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

Project Investigators:

Principal Investigator
Gitta Coaker
Department of Plant Pathology
University of California, Davis

Co-Principal Investigators
Johan Leveau
Department of Plant Pathology
University of California, Davis

Trevor Suslow
Department of Plant Science
University of California, Davis

SUMMARY:
The objective of the proposed research is to use a high-throughput DNA sequencing approach to identify bacteria present just prior to harvest 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. We have used a pyrosequencing approach to identify bacteria associated with field-grown Romaine lettuce during the winter, spring, and summer growing seasons over a two year period. This information is being used to provide a benchmark for important bacteria associated with leafy greens in the California and Arizona growing regions. This benchmark can be used in the future and compared to the microbial community from contaminated lettuce to highlight potential index and indicator species for *E. coli* O157:H7 contamination.

II. Main Body of Report

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

Project Investigators:

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.

Background
The contamination of lettuce and other leafy greens with enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 has become a serious concern in the major regions of domestic production (California and Arizona) due to several recent outbreaks associated with consumption of both whole and minimally processed product. Leafy greens producers recognize that the predominant period of risk for outbreaks with *E. coli* O157:H7 has been during the main summer and early fall season. There is an increased frequency of *E. coli* O157:H7 detection in July to late September; samples collected from the same locations at other times in the growing season are unlikely to have detectable levels of contamination. Of the more than 20 US outbreaks of *E coli* O157:H7 from lettuce and spinach since 1995, at least nine have been traced to the Central Coast region. In addition, two regionally limited outbreaks in 2008 with processed lettuce were traced back to this region. In contrast, only one outbreak has been linked to the Yuma district (O154 serotype), despite many similarities in production practices, the presence of presumptive risk

factors for contamination and harvesting conditions between the three geographical locations. Currently, the most viable correlative factor for these spikes in detection is consecutive days of temperatures above 24°C (75°F) in combination with overhead irrigation or seasonal rainfall.

An understudied factor in the failure or success of *E. coli* O157:H7 to establish or proliferate on field-grown lettuce leaves is the microbial community associated with these surfaces. These members of the microbial community may thus have utility as index or indicator species. An index organism acts as a benchmark for conditions that favor survival or growth of a pathogen; an indicator is a benchmark for evidence of recent fecal contamination and shares key attributes of persistence and growth with the target pathogen. While routine pre-harvest testing of lettuce and leafy greens for the presence of *E. coli* O157:H7 and related EHEC has greatly increased over the past three 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 early warning systems 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.

Procedures:

In the funded research, we sought to identify promising index and indicator species for *E. coli* O157:H7 contamination and survival on lettuce. High-throughput DNA sequencing technologies were used to profile microbial communities associated with Romaine lettuce leaves (phyllosphere) produced in the Salinas, Imperial, and Yuma districts during the winter to late summer transitions over two production cycles. We also performed culture-based analyses for each sample. This approach enabled us to establish a baseline of microbial communities associated with lettuce. This research is a first step towards taking a culture-independent approach to: (1) Identify which microbial species warrant additional investigation as potential index or indicator species. (2) Investigate how microbial communities and/or microbial ecology contribute to explaining the recurring pattern of outbreaks of *E. coli* O157:H7 on lettuce and leafy greens in the Central Coast region and the apparent absence of outbreak events from the Imperial and Yuma districts.

Field Sampling and DNA sequencing

In the first year of funding (2009-2010), we proposed to collect lettuce samples from 14 fields from Salinas Valley and 16 fields from Imperial and Yuma districts. We were able to collect samples from 40 fields from Salinas during the spring and summer season and 16 fields from Imperial/Yuma districts during the winter season. In the second year of funding (2010-2011), we proposed to collect the same number of samples as above. We were able to collect samples from 24 fields from Salinas during the spring and summer season, and 21 fields from Imperial/Yuma districts during the winter season.

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 and from each head, two outer most and two inner leaves from the fourth leaf circle were picked. The selection of fields during each season was achieved by communicating with cooperative growers. For each sample we have extracted 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.

We initially proposed to use high-throughput DNA sequencing to sequence a total of 176 samples corresponding to 88 fields (44 fields in year 1 and 44 fields in year 2). We have completed the sequencing and data analyses for all samples in year 1. We have sequenced the spring and summer samples for year 2 (24 fields). We have just completed obtaining samples from the 2011 winter season in Imperial and Yuma and are gearing up to complete sequencing from these samples (21 fields). We anticipate that all sequencing should be completed and analyzed within the next 2-3 months.

