

DEMOGRAPHICS AND TRANSFER OF *ESCHERICHIA COLI* WITHIN *BOS*
TAURUS POPULATIONS

A Thesis
presented to
the Faculty of California Polytechnic State University,
San Luis Obispo

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Biology

by
Joshua Ryan Dillard
September 2015

©2015

Joshua Ryan Dillard

ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE: Demographics and Transfer of *Escherichia coli* within *Bos taurus* Populations

AUTHOR: Joshua Ryan Dillard

DATE SUBMITTED: September 2015

COMMITTEE CHAIR: Michael W. Black, Ph.D.
Professor of Molecular Biology

COMMITTEE MEMBER: Christopher L. Kitts, Ph.D.
Biology Department Chair, Professor of Biology

COMMITTEE MEMBER: Alexander Dekhtyar, Ph.D.
Professor of Computer Science

ABSTRACT

Demographics and Transfer of *Escherichia coli* Within *Bos taurus* Populations

Joshua Ryan Dillard

In the United States, symptoms caused by pathogenic strains of *Escherichia coli* are on the rise. A major source of these pathogenic strains is the *E. coli* in the digestive tract of cattle. The purpose of this project was to determine if *E. coli* are transferred between individuals of the same species and if interspecies transmission is possible. Proximity of cattle was also studied as a contributing factor to the transfer of *E. coli*. To accomplish this goal, *E. coli* isolates from cattle and cohabitating ground squirrels were compared through a new method of bacterial strain typing called pyroprinting. Bulls from the Cal Poly Bull Test were sampled every summer from May to September when around 200 bulls from ranches across California are housed together to be tested and eventually auctioned off. The impact of cattle origin (ranch, city) and habitation (pen) on *E. coli* isolate strain type were evaluated via pyroprinting. The cattle were studied to see if transfer was related to proximity of cohabitation. Since the complete population of intestinal *E. coli* could not be sampled, transfer could not be directly seen. The probability of sharing *E. coli* in each time point was used to infer transfer. There was an increase in the probability of sharing *E. coli* from the May sample date to the September date, indicating that some form of transfer was occurring. There was an even greater increase in the probability of sharing *E. coli* when the bulls were housed in close proximity. Lastly, ground squirrels cohabitating in the area were found to house some of the same strains as the cattle. This makes transfer between squirrels and cattle a

possibility. Overall, this paper shows that the intestinal *E. coli* composition of bulls may be readily altered by the introduction of new bulls into a population.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor Dr. Michael Black for all of his support during my thesis, both in the laboratory and during editing. Besides my advisor, I would like to thank Dr. Chirstopher Kitts for all of his help with the experimental design and the many months of thesis editing together. I would also like to thank Dr. Alexander Dekhtyar for all of his help with the CPLOP database and clustering algorithms used in this experiment.

I would like to thank all the fellow lab mates who helped with pyroprinting and collecting the *E. coli* isolates, as well as Emily Neal and Dr. Jennifer VanderKelen for instructing me in the use of a pyrosequencer and other methods. Special thanks also goes out to Aldrin Montana for his assistance on my project by developing the hierarchical and ontological clustering methods used. Also, I thank my fellow graduate students at Cal Poly, my parents, and sister for all they physical and emotional support during my time in the Master's Program.

Last but not least, I would like to thank the W. M. Keck Foundation, the National Science Foundation, and CSUPERB for funding my project. Without their generosity this project would not be possible.

TABLE OF CONTENTS

	Page
LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER	
I. INTRODUCTION	1
A Background on <i>Escherichia coli</i>	1
<i>E. coli</i> pathogenicity and virulence factors	2
<i>E. coli</i> Prevalence in Cattle, Squirrels, and Environmental Contamination	5
Classification of <i>E. coli</i>	6
Methods of Strain Typing	8
Pyroprinting	10
Comparing Pyroprints	13
Purpose	14
Testing Hypotheses	14
II. METHODS	16
Collecting Samples	16
Confirming <i>E. coli</i>	17
Colony PCR and Pyroprinting	17
Matching Pyroprints and Clustering Isolates	19
Statistics	20
III. RESULTS	21
Sampling Bulls	21

A Sample Taken from a Single Bull	21
Isolates per Strain.....	22
Strain Sharing Between Bulls over Time.....	23
Geographical Origins of Strains.....	24
Strain Sharing Between Bulls from Different Ranches	26
Strain Makeup and Host Distribution across Ranches.....	29
Probability of <i>E. coli</i> Sharing.....	31
Probability of Sharing Increases after 4 Months of Cohabitation.....	33
Probability of Sharing Across Ranches	34
Strain Sharing Between Cohabiting Bulls	35
Squirrel Cohabitation Effects on <i>E. coli</i> Sharing.....	38
IV. DISCUSSION.....	40
REFERENCES	48
APPENDICES	
A. Pen Breakdown for Both Years	53
B. Pen Breakdown for Time Points	53
C. Ranch Distribution	54
D. Pen Distribution 2011	55
E. Pen Distribution 2012.....	55
F. Pen 7 Strain Distribution	56
G. Isolates across Time Points.....	56
H. Cross Time Point Strain Sharing.....	57

I. Probability of Sharing for Different Pens.....	57
J. Total Numbers Collected.....	58

LIST OF TABLES

Table		Page
1. Primers sequences and dispensation orders used during this study		18
2. Number of total strains, number of single isolate strains, and number of multi-isolate strains for the two different clustering methods		23
3. Probability of sharing overall using hierarchical clustering for the 2011 and 2012 sampling years.....		33
4. Probability of sharing overall for Hierarchical, Ontological, and Matching data for the 2012 sampling year		33
5. The probability of sharing for same bull and different bull for the 2012 sampling year. This table shows probability of sharing between isolates from the same bull in the left three columns and different bulls in the right three columns. Hierarchical, Ontological, and Matching clustering methods were used		34
6. The Probability of sharing for different ranches and for the same ranch different bull for the 2012 sampling year. The left three columns show probabilities for matching between different Ranches. The right three columns show probabilities for matching between different bulls from the same ranch of origin. The data shows Hierarchical, Ontological, and Matching methods.		35

7. The probability of sharing for same pen same ranch and same pen
different ranch for the 2012 sampling year. The left three columns
show probabilities for matching between bulls from the same ranch
of origin and housed in the same pen. The right three columns show
probabilities for matching between bulls from the different ranch of
origin and housed in the same pens. All data is shown for Hierarchical,
Ontological, and Direct Matching methods38

LIST OF FIGURES

Figure	Page
<p>1. ITS regions for pyrosequencing, noncoding regions surrounded by highly conserved rDNA genes. The arrows indicate the primers used to amplify ITS region. ITS 1 is the 16-23 region while ITS 2 is the 23-5 region. A streptavidin bead is attached to the reverse primer</p>	11
<p>2. Diagram of the seven copies of the ribosomal DNA with two SNP's present. The different colors represent the different nucleotides as they are dispensed during pyrosequencing. The white areas in ITS1-6 and ITS1-7 are where SNP's caused the reading frame to become out of sync</p>	11
<p>3. Pyrogram of 23-5 ITS region sample.....</p>	12
<p>4. Pearson Correlation of pyroprints in the 23-5 region. Red indicates a match of 99.0% or greater. Each box represents the column isolate pyroprint compared to the row isolate pyroprint</p>	13
<p>5. Mathematical equation used to determine a Pearson correlation. X is the peak height for the first isolate and y is the peak height for the second isolate at the same dispensation</p>	19
<p>6. Pie chart of the 13 strains found from 50 isolates taken from one bull. The number inside each slice indicates the number of isolates in that strain 1</p>	22
<p>7. Grouping of <i>E. coli</i> strains across the different sampling dates. The grey bar represents the different dates isolates from a strain were found</p>	

in. The total number of strains found sharing across specific dates is represented numerically on the side of the chart. Ontological clustering method is the chart on the left, while hierarchical clustering method is the chart on the right. The dotted line indicates the cutoff between cohorts24

8. Geographic origin of isolates within a strain. Geographical maps of California with a black star indicating the location of the ranches of origin found in that strain. The three maps on top are from strains produced with hierarchical clustering, whereas the three maps on the bottom are strains produced with ontological clustering.....25

9. The number of different ranches in which isolates from the same strain were detected overall for 2011 and 2012, May and September26

10. The number of different ranches in which isolates from the same strain were detected (2011)..... 27

11. The number of different ranches in which isolates from the same strain were detected (2012).....29

12. The number of different ranches in which isolates from the same strain were detected for the sampling year 2012 broken down by strain. Each color in the bar graph represents a different strain and the isolate size of that strain in the overall distribution. The black numbers on top of the bars indicate the total number of strain in that bar31

13. Photograph of the Cal Poly Bull Test Pens. The white numbers indicating pen numbers are shown over the geographic location of each pen on the map36

14. Percentage of sharing comparing nearby pens to distant pens. A&C (left pie graphs) show the percentile difference between May 2012 strains that share isolates with nearby pens (blue color) compared to strains that share isolates with distant pens (red color). B&D (right pie graph) show the same difference for the September 2012 time point37

15. Strains with isolates from both squirrels and bulls. Bull isolates are further identified by their date of their collection39

I. Introduction:

A Background on Escherichia coli:

Escherichia coli is a species in the *Enterobacteriaceae* family, and can be categorized further into individual strains. Different strains of *E. coli* are simply different variations of the same species. These variations contain slightly different genetic structures, giving different strains different attributes. Most strains of *E. coli* are commensal bacteria found in the intestines of mammals and birds. Within human intestines, probiotic strains of *E. coli* help to maintain healthy intestinal function (Vrese & Schrezenmeir, 2008). Different animal species vary in the number and type of *E. coli* strains they harbor. In general larger animals have a much larger diversity of *E. coli* strains within their intestines than smaller animals (Gordon & Cowling, 2003). This correlation is likely due to the fact that larger animals (specifically herbivores such as cattle) have a longer and more complex digestive system, creating more niches for bacteria to colonize.

When an animal defecates it releases *E. coli*, as well as many other microorganisms and viruses, into the environment through its fecal matter. This leads to *E. coli* contaminating water sources. Thus *E. coli* is known as a fecal indicator bacterium (FIB). This means that the presence of *E. coli* indicates a water sample is contaminated by fecal matter. Although *E. coli* is not well suited to thrive in water, it can survive in water sources for up to 12 weeks (Wang & Doyle, 1998). Due to *E. coli*'s ability to survive in water, and it being a FIB, when a water source is suspected of fecal contamination, it is generally tested for the presence of *E. coli* (Clermont et al., 2011). A water source contaminated with fecal matter will not only contain *E. coli*, but many other

potentially pathogenic bacteria and viruses previously housed in the intestinal track of animals. Due to *E.coli's* ability to survive in water, the absence of *E. coli* in a water source is used to indicate that no other fecal pathogens are present.

E. coli pathogenicity and virulence factors:

Many people have come to think of *E.coli* only as a pathogen, without recognizing the diversity of strains present in the species. Only by gaining virulence genes do some strains of *E. coli* change from normally commensal bacteria to a life threatening pathogen. Pathogenic strains of *E. coli* account for 16 percent of the nearly 30,000 cases of food borne illnesses each year in the United States. According to the Center for Disease Control, in 2009-2010, 46 individuals were hospitalized with serious complications due to pathogenic *E. coli* strains, leading to 3 deaths (“Center for Disease Control and Prevention,” 2015). Pathogenic strains of *E. coli* most commonly enter the body through ingestion of contaminated foods. Ground beef is the food most commonly contaminated with *E. coli*. It is also possible for foods, such as salads or raw appetizers, to get *E. coli* by being in close proximity to contaminated beef products in the restaurant. Foodborne illness is not the only type of infection, *E. coli* can also cause nosocomial infection found in hospitals.

Once a person is infected with a pathogenic strain of *E. coli* multiple symptoms may occur. The most common symptom is bloody diarrhea and stomach cramping as the bacteria bind to the epithelial cells of the intestines. However, *E. coli* may also cause urinary tract infections, as well as septicemia in some cases (Nicolle, 2008). If the urinary tract infection worsens it can lead to other secondary conditions such as hemolytic uremic syndrome (HUS) (Tawil, Sacher, Mandeville, & Meunier, 2012). HUS

is a life threatening condition that causes blood cells to lyse and eventually leads to kidney failure (Tawil et al., 2012).

The pathogenic properties of *E. coli* are caused by virulence genes within the *E. coli* genome. The effects of these virulence factors include: an increased ability to colonize, inhibition or evasion of host defenses, and an increased ability to obtain nutrients to live while inside the host cells. The two major virulence factors found in *E. coli* are toxins called Shiga Toxin 2 and Shiga Toxin 1. Shiga toxin 1 or *stx1* is a toxin very similar to the *stx* gene that was originally found in *Shigella dysenteriae*. However, neither of the Shiga Toxin genes found in *E. coli* are the original gene found in *S. dysenteriae*. Shiga Toxin expressing *E. coli* arose from less virulent progenitors after acquiring a *stx2* bacteriophage at a single site and *stx1* bacteriophage during another separate and single event (Shaikh & Tarr, 2003). The combination of these two *stx* genes created the Shiga Toxin producing *E. coli* known today. This virulence factor can be passed to other *E. coli* strains through phage mediated horizontal gene transfer. In a 2008 study (Sekse, Solheim, Urdahl, & Wasteson, 2008), researchers were able to show that a pathogenic O157:H7 strain can transfer its pathogenicity to a nonpathogenic strain in close proximity through transduction by *stx* phages. The study found that allowing sheep to consume food laced with O157:H7 would result in some of the native nonpathogenic strains of *E. coli* taking up the shiga toxin virulence factor.

