

AMR Resources at NCBI's Pathogen Portal

Michael Feldgarden

Michael.Feldgarden@nih.gov



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

NDARO

[Health](#) > [Pathogen Detection](#) > National Database of Antibiotic Resistant Organisms (NDARO)

Search page ^ v

National Database of Antibiotic Resistant Organisms (NDARO)

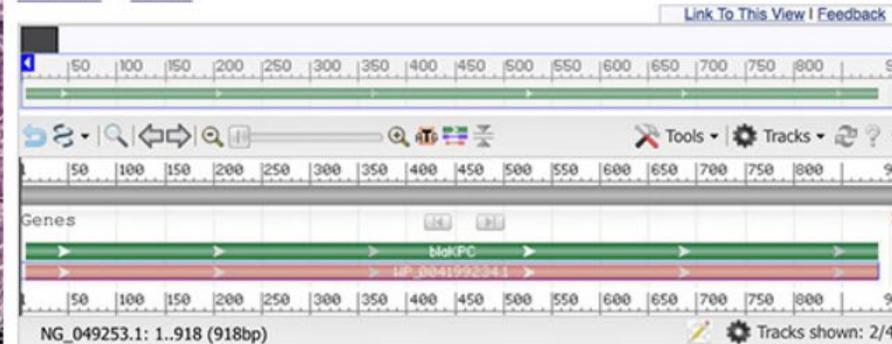
Welcome to the NCBI National Database of Antibiotic Resistant Organisms (NDARO), a collaborative, cross-agency, centralized hub for researchers to access AMR data to facilitate real-time surveillance of pathogenic organisms.



Klebsiella pneumoniae blaKPC gene for carbapenem-hydrolyzing class A beta-lactamase KPC-2, complete CDS

NCBI Reference Sequence: NG_049253.1

[GenBank](#) [FASTA](#)



From left to right: Multi-drug resistant *Salmonella enterica*, kpc2 carbapenem resistance gene

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: Combined BLAST/HMM AMR gene discovery

Curated AMR gene sources:

Domain experts

Large scale databases

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

Manual extraction from literature

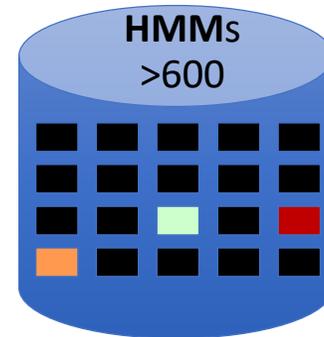
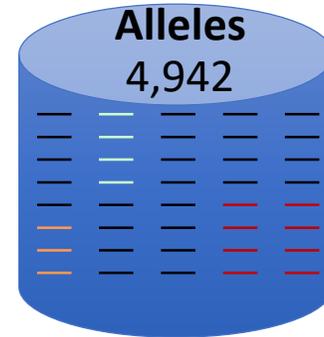
Curated HMM sources:

ResFams

Select
Set cutoffs

New NCBI-built HMMs

Group sequences
Align
Build HMM
Set cutoffs



Validation



Isolate Browser
Presence of known AMR genes can be visualized and downloaded



Surveillance alerts
Identified colistin resistant *E. coli* without traditional phenotyping (Vasquez et al., 2016)



AMR gene discovery
Identified novel plasmid-borne fosfomycin gene (Rehman et al., 2017)



PGAP Annotations
Standardized annotation for all researchers

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

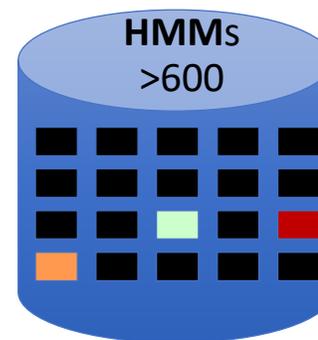
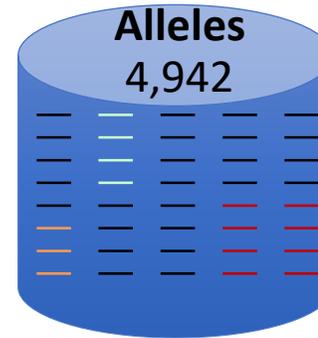
Ongoing curation of beta-lactamases,
Qnr, and MCR

