

Genomic and phenotypic assessment of *E. coli* O157:H7 REPEXH02 strains

Summary

Certain genetic subtypes of *E. coli* O157:H7 have been deemed reoccurring, emerging or persistent by the CDC. REPEXH02 has been linked to leafy greens grown in specific regions in California. The goal of this project is to determine features that may facilitate persistence of REPEXH02 in the Salinas and Santa Maria growing regions. *E. coli* O157:H7 will be isolated from soil, water, and wildlife samples, and will be sequenced along with historical *E. coli* O157:H7 isolates from the region. Experimental assays will measure resilience phenotypes. A suite of tools will be used to identify associations between genome data, phenotypes, and environmental parameters to determine the basis for persistence of REPEXH02. Understanding environmental factors associated with presence and persistence will aid in selecting appropriate mitigation strategies.

Objectives

1. Determine the evolutionary history of REPEXH02 strains in the broader context of *E. coli* O157:H7 clade 2 and identify unique gene content with potential to impact strain persistence.
2. Identify non-host reservoirs and environmental harborage sites of REPEXH02 strains utilizing environmental sample collection, phenotypic assays, and genome wide association analysis.
3. Develop a predictive model for the persistence of REPEXH02 strains based on the physicochemical properties of collected samples using AI models.

Methods

A total of 51 soil samples were collected from unique sites during sampling trips in March, November, and December 2025. Soil samples were dried and sent for physicochemical analysis to the UC Davis analytical lab. Soil samples were analyzed for heavy metal concentration, total nitrogen, phosphorus, and carbon, as well as sodium, magnesium, and organic matter. 37 REPEXH02 strains were evaluated for curli production and biofilm formation over 5 days. Strains were inoculated into LB broth without salt and incubated statically at 28°C for 48 and 120h. Biofilm density was measured with a crystal violet assay and reported as the absorbance at 570nm. A set of 2,035 O157:H7 genomes were assembled and analyzed for clade membership. Of these genomes, 553 were identified as clade 2. A maximum likelihood tree was constructed for the clade 2 genomes using 4,098 core genes.

Results to Date

Metal concentrations in soils were variable, with low levels of cadmium (Cd) detected in most soils (Figure 1). These data will be used for modeling in objective 3, along with genomic and phenotypic data. Biofilm formation was distinctly different for REPEXH02 strains that were capable of producing curli compared to those that do not (Figure 2). The difference in biofilm formation between the two groups was significantly different ($p < 0.05$) after 5 days of growth. The phylogenetic analysis indicated that REPEXH02 strains form two distinct clusters within clade 2 (Figure 3). To date, the nearest neighbors of REPEXH02 group 1 (Figure 3A) include human clinical isolates from 2011 and isolates from water collected in California in 2016. The nearest neighbors of REPEXH02 group 2 (Figure 3B) include clinical isolates from 2011, 2018, and 2022. Following REPEXH02's emergence, it has diversified into two well-supported clades. Persistence of both lineages over time in environmental samples suggests the possibility that divergence reflects adaptation to different environmental conditions.

Benefits to the Industry

Overall, this research will help define evolutionary history and determine the potential for environmental harborage or REPEXH02 strains. The predictive model will allow metrics such as water and soil physicochemical data to determine which areas should be the focus of more intensive sampling efforts or mitigation measures. More broadly, understanding potential selective pressures for emergence of new *E. coli* O157:H7 variants that have persistent traits will determine what types of information is most useful for making predictions, and will help focus resource allocation for mitigation to the most impactful activities.