The virulence factors that cause *E. coli* strains to become pathogenic can be found in waste sludge from around and in cattle feedlots. If cattle, like sheep, can have their nonpathogenic strains transformed into pathogenic ones, then the virulence factors found within the sludge in feedlots could be a major contributor to pathogenic *E. coli* strain

outbreaks. There are many different factors affecting the ability of the phage to properly distribute, or increase the likelihood of phage transfer occurring. One of the most important factors is temperature. For example, lambda prophage activation requires a threshold temperature of 20°C, and below this temperature the prophage cannot be activated (Gabig, Obuchowski, Srutkowska, & Wegrzyn, 1998; Yue, Du, & Zhu, 2012). This is important since many pathogenic strains of *E. coli* obtain their Shiga Toxin genes through horizontal gene transfer mediated by lambda prophages. *E. coli* O157:H7, for example, has its *stx* genes located in lambdoid prophages within its genome (Yue et al., 2012). So any increase in lambda prophage activation could increase distribution of *stx* virulence genes. An increase of the temperature to 37°C will allow the development of the lambda phages to occur normally (Gabig et al., 1998; Yue et al., 2012). Temperature is a determining factor for the production of *stx* containing phages. *Stx* proteins are more abundant in phages developed in higher temperatures. Likewise, the amount of protein continues to increase as UV radiation is added to the samples kept at temperatures ranging from 22-37°C (Yue et al., 2012). Thus, a synergistic effect is seen between the exposure of phages to UV radiation and the temperature of the environment (Yue et al., 2012). The Yue data were then used to determine why cattle pens had such a high occurrence of pathogenic *E. coli* strains. Due to the combination of a high heat environment, along with the UV radiation of the sun, which increases lambda phage production, these feedlots are optimal places for phage transfer. The sludge in these feedlots contained high amounts of *stx* lambda phages, which in turn greatly increased the likelihood of the *stx* gene being taken up by an originally nonpathogenic strain (Yue et al., 2012).

Many of these pathogenic strains do not affect the cows themselves; however, these same strains are able to cause disease in humans when affected cow meat was handled or eaten. The prevalence of cattle in the worldwide food supply creates an increased possibility of *E. coli* infection. This coupled with *E. coli*'s effective virulence factors and ease of genetic transmission, makes *E. coli* a constant source of worldwide contamination.

***E. coli* Prevalence in Cattle, Squirrels, and Environmental Contamination:**

E. coli can colonize the intestinal track of a variety of mammals, with different mammals often containing different strains. The size of the animal and complexity of the digestive system are contributing factors to the variation of strains found in that animal. Ruminates (cattle in particular) house the largest variety of different *E. coli* strains (Wells et al., 1991). Cattle also house a larger percent of pathogenic *E. coli* strains in their digestive tract compared to other animals (Osman, Mustafa, Elhariri, & Abdelhamed, 2013). During the Osman study, pathogenic *E. coli* strains were isolated from cattle, sheep, and goats. The majority (63.6%) of the pathogenic strains collected came from cattle, where as 27.3% came from goats and 9.1% from sheep. Pathogenic *E. coli* collected from from cattle fecal matter had similar serotypes and virulence genes to those found in human patients with gastrointestinal and other diseases (Osman et al., 2013). Osman et al suggest cattle are likely the largest source of pathogenic *E. coli* affecting humans. In a related study, runoff water from cattle slaughter houses and cattle carcasses have a high prevalence of pathogenic *E. coli* contamination (Ayaz, Gencay, & Erol, 2014). Of the 744 cattle sampled in the study, 4.2% of the cattle rectal swabs and 6.3% of the cattle carcasses contained pathogenic *E. coli* strains. In addition, 20.8% of all

slaughter house waste water samples contained pathogenic strains (Ayaz et al., 2014) Since *E. coli* can be easily transferred from cattle to a water source, and that water source has the ability to flow into cattle populated areas, looking at the ability of *E. coli* to be shared between cattle will help when determining the possibility of one individual bull picking up a pathogenic strain of *E. coli* from the water and sharing it with the other cattle.

Interspecies transmission of *E.coli* may also be possible. A very common infestation among cattle ranches is the abundance of ground squirrels. This makes squirrels viable candidates for interspecies transmission. Because of their small size and less complex digestive system, squirrels house a much lower number of different strains of *E. coli* in their digestive tract than cattle (Hassan, Ellender, & Wang, 2007). According to the Hassan study, squirrels can also share their intestinal flora with other species. If squirrels are able to house similar strains of *E. coli* as cattle, then this could lead to wider area of spread, as squirrels movement is not inhibited by fencing. This could lead to contamination of new environments. If squirrels are able to house cattle specific strains of *E.coli*, then it is possible for squirrels to carry those strains from the cattle pens of a farm to the crop portion of the farm. By comparing *E. coli* isolates in bulls to those in squirrels, it is possible to detect strains made up of both bull and squirrel isolates. Confirmation of these shared stains would indicate the possibility of cross contamination of soil, water, or distant bull from a squirrel transferred strain.

Classification of E. coli:

Looking at *E. coli* strain transfer is important, and to do so one must first determine what constitutes a strain. *E. coli* is a versatile bacterial species able to survive

for many weeks in water and sediment (Clermont et al., 2011). Each different strain of *E.coli* has a different ability to adapt to particular environments. *E. coli* contains a core genome of less than 2000 conserved genes, and the species as a whole includes around 10,000 genes in total (Clermont et al., 2011; Rasko et al., 2008; Touchon et al., 2009). Because of this potential for variance, *E. coli* has a large number of different strains with a variety of possible gene combinations. In a perfect world, the entire genome of *E. coli* would be sequenced and only the *E. coli* with complete matching genomes would be considered the same strain. However, total genome sequencing is very time consuming and expensive, so less thorough ways of determining strains are more commonly used.

E. coli are organized into one of 6 phylogenetic groups :A, B1, B2, C, D, E, and F (Clermont et al., 2011) based on genetic data and morphology. Determining how to classify a strain is much more difficult. There is disagreement around what constitutes a strain. Some studies infer that only two cells with a direct genetic match (a clone) are truly the same strain. While others studies state that within two cells of the same strain there may be single nucleotide polymorphisms that make the two cells slightly different, but they would still belong to the same strain. There are two major ways of classifying a strain, phenotyping and genotyping. Phenotyping uses morphology, serology, biochemical testing, and antibiotic susceptibility to determine a strain. However, phenotyping is not discriminating enough to distinguish closely related strains (Li, Raoult, & Fournier, 2009). Accepted methods for classification involve some form of genotyping. Three categories of strain typing have resulted from these differences of opinion. These are DNA banding pattern typing, DNA sequence typing, and DNA hybridization (Li et al., 2009). The Li paper concluded DNA sequence typing was the

optimal method of determining a strain, however different strain typing methods are used by different researchers.

Methods of Strain Typing:

Strain typing is used frequently for microbial source tracking, pathogen identification, and bacterial transmission studies. For this reason, there are multiple procedures used to accomplish strain typing. There are many different forms of strain typing from immunotyping serology stated above to molecular methods that determine strain differences by comparing non-coding and coding regions of the bacterial DNA.

Serotyping is a method using antigens found on the surface of bacteria and categorizes bacteria with similar surface antigens into the same serogroup. This is an inexpensive way of categorizing *E. coli*. For example, O157:H7 refers to two different surface antigens on this pathogenic strain of *E.coli*. On the other hand, it is not as discriminating as methods using genetic information. Strains can have different genetic make ups, yet still contain the same surface antigens, therefore, serotyping has the tendency to combine two different strains into the same serogroup (Ren et al., 2008).

Ribotyping uses genetic information instead of antibodies to differentiate strains. This method creates genetic fingerprints by comparing the ribosomal ribonucleic acids (rRNA), hence the name ribotyping. Ribotyping works by using DNA restriction enzymes to cut genes coding for rRNA into pieces. After the DNA is digested with a specific restriction enzyme, the fragments are separated via gel electrophoresis. The separated DNA pieces are then transferred to a nylon filter and hybridized to labeled DNA probes. The labeled DNA fragments create distinct band patterns for each different strain (Grimont & Grimont, 1986). However small changes in nucleotides that don't

affect the overall length of the fragments would go unnoticed, so like serotyping, ribotyping tends to be less discriminating between strains than other molecular methods (Kristjánsson et al., 1994).

Restriction endonuclease analysis, REA, is a process in which DNA fragments are created through digestion with restriction enzymes much like ribotyping. However, unlike ribotyping, the resulting pieces are then used to create a restriction map of the position of all the restriction sites within a sequence of DNA. Whereas ribotyping takes into account the patterns of bands, REA creates a complete restriction map showing the position of restriction sites on a specific DNA sequence. Comparing these maps allows similarities between strains to be determined. These gels can be more difficult to interpret and the restriction maps can be difficult to digitize for database analysis so, human error is much more likely to occur with REA (Kristjánsson et al., 1994).

Another molecular strain typing method is amplified fragment length polymorphism (AFLP). This process uses a procedure similar to REA where restriction enzymes digest cellular DNA; however AFLP then ligates half site specific adaptors to the restriction ends. Due to these specific adaptors selective amplification is possible. Once a specific fragment is amplified, electrophoretic separation is used to separate the amplified fragments on a gel. These amplified fragment bands are then visualized and compared to determine similarity (Vos et al., 1995). The highly sensitive ability of AFLP to detect DNA polymorphisms and high level of reproducibility makes this a great method for genetic fingerprinting (Vos et al., 1995). However, it is more expensive and time consuming than REA or ribotyping.

Multilocus sequence typing (MLST) is another means of differentiating between bacterial strains. There are multiple variations of MLST analysis that use different numbers of housekeeping genes. In one variant of MLST, 11 housekeeping genes are sequenced to directly measure the genetic variation. The results can then be compared to determine phylogenetic relationships (Maiden et al., 1998). MLST has a higher sensitivity to genetic differences than ribotyping or REA, and the data is easily loaded into a database. MLST's drawbacks are in the high cost of operation and the large time commitment necessary to complete the process.

After taking all these procedures into consideration, this project used a new method of bacterial strain typing called pyroprinting which combines the speed of ribotyping with the high discriminatory power of MLST.

Pyroprinting:

Pyroprinting is a new, highly reproducible method of bacterial strain typing. Pyroprinting uses DNA pyrosequencing of multiple polymorphic loci and, in doing so, increases the ability to discriminate between two closely related strains (Black et al., 2014). Pyroprinting begins with PCR of the target regions. The target regions for this method are the internal transcribed spacers (ITS) (small regions in between coding DNA regions) in the rRNA operons. The ITS regions could contain large sequencing differences without effecting major cellular processes. The highly variable DNA sequences of these ITS regions make an optimal target for pyroprinting.

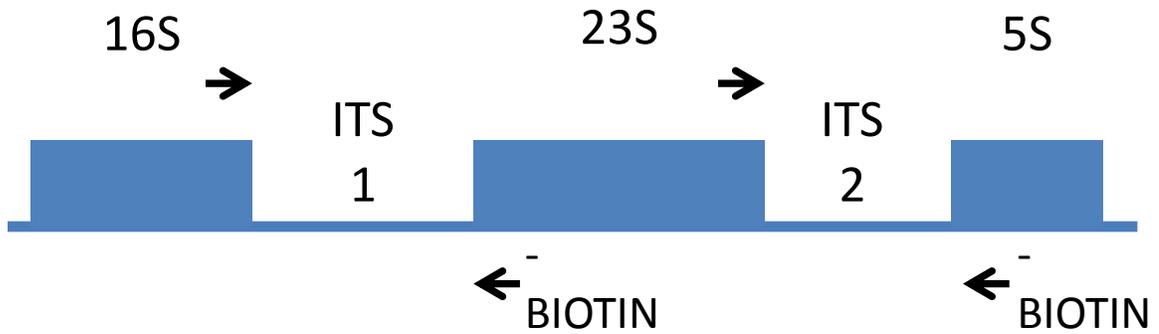


Figure 1: ITS regions for pyrosequencing, noncoding regions surrounded by highly conserved rDNA genes. The arrows indicate the primers used to amplify ITS region. ITS 1 is the 16-23 region while ITS 2 is the 23-5 region. A streptavidin bead is attached to the reverse primer.

There are seven copies of the rRNA operon in *E. coli*. These seven copies all have slightly different genetic sequences due to single nucleotide polymorphisms (SNPs). All seven copies are sequenced during pyrosequencing since the primers bind to the identical coding regions that flank the ITS regions. During pyrosequencing the nucleotides are dispensed in a specific order. During this the seven copies can be read at different rates depending on their sequence (Figure 2).

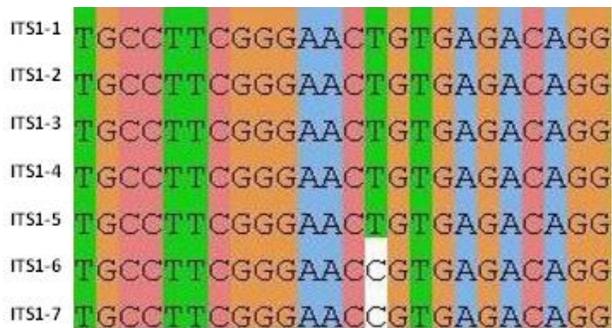


Figure 2: Diagram of the seven copies of the ribosomal DNA with two SNPs present. The different colors represent the different nucleotides as they are dispensed during pyrosequencing. The white areas in ITS1-6 and ITS1-7 are where SNPs caused the reading frame to become out of sync.

Due to this determining an exact DNA sequence is not possible, but what is generated instead is a pattern of light peaks (Figure 3), generated as the nucleotides bind to the target DNA. These patterns of light peaks are used to compare the strains.

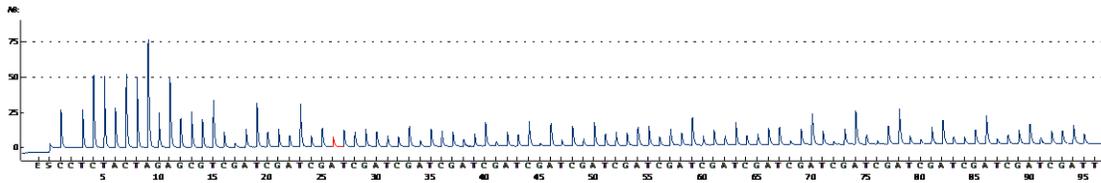


Figure 3: Pyrogram of 23-5 ITS region sample.