ResFams

Select
Set cutoffs

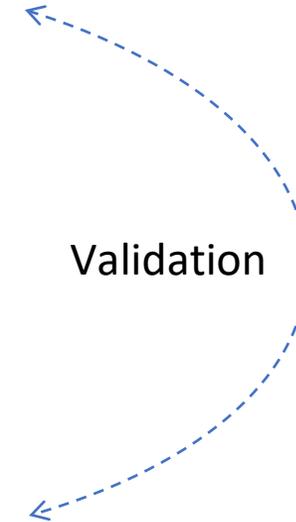
New HMMs

Group sequences
Align
Build HMM
Set cutoffs



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

Validation



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**

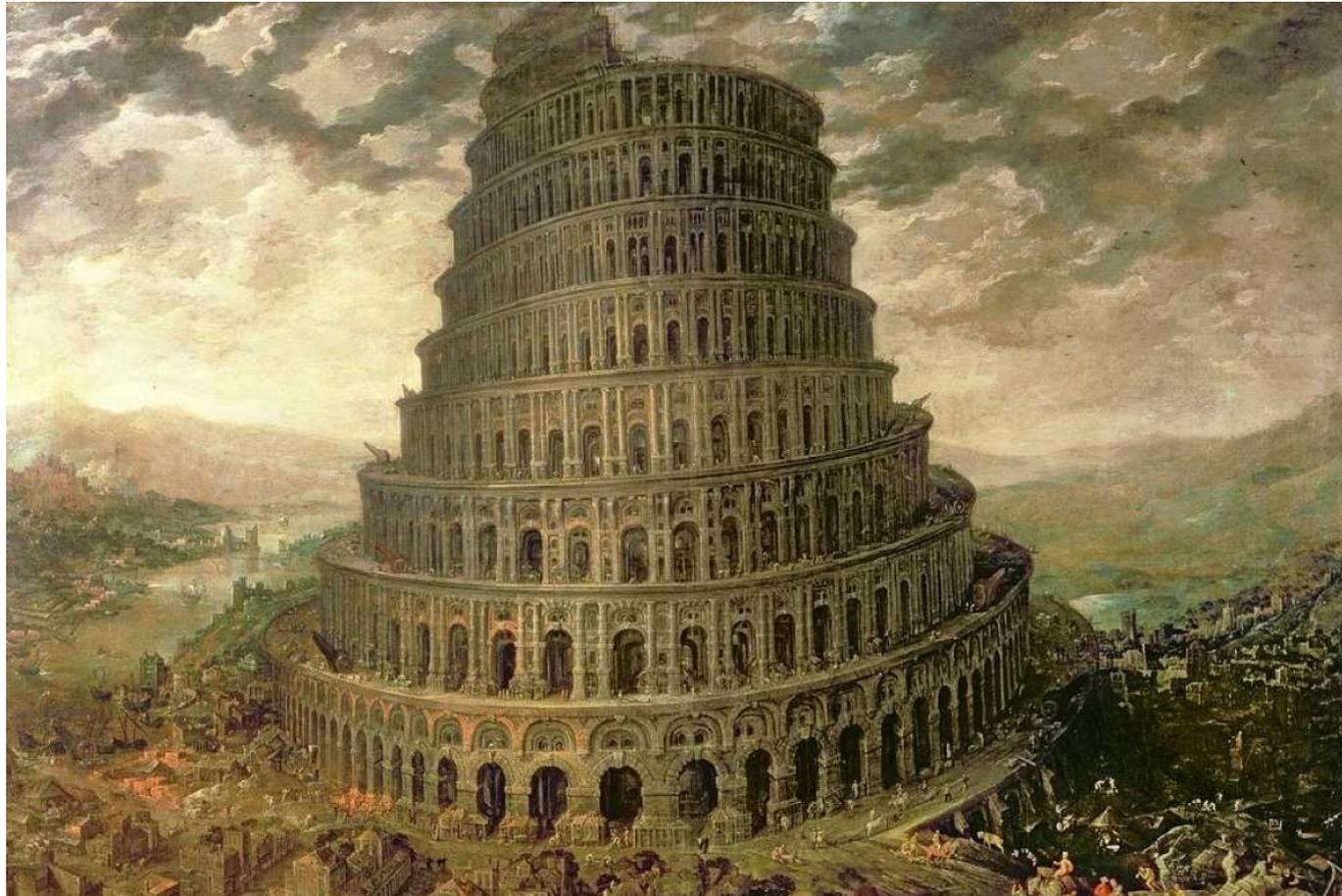
<http://ncbi.nlm.nih.gov/bioproject/PRJNA313047>

