BCB Dissertation Proposal of David C. Brown

•••

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

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

Welcome & Agenda

Meeting Procedure

- *Presentation*
  - ~45 min. in length.
- *Questions*
  - Will follow the presentation.
  - Preceded by break, if desired.
- *Deliberation*
  - Use of the breakout room.

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

David C. Brown, B.S. Biology
Lab of Dr. Dan Janies
BiG, CCI - UNC Charlotte

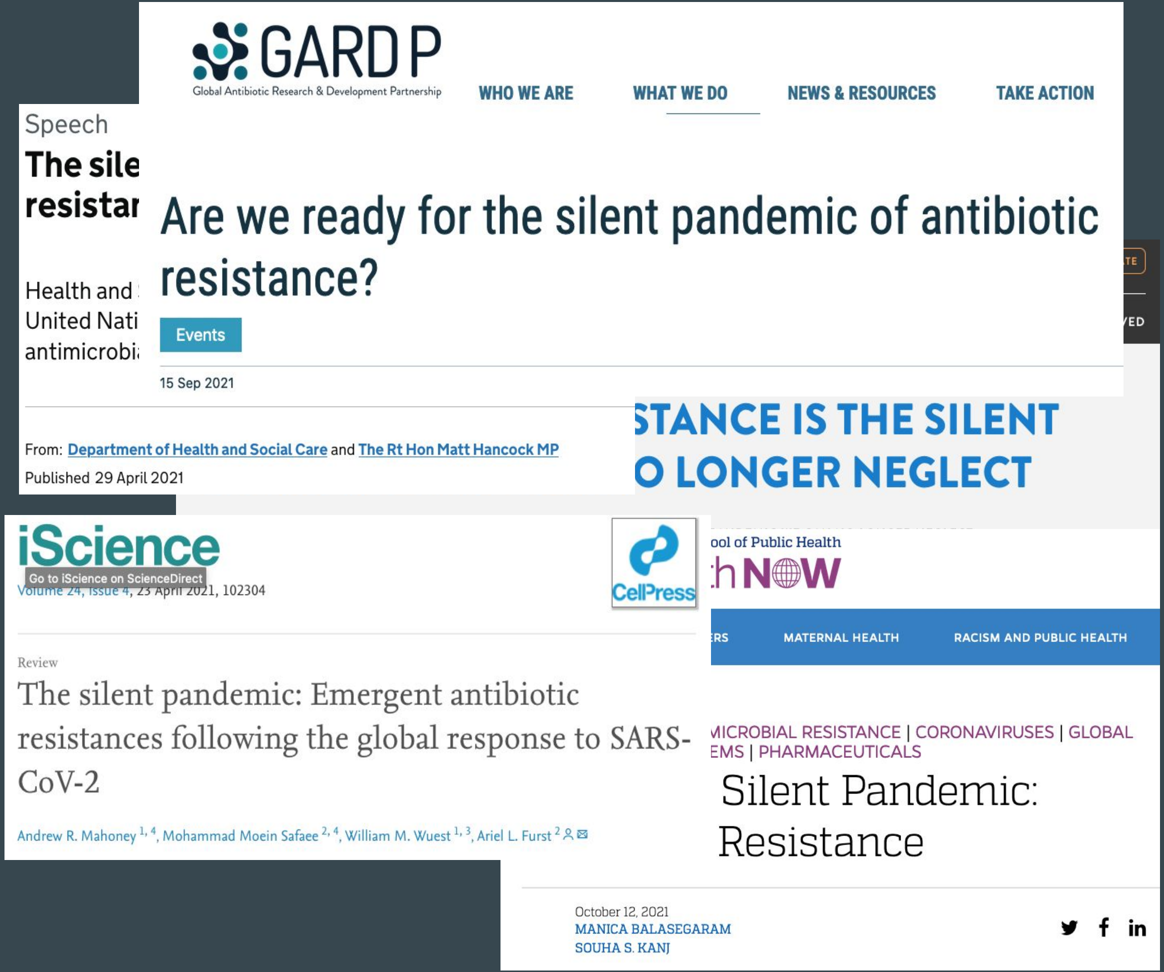
Modified image of CRE. Accessed Dec. 12, 2021 at https://www.cdc.gov/hai/organisms/cre/cre-patients.html.

Direction of Dissertation Investigations

To elucidate the evolutionary history and proliferation of global multidrug resistance (MDR) in *Escherichia coli* (*E. coli*).

Background

- Global concern is growing over the weakened efficacy of current antibiotic therapies.
- Antimicrobial resistance (AMR) is a silent pandemic.
- A microbe that exhibits resistance to three or more classes of antibiotics is termed multidrug resistant (MDR) (CDC, 2021).



Speech

The sile
resistar

Health and
United Nati
antimicrobi

From: Department of Health and Social Care and The Rt Hon Matt Hancock MP

Published 29 April 2021

Are we ready for the silent pandemic of antibiotic resistance?

15 Sep 2021

STANCE IS THE SILENT
O LONGER NEGLECT

iScience

Volume 24, Issue 4, 23 April 2021, 102304

Review

The silent pandemic: Emergent antibiotic resistances following the global response to SARS-CoV-2

Andrew R. Mahoney [1, 4], Mohammad Moein Safaee [2, 4], William M. Wuest [1, 3], Ariel L. Furst [2]

MICROBIAL RESISTANCE | CORONAVIRUSES | GLOBAL
EMS | PHARMACEUTICALS

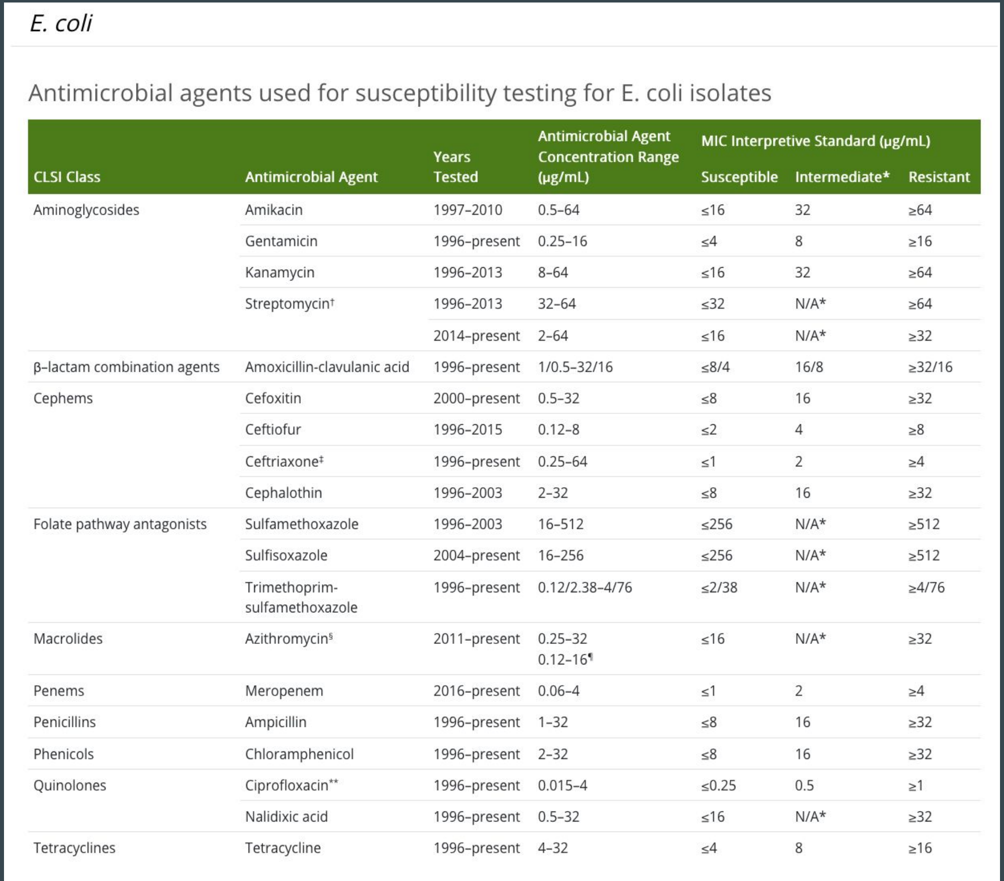
Silent Pandemic:
Resistance

October 12, 2021
MANICA BALASEGARAM
SOUHA S. KANJ

Images from Google search for "silent pandemic resistance". Accessed Dec. 12, 2021.

Background cont.

- MDR bacteria drive the silent AMR pandemic.
- MDR is a broad definition that lacks distinctions.
    - i.e., there are at minimum 120 identifiable categories of MDR for *E. coli.*
- There are a multitude of evolutionary pathways towards MDR.



E. coli

Antimicrobial agents used for susceptibility testing for E. coli isolates

				MIC Interpretive Standard (µg/mL)		
CLSI Class	Antimicrobial Agent	Years Tested	Antimicrobial Agent Concentration Range (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

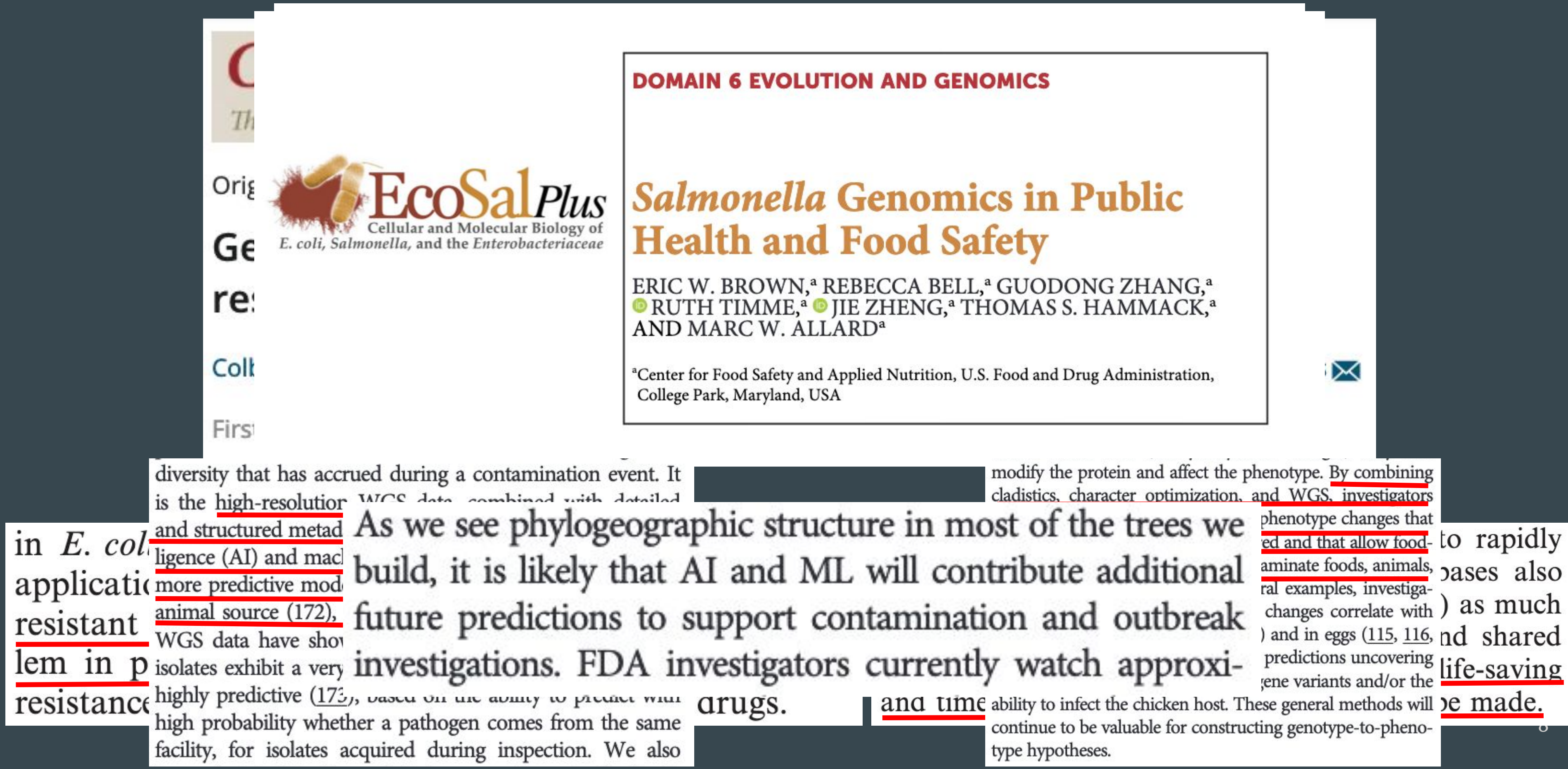
Image of the antibiotic classes tested for resistance in *E. coli.* Accessed Dec. 12, 2021 at https://www.cdc.gov/narms/antibiotics-tested.html.

Current Analysis Issues

- 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 can cloud the vertical (ancestor to descendant) signal commonly sought in phylogenetic analyses.

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

Previous Work & Current Issues



C
Th
Orig
Ge
re
Coll
Firs

EcoSal Plus
Cellular and Molecular Biology of
E. coli, Salmonella, and the *Enterobacteriaceae*

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

in *E. col*
applicati
resistant
lem in p
resistance

diversity that has accrued during a contamination event. It
is the high-resolution
and structured metad
ligence (AI) and mac
more predictive mod
animal source (172),
WGS data have sho
isolates exhibit a very
highly predictive (173
high probability whether a pathogen comes from the same
facility, for isolates acquired during inspection. We also

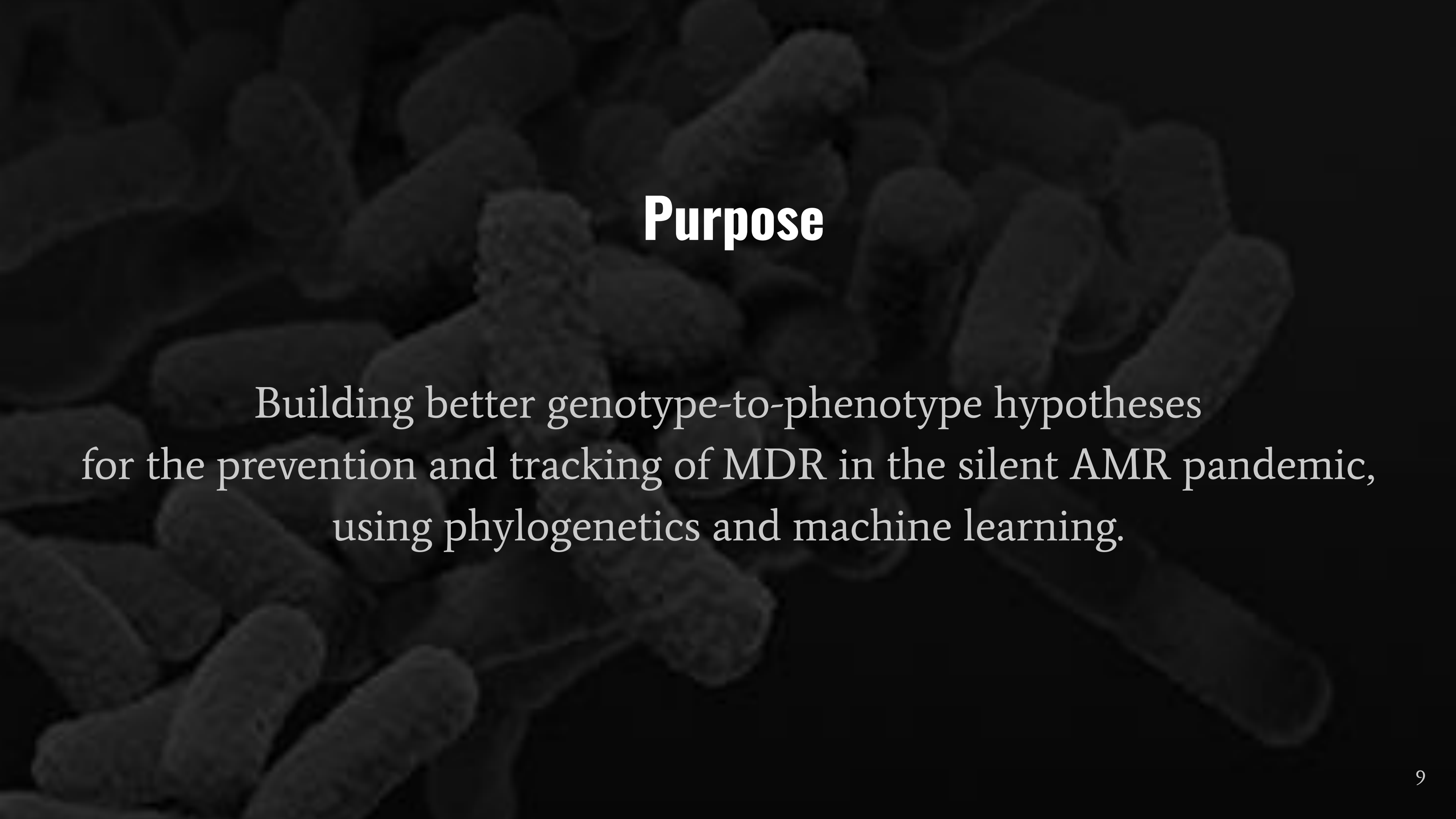
As we see phylogeographic structure in most of the trees we build, it is likely that AI and ML will contribute additional future predictions to support contamination and outbreak investigations. FDA investigators currently watch approxi-

drugs.

and time

modify the protein and affect the phenotype. By combining
cladistics, character optimization, and WGS, investigators
phenotype changes that
ed and that allow food-
aminate foods, animals,
ral examples, investiga-
changes correlate with
) and in eggs (115, 116,
predictions uncovering
gene variants and/or the
ability to infect the chicken host. These general methods will
continue to be valuable for constructing genotype-to-pheno-
type hypotheses.

to rapidly
bases also
) as much
nd shared
life-saving
be made.



