

AMR Resources at NCBI's Pathogen Portal

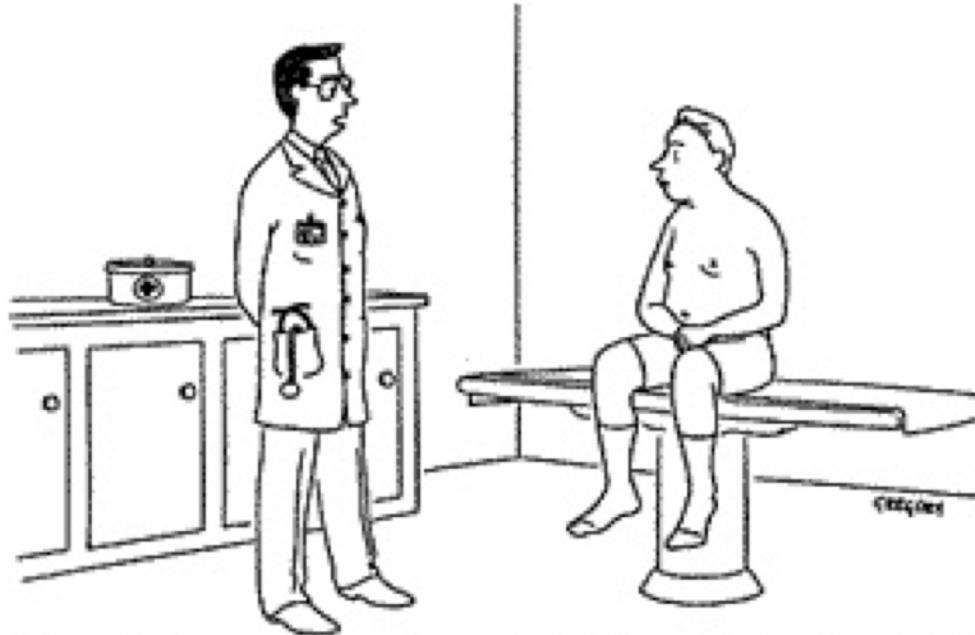
Michael Feldgarden

Michael.Feldgarden@nih.gov



U.S. National Library of Medicine
National Center for Biotechnology Information

End of an Era



"Your infection may be antibiotic-resistant, but let's see how it responds to intensive litigation."

NCBI's Role

NATIONAL ACTION PLAN FOR COMBATING ANTIBIOTIC-RESISTANT BACTERIA

MARCH 2015



“Create a repository of resistant bacterial strains (an “isolate bank”) and **maintain a well-curated reference database that describes the characteristics of these strains.**”

“Develop and maintain a **national sequence database of resistant pathogens.**”

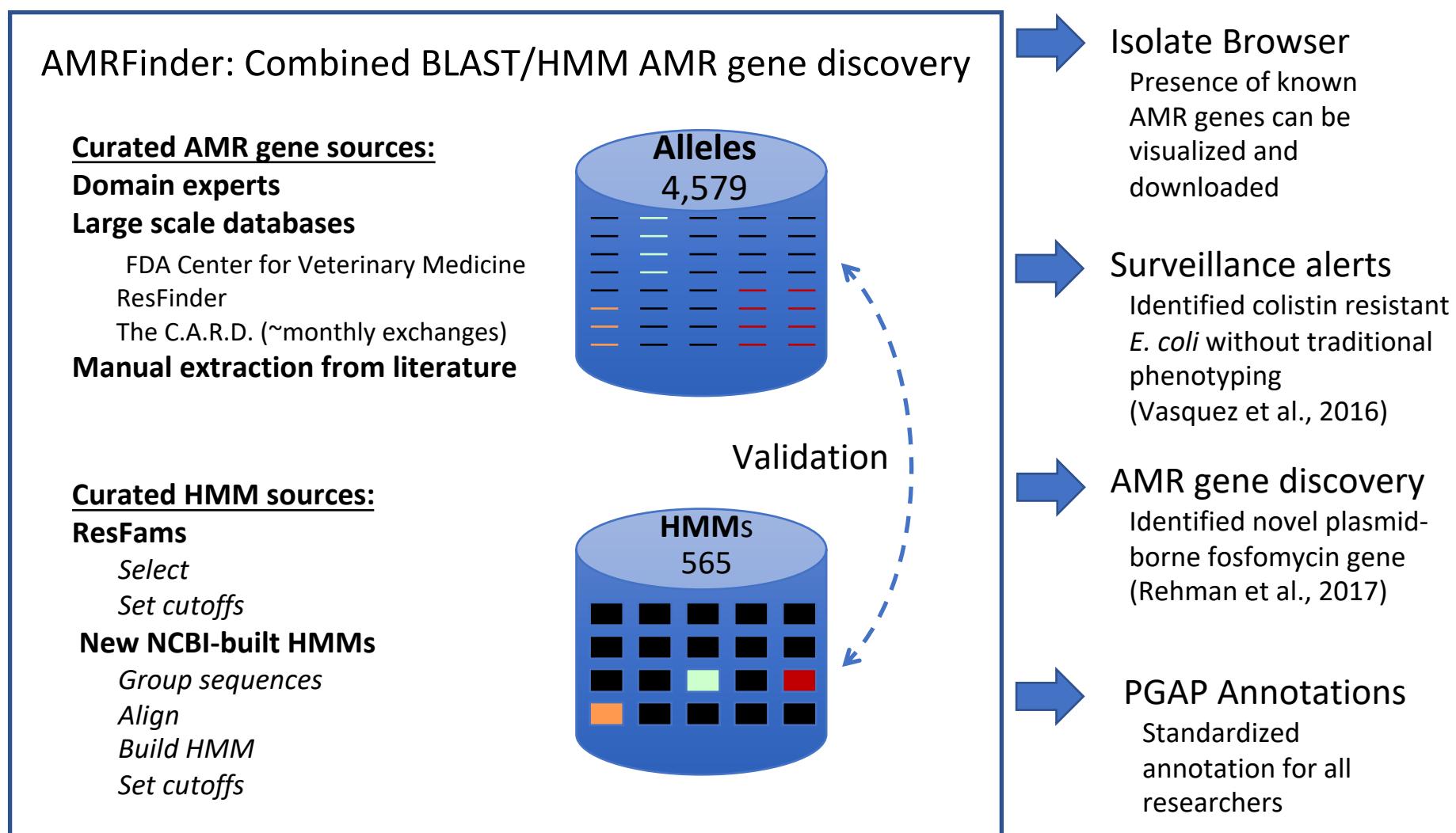
Automating AMR Gene Detection

Pathogen Detection Pipeline (SRA):