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 (Partridge *et al.*, 2018)

J Antimicrob Chemother
doi:10.1093/jac/dky262

Journal of
Antimicrobial
Chemotherapy

pd-help@ncbi.nlm.nih.gov

Proposal for assignment of allele numbers for mobile colistin resistance (*mcr*) genes

Sally R. Partridge^{1*}, Vincenzo Di Pilato², Yohei Doi³, Michael Feldgarden⁴, Daniel H. Haft⁴, William Klimke⁴, Samir Kumar-Singh⁵, Jian-Hua Liu⁶, Surbhi Malhotra-Kumar⁷, Arjun Prasad⁴, Gian Maria Rossolini^{2,8}, Stefan Schwarz⁹, Jianzhong Shen¹⁰, Timothy Walsh¹¹, Yang Wang¹⁰ and Basil Britto Xavier⁷

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

The Joy of Being Allele Curators

- NCBI, at request of experts, is assigning alleles for the following genes:
 - Many beta-lactamases ('Lahey')
 - Qnr ('Lahey')
 - MCR
- Submitting an allele to GenBank with an incorrect name creates many problems
 - People might use the GenBank record with incorrect nomenclature, not RefSeq curated nomenclature
 - Collisions spoil the literature
- Editors, reviewers, and authors should check allele assignments with NCBI (and for other AMR genes with the assigning authority)
- Need more groups to standardize nomenclature for other genes
- contact **pd-help@ncbi.nlm.nih.gov**
 - CARD github site