Purpose

Building better genotype-to-phenotype hypotheses
for the prevention and tracking of MDR in the silent AMR pandemic,
using phylogenetics and machine learning.

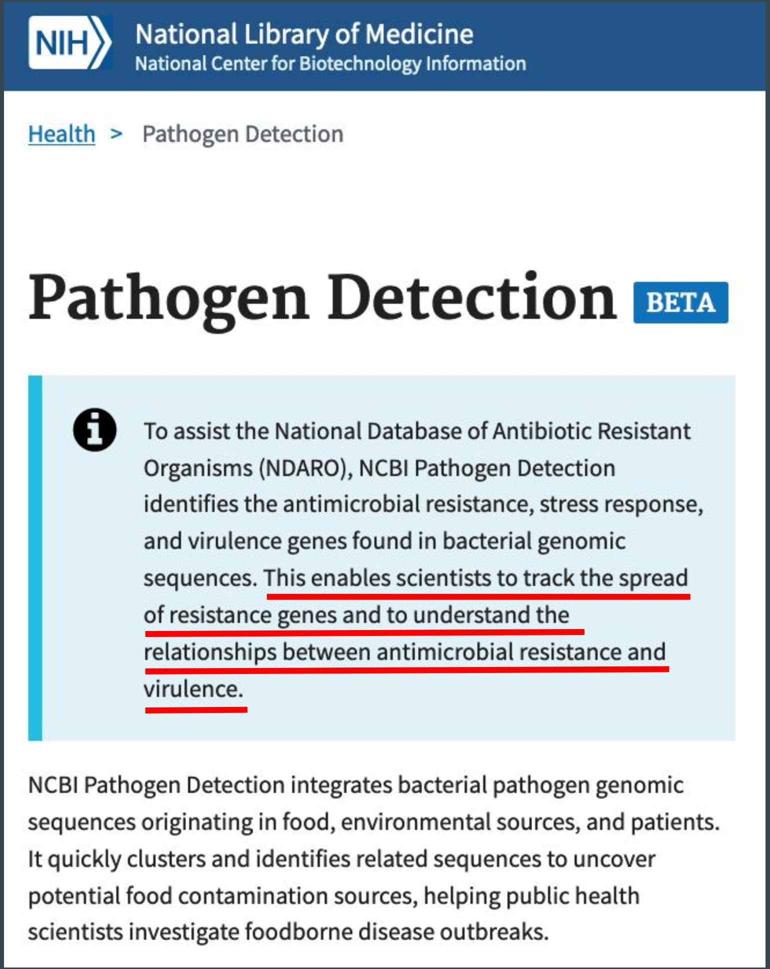
Research Questions

- **Chapter 2: Mapping AMR Traits to a Whole Genome Sequence* Phylogeny**
  - *Research Question:* How well does defective mismatch repair (MMR) explain the development of MDR?
- **Chapter 3: Correlational Assessment of WGS for Prediction of MDR Phenotypes**
  - *Research Question:* How well does WGS data predict categories of MDR?
- **Chapter 4: Combined Phylogenetic and Machine Learning Techniques**
  - *Research Question:* Can pretreating input data before phylogenetic analysis yield more clear hypotheses for traits that are of uncertain origin like MDR?

* Whole genome sequences defined on slide 12. "Complete Genome".

Data Source

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


NIH
National Library of Medicine
National Center for Biotechnology Information
Health > Pathogen Detection

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.

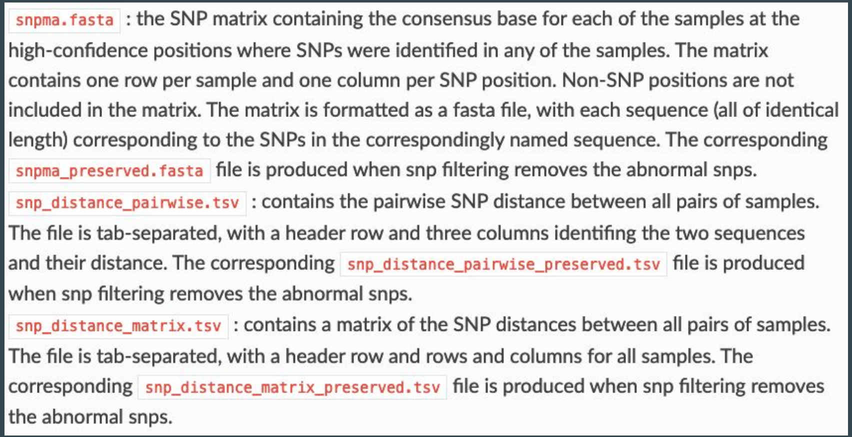
NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

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

Data Collection & Processing

- Use of the same *E. coli* data set as the *Cladistics* paper (Ford et al.).
- Filtered to 10,018 assembled isolates.
  - Only include "Complete Genome"
  - WGS downloaded from NCBI Genome
- Pass WGS as input to the CFSAN SNP Pipeline for processing.

Relevant Outputs of the CFSAN SNP Pipeline



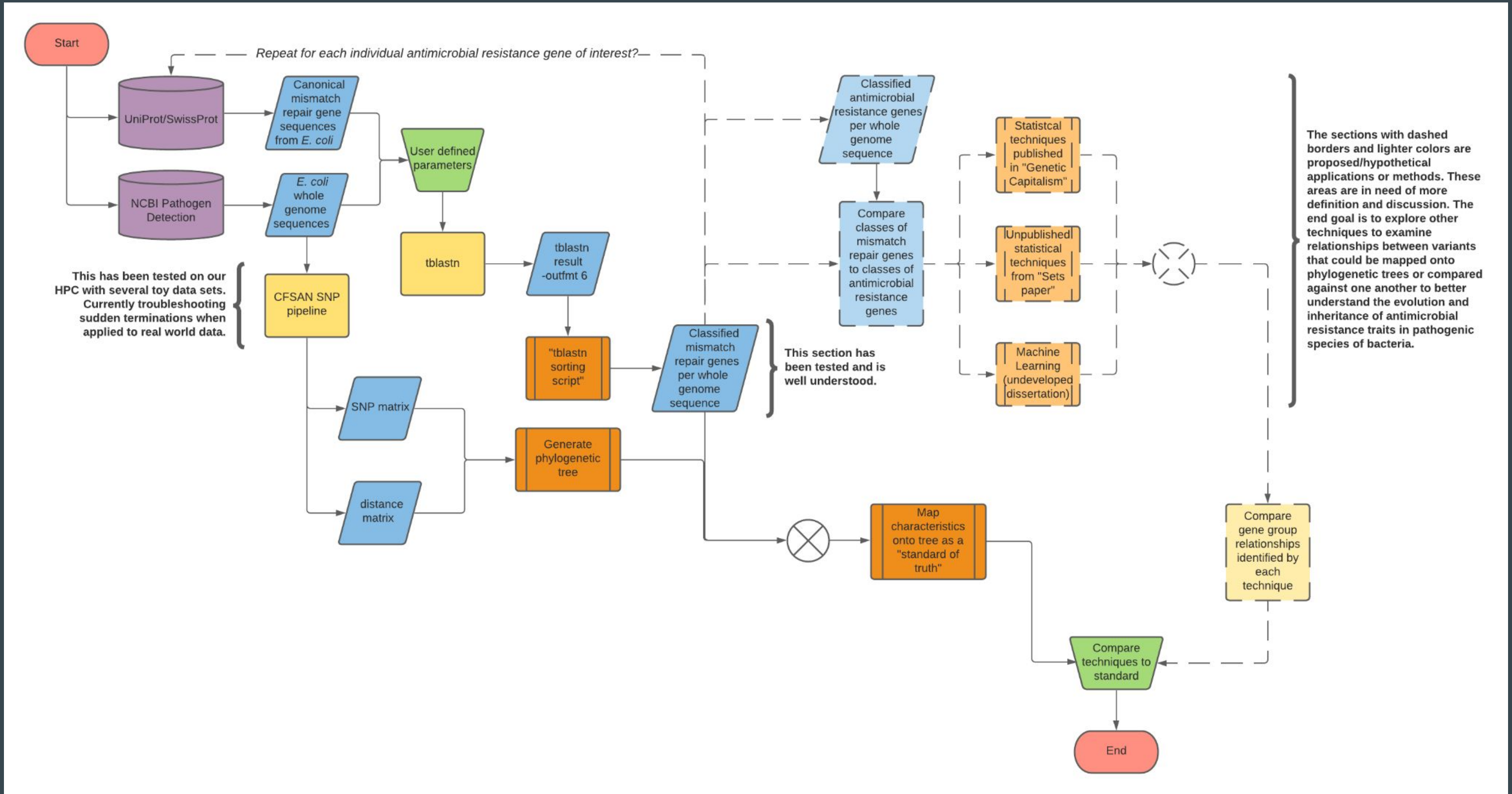
`snpma.fasta` : the SNP matrix containing the consensus base for each of the samples at the high-confidence positions where SNPs were identified in any of the samples. The matrix contains one row per sample and one column per SNP position. Non-SNP positions are not included in the matrix. The matrix is formatted as a fasta file, with each sequence (all of identical length) corresponding to the SNPs in the correspondingly named sequence. The corresponding `snpma_preserved.fasta` file is produced when snp filtering removes the abnormal snps.

`snp_distance_pairwise.tsv` : contains the pairwise SNP distance between all pairs of samples. The file is tab-separated, with a header row and three columns identifing the two sequences and their distance. The corresponding `snp_distance_pairwise_preserved.tsv` file is produced when snp filtering removes the abnormal snps.

`snp_distance_matrix.tsv` : contains a matrix of the SNP distances between all pairs of samples. The file is tab-separated, with a header row and rows and columns for all samples. The corresponding `snp_distance_matrix_preserved.tsv` file is produced when snp filtering removes the abnormal snps.

Screenshot from CFSAN SNP Pipeline Usage Outputs. Accessed Dec. 14, 2021 at https://snp-pipeline.readthedocs.io/en/latest/usage.html. More information at https://snp-pipeline.readthedocs.io/en/latest/dataflow.html.

General Workflow Process



Start

Repeat for each individual antimicrobial resistance gene of interest?

UniProt/SwissProt

NCBI Pathogen Detection

Canonical mismatch repair gene sequences from *E. coli*

E. coli whole genome sequences

User defined parameters

tblastn

tblastn result -outfmt 6

"tblastn sorting script"

CFSAN SNP pipeline

SNP matrix

distance matrix

Generate phylogenetic tree

Classified mismatch repair genes per whole genome sequence

Classified antimicrobial resistance genes per whole genome sequence

Compare classes of mismatch repair genes to classes of antimicrobial resistance genes

Statistcal techniques published in "Genetic Capitalism"

Unpublished statistical techniques from "Sets paper"

Machine Learning (undeveloped dissertation)

Compare gene group relationships identified by each technique

Map characteristics onto tree as a "standard of truth"

Compare techniques to standard

End

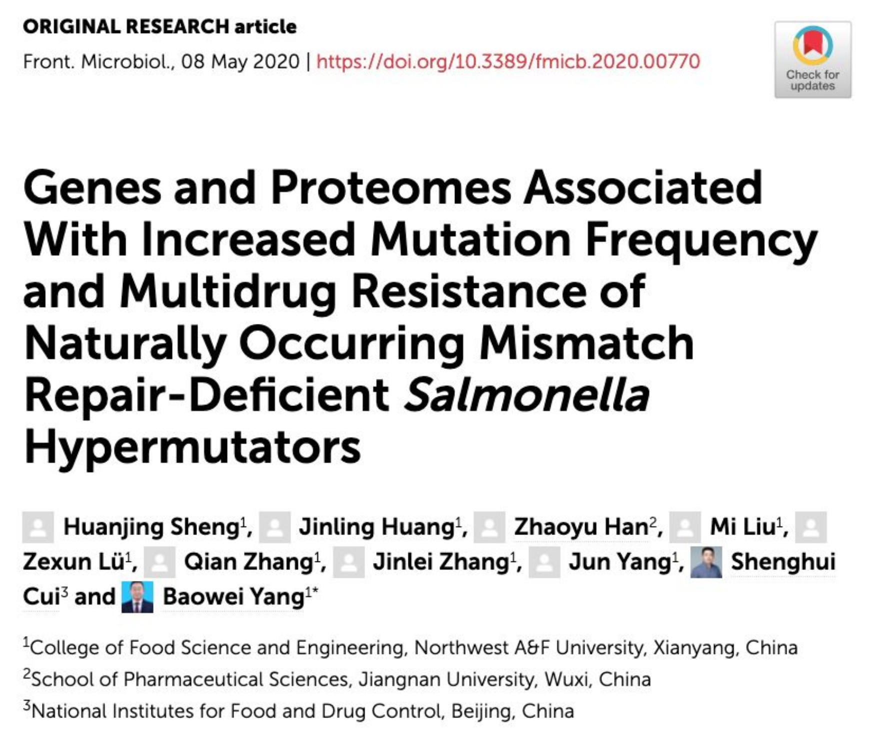
This has been tested on our HPC with several toy data sets. Currently troubleshooting sudden terminations when applied to real world data.

This section has been tested and is well understood.

The sections with dashed borders and lighter colors are proposed/hypothetical applications or methods. These areas are in need of more definition and discussion. The end goal is to explore other techniques to examine relationships between variants that could be mapped onto phylogenetic trees or compared against one another to better understand the evolution and inheritance of antimicrobial resistance traits in pathogenic species of bacteria.

Chapter 2 - Mapping Traits to WGS Phylogeny

- 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 system in bacteria is Mutator S (MutS).



ORIGINAL RESEARCH article

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

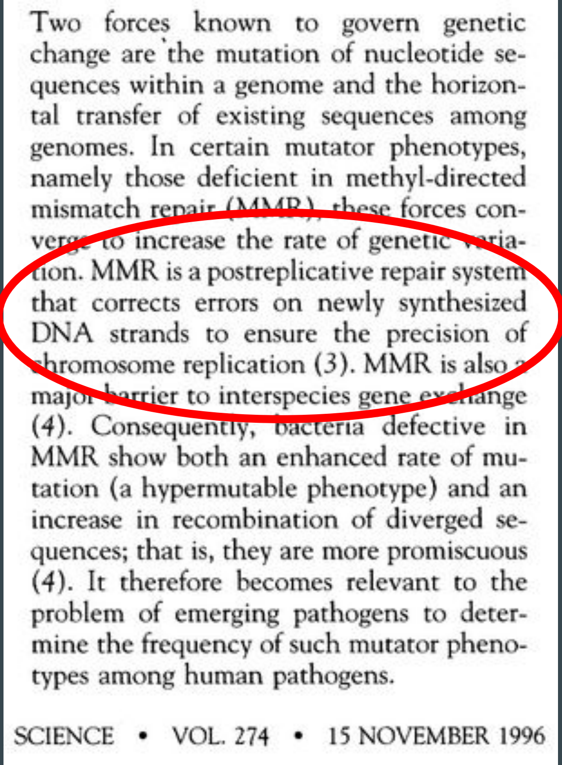
Genes and Proteomes Associated With Increased Mutation Frequency and Multidrug Resistance of Naturally Occurring Mismatch Repair-Deficient *Salmonella* Hypermutators

Huanjing Sheng[1], Jinling Huang[1], Zhaoyu Han[2], Mi Liu[1], Zexun Lü[1], Qian Zhang[1], Jinlei Zhang[1], Jun Yang[1], Shenghui Cui[3] and Baowei Yang[1*]

[1]College of Food Science and Engineering, Northwest A&F University, Xianyang, China
[2]School of Pharmaceutical Sciences, Jiangnan University, Wuxi, China
[3]National Institutes for Food and Drug Control, Beijing, China

Image from Sheng et al., 2020. Accessed Dec. 14, 2021 at https://www.frontiersin.org/articles/10.3389/fmicb.2020.00770/full.