GenomeTrakr
PulseNet
PHE
FDA/CDC Antimicrobial Resistant
Isolate Bank
State laboratories
Clinical laboratories

Genbank assemblies:

General submissions



Routine Genomic Surveillance Finds the Fourth US *mcr-1*⁺ Isolate

- A stool sample was collected from a pediatric patient with diarrhea
- Non-Shiga toxin-producing *E. coli* O157 was isolated and sequenced
- The sequence data was uploaded to SRA

NCBI

- The Pathogen Detection pipeline assembled and annotated the genome
- The antimicrobial resistance module detected an *mcr-1* gene encoded by this sample (*mcr-1* is a recently discovered colistin resistance gene that has been found on mobile elements - Liu et al., 2016)
- CDC was notified
- CDC published this information recently in MMWR - note it is the fourth isolate in the US to have been found to encode *mcr-1* (Vasquez et al., 2016) and the first to have been discovered by the Pathogen Detection system (in all three previous cases the data were not submitted prior to the announcement)

Novel Gene Discovery: FosA7

- Fosfomycin is used to treat uncomplicated UTI (single dose therapy)
- Typically not screened in AST panels
- In 2017, AMRFinder identified a possible Fos gene (fosfomycin resistance) that was widespread (many hundreds of isolates) in NCBI's Pathogen Detection System
- Suggested to collaborators that laboratory work was needed
- Work by Rehman et al. 2017 indicates this is a novel plasmid-borne fosfomycin resistance gene, *fosA7*
 - glutathione transferase
- Found in ~2.5% of *Salmonella* in NCBI's Pathogen Detection System

Building an AMR Database

Domain experts

Bush and Jacoby (beta-lactamases)
Marilyn Roberts (MLS/tetracycline)
Pasteur Institute (beta-lactamases)

Large scale databases

FDA Center for Veterinary Medicine
ResFinder
The C.A.R.D. (~monthly exchanges)

Manual extraction from literature

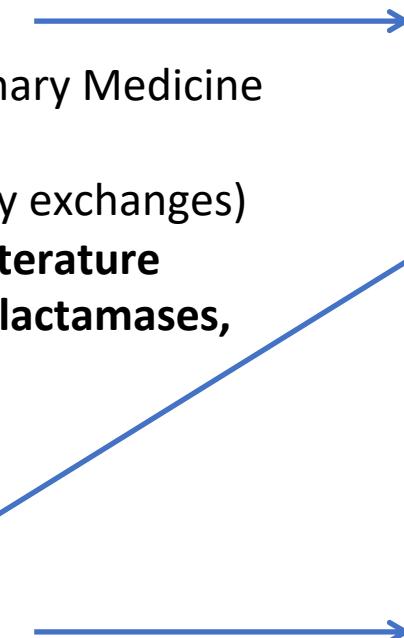
Ongoing curation of beta-lactamases,
Qnr, and MCR

ResFams

Select
Set cutoffs

New HMMs

Group sequences
Align
Build HMM
Set cutoffs



allele = unique protein (*blaTEM-1*)
gene = set of related proteins (*sul1*)

The Role of Manual Curation

- Only full-length genes are included
 - important for identifying best hit
- Start sites are curated
 - *attC* sites are removed
 - leader peptides verified
- Protein names are standardized in format for bioinformatic ease
- Gene symbol (name) hierarchy is curated
- HMM cutoffs are verified to include known AMR genes, and *exclude* related sequences that do not affect AMR
 - Haft, DH, *et al.*, 2018 **PMID: 29112715**

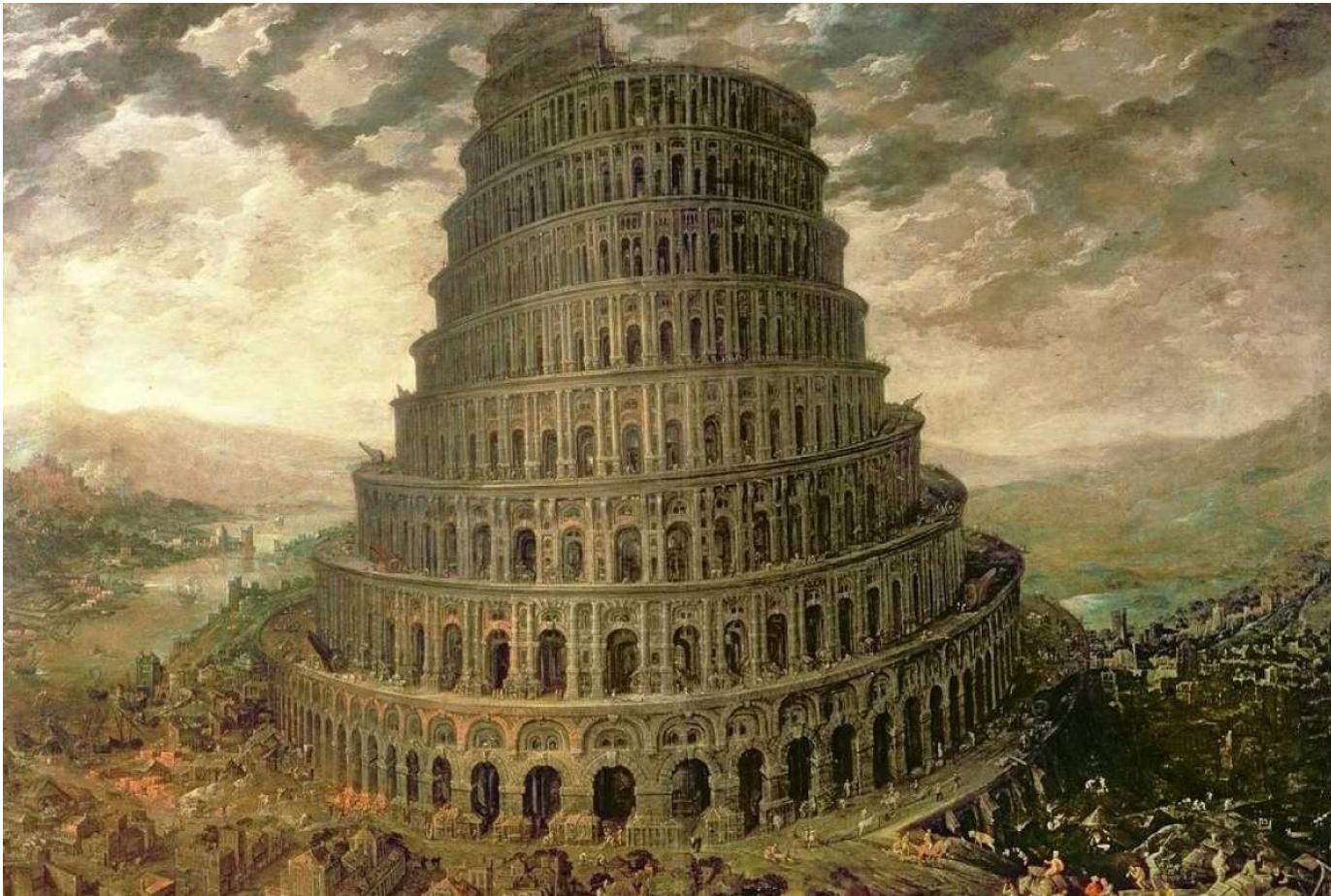
<http://ncbi.nlm.nih.gov/bioproject/PRJNA313047>

