

# Dissertation Defense of David C. Brown

Department of Bioinformatics and Genomics - College of Computing and Informatics,  
The University of North Carolina at Charlotte

Committee Chair: Dr. Daniel Janies  
Committee: Dr. Jun-tao Guo, Dr. Alex Dornburg, Dr. Adam Reitzel



# Welcome & Agenda

## Defense Meeting Procedure

- *Presentation*
  - ~45 min. in length.
- *Public Questions*
  - Follow the presentation.
- *Private Questions & Deliberation*
  - Adjourn to the conference room.

# Investigating Multidrug Resistance in *Escherichia coli* with Phylogenetics and Machine Learning

David C. Brown

# Global concern is increasing about the weakened efficacy of current antibiotic therapies.

Speech

## The silent pandemic of antimicrobial resistance

Health and Social Care Secretary Matt Hancock spoke at the United Nations high-level interactive dialogue on antimicrobial resistance (AMR).

From: [Department of Health](#)  
Published 29 April 2021

UNITED NATIONS FOUNDATION

Search

BLOG EVENTS MEDIA CAREERS DONATE

WHAT WE DO WHY WE DO IT WHO WE ARE GET INVOLVED

## ANTIMICROBIAL RESISTANCE IS THE SILENT PANDEMIC WE CAN NO LONGER NEGLECT

BLOG > GLOBAL HEALTH > ANTIMICROBIAL RESISTANCE IS THE SILENT PANDEMIC WE CAN NO LONGER NEGLECT

BY [KATE DODSON](#) ON NOVEMBER 22, 2021

GLOBAL HEALTH

f t in

GARDP  
Global Antibiotic Research & Development Partnership

WHO WE ARE WHAT WE DO NEWS & RESOURCES TAKE ACTION

## Are we ready for the silent pandemic of antibiotic resistance?

Johns Hopkins Bloomberg School of Public Health

Global Health **NOW**

CORONAVIRUSES DISASTERS MATERNAL HEALTH RACISM AND PUBLIC HEALTH

GHN EXCLUSIVE | ANTIMICROBIAL RESISTANCE | CORONAVIRUSES | GLOBAL HEALTH | HEALTH SYSTEMS | PHARMACEUTICALS

## A Second, Silent Pandemic: Antibiotic Resistance

October 12, 2021  
[MANICA BALASEGARAM](#)  
[SOUHA S. KANJ](#)

t f in

# Introduction

- A microbe that exhibits antimicrobial resistance (AMR) to three or more classes of antibiotics is termed multidrug resistant (MDR) (CDC, 2021)
- MDR bacteria drive the silent antimicrobial resistance (AMR) pandemic.

Antimicrobial agents used for susceptibility testing for *E. coli* isolates

CLSI Class	Antimicrobial Agent	Years Tested	Antimicrobial Agent Concentration Range (µg/mL)	MIC Interpretive Standard (µg/mL)		
				Susceptible	Intermediate*	Resistant
Aminoglycosides	Amikacin	1997–2010	0.5–64	≤16	32	≥64
	Gentamicin	1996–present	0.25–16	≤4	8	≥16
	Kanamycin	1996–2013	8–64	≤16	32	≥64
	Streptomycin†	1996–2013	32–64	≤32	N/A*	≥64
		2014–present	2–64	≤16	N/A*	≥32
β-lactam combination agents	Amoxicillin-clavulanic acid	1996–present	1/0.5–32/16	≤8/4	16/8	≥32/16
Cephems	Cefoxitin	2000–present	0.5–32	≤8	16	≥32
	Ceftiofur	1996–2015	0.12–8	≤2	4	≥8
	Ceftriaxone‡	1996–present	0.25–64	≤1	2	≥4
	Cephalothin	1996–2003	2–32	≤8	16	≥32
Folate pathway antagonists	Sulfamethoxazole	1996–2003	16–512	≤256	N/A*	≥512
	Sulfisoxazole	2004–present	16–256	≤256	N/A*	≥512
	Trimethoprim-sulfamethoxazole	1996–present	0.12/2.38–4/76	≤2/38	N/A*	≥4/76
Macrolides	Azithromycin§	2011–present	0.25–32 0.12–16¶	≤16	N/A*	≥32
Penems	Meropenem	2016–present	0.06–4	≤1	2	≥4
Penicillins	Ampicillin	1996–present	1–32	≤8	16	≥32
Phenicol	Chloramphenicol	1996–present	2–32	≤8	16	≥32
Quinolones	Ciprofloxacin**	1996–present	0.015–4	≤0.25	0.5	≥1
	Nalidixic acid	1996–present	0.5–32	≤16	N/A*	≥32
Tetracyclines	Tetracycline	1996–present	4–32	≤4	8	≥16

# Current Issues



Cladistics 36 (2020) 345–347

---

*Cladistics*

---

10.1111/cla.12428

## Epidemiology needs more interdisciplinary teams with expertise in molecular systematics, public health and food safety

Marc W. Allard\* and Eric W. Brown

*Center of Food Safety and Applied Nutrition, U. S. Food and Drug Administration, College Park, MD, 20740, USA*

Received 7 May 2020; Revised 10 June 2020; Accepted 17 June 2020

Big data methods and tools (Janies, 2019) to rapidly parse through these growing genomic databases also are greatly needed (Ford et al. their Figure 1) as much of the data are globally publicly released and shared almost entirely in real-time so that critical life-saving and timely phylogenetic-based decisions can be made.

in *E. coli* lineages. Their results also may have direct applications for optimal drug use to combat antibiotic resistant strains and in understanding a crucial problem in public health due to the rise of antimicrobial resistance and the limited choice of available drugs.

## Current Issues - cont.

- MDR is a phenotypic description, not a specific or unique AMR genetic trait.
- Many potential combinations of AMR genes could lead to an MDR phenotype.
- Reticulate evolutionary mechanisms like horizontal gene transfer (HGT) can cloud the vertical (ancestor to descendant) signal commonly sought in phylogenetic analyses.

### Purpose

*How to build appropriate genotype-to-phenotype hypotheses for MDR given these constraints?*

# Literature Review



## 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

diversity that has accrued during a contamination event. 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 (172), and or geographic location. Published WGS data have shown that most *Salmonella* and *Listeria* isolates exhibit a very strong phylogeographic signal that is

modify the protein and affect the phenotype. By combining cladistics, character optimization, and WGS, investigators may be able to identify genotype-to-phenotype changes that specific bacterial lineages have acquired and that allow food-borne pathogens to survive and contaminate foods, animals, and the environment (177). In several examples, investigators have predicted which genomic changes correlate with outbreaks in Italian-style meats (178) and in eggs (115, 116,



# Literature Review - cont.

- Bacteria with deficient mismatch repair mechanisms have been described as having a "hypermutable phenotype".
- This highly adaptive phenotype is suggested to drive the acquisition of antimicrobial resistance.
- A main component of the methyl-directed mismatch repair (MMR) system in bacteria is Mutator S (gene *mutS*, protein MutS).

## ORIGINAL RESEARCH article

Front. Microbiol., 08 May 2020 | <https://doi.org/10.3389/fmicb.2020.00770>



## Genes and Proteomes Associated With Increased Mutation Frequency and Multidrug Resistance of Naturally Occurring Mismatch Repair-Deficient *Salmonella* Hypermutators

Huanjing Sheng<sup>1</sup>, Jinling Huang<sup>1</sup>, Zhaoyu Han<sup>2</sup>, Mi Liu<sup>1</sup>, Zexun Lü<sup>1</sup>, Qian Zhang<sup>1</sup>, Jinlei Zhang<sup>1</sup>, Jun Yang<sup>1</sup>, Shenghui Cui<sup>3</sup> and Baowei Yang<sup>1\*</sup>

<sup>1</sup>College of Food Science and Engineering, Northwest A&F University, Xianyang, China

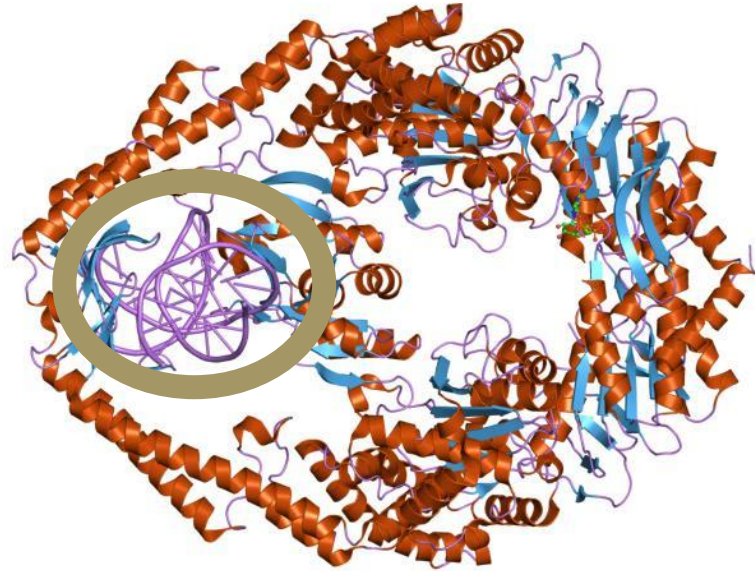
<sup>2</sup>School of Pharmaceutical Sciences, Jiangnan University, Wuxi, China