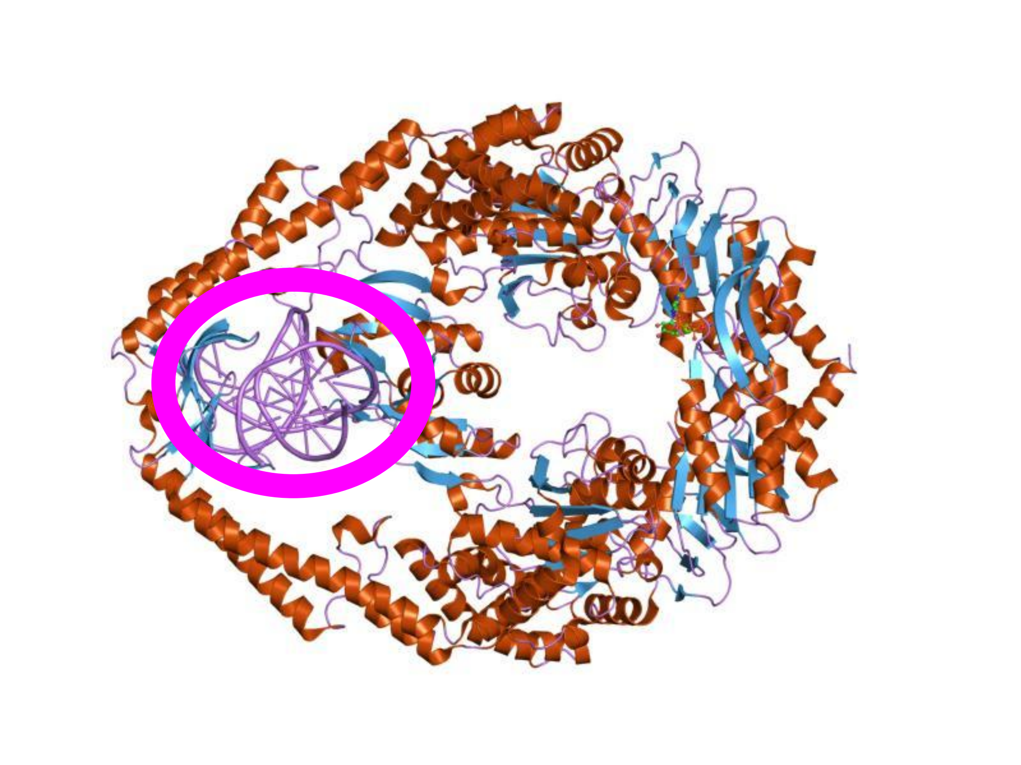
Chapter 2 - Mapping Traits to WGS Phylogeny - cont.



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.

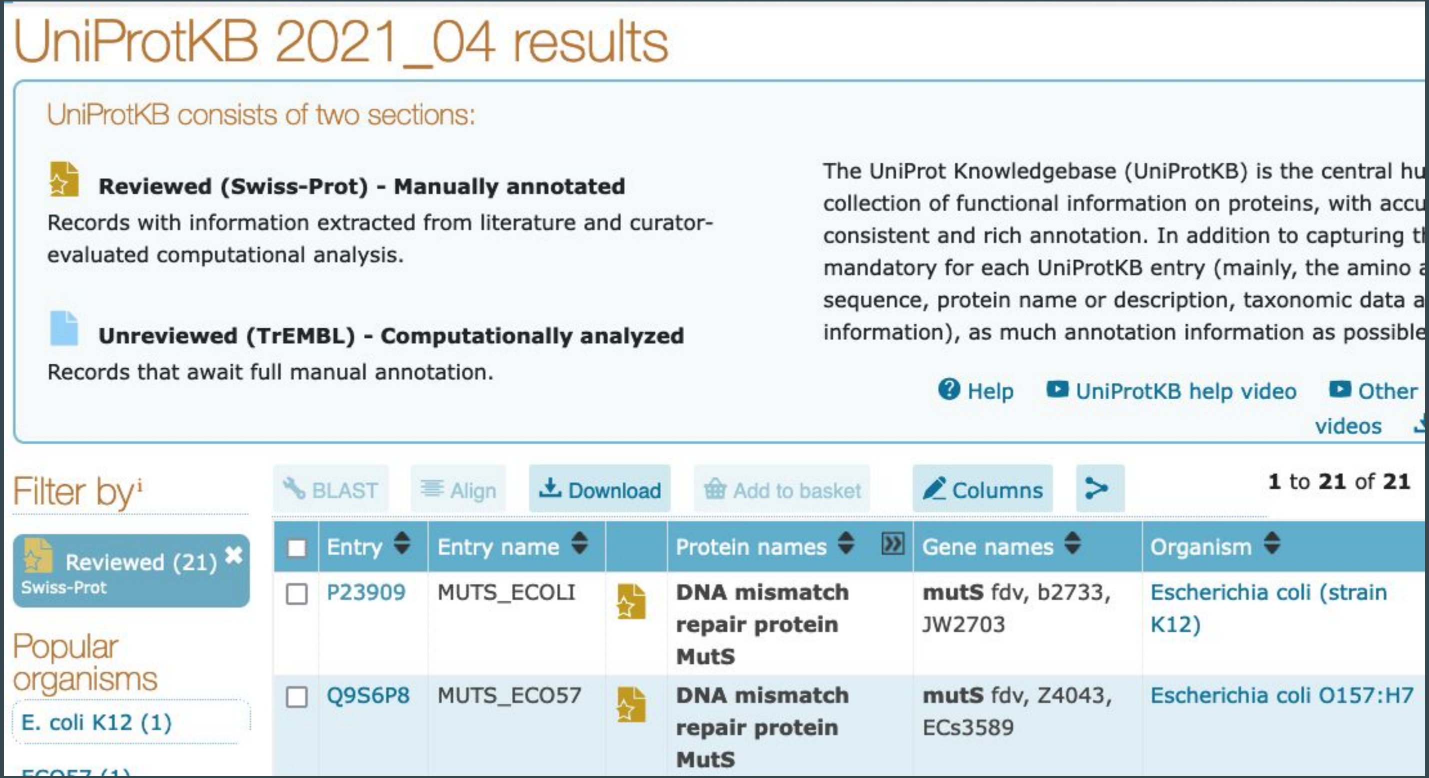
SCIENCE • VOL. 274 • 15 NOVEMBER 1996

Image from LeClerc et al., 1996.



Older image of the top view of MutS protein (PDB 1oh6) as hosted at EBI (https://www.ebi.ac.uk/pdbe/entry/pdb/1oh6). Accessed Dec. 13, 2021.

Chapter 2 - Characterizing MMR



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

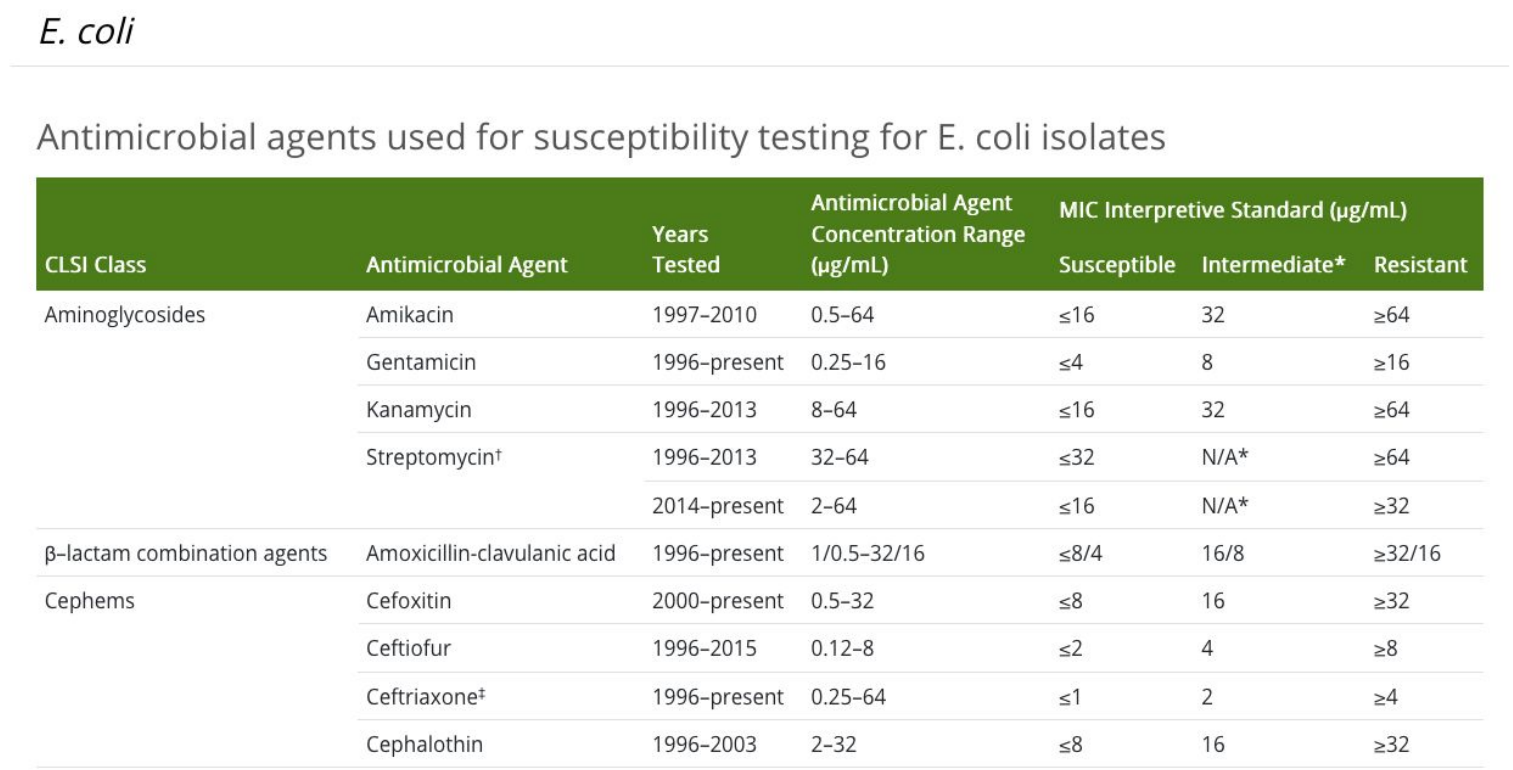
Records that await full manual annotation.

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

☐	Entry	Entry name		Protein names	Gene names	Organism
☐	P23909	MUTS_ECOLI	★	**DNA mismatch repair protein MutS**	**mutS** fdv, b2733, JW2703	Escherichia coli (strain K12)
☐	Q9S6P8	MUTS_ECO57	★	**DNA mismatch repair protein MutS**	**mutS** fdv, Z4043, ECs3589	Escherichia coli O157:H7

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.

Chapter 2 - Characterizing MDR



E. coli

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

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.

Chapter 2 - Characterizing MDR - cont.

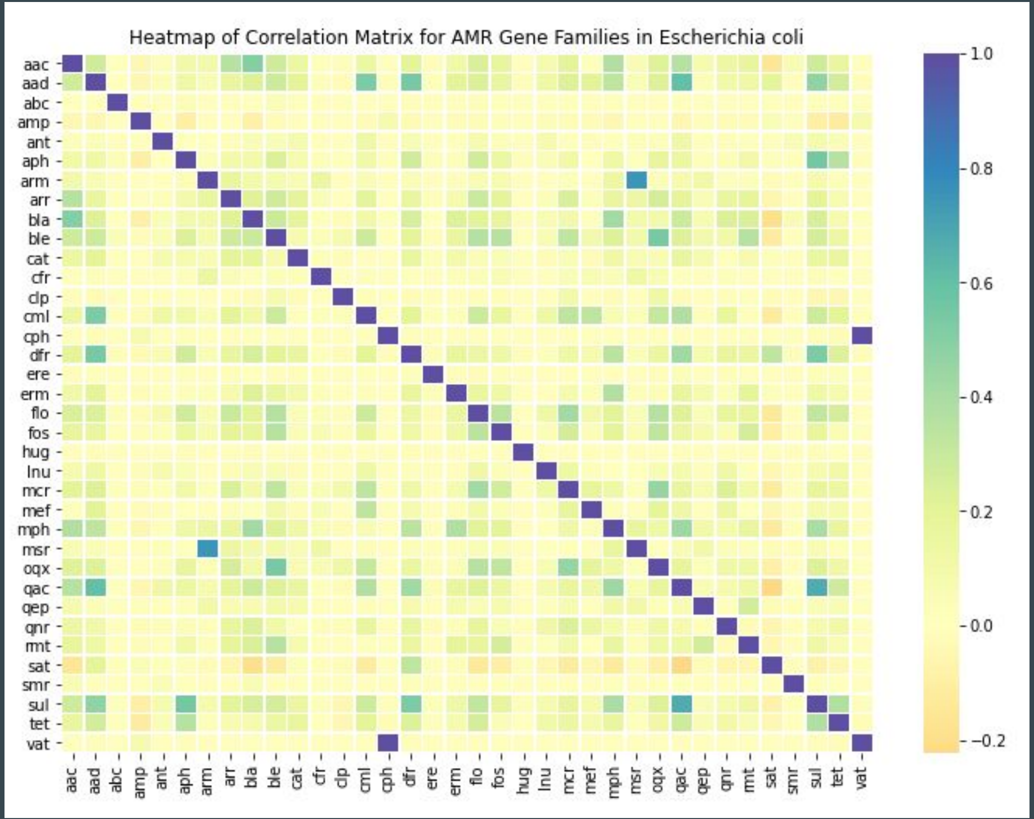
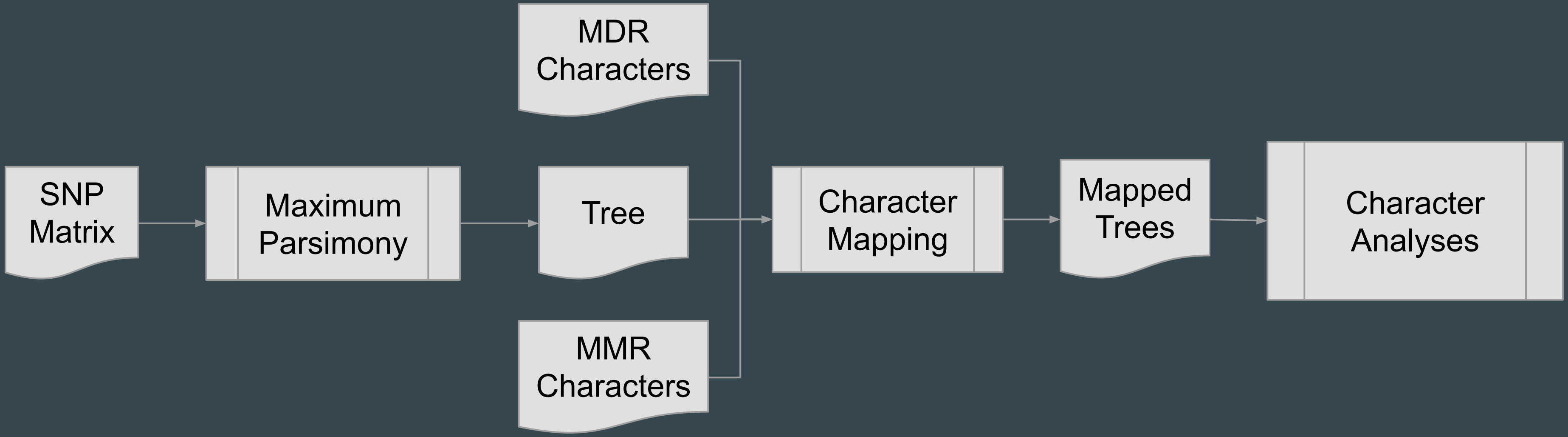


Image from work-in-progress analysis. Correlation matrix for presence/absence of 36 AMR gene families in the 10,018 *E. coli* data set.

Chapter 2 - Hypothesis & Workflow

Certain Mutator S variants and/or their lineages are correlated with specific, grouped classes of antimicrobial resistance genes.



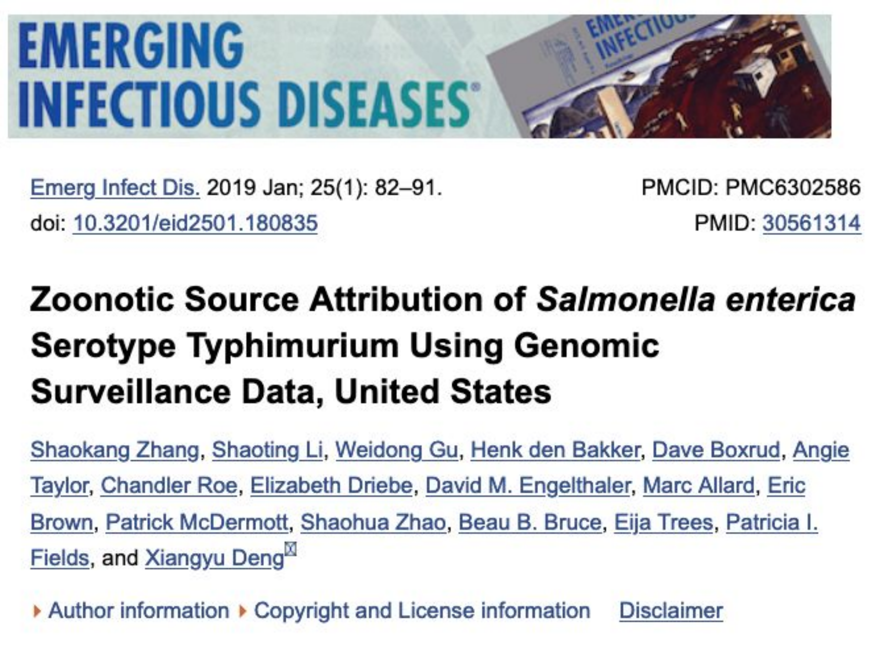
Chapter 2 - Summary

- **Input**
  - WGS SNP matrix
- **Output**
  - Characters Mapped to Phylogenetic Tree
- **Analysis**
  - Evaluate MDR/MMR Characters
    - *Character Counts*
      - TNT function calls
    - *Character Correlations*
      - Phylogenetic least squares, concentrated changes, etc.
- **Resource**
  - Ford et al., 2020

- **Potential Results**
  - *Character Counts*
    - Prevalence and diversity of some classes of antibiotics should increase over time.
    - Some MutS variants could exhibit larger counts of MDR.
  - *Character Correlations*
    - Some MDR characters could correlate more strongly with each other than with others.
    - Implies that AMR acquisition (the path to MDR) is not independent.