The Utility of HMMs: ‘Beta-lactamases’ in GenBank

- Examined **GenBank** protein sequences that had ‘beta-lactamase’ in product name and not described as partial or synthetic constructs:
 - Only **11%** of sequences (108,386/1,030,160) appear to be beta-lactamases
 - Only **20%** of unique proteins (27,682/137,297) appear to be beta-lactamases
- Examined 21 putative metallo- β -lactamases from metagenomic data that had been functionally characterized:
 - AMRFinder correctly identified the 18 functional metallo- β -lactamases
 - AMRFinder correctly did not call the 3 non-functional proteins as beta-lactamases

Berglund *et al.* 2017. Identification of 76 novel B1 metallo- β -lactamases through large-scale screening of genomic and metagenomic data. *Microbiome* 5:134

Lessons Learned from Database Construction: Nomenclature



Lessons Learned from Database Construction: Nomenclature

- Aminoglycoside modifying enzymes (AMEs)
 - genes vs. alleles
 - aac(6')-Ib
 - aac(3)-I
 - two non-overlapping, partially complete nomenclatures
- OXA beta-lactamase nomenclature
- MCR: NCBI is now curating these genes

<https://www.ncbi.nlm.nih.gov/pathogens/submit-beta-lactamase/>

AME nomenclature

"Two nomenclature schemes have been proposed in the past, but the dizzying rate of discovery of new genes together with the appearance of enzymes with new characteristics superseded the criteria defined. We suggest that members of the community should engage in a debate to come up with a consensus new nomenclature.

We suggest that returning to a simpler nomenclature with the support of an internet repository site could facilitate the naming of the genes, avoid duplications, and facilitate further changes when new enzymes with new, and may be unexpected, characteristics are discovered."

Ramirez and Tomalsky, 2010



AME nomenclature

AME family	No. of Proteins (including fusions)
aac	306
ant	229
aph	79

N-Acetyltransferases (AAC) – catalyzes acetyl CoA-dependent acetylation of an amino group

O-Adenyltransferases (ANT) – catalyzes ATP-dependent adenylation of hydroxyl group

O-Phosphotransferases (APH) – catalyzes ATP-dependent phosphorylation of a hydroxyl group



Many Synonyms in AME Nomenclature

ANT(9)-Ia	<i>ant(9)-Ia, aad(9), spc</i>
ANT(9)-Ib	<i>ant(9)-Ib, aad(9), spc</i>
ANT(4')-Ia C	<i>ant(4')-Ia, aadD2, aadD, ant(4',4'')-I</i>
ANT(4')-IIa	<i>ant(4')-IIa</i>
ANT(4')-IIb	<i>ant(4')-IIb</i>
ANT(2'')-Ia	<i>ant(2'')-Ia, aadB</i>
ANT(3'')-Ia	<i>ant(3'')-Ia, aadA, aadA1, aad(3'')(9) aadA2</i>

CARD is a good resource for
AME synonyms



U.S. National Library of Medicine



NCBI

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



Soon:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.



AME nomenclature

General agreement from previous roundtable discussions:

- stop using names of the aac(6')-Ib type
- assign numbers to unique proteins
- an indication of phenotype is not to be part of gene/protein name
- use something like aacA1-1 where -1, -2 etc indicate different proteins
- may have forum at ASM Microbe 2019 to discuss this

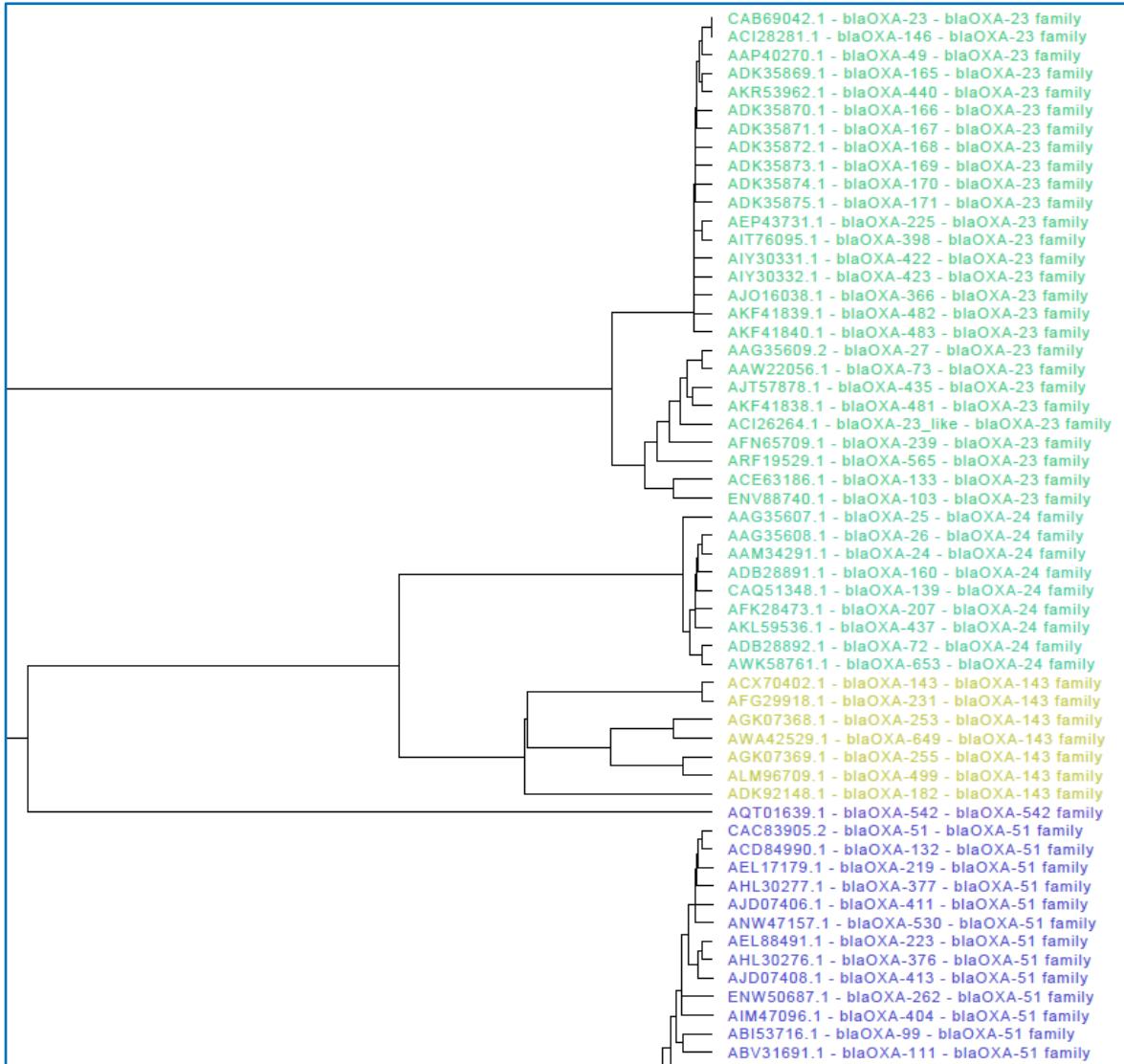
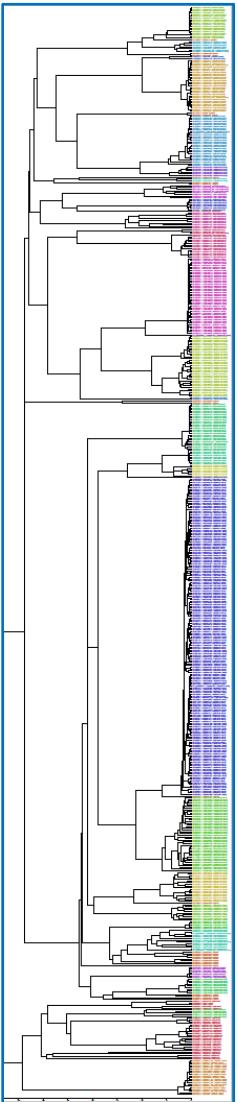


