Deconstructing Polygenic Inheritance Using Bioinformatics and Machine Learning

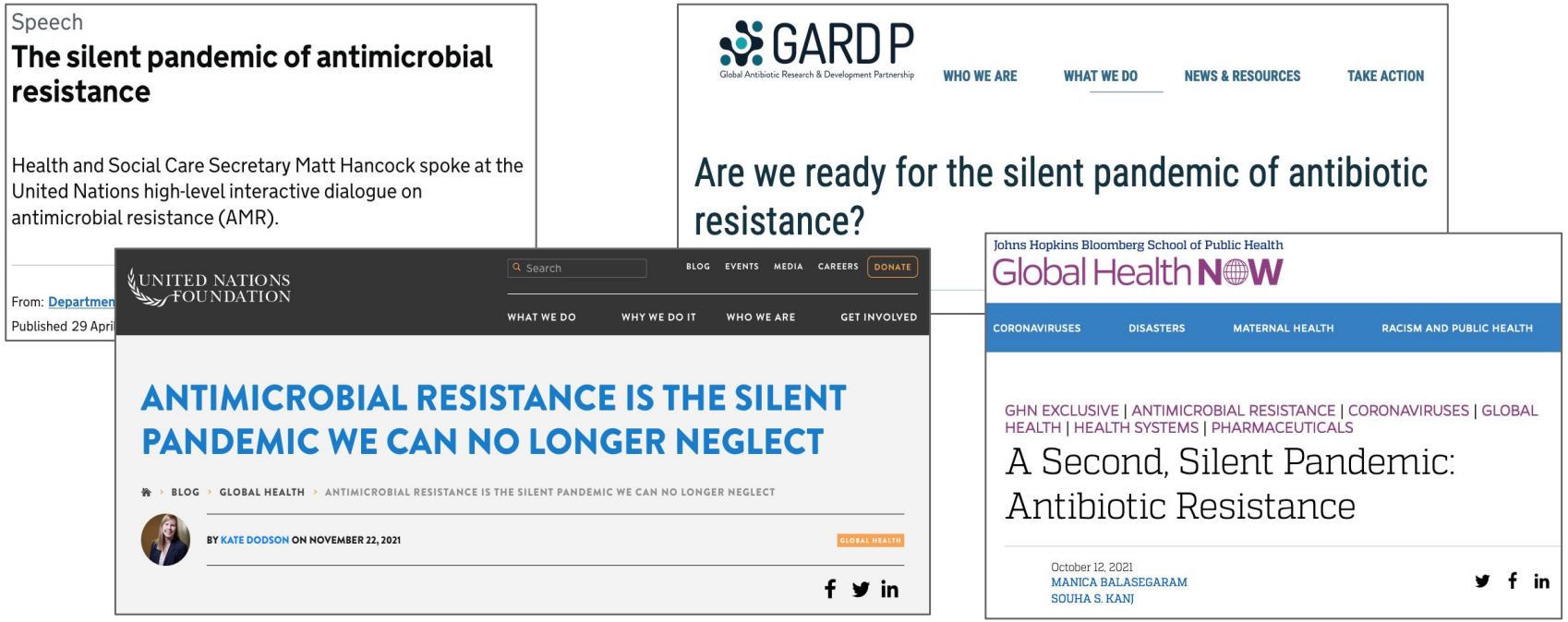
David C. Brown
davidcbrown345@gmail.com

Ph.D. Bioinformatics and Computational Biology
University of North Carolina at Charlotte

Dissertation

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

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



Selected screen captures from Google search results for "silent pandemic resistance". Accessed Dec. 12, 2021.

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.

Image of the antibiotic classes tested for resistance in *E. coli*. Accessed Dec. 12, 2021 at https://www.cdc.gov/narms/antibiotics-tested.html.

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
Phenicols	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

Call for Action

DOMAIN 6 EVOLUTION AND GENOMICS

Salmonella Genomics in Public Health and Food Safety

ERIC W. BROWN,[a] REBECCA BELL,[a] GUODONG ZHANG,[a] RUTH TIMME,[a] JIE ZHENG,[a] THOMAS S. HAMMACK,[a] AND MARC W. ALLARD[a]

[a]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,

Current Issues

WILEY-BLACKWELL

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

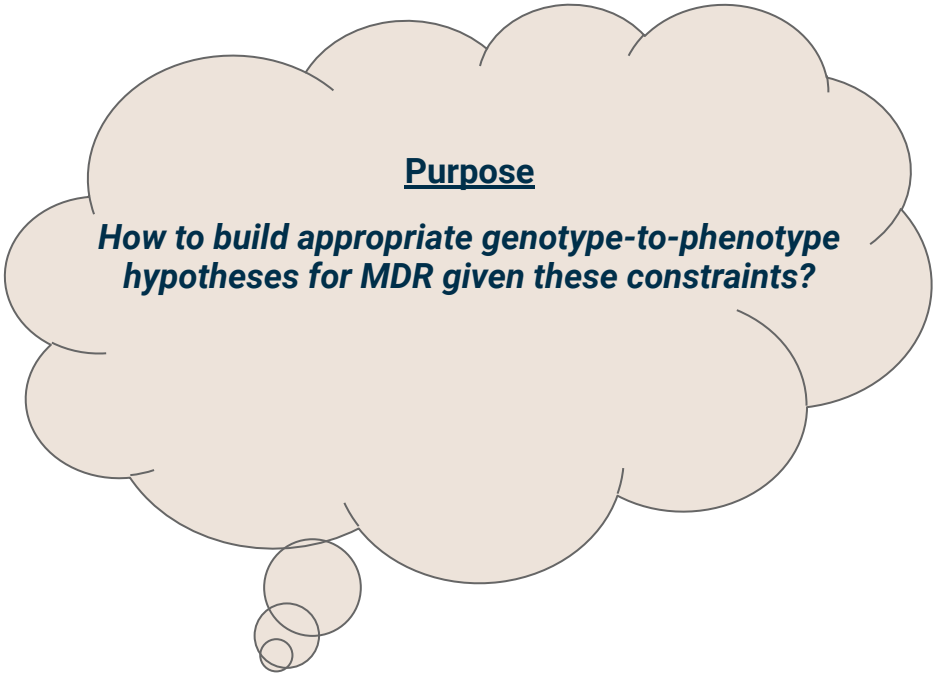
Received 7 May 2020; Revised 10 June 2020; Accepted 17 June 2020

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.