During pyrosequencing, the DNA of the ITS region is bound to a streptavidin bead via the biotinylated DNA fragment (Figure 1) containing the specific binding sequence. The DNA is then denatured and washed so that the bead-bound single strand is all that remains. Next, a specific primer and DNA polymerase are added to the solution to start replication as soon as nucleotides are present. Other enzymes are also introduced such as apyrase, ATP sulfurylase, and luciferase. The pyrosequencer releases one type of dinucleotide at a time from a cartridge, into the DNA solution. If the nucleotide is the next complementary nucleotide in the DNA sequence, it gets incorporated into the growing DNA strand and a pyrophosphate is released as a byproduct. The ATP sulfurylase then reacts with the pyrophosphate and adenosine phosphosulfate (APS) to generate ATP. The ATP then reacts with luciferin, in conjunction with luciferase, to create a single photon of light per molecule of ATP (Black et al., 2014; Ronaghi, 2001). Pyroprinting is completed separately for both the 23-5 ITS region (ITS-2) and the 16-23 ITS region (ITS-1) and continues for a total of 95 (ITS-2) or 94 (ITS-1) dispensations depending on the locus. When finished, the light generated is measured at each dispensation and relayed as a peak height graph (Figure 3).

Comparing Pyroprints:

Since the output light peaks different from sample to sample (due to different total amounts of DNA) the results must be standardized before they are compared. Standardization is important since the differences in light peak patterns are being compared rather than the numerical peak heights. A Pearson correlation was used for comparison since it standardizes the data sets before they are measured (Figure 4). A correlation of 99.0% was determined to be sufficient for a match between two pyroprints (Black et al., 2014). Therefore, if two pyroprints are correlated at 99.0% or above in both ITS regions they are considered to be a match. This means a true strain match must return a correlation of 99.0% or higher in both ITS-1 and ITS-2. Pyroprinting is useful in determining differences between extremely similar strains, and the cost effectiveness of pyroprinting allows for easy generation of pyroprints for a large variety and number of bacterial isolates. The high rate of reproducibility and ease of use made pyroprinting our optimal choice when choosing a method of bacterial strain typing.

	Cw-434	Cw-435	Cw-436	Cw-437	Cw-438
Cw-434	100				
Cw-435	99.4	100			
Cw-436	99.0	98.6	100		
Cw-437	98.9	98.7	99.9	100	
Cw-438	96.6	97.0	95.6	95.7	100

Figure 4: Pearson Correlation of pyroprints in the 23-5 region. Red indicates a match of 99.0% or greater.

Each box represents the column isolate pyroprint compared to the row isolate pyroprint.

Purpose:

This paper will show that transfer of *E. coli* strains is occurring between the studied populations of bulls by using the probability of bulls sharing *E. coli* to indicate transfer in bull populations. If the probability of two isolates being shared increases over time, it is an indication that *E. coli* transfer is occurring.

Testing Hypotheses:

This study investigated the transmission of *E. coli* within a population of bulls, along with the cohabitating squirrels in the pen environment. We hypothesized *E. coli* transfer occurs between cohabitating cattle. Yet, due to the small number of isolates taken from each bull, it is impossible to directly visualize strain transference, so the probability of sharing was used to infer transfer. An increase in this probability over time implies that transfer of *E. coli* is occurring.

- Primary Hypothesis: Transfer of *E. coli* occurs between cohabiting cattle.
 - Prediction I: Probability of sharing increases after four months of cohabitation.
 - Prediction II: Probability of sharing between bulls with the same ranch of origin is greater than between bulls with different ranches of origin at the start of the experiment.
 - Prediction III: Probability of sharing between bulls of same ranch of origin will decrease after 4 months of cohabitation with bulls from different ranches of origin.

- Prediction IV: Probability of sharing between bulls of different origins housed in the same or neighboring pens is greater than those housed in distant pens.
- Secondary Hypothesis: Cattle and cohabitating squirrels can host the same strains of *E. coli*.

II. Methods:

Collecting Samples:

Bulls were sampled for *E. coli* at the Cal Poly Bull Test, in which Cal Poly receives bulls from different ranches across California and houses them in their cattle fields for four months. During this time, the bulls are housed in pens while their fitness and worth are determined. After four months the bulls are then sold to meat production plants or for breeding purposes.

Four testing dates were used covering two annual bull tests. The first cohort of 180 bulls was sampled on May 28th, 2011, 16 days after arrival. On September 3rd, 2011, 167 of the same cohort of bulls were resampled (13 of the bulls were killed by illness or sold off before the second sampling). On May 12th, 2012 the second cohort of bulls containing 200 bulls was sampled on their first day of arrival. On September 1st, 2012 192 of the second bull cohort was resampled. Squirrels were randomly sampled throughout the time frame of this experiment.

Fecal samples were taken directly from the bulls, streaked out for single isolates onto McConkey agar plates (Zimbardo & Power, 2009), and incubated at 35°C for 24 hours. Four different isolated colonies from the original samples were streaked out for isolation onto new McConkey agar plates. These plates were then placed into an incubator at 35°C and allowed to grow for 24-48 hours. The entire sampling and *E. coli* isolation process was repeated for each of the three remaining test dates. In conjunction with the bull sampling 49 squirrels were harvested and euthanized, after which 10 isolates were taken rectally from each squirrel. *E. coli* isolates were obtained using the same methods used for the bulls.

Confirming E. coli:

After the McConkey plates were incubated they were then checked for the presence of isolated pink colonies. Half of a single isolated colony was plated onto an EMB plate (Zimbrow & Power, 2009) while the other half of the same isolated colony was streaked out on LB plates (Zimbrow & Power, 2009). These plates were incubated overnight at 35°C. A metallic sheen on the surface of the bacterial colonies on EMB plates is a positive confirmation for *E. coli*. Isolates that showed a positive confirmation on EMB were spot inoculated onto a citrate plate and inoculated into tryptone broth. The tryptone broth and citrate plates were placed into the incubator at 35°C for 24 hours. Simmon's Citrate plates test for the ability of the bacteria to use citrate as their only source of carbon and ammonium as their only nitrogen source. A negative reading on Citrate agar is a positive confirmation for *E. coli*. Tryptone broth tests for the ability of bacteria to degrade tryptophan and produce indole. Indole can be seen by adding Kovac's solution to the test tubes. A bright red or pink coloration on the top of the tube is a positive result for the production of indole and a positive confirmation of *E. coli*.

Of the confirmed *E. coli* isolates a maximum of two isolates were placed into LB+ 10% glycerol media and stored in a -70°C freezer until used for pyroprinting.

Colony PCR and Pyroprinting:

A master mix was made containing 2X- Quick load TAQ polymerase (New England Biolabs, MA), water, forward primer [0.2 µM], and reverse primer with biotin [0.2 µM] (Table 1) for a final volume of 25 µL.

Primer Name	Sequence
ITS-2 Foward	5'-ATGAACCGTGAGGCTTAACCTT-3'
ITS-2 Reverse Biotin	Biotin 5'- CTACGGCGTTTCACTTCTGAGT-3'
ITS-1 Forward	5'- GGAACCTGCGGTTGGATCAC-3'
ITS-1 Reverse Biotin	Biotin 5'- CTTCATCGCCTCTGACTGCC-3'
ITS-2 Sequencing Primer	5'- CGTGAGGCTTAACCTT-3'
ITS-1 Sequencing Primer	5'- GGAACCTGCGGTTGGATCAC-
ITS-1 Dispensation Order	3'CCTCTACTAGAGCG 20(TCGA)TT
ITS-2 Dispensation Order	AACACGCGA 21(GATC)C

Table 1: Primers sequences and dispensation orders used during this study.

Cells from a single colony of a confirmed *E. coli* sample were resuspended in the reaction mixture. PCR cycling parameters were: 95°C for 2 minutes followed by 45 cycles of 95°C for 30 seconds (40 cycles for 16-23 ITS), 55°C for 30 seconds and 68°C for 4 minutes. PCR products (5µls) were confirmed using a 2% agarose gel with a master mix free of DNA as the negative control. PCR was repeated if a band appeared in the negative control.

PCR product (20 µls) was sequenced by adding 20 µl to the Pyromark Q24 binding buffer using protocols posted by the manufacturer (Qiagen, MD) see table 1 for primer sequences and dispensations. The light peak outputs were stored on an online database (cplop.org) managed through a cooperative effort between the Cal Poly Computer Science department and the Cal Poly Biology department.

Matching Pyroprints and Clustering Isolates:

Pyroprints from the ITS-1 and ITS-2 for all *E. coli* collected in the study were evaluated by direct Pearson correlation matching and two different isolate clustering methods. The direct matching method compared each isolate to every other isolate in the data set. Pairwise correlations over 99.0% for both ITS regions were documented as a match. A matching probability was then generated by comparing the number of matching pairs in a certain category to the total number of pairs in that category.

Pearson correlation was used to determine the similarity between the peak heights in a given dispensation between two different pyroprints. In the equation below (Figure 5), x and y represent the two different peak heights being compared, and the sum for all 94 or 95 dispensations (depending on the ITS region) (Figure 5).

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}}$$

Figure 5: Mathematical equation used to determine a Pearson correlation. X is the peak height for the first isolate and y is the peak height for the second isolate at the same dispensation.

Two clustering methods were used to group matching isolates into strains. The first was hierarchical clustering, in which all isolates are clustered in order of similarity, with each new isolate compared to existing clusters (Corpet, 1988). The second was ontological clustering, where isolates were clustered using predetermined factor groupings (Montana, 2013). The hierarchical clustering method starts out with all isolates separated and uses a best match to worst match order, to group isolates into clusters. Ontological clustering compares the isolates in an order based on attributes that

we determined would be most likely shared in the same strain. The order used: Host ID> Ranch ID> Pen ID> Sample date.

Statistics:

A two sample T-test was used to determine the relationship between two similar patterns of strain distribution. A Test for Two Variances was used to determine the relationship of the spread across times, locations, and pens. A Chi squared test was used to determine significance between the differences in the probability of sharing *E. coli*.

III. Results:

Sampling Bulls:

This thesis looked at the transfer of *E. coli* strains within a population of bulls. The first step was collecting *E. coli* isolates from 200 bulls during the Cal Poly Bull Test. Since this thesis looked at the *E. coli* present before and after cohabitation, it was important to sample the bulls twice: once when they came in, and once before they left. To do this, the bulls were sampled once in May (upon entry) and again in September (before leaving). This sampling was conducted for two back-to-back years (2011 and 2012). However, the May sampling date for the 2011 cohort was taken 2 weeks after arrival. The effects of this will be discussed further on in this thesis.

A Sample Taken from a Single Bull:

According to the literature, ruminants (bulls) have a high diversity of *E. coli* strains within their gut, due to their feeding habits and intestinal anatomy (Wells et al., 1991). To confirm this using pyroprinting, a study was conducted in which 30 isolates were taken from a single bull. These 30 isolates clustered into 13 different strains (Figure 6), of which 8 contained only a single isolate.

A Sample of 50 Isolates Taken from a Single Bull

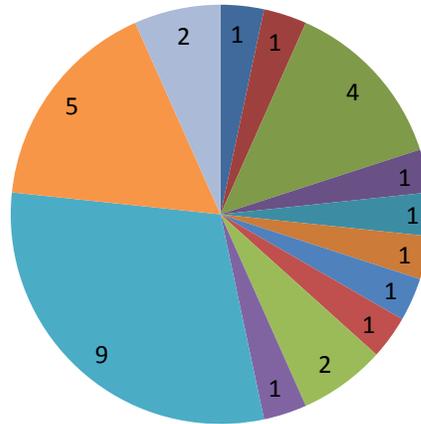


Figure 6: Pie chart of the 13 strains found from 50 isolates taken from one bull. The number inside each slice indicates the number of isolates in that strain.

Isolates per Strain:

The number of isolates contained in a strain is very important factor to this study. To compare a strain's ranches, pens, and even sampling dates, the strains must contain at least two isolates. This means single isolates that clustered with no other isolate were of little use for this project. Thus, all the strains containing only one isolate were left out of all strain-based analyses. Out of all the strains collected, 50.78% of the ontological strains and 53.82% of the hierarchical strains contained only one isolate, meaning the useable strains for the ontological and hierarchical data analysis were 49.22% and 46.18% respectively (Table 2).

Of the data containing more than one isolate, there were a total of 290 strains produced by hierarchical clustering and 315 strains by ontological clustering. The largest number of isolates per strain was 18 for hierarchical and 13 for ontological clustering.

	Hierarchical	Ontological
Number of Strains	628	640
Number of Single Isolate Strains	338	325
Number of Multi-Isolate Strains	290	315

Table 2: Number of total strains, number of single isolate strains, and number of multi-isolate strains for the two different clustering methods.

Strain Sharing Between Bulls over Time:

Only 16% of strains from hierarchical clustering and 14% of strains from ontological clustering shared isolates across the time points. A large difference is seen between the two clustering types regarding transfer between May 2011 and May 2012 (Figure 7), as ontological clustering produced 20, whereas hierarchical clustering produced only 1. This difference occurs because sample date is an attribute in ontological clustering. Overall, the two clustering methods differ the most across cohorts (ie. May 2011-May 2012, September 2011-May 2012, and September 2011-September 2012) (Figure 7).