Lessons Learned from Database Construction

- Nomenclature
 - AME
 - genes vs. families
 - aac(6')-Ib
 - aac(3)-I
 - two non-overlapping, partially complete nomenclatures
 - OXA beta-lactamase nomenclature
 - MCR: NCBI is now curating these genes

<https://www.ncbi.nlm.nih.gov/pathogens/submit-beta-lactamase/>

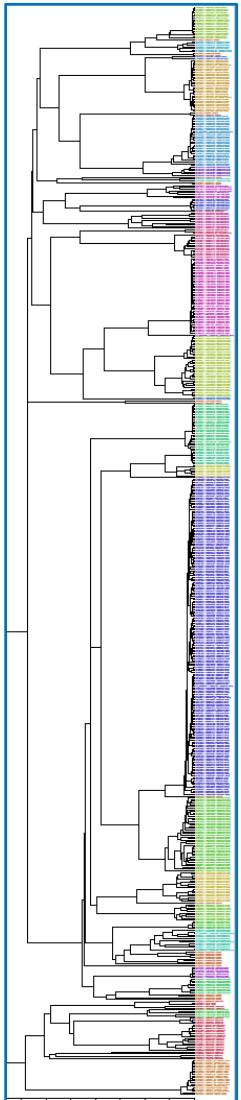
OXA Nomenclature: Too Many Things Called OXA



Only common biochemical feature is oxacillinase activity

Potentially important component of AMR in *Acinetobacter* (and other species)





OXA Nomenclature: Too Many Things Called OXA

<https://www.ncbi.nlm.nih.gov/nuccore/1035503506/>

gene

101..922
/gene="blaOXA"
/locus_tag="A7J11_03003"
/allele="blaOXA-103"



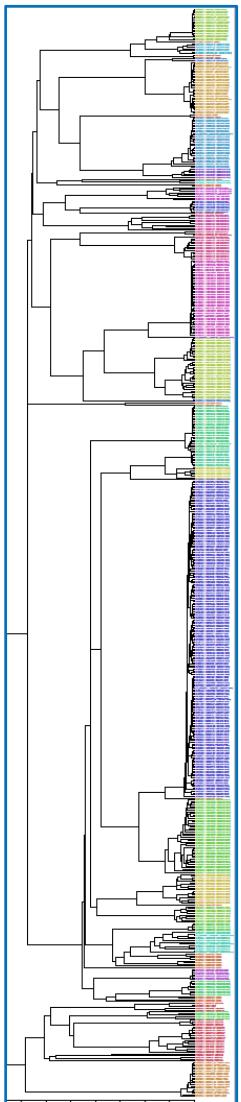
CDS

101..922
/gene="blaOXA"
/locus_tag="A7J11_03003"
/allele="blaOXA-103"
/codon_start=1
/transl_table=11
/product="OXA-23 family carbapenem-hydrolyzing class D beta-lactamase OXA-103"
/protein_id="WP_005025422.1"



1. gene and allele nomenclature stay as is
2. protein product name has additional family association





OXA Nomenclature: Too Many Things Called OXA

(proposal by Dan Haft @NCBI)

class D beta-lactamases, nearly all of which are called OXA & extremely diverse

1. did not want to break existing genetic nomenclature
 - all genes *blaOXA-XXX*
2. class D beta-lactamases fall into families, the larger families are well known
 - ex: OXA-51
3. tentative proposal for shorthand in scientific literature (tables, etc.)
 - put family association in parentheses
 - ex. OXA-75(51) versus OXA-51 family carbapenem-hydrolyzing class D beta-lactamase OXA-75



Lessons Learned from Database Construction

- Nomenclature
 - AME
 - genes vs. families
 - aac(6')-Ib
 - aac(3)-I
 - two non-overlapping, partially complete nomenclatures
 - OXA beta-lactamase nomenclature
 - MCR: NCBI is now curating these genes
- Many genes have not been phenotyped