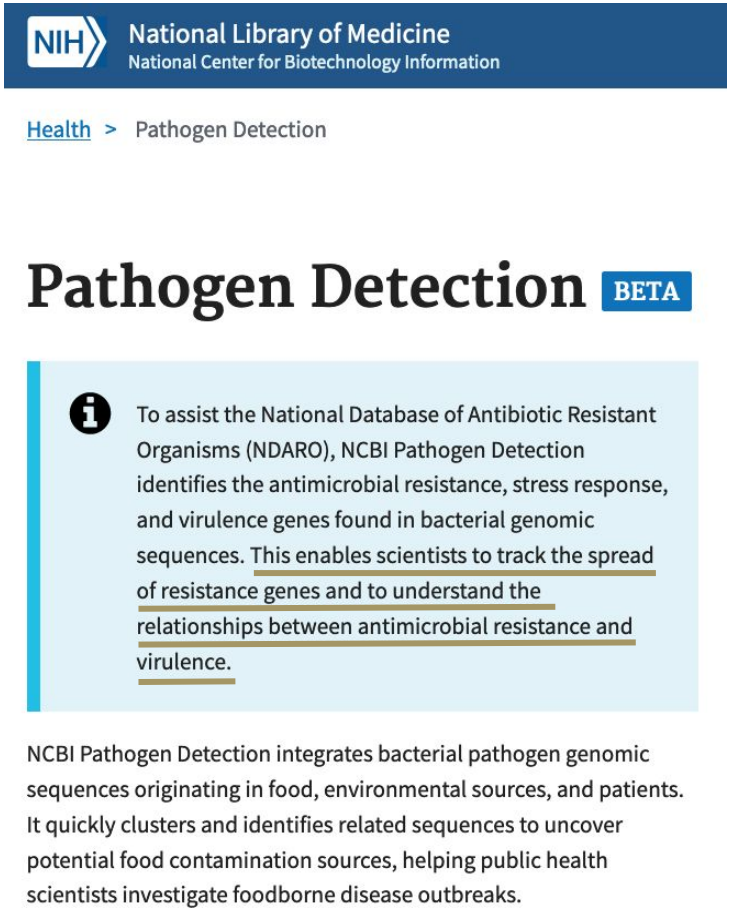
Current Issues - cont.

- 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.



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.



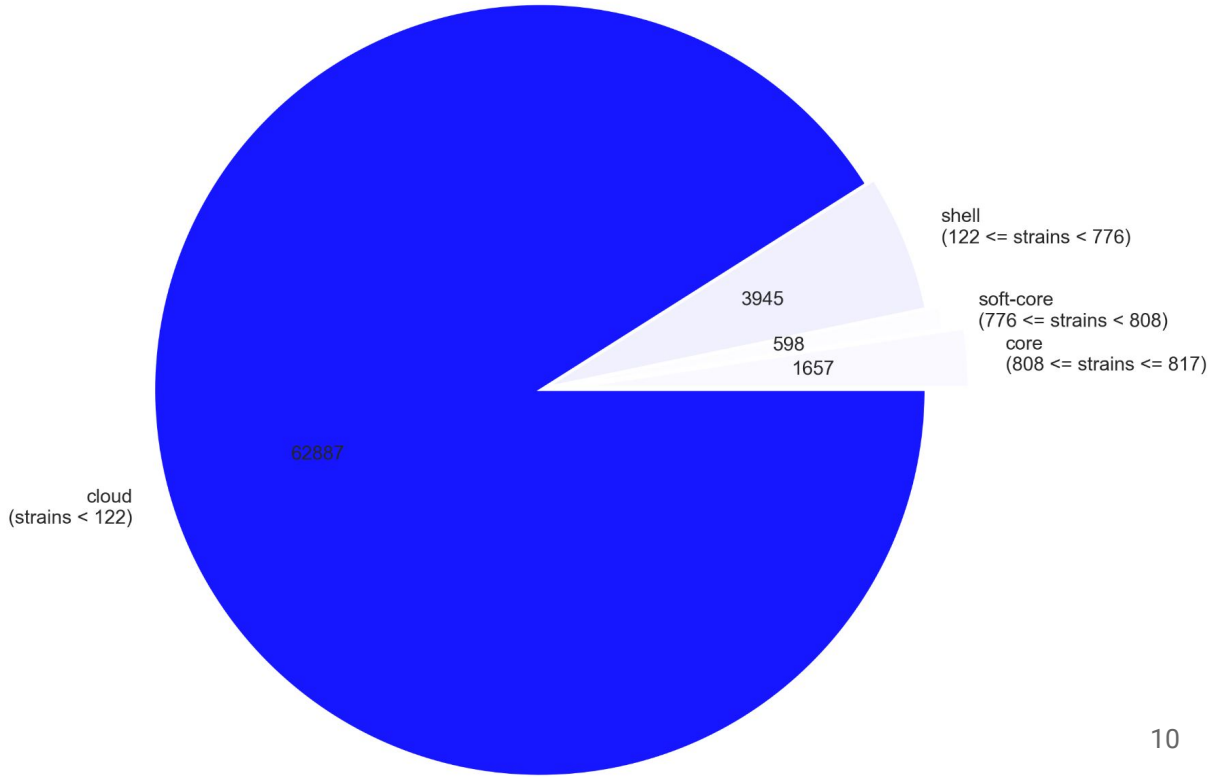
Screenshot from NCBI Pathogen Detection. Accessed Dec. 12, 2021 at https://www.ncbi.nlm.nih.gov/pathogens.

