Investigating Changes in Prevalent Human *Escherichia coli* Strains with Different Collection Methods Using Pyroprinting: A Novel MST Method

http://cplop.cosam.calpoly.edu/

W. Tang, C. J. Sabatini, **E. R. Neal**, K. McGaughey, A. Goodman, A. Dekhtyar, M. Black, C. Kitts

California Polytechnic State University San Luis Obispo, CA

http://cplop.cosam.calpoly.edu/

Abstract

Microbial Source Tracking (MST) is the science of identifying the source of a bacterial species to its host species. This practice is used frequently in applied science to identify sources of contamination in the environment like beaches or lakes, for example, but MST techniques can also be used to study microbial population dynamics in specific organisms. In this study, a new MST method was developed to investigate the variation of *Escherichia coli* strains in humans. The new approach uses pyrosequencing to generate DNA fingerprints (or pyroprints) based on the sequences from two polymorphic regions within the ribosomal RNA operon of *E. coli.* Seven copies of the ribosomal RNA operons are present in the *E. coli* genome and each possesses two highly variable regions called Intergenic Transcribed Spacers (ITS). These regions of DNA are non-coding, and thus are able to accumulate nucleotide changes. The pyroprinting MST method developed by our lab capitalizes on these differences to create patterns of data by simultaneously sequencing all seven copies of each ITS: the 16-23S rRNA region (ITS 1) and 23-5S rRNA region (ITS 2). A "match" between *E. coli* isolates is determined when the patterns are nearly identical at both ITS 1 and 2. A cluster developed by matching isolates is considered to be a unique strain.

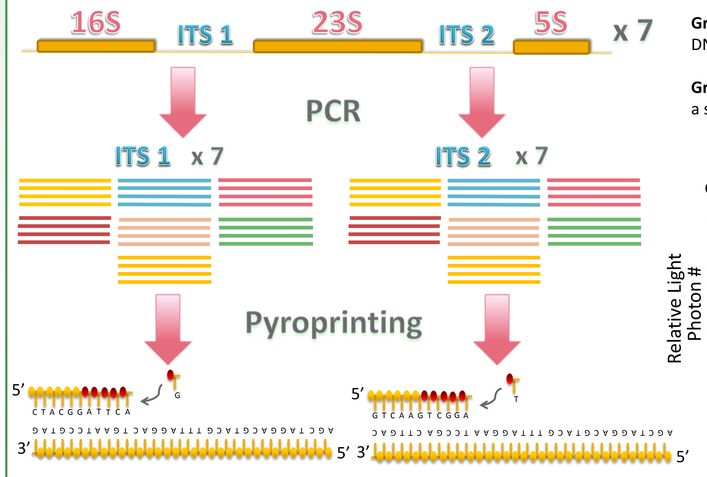
Specifically, this study investigated *E. coli* strain diversity in three humans over a six-month time period and between different sampling methods. *E. coli* isolates were collected and confirmed from a series of anal and fecal swabs taken once a month for each of the six months. The swab series consisted of four distinct sampling strategies. The results obtained indicated that different sampling techniques may detect different strains. The data obtained from the temporal study confirmd the hypothesis that individuals host one or two dominant strains. Interestingly, the population changed such that different strains were dominant from month to month. Additionally, minor strains and number of minor strains varied monthly for most individuals. Lastly, the amount of diversity of *E. coli* populations was different between individuals.

Introduction

Escherichia coli is a well-studied commensal bacterium inhabiting the mammalian intestinal tract and is an indicator of fecal contamination when found in the environment. Thus, *E. coli* is commonly selected for Microbial Source Tracking (MST) in which a bacterial contaminant is matched to its host species by identifying and differentiating between strains. Current MST methods can be very expensive, in both money and labor, and many have issues in reproducibility when identifying different *E. coli* strains. In order to address many of the common complaints regarding MST methods, our lab has developed a novel pyrosequencing based method to distinguish between different *E. coli* strains. The method designates *E. coli* isolates into strains using the DNA sequences of the seven ribosomal RNA operons. The fingerprints (or pyroprints) that are obtained for comparison inspired the method's name: Pyroprinting.