<https://www.ncbi.nlm.nih.gov/pathogens/submit-beta-lactamase/>

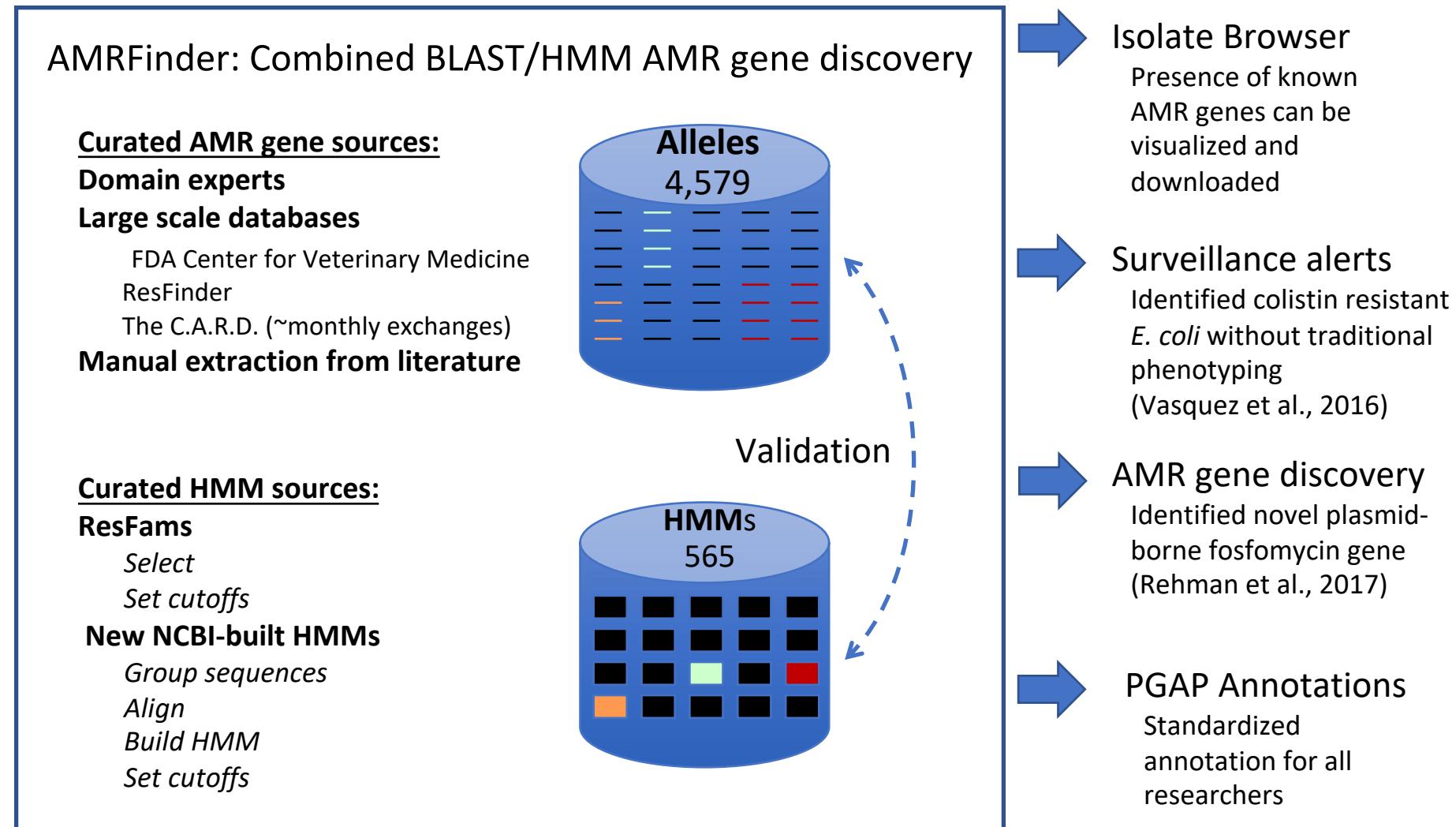
AMRFinder: Automating AMR Gene Detection

Pathogen Detection Pipeline (SRA):

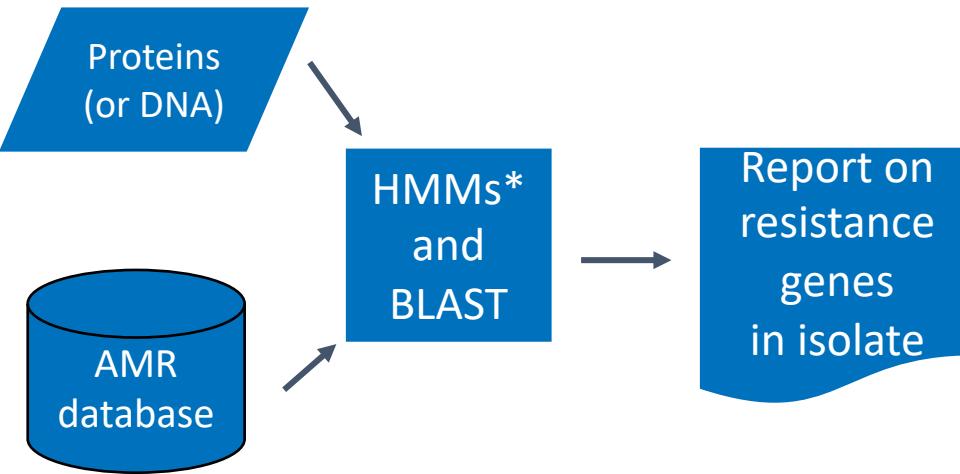
GenomeTrakr
PulseNet
PHE
FDA/CDC Antimicrobial Resistant Isolate Bank
State laboratories
Clinical laboratories

Genbank assemblies:

General submissions



AMRFinder Overview



4,579 resistance proteins
565 HMMs
34 drug classes resisted
~50% beta-lactamases

Hierarchical Structure

	Functional determination
Exact match	<p>Protein name</p> <p>KPC-2</p>
HMM score > cutoff of KPC family	<p>Likely resistance to carbapenems and other beta-lactam antibiotics.</p> <p>KPC family</p>
HMM score > cutoff	<p>Class A β-lactamase</p>
HMM score > cutoff	<p>not beta-lactamase</p> <p>Prevents false-positive identification as a beta-lactamase. Not reported.</p>

- AMRFinder now can search unannotated genomes

<https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/>

Need Comprehensive Test Case: NARMS Data

- National Antimicrobial Resistance Monitoring System
 - tracks changes in the antimicrobial susceptibility of enteric bacteria found in ill people (CDC), retail meats (FDA), and food animals (USDA) in the United States
- Thousands of food-borne pathogens
- Both genomic and AST data are available
 - AST data generated using standard NARMS Sensititre™ panels
- Allows us to confirm genotype predictions using phenotype data (AST)