<sup>3</sup>National Institutes for Food and Drug Control, Beijing, China

# Literature Review - cont.

Two forces known to govern genetic change are the mutation of nucleotide sequences within a genome and the horizontal transfer of existing sequences among genomes. In certain mutator phenotypes, namely those deficient in methyl-directed mismatch repair (MMR), these forces converge to increase the rate of genetic variation. MMR is a postreplicative repair system that corrects errors on newly synthesized DNA strands to ensure the precision of chromosome replication (3). MMR is also a major barrier to interspecies gene exchange (4). Consequently, bacteria defective in MMR show both an enhanced rate of mutation (a hypermutable phenotype) and an increase in recombination of diverged sequences; that is, they are more promiscuous (4). It therefore becomes relevant to the problem of emerging pathogens to determine the frequency of such mutator phenotypes among human pathogens.

SCIENCE • VOL. 274 • 15 NOVEMBER 1996



## Research Questions

Does defective MMR explain the development of MDR in *E. coli*?

How well does WGS *E. coli* data predict categories of MDR?

# Data Source

- NCBI Pathogen Detection (NCBI PD) is a project to monitor bacterial pathogens.
- Used by the GenomeTrakr program for international food safety surveillance.
- Represents value in the study of the silent AMR pandemic.

## Pathogen Detection BETA



To assist the National Database of Antibiotic Resistant Organisms (NDARO), NCBI Pathogen Detection identifies the antimicrobial resistance, stress response, and virulence genes found in bacterial genomic sequences. This enables scientists to track the spread of resistance genes and to understand the relationships between antimicrobial resistance and virulence.

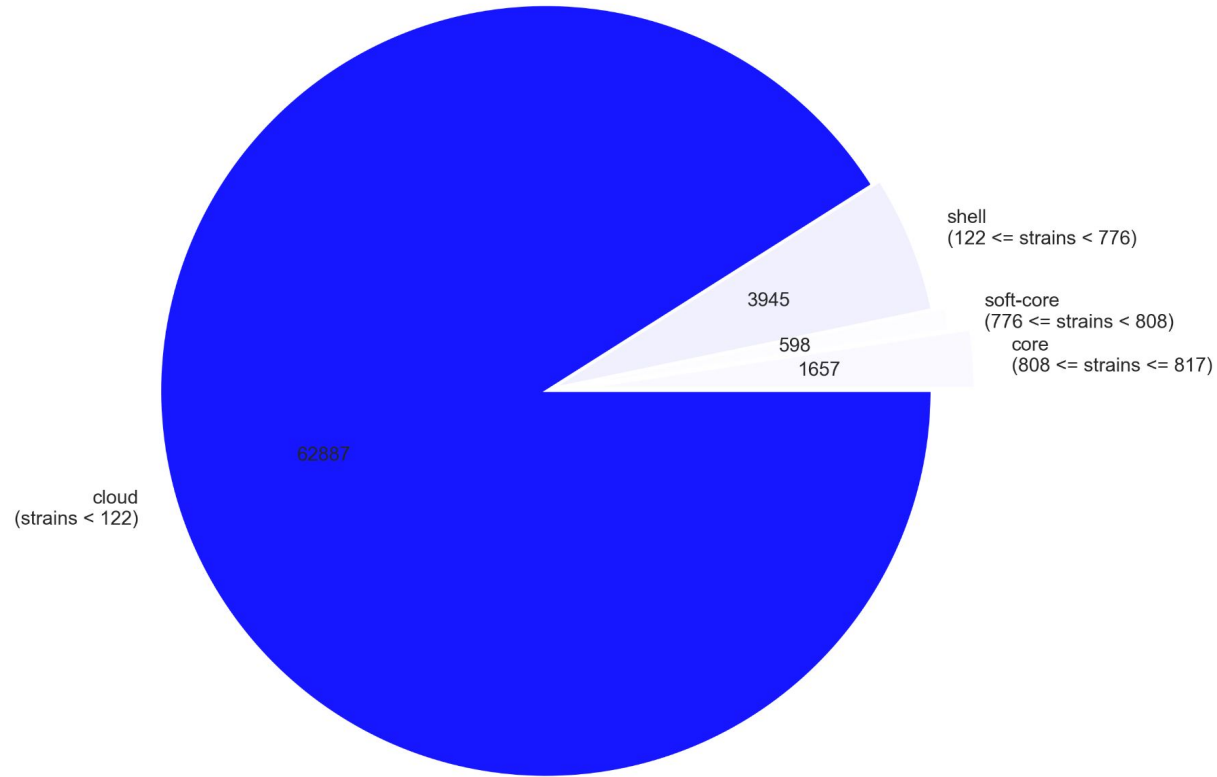
NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

# Data Processing

- Used the same 29,255 isolate *E. coli* data set and metadata table as Ford et al. (2020).
- Filtered the data set for:
  - NCBI:txid562 (*E. coli*).
  - Complete Genome and Chromosome levels of assembly.
  - A RefSeq identifier (GCF\_#####).
- Requested 911 isolates from NCBI FTP site
  - Downloaded 875 WGS FASTA files
- Quality controlled with CheckM to **817** *E. coli* of high confidence, completeness, and quality.

# Data Processing - cont.

- Annotated the 817 *E. coli* with Prokka
- Constructed the pangenome with Roary
  - Core gene alignment file for phylogenetics
  - Gene presence absence file for ML



# Characterizing MDR - 3 or more Resistance Categories

*E. coli*

Antimicrobial agents used for suscep

CLSI Class	Antimicrobial Agent
Aminoglycosides	Amikacin
	Gentamicin
	Kanamycin
	Streptomycin <sup>†</sup>
β-lactam combination agents	Amoxicillin-clavulanic acid
Cephems	Cefoxitin
	Ceftiofur
	Ceftriaxone <sup>†</sup>
	Cephalothin

Resistance Category	3 Letter Resistance Gene Prefix
aminoglycosides	<i>aac, aad, ant, aph, arm, rmt</i>
beta-lactam combination agents	<i>bla</i>
cephems	<i>abc</i>
folate pathway antagonists	<i>dfr, sul</i>
macrolides	<i>ere, erm, mef, mph, msr</i>
nucleosides	<i>sat</i>
penicillins	<i>amp</i>
phenicols	<i>cat, flo</i>
quinolones	<i>qep</i>
tetracyclines	<i>tet</i>
others	<i>arr, ble, cml, fos, lnu, mcr, oqx, qac</i>



# Characterizing MMR - MutS Protein Variants

## UniProtKB 2021\_04 results

UniProtKB consists of two sections:



### Reviewed (Swiss-Prot) - Manually annotated

Records with information extracted from literature and curator-evaluated computational analysis.



### Unreviewed (TrEMBL) - Computationally analyzed

Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub collection of functional information on proteins, with accurate and consistent and rich annotation. In addition to capturing the mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and other information), as much annotation information as possible

[? Help](#) [UniProtKB help video](#) [Other videos](#)

Filter by<sup>i</sup>



Reviewed (21) ✕

Swiss-Prot

Popular organisms

[E. coli K12 \(1\)](#)

[EC057 \(1\)](#)

[ECOL6 \(1\)](#)

[BLAST](#)

[Align](#)

[Download](#)

[Add to basket](#)

[Columns](#)



1 to 21 of 21

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism
<input type="checkbox"/>	P23909	MUTS_ECOLI		<b>DNA mismatch repair protein MutS</b>	<b>mutS</b> fdv, b2733, JW2703	Escherichia coli (strain K12)
<input type="checkbox"/>	Q9S6P8	MUTS_ECO57		<b>DNA mismatch repair protein MutS</b>	<b>mutS</b> fdv, Z4043, ECs3589	Escherichia coli O157:H7
<input type="checkbox"/>	A7ZQH3	MUTS_ECO24		<b>DNA mismatch repair protein</b>	<b>mutS</b> EcE24377A_3030	Escherichia coli O139:H28 (strain

Image of the reviewed MutS variants for *E. coli* found in the UniProt database. Accessed Dec. 12, 2021 at <https://www.uniprot.org/uniprot/?query=organism%3A%22escherichia+coli%22+gene%3Amuts+reviewed%3Ayes&sort=score>.



