

# UC DAVIS

---

## CENTER FOR PRODUCE SAFETY

### CPS - CALIFORNIA LEAFY GREENS RESEARCH PROGRAM

### FINAL PROJECT REPORT, DUE APRIL 30, 2010

#### Project Title

A high-throughput, culture-independent approach to identify index and indicator species for *E. coli* O157:H7 contamination

#### Project Period

April 1, 2009 through March 31, 2010

#### Principal Investigator

Gitta Coaker  
Department of Plant Pathology  
University of California, Davis  
glcoaker@ucdavis.edu  
530.752.6541

#### Co-Principal Investigators

Johan Leveau  
Department of Plant Pathology  
University of California, Davis  
jleveau@ucdavis.edu  
530.752.5046

Trevor Suslow  
Department of Plant Science  
University of California, Davis  
tvsuslow@ucdavis.edu  
530.754.8313

#### Objective

Objective 1: Use high-throughput DNA sequencing technologies to compare microbial communities associated with lettuce produced in the Salinas, Imperial, and Yuma districts during the season cycle.

## **Original Project Objectives:**

This project had a single objective, listed below:

Use high-throughput DNA sequencing technologies to compare microbial communities associated with lettuce produced in the Salinas, Imperial, and Yuma districts during the seasonal cycle.

## **Outcomes and Accomplishments of the Project:**

### ***Sample collection and high-throughput sequencing***

The objective of the proposed research is to use a high-throughput DNA sequencing approach to identify bacterial species present on Romaine lettuce leaves grown in the Salinas, Imperial, and Yuma Districts. We hypothesize that the microbial community associated with lettuce leaves can play a key role in the ability of *E. coli* O157:H7 to survive and persist on leaves. Some bacteria may be inhibitory, while other bacteria may promote survival of human pathogens. In addition, it is likely that other, more abundant, bacteria may be useful as index or indicators for the establishment of *E. coli* O157:H7 on plants in the field. While routine pre-harvest testing of lettuce and leafy greens for the presence of *E. coli* O157:H7 and related enterohaemorrhagic *E. coli* (EHEC) has greatly increased over the past two years, there is both the desire and increasing data-based justification to move to a more seasonally focused and predictive screening system. Based on post-contamination survival and growth potential under ideal conditions, it is untenable that an early warning system for *E. coli* O157:H7 outbreaks will be based on direct detection of EHEC. Instead, development of such systems must rely on other, indirect, indicators for the establishment of *E. coli* O157:H7 on plants in the field. We have used a high-throughput sequencing approach to identify bacteria associated with field-grown Romaine lettuce during different seasons and locations. In the coming year, we are focusing on collecting samples from fields that have identifiable risk factors or actual positives for *E. coli* O157:H7 or related EHEC. This information will be used to provide a benchmark for important bacteria associated with leafy greens and highlight particular bacterial species with a promise as index/indicator organisms as well as biocontrol agents.