Experimental Design

- Only examined those isolates that passed assembly validation checks
- Excluded isolates that had genotype-phenotype conflicts in three or more drug classes (0.5% of isolates)
- **Examined 6,242 isolates:**
 - 294 *C. coli* isolates
 - 476 *C. jejuni* isolates
 - 47 *E. coli* isolates
 - 5,245 *S. enterica* isolates
- For *Campylobacter*, **point mutations for macrolide and fluoroquinolone resistance were also assessed**, as previous work indicated they are the dominant mode of resistance (Zhao et al., 2015)
- For *S. enterica*, **point mutations were assessed for fluoroquinolone resistance** too (Tyson et al., 2017)

AMRFinder-Resfinder Comparison

	AMRFinder	Resfinder
misclassification	0	247
underspecification	5	0
overspecification	0	977

- Gene calls for near identical regions on genomes were compared (+/- 40 bp) between AMRFinder and Resfinder using default settings
- 88% of calls were identical between the two systems ($N_{\text{total}} = 14,023$)
- Misclassifications were due to missing sequences or differences in nucleotide and protein distances
- Overspecifications were either due to novel or partial sequences
- Incorrect assignments can be important:
 - Aminoglycoside modifying enzymes were miscalled:
 - 22 instances where *aac(6')-Ib** was miscalled as *aac(6')-Ib-cr*
 - *aac(6')-Ib-cr* confers amikacin, tobramycin, and ciprofloxacin resistance*, while *aac(6')-Ib** do not confer resistance to one or more of these drugs

Overall Consistency Was High

N = 87,679

	# resistant observations	# susceptible observations
# predicted resistant	13,122	781
# predicted sensitive	622	73,154

PPV = 0.955

NPV = 0.992

Campylobacter coli Has High Overall Consistency

Antibiotic	# isolates susceptible ^a	# isolates resistant ^a	% consistent ^b	% resistant	PPV ^c	NPV
azithromycin	265	29	100.0%	9.9%	0.763	1
ciprofloxacin	207	87	100.0%	29.6%	1	1
clindamycin	248	29 (17)	94.2%	15.6%	NC	0.844
erythromycin	265	29	100.0%	9.9%	0.763	1
florfenicol	294	0	100.0%	0.0%	NC	1
gentamicin	288	6	100.0%	2.0%	1	1
nalidixic acid	201 (3)	87 (3)	98.0%	30.6%	1	0.986
telithromycin	257 (16)	21	94.6%	7.1%	0.553	1
tetracycline	80 (3)	210 (1)	98.6%	71.8%	0.989	0.988

PPV = 0.904

NPV = 0.982

- Macrolides and ketolides had low PPV

Campylobacter jejuni Has Very High Consistency

Antibiotic	# isolates susceptible	# isolates resistant	% consistent	% resistant	PPV ^c	NPV
azithromycin	476	0	100.0%	0.0%	0	1
ciprofloxacin	386 (1)	86 (3)	99.2%	18.7%	0.989	0.992
clindamycin	470	0 (6)	98.7%	1.3%	NC	0.987
erythromycin	476	0	100.0%	0.0%	0	1
florfenicol	476	0	100.0%	0.0%	NC	1
gentamicin	475	0 (1)	99.8%	0.2%	NC	0.998
nalidixic acid	385 (3)	86 (2)	98.9%	18.7%	0.977	0.992
telithromycin	476	0	100.0%	0.0%	0	1
tetracycline	145 (4)	325 (2)	98.7%	68.9%	0.988	0.986

PPV = 0.971

NPV = 0.992

- Very high consistency between genotype and phenotype

S. enterica Has High Consistency

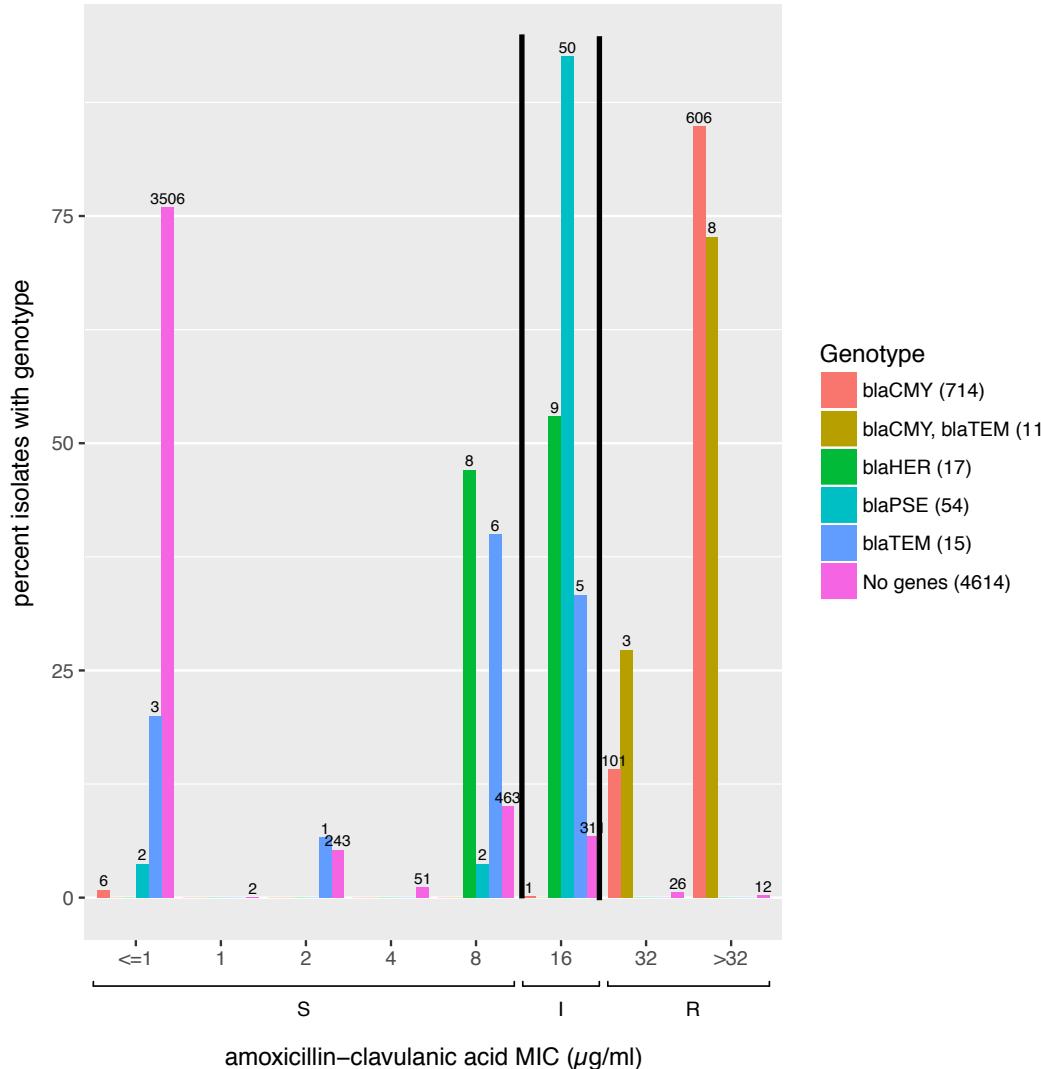
Antibiotic	# isolates susceptible ^a	# isolates resistant ^a	% consistent ^b	% resistant	PPV ^c	NPV
amikacin	2820	0	100.0%	0.0%	NC	1
AMC	4622 (7)	718 (38)	99.2%	14.0%	0.99	0.992
ampicillin	3734 (27)	1620 (44)	98.7%	30.7%	0.984	0.988
azithromycin	2592 (5)	6 (1)	99.8%	0.3%	0.545	0.999
cefoxitin	4686 (67)	658 (14)	98.5%	12.4%	0.908	0.997
ceftiofur	4093 (13)	697 (13)	99.5%	14.7%	0.982	0.997
ceftriaxone	4652 (8)	744 (21)	99.5%	14.7%	0.989	0.996
CHL	5214 (5)	202 (4)	99.8%	3.8%	0.976	0.999
ciprofloxacin ^d	5283 (14)	114 (14)	99.5%	2.4%	0.891	0.997
cotrimoxazole	5343 (8)	69 (5)	99.8%	1.4%	0.896	0.999
gentamicin	4692 (109)	571 (53)	97.0%	11.5%	0.84	0.989
kanamycin	3382 (23)	412 (67)	97.7%	12.3%	0.947	0.981
meropenem	609	0	100.0%	0.0%	NC	1
nalidixic acid	5294 (35)	15 (81)	97.9%	1.8%	0.3	0.985
streptomycin	3291 (254)	1084 (76)	93.9%	33.7%	0.877	0.977
sulfonamide	3763 (35)	1572 (55)	98.3%	30.0%	0.978	0.986
tetracycline	2558 (42)	2776 (49)	98.3%	52.1%	0.985	0.981