This study used our new method to investigate the success of different sampling strategies for obtaining *E. coli* from humans. Success was determined by the variation of strains detected from each sampling type. Ideally, different types of sampling methods should detect similar proportions of strains. The secondary purpose of this study was to track strain distribution of intestinal *E. coli* populations over a six-month period in select human subjects. We hypothesized that each individual would host one or two dominant strains over the entire six month period, but that minor strains would vary.

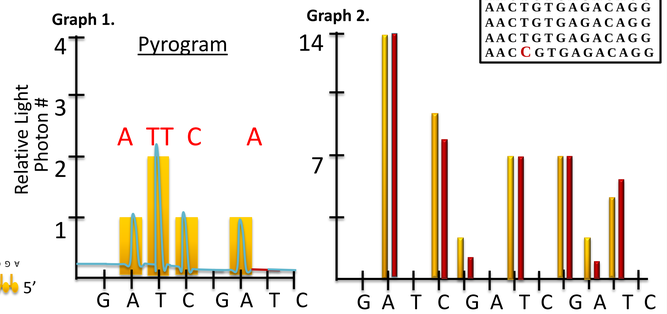
Methods



Graph 1. An example pyrogram from a single DNA sequence.

Graph 2. Example pyroprints for two strains with a single nucleotide difference.

Example Strain 1	AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACCGTGAGACAGG AACCGTGAGACAGG
Example Strain 2	AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACTGTGAGACAGG AACCGTGAGACAGG



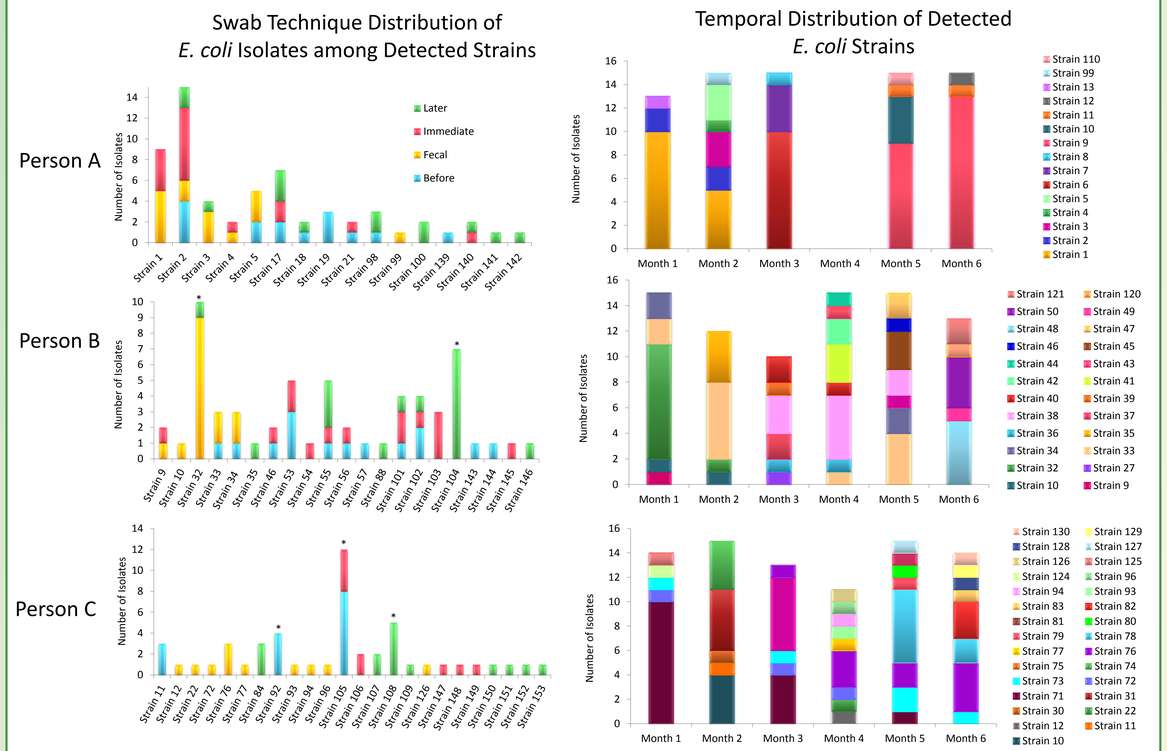
***E. coli* Sampling & Confirmation.** *E. coli* were sampled from three individuals once a month for six months. The individuals are two females and one male, ranging in age from 20-25. Four swabbing techniques were used to obtain intestinal *E. coli*. Anal swabs were collected **before**, **immediately** after, and several hours **later** after defecation. Lastly, collected feces were sampled with a sterile swab post homogenization. Inoculated swabs were streaked onto agar plates for single isolated colonies. From these initial plates, 15 single colonies from each swab technique were selected for a series of metabolic tests to confirm *E. coli*.