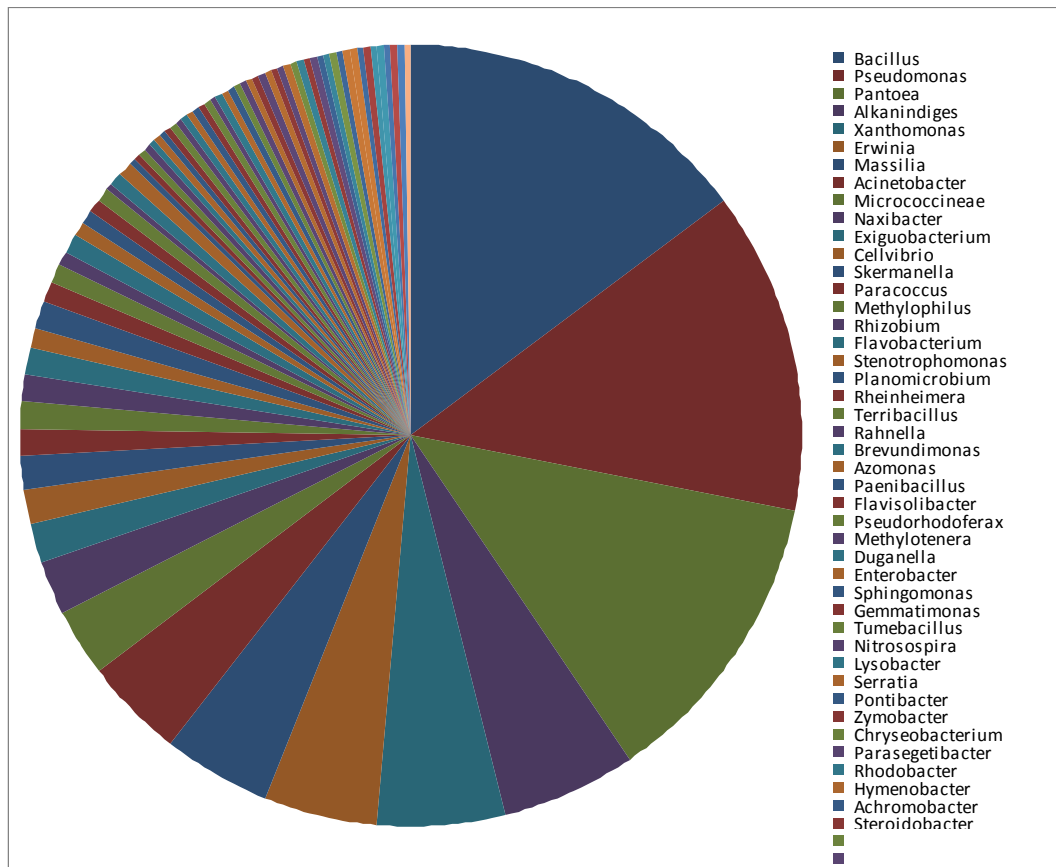
We proposed to collect lettuce samples from 14 fields from Salinas and 16 fields from Imperial and Yuma Districts. We were able to collect samples from 27 fields from Salinas during the spring and summer season and 16 fields from Imperial/Yuma Districts during the winter season. For each field, two samples were collected. Per field, the two samples were taken from the opposing corners within a two-acre plot central to the field. One sample consists of four pooled lettuce heads. The selection of fields during each season was achieved by communicating with cooperative growers. All the sampling was coordinated through Trevor Suslow, a Co-PI on this proposal. For each sample we have extracted high quality microbial DNA from the surface of Romaine lettuce leaves and quantified culturable bacterial population sizes. We have also collected information on climatic conditions from public databases (temperature profiles (air and soil), rainfall, wind speed and direction, relative humidity) as well as site-specific irrigation information. A total of 44 samples (22 fields) were selected for high-throughput DNA sequencing and we are currently in the process of sending an additional 44 samples for sequencing. We had originally anticipated sequencing all 88 samples at once in February of 2010. However, the opportunity presented itself to sequence half the samples at an earlier date. We decided to sequence 44 of the samples representing the majority of the Salinas samples in December 2009 in order to be able to start analyzing the resulting information. We have the second set of 44 samples extracted and are currently waiting for an open slot to sequence these samples at the

University of Nebraska, CAGE facility. We anticipate that the remaining DNA samples will be sequenced in the next 2-4 weeks.

During this funding period, we have also troubleshooted conditions to effectively isolate high-quality bacterial DNA from Romaine lettuce leaves. It is very important to have high quality DNA in order to be able to complete DNA sequencing. We PCR amplified the 16S ribosomal RNA (rRNA) gene for high-throughput sequencing. Highly variable portions of 16S rRNA gene provide unique signatures for each bacterial species and useful information about relationships between them. Universal PCR primers can be designed to recognize these conserved bacterial 16S rRNA gene sequences (16S rDNA) and used to amplify variable, diagnostic regions of the gene. In order to identify the microbial communities associated with lettuce, we amplified 16S rDNA from our DNA samples.