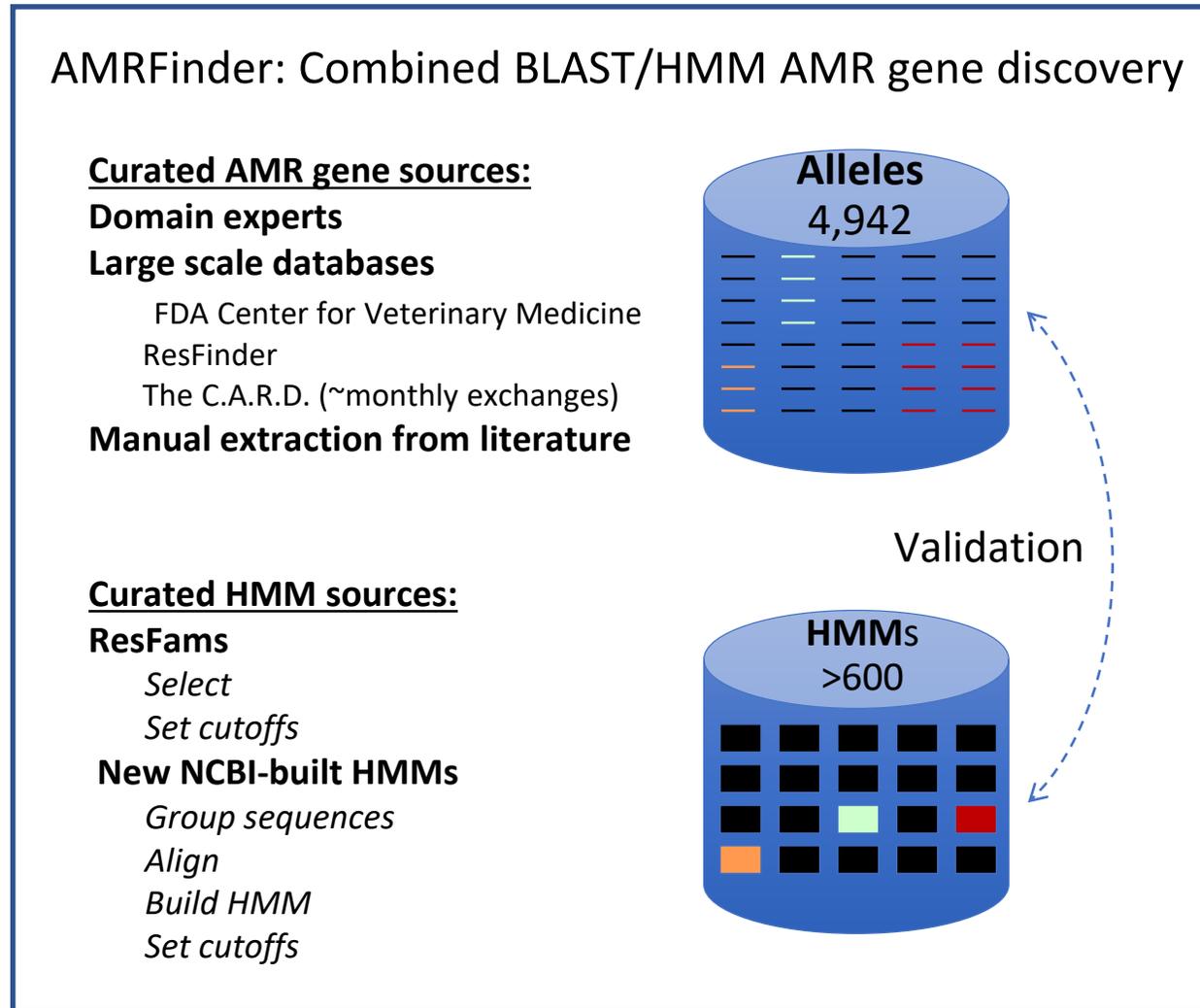
AMRFinder: Automating AMR Gene Detection

Pathogen Detection Pipeline (SRA):

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

Genbank assemblies:

General submissions



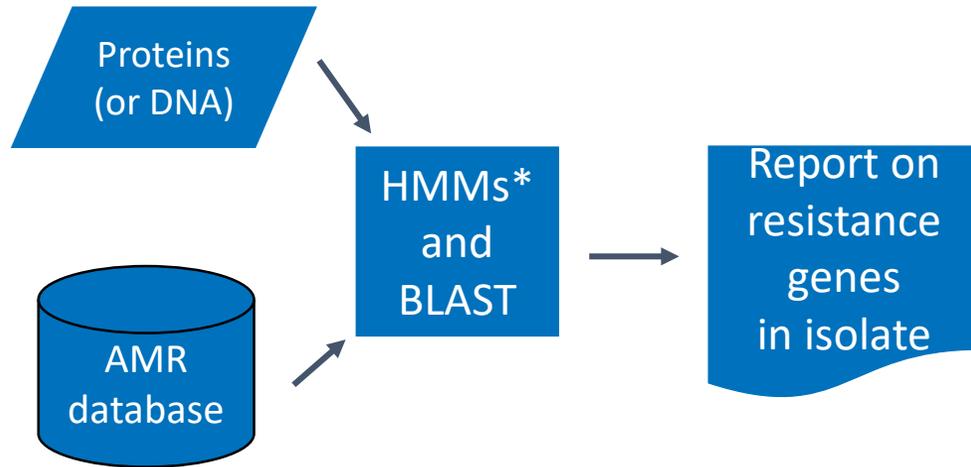
Isolate Browser
Presence of known AMR genes can be visualized and downloaded

Surveillance alerts
Identified colistin resistant *E. coli* without traditional phenotyping (Vasquez et al., 2016)

AMR gene discovery
Identified novel plasmid-borne fosfomycin gene (Rehman et al., 2017)

PGAP Annotations
Standardized annotation for all researchers

AMRFinder Overview



4,942 resistance proteins
>600 HMMs
34 drug classes resisted
~50% beta-lactamases

