

# Organizational Meeting for Dissertation Committee



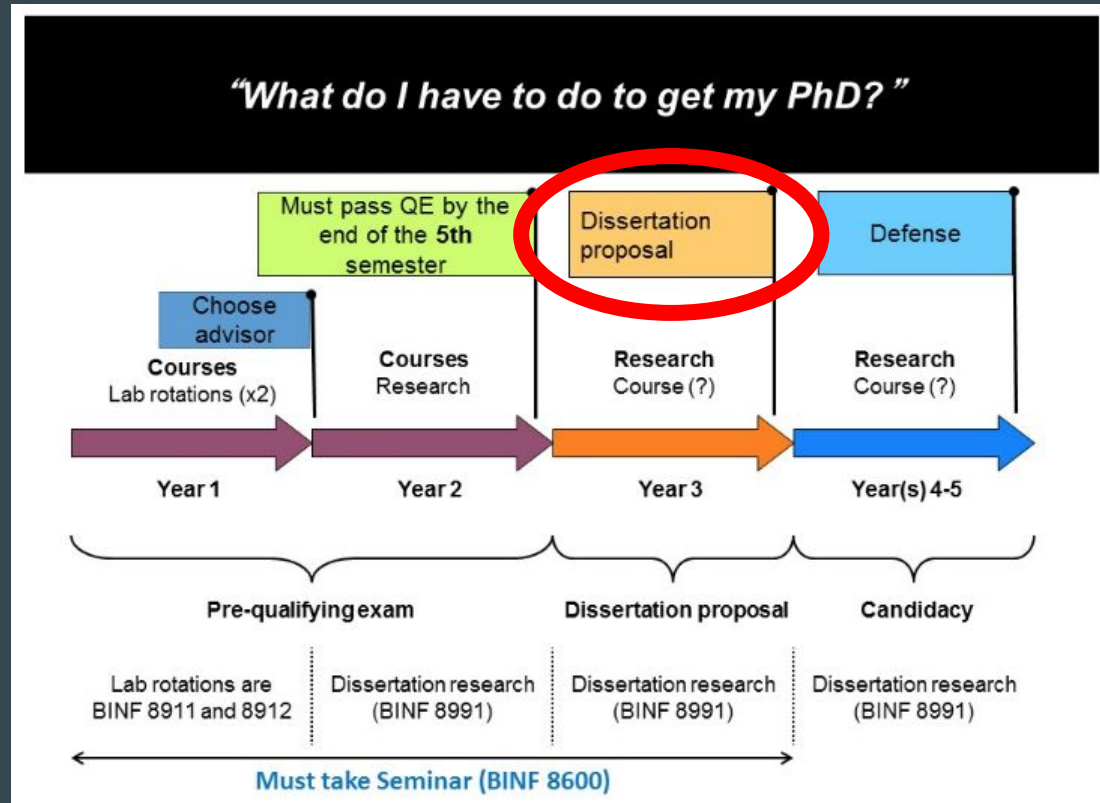
David Brown  
Lab of Dr. Dan Janies  
Charlotte Department of Bioinformatics and Genomics

# Welcome & Introduction

- David Brown
  - Current Education
    - Rising 4th year Ph.D. student
    - Bioinformatics and Computational Biology (Ph.D. BCB)
  - Previous Education
    - B.S. Biology - UNCW (2017)
  - Interest
    - Application of computational techniques to whole genome sequencing (WGS) data to analyze, understand, and predict the genotypic and phenotypic characteristics of human pathogens.



# BCB Ph.D. Timeline



# Dissertation Proposal

Domain Knowledge

Domain Skill Set

Domain Need

My Research Interest

# Research - Domain Knowledge

- Joined the Janies' Lab in 2018
- Began working on research into antimicrobial resistance (AMR) in *Escherichia coli*
- Specifically interested in investigating multidrug resistance (MDR)



Original Article | [Full Access](#)

## Genetic capitalism and stabilizing selection of antimicrobial resistance genotypes in *Escherichia coli*

Colby T. Ford, Gabriel Lopez Zenarosa, Kevin B. Smith, David C. Brown, John Williams, Daniel Janies [✉](#)