One unexpected problem we encountered was that all published PCR primers for high-throughput DNA sequencing of 16S rDNA also amplify plant DNA. When we used these primers on our samples extracted from Romaine lettuce, the majority of the PCR product belonged to lettuce and not to bacteria. There is extensive similarity between the bacterial 16S rRNA gene and the lettuce chloroplast 16S rRNA and mitochondrial 18S rRNA genes. Therefore, we spent a considerable amount of time designing and testing various primer sets that will only amplify bacterial DNA and will not amplify lettuce DNA. We were successful in designing and testing a primer set that met these conditions (799F/1492R). We anticipate that this primer set will be very useful for other scientists who are interested in analyzing microbial diversity on plant surfaces and we are currently in the process of writing a manuscript describing a set of primers that are specific to bacteria, lettuce chloroplast, and lettuce mitochondrial DNA. We have also shared this information with all currently funded CPS researchers, and some researchers have decided to use these primer sets in their own food safety programs. The facility where we have conducted the high-throughput pyrosequencing (University of Nebraska CAGE facility) has also been getting excellent results with this primer set and has been recommending that all their customers working with plants use this primer set.

To date, we have completed sequencing and initial analyses of 44 samples corresponding to 22 fields harvested in the late spring and summer season in Salinas. Samples were sequenced using 454 pyrosequencing on the Genome Sequencer FLX System. We were able to obtain 411,565 high quality sequences for the 44 samples. The median number of sequences we obtained per sample was 9,353. We identified 413 different bacterial genera and found that the five most abundant genera constitute over half of the entire bacterial population (Figure 1). We have also performed statistical analysis such as rarefaction curves that plot the genera observed as a function of sequence sampled. This was done to evaluate the coverage of bacterial diversity obtained by 454 pyrosequencing across 44 lettuce samples. The slope of the rarefaction curve can then be analyzed to determine if we were successful in sequencing the majority of bacterial genera present in the sample or if deeper sequencing will reveal many more bacterial genera. Our results suggest that we have captured most of the major bacterial genera that are associated with lettuce leaves.

