CPS 2018 RFP
FINAL PROJECT REPORT

Project Title
Illuminating the role of whole genome sequencing in produce safety

Project Period
January 1, 2019 – December 31, 2020 (extended to March 31, 2021)

Principal Investigator
Kerry Cooper
University of Arizona
School of Animal and Comparative Biomedical Sciences
1177 E. 4th St.
Tucson, AZ 85721-0090
T: 520-621-3342
E: kcooper@email.arizona.edu

Objective

1. Determine mutational rates of pathogens during persistent colonization in different agricultural environments under distinctive geographical environmental conditions.

Funding for this project provided by the Center for Produce Safety through:
CDFA SCBGP grant# 18-0001-078-SC

FINAL REPORT

Abstract

Due to its higher resolution, whole genome sequencing (WGS) is rapidly becoming the new gold standard for foodborne outbreak investigations. Nonetheless there is a continued effort to refine the implementation of WGS, thus improving accuracy while reducing investigation time. Foodborne pathogens often cycle around defined regions, which was a major factor behind the creation of the FDA's GenomeTrakr network. As the pathogens continue to cycle through the environment in those regions, however, the genomes may accrue mutations that could alter the regional genomic pathogen profile, which could limit the positive impact of GenomeTrakr on investigations. The goal of this project was to determine the mutation rates of *Salmonella* Typhimurium, *Listeria monocytogenes*, and *Escherichia coli* O157:H7 during long-term persistence in agricultural soil and irrigation water maintained under temperature, humidity and sunlight hours conditions of either Salinas, California, or Yuma, Arizona. The project found that no mutations occurred in *S.* Typhimurium, *L. monocytogenes*, or *E. coli* O157:H7 during long-term colonization in soil or irrigation water under either environmental condition. However, cycling between new environments every few weeks or monthly did result in mutation rates of up to 2.48 mutations/year in *E. coli* O157:H7, 2.48 mutations/yr in *S.* Typhimurium, and 6.19 mutations/yr in *L. monocytogenes*, although the results varied between the soil and irrigation water environments and the Salinas-CA and Yuma-AZ conditions. The results of this project indicate a re-occurring or persistent foodborne pathogen (associated with a produce commodity) that is continuously isolated over time and reveals at least one mutation would be assumed to be cycling between different environments, whereas no mutations would suggest that the pathogen is simply surviving in the environments.

Background

Whole genome sequencing (WGS) is used by the U.S. Centers for Disease Control and Prevention (CDC), U.S. Food and Drug Administration (FDA), and U.S. Department of Agriculture (USDA) as the major genomic source tracking method during an outbreak investigation. The technique involves sequencing the genomes of different clinical strains that are potentially associated with the same outbreak based on initial epidemiological data, and then comparing the different genomes for mutations, single nucleotide polymorphisms (SNPs) or allele differences. As the implementation of WGS for outbreak investigations is still being refined and validated, the FDA currently suggests strains possessing less than 20 SNPs be included in the initial outbreak investigation. The higher the number of allowable SNPs for an outbreak investigation means potentially including many more cases in the initial investigation and thus, potentially having significantly more epidemiological and traceback data for investigators to review, slowing investigations. The faster an outbreak source can be identified, the more rapidly any contaminated product can be pulled off the shelves, reducing both the risk to consumers and the time frame of the government declared outbreak. A big factor in improving the speed of WGS for source tracking is to genetically characterize the pathogens in distinct geographic regions, as pathogens from a defined region are potentially distinct from pathogens in another defined region. This is a major factor behind the development of the FDA's GenomeTrakr network, which is a genome database of foodborne pathogens from different states, regions, and countries. Thus, during an outbreak a clinical strain of a foodborne pathogen can be matched to a particular area, and investigators can narrow down the search faster. It is known that foodborne pathogens remain prevalent in defined regions, cycling through environments within that area, and occasionally slipping through

surveillance and causing outbreaks. However, one potential issue in building this database, year after year, is that we do not know how these pathogens will genetically alter during prolonged periods in these agricultural environments. The pathogens may continue to replicate and mutate in those environments, thus changing the regional profile of that pathogen over time, and potentially limiting the efforts of GenomeTrakr to accelerate investigations.

The FDA released an investigative report on the Fall 2020 *Escherichia coli* O157:H7 outbreak associated with U.S.-grown leafy greens, which states that the strain is linked via WGS to the leafy greens outbreaks that occurred in 2018 and 2019. In fact, WGS has shown that this is not a unique situation with *E. coli* O157:H7 outbreaks and leafy greens, as it has revealed there are several strains that are consistently associated with clinical cases across the United States. It has become so common that the CDC has recently adopted the new classification of reoccurring, emerging, and persisting strains of *E. coli* O157:H7. While the above example is a reoccurring *E. coli* O157:H7 strain associated with leafy greens in California, the 2018 romaine lettuce outbreak strain associated with the Yuma, AZ, growing region has been associated with over 600 clinical cases over the last 4 years and deemed a persistent *E. coli* O157:H7 strain by the CDC. Therefore, this project is essential to understanding how foodborne pathogens, like *E. coli* O157:H7, persist in the environment, and to determine the most effective WGS methods for source tracking the pathogens over time. The objective of this project was to determine the mutational rates of pathogens surviving in agricultural soil and irrigation water over a prolonged period under different geographical conditions.

Research Methods and Results

Sample collection: Irrigation water and soil samples were collected three times throughout the project: (1) Yuma-AZ soil and irrigation water (October 10, 2019); (2) Salinas-CA soil and irrigation water (November 19, 2019); and (3) Yuma-AZ irrigation water only (March 11, 2020). At the point of collection, a pre-inoculation baseline sample was collected to determine the initial soil or irrigation water microbiome from the different regions, and also to confirm the samples were free of existing foodborne pathogens prior to project setup. The agricultural microbiomes were processed using either 10 grams (g) of agricultural soil or 5 liters (L) of agricultural irrigation water from each location. Each sample was processed to extract DNA, an Illumina library was prepared, and sequenced as described below.

Pathogen inoculation and microcosms: The selected pathogens for the project were all associated with fresh produce, including *E. coli* O157:H7 strain TW14359 (2006 spinach-associated outbreak), *Salmonella enterica* serotype Typhimurium strain RM15963 (isolated from irrigation water in Salinas, CA), and *Listeria monocytogenes* strain RM20669 (2014 stone fruit–associated outbreak). Individual environmental microcosms containing either 10 g of agricultural soil, 100 milliliters (ml) of irrigation water, or 100 ml of buffered peptone water (control) were established for each pathogen, condition (long-term, cycling, un-inoculated), environment (soil, water, control), geographical environment (Yuma, AZ and Salinas, CA), and time point (every two weeks for 12 weeks; then every four weeks for 16 time points) for the project (n=572). Soil microcosms were inoculated with either 10^8 colony forming units (CFU)/g (soil) or 10^8 CFU/ml (irrigation water or control) of *S.* Typhimurium, *E. coli* O157:H7 or *L. monocytogenes* by adding the corresponding pathogen inoculum to 10 g of agricultural soil, followed by briefly shaking the microcosm to homogenize the sample. Irrigation water microcosms and control microcosms were inoculated by adding either 99 ml of irrigation water or 99 ml of buffered peptone water to the microcosm and then adding 1 ml of the appropriate pathogen inoculum. Inoculated microcosms were then shaken briefly to homogenize the sample. In total, microcosms for each pathogen under the two regional conditions were set up for the following environments: (1) soil long-term; (2) soil cycling [re-

isolated pathogen used to inoculate new soil microcosm every sampling period]; (3) irrigation water long-term; (4) irrigation water cycling [re-isolated pathogen used to inoculate new irrigation water microcosm every sampling period]; (5) soil/water cycling [re-isolated pathogen used to inoculate new soil or irrigation water microcosms every sampling period by switching soil to irrigation water to soil for the duration of the project]; (6) control long-term; (7) control cycling [re-isolated pathogen used to inoculate new buffered peptone water every sampling period]; (8) un-inoculated soil [microbial communities control]; (9) un-inoculated irrigation water [microbial communities control]. All microcosms were placed in the corresponding geographical location chamber as described below.

Environmental conditions: Darwin plant growth chambers were used to mimic the natural agricultural soil and irrigation water conditions for each geographical location: Yuma-AZ and Salinas-CA. Prior to inoculation, the regional environmental conditions for both Salinas, CA, and Yuma, AZ, were also calculated by taking the monthly average for the amount of sunlight, humidity and temperature for each region across the last five years using historical data from the website Weather Underground (https://www.wunderground.com). Upon the initial inoculation of the microcosms, the plant growth chambers conditions (light, humidity, and temperature) were set to average conditions for the beginning of the growing season for each corresponding regional condition, thus Yuma-AZ (September) and Salinas-CA (January) and both conditions were changed every 30 days to the conditions for the next month for the duration of the project.

Sample processing: Samples were removed for processing every 2 weeks throughout the course of the project for both shotgun metagenomics sequencing and pathogen culturing. Soil microcosms for shotgun metagenomics sequencing were processed by removing 2 g of soil and freezing back at -20°C until DNA extraction. The remaining 8 g of soil was mixed with sterile phosphate buffered saline, vortexed, sonicated using a heated ultrasonic cleaning bath, and the sediment permitted to settle. Irrigation water samples were processed by filtering the entire 100 ml sample using a sterile 0.22-um filter, aseptically rolling the filter into a sterile conical tube, washing with sterile phosphate buffered saline, and then vortexing to dislodge the microbes. One milliliter of each processed soil or irrigation water sample was removed and stored at -20°C until DNA extraction, while another 1 ml was serially diluted 10-fold (10^{-2} to 10^{-5}) and then plated on the appropriate selective media for the corresponding pathogen. Samples were incubated at 37°C for 24–48 hours, then colonies counted to quantify the viable and culturable pathogen present in each corresponding microcosm. Samples positive for chromogenic appearance on selective media were further confirmed via polymerase chain reaction of target pathogen genes. Plates with the lowest dilution concentration were subjected to a “plate wash” by adding sterile phosphate buffered saline to the plate and re-suspending the bacteria from the agar. One milliliter of “plate wash” sample was frozen back at -20°C until DNA extraction.

Illumina sequencing library preparation and sequencing: Sample DNA was extracted using the Qiagen PowerSoil Pro Kit per the manufacturer’s instructions. Next, Illumina sequencing libraries were prepared from the extracted DNA using the Illumina DNA Prep Tagmentation kits per the manufacturer’s instructions, and each sample was individually barcoded using the Illumina Unique Dual Indexes for multiplex sequencing. All prepared and barcoded samples were pooled together in equal molar ratios, and then were sequenced on a single lane of a flow cell on an Illumina NovaSeq S4 sequencer by either the genomics company Psomagen (Rockville, MD) or the genomics core facility at North Carolina State University (Raleigh, NC).

Bioinformatic analysis: The microbial communities for each of the samples were determined by assembling the sequencing reads using the program MetaSPades with the default settings. Next, assembled contigs were taxonomically identified by using the program Kraken and compared

against an 8 gigabases MiniKraken database. Identification of any potential mutations or SNPs in the target pathogens was determined by mapping the sequencing reads of the target sample against the complete genome of either *E. coli* O157:H7 str. TW14359, *S.* Typhimurium str. RM15963, or *L. monocytogenes* str. RM20669, depending on the sample. Mapping of the reads and mutation identification was conducted using a pipeline established in the program Geneious Prime (v2021.1.03). SNPs, nucleotide insertions, or deletions were identified using the following parameters: at least 10x coverage, 97% minimum variant frequency, maximum variant p-value 10^{-6}, and minimum strand-bias p-value 10^{-5} for each of the samples. Potential mutations were then visibly inspected by the PI to conclude if they were accurate mutations, repeat regions, or sequencing artifacts, and the location of final accurate mutations were determined using the original genome map for the pathogen. Mutation rate for the different pathogens in various environments under either the Salinas-CA or Yuma-AZ conditions were calculated by taking the total number of mutations that occurred in the pathogen in that environment and condition divided by the 42 weeks of the study then multiplied by 52 weeks in a year to give the number of mutations per year (number of mutations/42 weeks of project x 52 weeks = mutations/yr).

E. coli O157:H7 mutation results. No mutations were detected in *E. coli* O157:H7 in any of the long-term environments, including the controls (buffered peptone water) kept under either the Salinas-CA or Yuma-AZ conditions throughout the project. However, cycling through soil every few weeks under the Salinas conditions resulted in an SNP in the *rpoS* gene at week 22 of the study, and an additional single base deletion in an intergenic region at week 42 (mutation rate = 2.48 mutations/yr). Soil cycling under the Yuma conditions did not result with an SNP in the *rpoS* gene until week 42 of the study (mutation rate = 1.24 mutations/yr). Cycling through irrigation water every few weeks resulted in *E. coli* O157:H7 under the Salinas conditions with a single SNP in the *rpoS* gene at 42 weeks (mutation rate = 1.24 mutations/yr), while there were no mutations that occurred during water cycling of *E. coli* O157:H7 under the Yuma conditions. Additionally, there were no mutations in *E. coli* O157:H7 cycling between soil and irrigation water every few weeks during the entire 42 weeks under either the Salinas or Yuma conditions. The cycling in the control environments every few weeks resulted in three mutations occurring by week 42 under the Salinas conditions (mutation rate = 3.71 mutations/yr), including the *rpoS* gene, single base intergenic deletion, and an SNP in the *wbdN* gene. In contrast, under the Yuma conditions only the SNP in the *rpoS* gene was detected in the control cycling at week 42 (mutation rate = 1.24 mutations/yr) of the study (**Table 1**). The SNP in the *rpoS* gene is non-synonymous and results in the methionine of the start codon being changed to threonine, which results in a 93 amino acid truncation of rpoS. The *rpoS* gene is a stress regulator of *E. coli* O157:H7 that has a critical role in surviving under stress conditions, and alterations to this regulator would have a large impact on numerous other genes that could alter the ability of *E. coli* O157:H7 to survive or thrive in the different environments as it cycles through those environments. The SNP in the *wbdN* gene was also non-synonymous and resulted in an aspartic acid change to glycine or an acidic amino acid to an aliphatic amino acid, but the impact of this change is not known without further experiments. The *wbdN* gene is part of the *E. coli* O157:H7 O antigen gene cluster, so minor alterations to the O antigen could result in better environmental survival. However, the exact impact of either mutation in *E. coli* O157:H7 can only be confirmed through further studies.