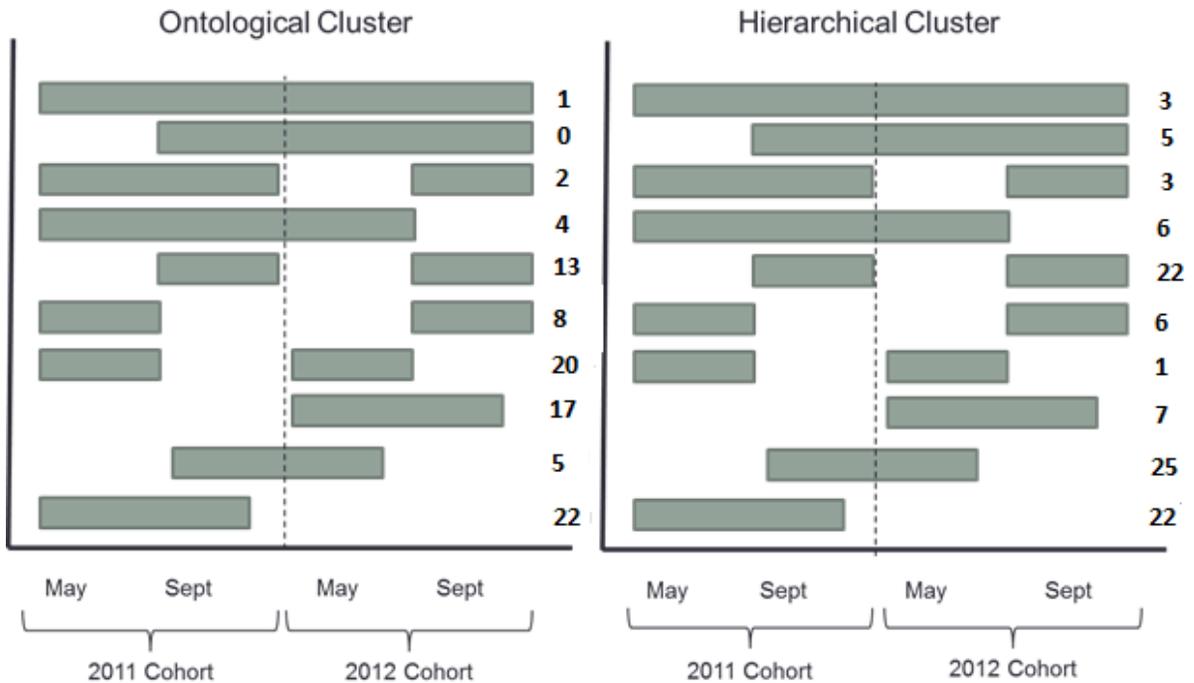


Figure 7: Grouping of *E. coli* strains across the different sampling dates. The grey bar represents the different dates isolates from a strain were found in. The total number of strains found sharing across specific dates is represented numerically on the side of the chart. Ontological clustering method is the chart on the left, while hierarchical clustering method is the chart on the right. The dotted line indicates the cutoff between cohorts.

Geographical Origins of Strains:

To gather more information on strain origin, three of the strains found in May 2012 using ontological cluster (Strains 7654, 7657, and 7661) and hierarchical clustering (Strains 188, 69, and 74) method were taken and the ranch of origin of all contained isolates was determined. We wanted to see if ranches house unique strains of *E. coli* resulting in bulls having strains specific to the ranch in which they originated. However, only one of the strains sampled (strain 69) was grouped by origin, whereas the others contained isolates from different geographical locations across California (Figure 8). Between the hierarchical and ontological clustering there were only three isolates that

overlapped. These three isolates were found in large strains with many isolates in both hierarchical and ontological methods. These data show that it is possible for ubiquitous strains of *E. coli* to exist.

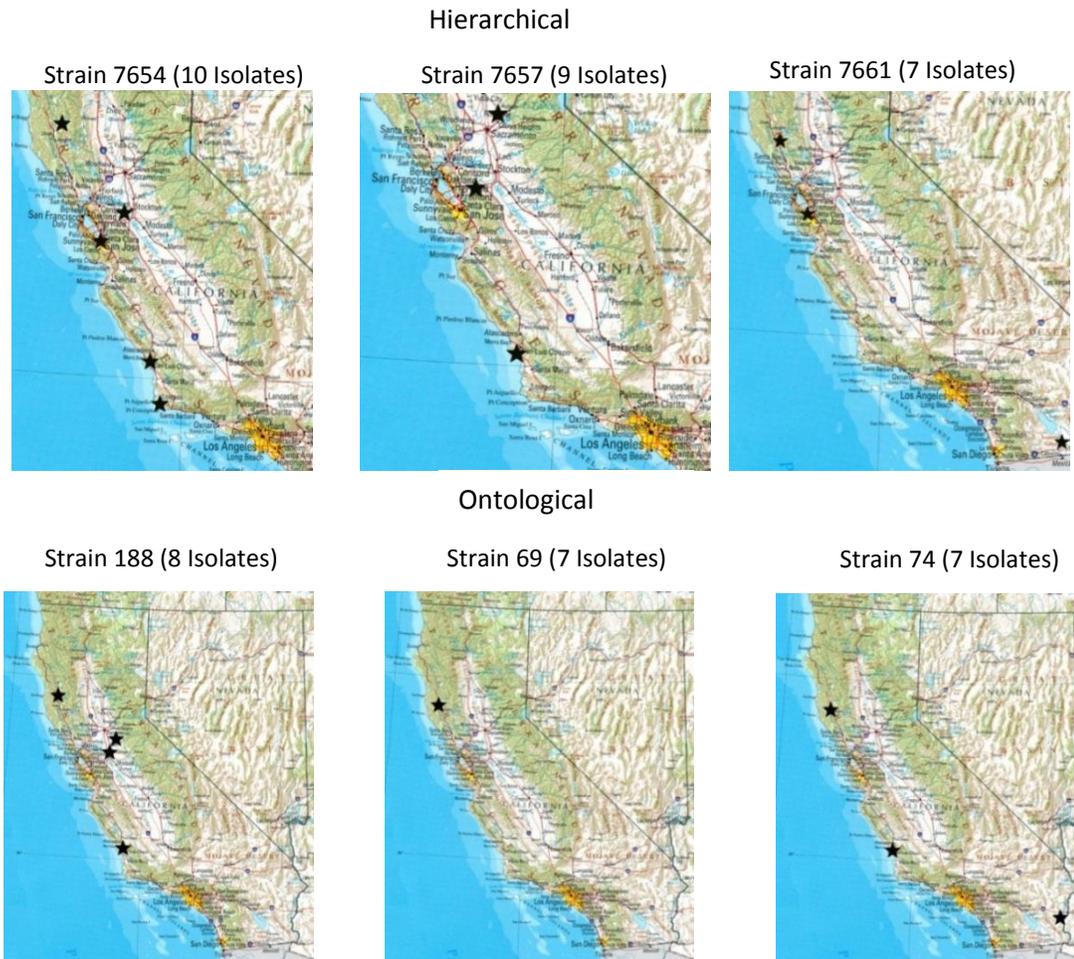


Figure 8: Geographic origin of isolates within a strain. Geographical maps of California with a black star indicating the location of the ranches of origin found in that strain. The three maps on top are from strains produced with hierarchical clustering, whereas the three maps on the bottom are strains produced with ontological clustering.

Strain Sharing Between Bulls from Different Ranches:

The next step was to test the extent of strain sharing across ranches. It was expected that some sharing would be occurring between bulls from the same ranch at the start of the experiment due to these bulls cohabitating during their time at that ranch. Therefore, most the strains found in May samplings should contain isolates found in bulls from only a single ranch. Conversely, in September there should be a greater spread of isolates across ranches due to the strain sharing that occurred over the bull's time of cohabitation. The number of strains found within a bull from any given ranch will indicate the overarching trend of strain sharing in ranches (Figure 9).

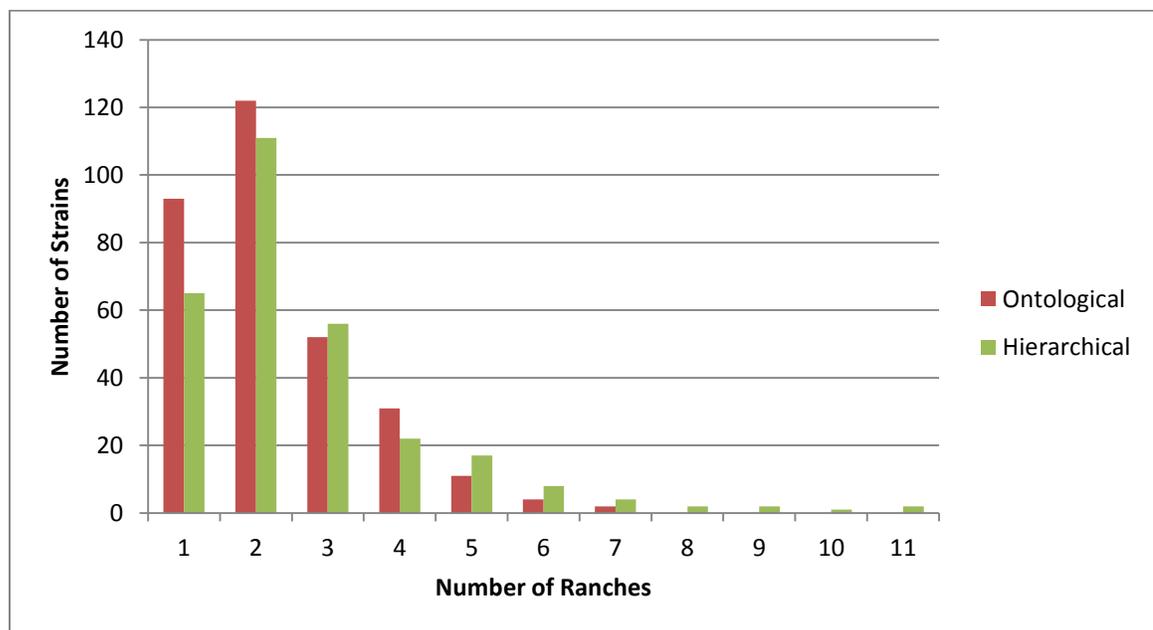


Figure 9: The number of different ranches in which isolates from the same strain were detected overall for 2011 and 2012, May and September.

Most strains of *E. coli* were detected in cattle from one or two different ranches (Figure 9). The largest number of different ranches was 11 using ontological clustering and 7 using hierarchical clustering (Figure 9). To determine the effects of cohabitation,

the data above was broken down into the separate sampling years and the two sampling points within each year (When the bulls arrived in May and when they left in September).

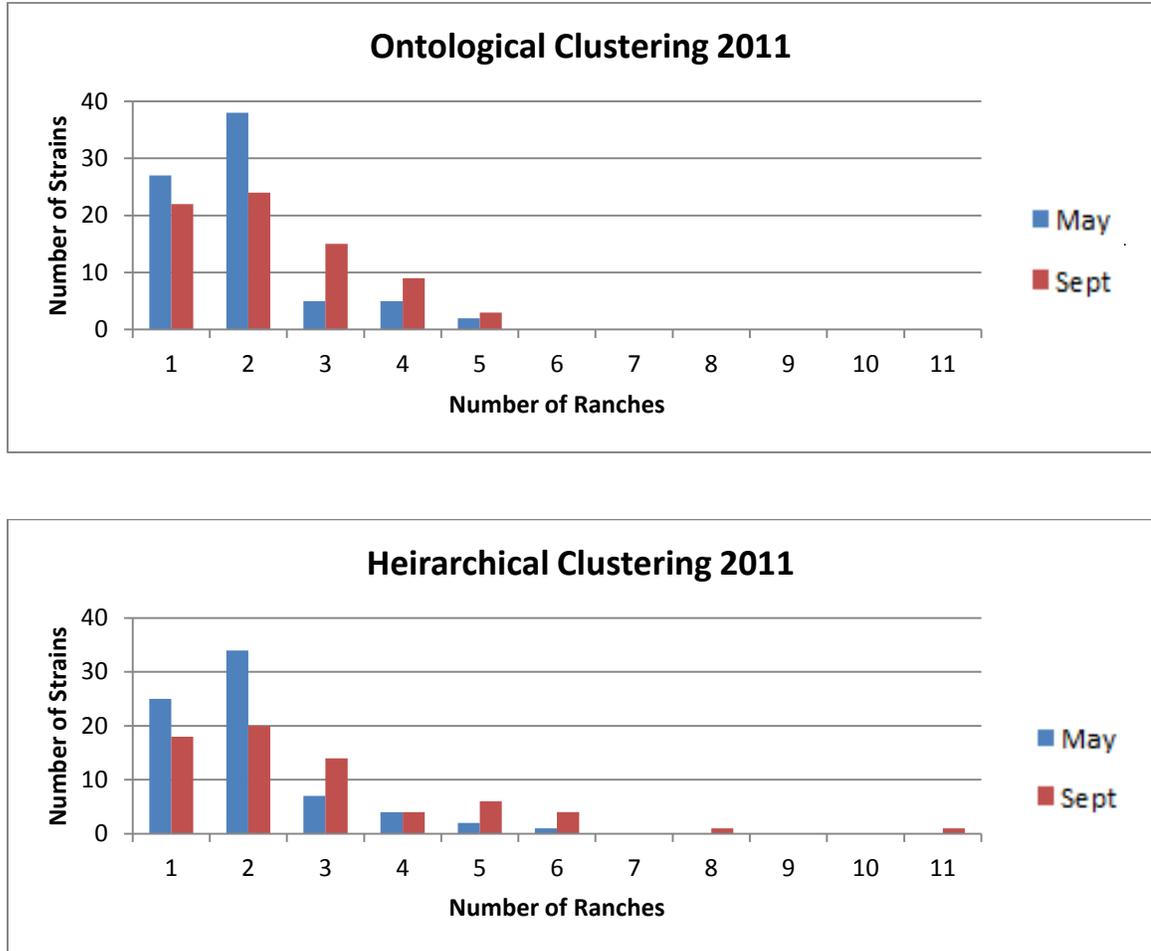


Figure 10: The number of different ranches in which isolates from the same strain were detected (2011).

With both clustering methods the distribution of isolates within a strain in September were spread across more ranches than in May (Figure 10). This trend appears in both the ontological and hierarchical clustering methods. The May sampling date for 2011 was on May 28th, 2011, and at this time the bulls had already been cohabitating for 2 weeks. Therefore, it is possible that strain sharing already occurred during those 2 weeks. This flaw in sampling is one possible explanation for the fact that no significant

differences were found between the distributions of strains in May and September for either ontological ($P= 0.951$) or hierarchical ($P= 0.664$) clustering methods in 2011.