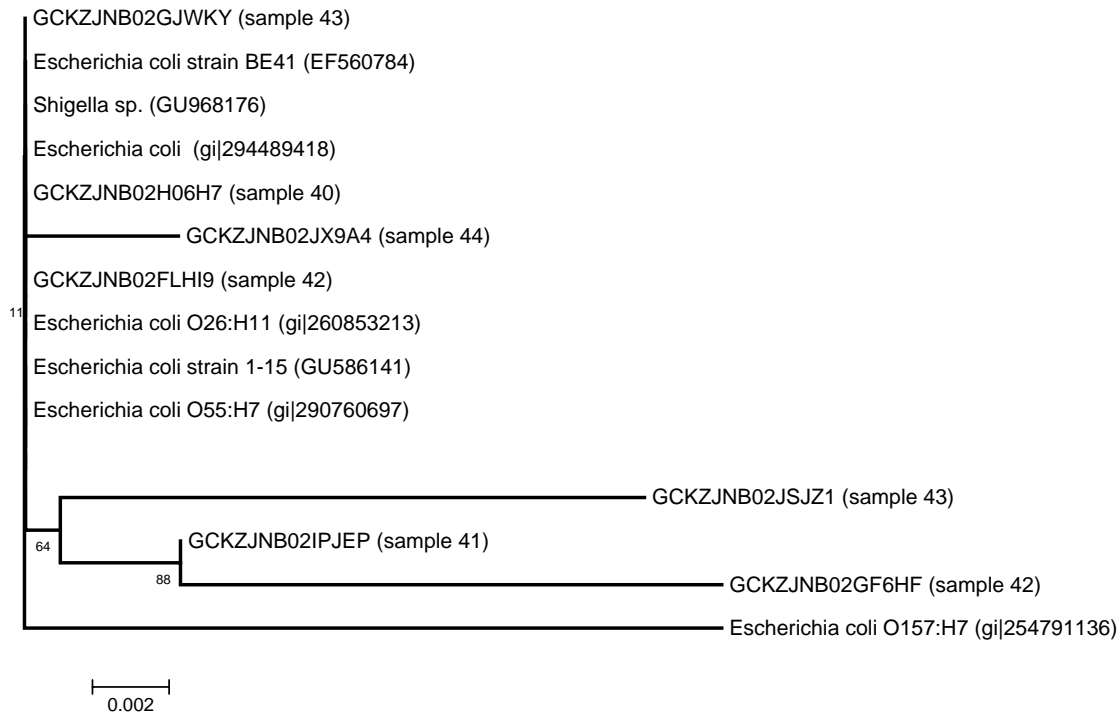


**Figure 1. Bacterial genera identified by pyrosequencing of 44 lettuce samples corresponding to 22 fields.** Note that the 5 most abundant genera make up over half of the population.

We are currently conducting more detailed phylogenetic analysis on the top 10 genera to determine if we can identify distinct bacterial species present within each genus and are investigating if these species change over time (i.e. one species absent under lower risk in the spring season, but present under higher risk in the summer season). This will be done by comparing the sequences we obtained for these genera with other known species from the same genera present in public databases. We anticipate completing this analysis within the next two weeks. Once the major species have been identified, we will analyze the data to determine if there is a correlation between the climatic conditions, geographic location, microbial community composition, and individual species abundance.

Interestingly, two of the most abundant genera identified in lettuce samples were coliform bacteria: *Pantoea* and *Erwinia*. Our initial analyses reveal that there are roughly 15-20 different species of *Pantoea* and *Erwinia* present in different lettuce samples. We also identified the *Escherichia* genus in 5 lettuce samples. However, the proportion of *Escherichia* was very low (0.004%) in total number of sequences obtained from 44 samples. The low-level of *E. coli* is consistent with observations from Industry, both for generic, O157, and non-O157 STEC/EHEC. In order to find the species designation for these *Escherichia*, we performed a more detailed phylogenetic analysis. Our preliminary analysis identified several of them belonging to generic *E. coli* with DNA identity of 99-100% (Figure 3). None of the *Escherichia* sequences were identical to that of *E. coli* O157. However, several of them are similar (98%) to other pathogenic strains of *E. coli* such as O55:H7 (isolated from infant with diarrhea) and *E. coli* strain 1-15 (isolated from frozen chicken). We cannot unambiguously differentiate between generic *E. coli* and STEC/EHEC *E. coli* based on the sequence data

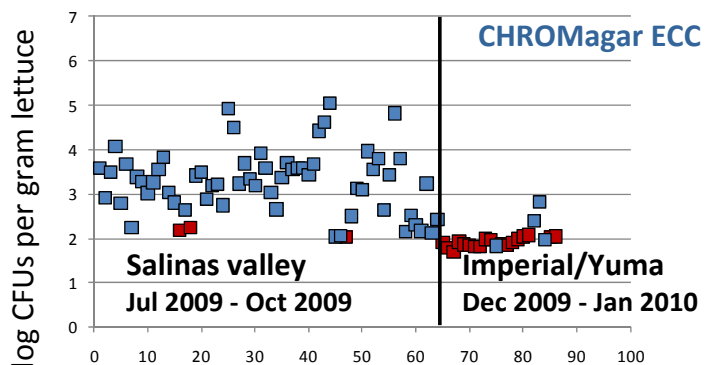
obtained. We plan to use specific PCR primers to determine if any of these samples contain STEC/EHEC during the second year.