Salmonella Typhimurium mutation results. The only long-term mutation in *S.* Typhimurium included an intergenic SNP at 22 weeks in the control kept under the Yuma conditions (mutation rate = 1.24 mutations/yr). Soil cycling resulted in no mutations in *S.* Typhimurium under the Yuma conditions, but an SNP in the *ompR* gene and an intergenic SNP at week 42 (mutation rate = 2.48 mutations/yr) under the Salinas conditions. Cycling in irrigation water every few weeks resulted in the opposite, with no mutations under the Salinas conditions, but there was an SNP in the *ompR* gene and an intergenic SNP at week 12 under the Yuma conditions (mutation rate = 2.48

mutations/yr). Furthermore, there were also no mutations detected in *S.* Typhimurium cycled through soil and irrigation water every few weeks under the Salinas conditions, but there was an SNP in the *ompR* gene in *Salmonella* kept under the Yuma conditions that occurred by week 12 of the study (mutation rate = 1.24 mutations/yr). Finally, control cycling under both conditions resulted in an SNP in the *ompR* gene and an intergenic SNP, but that occurred at week 12 under Yuma conditions and week 42 under the Salinas conditions for *S.* Typhimurium (mutation rate = 2.48 mutations/yr) in the study (**Table 2**). The SNP in the *ompR* gene was a non-synonymous mutation that resulted in hydroxylic serine amino acid being changed to an essential aromatic phenylalanine amino acid, but again without further study the exact impacts on the gene are unknown. However, the *ompR* gene does have a critical role in dealing with osmotic stress for *Salmonella*, and therefore could be critical for the pathogen to survive and cycle through these different agricultural environments.

<u>*Listeria monocytogenes* mutation results</u>. The only long-term mutations that occurred in *L. monocytogenes* were in the control environment under the Yuma conditions, which included three intergenic SNPs, an SNP in the GntR family transcriptional regulator gene, an SNP in PP2C family protein-serine/threonine phosphatase gene, and a 3-base insertion in a putative drug-efflux transporter gene at week 42 of the study (mutation rate = 7.43 mutations/yr). Soil cycling resulted in only an intergenic SNP at week 42 in *L. monocytogenes* kept under the Yuma conditions (mutation rate = 1.24 mutations/yr), but five mutations including an intergenic insertion, SNP in a hypothetical protein gene, SNP in a maltose phosphorylase gene, SNP in an alanine racemase gene, and a SNP in a MarR family transcriptional regulator gene at week 42 (mutation rate = 6.19 mutations/yr) under the Salinas conditions. Cycling in irrigation water every few weeks resulted in no mutations in *L. monocytogenes* kept under Yuma conditions, but an intergenic SNP, a SNP in PP2C family protein-serine/threonine phosphatase gene, and a 3-base insertion in a putative drug-efflux transporter gene at week 22 under the Salinas conditions (mutation rate = 3.71 mutations/yr). Furthermore, there were no mutations detected in *L. monocytogenes* cycled through soil and irrigation water every few weeks under either the Salinas or Yuma conditions. Finally, control cycling under the Yuma conditions resulted in an intergenic SNP, an SNP in PP2C family protein-serine/threonine phosphatase gene, and a 3-base insertion in a putative drug-efflux transporter gene (mutation rate = 3.71 mutations/yr) at 22 weeks, but five mutations including an intergenic deletion, SNP in a hypothetical protein gene, SNP in a maltose phosphorylase gene, SNP in an alanine racemase gene, and an SNP in a MarR family transcriptional regulator gene at week 42 (mutation rate = 6.19 mutations/yr) under the Salinas conditions (**Table 3**). The SNP in the GntR family transcriptional regulator gene results in an early stop codon that produces a 185 amino acid truncation of the protein product, which would suggest it would be non-functional and shut down as a transcriptional regulator. The SNP in the PP2C family protein-serine/threonine phosphatase gene results in a non-synonymous mutation that changes an aliphatic glycine to an acidic glutamic acid amino acid. The 3-base insertion in the putative drug-efflux transporter gene results in the insertion of an essential aliphatic isoleucine into the gene. The SNP in the hypothetical protein is also a non-synonymous mutation that results in an early stop codon that produces a 263 amino acid truncation of the protein. The maltose phosphorylase gene had a non-synonymous mutation that resulted in a basic arginine changing to an aliphatic proline amino acid, whereas the alanine racemase gene also had a non-synonymous SNP that resulted in an early stop codon and a 357 amino acid truncation of the protein. Maltose phosphorylase is an enzyme that is involved in the metabolism of starch and sucrose. The alanine racemase catalyzes the conversion of L-alanine to D-alanine, which is an essential component of the peptidoglycan of the cell wall of gram-positive bacteria. The non-synonymous SNP in the MarR family transcriptional regulator gene resulted in a change from an essential sulfur-containing methionine to an essential aliphatic isoleucine amino acid. Although the direct impact on the different genes that these various non-synonymous mutations have is unknown without further studies. Nevertheless, it

appears that the combination of these non-synonymous mutations would result in alterations to different biochemical pathways and transcriptional regulators that could represent *L. monocytogenes* adapting to the environment(s) that it is cycling through such as soil or irrigation water.

Microbial community results. While not an objective of this project, the utilization of shotgun metagenomics allows the determination of microbial communities present in the environments and how those communities shift after the introduction of the various pathogens. The soil from a leafy green field in Yuma, AZ (**Figure 1**) and kept under the Yuma conditions for two weeks after the introduction of *E. coli* O157:H7 resulted in several shifts in the microbial community genera, including a 3% increase in *Streptomyces*, 1% decrease in *Pimelobacter*, 2% decrease in *Staphylococcus*, and 1% increase in *Bacillus* (**Figure 2**). The introduction of *S.* Typhimurium to the soil resulted in similar shifts, including a 1% increase in *Streptomyces*, 1% decrease in *Pimelobacter*, and a 2% decrease in *Staphylococcus* after two weeks (**Figure 3**). The introduction of *L. monocytogenes* to the Yuma soil resulted in quite different community responses such as a 3% decrease in *Streptomyces*, 1% decrease in *Staphylococcus*, and 1% increase in *Lysobacter* (**Figure 4**). The microbial communities from Yuma irrigation water from an open cement-lined canal (**Figure 5**) also had slight shifts following the introduction of the different pathogens: two weeks after the introduction of *E. coli* O157:H7 resulted in a 1% increase in *Pseudomonas* and 2% decrease in *Candidatus Nanopelagicus* (**Figure 6**), whereas *S.* Typhimurium also resulted in a 10% increase in *Staphylococcus* (**Figure 7**), and *L. monocytogenes* had similar shifts compared to *E. coli* O157:H7 (**Figure 8**). Soil from a leafy green field in Salinas, CA, after two weeks under the Salinas conditions of the project (**Figure 9**) had major shifts after the introduction of *E. coli* O157:H7, including a 1% increase in *Stenotrophomonas*, 2% increase in *Pseudomonas*, 4% decrease in *Nocardioides*, 3% decrease in *Arthrobacter*, and 5% decrease in *Streptomyces* (**Figure 10**). Whereas the introduction of *S.* Typhimurium resulted in a 3% increase in *Bacillus*, 1% decrease in *Nocardioides*, and 1% decrease in *Arthrobacter* after two weeks (**Figure 11**), and inoculation with *L. monocytogenes* had a 1% decrease in *Micromonospora*, 3% decrease in *Streptomyces*, but a 1% increase in *Nocardioides* (**Figure 12**). Irrigation water from an irrigation pipe in Salinas, CA after two weeks under the Salinas conditions used in the project (**Figure 13**) had huge community shifts two weeks after the introduction of *S.* Typhimurium, including a 17% decrease in *Pseudomonas*, 20% decrease in *Delftia*, 11% decrease in *Acinetobacter*, 6% increases in *Rhodoferax* and *Comamonas*, and a 4% increase in *Acidovorax* (**Figure 14**). Two weeks after the introduction of *L. monocytogenes* to Salinas irrigation water also caused some community shifts like a 3% decrease in *Delftia* and *Sphingopyxis*, 11% decrease in *Acinetobacter*, 2% increase in *Pseudomonas*, and 3% increase in *Brevundimonas* (**Figure 15**), and unfortunately the WGS did not generate any data for the Salinas irrigation water inoculated with *E. coli* O157:H7 at 2 weeks. Although this analysis is very preliminary and significantly more is required, it does provide some interesting insights into those microbial community members that might be positively or negatively impacted by the presence of the different pathogens in these environments.

Survivability of pathogens in different microcosms results. The results of the culturable counts for the microcosms kept under the Yuma-AZ conditions corresponding to the September to mid-June time points found none of the three pathogens were recovered from any of the long-term experimental microcosms (soil and irrigation water). Conversely, nearly all the control microcosms for all three pathogens maintained culturable counts >10^6 CFU/ml for all 42 weeks, except for the reduction of the *Salmonella* control at week 42 and the week 12 *E. coli* O157:H7 control that was non-culturable but returned to higher levels by week 16 (**Figure 16**). For the first 12 weeks, even the cycling-every-two-weeks microcosms did not have culturable pathogens in either soil or irrigation water under the Yuma-AZ conditions corresponding to September to November time

points. However, during the December to mid-June time points, there were low levels of culturable cells detected in the re-inoculated cycling microcosms for all three pathogens under the Yuma-AZ conditions, but the results were extremely variable from month to month (**Figures 17–19**). Results for the long-term microcosms under Salinas-CA conditions were different than those under the Yuma-AZ conditions, and represented the January to mid-October time points. *E. coli* O157:H7 in soil microcosms became non-culturable by week 10, and then spiked back up to low levels of culturable cells at weeks 22 and 26 before becoming non-culturable again at week 32. The culturable *E. coli* O157:H7 cells in irrigation water were >10^3 CFU/ml at weeks 2–16, were non-culturable at week 22, and returned to >10^4 CFU/ml levels from weeks 26–42. The control microcosms maintained at least 10^3 CFU/ml for the entire 42 weeks (**Figure 20**). *Salmonella* was culturable in the long-term soil microcosms until week 16, but only until week 12 in the long-term irrigation water microcosms. This result contrasts with the *Salmonella* controls, which maintained between 10^6 and 10^8 CFU/ml for the entire 42 weeks of the study (**Figure 21**). *L. monocytogenes* was culturable in the long-term soil microcosms until week 10 and was culturable in long-term irrigation water microcosms until week 16. Conversely, the *L. monocytogenes* controls remained at >10^5 CFU/ml in the control microcosms for all 42 weeks, under the Salinas-CA conditions (**Figure 22**). Unlike *E. coli* O157:H7, *Salmonella* and *L. monocytogenes* did not have any culturable cell count spikes during the 42 weeks of the study. The cycling soil and irrigation water microcosms kept under the Salinas-CA conditions for all three pathogens had culturable *E. coli* O157:H7, *Salmonella* or *L. monocytogenes* for all 42 weeks. Nevertheless, there were variabilities among the cycling microcosms for each of the pathogens, but the pathogens clearly remained culturable in each environment for several weeks under Salinas-CA conditions (**Figures 23–25**).

Soil chemistry results. Several produce industry representatives asked about differences in the soil chemistry between the two geographical regions in the project, and whether this could be impacting the difference in the survivability of the different pathogens between Yuma and Salinas. Therefore, the PI had each of the soil samples from the two locations analyzed for major chemical composition of the soils at the University of Arizona Laboratory for Emerging Contaminants (ALEC). The results of the soil analysis are presented in **Table 4**, and show significant differences between the two locations for several different chemicals. In particular, sodium levels in soil from Yuma, AZ, were three times higher than in soil from Salinas, CA. However, the specific impact on pathogen survivability of the different chemicals in the soils will need additional studies, but this does provide some preliminary data to suggest that chemical composition difference could impact pathogen survival in agricultural soils.

Outcomes and Accomplishments

*Positive Outcome 1***.** The PI's laboratory examined the most selective media for isolating the three pathogens utilized in this project. The parameters of the project require that the pathogens be isolated via direct plating from the soil and irrigation water without any type of enrichment process, as additional culturing could induce mutations in the pathogens and potentially alter the results. The best selective media for direct plating for each pathogen was determined to be: for: Xylose lysine tertitol agar (XLT) for *Salmonella*; Cefixime tellurite sorbitol MacConkey agar (CT-SMAC) for *E. coli* O157:H7; and RAPID'L.mono agar for *L. monocytogenes*. Each of these media allowed for direct plating that minimized passing cultures numerous times for re-isolation of the pathogens from the environmental samples, which reduced the risk of inducing mutations by growth on artificial media instead of the agricultural environments.

Positive Outcome 2. The PI's laboratory determined the most effective method to maximize screening of the pathogen population in each of the different agricultural environmental samples

for the presence of any mutations. The method included taking a two-step approach to determine the mutation rates of the pathogens in the samples, including: (1) collection of a "plate wash" of the lowest dilution of the pathogen on selective media; and (2) shotgun metagenomic sequencing of each sample directly without plating the pathogen on selective media. The project sampling was changed from collecting five individual colonies of the different re-isolated pathogens from a sample to collecting all the bacterial colonies from the lowest dilution plate of the selective media, which obtained the maximum amount of the pathogen population from the sample. Pathogens were collected directly off the selective media, allowing us to obtain higher concentrations of the pathogens for sequencing. This method required growth on the artificial media that could induce additional mutations unrelated to the agricultural environment. Therefore, the total DNA from the different agricultural samples was also directly isolated without plating and then sequenced using a shotgun metagenomics approach, which should reveal only mutations induced in each of the environments. This second approach ran the risk of not having enough pathogen DNA to accurately detect the mutations, but it also allowed for the identification of the microbial communities present in the different samples and how these communities responded to the presence of the pathogen. Overall, this two-step approach allows each step to cover the weakness of the other step, thus it provided the most accurate method for accomplishing the objective of the project.