# Methodology - Correlating evolution for MutS & MDR

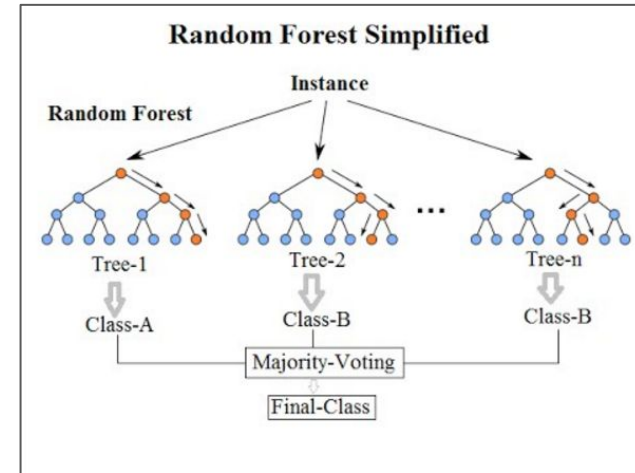
- Built phylogenetic tree with RAxML on the core genome file from Roary.
- Analyzed MutS and MDR characters with Mesquite and R, specifically *fitPagel()* from *phytools*.

Isolate	MutS-	MDR+
GCF_000000001	0	0
GCF_000000002	1	1
GCF_000000003	0	1
GCF_000000004	1	0
GCF_000000005	0	0

# Methodology - Identifying best predictors of MDR

- Trained random forest classifiers with scikit-learn library in Python 3
  - Feature Matrix: gene presence/absence file from Roary
  - Label Matrix: presence/absence of AMR genes from NCBI PD, compressed into resistance categories

Isolate	Gene 1	Gene 2	Gene 3	Gene n	ami +	fol +	mac +	tet +
GCF_000000001	1	0	1	0	0	0	1	1
GCF_000000002	1	1	0	1	1	1	1	0
GCF_000000003	1	1	1	1	0	1	1	1
GCF_000000004	1	1	1	1	1	0	1	0
GCF_000000005	1	0	1	1	0	0	0	1



# Limitations

- **Data**

- Lack of expression data collected by GenomeTrakr.
- Heavy reliance on accurate metadata from NCBI PD.

- **Methods**

- Did not employ a reference sequence.
- Did not count mutations.
- High computational resource demands.

# First Research Question and Hypothesis

Does defective MMR explain the development of MDR in *E. coli*?

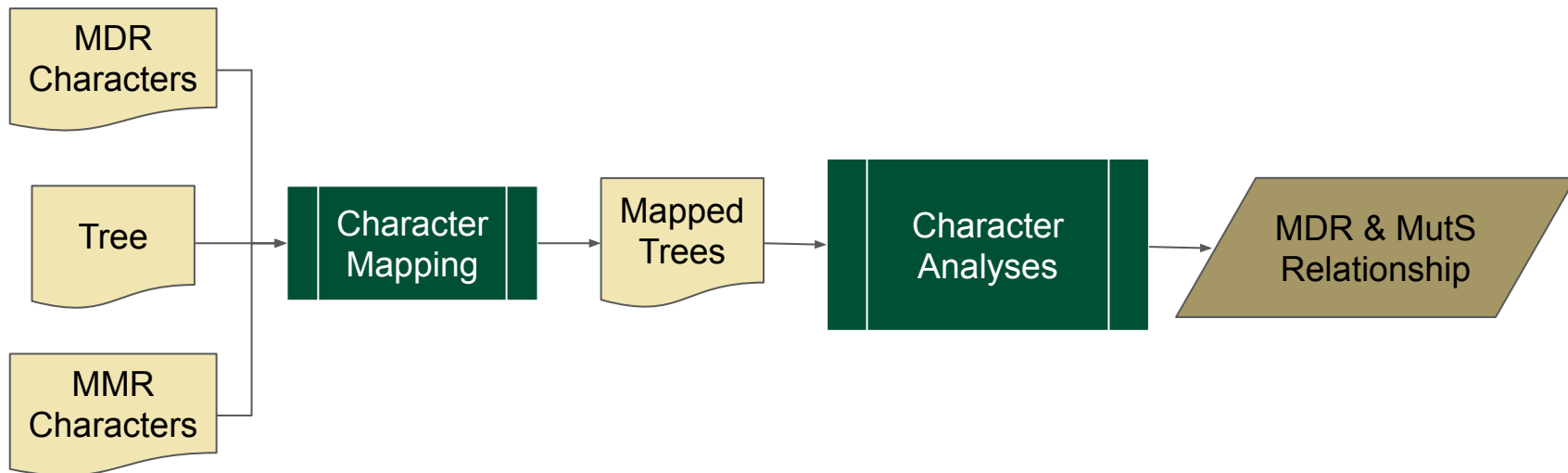
$H_0$

No relationship (independence) between repair deficient MutS variants and MDR.

$H_1$

Repair deficient MutS variants correlate with MDR.

# Evolutionary Relationship of MDR & MutS - Workflow



# No *mutS* Genes Match LeClerc's Deficient Definition

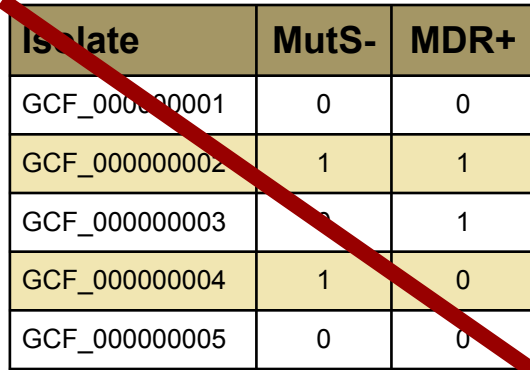
```
FEATURES             Location/Qualifiers
    source             1..69
                        /organism="Escherichia coli str. K-12 substr. MG1655"
                        /mol_type="genomic DNA"
                        /strain="K-12"
                        /sub_strain="MG1655"
                        /host="Homo sapiens"
                        /db_xref="taxon:511145"
                        /country="USA: Boston"
                        /collection_date="2017-01-15"
                        /note="C321.deltaA"
    gene               1..69
                        /locus_tag="CXP41_14995"
                        /pseudo
    CDS                1..69
                        /locus_tag="CXP41_14995"
                        /inference="COORDINATES: similar to AA
                        sequence:RefSeq:WP_007663512.1"
                        /note="incomplete; partial on complete genome; missing
                        start; Derived by automated computational analysis using
                        gene prediction method: Protein Homology."
                        /pseudo
                        /codon_start=1
                        /transl_table=11
                        /product="DNA mismatch repair protein MutS"

ORIGIN
    1 ctcgaaggct ttaatttgca agctcgtcag gcgctggagt ggatttatcg cttgaagagc
   61 ctggtgtaa

//
```

Screen capture of NCBI Assembly record annotations for MutS protein in RefSeq isolate GCF\_002846385 (GenBank: CP025268.1; Assembly: GCA\_002846385; BioProject: PRJNA421841). Accessed March 20, 2023.

# Change to Protein Space: MutS Consensus Sequence



Isolate	MutS-	MDR+
GCF_000000001	0	0
GCF_000000002	1	1
GCF_000000003	0	1
GCF_000000004	1	0
GCF_000000005	0	0

Isolate	pos_1	pos_2	pos_3	pos_n	MDR+
GCF_000000001	1	0	1	0	0
GCF_000000002	1	1	0	1	1
GCF_000000003	1	1	1	1	1
GCF_000000004	1	1	1	1	0
GCF_000000005	1	0	1	1	0