In 2012 the bulls were initially sampled on May 12th, 2012, the date the bulls first arrived, so there was no cohabitation between bulls from different ranches before this date (Figures 11A&B). Since the 2012 samples were initially sampled the moment they entered their cohabitation pens, no sharing between bulls at the Cal Poly Bull Test had occurred yet. Because of this, there is less strains with isolates from different ranches during the May sampling. Even though the visual trend seems more apparent in 2012 compared to 2011, the spread across ranches for both ontological ($p=0.819$) and hierarchical ($p=0.637$) methods between these two dates were not significant.

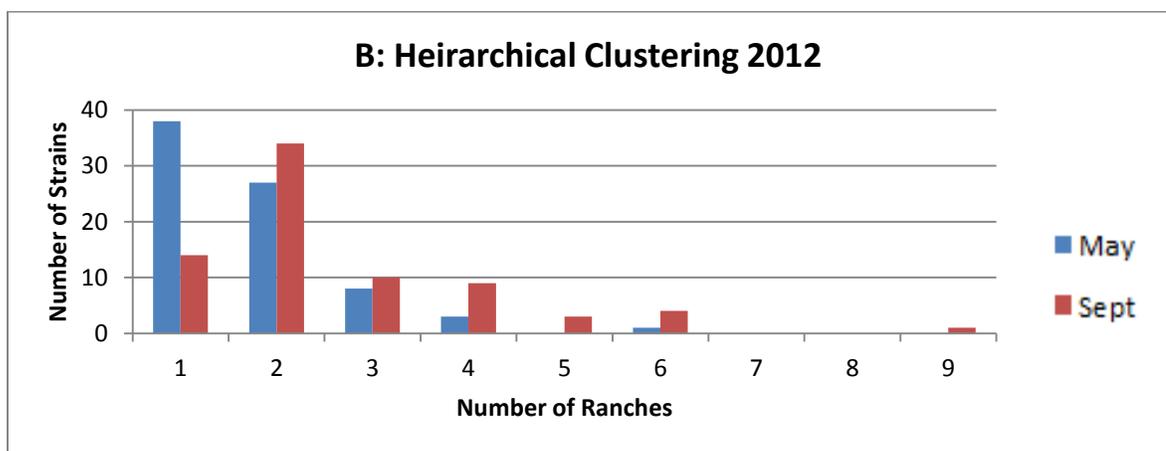
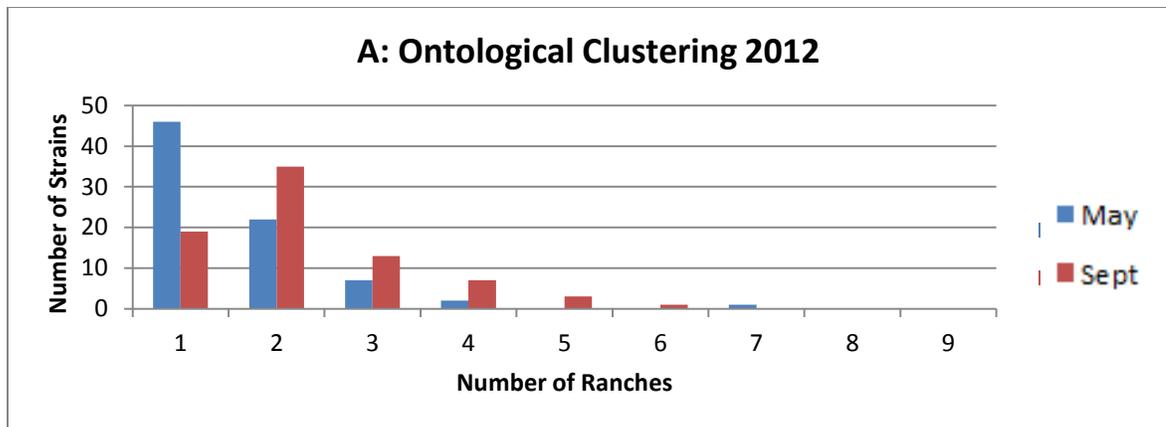


Figure 11: The number of different ranches in which isolates from the same strain were detected (2012).

Strain Makeup and Host Distribution across Ranches:

Careful analysis revealed many of the strains with isolates found only in a single ranch contained two isolates per strain, whereas the mixed ranch strains contained more isolates. For example three strains contained isolates from five ranches in September 2012 and four strains were shared across six ranches for the same time point (Figure 11B). Thus, distribution looks better if the isolates are viewed. The trend in sharing can be visualized when looking at the strains broken into their corresponding isolates (Figure 12). This view emphasizes the effects of isolate count on distribution across ranches. In May there are 46 strains each containing isolates from only in one ranch, while in

September there are 19. Yet, when the isolates contained in those strains are evaluated, there are over 120 isolates from one ranch in May and only 40 isolates in September. Also, if you take a strain with isolates found across 3 separate ranches you can see a difference of 6 strains, but a difference of 40 isolates, indicating that by comparing isolate count instead of strain count, a larger difference could be detected. This difference in number did decrease the p value (ontological $p=0.463$, hierarchical $p=0.073$), but not by enough to be significant. Since these trend values were not significant, a new approach was used that calculates the probability of sharing in order to determine transfer.

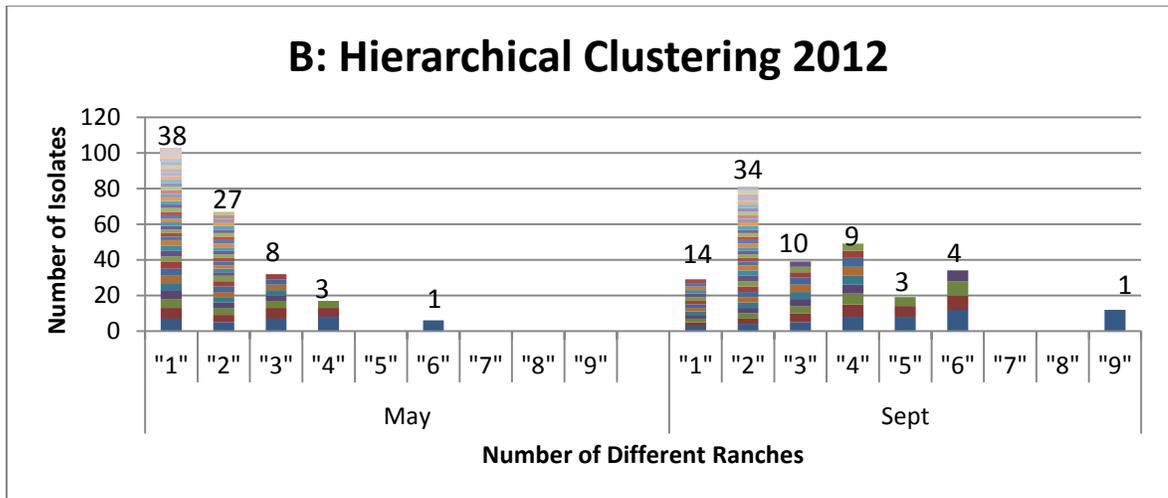
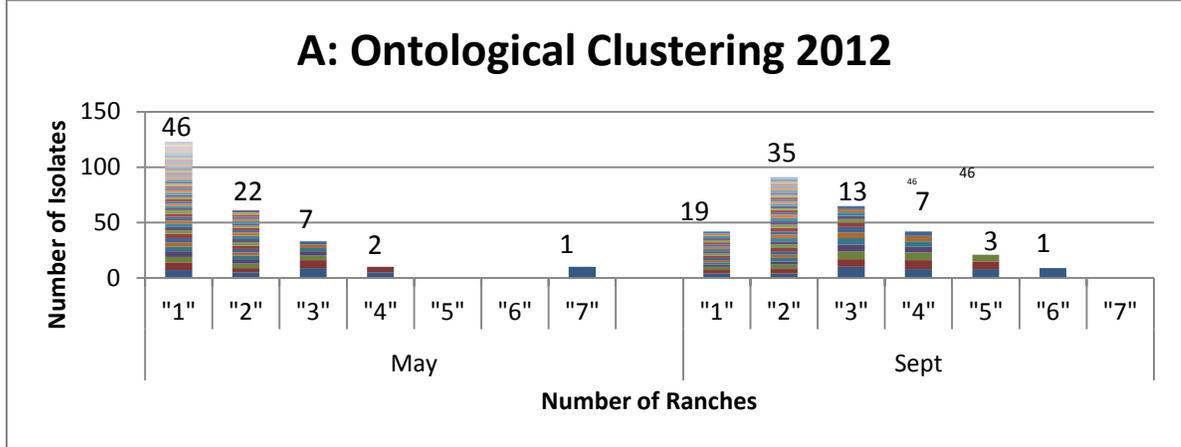


Figure 12: The number of different ranches in which isolates from the same strain were detected for the sampling year 2012 broken down by strain. Each color in the bar graph represents a different strain and the isolate size of that strain in the overall distribution. The black numbers on top of the bars indicate the total number of strain in that bar.

Probability of E. coli Sharing:

As shown previously, even though distribution trends can be seen across ranches, the p values indicated that there was no significant difference between May and September samples. By using a different method to determine the probability of two isolates matching (or being clustered), the sharing of *E. coli* can be addressed. A

significant difference can be seen between May and September samples by looking at the probability two isolates will match instead of directly using number of strains. The probability of sharing was determined by taking the total number of matches divided by the number of possible matches. This gives the probability that any two isolates represent the same strain. This method results in a larger sample size and increase the ability to determine significance. For example, there are 46 single ranch strains in May compared to 19 single ranch strains in September. However, using this new method you would be comparing a much larger sample size, 292 possible matching pairs of isolates in May and 508 possible matching pairs in September. The three possible ways of determining a sharing probability are by looking at pair matches in hierarchical clustering, ontological clustering, and direct isolate matching. Even though the overall possible number of pairs (274,170) is the same for all methods, when comparing the total number of matches, direct isolate match gives the largest number of matching pairs at 3072, followed by hierarchical (945) and ontological (919).

There is no significant difference between the probability of sharing for the 2011 data and the 2012 data (chi 1.35, p= 0.246) (Table 3). Since the 2012 bulls were collected on entry whereas the 2011 bulls were not, the 2012 data is more accurate. Due to the fact that 2012 was collected on entry and that there was no difference between the 2011 and 2012 data, the 2012 data only was used in further sharing probability analysis.

Hierarchical Clustering	2011	2012
All Year	0.41%	0.34%
May	0.50%	0.45%
September	0.86%	0.71%
May-September	0.15%	0.11%

Table 3: Probability of sharing overall using hierarchical clustering for the 2011 and 2012 sampling years.

Probability of Sharing Increases after 4 Months of Cohabitation:

The overall probability of two *E. coli* isolates coming from the same strain is 1.5 to 2 times higher in September than in May (Table 4). This difference was significant for Hierarchical, Ontological, and Matching (chi squared: 58.32, 22.5, 488.06 respectively, with a p value of <0.0001 for all three methods). There was also a much lower, but present matching across the May and September time points, which correlates with the results shown previously (Figure 7).

Sampling Time	Hierarchical	Ontological	Matching
All Year	0.34%	0.34%	1.12%
May	0.45%	0.52%	0.98%
September	0.71%	0.67%	2.40%
May-September	0.11%	0.07%	0.52%

Table 4: Probability of sharing overall for Hierarchical, Ontological, and Matching data for the 2012 sampling year.

Looking at just isolates from the same bull resulted in a much higher probability of sharing with a probability of 22.78% in May and 23.78% in September (Table 5).

There was found to be no significant difference between May and September samples (p=

0.744, 0.925, and 0.847 for Hierarchical, Ontological, and Matching methods respectively) and almost no sharing between May to September in the two isolates taken from the same bull. In different bulls there is 1.3 fold increase (0.44% to 0.59% using ontological clustering) to a 2.4 fold increase (0.90% to 2.33% using direct matching) in sharing from the May to September ($p < .0001$). Even though the probability of sharing changes in different bulls from May to September, the probability doesn't change across two isolates taken from the same bull.

	Same Bull			Different Bull		
	Hierarchical	Ontological	Matching	Hierarchical	Ontological	Matching
All Year	8.47%	11.07%	11.07%	0.31%	0.29%	1.08%
May	22.78%	32.22%	30.56%	0.39%	0.44%	0.90%
September	23.78%	30.81%	28.65%	0.65%	0.59%	2.33%
May-September	0.30%	0.00%	1.04%	0.10%	0.07%	0.52%

Table 5: The probability of sharing for same bull and different bull for the 2012 sampling year. This table shows probability of sharing between isolates from the same bull in the left three columns and different bulls in the right three columns. Hierarchical, Ontological, and Matching clustering methods were used.

Probability of Sharing Across Ranches:

The probability of sharing between bulls with the same ranch of origin was greater than between bulls with different ranches of origin at the start of the experiment (Table 6). While the matching probabilities across different ranches increase significantly ($p < 0.0001$), over two fold between May and September (the largest being direct matching change from 1.03% to 2.30%), the probability of sharing decreases nearly four times ($p < 0.0001$) from May to September in different bulls from the same

ranch of origin (with the largest being 3.53% to 0.93%) (Table 6). This shows that the probability of sharing between bulls with the same ranch of origin decreased after four months of cohabitation. By September the probability of sharing for same ranch of origin (0.64%) and different ranch of origin (0.93%) is much closer than in May.

Sample Time	Different Ranch			Same Ranch Different Bull		
	Hierarchical	Ontological	Matching	Hierarchical	Ontological	Matching
All Year	0.27%	0.23%	1.18%	1.21%	1.51%	2.27%
May	0.22%	0.25%	1.03%	3.53%	4.00%	4.38%
September	0.64%	0.53%	2.30%	0.93%	1.78%	2.83%
May-September	0.10%	0.07%	0.55%	0.20%	0.14%	0.94%

Table 6: The Probability of sharing for different ranches and for the same ranch different bull for the 2012 sampling year. The left three columns show probabilities for matching between different Ranches. The right three columns show probabilities for matching between different bulls from the same ranch of origin. The data shows Hierarchical, Ontological, and Matching methods.