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 susce

CLSI Class	Antimicrobial Agent
Aminoglycosides	Amikacin
	Gentamicin
	Kanamycin
	Streptomycin†
β–lactam combination agents	Amoxicillin-clavulanic acid
Cephems	Cefoxitin
	Ceftiofur
	Ceftriaxone‡
	Cephalothin

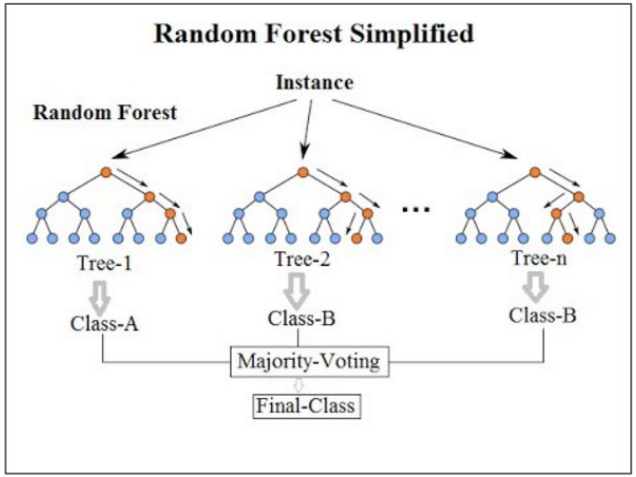
Resistance Category	3 Letter Resistance Gene Prefix
aminoglycosides	*aac, aad, ant, aph, arm, rmt*
beta-lactam combination agents	*bla*
cephems	*abc*
folate pathway antagonists	*dfr, sul*
macrolides	*ere, erm, mef, mph, msr*
nucleosides	*sat*
penicillins	*amp*
phenicols	*cat, flo*
quinolones	*qep*
tetracyclines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

Modified image of the 10 antibiotic classes tested for resistance in *E. coli.* Accessed Dec. 12, 2021 at https://www.cdc.gov/narms/antibiotics-tested.html.

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



Right image basic conceptualization of a random forest from Wikipedia. Accessed April 4, 2023 at https://upload.wikimedia.org/wikipedia/commons/7/76/Random_forest_diagram_complete.png.

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 Mutator S explain the development of MDR in *E. coli*?

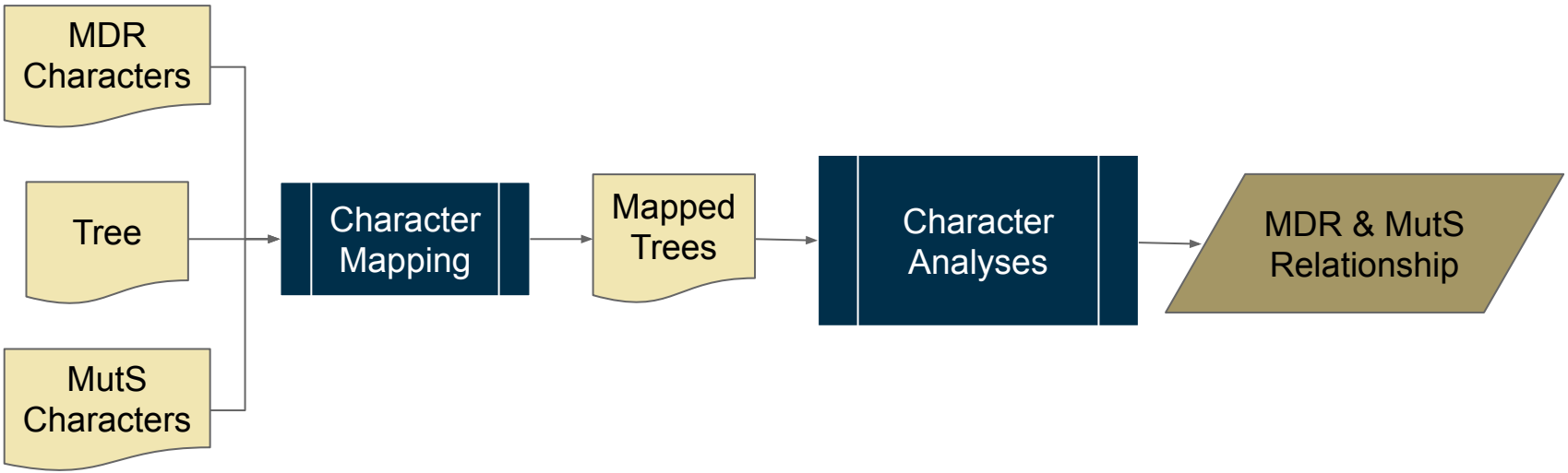
H_0

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

H_1

Repair deficient MutS variants correlate with MDR.