PPV = 0.94

NPV = 0.992

- For streptomycin, PPV is only 0.877; might result from lack of gene expression

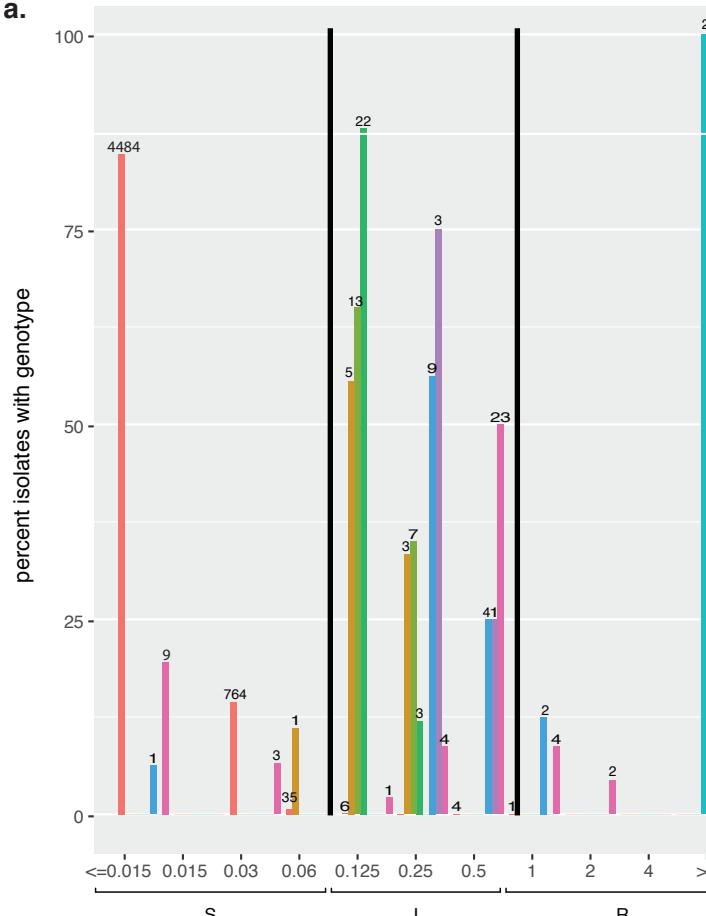
Multiple Beta-Lactamases Confer Decreased Susceptibility to Amoxicillin-Clavulanic Acid in *S. enterica*



- As expected, *blaCMY+* isolates were resistant
- Many beta-lactamases displayed either decreased or intermediate susceptibility
- 50/52** *blaCARB+* isolates displayed intermediate susceptibility
- 9/17** *blaHER+* isolates displayed intermediate susceptibility
- 5/15** *blaTEM+* isolates displayed intermediate susceptibility
- Observed in smaller sample by Tyson et al., 2017
- In *E. coli*, overexpression of TEM beta-lactamase is associated with clinical resistance (Di Conta et al., 2014)
- Similar phenomenon could be at work in *S. enterica*

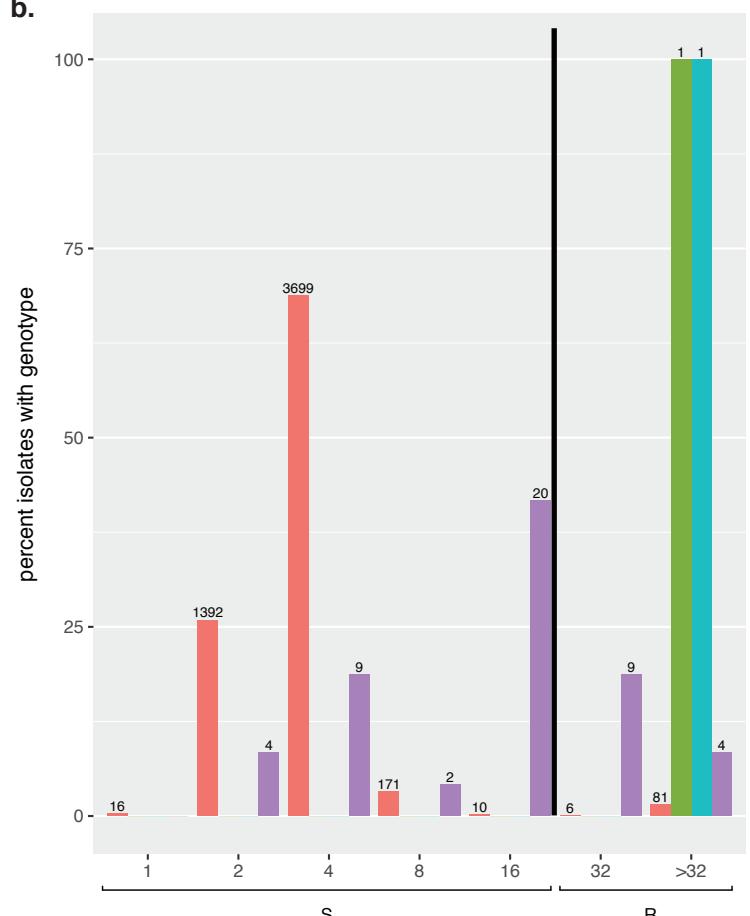
Resistance Markers for Fluoroquinolones Are Linked with Intermediate Susceptibility, but Not Clinical Resistance in *S. enterica*

a.