Chapter 3 - Correlational Assessment of WGS for Prediction of MDR Phenotypes

- Similar data has been demonstrated to predict host species for *Salmonella* isolates.
- The US FDA has called for more research into improving predictive models based on WGS data.
- Multidrug-resistance phenotypes can be thought of as similar to host specificity, due to required underlying genes for host colonization, pathogenicity, and virulence.



Emerg Infect Dis. 2019 Jan; 25(1): 82–91. PMCID: PMC6302586

doi: 10.3201/eid2501.180835 PMID: 30561314

Zoonotic Source Attribution of *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United States

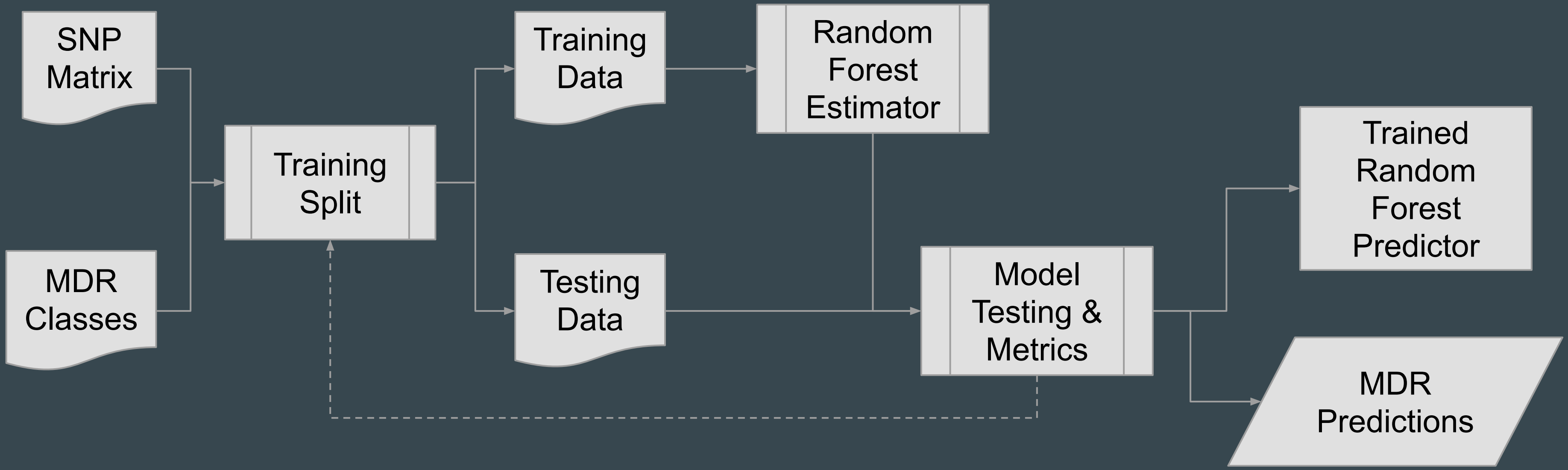
Shaokang Zhang, Shaoting Li, Weidong Gu, Henk den Bakker, Dave Boxrud, Angie Taylor, Chandler Roe, Elizabeth Driebe, David M. Engelthaler, Marc Allard, Eric Brown, Patrick McDermott, Shaohua Zhao, Beau B. Bruce, Eija Trees, Patricia I. Fields, and Xiangyu Deng☒

▸ Author information ▸ Copyright and License information Disclaimer

Image from Zhang et al., 2019. Accessed Dec. 14, 2021 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302586/.

Chapter 3 - Hypothesis & Workflow

Certain patterns of WGS SNPs can both define and predict specific types of multidrug-resistance using supervised machine learning.



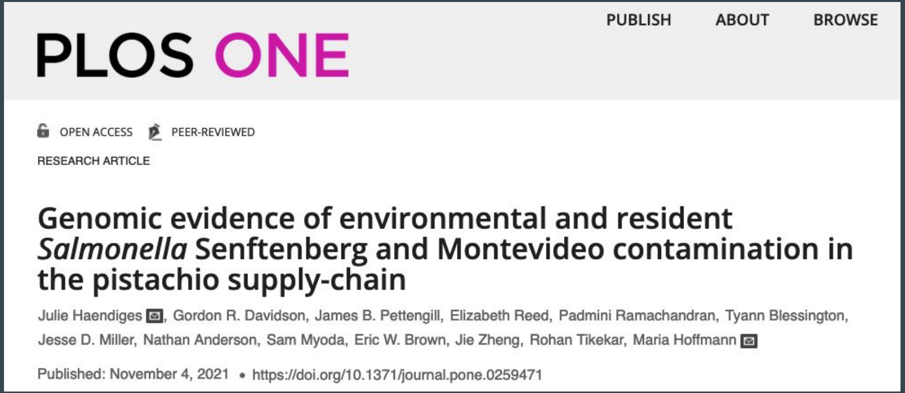
Chapter 3- Summary

- **Input**
  - WGS SNP matrix
- **Output**
  - Trained Random Forest
  - Predictions for MDR Classes
  - Assessment Metrics for Trained Model
- **Analysis**
  - Perform Supervised Machine Learning
    - Multiclass Classification
      - *Random Forest Estimator*
- **Resources**
  - Deng et al., 2021.
  - Zhang et al., 2019.

- **Potential Results**
  - *Random Forest Estimator*
    - Certain patterns of WGS SNP information should predict types of MDR.
    - Could reveal novel virulence genes or potential compensatory mutations necessary for acquiring a type of MDR.

Chapter 4 - Combined Phylogenetic and Machine Learning Techniques for Predicting MDR Associations

- Large sample sizes and high genetic variation represent major challenges for big data genomics at scale.
- Reticulate evolution and horizontal gene transfer further complicate analyses, especially for highly similar samples.
- Techniques agnostic of evolutionary models could therefore be useful.



RESEARCH ARTICLE

Genomic evidence of environmental and resident *Salmonella* Senftenberg and Montevideo contamination in the pistachio supply-chain

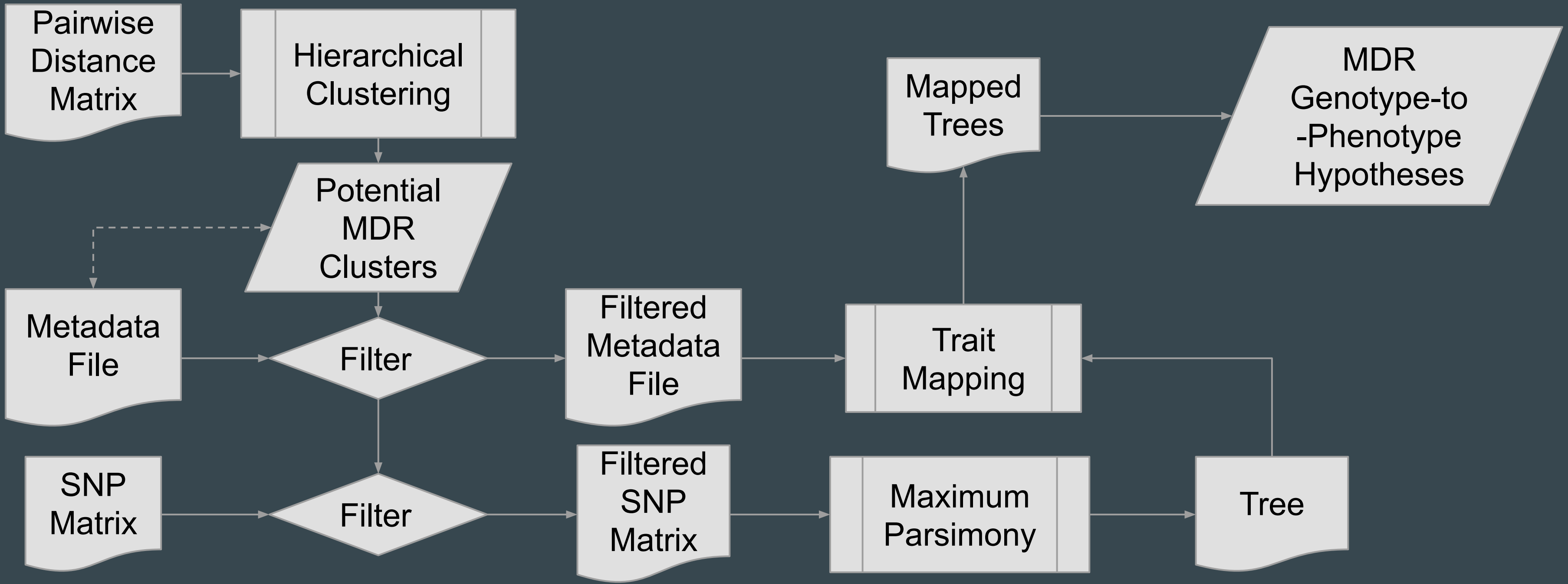
Julie Haendiges, Gordon R. Davidson, James B. Pettengill, Elizabeth Reed, Padmini Ramachandran, Tyann Blessington, Jesse D. Miller, Nathan Anderson, Sam Myoda, Eric W. Brown, Jie Zheng, Rohan Tikekar, Maria Hoffmann

Published: November 4, 2021 • https://doi.org/10.1371/journal.pone.0259471

Image from Haendiges et al., 2021. Accessed Dec. 14, 2021 at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0259471#sec002.

Chapter 4 - Hypothesis & Workflow

Functional clusters (via machine learning) can be used to clarify phylogenetic hypotheses while remaining agnostic of evolutionary models.



Chapter 4 - Summary

- **Input**
  - Pairwise Distance Matrix
  - WGS SNP Matrix
  - Metadata File
- **Output**
  - Clusters
  - Mapped Phylogenetic Trees
- **Analysis**
  - Unsupervised Machine Learning
    - *Hierarchical Clustering*
  - Evaluate Characters & Traits
    - *Mapped Phylogenetic Trees*
- **Resources**
  - Deng et al., 2021.
  - Haendiges et al., 2021.

- **Potential Results**
  - *Hierarchical clustering*
    - Clustering should reveal different groups of MDR.
    - These MDR groups will likely have different evolutionary histories, as evidenced through geographical or host-related trajectories.
  - *Mapped Phylogenetic Trees*
    - Streamlined trees should have more clear genotype-to-phenotype hypotheses visible, as demonstrated by other researchers.

Potential Issues & Mitigation

- <u>Computing Risks</u>
  - *Examples:*
    - HPC Outages, Core Availability
  - *Mitigation*:
    - Contact with UNC Charlotte Office of ONE IT for HPC
- <u>Data Risks</u>
  - *Examples:*
    - Data Set Size, Sample Selection
  - *Mitigation*:
    - Contact NCBI-PD Team
    - Follow Published Research Methods
- <u>Method Risks</u>
  - *Examples:*
    - Novel Applications
  - *Mitigation*:
    - Check Online Documentation
    - Follow Published Research Methods
- <u>Model Risks</u>
  - *Examples:*
    - Model Assumptions, Model Biases
  - *Mitigation*:
    - Check Online Documentation
    - Follow Published Research Methods

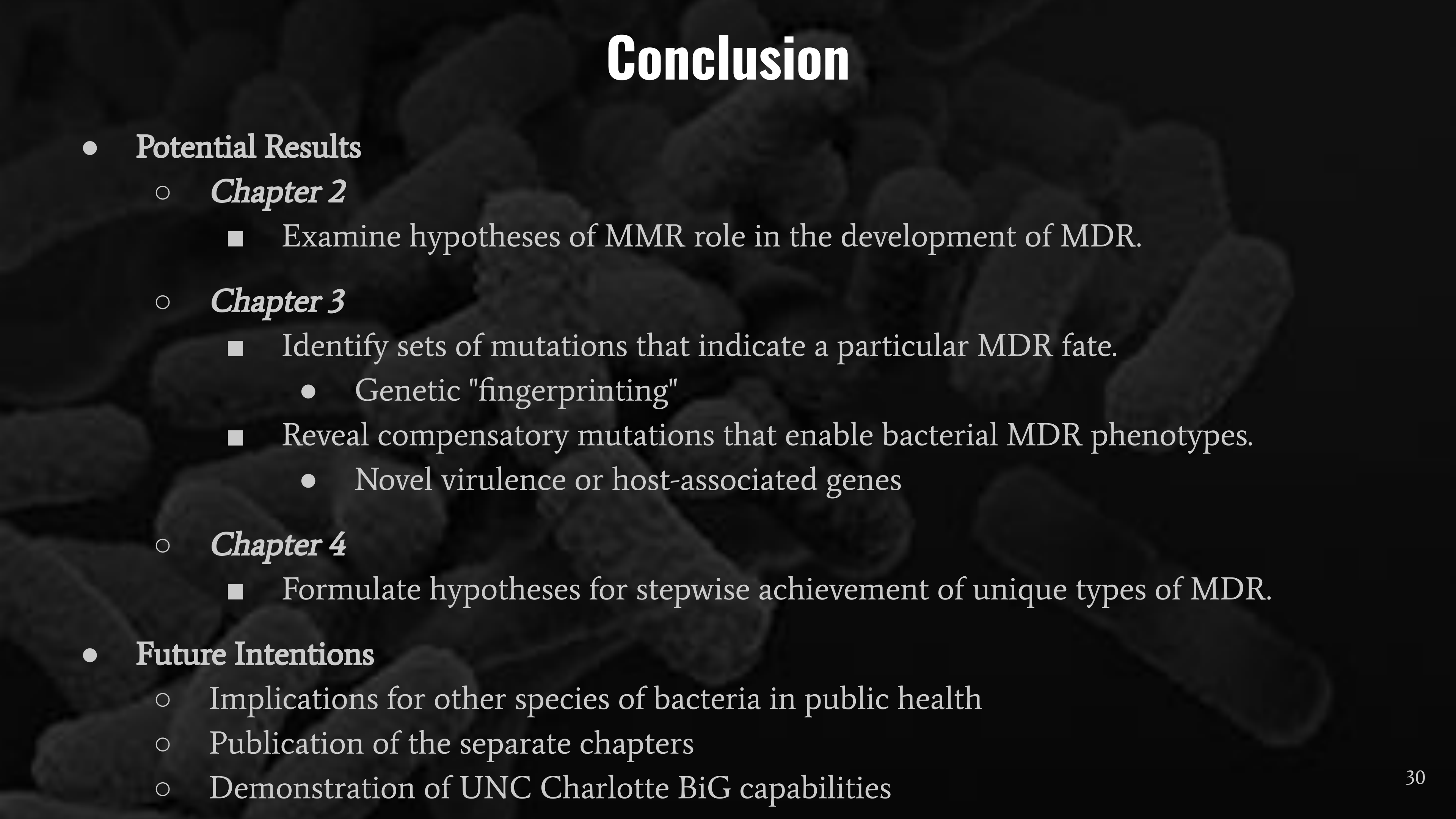
Timeframe											
Research Activities	**Prev.**	**Dec 2021**	**Jan 2022**	**Feb 2022**	**Mar 2022**	**Apr 2022**	**May 2022**	**Jun 2022**	**Jul 2022**	**Aug 2022**	**Sep 2022**
Gather all Data, Pipelines, and Control Code	■	■	■								
Troubleshoot Code & Pipeline Issues	■	■	■								
Run CFSAN Pipeline on WGS		■	■								
Investigation 1: Make Tree & Map Characters (MDR/MMR)			■	■							
Draft Investigation 1 Results & Send to Advisor				■	■	■					
Investigation 2: Predict MDR from WGS SNP Matrix			■	■							
Draft Investigation 2 Results & Send to Advisor				■	■	■					
Investigation 3: Combine Phylogenetics & Machine Learning					■	■	■				
Draft Investigation 3 Results & Send to Advisor						■	■	■			
Present Preliminary Results to Committee								■			
Check Analyses							■	■	■		
Adjust or Repeat Analyses Based on Feedback							■	■	■		
Combine & Polish Document									■	■	
Create Slide Deck										■	■
Final Defense											■

Research Summary

- **Chapter 2: Mapping AMR Traits to a Whole Genome Sequence Phylogeny**
    - *Research Question:* How well does defective mismatch repair (MMR) explain the development of MDR?
    - *Methods*: Phylogenetics and cladistics

- **Chapter 3: Correlational Assessment of WGS for Prediction of MDR Phenotypes**
    - *Research Question:* How well does WGS data predict categories of MDR?
    - *Methods:* Supervised machine learning

- **Chapter 4: Combined Phylogenetic and Machine Learning Techniques**
    - *Research Question:* Can pretreating input data before phylogenetic analysis yield more clear hypotheses for traits that are of uncertain origin like MDR?
    - *Methods:* Unsupervised machine learning and phylogenetics



Conclusion

- **Potential Results**
  - ***Chapter 2***
    - Examine hypotheses of MMR role in the development of MDR.
  - ***Chapter 3***
    - Identify sets of mutations that indicate a particular MDR fate.
      - Genetic "fingerprinting"
    - Reveal compensatory mutations that enable bacterial MDR phenotypes.
      - Novel virulence or host-associated genes
  - ***Chapter 4***
    - Formulate hypotheses for stepwise achievement of unique types of MDR.
- **Future Intentions**
  - Implications for other species of bacteria in public health
  - Publication of the separate chapters
  - Demonstration of UNC Charlotte BiG capabilities

Significance

- **Intellectual Merit**
  - Refine current understandings of multidrug resistance in enteric bacteria.
  - Demonstrate newly available tools and data sets:
    - CFSAN SNP Pipeline, NCBI Pathogen Detection

- **Broader Impacts**
  - Improvements for current disease surveillance.
  - Insights for anticipating the direction of the silent pandemic.
  - Prolong the efficacy of current antibiotics.