Identifying Phylogenetic Relationship Between Multidrug Resistance and Mutator S



Change to Protein Space: MutS Consensus Sequence

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

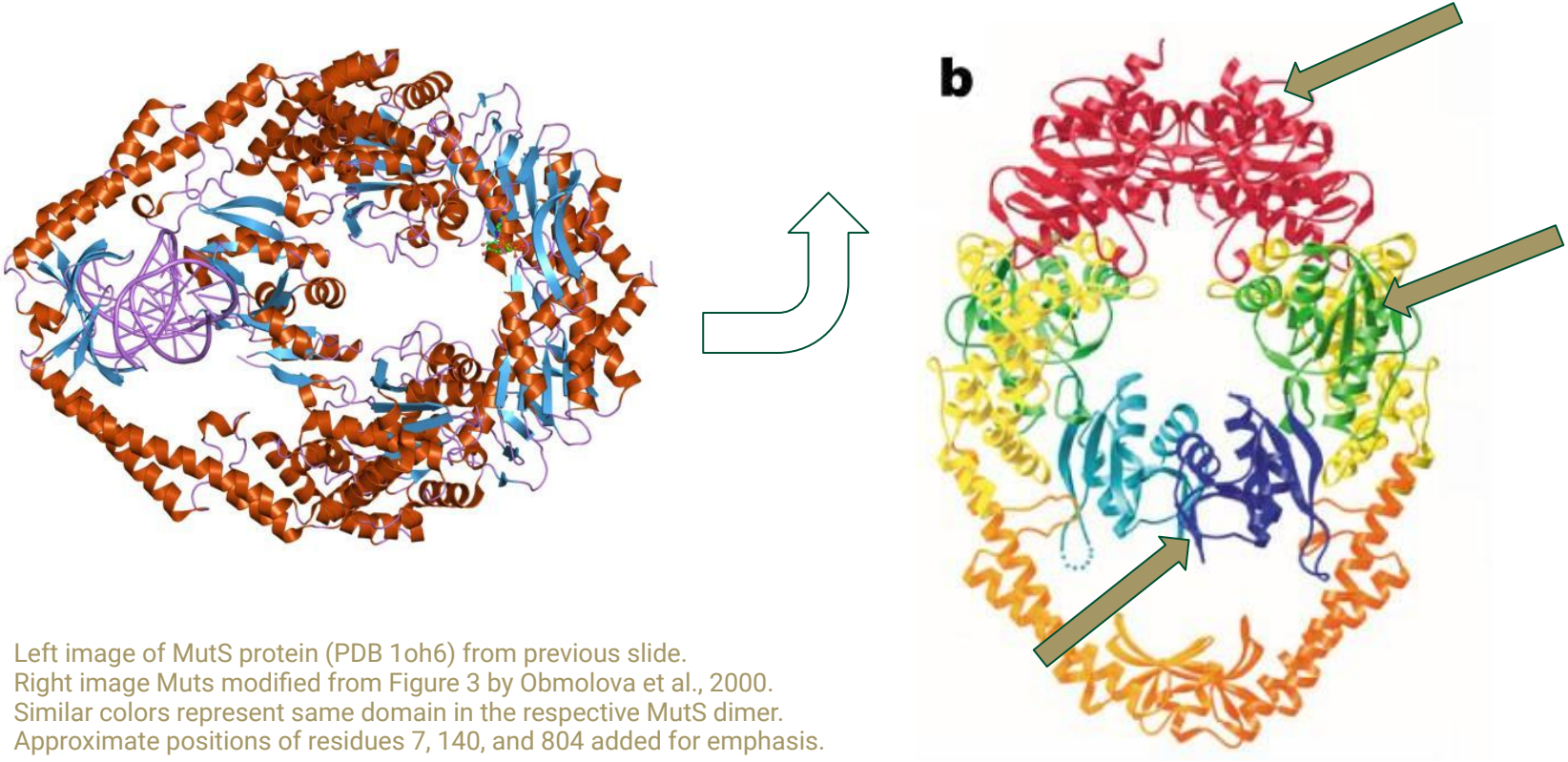
Isolate	pos_1	pos_2	pos_3	pos_*n*	MDR+
GCF_000000001	1	0	1	0	0
GCF_000000002	1	1	0	1	1
GCF_000000003	1	1	1	1	1
GCF_000000004	1	1	1	1	0
GCF_000000005	1	0	1	1	0

Weighted AIC Supports Independent Model of Correlation Between MutS Residues and MDR as Calculated by Pagel's 1994 Algorithm

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).

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: RQ1

Does Mutator S explain the development of MDR in *E. coli*?

- No observation of literature 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?