Strain Sharing Between Cohabiting Bulls:

All of the bulls in this study were housed in one of seven different pens located in the Cal Poly Bull Unit (Figure 13). Prediction IV addresses whether bulls will share *E. coli* strains more readily with bulls found in the same pen and nearby pens rather than distant pens. This would seem logical since some pens are physically touching, making direct-contact spread more likely. A “nearby pen” was defined as the pen the bull resided in or any pen touching the bull’s pen. A “distant pen” was defined as any pen the bull was not part of or next to.

Using the pen map (Figure 13), the number of strains found containing multiple isolates from the same pen or “nearby pens” were compared against the number of strains with isolate sharing between bulls in “distant pens”. The bulls were not randomly distributed into the pens, but placed there in order of arrival. Thus, we would expect bulls to have more sharing between same or neighboring pens since this is where their ranch mates are.

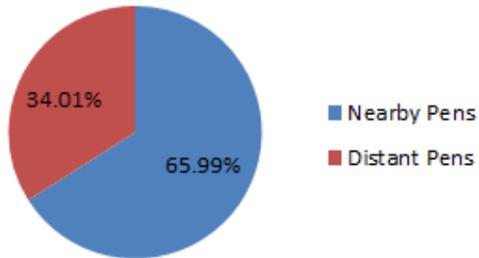


Figure 13: Photograph of the Cal Poly Bull Test Pens. The white numbers indicating pen numbers are shown over the geographic location of each pen on the map.

A higher percentage of strain sharing was observed between bulls in nearby pens compared to those residing in distant pens across all sampling times (Figure 14A-D). This trend held for both hierarchical and ontological clustering methods. However, the predicted difference between September and May was only found in the ontological clustering method (Figure 14).

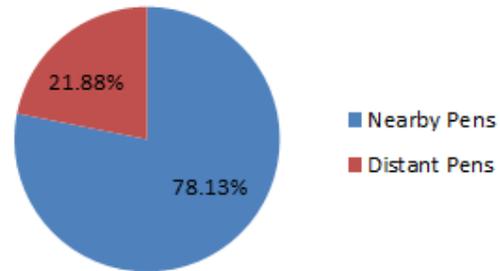
A: Ontological May 2012

Total Number of Isolates: 362



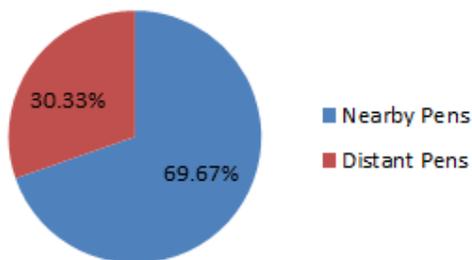
B: Ontological Sept 2012

Total Number of Isolates: 379



C: Hierarchical May 2012

Total Number of Isolates: 362



D: Hierarchical Sept 2012

Total Number of Isolates: 379

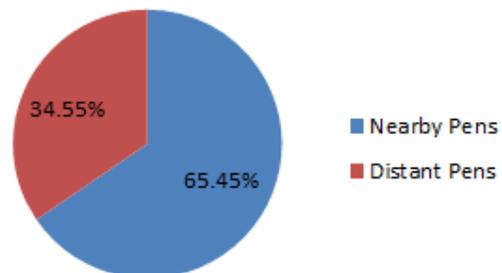


Figure 14: Percentage of sharing comparing nearby pens to distant pens. A&C (left pie graphs) show the percentile difference between May 2012 strains that share isolates with nearby pens (blue color) compared to strains that share isolates with distant pens (red color). B&D (right pie graph) show the same difference for the September 2012 time point.

The probability of sharing *E. coli* between bulls housed in the same pens from the same ranch shows a twofold decrease in sharing probability from May to September (Table 7). Conversely there was a fourfold (0.42% to 1.97%) to fivefold (0.49% to 2.42%) increase in strain sharing probability (depending on the method) from May to September in bulls from different ranches housed in the same pen ($p = <0.0001$) (Table

7). Overall, as you put the bulls into situations where they might more easily transfer a new stain, the probability of sharing increases.

Sampling Time	Same Pen Same Ranch			Same Pen Different Ranch		
	Hierarchical	Ontological	Matching	Hierarchical	Ontological	Matching
All Year	2.01%	2.67%	2.11%	0.62%	0.71%	1.66%
May	5.48%	6.69%	4.44%	0.42%	0.49%	1.02%
September	2.06%	3.17%	2.12%	1.97%	2.42%	4.63%

Table 7: The probability of sharing for same pen same ranch and same pen different ranch for the 2012 sampling year. The left three columns show probabilities for matching between bulls from the same ranch of origin and housed in the same pen. The right three columns show probabilities for matching between bulls from the different ranch of origin and housed in the same pens. All data is shown for Hierarchical, Ontological, and Direct Matching methods.

Squirrel Cohabitation Effects on E. coli Sharing:

Next, this thesis looked at the ability of ground squirrels and bulls to house the same *E.coli*. Squirrels roam around the penning area of the Cal Poly Bull Unit and may be participating in the exchange of *E.coli*, as squirrels could pick up bull strains or squirrels to transfer their existing strains to bulls. A total of 49 squirrel isolates were collected from around the bull pens; 5 to 10 *E. coli* isolates from each squirrel depending on who sampled them. Since squirrels have no ranch of origin or pen, only hierarchical clustering was used to build strains. The squirrel samples were taken when an animal was caught, which was spread out over the 2011 testing year. These samples were then compared to existing bull samples in the CPLOP database to determine if any strains were made up of *E. coli* isolates from both bulls and squirrels. Only six different strains

were identified with isolates from both bulls and squirrels (Figure 18). Of these six different strains, there were only 7 squirrel *E. coli* isolate, or 14% of the squirrel isolates collected out of the six shared strains, five of them contained only one squirrel isolate to multiple bull isolates, with the exception of Strain 7654, which was a 2:3 squirrel to bull ratio. Strain 7653 was the only strain that contained a squirrel isolate that matched a bull isolate from 2012, all the other combined strains contained bull isolates from 2011.

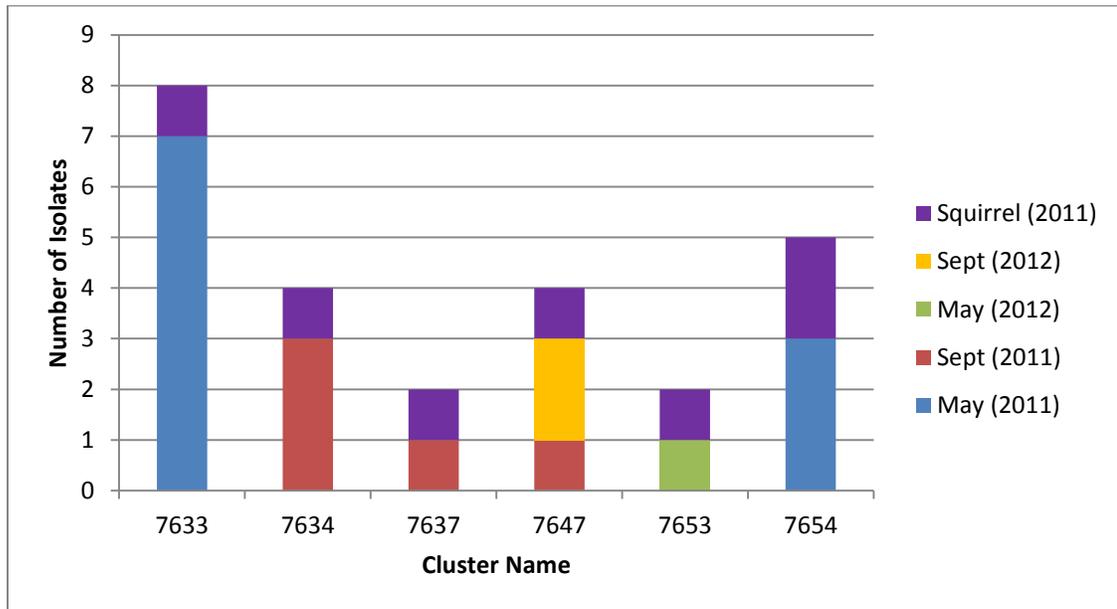


Figure 15: Strains with isolates from both squirrels and bulls. Bull isolates are further identified by their date of their collection.

IV. Discussion:

Throughout this paper the transmission of *E.coli* in bulls was evaluated. The probability of sharing *E. coli* was used to imply transmission. Since the only thing that changed for the bulls over the four months was the presence of other bulls being housed in close proximity and the new environment, one of these factors must induce transfer.

The overall hypothesis of this paper was that *E. coli* transfer occurs between cohabitating bulls. Multiple testing procedures were done to test the predictions. Before any of the predictions could be tested, it had to be determined that strains were being shared across May to September time points, because without this sharing none of the other predictions could be supported. Sharing across multiple time points was a common occurrence in strains containing more than 1 isolate (Figure 6). There was, however, a large dissimilarity between hierarchical and ontological clustering. Ontological clustering found no shared strains between September 2011, May 2012, and September 2012, whereas hierarchical clustering found 5 strains shared across said dates. This can be explained by the fact that “Host ID” was the first attribute used and “sampling date” was the last attribute used and in the ontological clustering method; therefore it is likely isolates from those five strains were clustered before reaching “sample date” (ie: Host ID>Ranch>Pen>Sample Date) in ontological clustering.

Overall, as expected, the greatest amount of strain sharing across dates occurred between the May and September time points within the same cohort. However, there was also a very large amount of strain sharing for both May 2011 and 2012 time points. This could be due to the fact that in May the bulls were first arriving from their separate

ranches, and the ranches were the same across both years of testing. It is possible that similar ranch-specific strains were brought to Cal Poly from those ranches during both years of testing. However, this cannot be definitively concluded from the data found in this study due to the small sample size of two isolates per bull. Using the current experimental methods, a maximum of two *E.coli* strains from an individual bull could be detected at any sample date. However, other related papers state that ruminants, cattle in particular, house a large diversity of *E. coli* strains and many more isolates would be required to determine full *E. coli* diversity within cattle (Anderson, Whitlock, & Harwood, 2006). Because of this, many *E. coli* strains in each bull tested were not represented. Similarly, our study of 30 isolates from one bull (Figure 6) detected 13 different strains present. Furthermore, in this study a single strain made up over 25% of the isolates, suggesting that the small samples would likely only contain the most abundant strains.

In a related project completed by another student in the lab, 100 to 150 isolates were taken from each of two cows. These isolates were then pyroprinted and clustered into strains to determine the number of isolates found in bovine fecal samples that would be required to see the full strain diversity inhabiting the animal. When an ACE richness estimator was completed, it was determined that over 400 isolates would have to be taken to sample all the strain diversity in these cows (Stivers, 2015). This shows that even the 100-150 isolates taken from each host animal during the study were not enough to represent the strain diversity. The need for such a high number of isolates suggests the reason for lack of significance in the trend comparison results of this thesis. The results obtained using only two isolates per bull simply do not adequately take into account the

strain diversity of the sampled bulls. Therefore, this study may have missed a great deal of strain sharing.

Prediction I stated that the probability of sharing will increase with cohabitation (from May to September). Our results supported this prediction. There was an increase in the probability of sharing, for all clustering types, from May to September (Table 4). This increase in sharing can be used to indicate that *E. coli* strain transfer occurred during the four months of cohabitation. However, the cohabitating bulls were not the only difference between the months. Alternatively this sharing could be due to all the bulls eating the same (possibly new) type of food, or the fact that all the bulls are now in the same environment. Although the increase in probability of sharing does indicate transfer is occurring, it is impossible to know where the strains are being taken from (bull, environment, or food source).

The four isolates from the same bull were compared to see if the probability of sharing changed from May to September. There was an increase from May and September of less than 1% for the isolates from the same bull. When chi squared was used the difference was found to be not significant. This suggests that, even though transfer has occurred, the probability of the isolates from one bull being the same strain didn't change across time points. There were very few times that isolates were found to be conserved within a bull from May to September. (Between 1.04% and 0% depending on the method) (Table 3). Since the probability of sharing stayed the same between the two isolates from the same bull, even though the strains differ across time points, this suggests that most the isolates sampled from a bull in September were either from a

different source (another bull or the environment) or were present but not sampled in May due to the small sample size.

The second prediction was that the probability of sharing between bulls with the same ranch of origin would be greater than bulls with different ranches upon their initial arrival. This was thought to be likely since the *E. coli* would be constantly shared at their ranch of origin. Although there were ubiquitous strains containing isolates from various ranches (Figure 8), these strains were few in number and generally included a large number of isolates. Despite the presence of these ubiquitous strains, the probability of sharing was greater for bulls from the same ranch in May than in September. As predicted the probability of sharing was greater in bulls from the same ranch in May than in bulls of different ranches in May (Table 6).

The third prediction states that the probability of sharing between bulls of the same ranch of origin will decrease after four months of cohabitation with bulls from different ranches of origin. When the probability of sharing was compared for bulls from the same ranch, a decrease of two to four fold ($p < .0001$) in sharing from May to September was found. On the other hand, there was two fold increase in probability of sharing for bulls from the different ranch from May to September depending on the method used ($p < .0001$) (Table 6). This decrease is due to the dilution of strains within those bulls over time. Many bulls from the same ranch contain similar strains, but when these bulls share their strains with bulls from different ranches a larger number of different strains are now present, which dilutes the overall strain makeup within those bulls. This dilution effect decreases the probability of sharing in bulls from the same ranch due to the larger variety of *E.coli* strains now within the bull. Yet, even after the

two to four fold decrease in the bulls from the same ranch, the bulls from different ranches still had a 0.5-2 times lower probability of sharing after four months of cohabitation (Table 6). However, since sharing probability increased in different ranches and decreased in the same ranch from May to September, sharing between all bulls regardless of their ranch of origin is occurring.