Questions

Thank you for your kind attention.

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. 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

5. 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

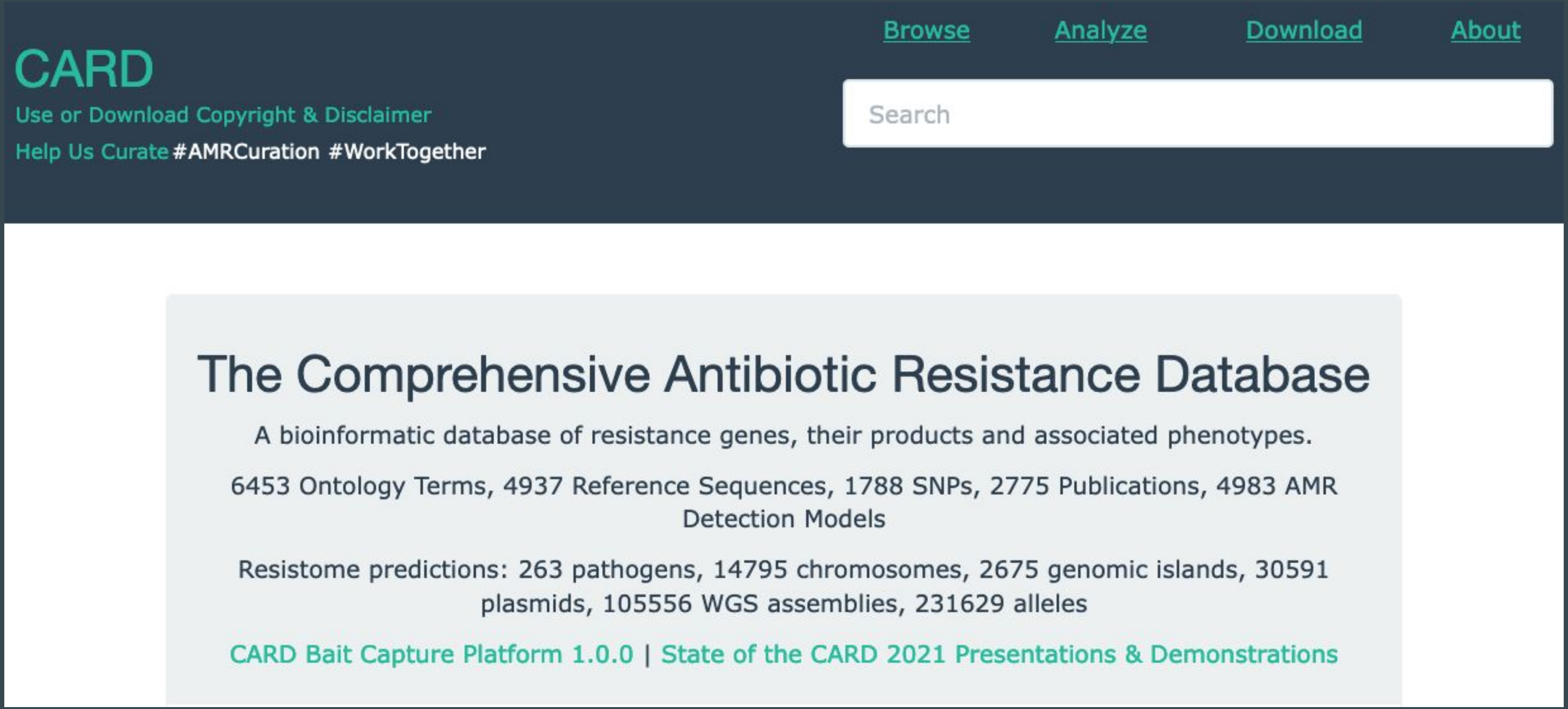
6. Shapiro-Ilan, D. I., Fuxa, J. R., Lacey, L. A., Onstad, D. W., & Kaya, H. K. (2005). Definitions of pathogenicity and virulence in invertebrate pathology. Journal of invertebrate pathology, 88(1), 1–7. https://doi.org/10.1016/j.jip.2004.10.003

7. 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

8. 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

Potential Q&A Materials

Comprehensive Antibiotic Resistance Database - CARD



Browse
Analyze
Download
About
CARD
Use or Download Copyright & Disclaimer
Help Us Curate #AMRCuration #WorkTogether
Search

The Comprehensive Antibiotic Resistance Database

A bioinformatic database of resistance genes, their products and associated phenotypes.

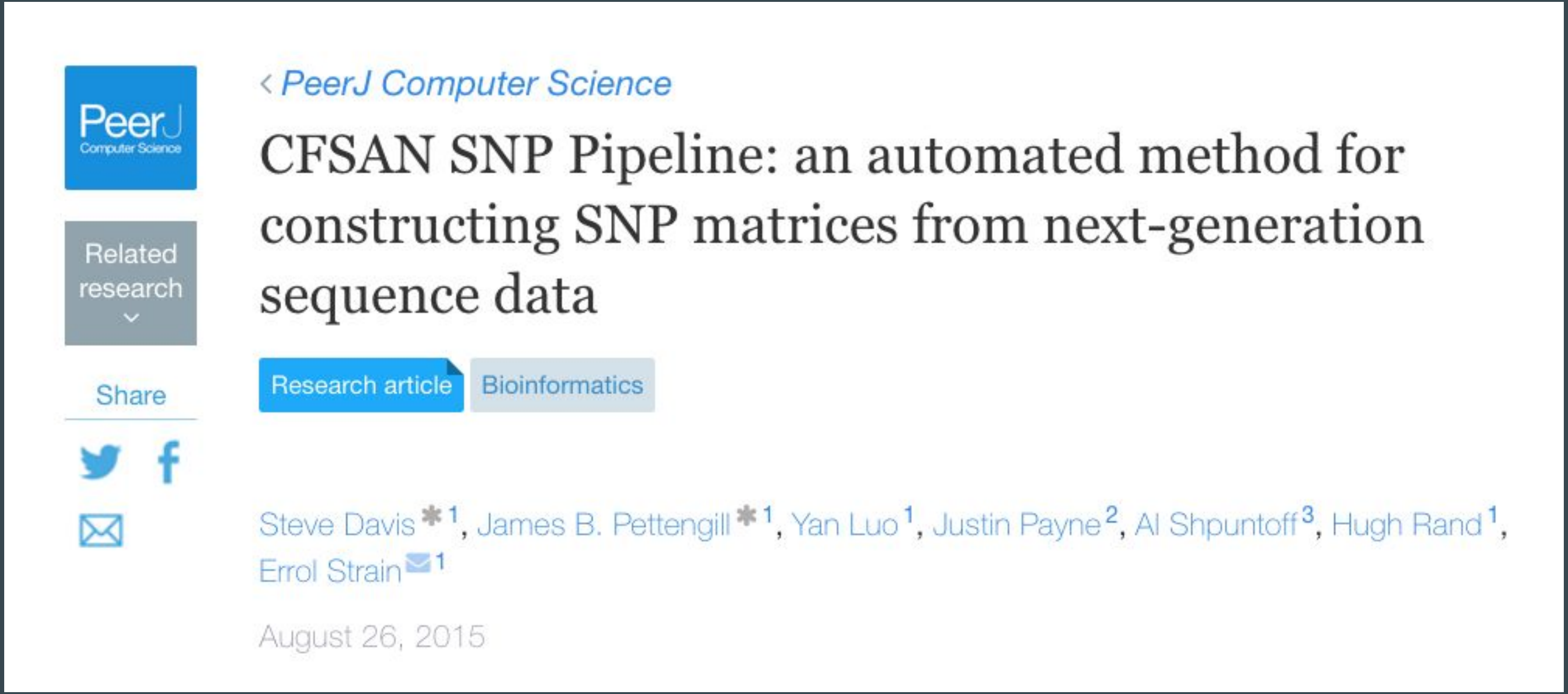
6453 Ontology Terms, 4937 Reference Sequences, 1788 SNPs, 2775 Publications, 4983 AMR Detection Models

Resistome predictions: 263 pathogens, 14795 chromosomes, 2675 genomic islands, 30591 plasmids, 105556 WGS assemblies, 231629 alleles

CARD Bait Capture Platform 1.0.0 | State of the CARD 2021 Presentations & Demonstrations

Image from CARD website. Accessed Dec. 13, 2021 at https://card.mcmaster.ca/.

CFSAN SNP Pipeline



< *PeerJ Computer Science*

CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data

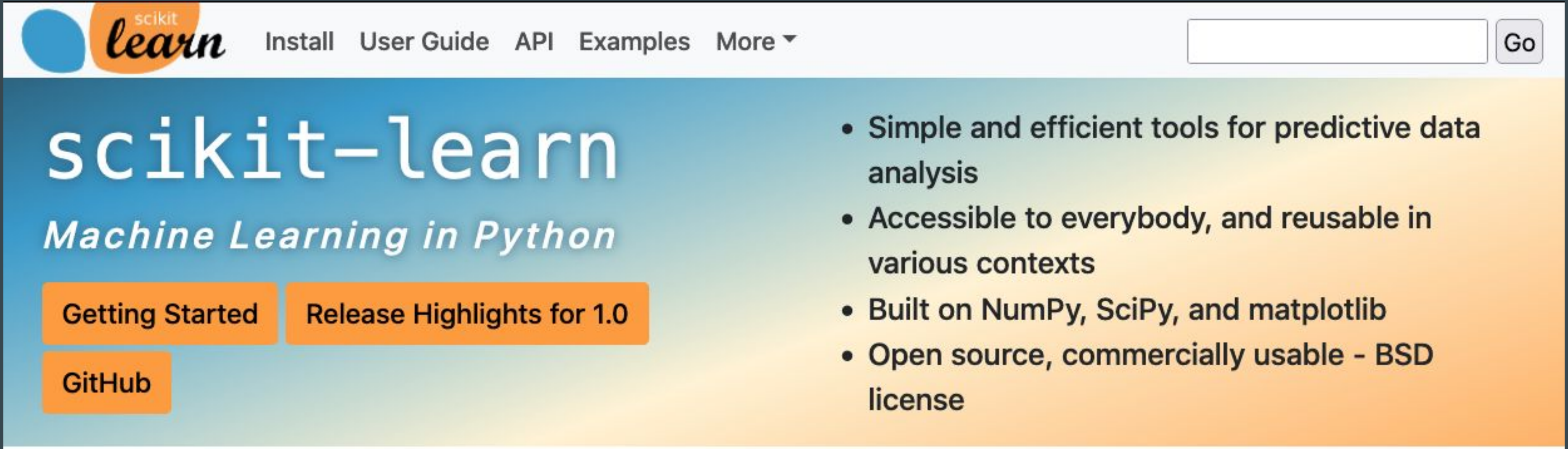
Research article Bioinformatics

Steve Davis *[1], James B. Pettengill *[1], Yan Luo [1], Justin Payne [2], Al Shpuntoff [3], Hugh Rand [1], Errol Strain ✉[1]

August 26, 2015

Image from Davis, et al. (https://doi.org/10.7717/peerj-cs.20). Accessed Dec. 13, 2021 at https://peerj.com/articles/cs-20/. Further documentation at https://snp-pipeline.readthedocs.io/en/latest/.

scikit-learn



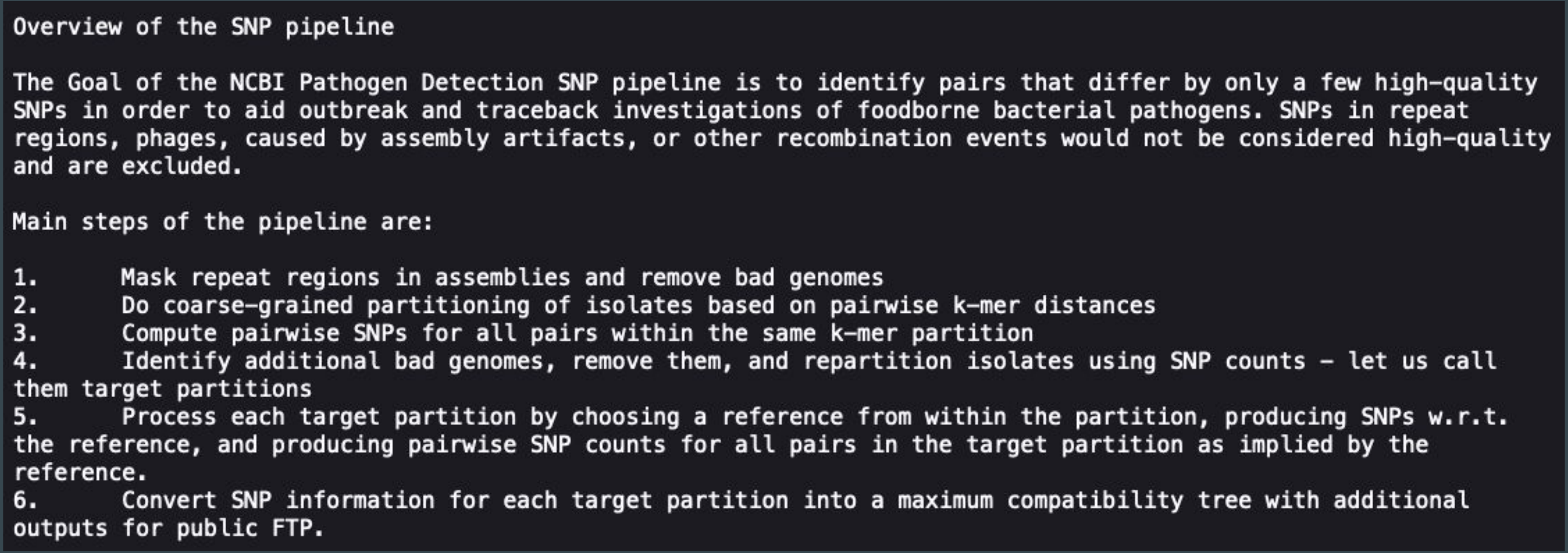
scikit-learn

Machine Learning in Python

- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Image from scikit-learn website. Accessed Dec. 13, 2021 at https://scikit-learn.org/stable/.