First published: 29 June 2020 | <https://doi.org/10.1111/cla.12421>

# GenomeTrakr

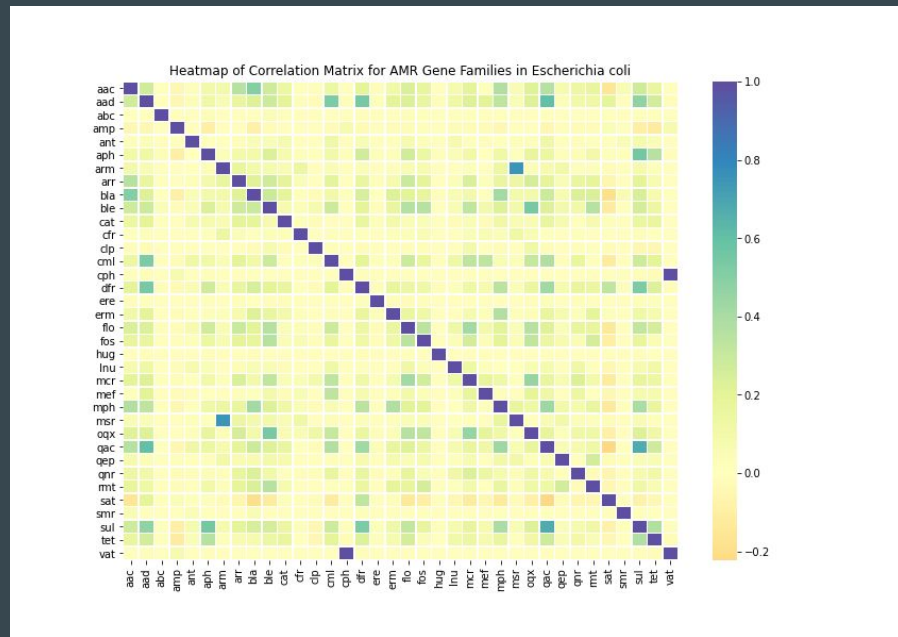
# 2021

**Tuesday, October 12 - Wednesday, October 13**

"GenomeTrakr is a laboratory network that utilizes whole genome sequencing (WGS) for pathogen identification. Led by FDA's Center for Food Safety and Applied Nutrition (CFSAN), the GenomeTrakr network is comprised of public health and university laboratories that collect and share genomic and geographic data from foodborne pathogens. The data, which are housed in public databases at the National Center for Biotechnology Information, can be accessed by researchers and public health officials for real-time comparison and analysis that can speed foodborne illness outbreak investigations and reduce foodborne illnesses and death."


# Research - Domain Skill Set

- CyberONR
  - Taught undergraduates about coding and machine learning (ML) to solve problems
  - Introduced multi-label classification ML algorithms
- Similar techniques could be used to predict labels such as AMR genotypes, as seen at right.



# Research - Domain Relevant and Recent Need

- "It is the high-resolution WGS data, combined with detailed and structured metadata, that may be used by artificial intelligence (AI) and machine learning (ML) tools to make even more predictive models for the accurate prediction of food, animal source, and or geographic location."
- Vertical evolution vs mobile genetic elements (MGE) and horizontal gene transfer (HGT) in enteric bacteria.



**DOMAIN 6 EVOLUTION AND GENOMICS**

***Salmonella* Genomics in Public Health and Food Safety**

ERIC W. BROWN,<sup>a</sup> REBECCA BELL,<sup>a</sup> GUODONG ZHANG,<sup>a</sup>  
● RUTH TIMME,<sup>a</sup> ● JIE ZHENG,<sup>a</sup> THOMAS S. HAMMACK,<sup>a</sup>  
AND MARC W. ALLARD<sup>a</sup>

<sup>a</sup>Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration,  
College Park, Maryland, USA