Results and Discussion:

Results from High Throughput DNA Sequencing

We have finished analyzing DNA sequence data for all samples in year 1. Some general results from these analyses are described in this section. We focused on quantification at different taxonomic levels, particularly at the genus level. We also investigated correlations between bacterial abundance and location, presence of other bacteria, as well as environmental conditions.

We obtained 818,055 bacterial sequences from 88 samples of year 1 and found that the dominant taxonomic groups across all samples were *Proteobacteria*, *Firmicutes*, *Bacteroidetes,* and *Actinobacteria.* Significant variations in relative abundances of bacterial taxa were observed across lettuce samples collected from different geographical regions. Bacterial communities were fairly diverse and varied significantly at genus-level across the lettuce samples. Altogether, 728 distinct bacterial genera were recovered, with an average of about 129 genera per sample. Table 1 describes the 10 most dominant genera found in the Salinas Valley and Imperial and Yuma Disctricts. Out of 728 distinct genera detected on lettuce samples, about 98% were represented in <1% relative abundance. These findings indicate that though the lettuce phyllosphere might contain some abundant genera (*Pseudomonas, Erwinia,* and *Pantoea*), each lettuce samples also harbors a considerable portion of low-frequency genera which makes phyllosphere microbiota distinct and probably specific to each sample. We found specific differences in the distribution of bacterial taxa across different lettuce production regions. For instance, *Bacillus* were markedly overrepresented in Salinas Valley samples where as this genus was present in quite low abundance in majority of the samples from Imperial/Yuma. More specifically, *Bacillus* was predominant in Marina region samples followed by King City, Soledad, Gonzales, and Salinas. In contrast, *Pseudomonas* was present in considerably higher number in majority of the samples from Imperial/Yuma compared to Salinas Valley samples.

Table 1. Dominant genera present in the lettuce phyllosphere.

Genera	% abundance
Pseudomonas	22
Bacillus	10
Pantoea	8
Massilia	7
Xanthomonas	6
Alkanindiges	4
Duganella	4
Erwinia	3
Acinetobacter	3
Naxibacter	2

We were also able to identify relationships between the relative abundances of these major genera and solar radiation and soil temperature. For example, the abundance of *Pseudomonas* decreased as soil temperature increased (R^2 =0.8) while *Pantoea* showed the opposite pattern (R^2 =0.6) (Figure 1a and b). Furthermore, several microbe-microbe interactions interactions were also found to exist between major taxa. For example, a comparison of the relative proportions of *Pantoea* and *Massilia* (R^2 =0.7) suggests a negative correlation between these bacteria across all sampling locations (Figure 2).

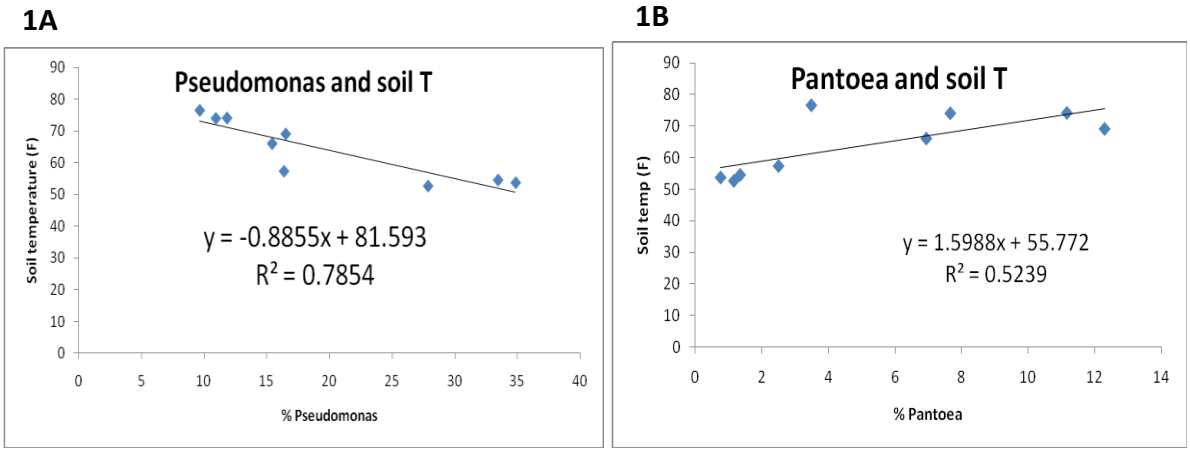

Figure 1. Correlations exist between *Pseudomonas* and *Pantoea* and the soil temperature.