**Figure 2. Phylogenetic analysis of *E. coli* related sequences retrieved from lettuce samples with known species present in database.**

### **Culturable Bacteria**

Total culturable bacterial populations were also enumerated for each sample. Samples were plated on 1/10 Tryptic Soy Agar, King’s B Agar, and coliform specific ECC CHROMagar in order to quantify total culturable bacterial populations. Overall, the bacterial populations on 1/10 Tryptic Soy Agar and King’s B Agar were relatively stable within a region. We found lower bacterial population sizes in the winter season in from samples collected from either Imperial or Yuma compared to samples collected from Salinas in the spring or summer season (data not shown). The bacteria that grew on ECC CHROMagar (coliform specific) were present in lower numbers, but there was greater variation in coliform populations Figure 3. Coliform populations were often below the detection limit in the winter season from samples collected from either Imperial or Yuma. Based on these results, coliforms may have more promise as indicator or index organisms than other bacterial species, because their presence and abundance more closely follows the seasonality of *E. coli* outbreaks. We will closely analyze our pyrosequencing dataset to look at coliform bacterial prevalence in the coming months.

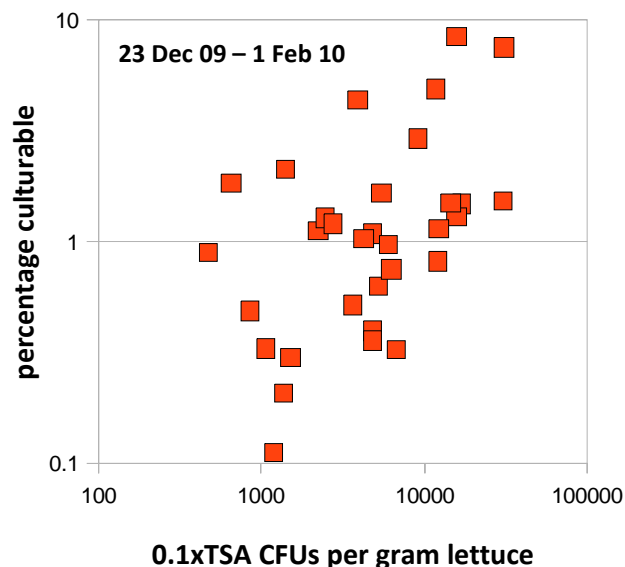


**Figure 3. Coliform population sizes are higher in the Salinas Valley growing seasons than in Imperial/Yuma growing seasons. Red indicates samples for which the numbers of ECC fell below the level of detection.**

### **Quantitative Real-Time PCR**

Although we did not propose to conduct quantitative real-time PCR (Q-PCR), we were able to obtain significant cost savings by having the postdoctoral researcher supported by this grant (Dr. Gurdeep Rastogi) learn how to analyze and process the DNA sequences himself as opposed to outsource this aspect of the research (\$4K Vs \$10K cost). Dr. Rastogi traveled to the University of Nebraska and attended a hands-on workshop to learn how to process pyrosequencing data. We used this savings to cover the cost to start analyzing all our samples by Q-PCR.

We used specific Q-PCR primers to estimate total bacterial populations for all lettuce samples collected from Salinas, Imperial, and Yuma districts in 2009 and early 2010. The results validate our choice to use a culture-independent approach to identify and quantify bacterial populations on lettuce. Q-PCR results indicate that only between 0.1-10% of the total bacterial population associated with lettuce leaves is culturable in standard growth medium such as 1/10 Tryptic Soy Agar and King's B Agar (Figure 4). We have also performed Q-PCR to estimate the proportion of major taxonomic groups of bacteria across all samples (*Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, and *Firmicutes*). The population of specific groups of bacteria such as *Proteobacteria* and *Firmicutes* was occasionally found to vary between two samples of the same field. Pyrosequencing analyses showed that *Proteobacteria* (70%), *Firmicutes* (19%), *Bacteroidetes* (5%) and *Actinobacteria* (2%) were the major taxonomic groups in all 44 lettuce samples. We plan to conduct Q-PCR on all major taxonomic groups within the next year and results will be correlated to the climatic conditions, and inter-taxa relationships will be inferred.



**Figure 4. Only 0.1%-10% of the bacteria on leaves is culturable.** Percent culturable bacteria were calculated by dividing colony counts on 1/10 Tryptic Soy Agar by the real-time PCR estimate of total bacterial abundance.