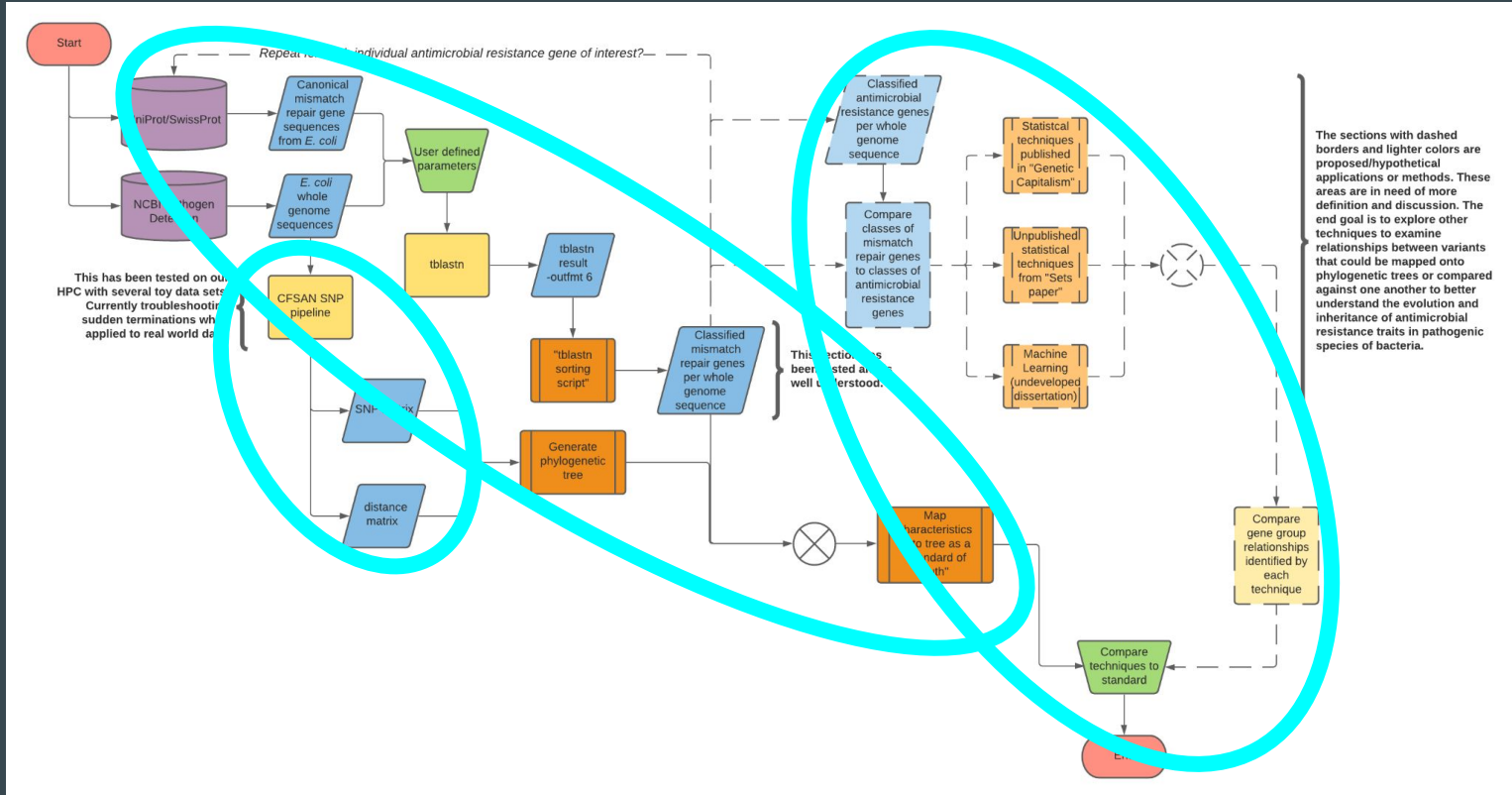
# Dissertation Proposal - Direction

- **Domain Knowledge**
  - B.S. Biology
  - Publication experience with *Escherichia coli* genetics
- **Domain Skill Set**
  - Phylogenetics of *Escherichia coli* using WGS techniques
  - Understanding of machine learning methods
- **Domain Need**
  - Predictive modelling for prevention of disease and protection of public health.
- **My Research Interest**
  - The combining of phylogenetic tools with machine learning techniques.

# Hypotheses - Understanding AMR in *Escherichia coli*

1. Which previously identified (Leclerc, 1997) hypermutable phenotype lineages of methyl-directed mismatch repair (MMR) genes are observed in the data?
2. What antimicrobial resistance drug classes or SNPs are associated with specific hypermutator lineages of MMR?
3. What is the predictive ability of machine learning algorithms trained on WGS SNP data to accurately and effectively predict multi-label classes (AMR gene sets or MDR traits)?
4. Can ensemble machine learning algorithms techniques predict features (location, source, etc.) relevant to the understanding of the global spread of pathogenic genetic lineages?

# Process for Enhancing Phylogenetic Trees with Other Techniques



# Hypotheses - Understanding AMR in *Escherichia coli*

1. Which previously identified (Leclerc, 1997) hypermutable phenotype lineages of methyl-directed mismatch repair (MMR) genes are observed in the data?
2. What antimicrobial resistance drug classes or SNPs are associated with specific hypermutator lineages of MMR?
3. What is the predictive ability of machine learning algorithms trained on WGS SNP data to accurately and effectively predict multi-label classes (AMR gene sets or MDR traits)?
4. Can ensemble machine learning algorithms techniques predict features relevant to the understanding of the global spread of pathogenic genetic lineages?

# Questions

Thank you for your time and kind attention.