The final prediction tested was that bulls in closer proximity are more likely to share strains than are bulls at distant proximities. The number of shared strains between bulls of neighboring pens and bulls in distant pens was compared. Due to the relationship between proximity of bulls and frequency of transmission, the nearby pen bulls shared more strains with one another than with those in distant pens ($P = .023$) (Figure 14). This outcome may have also been affected due to the bull arrival order. Since bulls were penned as they arrived, most bulls from the same ranch were penned together. However, to test if the sharing was being skewed by the high number of bulls from the same ranch in each pen, only bulls from different ranches within the same pen were tested for sharing. There was an increase in sharing across bulls from different ranches housed in that same pen from May to September. Bulls housed in the same pen from the same original ranch had their probability of sharing reduced by half in September when compared to May (Table 7). It was also found that bulls housed in the same pen, but from different ranches had an increase of four fold in probability of sharing in September when compared to May. The four fold increase confirms prediction IV, that proximity plays a major role in *E.coli* strain transfer.

The secondary hypothesis states that it is possible for bulls and squirrels to share strains of *E.coli*. After comparing the isolates collected from squirrels with the bull

isolates, only 7 out of 49 isolates were clustered into strains with a combination of squirrel and bull isolates (Figure 15). A 1.79% probability of a squirrel sharing a strain with a bull was found however, 49 isolates may not be enough to accurately determine sharing probabilities. Despite the small number of strains containing both bull and squirrel isolates, this still confirms the ability of squirrels to house and possibly even transfer the same strain of *E. coli* that are found in bulls. This is important because, unlike bulls, squirrels only house a small variety of strains (Hassan et al., 2007). This means it may be possible for a squirrel to pick up a strain of *E. coli* and take that strain off site (as squirrels movement is not inhibited by fences) possibly transmitting it to another nearby bull population. If the strain was pathogenic, then the new population of bulls could become carriers of pathogenic *E. coli* strains without direct contact with an infected bull.

Throughout this study, two different methods of clustering isolates into strains and one direct isolate matching method were evaluated. The direct matching method generates the most matching pairs. This is due to the fact that it considers any two isolates that match above 99% to be from the same strain. To determine the usefulness of direct matching, a few things must be considered. Direct matching is by far the most relaxed method when it considers a strain (match). Direct matching results in the largest number of matches due to the fact that any match above 99.0% is a match however, if determining genetic similarity in a strain is important then direct matching may not be the best method. This is because in some instances a problem can arise when considering three isolates. Isolate 1 matches isolate 2 and isolate 2 matches isolate 3 however; isolate 1 does not match isolate 3. There may be two different strains represented here, yet it is

difficult to determine which isolate belongs to which strain. Clustering methods can be used to resolve this problem. One such method is hierarchical clustering, which clusters isolates into strains based solely on their similarity. Hierarchical clustering clusters isolates based on the best match. Therefore, if there are three isolates the best two will be clustered into a strain first and the third will then have to match the average of that strain. Due to this hierarchical clustering is stricter when clustering isolates than direct matching however, it is not as strict as ontological clustering. Ontological clustering, on the other hand, group isolates into clusters based on attributes we determined would be most likely shared in a strain. In the majority of the ranch and pen comparisons, both the ontological and hierarchical clustering methods show similar trends and patterns. The only major exception noted was the data showing pen sharing between neighboring pens and distant pens. Ontological clustering is more likely to cluster isolates from the same ranch or pen into the same strain, therefore this method would work best when comparing across multiple pen or ranch groupings. Strains containing isolates from multiple ranches or pens would be less likely to be clustered together due to the fact that they are outside of the grouping order. Therefore, in order for strains containing isolates from multiple ranches or pens to exist they must have a very similar pyroprints. For example, if an isolate from a different bull and a different ranch is to match, it would have to match the average of all the isolates already clustered from the same bull and same ranch. This allows for a very strict strain definition when comparing across attributes. Overall, each method has different strengths and weaknesses, and the method used should be determined individually based on the study.

With the data acquired, it can be said that though direct matching yields higher values, that does not denote accuracy in strain typing. Both Hierarchical and Ontological clustering methods are useful when strain typing due to their harsher strain requirements than direct matching. However, the method of clustering that should be used depends on what is being compared. Ontological clustering will increase the number of strains clustered from the same pen, ranch, or bull due to the bias toward clustering isolates with the same attributes, whereas hierarchical clustering will cluster solely based on isolate similarity. If an experiment has factors believed to correlate at a higher genetic similarity (ie: same bull, location, or pen), then ontological clustering would be optimal however, if attributes for ontologies exist then hierarchical clustering would be a useful method.

In summary, due to the small sample size per bull, it is impossible for any true statistical difference to be determined when comparing trends. Yet, through the use of probability of sharing it can be implied that transfer occurs with cohabitation. For the different experiments all three matching methods supported our hypothesis. Therefore, we suggest any further research on this subject should use all three methods. It is not clear if the bulls are sharing each other's *E. coli* or if the bulls are sharing the same *E. coli* they picked up from the environment deposited from some other source. To determine this, a more in-depth study investigating full *E. coli* strain diversity within each bull's digestive tract would have to be done. It would require a budget and manpower far greater than this lab possesses to investigate transfer using full strain diversity.

REFERENCES

- Anderson, M. A., Whitlock, J. E., & Harwood, V. J. (2006). Diversity and Distribution of *Escherichia coli* Genotypes and Antibiotic Resistance Phenotypes in Feces of Humans, Cattle, and Horses. *Applied and Environmental Microbiology*, 72(11), 6914–6922. <http://doi.org/10.1128/AEM.01029-06>
- Ayaz, N. D., Gencay, Y. E., & Erol, I. (2014). Prevalence and molecular characterization of sorbitol fermenting and non-fermenting *Escherichia coli* O157:H7+/H7– isolated from cattle at slaughterhouse and slaughterhouse wastewater. *International Journal of Food Microbiology*, 174, 31–38. <http://doi.org/10.1016/j.ijfoodmicro.2014.01.002>
- Black, M. W., VanderKelen, J., Montana, A., Dekhtyar, A., Neal, E., Goodman, A., & Kitts, C. L. (2014). Pyroprinting: a Rapid and Flexible Genotypic Fingerprinting Method for Typing Bacterial Strains.
- Center for Disease Control and Prevention. (2015, April 1). Retrieved April 1, 2015, from <http://www.cdc.gov/ecoli/2014-outbreaks.html>
- Clermont, O., Olier, M., Hoede, C., Diancourt, L., Brisse, S., Keroudean, M., ... Denamur, E. (2011). Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. *Infection, Genetics and Evolution*, 11(3), 654–662. <http://doi.org/10.1016/j.meegid.2011.02.005>
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 16(22), 10881–10890. <http://doi.org/10.1093/nar/16.22.10881>

- Gabig, M., Obuchowski, M., Srutkowska, S., & Wegrzyn, G. (1998). Regulation of replication of lambda phage and lambda plasmid DNAs at low temperature. *Molecular & General Genetics: MGG*, 258(5), 494–502.
- Gordon, D. M., & Cowling, A. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology*, 149(12), 3575–3586. <http://doi.org/10.1099/mic.0.26486-0>
- Grimont, F., & Grimont, P. A. D. (1986). Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Annales de l'Institut Pasteur / Microbiologie*, 137(1, Supplement 2), 165–175. [http://doi.org/10.1016/S0769-2609\(86\)80105-3](http://doi.org/10.1016/S0769-2609(86)80105-3)
- Hassan, W. m., Ellender, R. d., & Wang, S. y. (2007). Fidelity of bacterial source tracking: *Escherichia coli* vs *Enterococcus* spp and minimizing assignment of isolates from nonlibrary sources. *Journal of Applied Microbiology*, 102(2), 591–598. <http://doi.org/10.1111/j.1365-2672.2006.03077.x>
- Kristjánsson, M., Samore, M. H., Gerding, D. N., DeGirolami, P. C., Bettin, K. M., Karchmer, A. W., & Arbeit, R. D. (1994). Comparison of restriction endonuclease analysis, ribotyping, and pulsed-field gel electrophoresis for molecular differentiation of *Clostridium difficile* strains. *Journal of Clinical Microbiology*, 32(8), 1963–1969.
- Li, W., Raoult, D., & Fournier, P.-E. (2009). Bacterial strain typing in the genomic era. *FEMS Microbiology Reviews*, 33(5), 892–916. <http://doi.org/10.1111/j.1574-6976.2009.00182.x>

- Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., ...
Spratt, B. G. (1998). Multilocus sequence typing: A portable approach to the
identification of clones within populations of pathogenic microorganisms.
Proceedings of the National Academy of Sciences of the United States of America,
95(6), 3140–3145.
- Montana, A. (2013). *Algorithms for library-based microbial source tracking*. California
Polytechnic State University San Luis Obispo. Retrieved from
<http://digitalcommons.calpoly.edu/theses/959/>
- Nicolle, L. E. (2008). Uncomplicated Urinary Tract Infection in Adults Including
Uncomplicated Pyelonephritis. *Urologic Clinics of North America*, 35(1), 1–12.
<http://doi.org/10.1016/j.ucl.2007.09.004>
- Osman, K. M., Mustafa, A. M., Elhariri, M., & AbdElhamed, G. S. (2013). The
Distribution of Escherichia coli Serovars, Virulence Genes, Gene Association and
Combinations and Virulence Genes Encoding Serotypes in Pathogenic E. coli
Recovered from Diarrhoeic Calves, Sheep and Goat. *Transboundary & Emerging
Diseases*, 60(1), 69–78. <http://doi.org/10.1111/j.1865-1682.2012.01319.x>
- Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P.,
... Ravel, J. (2008). The Pangenome Structure of Escherichia coli: Comparative
Genomic Analysis of E. coli Commensal and Pathogenic Isolates. *Journal of
Bacteriology*, 190(20), 6881–6893. <http://doi.org/10.1128/JB.00619-08>
- Ren, Y., Liu, B., Cheng, J., Liu, F., Feng, L., & Wang, L. (2008). Characterization of
Escherichia coli O3 and O21 O antigen gene clusters and development of

- serogroup-specific PCR assays. *Journal of Microbiological Methods*, 75(2), 329–334. <http://doi.org/10.1016/j.mimet.2008.07.010>
- Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research*, 11(1), 3–11. <http://doi.org/10.1101/gr.150601>
- Sekse, C., Solheim, H., Urdahl, A. m., & Wasteson, Y. (2008). Is lack of susceptible recipients in the intestinal environment the limiting factor for transduction of Shiga toxin-encoding phages? *Journal of Applied Microbiology*, 105(4), 1114–1120. <http://doi.org/10.1111/j.1365-2672.2008.03845.x>
- Shaikh, N., & Tarr, P. I. (2003). Escherichia coli O157:H7 Shiga Toxin-Encoding Bacteriophages: Integrations, Excisions, Truncations, and Evolutionary Implications. *Journal of Bacteriology*, 185(12), 3596–3605. <http://doi.org/10.1128/JB.185.12.3596-3605.2003>
- Stivers, A. T. (2015). A Preliminary Report on Escherichia coli Strain Diversity in Cows. *Senior Project*.
- Tawil, N., Sacher, E., Mandeville, R., & Meunier, M. (2012). Surface plasmon resonance detection of E. coli and methicillin-resistant S. aureus using bacteriophages. *Biosensors and Bioelectronics*, 37(1), 24–29. <http://doi.org/10.1016/j.bios.2012.04.048>
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., ... Denamur, E. (2009). Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths. *PLoS Genet*, 5(1), e1000344. <http://doi.org/10.1371/journal.pgen.1000344>

- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., ... Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23(21), 4407–4414. <http://doi.org/10.1093/nar/23.21.4407>
- Vrese, M. de, & Schrezenmeir, J. (2008). Probiotics, Prebiotics, and Synbiotics. In U. Stahl, U. E. B. Donalies, & E. Nevoigt (Eds.), *Food Biotechnology* (pp. 1–66). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/10_2008_097
- Wang, G., & Doyle, M. P. (1998). Survival of Enterohemorrhagic Escherichia coli O157:H7 in Water. *Journal of Food Protection*, 61(6), 662–667.
- Wells, J. G., Shipman, L. D., Greene, K. D., Sowers, E. G., Green, J. H., Cameron, D. N., ... Ostroff, S. M. (1991). Isolation of Escherichia coli serotype O157:H7 and other Shiga-like-toxin-producing E. coli from dairy cattle. *Journal of Clinical Microbiology*, 29(5), 985–989.
- Yue, W.-F., Du, M., & Zhu, M.-J. (2012). High Temperature in Combination with UV Irradiation Enhances Horizontal Transfer of stx2 Gene from E. coli O157:H7 to Non-Pathogenic E. coli. *PLoS ONE*, 7(2), e31308. <http://doi.org/10.1371/journal.pone.0031308>
- Zimbro, M. J., & Power, D. A. (2009). *Difco & BBL manual: manual of microbiological culture media*. Sparks, MD: Becton Dickinson and Co.