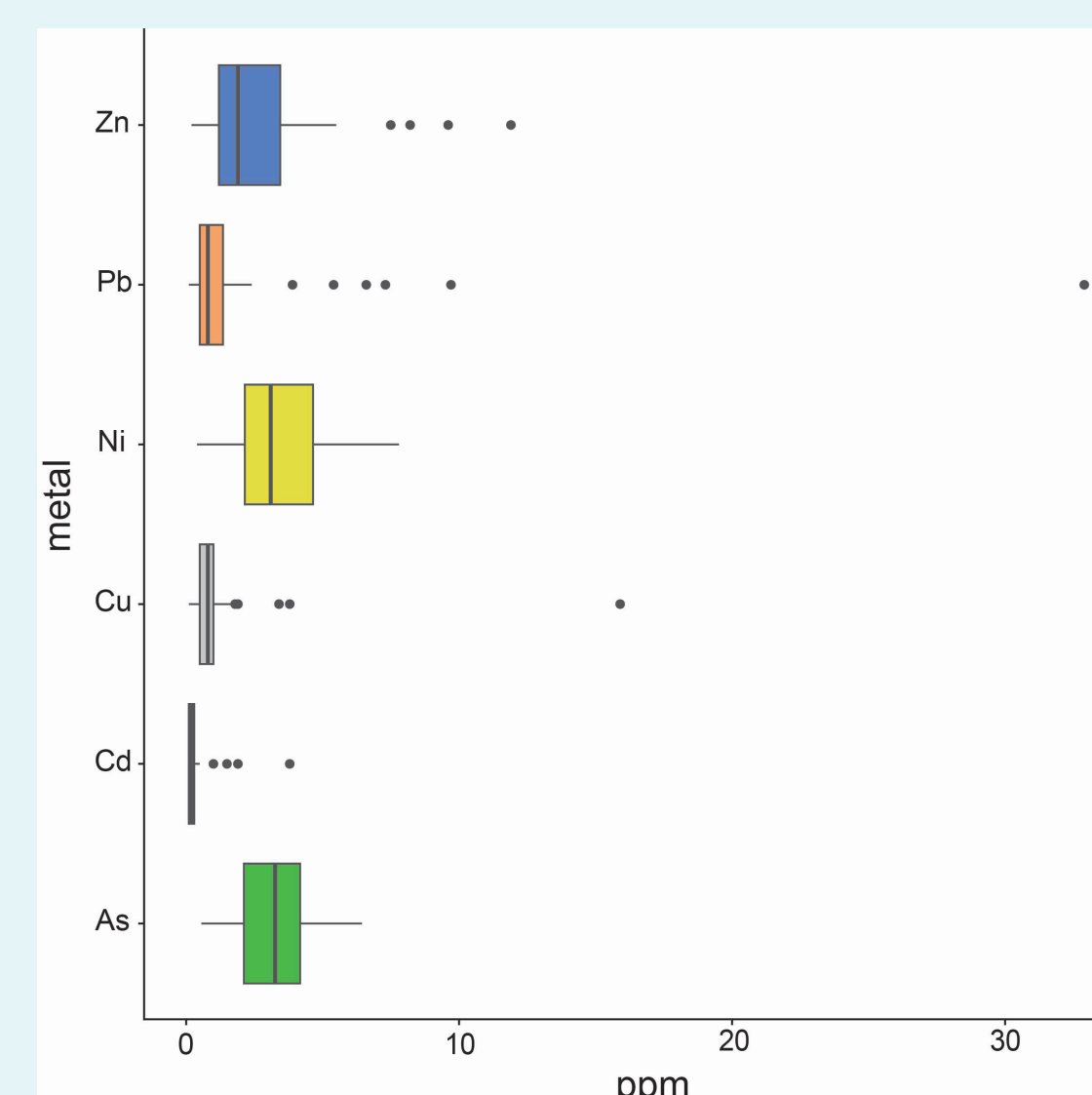


Figure 1: Distribution of metal concentrations (in ppm) for 51 soil samples collected from multiple leafy greens growing regions in 2025.