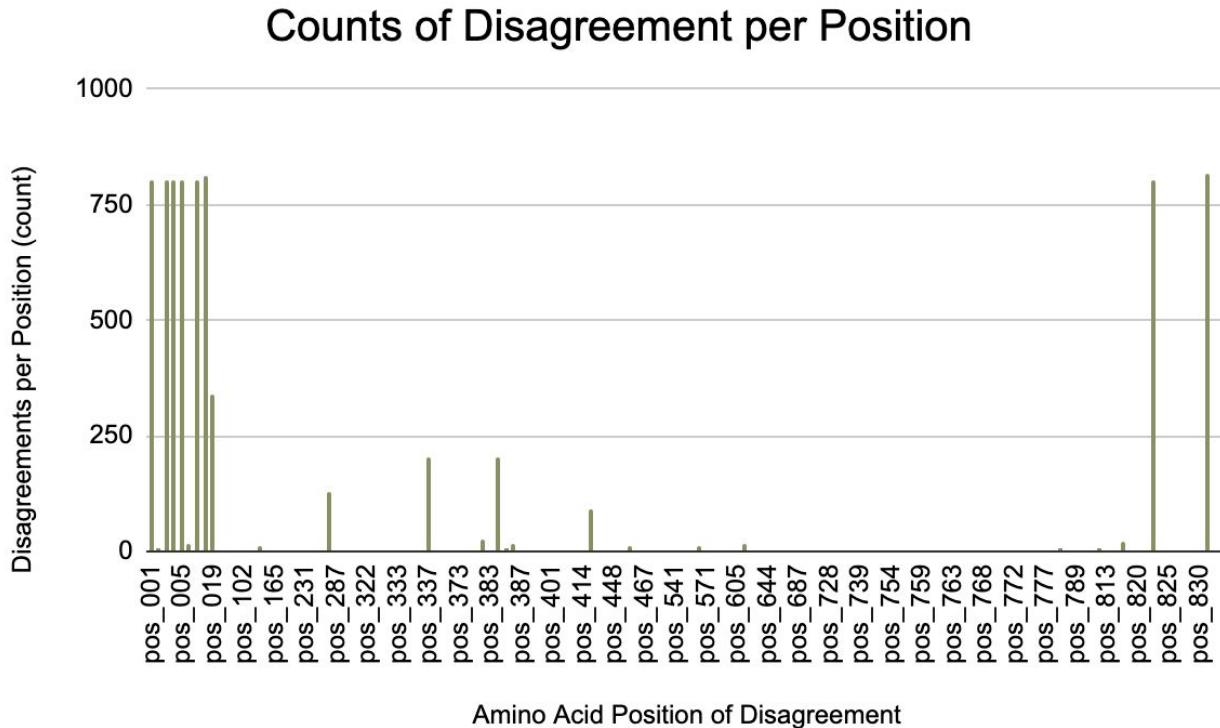
# Weighted AIC Supports Independent Model of Correlation Between MutS Residues and MDR

SNP_pos	independent_aic	weighted_independent_aic	dependent_x_aic	weighted_dependent_x_aic	dependent_y_aic	weighted_dependent_y_aic	dependent_both_aic	weighted_dependent_both_aic	aic_sum_check
pos_001	972.674191	0.73850	976.425098	0.11320	976.157236	0.12943	980.008335	0.01887	1.00000
pos_002	937.020127	0.72062	940.029061	0.16007	941.019181	0.09757	944.022361	0.02174	1.00000
pos_003	964.976440	0.76582	968.933036	0.10592	968.806119	0.11286	972.788963	0.01540	1.00000
pos_004	956.967060	0.77511	960.956773	0.10544	960.963009	0.10511	964.946896	0.01434	1.00000
pos_005	972.674191	0.73850	976.425098	0.11320	976.157236	0.12943	980.008335	0.01887	1.00000
pos_007	983.221111	0.15289	998.103538	0.00009	980.192509	0.69507	983.233381	0.15195	1.00000
pos_008	972.674191	0.73850	976.425098	0.11320	976.157236	0.12943	980.008335	0.01887	1.00000
pos_009	935.186074	0.76140	938.655002	0.13438	939.425851	0.09140	943.355416	0.01281	1.00000
pos_019	1084.630053	0.73699	1087.917985	0.14240	1088.601191	0.10119	1091.902770	0.01942	1.00000
pos_026	895.860034	0.69515	899.853922	0.09437	898.503629	0.18537	902.501635	0.02511	1.00000
pos_073	895.860034	0.69515	899.853922	0.09437	898.503629	0.18537	902.501635	0.02511	1.00000
pos_086	906.358058	0.68833	909.673777	0.13116	909.361772	0.15330	912.819764	0.02721	1.00000
pos_102	912.172833	0.56689	915.020106	0.13653	914.699345	0.16028	915.023333	0.13631	1.00000
pos_110	906.619284	0.70849	910.029034	0.12880	909.899289	0.13743	913.285171	0.02528	1.00000
pos_140	1048.579427	0.00000	913.056810	0.16496	910.077542	0.73168	913.991884	0.10336	1.00000
pos_156	909.546683	0.74668	912.946782	0.13640	913.586195	0.09908	917.014972	0.01784	1.00000
pos_165	906.648843	0.82548	910.107767	0.14642	1044.494424	0.00000	913.409630	0.02809	1.00000
pos_167	906.136171	0.68873	909.625882	0.12030	909.000245	0.16448	912.652268	0.02649	1.00000

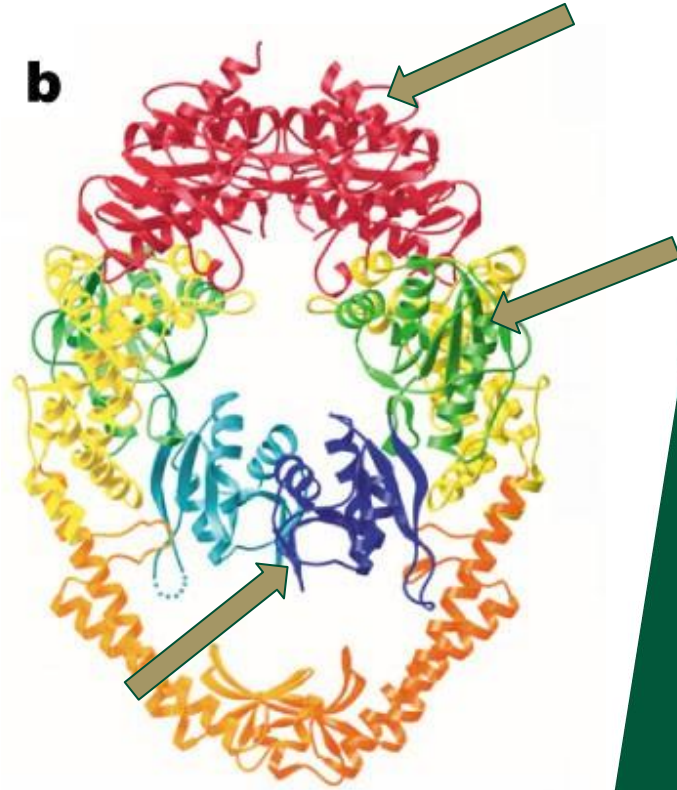
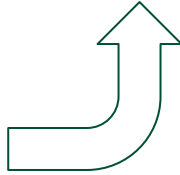
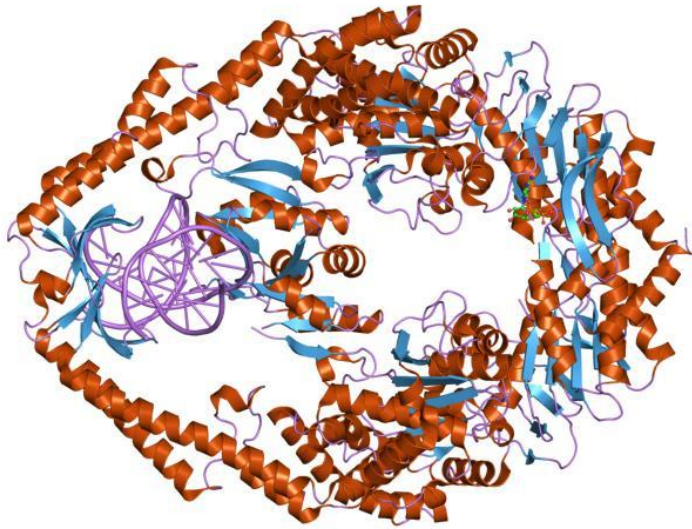
Only exceptions were the residue variants at positions 7 (domain I - mismatch binding), 140 (domain II - connector), and 804 (near final HTH motif).



# Position & Frequency of Residue Variants in MutS



# Locations Where MDR Depends on MutS Residues



Left image of MutS protein (PDB 1oh6) from previous slide.  
Right image Muts modified from Figure 3 by Obmolova et al., 2000.  
Similar colors represent same domain in the respective MutS dimer.  
Approximate positions of residues 7, 140, and 804 added for emphasis.

# Summary & Discussion

## Does defective MMR explain the development of MDR in *E. coli*?

- No observation of LeClerc's definition of hypermutable *mutS*.
- Generally independent evolutionary correlation between MutS residue variants and MDR. ( $H_0$ )
- Three positions support a dependent evolutionary correlation for MutS and MDR, potentially indicating novel definitions for repair deficient MutS or MDR development in *E. coli*.

## Second Research Question and Hypothesis

How well does WGS *E. coli* data predict categories of MDR?