*Negative Outcome 1***.** The project had a six-month delay in getting the original long-term agricultural environmental samples inoculated, due to a major delay in receiving the Darwin plant growth chambers. The Darwin plant growth chambers were critical to maintaining temperature, light, and humidity under the different conditions for the project. The manufacturer originally promised delivery in early March 2019 but proceeded to delay the delivery date until September 2019, at which point the experiments were started immediately. The timeline for the project was adjusted to accommodate this delay in the start of the major experiments and still complete the project on time. Additionally, the PI used the delay to troubleshoot many of the protocols for the project, and then modify the project according to the troubleshooting results to generate the most accurate data for the project.

Negative Outcome 2. The project PI, like most researchers around the world, was heavily challenged and delayed by the COVID-19 pandemic. The PI's research laboratory was completely shut down from March 2020 until the end of June 2020, which resulted in an over four-month delay in most of the experiments for the project. Additionally, there were social distancing protocols required by the university from July 2020 to September 2020, which only allowed one researcher in the laboratory at a time, further slowing progress. The timeline for the project was adjusted to accommodate the pandemic-related delay, as the PI's research laboratory was still able to get all appropriate samples processed and sequenced for the project by the original end of project (December 2020). However, the project was granted a no-cost extension until March 31, 2021, to conduct the computational metagenomic analysis of all the sequencing data.

Negative Outcome 3. At the week 46 time point of the study, the Darwin plant growth chambers keeping the inoculated agricultural environmental samples under the Salinas-CA conditions flooded due to a leak in the water line for maintaining the humidity of the chamber. The flooding compromised the remaining samples in those chambers, and thus required a biosafety clean-up. Due to the compromised Salinas samples at the week 46 time point, all the Salinas-CA samples for weeks 46 and 52 were lost. Therefore, the PI decided to end the project at the week 42 time point for both conditions (Salinas-CA and Yuma-AZ) to keep the samples balanced between the conditions and not rely on any potentially compromised samples.

First Accomplishment: Collection of different agricultural environmental samples from different leafy greens fields in Yuma, AZ and Salinas, CA, and establishing the pathogen inoculated environmental microcosms under distinctive geographical conditions.

Collecting soil and irrigation water samples from leafy greens fields in both Yuma, AZ, and Salinas, CA, and then establishing the inoculated environment microcosms was the critical first step in completing this project. Failure to achieve this accomplishment would have prevented the generation of the needed samples for WGS and generating the overall data required to meet the objective of the project.

Second Accomplishment: Processing inoculated agricultural soil and irrigation water microcosm samples for the long-term survival and environmental cycling every few weeks for 42 weeks post pathogen inoculation; and culturing each sample for viable and culturable pathogen levels and extracting DNA for sequencing of the different samples to determine the mutation rates of the pathogens in the different environments kept under different geographical conditions.

This accomplishment included processing the inoculated microcosm samples for each of the long-term survival and environmental cycling for the different time points up through 42 weeks post pathogen inoculation. However, the survivability of the pathogen varied between months for each of the locations. During this project a total of 988 samples were collected, processed and frozen back for potential sequencing. Completion of this accomplishment was central to the entire project, as it provided the necessary samples for the WGS to determine the mutations that occurred in those pathogens that survived in the different environments. It also provided some insight into the eventual results of the project, as the failure for the pathogens to survive long-term under the Yuma-AZ conditions indicated that these pathogens would most likely not be present in some of the sequence data, and thus, there would be few if any mutations detected in the pathogens under these conditions. Overall, the survival results from this accomplishment provide preliminary scientific data indicating that these three pathogens will not survive long-term in the environment in Yuma, AZ, while they will survive for at least several months in Salinas, CA.

Third Accomplishment: Illumina sequencing of “plate wash” and shotgun metagenomic sequencing samples collected during the project, and bioinformatic analysis of each sequenced sample to determine the presence of any mutations in the three different pathogens kept in the different environments under the two geographical conditions.

This accomplishment was the actual completion of the entire objective of the project. In total, 419 samples were submitted for WGS, and then analyzed for the presence of mutations that developed in the pathogen that was inoculated into the microcosm. There are 569 samples remaining that are available for future sequencing and analysis if needed to address additional questions or issues. Additionally, the composition of the microbial communities was determined for the different soil and water samples, which provides compositional changes in the presence of the pathogens. Overall, the results of this accomplishment elucidate how these pathogens persist and adapt in different agricultural environments under various geographical conditions. Furthermore, results confirm that utilization of the GenomeTrakr database for quickly identifying regions associated with a pathogen during outbreak investigations should be highly effective for outbreak investigations where the pathogen is persisting long-term in environment. Yet, it will be less effective for pathogens that are cycling through different environments before there is concern of additional mutations (when compared to the GenomeTrakr database) placing the strain outside the <20 single nucleotide polymorphism cut-off.

Summary of Findings and Recommendations

Unfortunately, in many cases, the pathogen-inoculated soil and irrigation water samples kept under Yuma-AZ conditions did not have high levels of pathogen DNA that could be used to accurately identify any mutations present in the pathogens. In fact, many of these samples did not have any pathogen DNA present in the sequence data after analysis, which is appropriate considering that the pathogens survived less than two weeks in any of the long-term samples or in most of the environmental cycling samples under the Yuma-AZ conditions; the lack of pathogen DNA present in the sample for sequencing resulted in the failure to get quality pathogen DNA for sequencing in these samples. However, this finding is a good indication to the produce industry that *E. coli* O157:H7, *S.* Typhimurium, and *L. monocytogenes* do not survive very well under Yuma-AZ conditions at least from September to November, but will survive slightly longer from December to mid-June, at least when cycling between environments. This was evident in the development of single mutations at week 42 in soil cycling for *E. coli* O157:H7 and *L. monocytogenes*. *Salmonella* was non-culturable in Yuma conditions at week 12 (November conditions), but WGS found mutations developed in the water cycling and soil/water cycling indicating that *Salmonella* might have been viable but non-culturable (VBNC) at that time point. However, further studies would be required to determine if VBNC pathogens, particularly *Salmonella*, can replicate and incur mutations in cycling through these environments.

In contrast, pathogen-inoculated soil and irrigation water samples kept under Salinas-CA conditions did yield quality levels of pathogen DNA from the sequence data, which allowed for determination of any mutations that occurred during the long-term survival in these environments. Results from the long-term survival environments found that there were no mutations in the pathogen population even after 42 weeks, which indicates that the pathogens are surviving in these environments but not actively replicating. Therefore, if the pathogens are simply surviving in the environment in the long-term there is minimal risk of pathogens developing high numbers of mutations that would alter the data present in the GenomeTrakr database. However, the addition of cycling between new environments every few weeks or monthly during the study revealed a different result, as there were mutations detected in the pathogens in the different environments including both cycling through soil and irrigation water every two weeks.

The goal of this project was to study the mutational rates of the pathogens in different agricultural environments to improve WGS for rapid outbreak identification and source tracking. The project found that no mutations occurred in *E. coli* O157:H7 during long-term colonization in soil or irrigation water under either condition studied. However, *E. coli* O157:H7 did have mutation rates from 1.24 to 2.48 mutations/yr in cycling through irrigation water and soil under the Salinas conditions, respectively. Under the Yuma conditions the only mutation rate was 1.24 mutations/yr during cycling through soil samples every few weeks. Like *E. coli* O157:H7, *S.* Typhimurium did not have any mutations develop during long-term colonization in soil or irrigation water under either condition. Interestingly, *S.* Typhimurium did have mutation rates from 1.24 to 2.48 mutations/yr in cycling through irrigation water and soil/irrigation water under Yuma conditions, respectively; in contrast, under the Salinas conditions the only mutation rate was 2.48 mutations/yr during cycling through soil samples every few weeks. *L. monocytogenes* also did not develop any mutations during long-term colonization of either soil or irrigation water under either of the conditions. However, *L. monocytogenes* did have the highest mutation rates of the three pathogens, ranging from 3.71 to 6.19 mutations/yr in water cycling and soil cycling under the Salinas conditions, respectively; under the Yuma conditions the mutation rate was 1.24 mutations/yr during cycling between soil samples every few weeks. Thus, these three pathogens do not appear to develop any mutations during long-term colonization of either agricultural soil or irrigation water regardless of the conditions that the pathogens are under. On the other hand, cycling between environments every few weeks or monthly, particularly different soils, results in mutations occurring in all three of the pathogens.

Overall, the results indicate a re-occurring or persistent foodborne pathogen (associated with leafy greens or other produce commodities) that is continuously isolated and reveals at least one mutation between samples at different time points would be assumed to be cycling between different hosts and/or environments, whereas no mutations would suggest that the pathogen is simply surviving in the environments and not being transferred between different hosts and/or environments. These results will assist regulatory agencies and the produce industry in determining if there is potentially a single source or multiple sources of the pathogen around fields or other points of possible product contamination during outbreak investigations or routine surveillance.

Finally, the project also was able to generate preliminary data on identifying and quantifying the microbial communities of soil and irrigation water samples from Yuma, AZ, and Salinas, CA, and the shifts in the communities that occurred due to the addition of the different pathogens into those corresponding environments. Additionally, the project generated some preliminary data on the survivability of the three different pathogens under the different conditions, including those months in either Salinas or Yuma that have the highest and lowest levels of pathogen survival, thus representing the months with the highest risk of contamination. Furthermore, the project generated some preliminary data on the impact of the soil chemistry on the survival of the pathogens in soils from different geographical locations. However, none of these were the objectives of the project, therefore all aspects are preliminary data and hard conclusions cannot be accurately drawn from the data without more detailed studies in the future.

APPENDICES

Publications and Presentations

Oral presentations

1. 2020 Center for Produce Safety Research Symposium Webinar Series
2. December Yuma Safe Produce Council Research Meeting
3. 2021 Center for Produce Safety Research Symposium Webinar Series

Poster presentations

1. 2019 Center for Produce Safety Research Symposium (Austin, TX)
2. 2020 University of Arizona Food Safety Consortium Conference
3. 2021 University of Arizona Food Safety Consortium Conference

Publication

A manuscript is currently being prepared for submission to a peer-review journal in 2021.

Budget Summary

Total research funds of $81,523 were awarded to this project. Some expenditures for the project included $26,414 for the purchase of two (2) Darwin plant growth chambers for maintaining the pathogen inoculated microcosms under the Salinas-CA and Yuma-AZ environmental conditions (humidity, temperature, and sunlight) throughout the project; $27,197 was used for salary and benefits for Victoria Obergh, who conducted all the microcosm experiments, DNA extractions, constructed all the Illumina sequencing libraries, and assisted in the final data analysis of the project; ~$7,400 was used for the Illumina DNA Library Prep Kit and Illumina Unique Dual Indexes, which were supplies needed for generating barcoded Illumina sequencing libraries for Illumina sequencing of the samples; $10,010 was used for conducting Illumina sequencing on two runs of a single lane of a flow cell on an Illumina NovaSeq S4 sequencing by the company Psomagen for one run and the genomics core facility at North Carolina State University for the other run; and ~$9,000 was used for purchasing different supplies for the project including but not limited to different types of selective media for the pathogens, DNA genomic isolation kit, cryotubes for storing the samples at -80°C or -20°C, EZ-Fit Filtration unit and 0.22-um filters, sterile containers for the microcosms, soil moisture and pH meters, carboys for irrigation water collection, and filter pipette tips. Due to shifts in the project design during year 1 of the study there were some minor additional costs to the project that utilized Dr. Cooper’s start-up funds, but otherwise there were enough funds to cover the cost of the project.

Tables 1–4 and Figures 1–25

Table 1. *E. coli* O157:H7 mutations

Salinas, CA conditions				
Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	*rpoS* gene	*rpoS* gene & intergenic deletion
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	No mutations	No mutations	*rpoS* gene
Soil/Water cycling**	No mutations	No mutations	No mutations	No mutations
Long term control	No mutations	No mutations	No mutations	No mutations
Control cycling	No mutations	No mutations	No mutations	*rpoS* gene, *wbdN* gene, & intergenic deletion

Yuma, AZ conditions				
Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	No mutations	*rpoS* gene
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	No mutations	No mutations	No mutations
Soil/Water cycling**	No mutations	No mutations	No mutations	No mutations
Long term control	No mutations	No mutations	No mutations	No mutations
Control cycling	No mutations	No mutations	No mutations	*rpoS* gene

*NOTE: Cycling conditions involved re-inoculated new soil or water every two weeks for the duration of the project, which represents moving into a new environment every two weeks.
**Soil/Water cycling involved switching between soil, irrigation water, soil, irrigation water, etc. every two weeks for the duration of the project.

Table 2. *Salmonella* Typhimurium mutations

Salinas, CA conditions

Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	No mutations	*ompR* gene & intergenic SNP
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	No mutations	No mutations	No mutations
Soil/Water cycling**	No mutations	No mutations	No mutations	No mutations
Long term control	No mutations	No mutations	No mutations	No mutations
Control cycling	No mutations	No mutations	No mutations	*ompR* gene & intergenic SNP

Yuma, AZ conditions

Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	No mutations	No mutations
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	*ompR* gene & intergenic SNP	No new mutations	No new mutations
Soil/Water cycling**	No mutations	*ompR* gene	No new mutations	No new mutations
Long term control	No mutations	No mutations	Intergenic SNP	No new mutations
Control cycling	No mutations	*ompR* gene & intergenic SNP	No new mutations	No new mutations

*NOTE: Cycling conditions involved re-inoculated new soil or water every two weeks for the duration of the project, which represents moving into a new environment every two weeks.
**Soil/Water cycling involved switching between soil, irrigation water, soil, irrigation water, etc. every two weeks for the duration of the project.