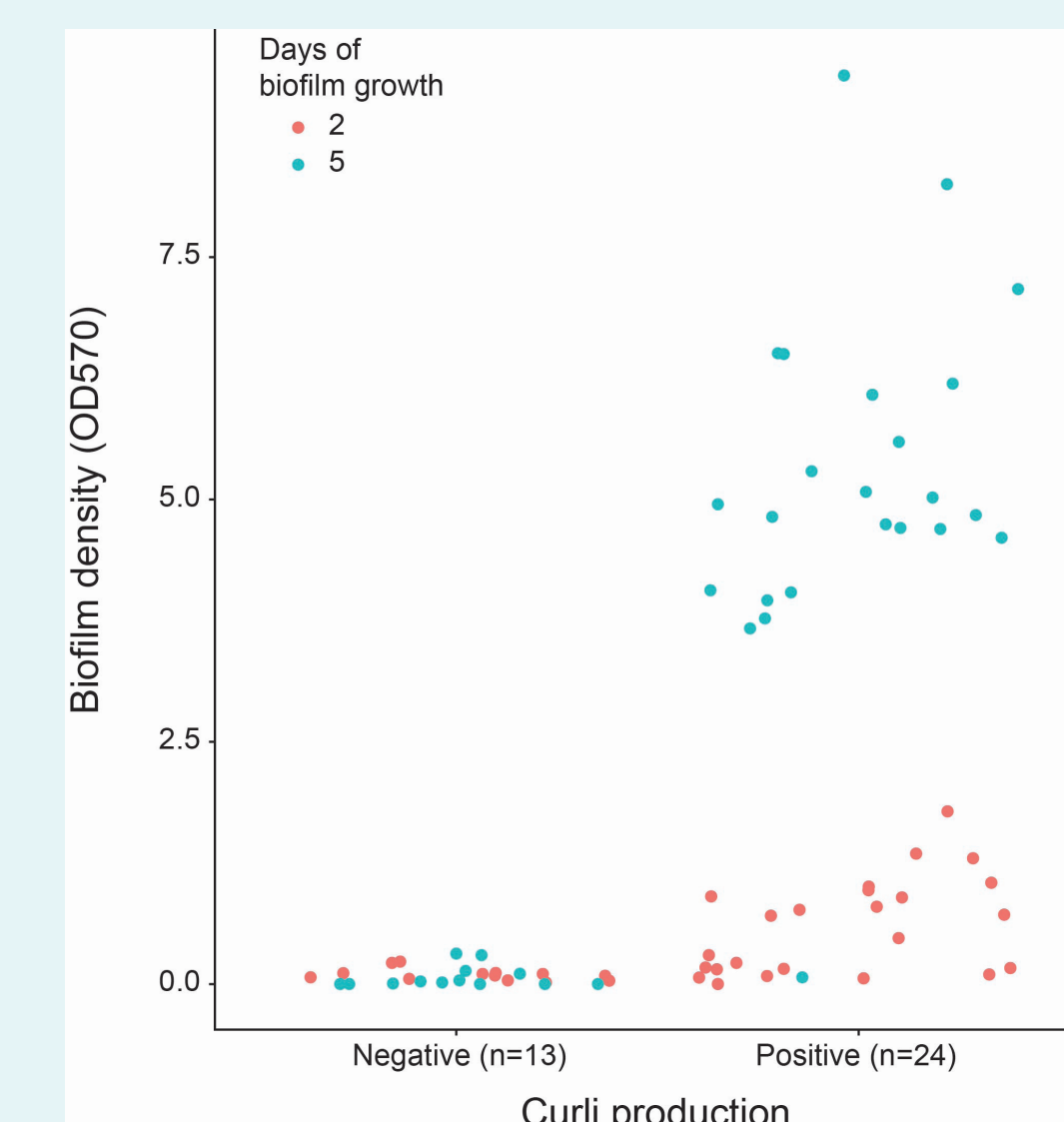


Figure 2: Biofilm formation data for 39 REPEXH02 strains. Biofilm density was measured after 2 days (orange circles) and 5 days (blue circles) of growth. Strains are grouped by whether or not they produced curli, which are cell surface structures known to contribute to biofilm formation.