### **Overall Conclusions and Future Directions**

Overall, we were able to successfully complete the proposed research and downstream analyses. We were also able to perform some additional analyses that were not originally proposed, such as the quantitative real-time PCR (described above). We have also obtained samples and analyzed samples for more fields than originally proposed in the original grant. We have processed all the samples and sequenced half of the samples. The remaining 44

samples are currently in a queue to be sequenced and we anticipate this will be completed in the next 2-4 weeks. The main findings are listed below:

- The microbial population present on lettuce leaves is extremely diverse and 90-99.9% of the population is not culturable in the laboratory.
- A culture-independent approach is necessary to identify microbial populations associated with lettuce leaves.
- The pyrosequencing approach used by our research group was highly successful in identifying and quantifying bacteria associated with lettuce. The DNA sequence information obtained from pyrosequencing can be used to identify bacteria at the species level.
- There are lower levels of culturable bacterial populations present in Imperial and Yuma districts in the winter season than in Salinas during the spring and summer season.
- There are much higher levels of coliform bacteria present in Salinas than in Yuma and Imperial districts in our samples. Often, coliforms were present below our limit of detection in Yuma and Imperial samples.
- Coliform bacteria were more sensitive to environmental changes (higher humidity, higher temperatures) and their population size changed significantly across the summer season in Salinas. Therefore, select coliform species may hold significant promise as index/indicator organisms for *E. coli* O157:H7 contamination.
- The information obtained in the last year can be used to provide a benchmark for important bacteria associated with leafy greens.

**Future directions:** In the next year, we will analyze an additional 88 samples from Salinas (spring and summer season), Imperial (winter season), and Yuma (winter season). In the coming year, we are focusing on collecting some samples from fields that have identifiable risk factors or actual positives for *E. coli* O157:H7 or related EHEC. This information will be used to highlight particular bacterial species with promise as index/indicator organisms as well as biocontrol agents. We are also particularly interested in testing promising bacterial species for utility as index/indicator organisms in future years. In the near term we are also interested in testing to determine if particular bacterial species can influence *E. coli* or *Salmonella* survival and persistence on leafy greens.

### ***How Grant Funds Were Spent***

Grant funds were used to support one postdoctoral researcher, Dr. Gurdeep Rastogi, for 9 months. Dr. Rastogi was jointly supervised by Dr. Gitta Coaker (PI) and Dr. Johan Leveau (Co-PI). Grant funds were used to support a technician in Dr. Trevor Suslow's research group for 6 months involved in collection of Romaine from grower fields. Grant funds (\$2,000) were used to defray the cost of travel to sampling sites as well as CPS meetings. Grant funds (\$16,000) were used to cover the cost of high-throughput DNA sequencing. Funds were used to cover the cost of training Dr. Rastogi in the analysis and processing of pyrosequencing data (\$4,000). The remaining funds (\$9,000) were used to cover the cost of reagents to extract DNA from lettuce leaves, plates and media to quantify culturable bacterial populations, and reagents and kits necessary to complete quantitative real-time PCR analyses.

### ***Publications and Presentations***

Met with industry members at the "Leafy Greens Industry Microbial Data/Trends and Practices Symposium" in Salinas, CA Sept 16-17, 2009.

California Leafy Greens Mid-Year Meeting. 20 min presentation “A culture-independent approach to identify index and indicator species of *E. coli* O157:H7” Seaside, CA. October 13, 2009.

Johan Leveau. Presentation at the Microbial Community and Produce Safety Research Interest Group, Center for Produce Safety, Davis CA, March 9, 2010

A high-throughput, culture-independent approach to identify potential index/indicator species for *E. coli* O157:H7 contamination. Johan Leveau, Gurdeep Rastogi, Trevor Suslow, Gitta Coaker. Presented at the California Leafy Greens Research Board Meeting, Harris Ranch, Coalinga, CA, March 16, 2010.

Paper:

Rastogi G., Tech J.J., Coaker G., and J.H.J. Leveau. A PCR-based toolbox for the quantification of bacterial abundances in plant environments. In preparation for submission to the Journal of Microbiological Methods.