Table 3. *L. monocytogenes* mutations

Salinas, CA conditions				
Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	No mutations	intergenic insertion; SNPs in hypothetical protein; maltose phosphorylase; alanine racemase; MarR family transcriptional regulator
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	No mutations	3 base insertion putative drug-efflux transporter; intergenic SNP; PP2C family protein-serine/threonine phosphatase SNP	No new mutations
Soil/Water cycling**	No mutations	No mutations	No mutations	No mutations
Long term control	No mutations	No mutations	No mutations	No mutations
Control cycling	No mutations	No mutations	No mutations	intergenic deletion; SNPs in hypothetical protein; maltose phosphorylase; alanine racemase; MarR family transcriptional regulator

Yuma, AZ conditions				
Condition	2 weeks	12 weeks	22 weeks	42 weeks
Long term soil	No mutations	No mutations	No mutations	No mutations
Soil cycling	No mutations	No mutations	No mutations	Intergenic SNP
Long term water	No mutations	No mutations	No mutations	No mutations
Water cycling	No mutations	No mutations	No mutations	No mutations
Soil/Water cycling**	No mutations	No mutations	No mutations	No mutations
Long term control	No mutations	No mutations	No mutations	Three Intergenic SNPs; GntR family transcriptional regulator SNP; 3 base insertion putative drug-efflux transporter; SNP PP2C family protein-serine/threonine phosphatase
Control cycling	No mutations	No mutations	Intergenic SNP; 3 base insertion putative drug-efflux transporter; SNP PP2C family protein-serine/threonine phosphatase	No new mutations

*NOTE: Cycling conditions involved re-inoculated new soil or water every two weeks for the duration of the project, which represents moving into a new environment every two weeks.
**Soil/Water cycling involved switching between soil, irrigation water, soil, irrigation water, etc. every two weeks for the duration of the project.

Table 4. Chemical composition of soils from leafy greens fields.

Chemical concentration	Salinas, CA soil	Yuma, AZ soil
Beryllium (ug/g)	0.900 ± 0.036	0.630 ± 0.024
Sodium (ug/g)	353.263 ± 10.5	1063.081 ± 24.84
Magnesium (ug/g)	12,290.145 ± 478.696	7,612.687 ± 238.952
Aluminum (ug/g)	19,532.45 ± 2,047.12	12,842.459 ± 1,556.167
Potassium (ug/g)	4,238.437 ± 225.472	2,480.85 ± 132.105
Calcium (ug/g)	11,738.755 ± 243.258	26,182.131 ± 852.122
Vanadium (ug/g)	53.380 ± 6.063	38.564 ± 4.74
Chromium (ug/g)	68.598 ± 3.365	20.816 ± 3.169
Manganese (ug/g)	482.345 ± 5.208	415.307 ± 11.762
Iron (ug/g)	27,340.178 ± 951.971	16,205.357 ± 1,116.458
Cobalt (ug/g)	13.631 ± 0.340	8.435 ± 0.33
Nickel (ug/g)	91.866 ± 2.186	22.322 ± 1.003
Copper (ug/g)	30.106 ± 0.781	25.411 ± 1.220
Zinc (ug/g)	113.335 ± 3.14	58.436 ± 3.583
Arsenic (ug/g)	5.354 ± 0.174	6.981 ± 0.525
Selenium (ug/g)	0.957 ± 0.205	0.488 ± 0.278
Molybdenum (ug/g)	0.56 ± 0.107	0.416 ± 0.025
Silver (ug/g)	0.086 ± 0.005	0.022 ± 0.002
Cadmium (ug/g)	1.452 ± 0.032	0.415 ± 0.06
Tin (ug/g)	0.118 ± 0.017	0.187 ± 0.043
Barium (ug/g)	147.501 ± 1.277	157.55 ± 11.776
Lead (ug/g)	11.295 ± 0.509	8.944 ± 0.26
Yttrium (ng/g)	12,698.928 ± 407.103	10,564.972 ± 859.474
Lanthanum (ng/g)	19,417.45 ± 158.025	17,625.097 ± 828.689
Cerium (ng/g)	39,788.945 ± 127.589	35,260.052 ± 1,547.805
Praseodymium (ng/g)	4,702.897 ± 192.188	4,383.724 ± 180.095
Neodymium (ng/g)	16,910.043 ± 406.391	15,203.162 ± 757.149
Samarium (ng/g)	3,434.599 ± 59.596	2,975.569 ± 78.92
Europium (ng/g)	746.217 ± 2.630	570.379 ± 63.172
Gadolinium (ng/g)	3,279.744 ± 32.84	2,710.612 ± 323.11
Terbium (ng/g)	426.651 ± 7.626	362.268 ± 38.928
Dysprosium (ng/g)	2,4923.255 ± 165.455	2,096.818 ± 319.179
Holmium (ng/g)	466.094 ± 12.131	406.657 ± 46.3
Erbium (ng/g)	1,223.029 ± 26.052	1,120.618 ± 140.809
Thulium (ng/g)	166.696 ± 1.649	155.306 ± 20.786
Ytterbium (ng/g)	990.153 ± 27.65	941.325 ± 137.976
Lutetium (ng/g)	142.819 ± 2.925	133.576 ± 16.8
Thorium (ng/g)	4,047.705 ± 397.885	2,377.607 ± 372.533
Uranium (ng/g)	2,266.929 ± 121.677	1,112.706 ± 175.681
Boron (ug/g)	1.474 3.057	9.622 ± 1.249
Silicon (ug/g)	362.145 ± 72.402	303.208 ± 41.356
Phosphorus (ug/g)	2,530.613 ± 277.387	2,045.458 ± 373.219
Titanium (ug/g)	478.642 ± 183.808	773.036 ± 133.655
Germanium (ug/g)	28.324 ± 0.397	16.653 ± 1.631
Zirconium (ug/g)	5.452 ± 1.358	10.441 ± 1.295
Niobium (ug/g)	0.327 ± 0.068	0.56 ± 0.078
Tantalum (ug/g)	0.007 ± 0.001	0.008 ± 0.001
Tungsten (ug/g)	0.108 ± 0.009	0.303 ± 0.26
Rhenium (ug/g)	0.003 ± 0.001	0.003 ± 0.001

Figure 1. Composition of identified microbial community members present in un-inoculated soil from leafy green field in Yuma, AZ after two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma soil samples. “Other” represents all members that are individually present in less than 0.1%.

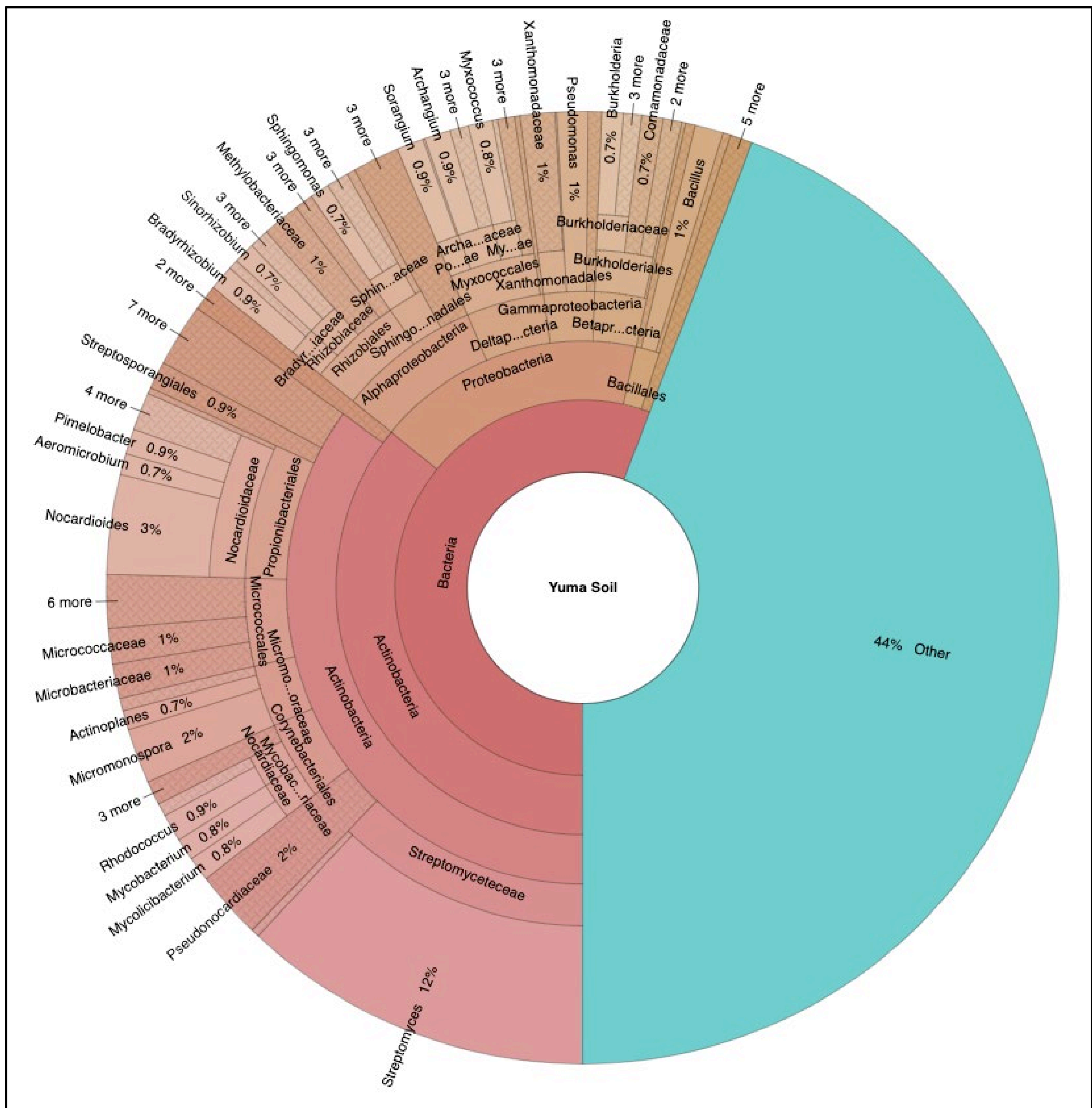

Figure 2. Composition of identified microbial community members present soil from leafy green field in Yuma, AZ after *Escherichia coli* O157:H7 inoculation and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma soil samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 3. Composition of identified microbial community members present soil from leafy green field in Yuma, AZ after *Salmonella* inoculation and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma soil samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 4. Composition of identified microbial community members present soil from leafy green field in Yuma, AZ after *Listeria* inoculation and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma soil samples. “Other” represents all members that are individually present in less than 0.1%.

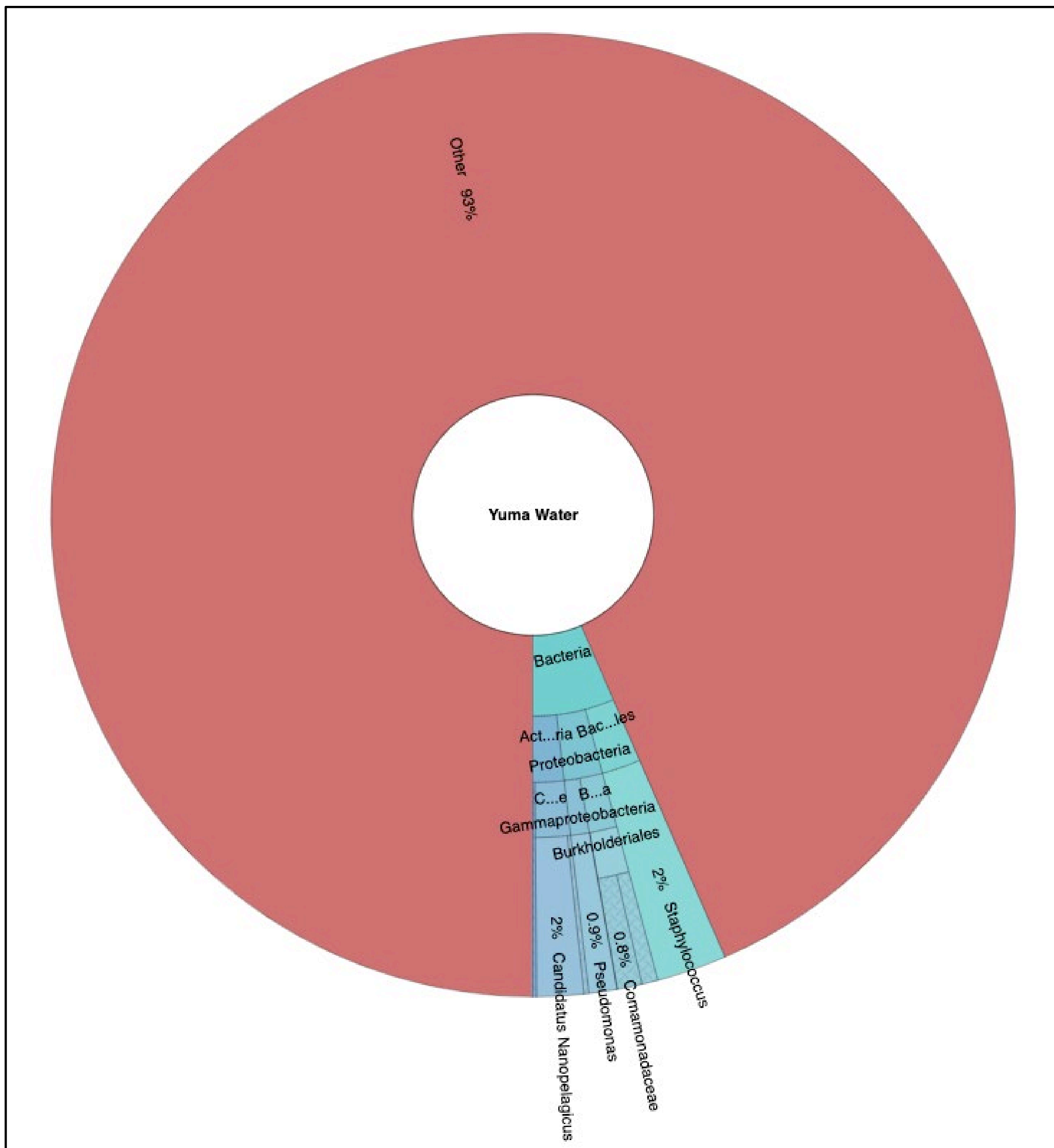

Figure 5. Composition of identified microbial community members present in un-inoculated irrigation canal water from leafy green field in Yuma, AZ and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