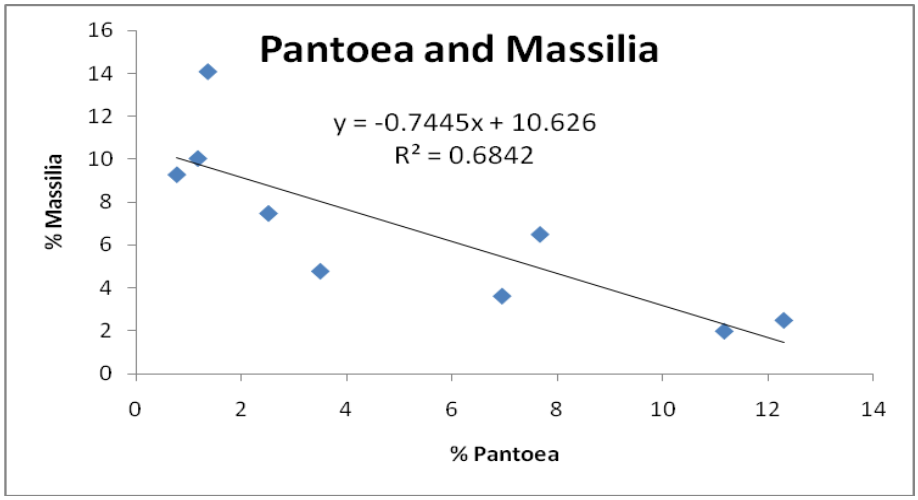

Figure 2. *Pantoea* and *Massilia* are negatively correlated with one another.

The presence of *E. coli* O157:H7 and other Shiga toxin-producing *E. coli* (STEC) strains in 88 lettuce samples from year 1 was tested by PCR targeting the Shiga toxin 2 gene (*stx2*) and *rfbE* genes (responsible for biosynthesis of the O157antigen). We were unable to identify any samples with a positive result, indicating that *E. coli* STEC prevalence was at a very low level (if present at all), in our samples. We also analyzed all DNA sequences for similarity to *E. coli.* Based on the taxonomic and phylogenetic analysis we detected very low levels (<0.02%) of *Escherichia/Shigella* affiliated reads in the whole sequencing data set. In 22 lettuce samples, we found *Escherichia/ Shigella* related reads in the range of 1-30 per sample. Since *Escherichia* and *Shigella* share high DNA similarities in their 16S rRNA sequences and pyrosequencing generated reads are short length (~ 450 bp) therefore it was difficult to predict whether these reads belonged to *Escherichia* or *Shigella* based on phylogenetic and taxonomic analysis. Importantly, the *stx2* and *rfbE* gene based PCR screening of phyllosphere DNA did not give any positive amplification suggesting that either these strains are not pathogenic, or they were present below the limit of detection using amplification of *stx2* and *rfbE.*

Results from Culturable Analyses

Total culturable bacterial populations were also enumerated for each sample. Samples were plated on 1/10 Tryptic Soy Agar (TSA), King's B (KB) agar, and coliform specific ECC CHROMagar in order to quantify total culturable bacterial populations. Overall, the bacterial populations on TSA, KB were relatively stable within a region. We found lower bacterial population sizes in the winter season from samples collected from either Imperial or Yuma during the winter season compared to samples collected from (Figure 3). The bacteria that grew on ECC CHROMagar (coliform specific) were present in lower numbers, but there was greater variation in coliform populations. 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.