NCBI Pathogen Detection - Unpublished Methods



Overview of the SNP pipeline

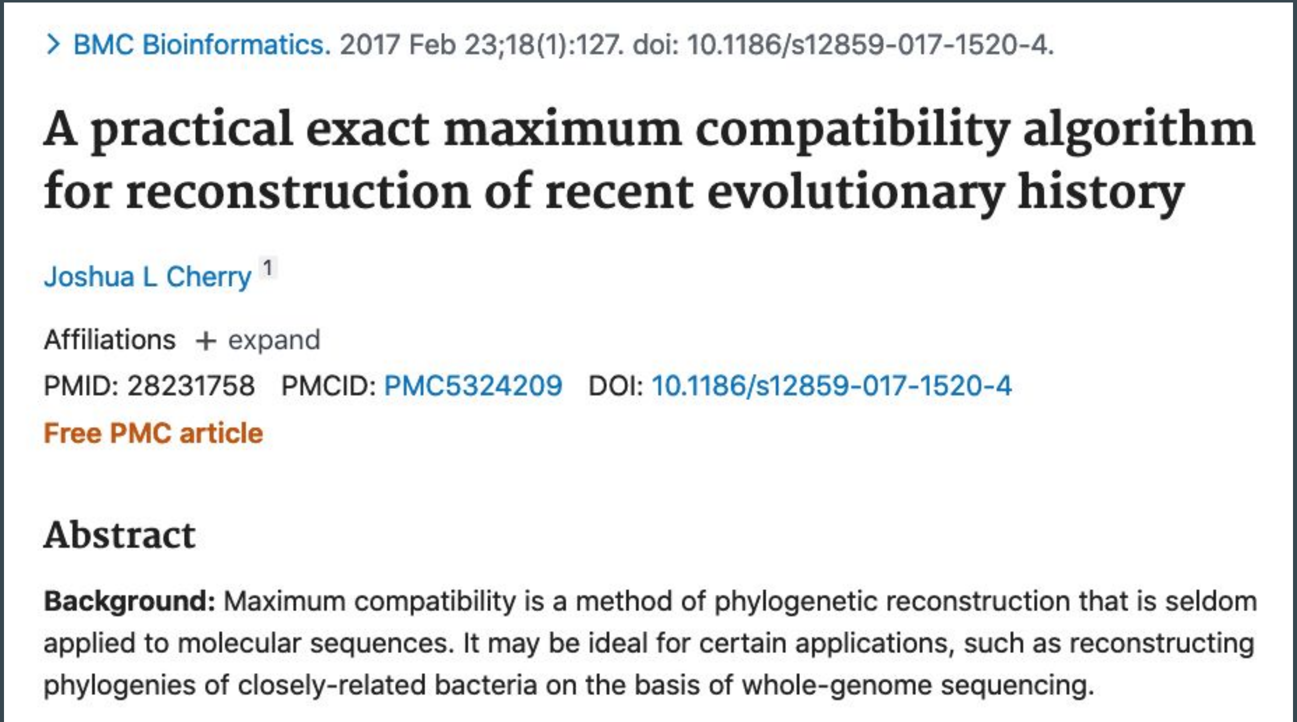
The Goal of the NCBI Pathogen Detection SNP pipeline is to identify pairs that differ by only a few high-quality SNPs in order to aid outbreak and traceback investigations of foodborne bacterial pathogens. SNPs in repeat regions, phages, caused by assembly artifacts, or other recombination events would not be considered high-quality and are excluded.

Main steps of the pipeline are:

1. Mask repeat regions in assemblies and remove bad genomes
2. Do coarse-grained partitioning of isolates based on pairwise k-mer distances
3. Compute pairwise SNPs for all pairs within the same k-mer partition
4. Identify additional bad genomes, remove them, and repartition isolates using SNP counts – let us call them target partitions
5. Process each target partition by choosing a reference from within the partition, producing SNPs w.r.t. the reference, and producing pairwise SNP counts for all pairs in the target partition as implied by the reference.
6. Convert SNP information for each target partition into a maximum compatibility tree with additional outputs for public FTP.

Image from NCBI FTP site. Accessed Dec. 14, 2021 at https://ftp.ncbi.nlm.nih.gov/pathogen/Methods.txt.

Maximum Compatibility



> BMC Bioinformatics. 2017 Feb 23;18(1):127. doi: 10.1186/s12859-017-1520-4.

A practical exact maximum compatibility algorithm for reconstruction of recent evolutionary history

Joshua L Cherry [1]

Affiliations + expand

PMID: 28231758 PMCID: PMC5324209 DOI: 10.1186/s12859-017-1520-4

Free PMC article

Abstract

Background: Maximum compatibility is a method of phylogenetic reconstruction that is seldom applied to molecular sequences. It may be ideal for certain applications, such as reconstructing phylogenies of closely-related bacteria on the basis of whole-genome sequencing.

Image from Cherry, 2017. Accessed Dec. 14, 2021 at https://pubmed.ncbi.nlm.nih.gov/28231758/.

Useful Deng et al., 2021 - Figure 1

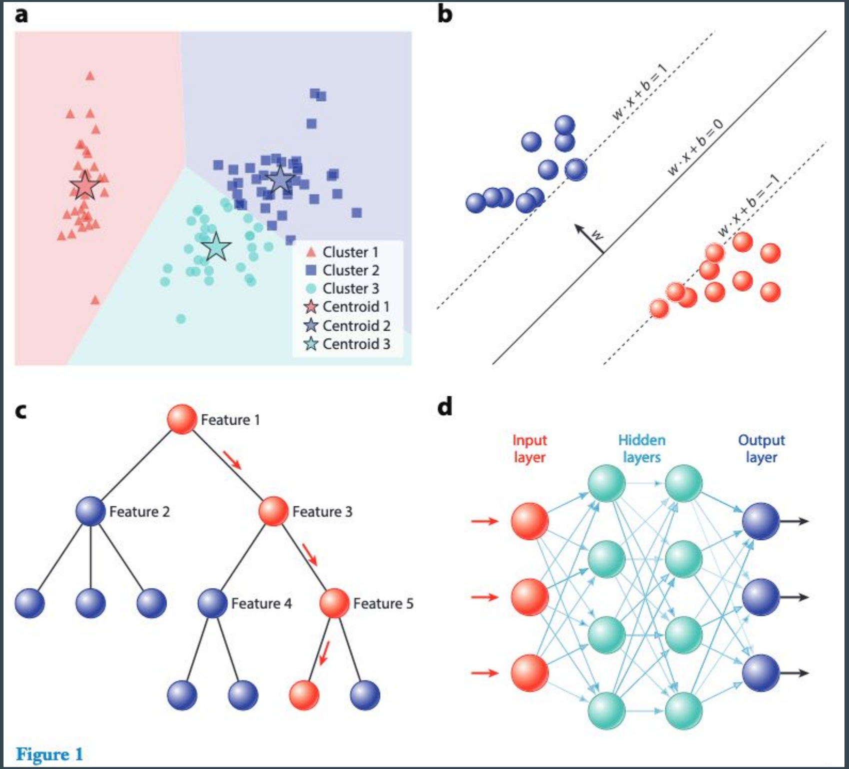
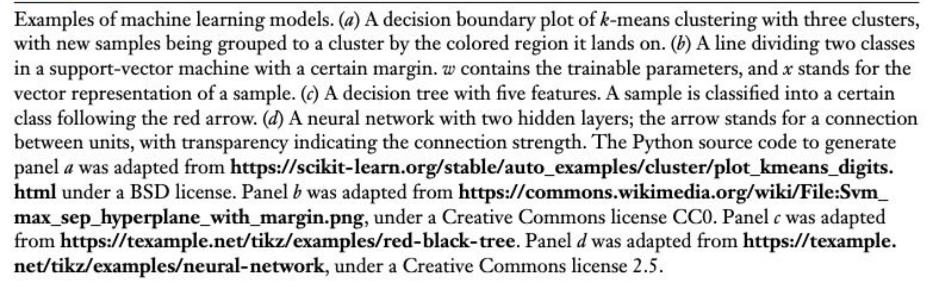


Figure 1



Examples of machine learning models. (*a*) A decision boundary plot of *k*-means clustering with three clusters, with new samples being grouped to a cluster by the colored region it lands on. (*b*) A line dividing two classes in a support-vector machine with a certain margin. *w* contains the trainable parameters, and *x* stands for the vector representation of a sample. (*c*) A decision tree with five features. A sample is classified into a certain class following the red arrow. (*d*) A neural network with two hidden layers; the arrow stands for a connection between units, with transparency indicating the connection strength. The Python source code to generate panel *a* was adapted from **https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_digits.html** under a BSD license. Panel *b* was adapted from **https://commons.wikimedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png**, under a Creative Commons license CC0. Panel *c* was adapted from **https://texample.net/tikz/examples/red-black-tree**. Panel *d* was adapted from **https://texample.net/tikz/examples/neural-network**, under a Creative Commons license 2.5.

Modified image from Deng et al., 2021. Accessed Dec. 14, 2021 at https://www.researchgate.net/profile/Abigail-Horn/publication/348655838_Emerging_Applications_of_Machine_Learning_in_Food_Safety/links/605d1277458515e8346fe9a0/Emerging-Applications-of-Machine-Learning-in-Food-Safety.pdf.

Useful Deng et al., 2021 - Table 2, excerpts.

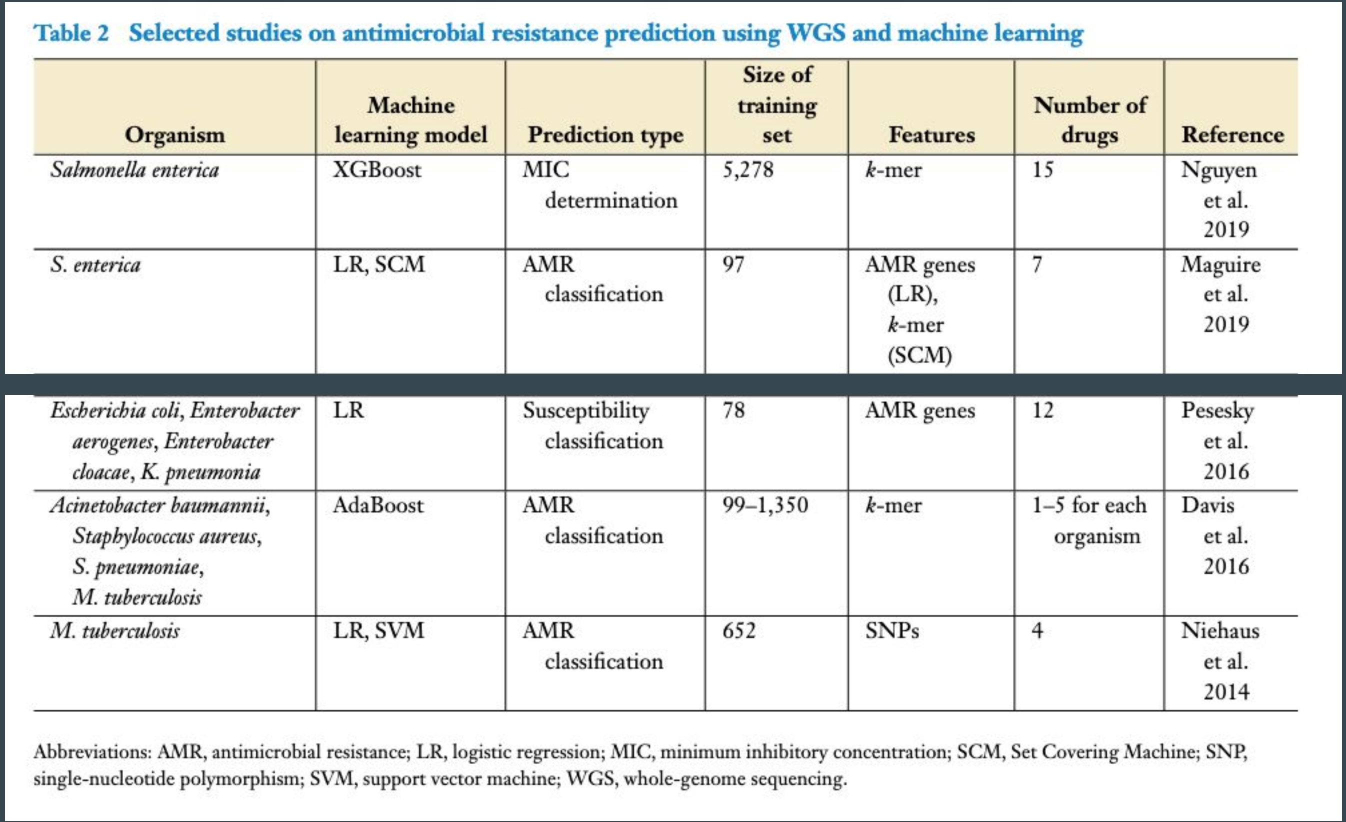


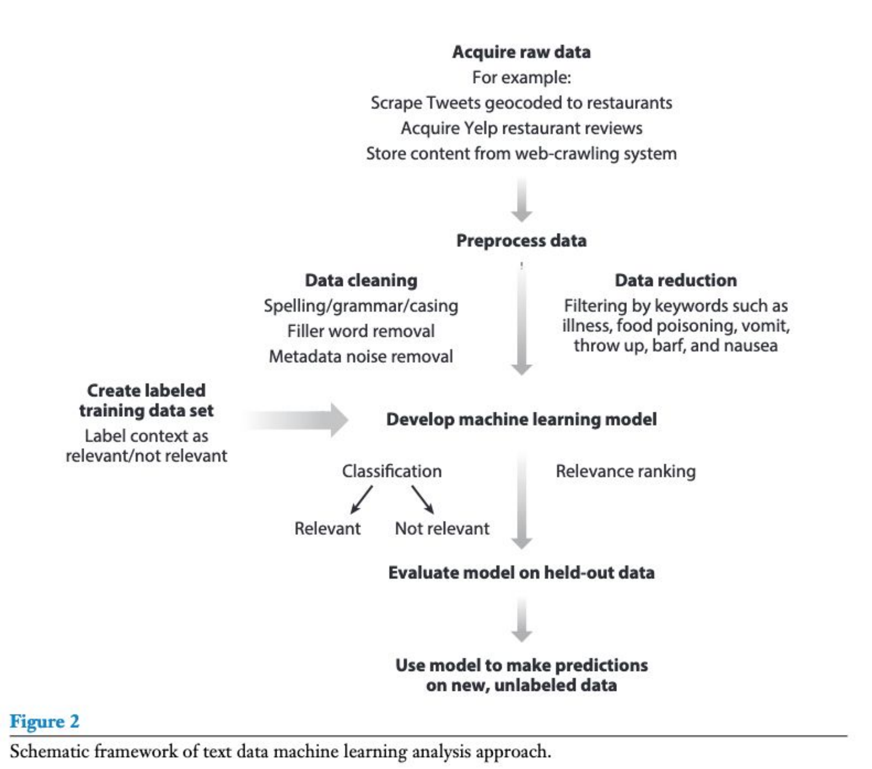
Table 2 Selected studies on antimicrobial resistance prediction using WGS and machine learning

Organism	Machine learning model	Prediction type	Size of training set	Features	Number of drugs	Reference
Salmonella enterica	XGBoost	MIC determination	5,278	*k*-mer	15	Nguyen et al. 2019
S. enterica	LR, SCM	AMR classification	97	AMR genes (LR), *k*-mer (SCM)	7	Maguire et al. 2019
Escherichia coli, *Enterobacter aerogenes*, *Enterobacter cloacae*, *K. pneumonia*	LR	Susceptibility classification	78	AMR genes	12	Pesesky et al. 2016
Acinetobacter baumannii, *Staphylococcus aureus*, *S. pneumoniae*, *M. tuberculosis*	AdaBoost	AMR classification	99–1,350	*k*-mer	1–5 for each organism	Davis et al. 2016
M. tuberculosis	LR, SVM	AMR classification	652	SNPs	4	Niehaus et al. 2014

Abbreviations: AMR, antimicrobial resistance; LR, logistic regression; MIC, minimum inhibitory concentration; SCM, Set Covering Machine; SNP, single-nucleotide polymorphism; SVM, support vector machine; WGS, whole-genome sequencing.

Modified image from Deng et al., 2021. Accessed Dec. 14, 2021 at https://www.researchgate.net/profile/Abigail-Horn/publication/348655838_Emerging_Applications_of_Machine_Learning_in_Food_Safety/links/605d1277458515e8346fe9a0/Emerging-Applications-of-Machine-Learning-in-Food-Safety.pdf.

Useful Deng et al., 2021 - Figure 2



Acquire raw data
For example:
Scrape Tweets geocoded to restaurants
Acquire Yelp restaurant reviews
Store content from web-crawling system

Preprocess data

Data cleaning
Spelling/grammar/casing
Filler word removal
Metadata noise removal

Data reduction
Filtering by keywords such as illness, food poisoning, vomit, throw up, barf, and nausea

Create labeled training data set
Label context as relevant/not relevant

Develop machine learning model

Classification
Relevant
Not relevant

Relevance ranking

Evaluate model on held-out data

Use model to make predictions on new, unlabeled data

Figure 2

Schematic framework of text data machine learning analysis approach.

Image from Deng et al., 2021. Accessed Dec. 14, 2021 at https://www.researchgate.net/profile/Abigail-Horn/publication/348655838_Emerging_Applications_of_Machine_Learning_in_Food_Safety/links/605d1277458515e8346fe9a0/Emerging-Applications-of-Machine-Learning-in-Food-Safety.pdf.