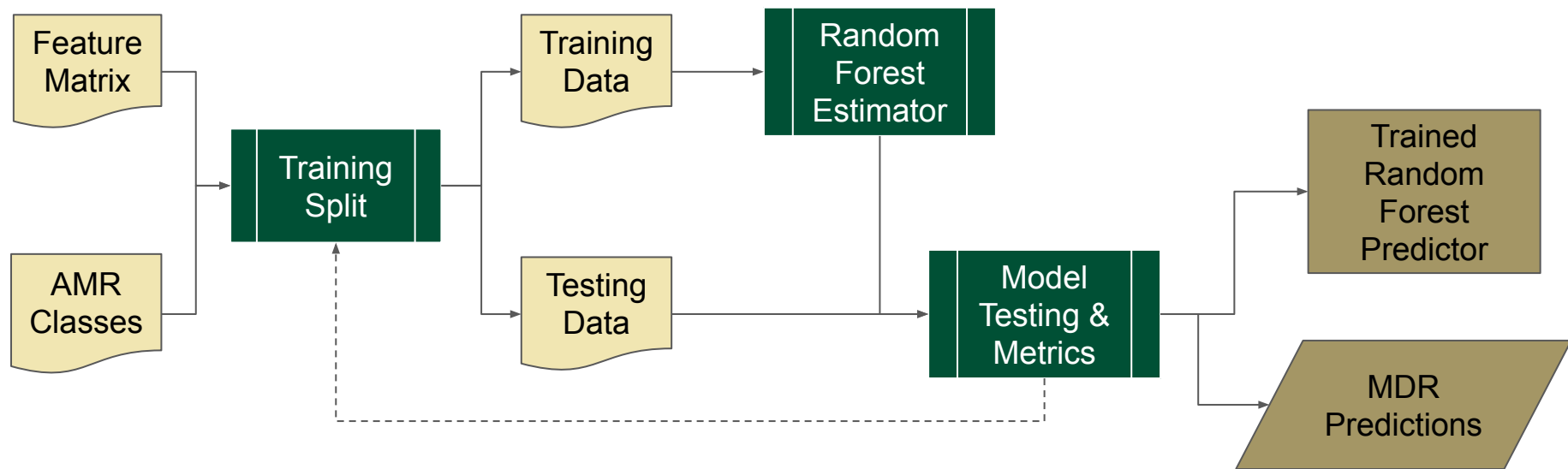
$H_0$

A random forest classifier trained on annotated *E. coli* gene features will reveal *mutS* or its paralog as the best predictor of MDR.

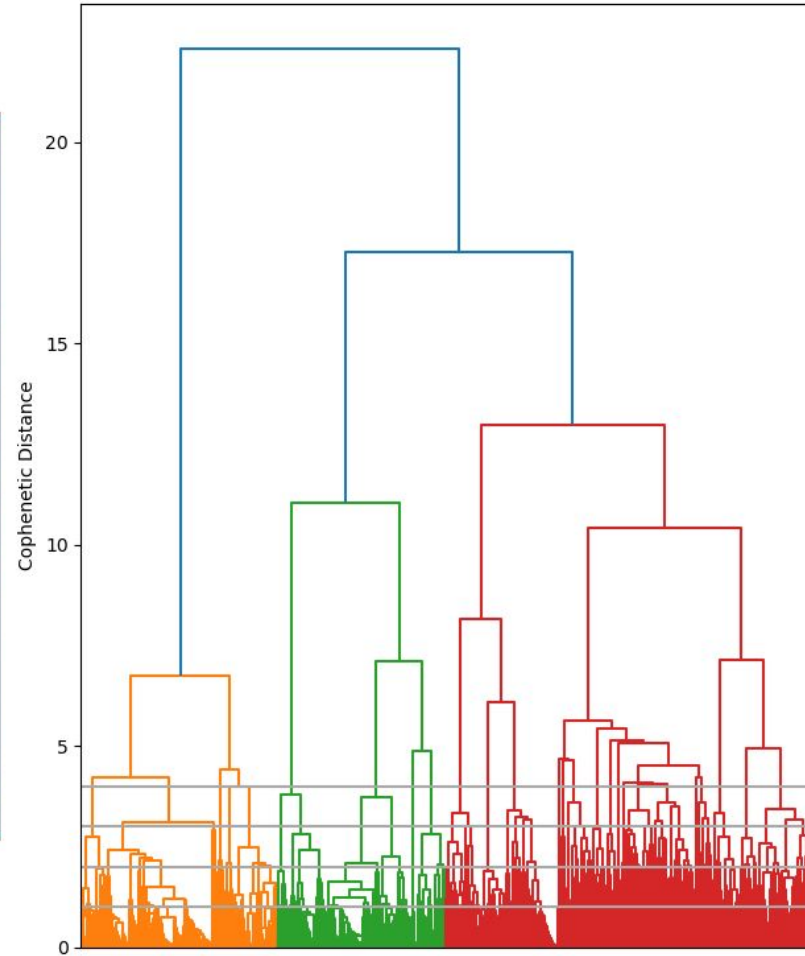
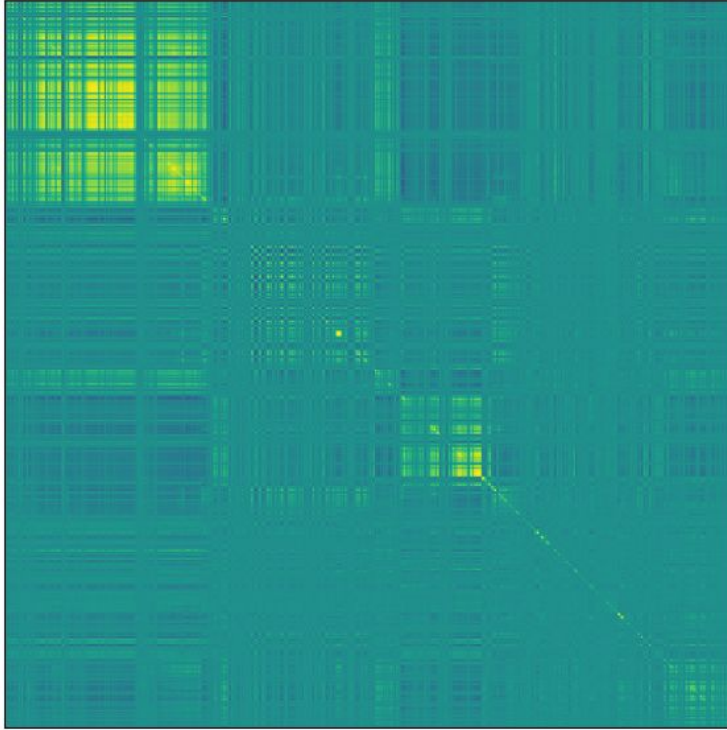
$H_1$

Gene features other than *mutS* best predict MDR.

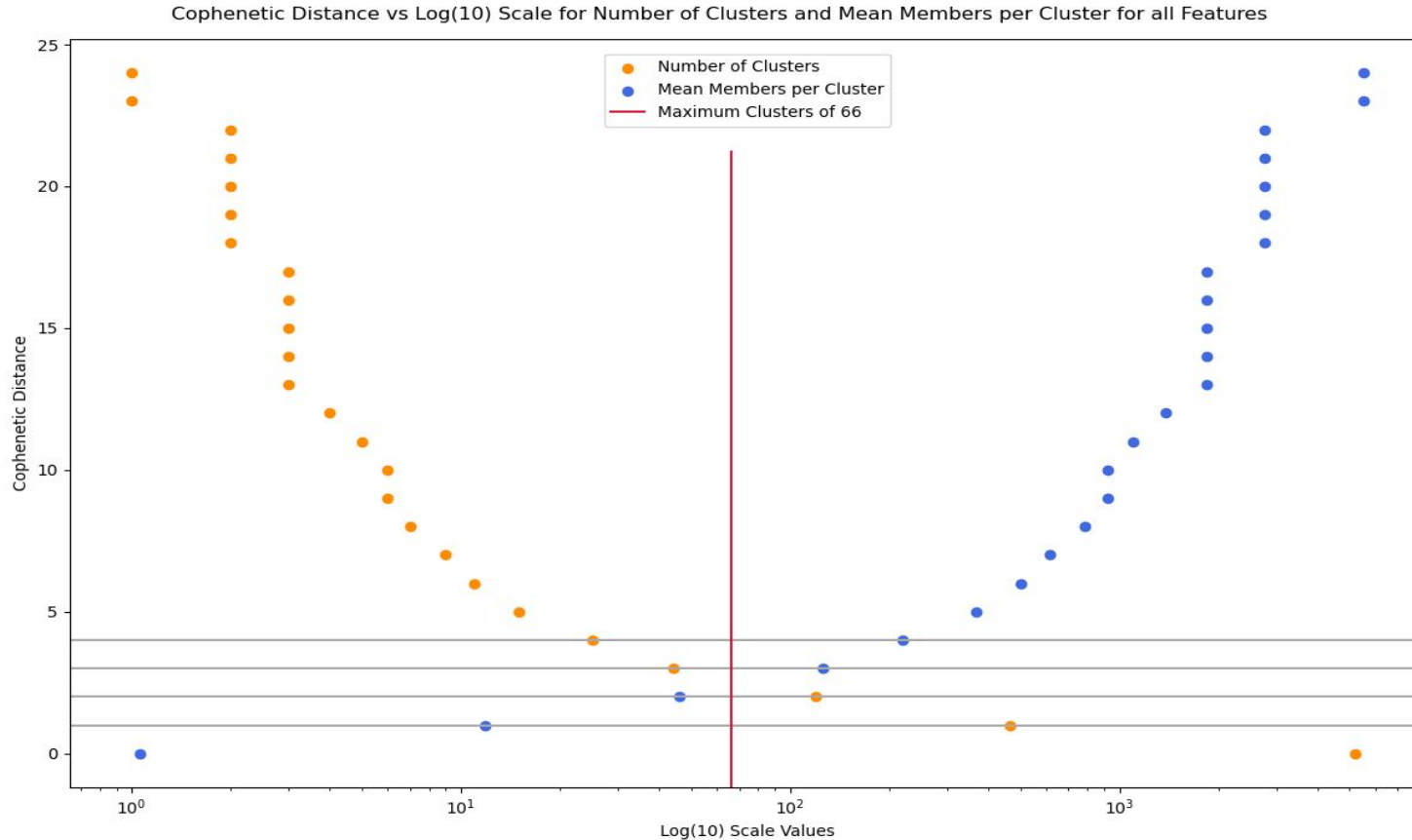
# Predicting MDR from WGS Annotations - Workflow