Hierarchical Structure

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

- 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)

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

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 2.0 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
 - **When compared to ResFinder 3.1, found equidistant hits—1 correct answer, multiple incorrect answers
- 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

Lessons from the NARMS Analysis

- Overall, high consistency in all four species (N = 6,242)
- Resfinder had a high number of possibly overspecified gene symbol assignments relative to AMRFinder
- A hierarchical gene structure and high quality *protein* database are important for accurate gene calls
- **Closest hit does not necessarily yield the correct AMR protein sequence**

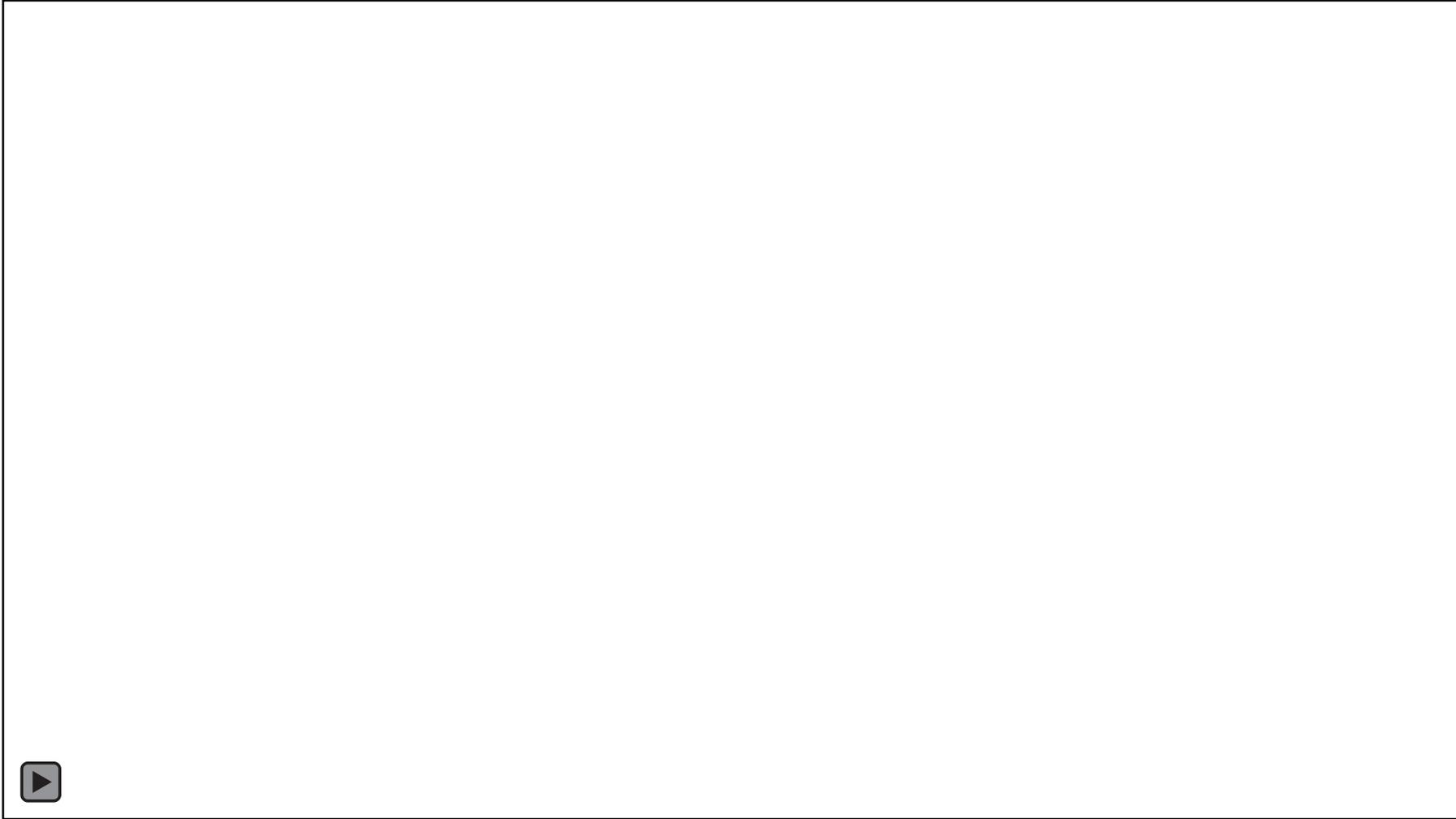
Finding KPC⁺/MCR⁺ isolates



Carbapenem sensitive, KPC-carrying isolates



Reference Gene Catalog



AMRFinderPlus

Protein identifier	Contig id	Start	Stop	Strand	Gene symbol	Sequence name	Scope	Element type	Element subtype	Class	Subclass
WP_000623108.1	NZ_CP022730.1	1035643	1037370	+	ptsI_V25I	Escherichia f	core	AMR	POINT	FOSFOMYCIN	FOSFOMYCIN
WP_000879199.1	NZ_CP022730.1	2471924	2473315	-	uhpT_E350Q	Escherichia f	core	AMR	POINT	FOSFOMYCIN	FOSFOMYCIN
WP_001296671.1	NZ_CP022730.1	3058367	3059500	-	blaEC	BlaEC family	core	AMR	AMR	BETA-LACTAM	BETA-LACTAM
WP_001281219.1	NZ_CP022730.1	845680	848307	-	gyrA_S83L	Escherichia q	core	AMR	POINT	QUINOLONE	QUINOLONE
WP_000027057.1	NZ_CP022732.1	27239	28099	-	blaTEM-1	class A broad	core	AMR	AMR	BETA-LACTAM	BETA-LACTAM