Molecular Analysis. Confirmed *E. coli* isolates were used in colony PCR to amplify both ITS 1 and ITS 2 in the ribosomal RNA operon of the *E. coli* genome. All seven copies of the rRNA operon in *E. coli* genome were amplified during PCR. Both ITS 1 and ITS 2 were pyroprinted using pyrosequencers. Normally during pyrosequencing, a sequence is determined by the heights and order of peaks of light produced when nucleotides dispensed in a particular order are incorporated to an extending strand of DNA. The amount of light produced is proportional to the number of nucleotides added in a row.

A pattern of peaks is produced based on the 7 different sequences; these are the pyroprints that are used for comparison in this method. Each *E. coli* isolate pyroprint is based on the results of two pyroprints: one for each ITS region. Unique *E. coli* pyroprints are considered to represent unique *E. coli* strains.

Statistics. Pearson Correlation was used to analyze differences between each *E. coli* isolate in peak heights at each nucleotide dispensation. A computer algorithm written by collaborating computer scientists was used to build strains using the correlation matrices for both ITS 1 and ITS 2. Isolates were clustered into a strain if the correlation between them was 99.5 or above. Any isolates with a correlation below 99.0 were never clustered together. Isolates with correlations between these values were clustered into a strain based on its average correlation with all other isolates within the strain. Once the strains were built, differences in swab techniques were analyzed statistically using binary logistic regression which determines the significance between differences in proportions. A p-value below α indicates there is a significantly different proportion of swab types in a strain.

Results



*Strains with a significant p-value less than α. Person A; α= 0.003. Person B; α= 0.002. Person C; α= 0.002.

Only fecal isolates were used for this temporal analysis.

Conclusions

- Sampling technique may influence the strains that are detected by pyroprinting.
  - Could be influenced by strain diversity (see Person A v. B & C)
  - These results could pose a question for other MST/detection methods.
- Over time individual hosts seem to harbor a diverse *E. coli* population
  - Type of strain detected as dominant changed frequently
  - Months that contain smaller, less represented strains may be months of transition
    - Less represented strains could be transient
- Human hosts are variable in the amount diversity and change within *E. coli* populations
  - Results between individuals in both swab type analysis and temporal analysis yields differences in distributions.
  - Variation seen both in overall diversity and temporal diversity
  - May be due to both genetic and environmental factors, but we hypothesize environment plays a larger role

A Future Direction

- Larger study regarding sampling method
  - Conduct study including other hosts
- Investigate *E. coli* strains sampled from a larger human populations:
  - Even though individual hosts are variable, do the strains detected in a large population of human hosts stay the same over time?

Acknowledgments

Special thanks to our supporters from the W.M. Keck Foundation and Cal Poly San Luis Obispo College-Based Fees. Thank you also to the Undergraduate Biotechnology Lab, the Environmental Biotechnology Institute, the entire M. Black lab, J. VanderKelen, and additional supporters at the Cal Poly Biological Sciences Department.