APPENDICES

Appendix A: Pen Breakdown for Both Years

Pen Name (2011)	Cal Poly Black Pen 1	Cal Poly Black Pen 2	Cal Poly Black Pen 3	Cal Poly Brown Pen 1	Cal Poly Brown Pen 2	Cal Poly Pen	Cal Poly Sick Pen
Number of Ranches	9	14	18	16	9	13	5
Pen Name (2012)	Cal Poly Pen 2	Cal Poly Pen 3	Cal Poly Pen 4	Cal Poly Pen 4X	Cal Poly Pen 5	Cal Poly Pen 6	Cal Poly Pen 7
Number of Ranches	10	7	5	1	9	6	9

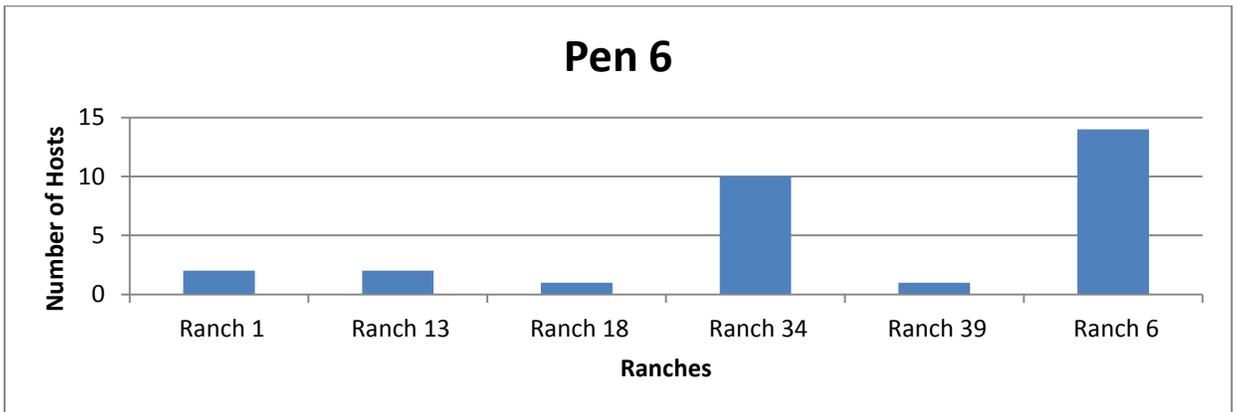
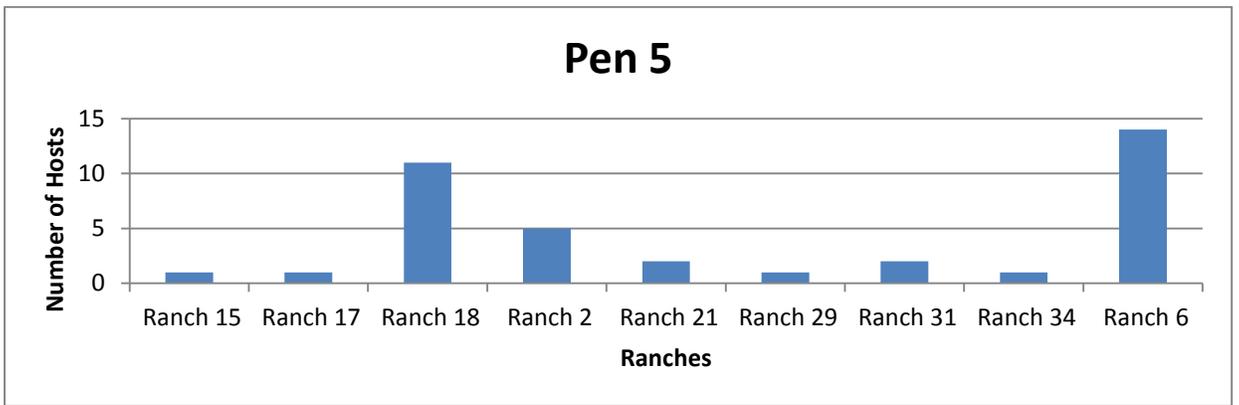
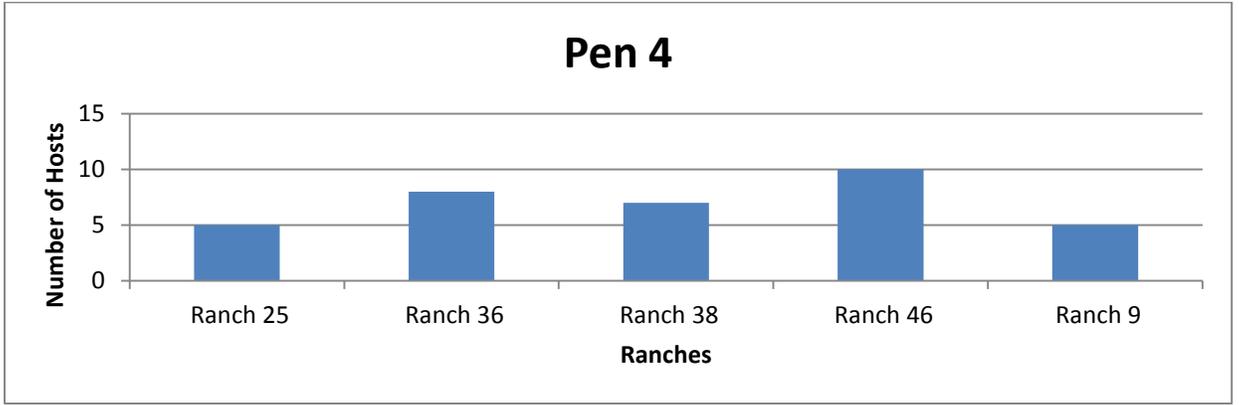
Breakdown of the number of ranches found in each pen overall for both years.

Appendix B: Pen Breakdown for Time Points

Pen (2011)	May	September	Pen (2012)	May	September
Cal Poly Black Pen 1	6	9	Cal Poly Pen 2	10	8
Cal Poly Black Pen 2	10	14	Cal Poly Pen 3	7	7
Cal Poly Black Pen 3	9	15	Cal Poly Pen 4	5	5
Cal Poly Brown Pen 1	6	10	Cal Poly Pen 4X	1	1
Cal Poly Brown Pen 2	3	6	Cal Poly Pen 5	9	6
Cal Poly Pen	8	7	Cal Poly Pen 6	5	5
Cal Poly Sick Pen	3	3	Cal Poly Pen 7	9	9

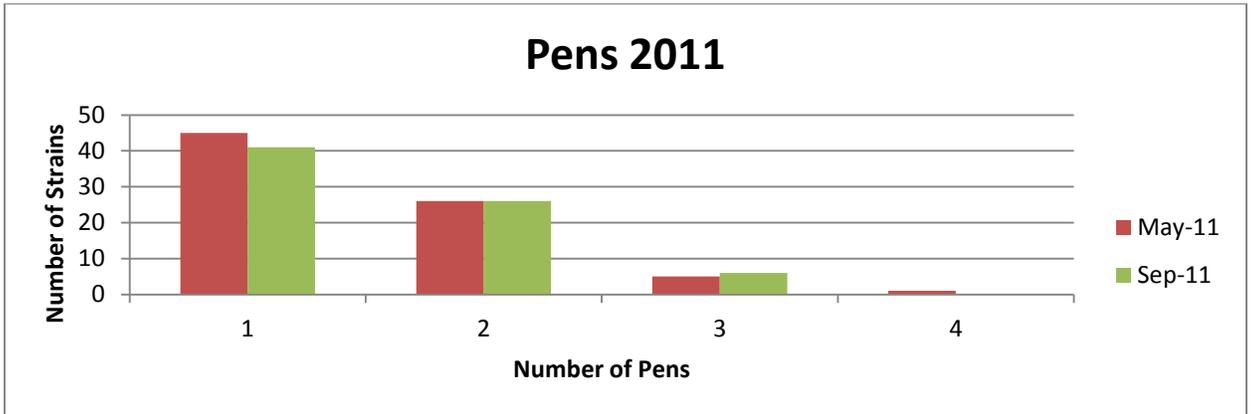
Breakdown of the number of ranches found in each pen shown at both time collection points.

Appendix C: Ranch Distribution



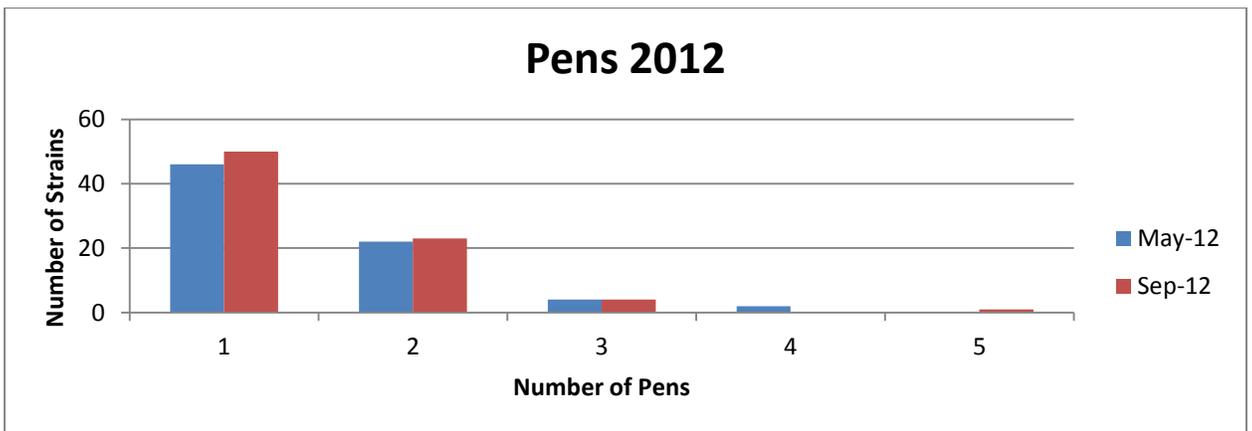
Host distribution across ranches for pens 4, 5, and 6.

Appendix D: Pen Distribution 2011



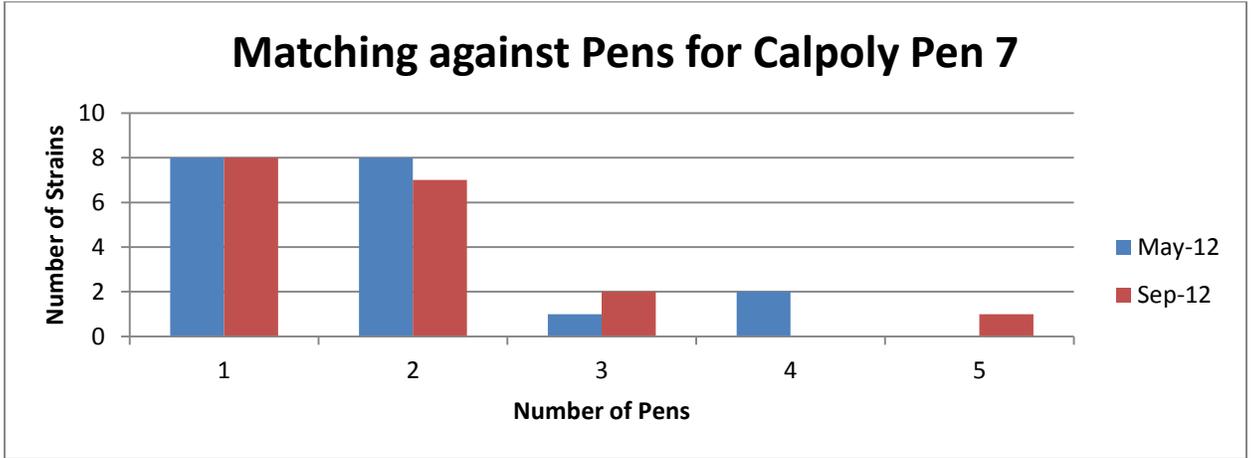
Distribution of strains across different pens for 2011 sampling date. Ontological clustering.

Appendix E: Pen Distribution 2012



Distribution of strains across different pens for 2012 sampling date. Ontological clustering.

Appendix F: Pen 7 Strain Distribution



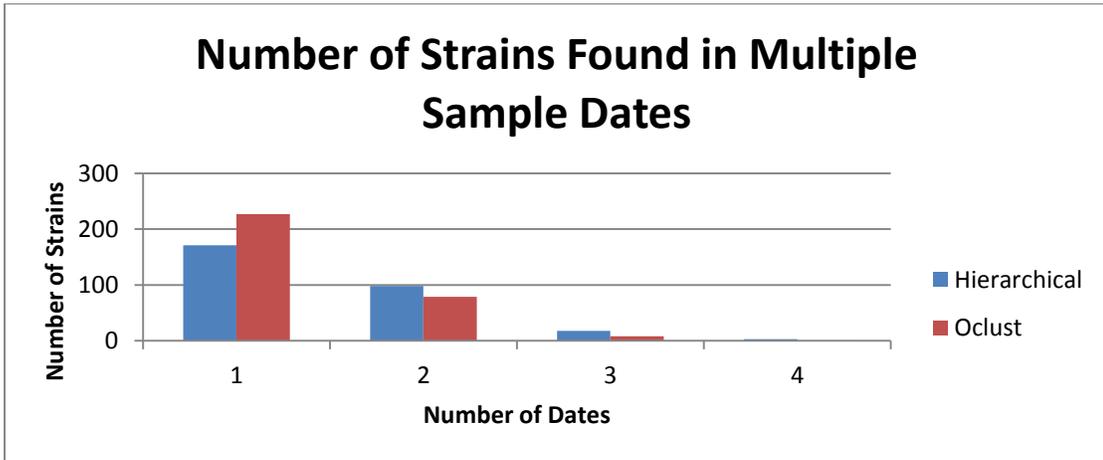
Distribution of strains across different pens for pen7 2012 sampling date for Ontological clustering.

Appendix G: Isolates across Time Points

	Hierarchical	Oclust
1	171	227
2	98	79
3	18	8
4	3	1

The different number of stains with isolates across multiple dates.

Appendix H: Cross Time Point Strain Sharing



Number of strains found across the 4 sampling dates graph.

Appendix I: Probability of Sharing for Different Pens

Sampling Times	Different Pens		
	Hierarchical	Ontological	Matching
All Year	0.23%	0.17%	0.94%
May	0.23%	0.26%	0.72%
September	0.45%	0.28%	2.01%

Probability of sharing for bulls in different pens.

Appendix J: Total Numbers Collected

	May 2011	Sept 2011	May 2012	Sept 2012	Total
	Sample	Sample	Sample	Sample	
Number of Bulls Sampled	189	181	190	179	739
Number of Isolates Produced	362	360	379	333	1434
Number of Pyroprints Run	724	720	758	666	2868
Total number of bulls, isolates, and pyroprints collected.					