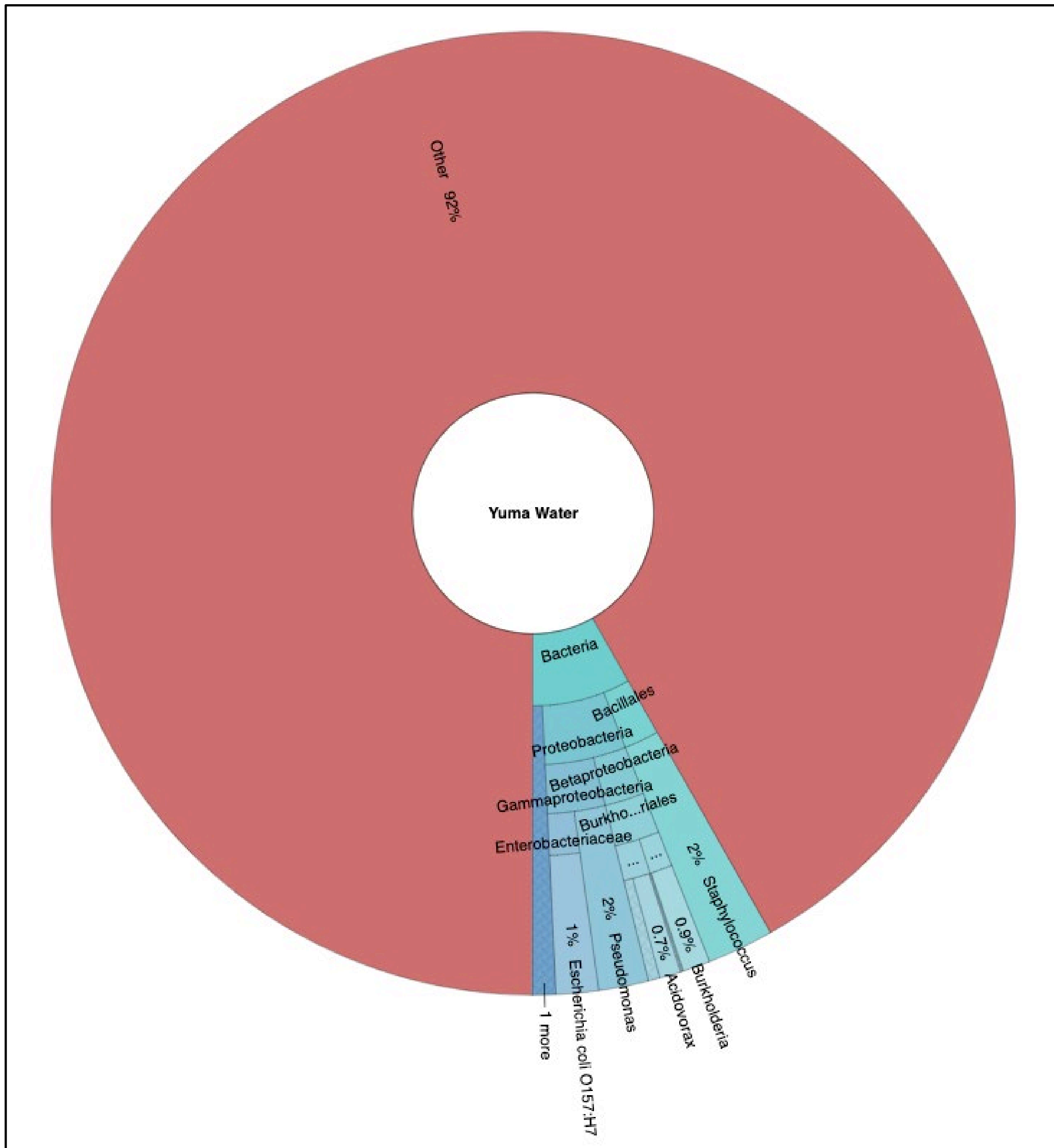

Figure 6. Composition of identified microbial community members present in irrigation canal water from leafy green field in Yuma, AZ inoculated with *Escherichia coli* O157:H7 and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

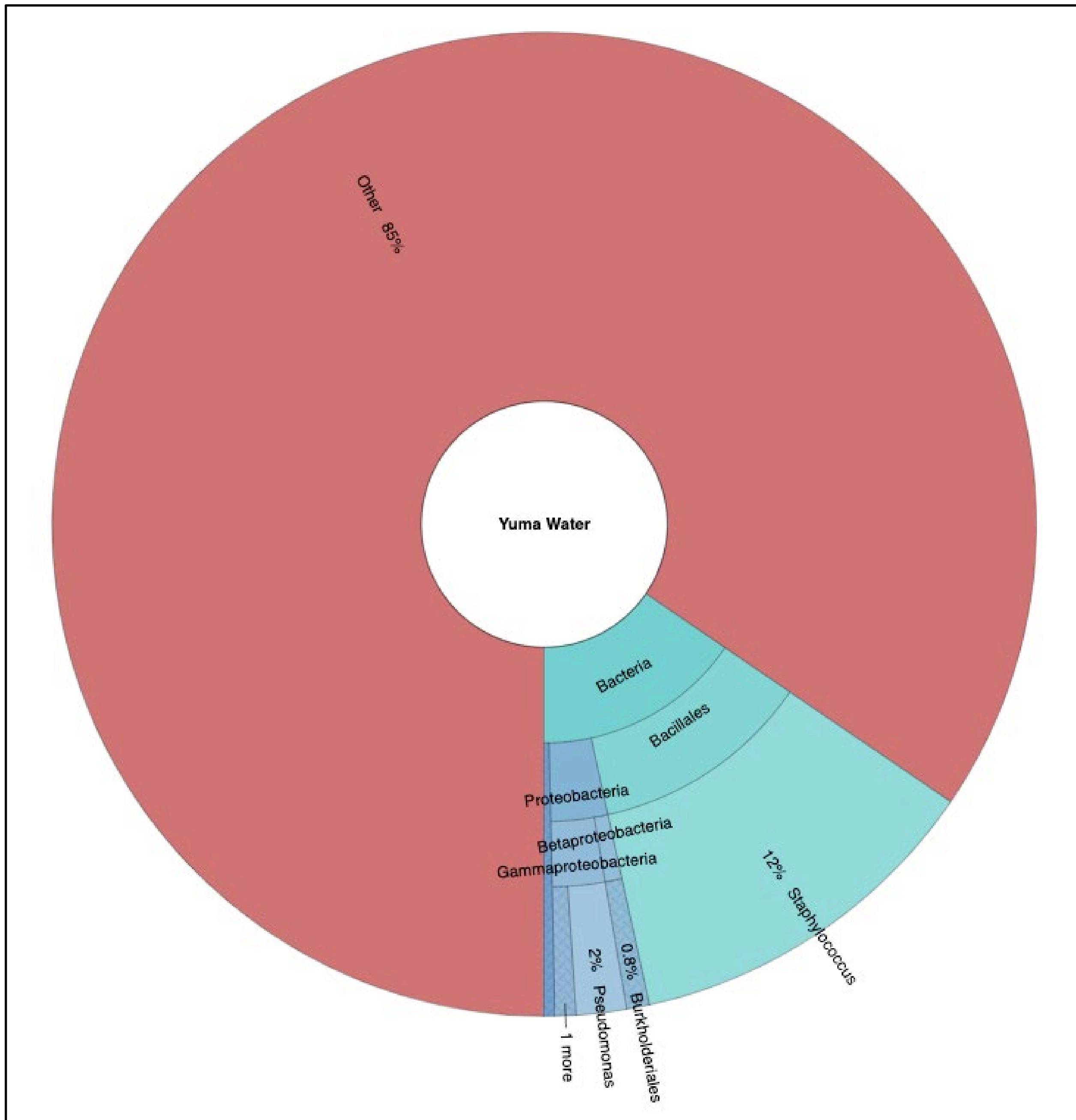

Figure 7. Composition of identified microbial community members present in irrigation canal water from leafy green field in Yuma, AZ inoculated with *Salmonella* and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma irrigation water samples. "Other" represents all members that are individually present in less than 0.1%.

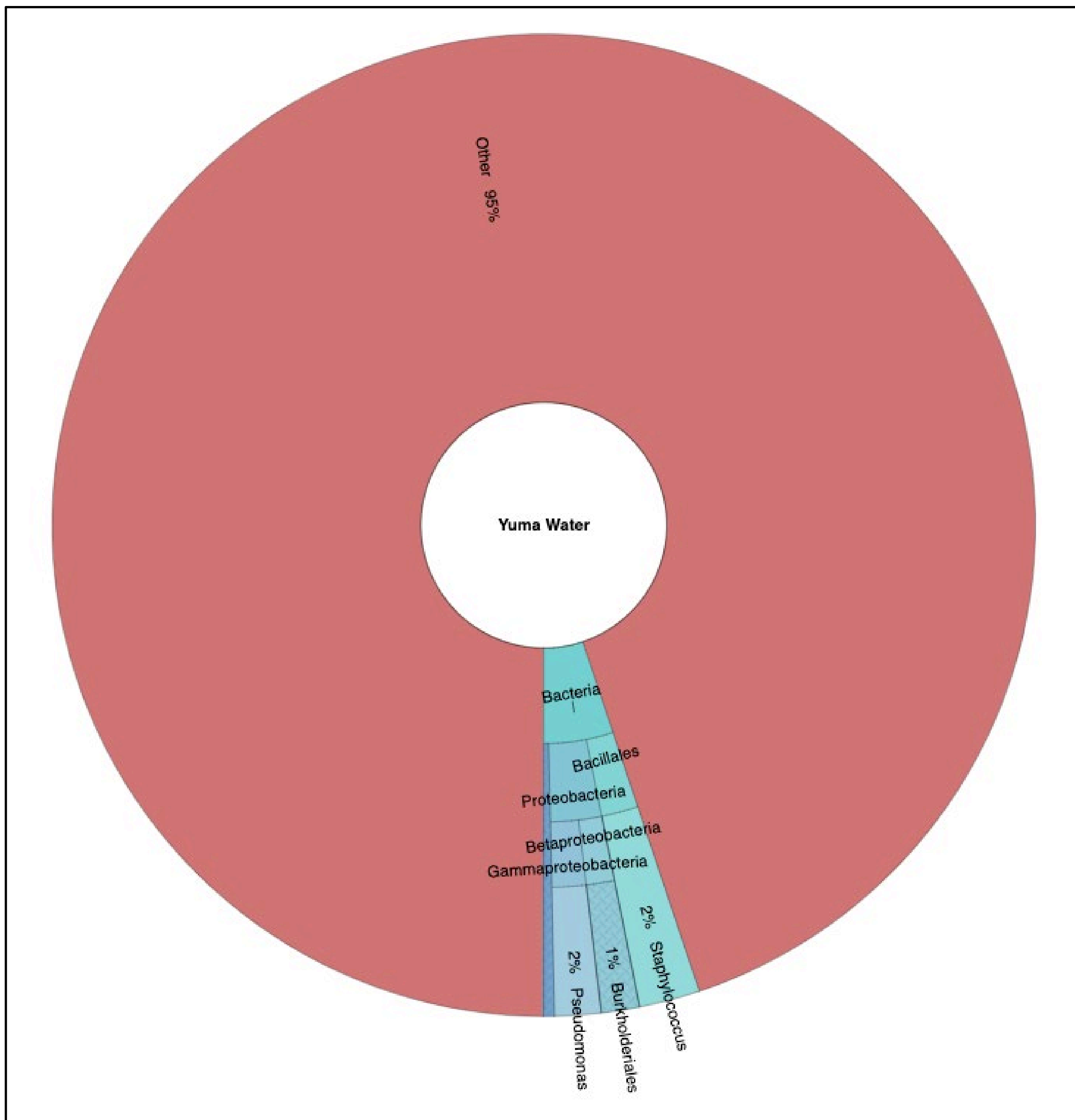

Figure 8. Composition of identified microbial community members present in irrigation canal water from leafy green field in Yuma, AZ inoculated with *Listeria* and two weeks kept under Yuma environmental conditions. All the same community members are presented in corresponding Yuma irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 9. Composition of identified microbial community members present in un-inoculated soil from leafy green field in Salinas, CA and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas soil samples. “Other” represents all members that are individually present in less than 0.1%.