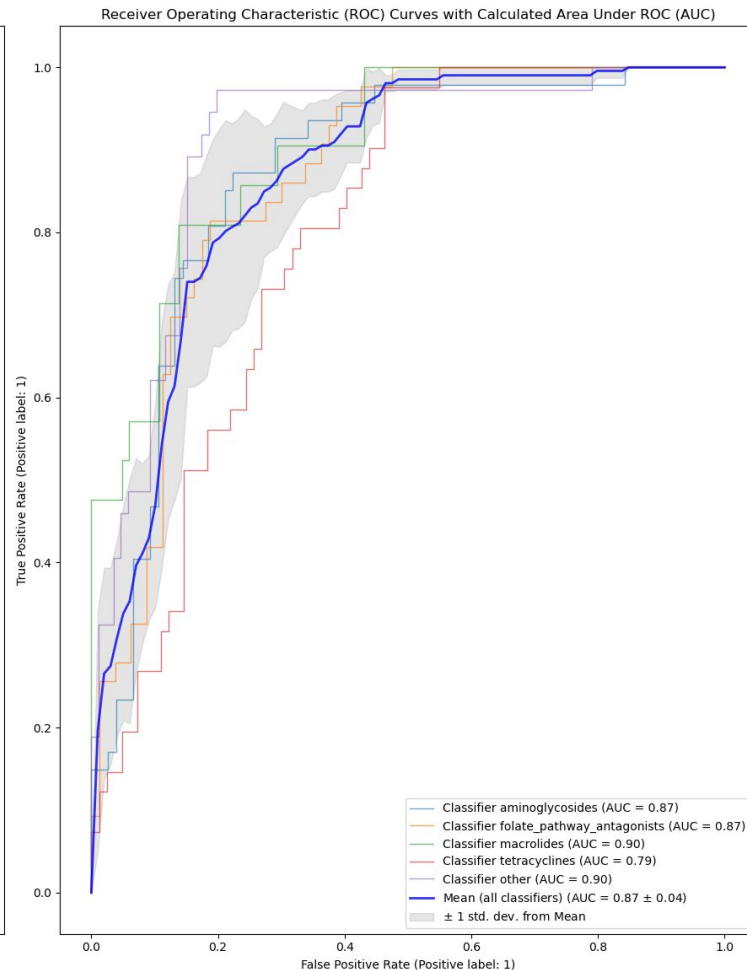
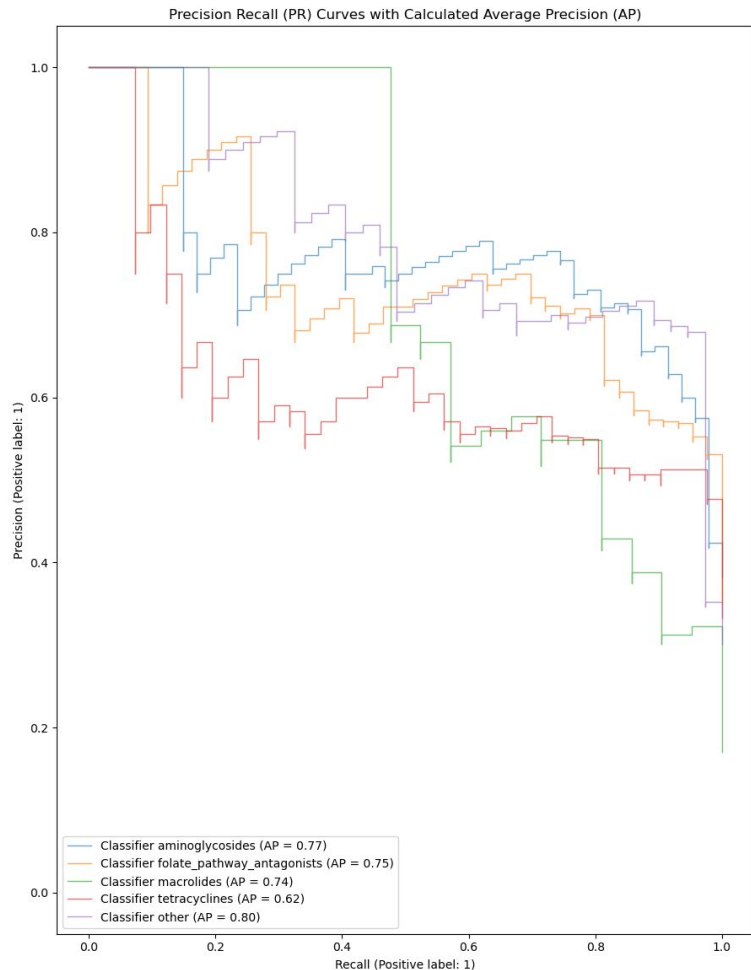
# Feature Correlation Matrix and Hierarchical Clustering