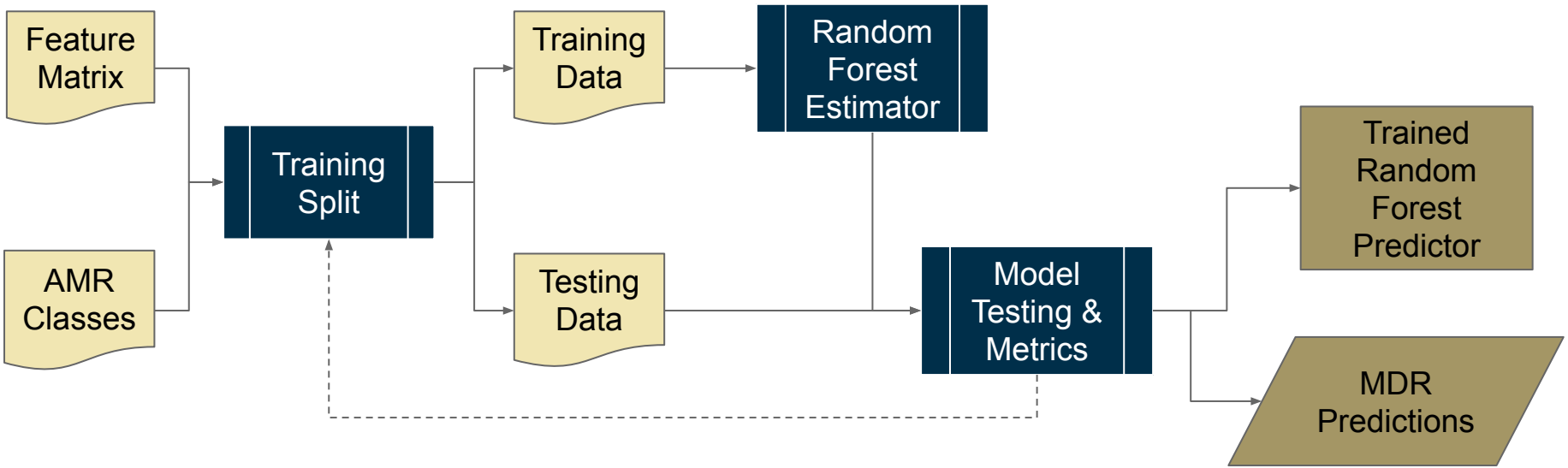
H_0

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

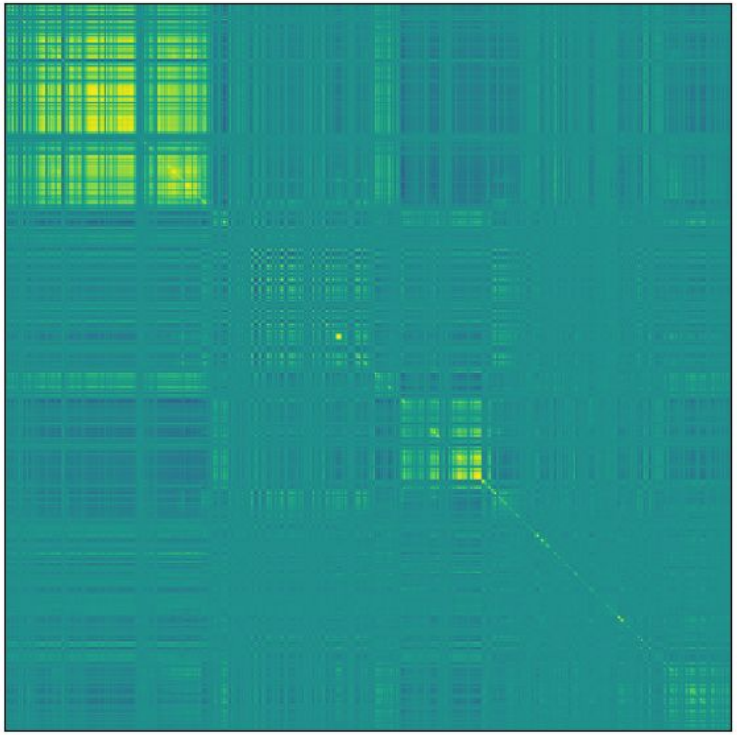
H_1

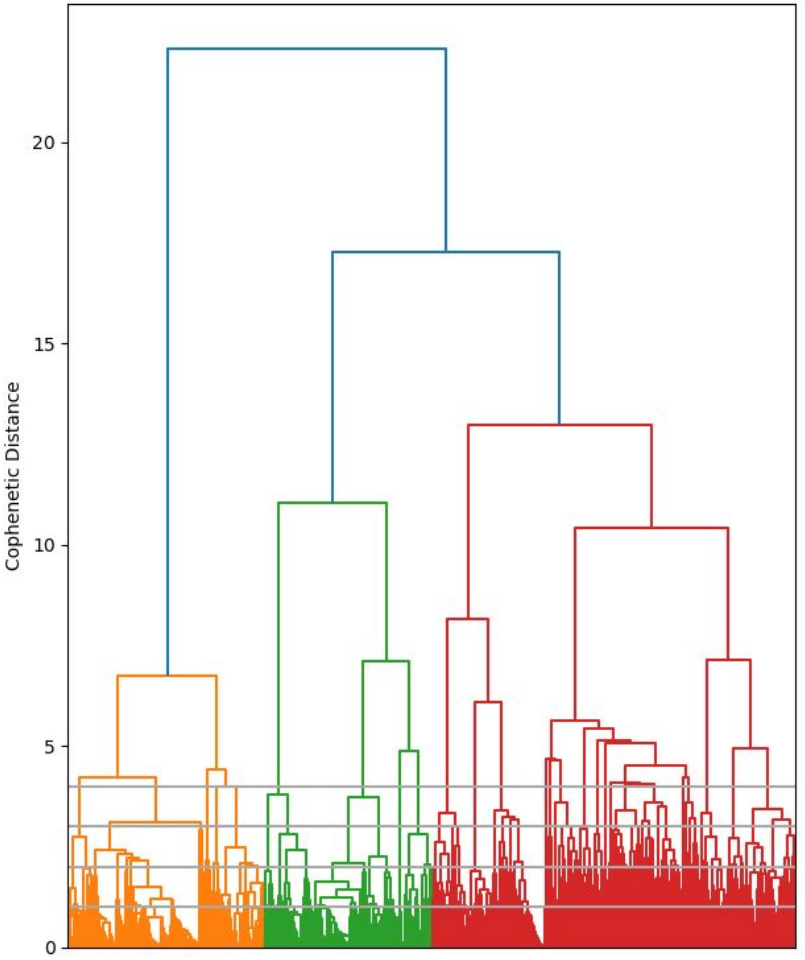
Gene features other than *mutS* best predict MDR.