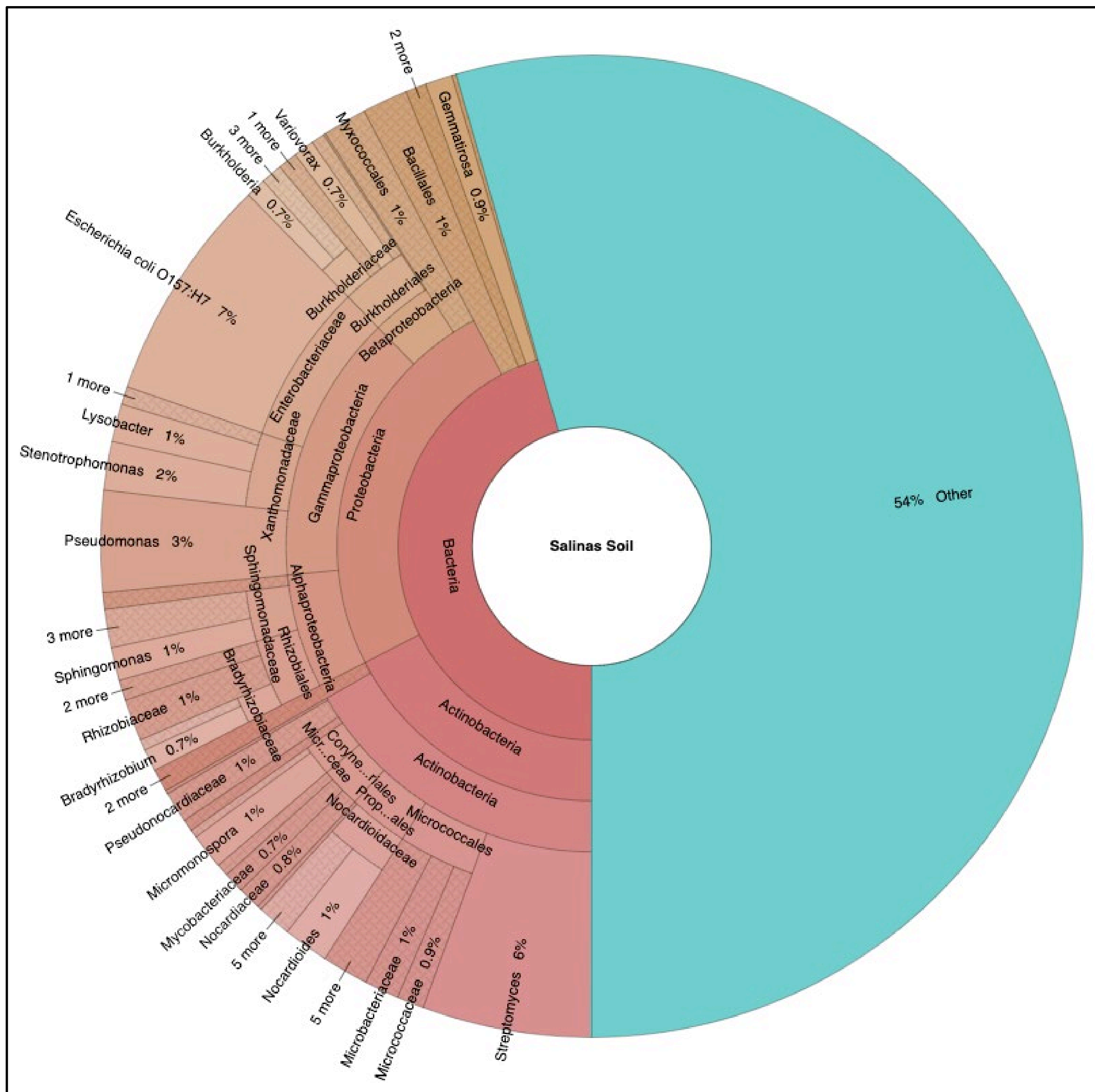

Figure 10. Composition of identified microbial community members present in soil from leafy green field in Salinas, CA inoculated with *Escherichia coli* O157:H7 and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas soil samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 11. Composition of identified microbial community members present in soil from leafy green field in Salinas, CA inoculated with *Salmonella* and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas soil samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 12. Composition of identified microbial community members present in soil from leafy green field in Salinas, CA inoculated with *Listeria* and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas soil samples. “Other” represents all members that are individually present in less than 0.1%.

Figure 13. Composition of identified microbial community members present in un-inoculated irrigation water from leafy green field in Salinas, CA and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

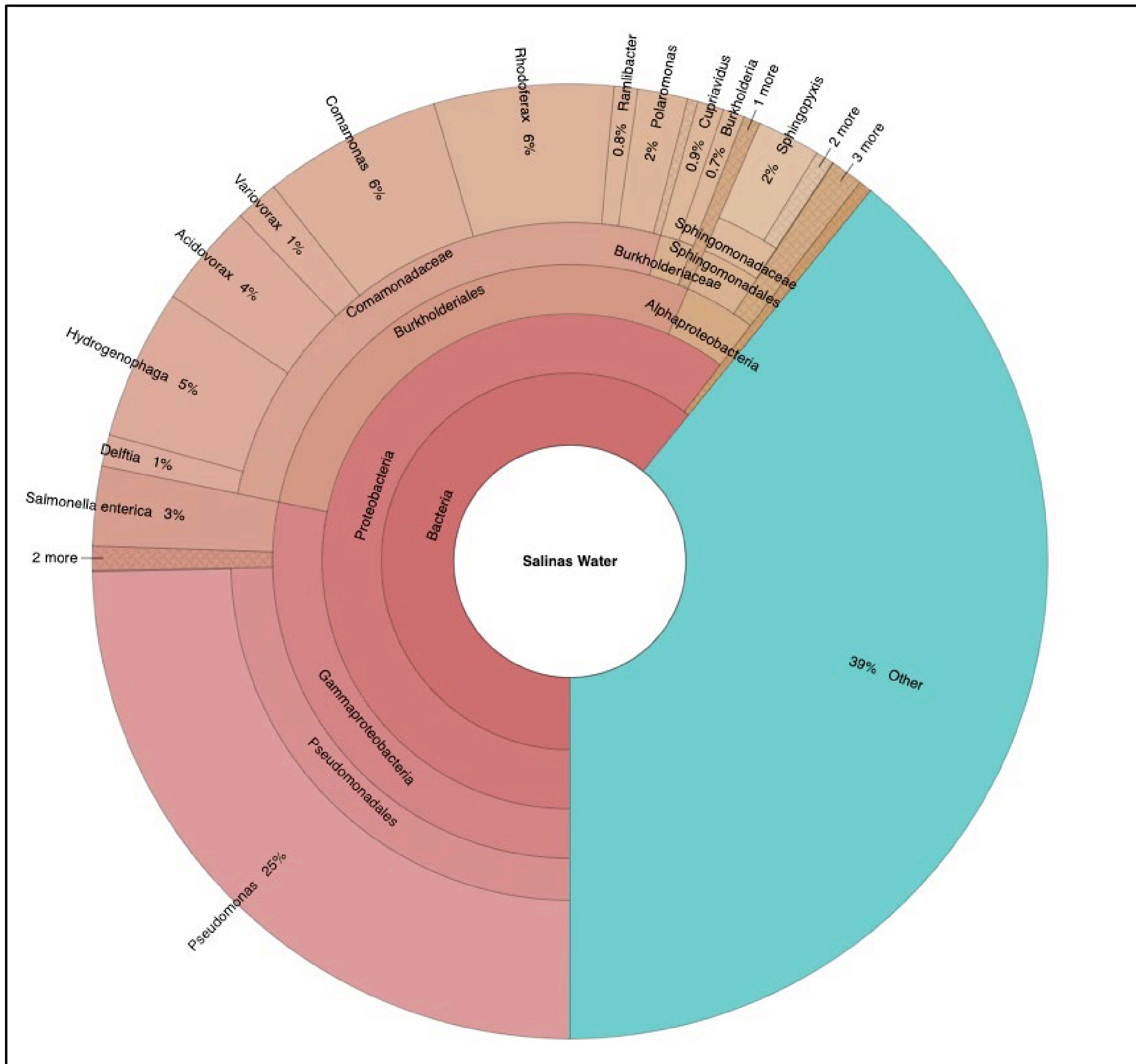

Figure 14. Composition of identified microbial community members present in irrigation water from leafy green field in Salinas, CA inoculated with *Salmonella* and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

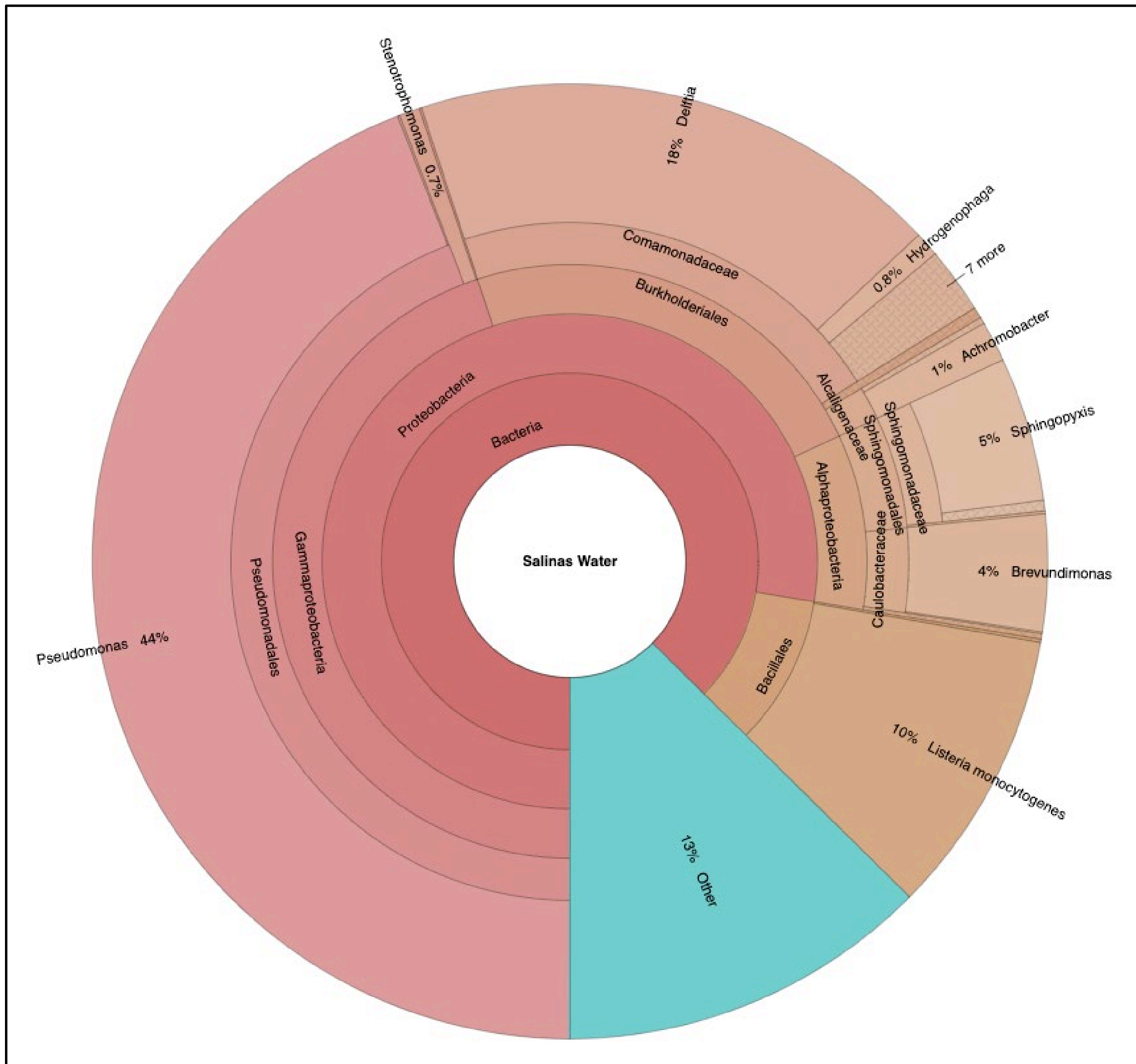

Figure 15. Composition of identified microbial community members present in irrigation water from leafy green field in Salinas, CA inoculated with *Listeria* and two weeks kept under Salinas environmental conditions. All the same community members are presented in corresponding Salinas irrigation water samples. “Other” represents all members that are individually present in less than 0.1%.

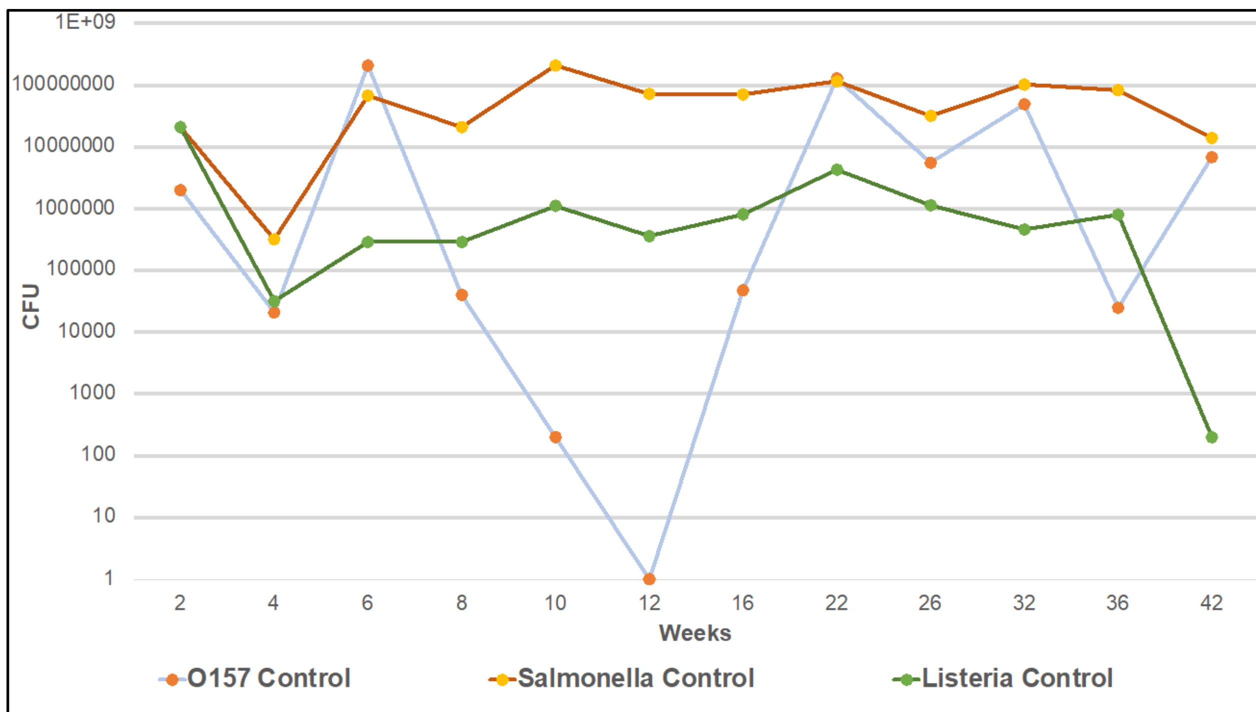

Figure 16. Quantification of culturable *Salmonella*, *Escherichia coli* O157:H7 and *Listeria* control microcosm cultures in buffered peptone water kept under the monthly average temperature, humidity and amount of daylight for Yuma, AZ, which equates to the months of September to mid-June.