Ciprofloxacin MIC (mg/L)

b.



- Most resistant markers conferred intermediate ciprofloxacin susceptibility, not resistance
- *qnr*+ isolates displayed decreased susceptibility to nalidixic acid
- Most point mutations (79/81) conferred resistance to nalidixic acid

Lessons from the NARMS Analysis

- Overall, high consistency in all four species (N = 6,242)
- Inconsistencies in *S. enterica* were mostly due to streptomycin susceptibility and decreased susceptibility that fell below resistance thresholds
- Beta-lactamase producing *S. enterica* displayed decreased/intermediate susceptibility to amoxicillin-clavulanic acid
- Resfinder had a high number of possibly overspecified gene symbol assignments relative to AMRFinder
 - After databases were locked down for analysis, both have larger databases:
 - Resfinder: 2,254 → 3,037 sequences (35% increase)
 - AMRFinder: 3,921 → 4,579 sequences (17% increase)
- A hierarchical gene structure and high quality *protein* database are important for accurate gene calls

Antibiogram Submissions

- We have received 6,564 antibiograms from the following collaborators:

- Broad GSC/NIAID
- Brigham & Women's Hospital
- CVM/FDA
- Department of Defense
- CDRH
- JCVI
- CDC

[ncbi.nlm.nih.gov/biosample/docs/
antibiogram/](http://ncbi.nlm.nih.gov/biosample/docs/antibiogram/)

pd-help@ncbi.nlm.nih.gov

- Ability to search for isolates with antibiograms in Biosample

“antibiogram[filter]”

Antibiogram Submissions

- Fields designed to find balance between comprehensiveness and ease of submission
- Data dictionaries based on outside expertise (ASM, CLSI) standardize input and minimize ‘data drift’

Antibiogram

Antibiotic	Resistance phenotype	Measurement sign	Measurement	Measurement units	Laboratory typing method	Laboratory typing platform	Vendor	Laboratory typing method version or reagent	Testing standard
tetracycline	resistant	==	10	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI
chloramphenicol	resistant	==	12	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI
ampicillin	resistant	==	6	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI
gentamicin	susceptible	==	20	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI
colistin	susceptible	==	13	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI
sulfamethoxazole	intermediate	==	14	mm	disk diffusion	Vitek GM-NEG card	Biomérieux	missing	CLSI

[ncbi.nlm.nih.gov/biosample/
docs/antibiogram/](http://ncbi.nlm.nih.gov/biosample/docs/antibiogram/)
pd-help@ncbi.nlm.nih.gov

Antibiogram Submissions

- Need more!
- Useful for confirming patterns using other data
- Useful for tool development

Resources Available/show and tell

- AMR_genotypes:fosA7
 - AMR_genotypes:blaKPC* AND AMR_genotypes:mcr*
 - AST_phenotypes:*penem=R AND AMR_genotypes:blaKPC*
 - bioproject_acc:PRJNA316321
 - AMRFinder output
-
- <https://www.ncbi.nlm.nih.gov/pathogens/isolates>
 - Questions? pd-help@ncbi.nlm.nih.gov

Conclusions

- Curation underpins quality automation and standardization
- Protein sequence–function–matters and can be important for phenotypic prediction
- Still outstanding nomenclature problems that can inhibit communication and research
- Prediction in NARMS food safety-related isolates was very accurate
- We need more phenotypic data:
 - antibiograms
 - functional characterization
- Hierarchical structure of gene organization limits overspecification errors
 - ‘Beta-lactamases’ versus beta-lactamases
 - **If it’s new, investigate it!**

Future Directions and Coming Attractions

- Next release of AMRFinder will have point mutations for *Campylobacter* sp., *E. coli*, and *Salmonella* (29 genes covered)
- AMRFinder will describe what drug or drug classes should be affected by that gene or mutation
- AMRFinder will be expanded to include: biocides, stress, heat, virulence
- AMR Reference Gene Table
 - describes phenotypes of resistance genes
 - linkouts to isolates with resistance gene (in Isolate Browser)
- AMR Gene Browser
 - output similar to AMRFinder
 - searchable

Acknowledgements

NCBI

Richa Agarwala
Azat Badretdin
Slava Brover
Joshua Cherry
Vyacheslav Chetvernin
Robert Cohen
Michael DiCuccio
Daniel Haft
Arjun Prasad
Douglas Slotta
Edward Rice
Kirill Rotmistrovsky
Stephen Sherry
Sergey Shiryev
Martin Shumway
Tatiana Tatusova
Igor Tolstoy
Chunlin Xiao
Leonid Zaslavsky
Alexander Zasyplkin
Alejandro A. Schaffer
Lukas Wagner
Aleksandr Morgulis

William Klimke
Kim Pruitt
James Ostell

FDA/CVM

Patrick McDermott

Greg Tyson

Shaohua Zhao

CDC

Jason Folster

USDA

Glenn Tillman

Cesar Morales

Mustafa Simmon

Jamie Wasilenko

NIH/NIAID

US ARMY/MEDCOM

FDA/CFSAN

JCVI

Broad Institute

Brigham & Women's Hospital

Wadsworth Institute

pd-help@ncbi.nlm.nih.gov

ncbi.nlm.nih.gov/pathogens

- This research was supported by the Intramural Research Program of the NIH, National Library of Medicine. <http://www.ncbi.nlm.nih.gov>

National Center for Biotechnology Information – National Library of Medicine – Bethesda MD 20892 USA

NCBI Resources

AMRFinder is publicly available for use in your pipeline:

<https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/>

Curated AMR gene download:

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047>

AMR HMM download:

<https://ftp.ncbi.nlm.nih.gov/hmm/NCBIfam-AMRFinder/>

Table of AMR gene accessions and names:

ftp://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/Data/

Isolate Browser:

<https://www.ncbi.nlm.nih.gov/pathogens/isolates>

Questions: pd-help@ncbi.nlm.nih.gov