Useful Ford et al., 2020 - Figure 1

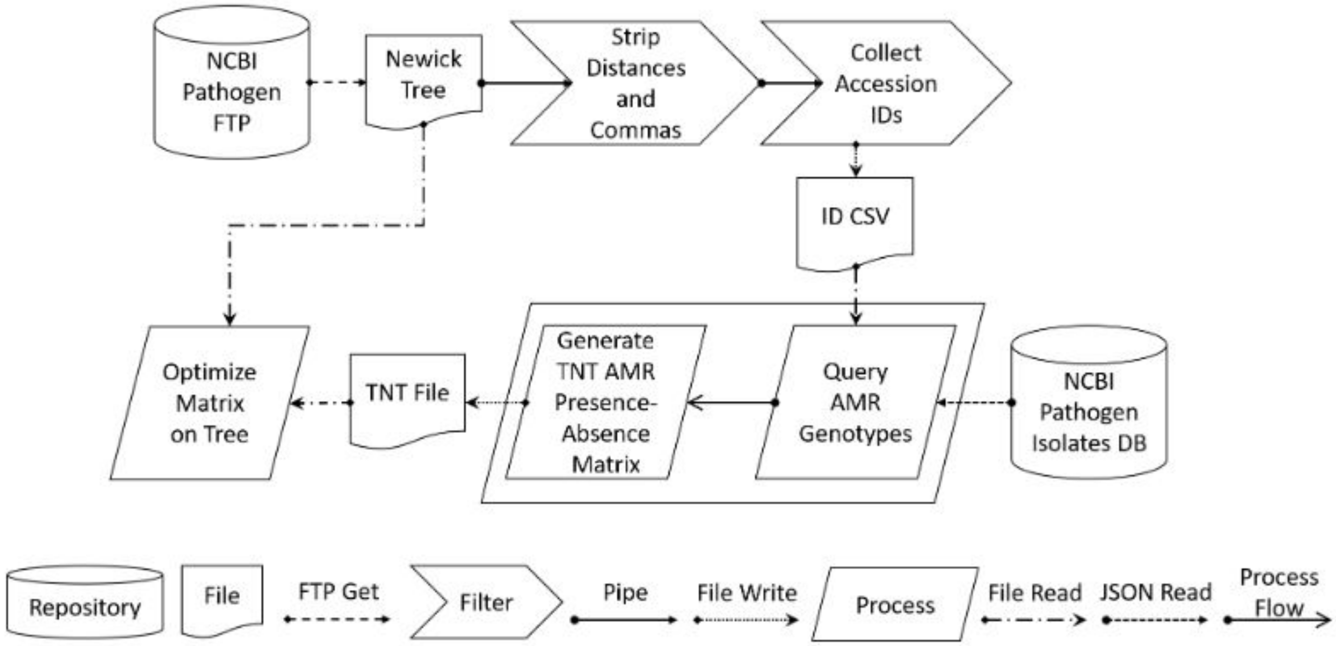
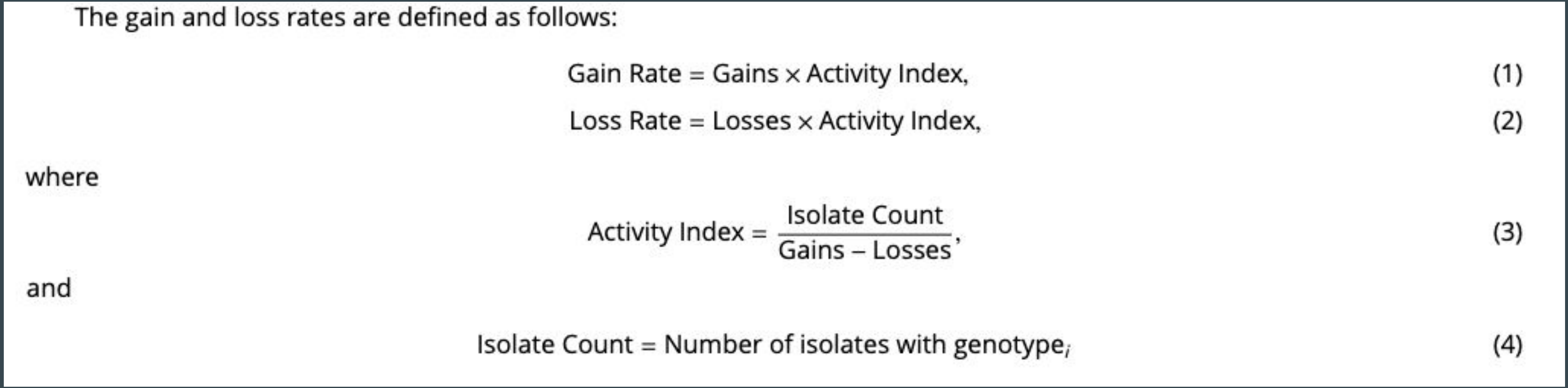


FIG 1 Workflow diagram of the TNT-based presence-absence creation and tree optimization. (Symbolic key at bottom.)

Image from Ford et al., 2020.

Useful Ford et al., 2020 - Method Excerpts



The gain and loss rates are defined as follows:

$$\text{Gain Rate} = \text{Gains} \times \text{Activity Index}, \tag{1}$$

$$\text{Loss Rate} = \text{Losses} \times \text{Activity Index}, \tag{2}$$

where

$$\text{Activity Index} = \frac{\text{Isolate Count}}{\text{Gains} - \text{Losses}}, \tag{3}$$

and

$$\text{Isolate Count} = \text{Number of isolates with genotype}_i \tag{4}$$

Image from Ford et al., 2020.

Useful Haendiges et al., 2021 - Figure 3

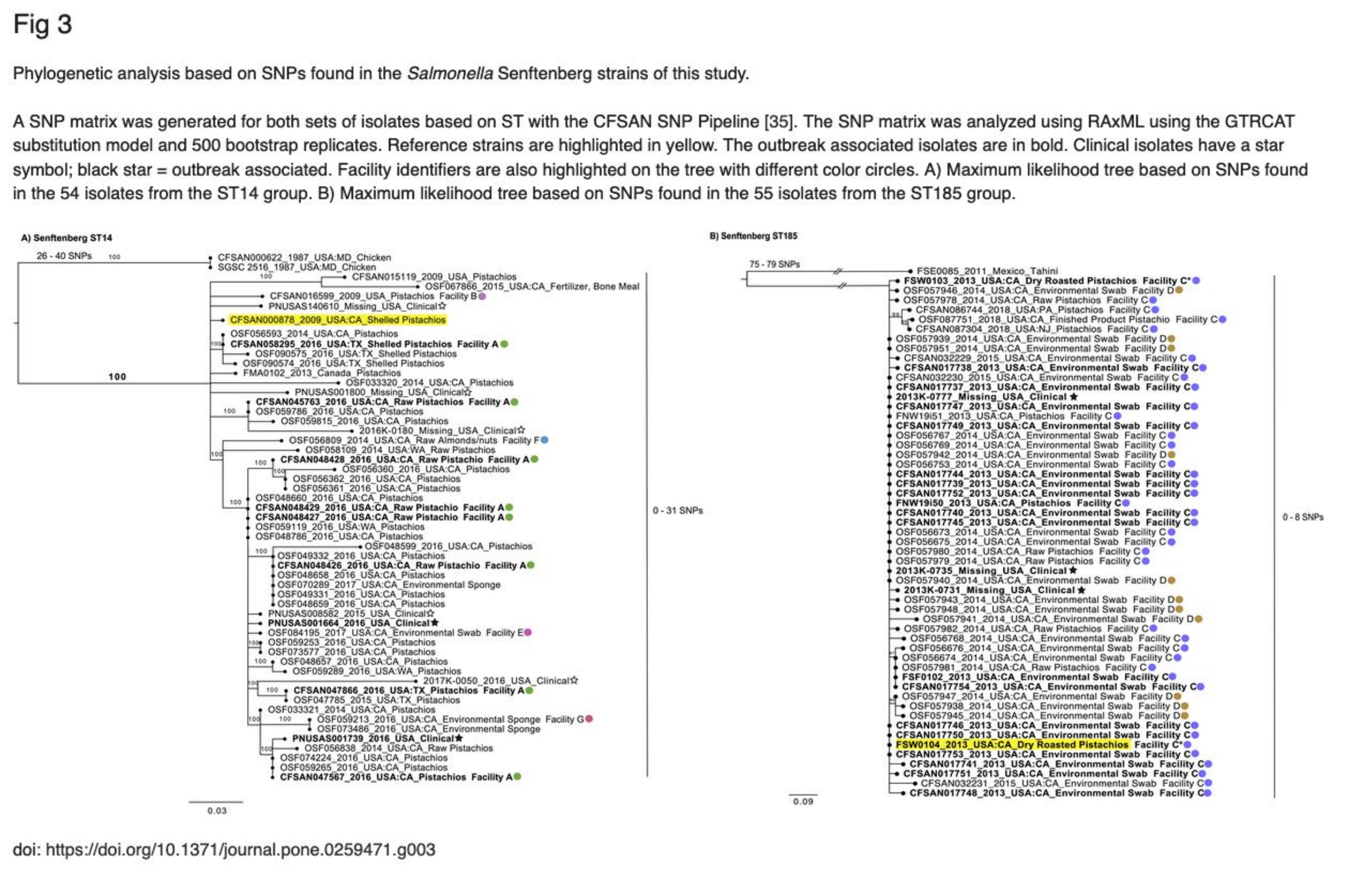


Fig 3

Phylogenetic analysis based on SNPs found in the *Salmonella* Senftenberg strains of this study.

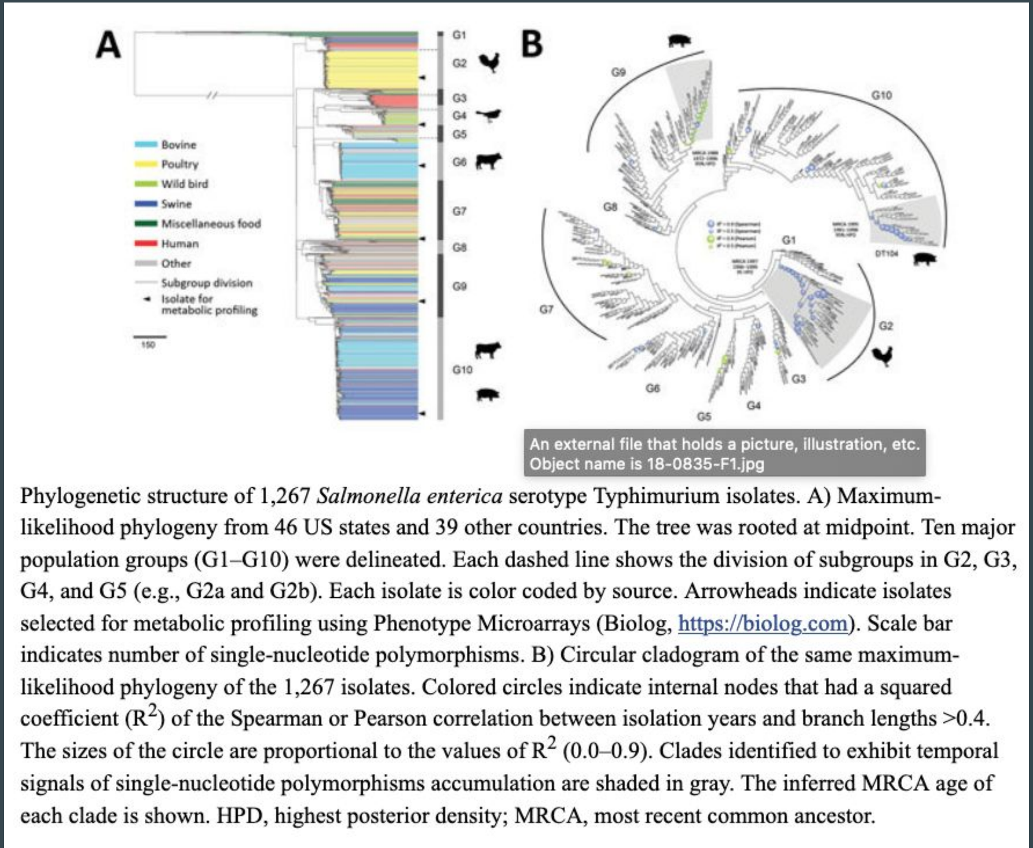
A SNP matrix was generated for both sets of isolates based on ST with the CFSAN SNP Pipeline [35]. The SNP matrix was analyzed using RAxML using the GTRCAT substitution model and 500 bootstrap replicates. Reference strains are highlighted in yellow. The outbreak associated isolates are in bold. Clinical isolates have a star symbol; black star = outbreak associated. Facility identifiers are also highlighted on the tree with different color circles. A) Maximum likelihood tree based on SNPs found in the 54 isolates from the ST14 group. B) Maximum likelihood tree based on SNPs found in the 55 isolates from the ST185 group.

A) Senftenberg ST14
26 - 40 SNPs
100
CFSAN000622_1987_USA:MD_Chicken
SGSC 2516_1987_USA:MD_Chicken
100
CFSAN015119_2009_USA_Pistachios
OSF067866_2015_USA:CA_Fertilizer, Bone Meal
CFSAN016599_2009_USA_Pistachios Facility B ●
PNUSAS140610_Missing_USA_Clinical ☆
CFSAN000878_2009_USA:CA_Shelled Pistachios
OSF056593_2014_USA:CA_Pistachios
100
CFSAN058295_2016_USA:TX_Shelled Pistachios Facility A ●
OSF090575_2016_USA:TX_Shelled Pistachios
OSF090574_2016_USA:TX_Shelled Pistachios
FMA0102_2013_Canada_Pistachios
100
OSF033320_2014_USA:CA_Pistachios
PNUSAS001800_Missing_USA_Clinical ☆
CFSAN045763_2016_USA:CA_Raw Pistachios Facility A ●
100
OSF059786_2016_USA:CA_Pistachios
OSF059815_2016_USA:CA_Pistachios
2016K-0180_Missing_USA_Clinical ☆
OSF056809_2014_USA:CA_Raw Almonds/nuts Facility F ●
OSF058109_2014_USA:WA_Raw Pistachios
100
100
CFSAN048428_2016_USA:CA_Raw Pistachio Facility A ●
OSF056360_2016_USA:CA_Pistachios
100
OSF056362_2016_USA:CA_Pistachios
OSF056361_2016_USA:CA_Pistachios
100
OSF048660_2016_USA:CA_Pistachios
CFSAN048429_2016_USA:CA_Raw Pistachio Facility A ●
CFSAN048427_2016_USA:CA_Raw Pistachio Facility A ●
OSF059119_2016_USA:WA_Pistachios
OSF048786_2016_USA:CA_Pistachios
OSF048599_2016_USA:CA_Pistachios
100
OSF049332_2016_USA:CA_Pistachios
CFSAN048426_2016_USA:CA_Raw Pistachio Facility A ●
OSF048658_2016_USA:CA_Pistachios
OSF070289_2017_USA:CA_Environmental Sponge
OSF049331_2016_USA:CA_Pistachios
OSF048659_2016_USA:CA_Pistachios
PNUSAS008582_2015_USA_Clinical ☆
PNUSAS001664_2016_USA_Clinical ★
OSF084195_2017_USA:CA_Environmental Swab Facility E ●
100
OSF059253_2016_USA:CA_Pistachios
OSF073577_2016_USA:CA_Pistachios
100
OSF048657_2016_USA:CA_Pistachios
OSF059289_2016_USA:WA_Pistachios
100
2017K-0050_2016_USA_Clinical ☆
100
CFSAN047866_2016_USA:TX_Pistachios Facility A ●
OSF047785_2015_USA:TX_Pistachios
OSF033321_2014_USA:CA_Pistachios
100
100
OSF059213_2016_USA:CA_Environmental Sponge Facility G ●
OSF073486_2016_USA:CA_Environmental Sponge
PNUSAS001739_2016_USA_Clinical ★
100
OSF056838_2014_USA:CA_Raw Pistachios
OSF074224_2016_USA:CA_Pistachios
OSF059265_2016_USA:CA_Pistachios
CFSAN047567_2016_USA:CA_Pistachios Facility A ●
0 - 31 SNPs
0.03
B) Senftenberg ST185
75 - 79 SNPs
FSE0085_2011_Mexico_Tahini
FSW0103_2013_USA:CA_Dry Roasted Pistachios Facility C* ●
OSF057946_2014_USA:CA_Environmental Swab Facility D ●
OSF057978_2014_USA:CA_Raw Pistachios Facility C ●
CFSAN086744_2018_USA:PA_Pistachios Facility C ●
OSF087751_2018_USA:CA_Finished Product Pistachio Facility C ●
CFSAN087304_2018_USA:NJ_Pistachios Facility C ●
OSF057939_2014_USA:CA_Environmental Swab Facility D ●
OSF057951_2014_USA:CA_Environmental Swab Facility D ●
CFSAN032229_2015_USA:CA_Environmental Swab Facility C ●
CFSAN017738_2013_USA:CA_Environmental Swab Facility C ●
CFSAN032230_2015_USA:CA_Environmental Swab Facility C ●
CFSAN017737_2013_USA:CA_Environmental Swab Facility C ●
2013K-0777_Missing_USA_Clinical ★
CFSAN017747_2013_USA:CA_Environmental Swab Facility C ●
FNW19i51_2013_USA:CA_Pistachios Facility C ●
CFSAN017749_2013_USA:CA_Environmental Swab Facility C ●
OSF056767_2014_USA:CA_Environmental Swab Facility C ●
OSF056769_2014_USA:CA_Environmental Swab Facility C ●
OSF057942_2014_USA:CA_Environmental Swab Facility D ●
OSF056753_2014_USA:CA_Environmental Swab Facility C ●
CFSAN017744_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017739_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017752_2013_USA:CA_Environmental Swab Facility C ●
FNW19i50_2013_USA:CA_Pistachios Facility C ●
CFSAN017740_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017745_2013_USA:CA_Environmental Swab Facility C ●
OSF056673_2014_USA:CA_Environmental Swab Facility C ●
OSF056675_2014_USA:CA_Environmental Swab Facility C ●
OSF057980_2014_USA:CA_Raw Pistachios Facility C ●
OSF057979_2014_USA:CA_Raw Pistachios Facility C ●
2013K-0735_Missing_USA_Clinical ★
OSF057940_2014_USA:CA_Environmental Swab Facility D ●
2013K-0731_Missing_USA_Clinical ★
OSF057943_2014_USA:CA_Environmental Swab Facility D ●
OSF057948_2014_USA:CA_Environmental Swab Facility D ●
OSF057941_2014_USA:CA_Environmental Swab Facility D ●
OSF057982_2014_USA:CA_Raw Pistachios Facility C ●
OSF056768_2014_USA:CA_Environmental Swab Facility C ●
OSF056676_2014_USA:CA_Environmental Swab Facility C ●
OSF056674_2014_USA:CA_Environmental Swab Facility C ●
OSF057981_2014_USA:CA_Raw Pistachios Facility C ●
FSF0102_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017754_2013_USA:CA_Environmental Swab Facility C ●
OSF057947_2014_USA:CA_Environmental Swab Facility D ●
OSF057938_2014_USA:CA_Environmental Swab Facility D ●
OSF057945_2014_USA:CA_Environmental Swab Facility D ●
CFSAN017746_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017750_2013_USA:CA_Environmental Swab Facility C ●
FSW0104_2013_USA:CA_Dry Roasted Pistachios Facility C* ●
CFSAN017753_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017741_2013_USA:CA_Environmental Swab Facility C ●
CFSAN017751_2013_USA:CA_Environmental Swab Facility C ●
CFSAN032231_2015_USA:CA_Environmental Swab Facility C ●
CFSAN017748_2013_USA:CA_Environmental Swab Facility C ●
0 - 8 SNPs
0.09

doi: https://doi.org/10.1371/journal.pone.0259471.g003

Image from Haendiges et al., 2021. Accessed Dec. 14, 2021 at https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0259471.g003.