# Solving Collinearity by Clustering Features



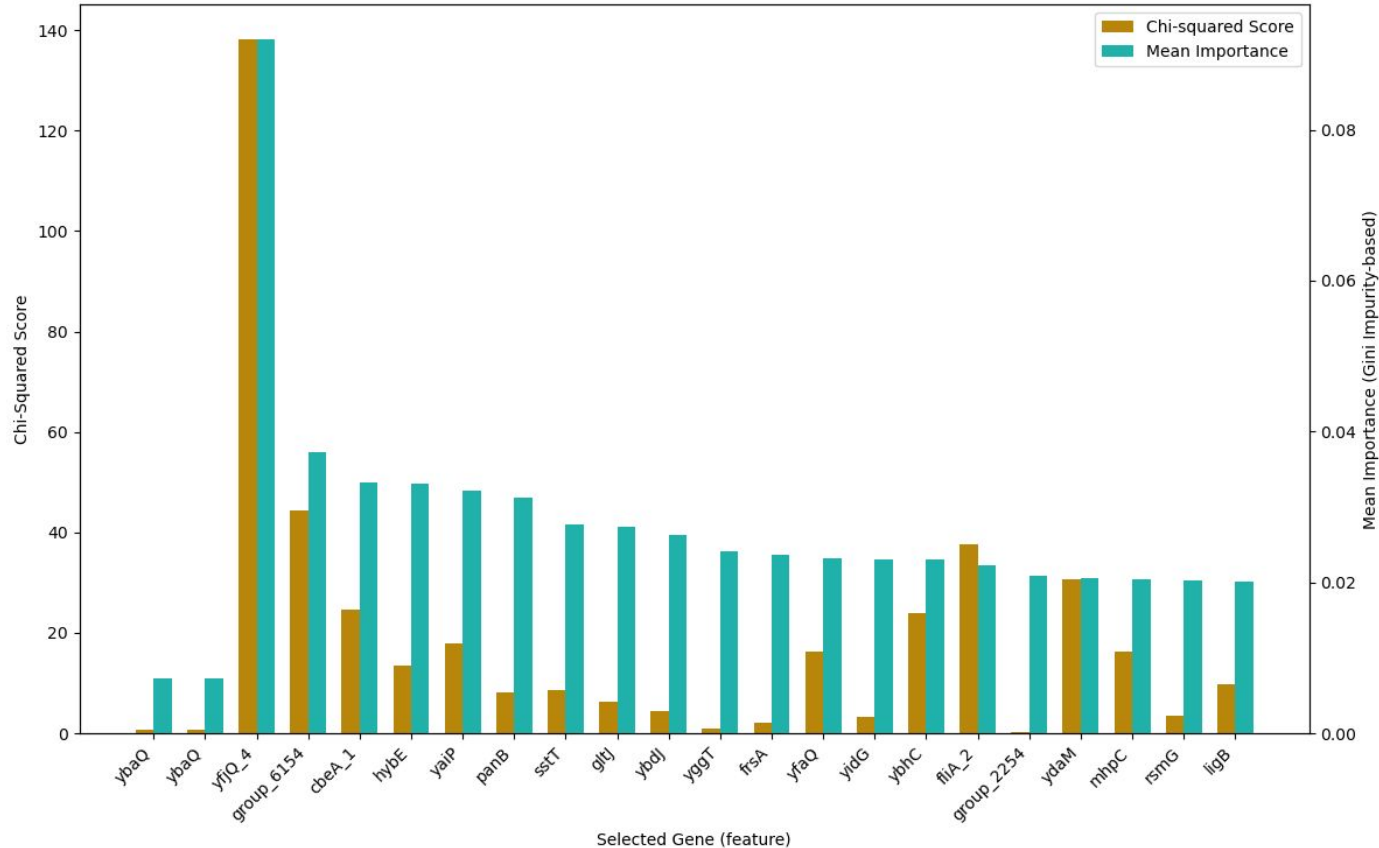
# Evaluating the Predictive Classifiers



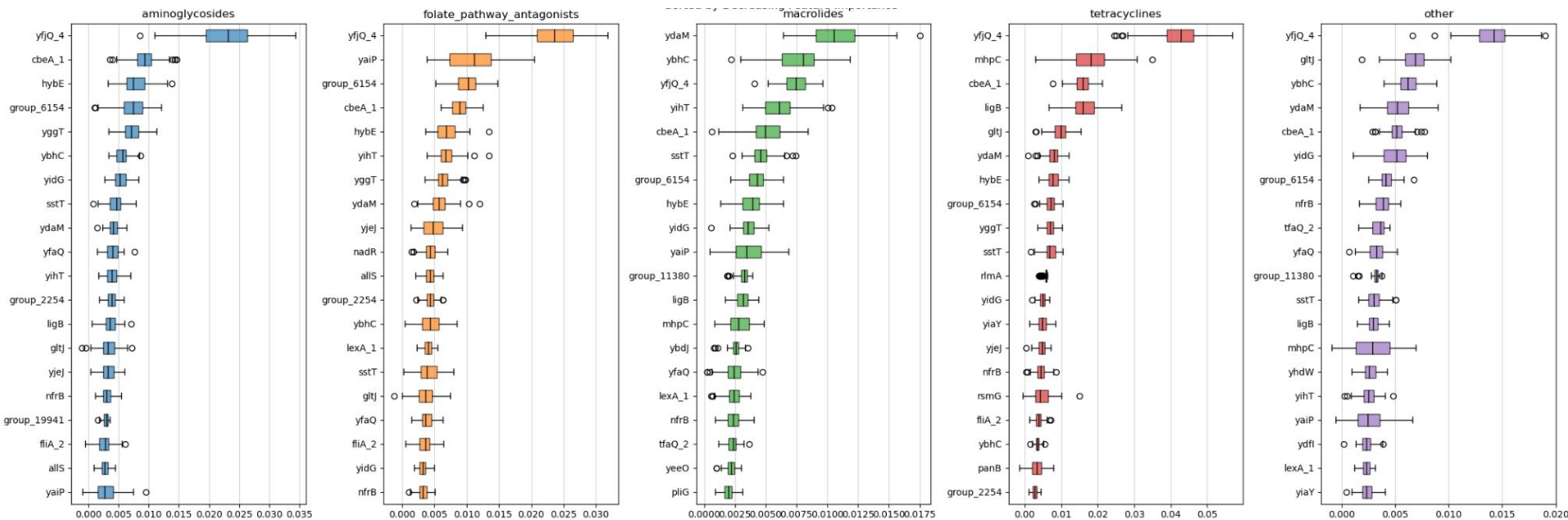


# Comparing Feature Importance Measures

Training Set - Chi Squared and Impurity-based (Gini) Feature Importance for Selected Features  
aminoglycosides



# Permutation Importance for Classifiers



# Summary & Discussion

## How well does WGS *E. coli* data predict categories of MDR?

- The *mutS* gene and its paralogs do not occur in the feature clusters that best predict any category of AMR.
- Gene features other than *mutS* best predict MDR. ( $H_1$ )
- Mean classifier AUC =  $0.87 \pm 0.04$ ;  $0.62 \leq AP \leq 0.80$

## Third Research Question

What are the gene contents of the feature clusters that best predict MDR?

- Avoid self-evidence from known AMR genes.
- Investigate three highly ranked feature clusters.
- Evaluate model explainability against literature.

# Top Ranked Cluster *yfjQ\_4* Annotations

Gene (Roary)	Annotation (Roary)	Protein names (independent BLAST)
<i>group_16177</i>	<i>hypothetical protein</i>	LPD25 domain-containing protein
<i>group_10103</i>	<i>hypothetical protein</i>	Protein TrbI
<i>group_10424</i>	<i>hypothetical protein</i>	TraK lipoprotein
<i>group_12760</i>	<i>hypothetical protein</i>	Protein TraQ
<i>group_7878</i>	<i>hypothetical protein</i>	Protein TraW
<i>group_19502</i>	<i>hypothetical protein</i>	Protein TraJ
<i>group_4098</i>	<i>hypothetical protein</i>	Protein TraB
<i>group_24638</i>	<i>hypothetical protein</i>	Terminase
<i>group_9019</i>	<i>hypothetical protein</i>	Protein TraE
<i>group_11161</i>	<i>hypothetical protein</i>	Protein TrbG
<i>group_5897</i>	<i>hypothetical protein</i>	Protein TraP
<i>group_4701</i>	<i>hypothetical protein</i>	Protein TrbJ
<i>group_9337</i>	<i>hypothetical protein</i>	Protein TraV
<i>group_17382</i>	<i>hypothetical protein</i>	Protein TraL
<i>group_2543</i>	<i>hypothetical protein</i>	Protein TrbB
<i>group_4818</i>	<i>hypothetical protein</i>	X polypeptide, ORF 19, ORF 169, P19 protein
<i>umuC_2</i>	SOS mutagenesis and repair	
<i>traM</i>	Relaxosome protein TraM	
<i>traY</i>	Relaxosome protein TraY	
<i>group_2526</i>	<i>hypothetical protein</i>	
<i>ccdB</i>	<i>hypothetical protein</i>	
<i>traA</i>	Pilin	
<i>yfjQ_4</i>	CP4-57 prophage; predicted protein	
<i>group_5903</i>	<i>hypothetical protein</i>	

# *ydaM* Includes Rac Prophage & Transcription Genes

www.nature.com/scientificreports

## SCIENTIFIC REPORTS

**OPEN** Physiological Function of Rac Prophage During Biofilm Formation and Regulation of Rac Excision in *Escherichia coli* K-12

Received: 29 May 2015  
Accepted: 07 October 2015  
Published: 04 November 2015

Xiaoxiao Liu<sup>1,\*</sup>, Yangmei Li<sup>1,2,\*</sup>, Yunxue Guo<sup>1</sup>, Zhenshun Zeng<sup>1,2</sup>, Baiyuan Li<sup>1,2</sup>, Thomas K. Wood<sup>3,4</sup>, Xingsheng Cai<sup>1</sup> & Xiaoxue Wang<sup>1</sup>

**HHS Public Access**  
Author manuscript  
*Trends Biochem Sci.* Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:  
*Trends Biochem Sci.* 2018 September ; 43(9): 668–684. doi:10.1016/j.tibs.2018.06.011.

### How macrolide antibiotics work

Nora Vázquez-Laslop and Alexander S. Mankin  
Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA

### Abstract

Macrolide antibiotics inhibit protein synthesis by targeting the bacterial ribosome. They bind at the nascent peptide exit tunnel and partially occlude it. Thus, macrolides have been viewed as ‘tunnel plugs’ that stop synthesis of every protein. More recent evidence, however, demonstrates that macrolides selectively inhibit translation of a subset of cellular proteins and that their action critically depends on the nascent protein sequence and on the antibiotic structure. Therefore, macrolides emerge as modulators of translation rather than global inhibitors of protein synthesis.