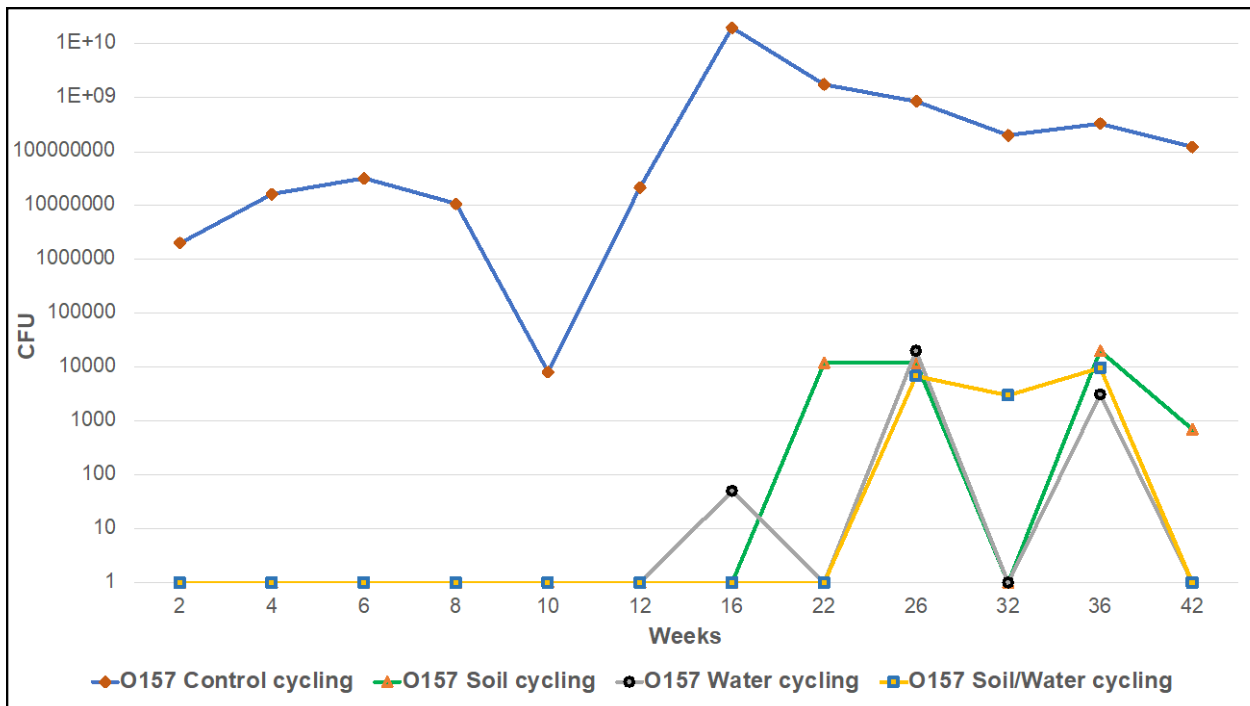

Figure 17. Quantification of culturable *Escherichia coli* O157:H7 that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Yuma, AZ, which equates to the months of September to mid-June.

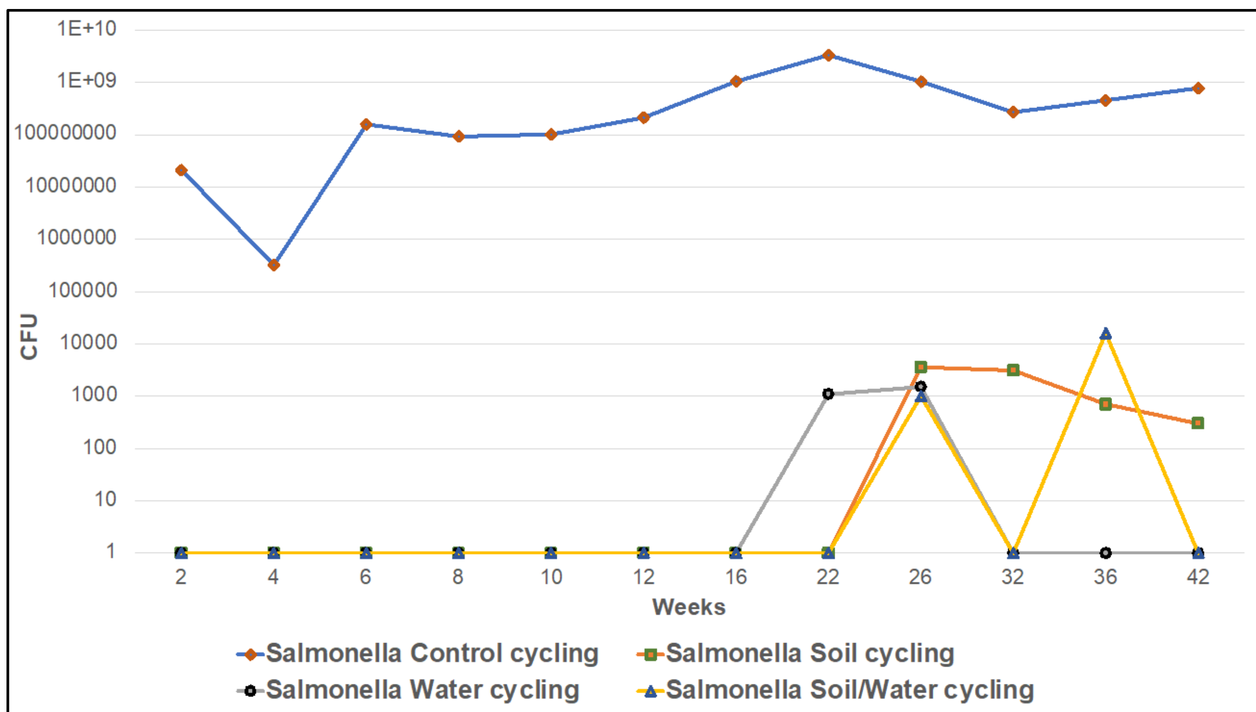

Figure 18. Quantification of culturable *Salmonella* that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Yuma, AZ, which equates to the months of September to mid-June.

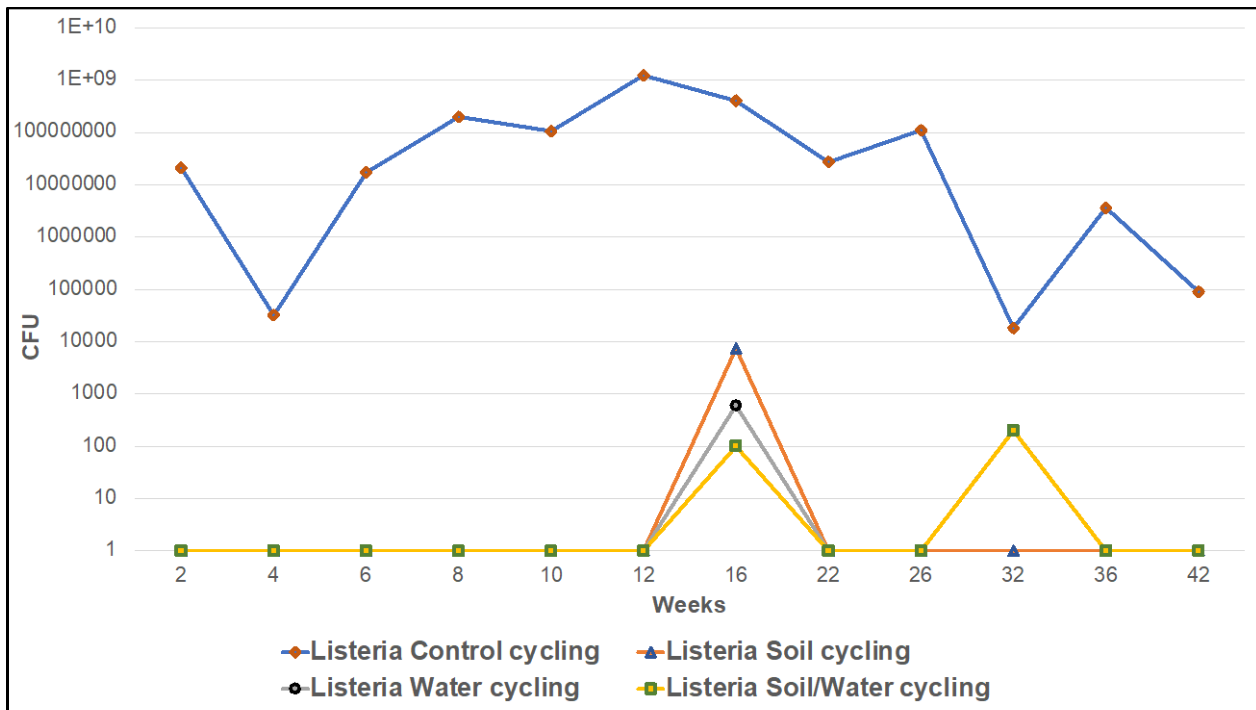

Figure 19. Quantification of culturable *Listeria* that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Yuma, AZ, which equates to the months of September to mid-June.

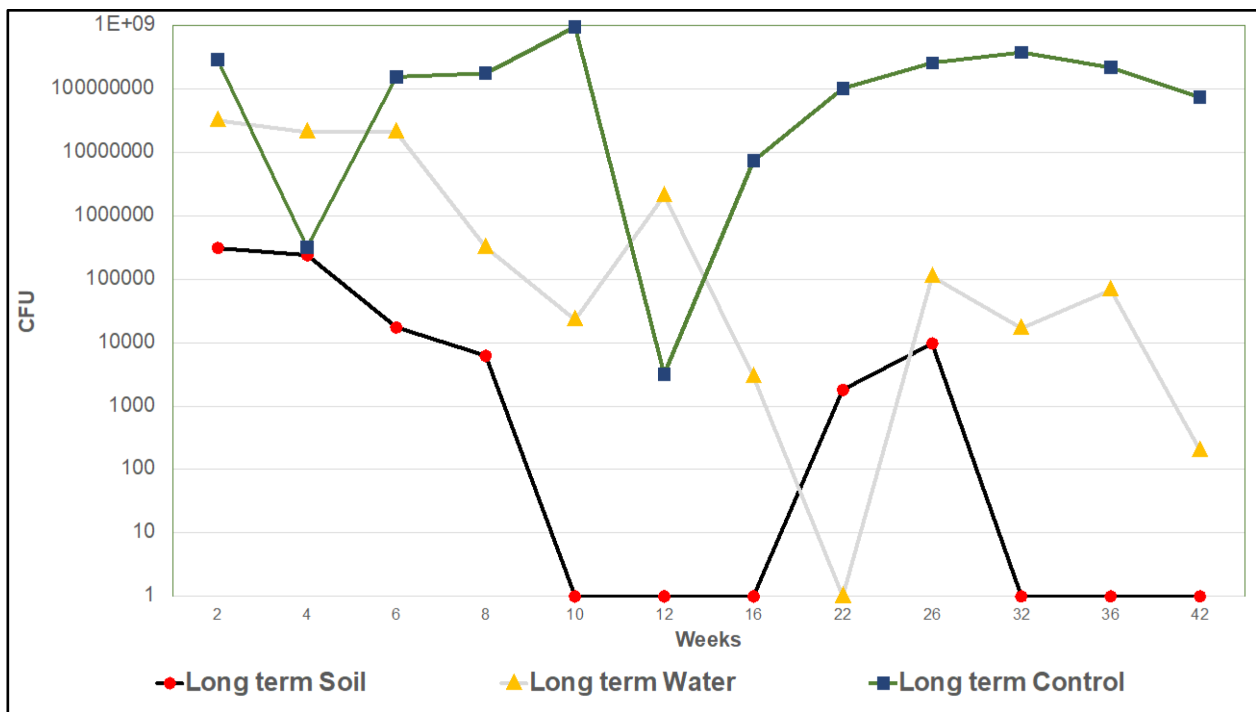

Figure 20. Quantification of culturable *Escherichia coli* O157:H7 in soil, irrigation water, and control microcosm cultures kept long-term under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.

