (e.g. You say you had to pay too much for the ice cream - what was the exact cost?)
- Challenging - asking questions that require the other team to defend their argument (e.g. Why do you believe that cost was too high? Do you have any evidence that other people think the cost is too high?)
- Countering - making statements or asking questions that contains an alternative argument (playing devil's advocate) (e.g. If ice cream cost less, do you think Cornell students would eat more ice cream, contributing to more unhealthy diets among the student population and possibly more healthcare costs for these individuals?)

***The Synthesizer*:** This person **summarizes points** that have been made by either team as the debate progresses. These summaries should provide synopses, draw connections between points to identify themes, and highlight aspects of the discussion that have not yet been addressed.

Again, while you will be assigned one main role during the debate, we expect everyone to actively provide information during the debate, justify their opinions, and challenge the views of others.

Appendix B

CHAT REFLECTION WORKSHEET

NAME:

CHAT NUMBER:

WHAT ROLE DID YOU PLAY?:

TRAFFIC DIRECTOR

1. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role was successfully played.

2. What was the result of the Traffic Director's action, and how did it improve the discussion?

3. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role could have been better played, or was totally missing from the discussion.

4. What could the Traffic Director have done differently in this example? How would this different action have changed the discussion for the better?

QUESTIONER

1. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role was successfully played.

2. What was the result of the Questioner's action, and how did it improve the discussion?

3. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role could have been better played, or was totally missing from the discussion.

4. What could the Questioner have done differently in this example? How would this different action have changed the discussion for the better?

SYNTHESIZER

1. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role was successfully played.

2. What was the result of the Synthesizer's action, and how did it improve the discussion?

3. Provide a quote from the chat (no name unless this was the role you played) that demonstrates a time when this role could have been better played, or was totally missing from the discussion.

4. What could the Synthesizer have done differently in this example? How would this different action have changed the discussion for the better?

Appendix C

CHAT 3 REFLECTION WORKSHEET

NAME:

WHAT ROLE DID YOU PLAY?:

1. Provide a specific example (quote or concise summary of exchanges) from the chat where you feel you made a good contribution, and describe specifically how your contribution helped the discussion. Keep the three chat roles in mind, especially the one you were assigned. (*If you were not able to attend the chat, provide a specific example where you feel someone else made a good contribution (do not include names) and describe how that person's contribution helped the discussion.)*

2. In which of the three chat roles do you think you have the most room for improvement? Provide a specific example from the current chat of where you could practice this role in the future, and describe what action your future self would take. How would this action help the discussion? *(If you were not able to attend the chat, which role could you take on to make the greatest contribution to your group? Provide a specific example of where you could practice this role in the future, and describe what action your future self would take. How would this action help the discussion?*

Appendix D

In this discussion forum, you will complete a group reflection on your chat. Why do a group reflection on your chat? Because effective team work and effective conversation results when everyone is on the same page about expectations and goals. The only way this can be achieved is having everyone in the group discuss these expectations and goals. After the debate, your TA will post a transcript to the "Chat 2" thread in the "Chats" discussion forum for your discussion group. Please reference this transcript when completing this assignment.

For your first post in your group reflection forum, fill in the following two sentences:

The most effective thing about our group's chat was…..because…..

The most ineffective thing about our group's chat was…..because…..

Your initial post is due by 6pm on Tuesday, Nov 1st.

You are expected to read through all the initial posts of your group members. Then, participate in a discussion that synthesizes these posts into key actions your group should take in the next chat. For example, by the end of the discussion, your group should come to conclusions along the lines of:

Next time, our group should continue…..because…..

Next time, our group should start/stop/change…..because…..

We are not looking for a specific number of replies in this forum. Rather, we are interested in your ability to **synthesize** what is being said and be **responsive** to your group-mates.

Your replies in the group discussion forum are due by 6pm on Thursday, Nov 3rd. Please review the attached rubric to see how you will be graded.