# Highly Ranked Cluster *cbeA\_1* Annotations

ami	fol	mac	tet	oth	Gene (Roary)	Non-unique	Annotation (Roary)	Protein names (BLAST)
1	3	4	2	4	<i>group_13421</i>		<i>hypothetical protein</i>	Zinc ribbon domain-containing protein
1	3	4	2	4	<i>group_4945</i>		<i>hypothetical protein</i>	RelB antitoxin
1	3	4	2	4	<i>group_22310</i>		<i>hypothetical protein</i>	Uncharacterized protein
1	3	4	2	4	<i>group_7475</i>		<i>hypothetical protein</i>	DUF3310 domain-containing protein
1	3	4	2	4	<i>group_9598</i>		<i>hypothetical protein</i>	Chemotaxis protein
1	3	4	2	4	<i>yfjJ</i>		<i>hypothetical protein</i>	Uncharacterized protein YagK
1	3	4	2	4	<i>group_7384</i>		<i>hypothetical protein</i>	Regulatory protein
1	3	4	2	4	<i>group_5273</i>		<i>hypothetical protein</i>	Phospholipase
1	3	4	2	4	<i>group_6844</i>		<i>hypothetical protein</i>	C8 domain-containing protein
1	3	4	2	4	<i>group_13325</i>		<i>hypothetical protein</i>	Lipoprotein
1	3	4	2	4	<i>group_4579</i>		<i>hypothetical protein</i>	
1	3	4	2	4	<i>cbeA_1</i>		CP4-44 prophage; antitoxin of the CbtA-CbeA toxin-antitoxin system	
1	3	4	2	4	<i>yeeS_1</i>		CP4-44 prophage; predicted DNA repair protein	
1	3	4	2	4	<i>php_1</i>		putative hydrolase	
1	3	4	2	4	<i>group_684</i>		<i>hypothetical protein</i>	
1	3	4	2	4	<i>group_257</i>	<i>yfjQ_1</i>	CP4-57 prophage; predicted protein	
1	3	4	2	4	<i>group_1508</i>	<i>yeeP_1</i>	CP4-44 prophage; predicted GTP-binding protein	

# Comparison of Best Predictors Across Classifiers

0-based Rank	aminoglycosides (ami)	folate pathway (fol)	macrolides (mac)	tetracyclines (tet)	others (oth)
0	<i>yfjQ_4</i>	<i>yfjQ_4</i>	<i>ydaM</i>	<i>yfjQ_4</i>	<i>yfjQ_4</i>
1	<i>cbeA_1</i>	<i>yaiP</i>	<i>ybhC</i>	<i>mhpC</i>	<i>gltJ</i>
2	<i>hybE</i>	<i>group_6154</i>	<i>yfjQ_4</i>	<i>cbeA_1</i>	<i>ybhC</i>
3	<i>group_6154</i>	<i>cbeA_1</i>	<i>yihT</i>	<i>ligB</i>	<i>ydaM</i>
4	<i>yggT</i>	<i>hybE</i>	<i>cbeA_1</i>	<i>gltJ</i>	<i>cbeA_1</i>



# Summary & Discussion

## What are the gene contents of the feature clusters that best predict MDR?

- Absence of known AMR gene determinants in best features.
- HGT, phage, and transcriptional regulation genes best predict MDR.
  - Induced mutable states (SOS) vs. structurally deficient hypermutators (MMR)
- Concerning presence of *tra* family (F plasmid) genes in cluster ***yfjQ\_4***.
- Unique combinations of feature clusters (gene sets) differentiate between categories of AMR.

## Conclusion

MDR does not correlate dependently with deficient MMR (MutS).

The proliferation of AMR genes into MDR phenotypes is driven by horizontal gene transfer and temporary states of increased mutation.

Phylogenetics and machine learning combine well to generate genotype-to-phenotype hypotheses for polygenic traits.

## Future Work

- Investigate the three exceptional positions in MutS for:
  - Protein conformational changes due to amino acid substitutions.
  - Count and compare mutations between isolates with those specific substitutions.
- Further explore the gene feature clusters that best predict MDR.
  - Find SNPs and amino acid substitutions from same genes that clustered differently.
  - Identify a genetic signature that indicates presence of mutations that allow MDR.

# Significance

- **Intellectual Merit**

- Refine current understandings of multidrug resistance in enteric bacteria.
- Demonstrate underutilized data sets from NCBI PD.
- Explore combining phylogenetic and machine learning approaches to build improved genotype-to-phenotype hypotheses.

- **Broader Impacts**

- Improve current disease surveillance analyses.
- Insights for anticipating the direction of the silent pandemic.
- Prolong the efficacy of current antibiotics.

Thank you for  
your kind  
attention.



## Acknowledgments

- *Funding Sources*
  - Dr. Daniel Janies
  - Department of Bioinformatics and Genomics
  - GAANN program
- *Committee Members*
  - Dr. Jun-tao Guo
  - Dr. Alex Dornburg
  - Dr. Adam Reitzel
- *Attendees*
  - Would not be here without your support.

# Questions

# Dissertation Defense of David C. Brown

Department of Bioinformatics and Genomics - College of Computing and Informatics,  
The University of North Carolina at Charlotte

Committee Chair: Dr. Daniel Janies  
Committee: Dr. Jun-tao Guo, Dr. Alex Dornburg, Dr. Adam Reitzel



# List of Abbreviations & Terms

- AMR
  - Antimicrobial resistance
- CDC
  - Centers for Disease Control and Prevention
- CFSAN
  - Center for Food Safety and Applied Nutrition
- CRE
  - Carbapenem-resistant Enterobacteriaceae
- *E. coli*
  - *Escherichia coli*
- FDA
  - U.S. Food and Drug Administration
- MDR
  - Multidrug resistance
- MMR
  - Methyl-directed mismatch repair
- MutS
  - MutatorS
- Pathogenicity
  - Potential ability to produce disease
- SNP
  - Single nucleotide polymorphism
- Virulence
  - Degree of disease producing power



# References

1. Centers for Disease Control and Prevention. (2019, March 15). Glossary of terms related to antibiotic resistance. Centers for Disease Control and Prevention. Retrieved December 1, 2021, from <https://www.cdc.gov/narms/resources/glossary.html>
2. Deng, X., Cao, S., & Horn, A. L. (2021). Emerging Applications of Machine Learning in Food Safety. *Annual Review of Food Science and Technology*, 12, 513–538. <https://doi.org/10.1146/annurev-food-071720-024112>
3. Ford, C.T., Zenarosa, G.L., Smith, K.B., Brown, D.C., Williams, J. and Janies, D. (2020), Genetic capitalism and stabilizing selection of antimicrobial resistance genotypes in *Escherichia coli*. *Cladistics*, 36: 348-357. <https://doi.org/10.1111/cla.12421>
4. Obmolova, G., Ban, C., Hsieh, P., & Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, 407(6805), 703–710. <https://doi.org/10.1038/35037509>
5. Haendiges, J., Davidson, G. R., Pettengill, J. B., Reed, E., Ramachandran, P., Blessington, T., Miller, J. D., Anderson, N., Myoda, S., Brown, E. W., Zheng, J., Tikekar, R., & Hoffmann, M. (2021). Genomic evidence of environmental and resident *Salmonella* Senftenberg and Montevideo contamination in the pistachio supply-chain. *PloS One*, 16(11), e0259471. <https://doi.org/10.1371/journal.pone.0259471>
6. LeClerc, J. E., Li, B., Payne, W. L., & Cebula, T. A. (1996). High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, 274(5290), 1208–1211. <https://doi.org/10.1126/science.274.5290.1208>
7. Shapiro-Ilan, D. I., Fuxa, J. R., Lacey, L. A., Onstad, D. W., & Kaya, H. K. (2005). Definitions of pathogenicity and virulence in invertebrate pathology. *Journal of invertebrate pathology*, 88(1), 1–7. <https://doi.org/10.1016/j.jip.2004.10.003>
8. Sheng H, Huang J, Han Z, Liu M, Lü Z, Zhang Q, Zhang J, Yang J, Cui S and Yang B (2020) Genes and Proteomes Associated With Increased Mutation Frequency and Multidrug Resistance of Naturally Occurring Mismatch Repair-Deficient *Salmonella* Hypermutators. *Front. Microbiol.* 11:770. doi: 10.3389/fmicb.2020.00770
9. Zhang, S., Li, S., Gu, W., den Bakker, H., Boxrud, D., Taylor, A., Roe, C., Driebe, E., Engelthaler, D. M., Allard, M., Brown, E., McDermott, P., Zhao, S., Bruce, B. B., Trees, E., Fields, P. I., & Deng, X. (2019). Zoonotic Source Attribution of *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United States. *Emerging infectious diseases*, 25(1), 82–91. <https://doi.org/10.3201/eid2501.180835>