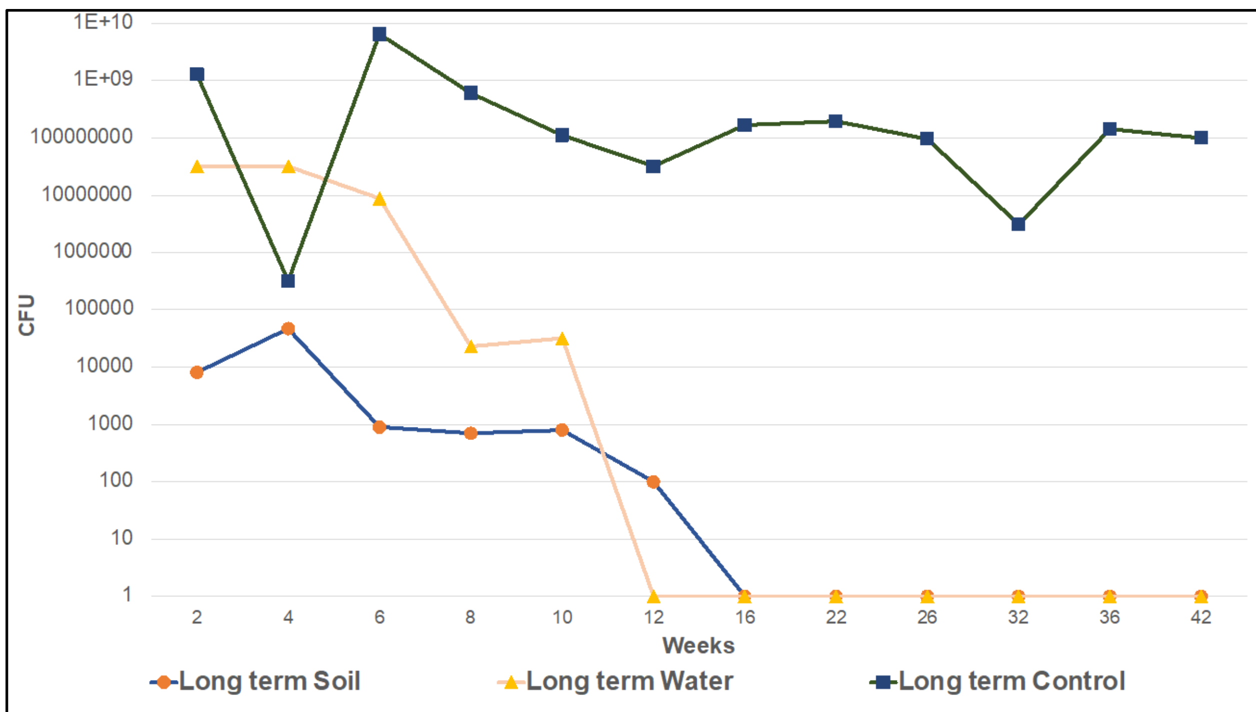

Figure 21. Quantification of culturable *Salmonella* in soil, irrigation water, and control microcosm cultures kept long-term under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.

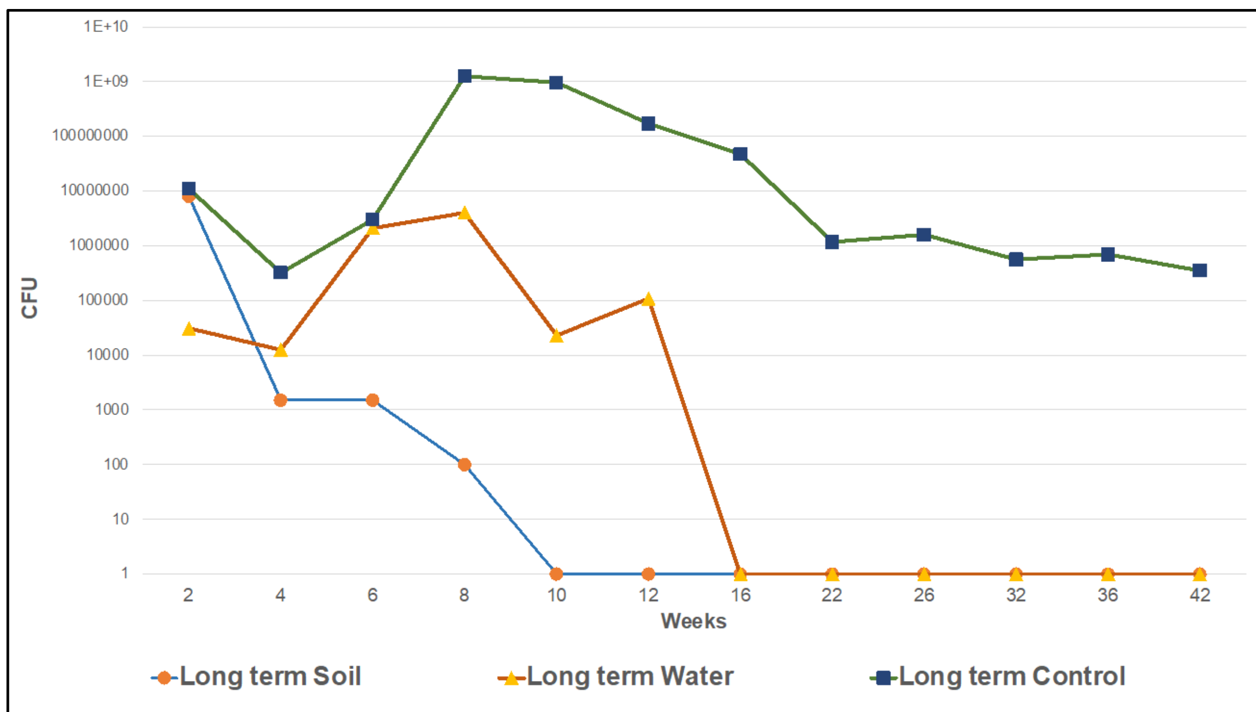

Figure 22. Quantification of culturable *Listeria* in soil, irrigation water, and control microcosm cultures kept long-term under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.

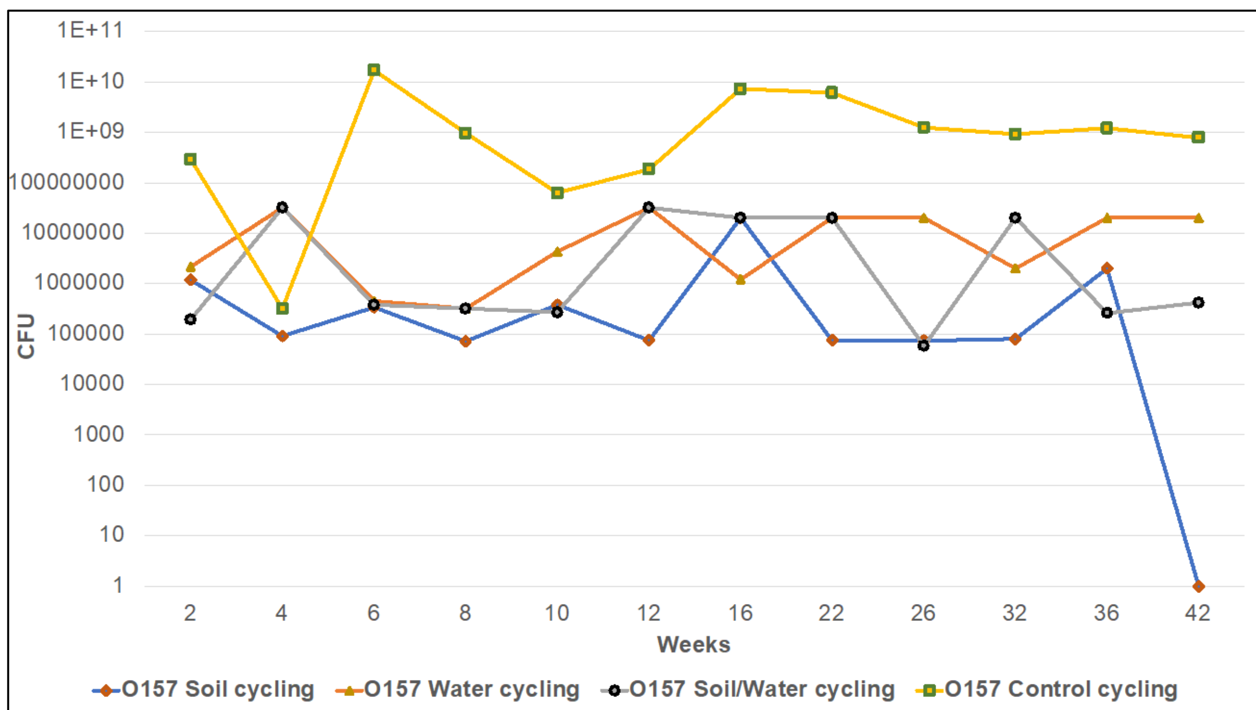

Figure 23. Quantification of culturable *Escherichia coli* O157:H7 that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.

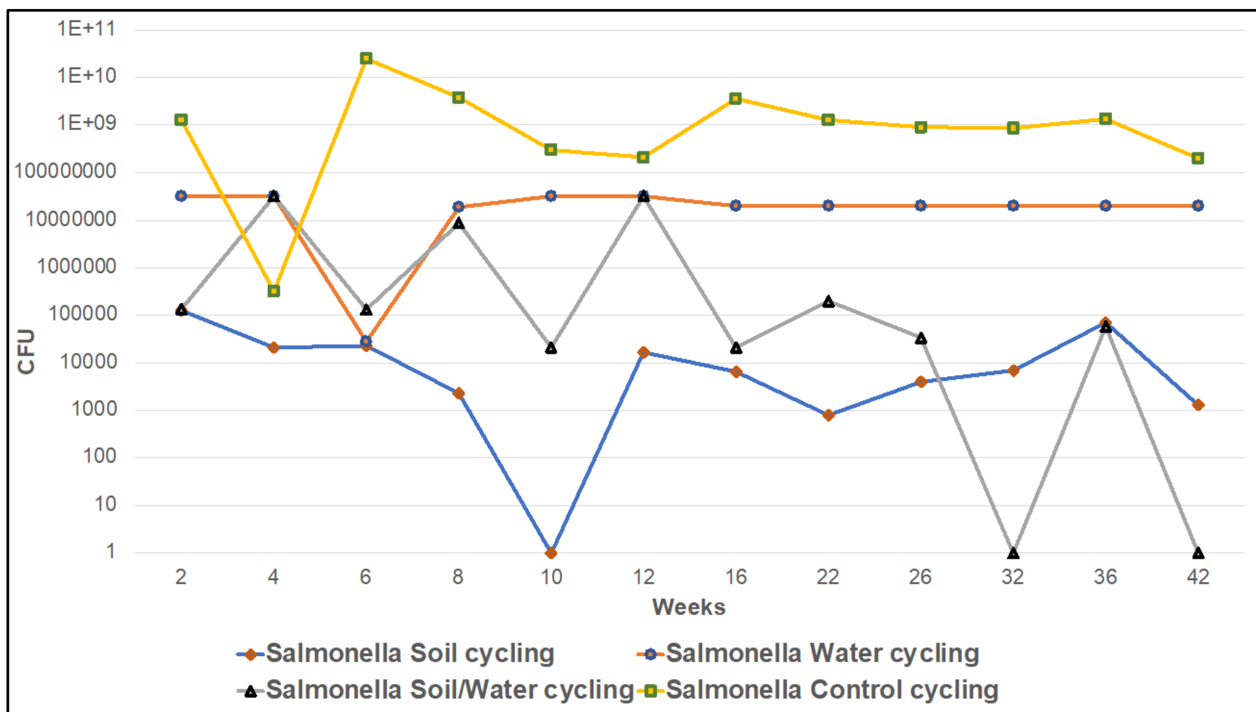

Figure 24. Quantification of culturable *Salmonella* that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.

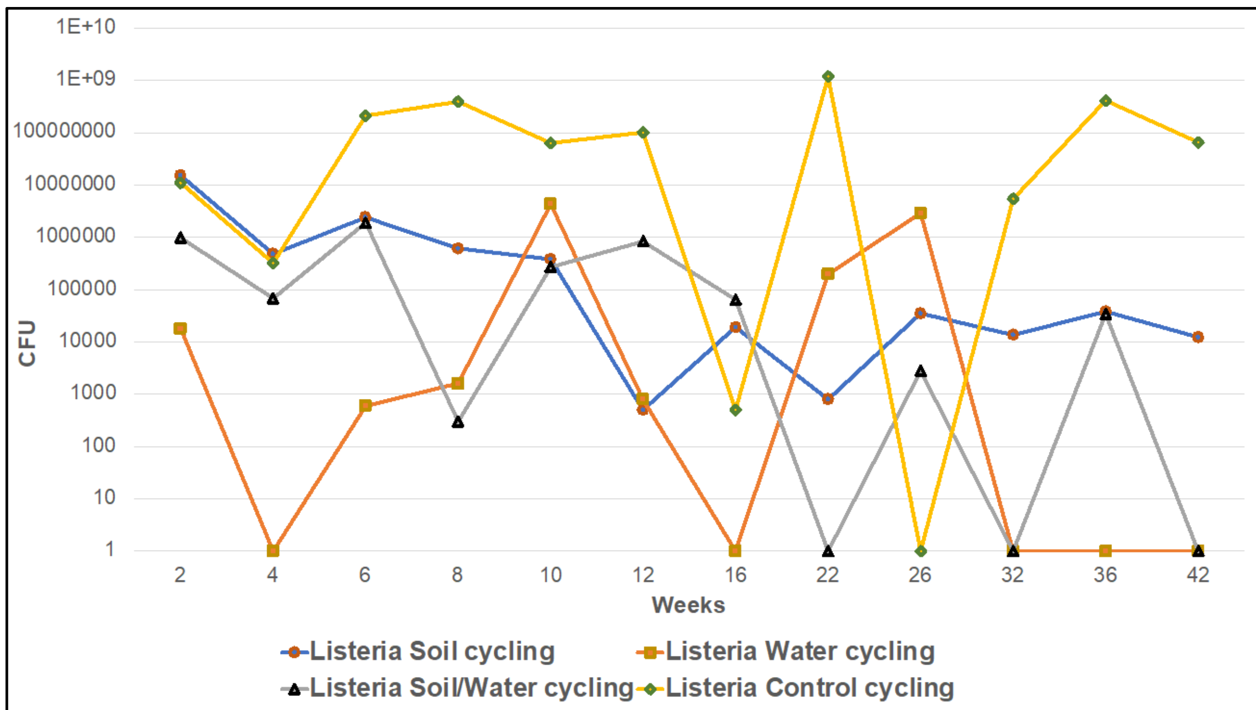

Figure 25. Quantification of culturable *Listeria* that were cycled between new soil, irrigation water, soil and water, or control microcosms every two weeks during the 42 weeks of the project, and kept under the monthly average temperature, humidity and amount of daylight for Salinas, CA, which equates to the months of December to mid-September.