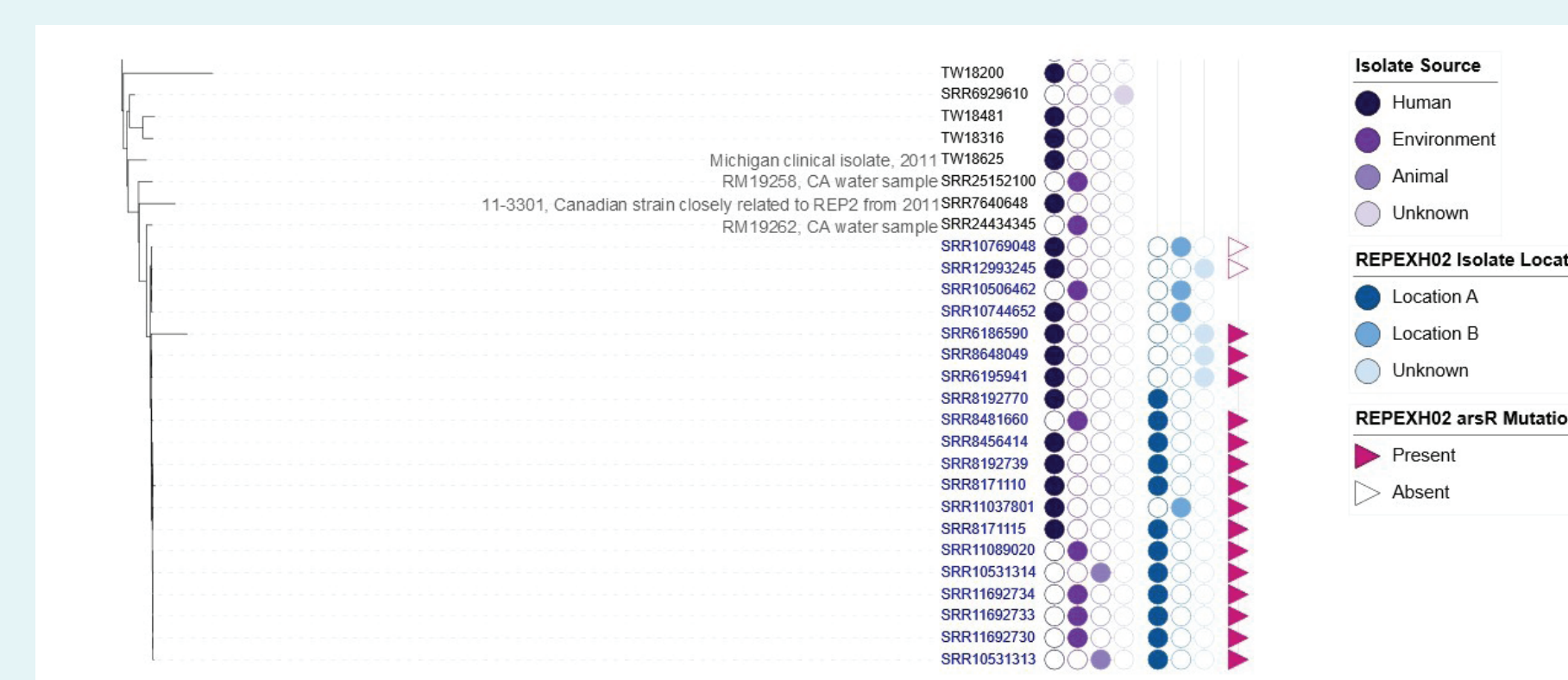
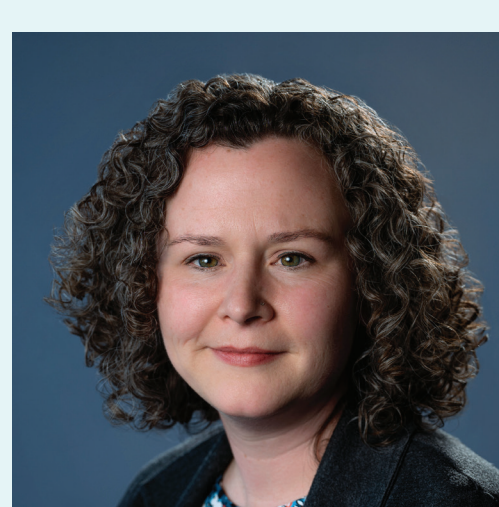


Figure 3A: Maximum likelihood tree of clade 2 genomes, focused view on REPEXH02 group 1 (A) and group 2 (B). REPEXH02 genomes are in blue font. Other genomes of note closely related to REPEXH02 are indicated in red font. Metadata for the genomes is presented on the right hand column in a series of colored circles and triangles. Isolate location and *arsR* mutation status are indicated only for the currently known REPEXH02 isolates.



Figure 3B



Contact

Teresa Bergholz, PhD
Michigan State University
tmb@msu.edu

Authors

Shannon Manning, Michelle Carter, Jiyoung Yi

Project funding dates

January 1, 2025 – December 31, 2026

Acknowledgements

Nolan Schinderle, Shaney Rump, Kyleen Hall