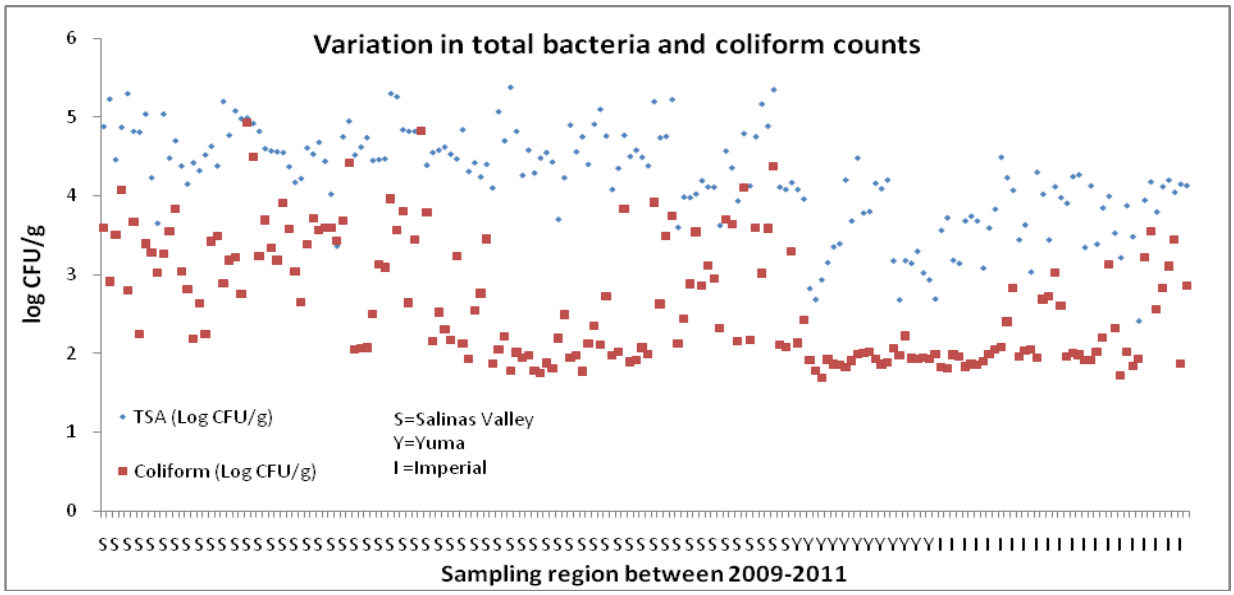

Figure 3. Culturable bacterial counts for all samples in 2009-2011. 2 log colony forming units/gram (CFU/g) is considered our limit of detection. If no coliform bacteria were identified in a particular sample, we inserted this value.

Outcomes and Accomplishments

The proposed research is a continuation of sampling and high-throughput DNA sequencing of bacterial populations that was funded during the April 1, 2009- March 31, 2010 cycle. The primary objective of previous research was to identify bacteria associated with Romaine lettuce grown in the CA Central Coast, Imperial District, and Yuma District. We have successfully completed the primary objective. We are still analyzing the latest round of sequencing data from year 2 as we have just finished sampling from the winter season in Yuma.

Summary of Findings and Recommendations

This study represents the largest analyses of microbial populations associated with leafy produce to date. We now have in-hand a baseline microbial population associated with leafy greens grown in Salinas, Imperial, and Yuma districts. Unfortunately, it was quite difficult to gain access to contaminated fields to obtain samples that were contaminated with *E. coli* or related pathogens. In the future, we aim to perform more detailed sequencing analyses of contaminated fields as part of a "rapid response" program. In this way, we should be able to identify how the microbial population of contaminated fields differs from the baseline microbial population described above.

APPENDICES

Publications and Presentations

Publications:

Rastogi G, Tech JJ, Coaker GL, Leveau JH. A PCR-based toolbox for the culture-independent quantification of total bacterial abundances in plant environments. J Microbiol Methods. 2010 Nov;83(2):127-32.

Presentations:

Coaker, G. 2010. Towards the Identification of Index and Indicator Species for *E. coli* Contamination. California Leafy Greens Research Board. Seaside, CA. Oct 5, 2010.

Coaker, G. 2010. Bacterial Recognition in Plants. 95th Annual Meeting of the Ecological Society of America. Pittsburg, PA. August 6, 2010.

Coaker, G. 2010. A high-throughput, culture-independent approach to identify index and indicator species for *E. coli* O157:H7 contamination. Center for Produce Safety Research Symposium. Davis, CA. June 23, 2010.

Leveau, J. 2010. Leaf surface microbiology: from plant health to food safety. International Seminar of Indonesian Society for Microbiology, Bogor, Indonesia.

Leveau, J. 2010. Molecular approaches to study microbial ecology on plant surfaces. Postharvest Biocontrol workshop, Leesburg VA.

Leveau, J. 2011. A culture-independent approach to identify indicator species for *E. coli* contamination. California Leafy Greens Research Program meeting, Harris Ranch, Coalinga CA.