Predicting MDR from WGS Annotations – Workflow

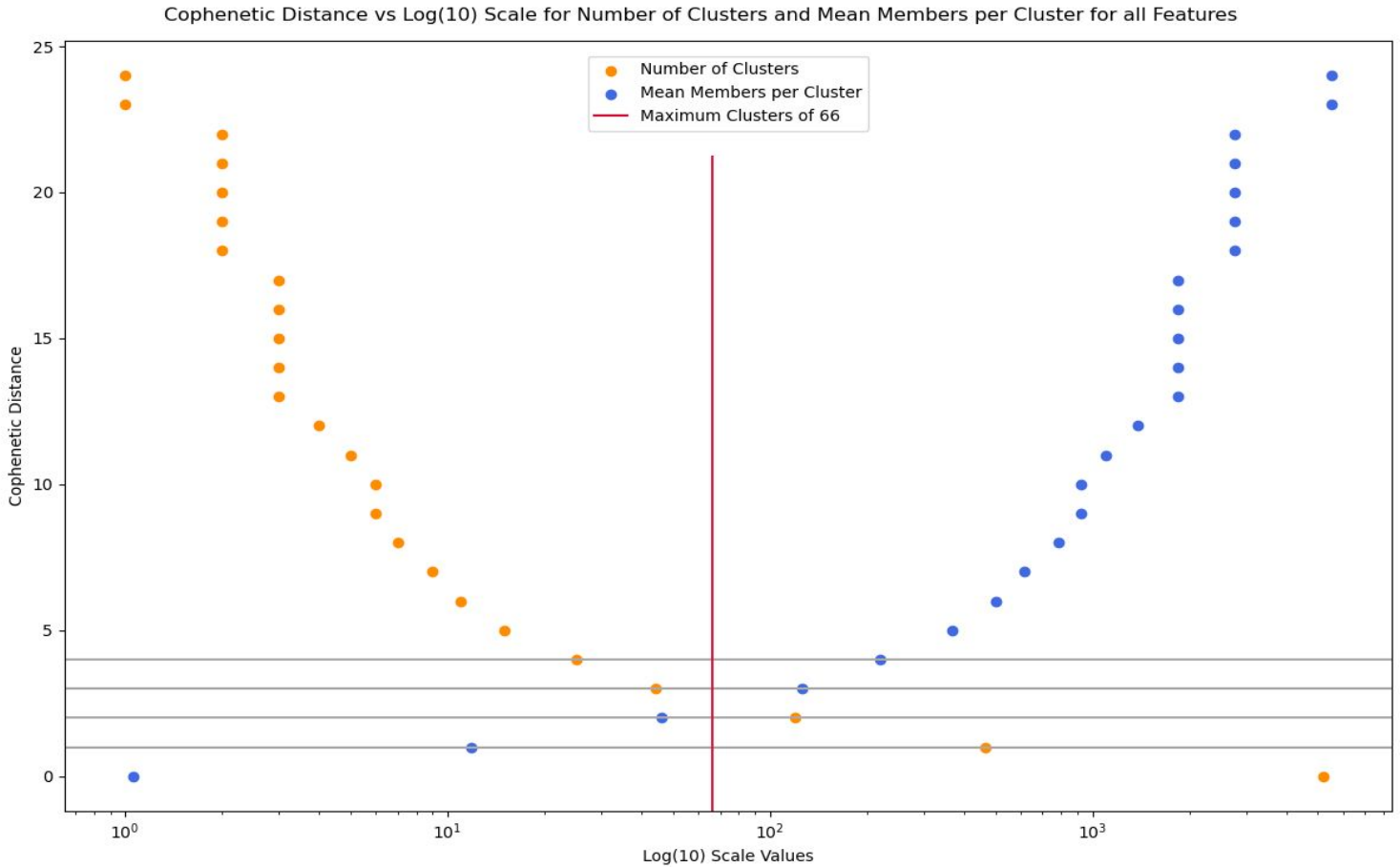


Feature Correlation Matrix and Hierarchical Clustering

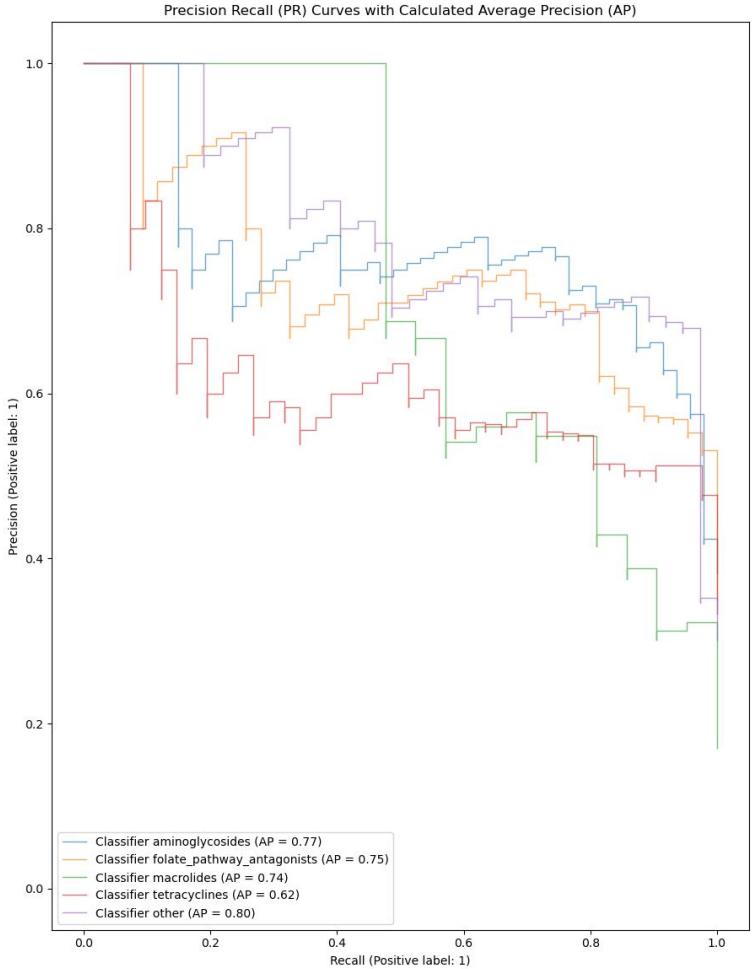


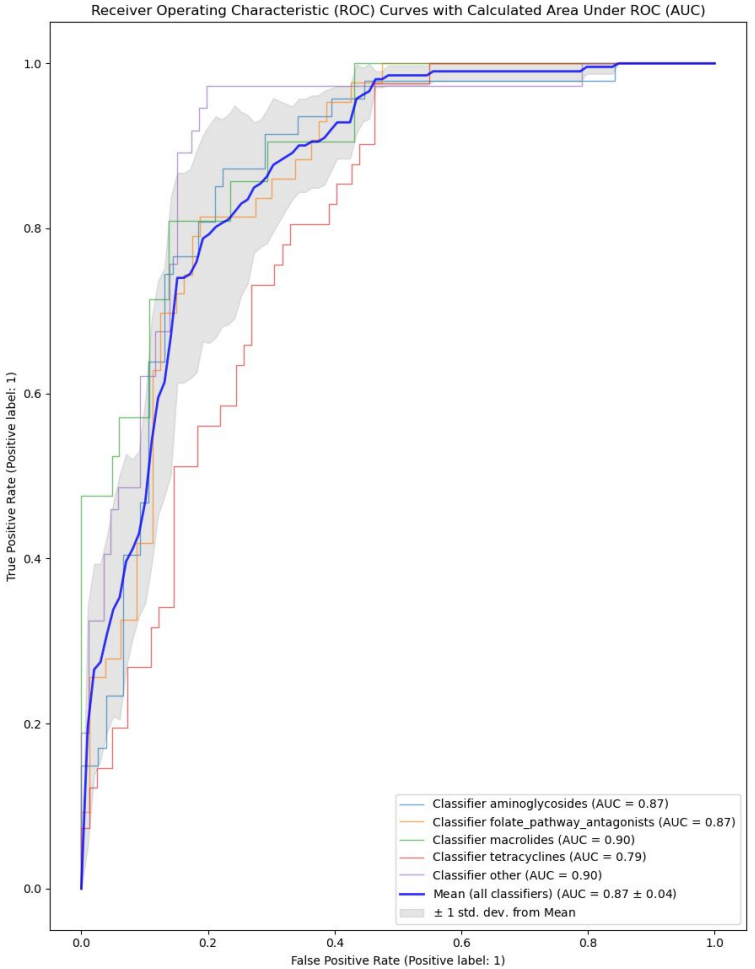


Solving Collinearity by Clustering Features

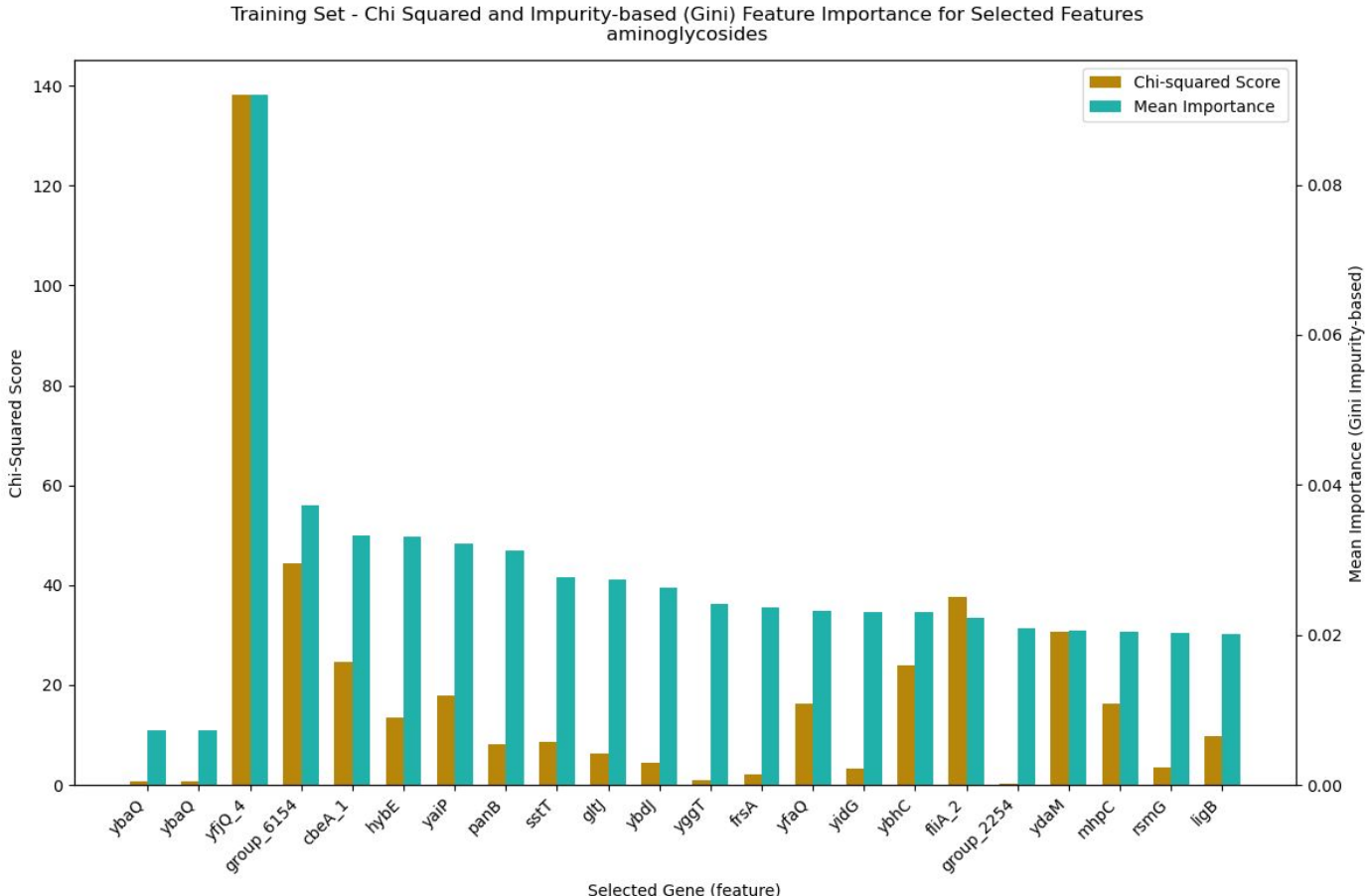


Evaluating the Predictive Classifiers

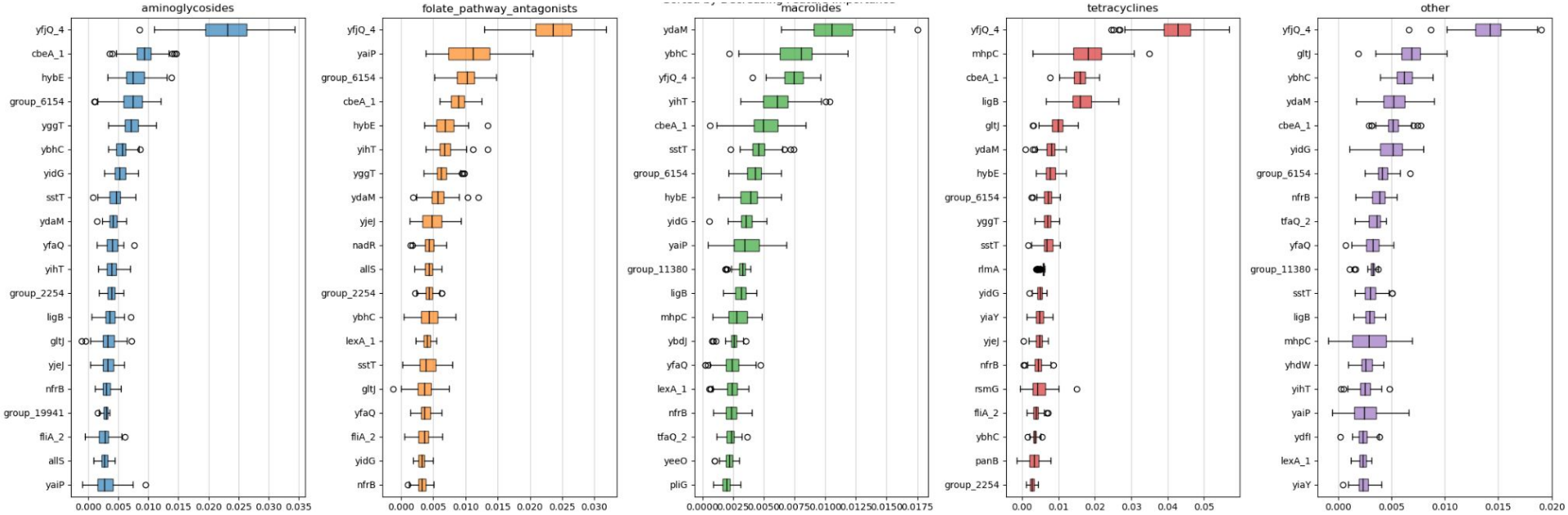




Comparing Feature Importance Measures



Permutation Importance for Classifiers



Summary: RQ2

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 ± 0.04; 0.62 ≤ AP ≤ 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.
- Develop a hypothesis for MDR polygenic inheritance.

Top Ranked Cluster *yfjQ_4* Annotations

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

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[1,*], Yangmei Li[1,2,*], Yunxue Guo[1], Zhenshun Zeng[1,2], Baiyuan Li[1,2], Thomas K. Wood[3,4], Xingsheng Cai[1] & Xiaoxue Wang[1]

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	*group_13421*		*hypothetical protein*	Zinc ribbon domain-containing protein