Useful Zhang et al., 2019 - Figure 1

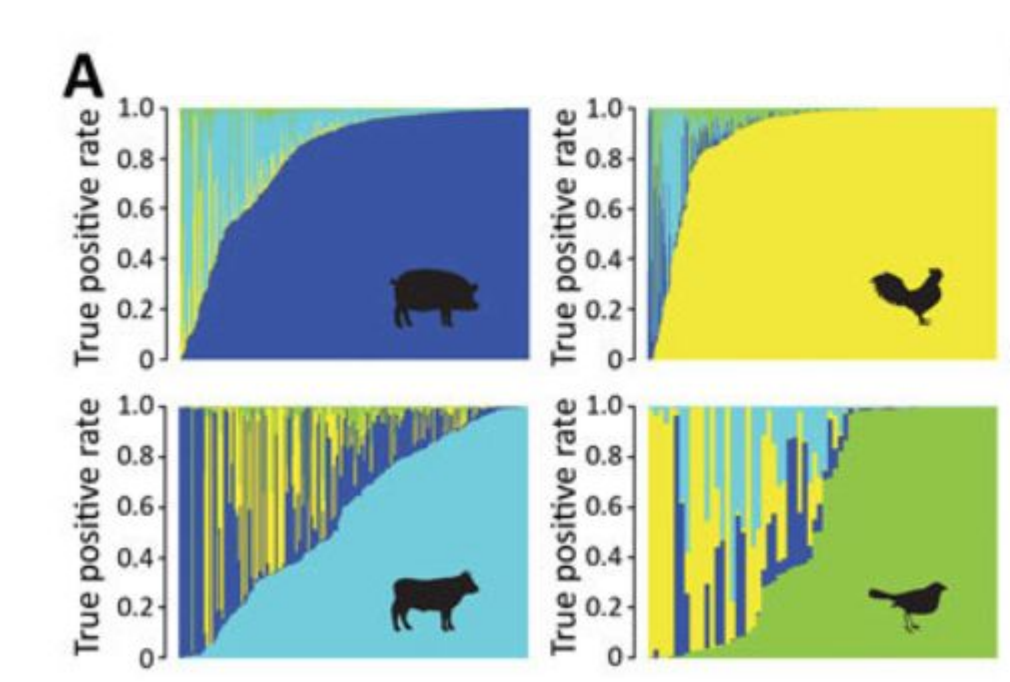


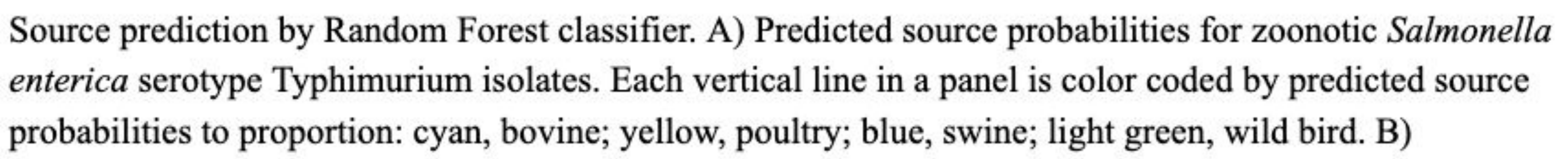
An external file that holds a picture, illustration, etc.
Object name is 18-0835-F1.jpg

Phylogenetic structure of 1,267 *Salmonella enterica* serotype Typhimurium isolates. A) Maximum-likelihood phylogeny from 46 US states and 39 other countries. The tree was rooted at midpoint. Ten major population groups (G1–G10) were delineated. Each dashed line shows the division of subgroups in G2, G3, G4, and G5 (e.g., G2a and G2b). Each isolate is color coded by source. Arrowheads indicate isolates selected for metabolic profiling using Phenotype Microarrays (Biolog, https://biolog.com). Scale bar indicates number of single-nucleotide polymorphisms. B) Circular cladogram of the same maximum-likelihood phylogeny of the 1,267 isolates. Colored circles indicate internal nodes that had a squared coefficient (R^2) of the Spearman or Pearson correlation between isolation years and branch lengths >0.4. The sizes of the circle are proportional to the values of R^2 (0.0–0.9). Clades identified to exhibit temporal signals of single-nucleotide polymorphisms accumulation are shaded in gray. The inferred MRCA age of each clade is shown. HPD, highest posterior density; MRCA, most recent common ancestor.

Image from Zhang et al., 2019. Accessed Dec. 14, 2021 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302586/.

Useful Zhang et al., 2019 - Figure 3A

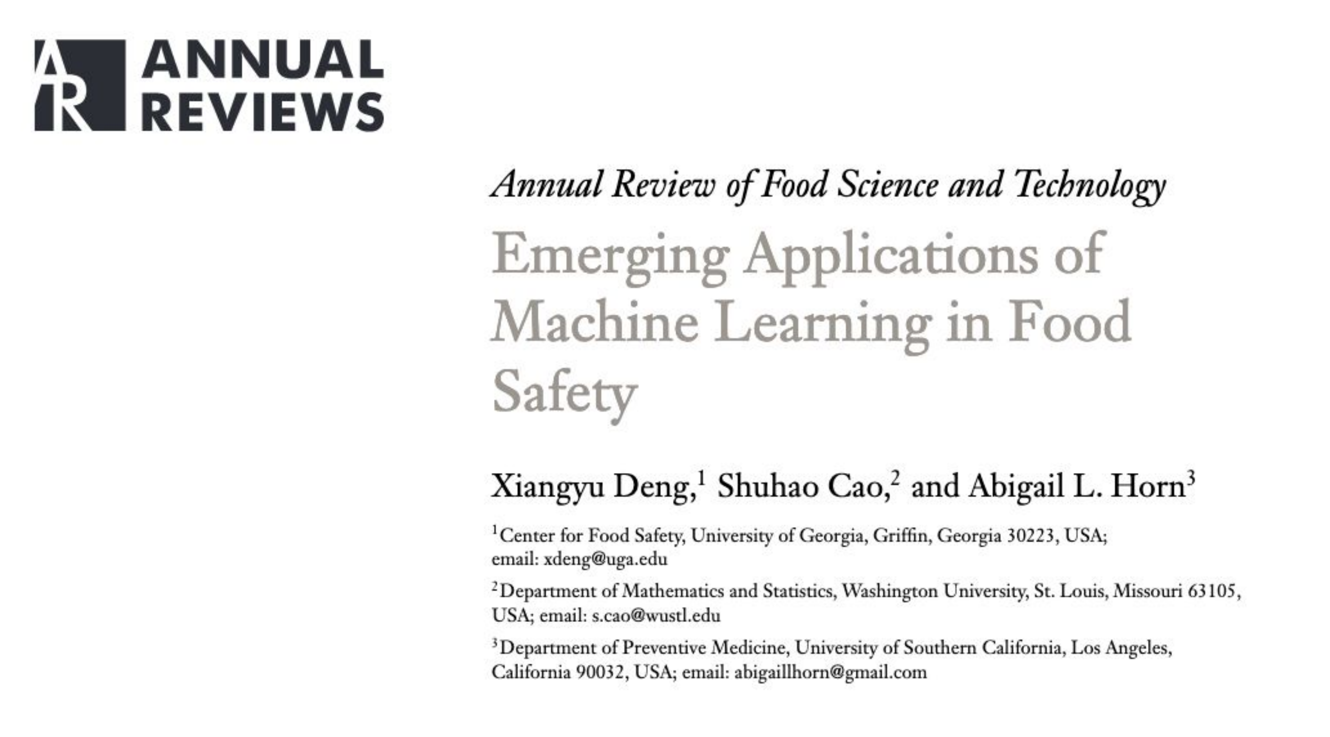




Source prediction by Random Forest classifier. A) Predicted source probabilities for zoonotic *Salmonella enterica* serotype Typhimurium isolates. Each vertical line in a panel is color coded by predicted source probabilities to proportion: cyan, bovine; yellow, poultry; blue, swine; light green, wild bird. B)

Image from Zhang et al., 2019. Accessed Dec. 14, 2021 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302586/.

Useful Machine Learning Guidelines



ANNUAL REVIEWS

Annual Review of Food Science and Technology

Emerging Applications of Machine Learning in Food Safety

Xiangyu Deng,[1] Shuhao Cao,[2] and Abigail L. Horn[3]

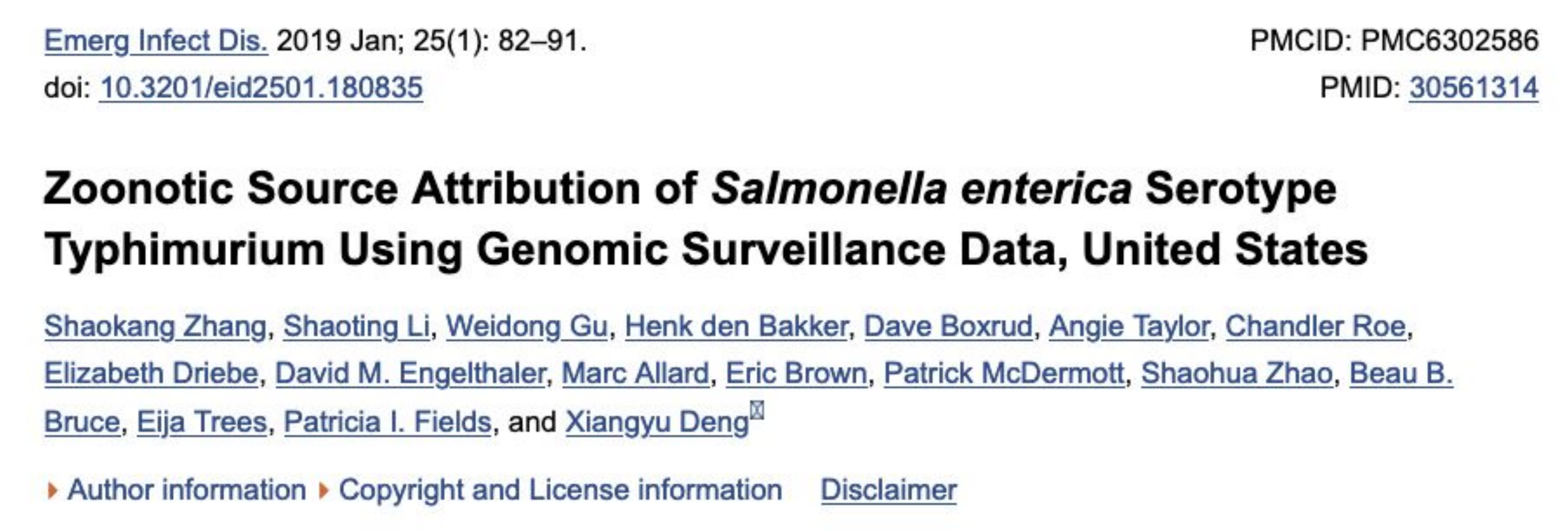
[1] Center for Food Safety, University of Georgia, Griffin, Georgia 30223, USA; email: xdeng@uga.edu

[2] Department of Mathematics and Statistics, Washington University, St. Louis, Missouri 63105, USA; email: s.cao@wustl.edu

[3] Department of Preventive Medicine, University of Southern California, Los Angeles, California 90032, USA; email: abigaillhorn@gmail.com

Image from Deng, et al. (https://doi.org/10.1146/annurev-food-071720-024112). Accessed Dec. 13, 2021 at https://www.researchgate.net/profile/Abigail-Horn/publication/348655838_Emerging_Applications_of_Machine_Learning_in_Food_Safety/links/605d1277458515e8346fe9a0/Emerging-Applications-of-Machine-Learning-in-Food-Safety.pdf.

Useful Machine Learning Guidelines - cont.



Emerg Infect Dis. 2019 Jan; 25(1): 82–91.
doi: 10.3201/eid2501.180835

PMCID: PMC6302586
PMID: 30561314

Zoonotic Source Attribution of *Salmonella enterica* Serotype Typhimurium Using Genomic Surveillance Data, United States

Shaokang Zhang, Shaoting Li, Weidong Gu, Henk den Bakker, Dave Boxrud, Angie Taylor, Chandler Roe, Elizabeth Driebe, David M. Engelthaler, Marc Allard, Eric Brown, Patrick McDermott, Shaohua Zhao, Beau B. Bruce, Eija Trees, Patricia I. Fields, and Xiangyu Deng✉

▸ Author information ▸ Copyright and License information Disclaimer

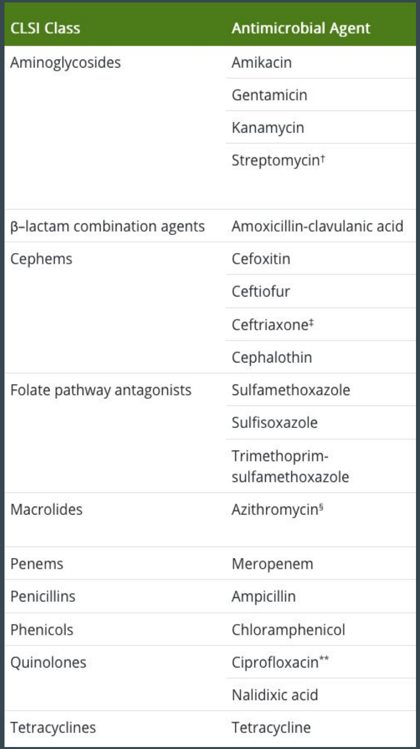
Image from Zhang, et al. (10.3201/eid2501.180835). Accessed Dec. 13, 2021 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302586/.

Immediate Next Steps

- Confirm CFSAN SNP Pipeline with the RHEL 08 upgrade.
- Avoid the Jan. 4th - 9th HPC outage.
- Calculate the CFSAN SNP Pipeline outputs.

CHAPTER 1 - Introduction and Background - Recap

- **Background**
  - Current silent pandemic of antibiotic resistance genes
  - Circumstances are worsening due to:
    - Proliferation in number of resistance genes
    - Increased amount of resisted drug classes
    - Geographic spread
  - Culminates in widespread multidrug-resistance
    - Defined as "resistant to at least one antibiotic in three or more drug classes."
      - https://www.cdc.gov/narms/resources/glossary.html
- **Data Validity**
  - Food-associated bacteria are widely multidrug resistant
  - Represents a diversity of sample sources
    - Globally
    - Clinical and/or environmental
  - *Escherichia coli*
    - History as a model organism in biology
    - Current research methods focus on *Salmonella*



CLSI Class	Antimicrobial Agent
Aminoglycosides	Amikacin
	Gentamicin
	Kanamycin
	Streptomycin†
β–lactam combination agents	Amoxicillin-clavulanic acid
Cephems	Cefoxitin
	Ceftiofur
	Ceftriaxone‡
	Cephalothin
Folate pathway antagonists	Sulfamethoxazole
	Sulfisoxazole
	Trimethoprim-sulfamethoxazole
Macrolides	Azithromycin§
Penems	Meropenem
Penicillins	Ampicillin
Phenicols	Chloramphenicol
Quinolones	Ciprofloxacin**
	Nalidixic acid
Tetracyclines	Tetracycline