1	3	4	2	4	*group_4945*		*hypothetical protein*	RelB antitoxin
1	3	4	2	4	*group_22310*		*hypothetical protein*	Uncharacterized protein
1	3	4	2	4	*group_7475*		*hypothetical protein*	DUF3310 domain-containing protein
1	3	4	2	4	*group_9598*		*hypothetical protein*	Chemotaxis protein
1	3	4	2	4	*yfjJ*		*hypothetical protein*	Uncharacterized protein YagK
1	3	4	2	4	*group_7384*		*hypothetical protein*	Regulatory protein
1	3	4	2	4	*group_5273*		*hypothetical protein*	Phospholipase
1	3	4	2	4	*group_6844*		*hypothetical protein*	C8 domain-containing protein
1	3	4	2	4	*group_13325*		*hypothetical protein*	Lipoprotein
1	3	4	2	4	*group_4579*		*hypothetical protein*	
1	3	4	2	4	*cbeA_1*		CP4-44 prophage; antitoxin of the CbtA-CbeA toxin-antitoxin system	
1	3	4	2	4	*yeeS_1*		CP4-44 prophage; predicted DNA repair protein	
1	3	4	2	4	*php_1*		putative hydrolase	
1	3	4	2	4	*group_684*		*hypothetical protein*	
1	3	4	2	4	*group_257*	*yfjQ_1*	CP4-57 prophage; predicted protein	
1	3	4	2	4	*group_1508*	*yeeP_1*	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	*yfjQ_4*	*yfjQ_4*	*ydaM*	*yfjQ_4*	*yfjQ_4*
1	*cbeA_1*	*yaiP*	*ybhC*	*mhpC*	*gltJ*
2	*hybE*	*group_6154*	*yfjQ_4*	*cbeA_1*	*ybhC*
3	*group_6154*	*cbeA_1*	*yihT*	*ligB*	*ydaM*
4	*yggT*	*hybE*	*cbeA_1*	*gltJ*	*cbeA_1*

Summary: RQ3

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***.
- **Hypothesis:** Unique combinations of features that exclude known AMR determinants, but include known virulence genes best predict different categories of AMR.

Conclusions

MDR does not correlate dependently with methyl-directed mismatch repair as represented by Mutator S.

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
  - UNC Charlotte, Department of Bioinformatics and Genomics
  - GAANN program
- *Committee Members*
  - Dr. Daniel Janies
  - Dr. Jun-tao Guo
  - Dr. Alex Dornburg
  - Dr. Adam Reitzel
- *Further References*
  - Dr. Colby Ford
  - Dr. Denis Jacob Machado
  - Dr. Mesbah Uddin

Contact information available upon request.

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