WP_000557454.1	NZ_CP022732.1	33071	33931	-	aac(3)-IIId	aminoglycosi	core	AMR	AMR	AMINOGLYCOSIDE	GENTAMICIN
WP_004201280.1	NZ_CP022735.1	200097	200570	+	dfrA14	trimethoprim	core	AMR	AMR	TRIMETHOPRIM	TRIMETHOPRIM
WP_000219391.1	NZ_CP022735.1	204276	205181	-	mph(A)	Mph(A) famil	core	AMR	AMR	MACROLIDE	MACROLIDE
WP_000027057.1	NZ_CP022735.1	206592	207452	+	blaTEM-1	class A broad	core	AMR	AMR	BETA-LACTAM	BETA-LACTAM
WP_001261740.1	NZ_CP022735.1	209673	210464	+	aadA2	ANT(3'')-Ia f	core	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN
WP_095866700.1	NZ_CP022735.1	213830	214621	+	sul3	sulfonamide-	core	AMR	AMR	SULFONAMIDE	SULFONAMIDE
WP_000800531.1	NZ_CP022735.1	215801	216133	-	qacl	quaternary a	core	AMR	AMR	QUATERNARY AMMONIUM	QUATERNARY AMMONIUM
WP_001206316.1	NZ_CP022735.1	216303	217094	-	aadA1	ANT(3'')-Ia f	core	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN
NA	NZ_CP022735.1	217190	218188	-	cmlA1	chlorampher	core	AMR	AMR	PHENICOL	CHLORAMPHENICOL
WP_000018329.1	NZ_CP022735.1	219098	219913	+	aph(3')-Ia	aminoglycosi	core	AMR	AMR	AMINOGLYCOSIDE	KANAMYCIN
WP_001082319.1	NZ_CP022735.1	222139	222942	+	aph(3'')-Ib	aminoglycosi	core	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN
WP_000480968.1	NZ_CP022735.1	222942	223778	+	aph(6)-Id	aminoglycosi	core	AMR	AMR	AMINOGLYCOSIDE	STREPTOMYCIN
WP_062896568.1	NZ_CP022735.1	225957	227171	+	floR	chlorampher	core	AMR	AMR	PHENICOL	CHLORAMPHENICOL/FLORFENICOL
WP_000804064.1	NZ_CP022735.1	232060	233259	+	tet(A)	tetracycline	core	AMR	AMR	TETRACYCLINE	TETRACYCLINE
WP_049589868.1	NZ_CP022735.1	41127	42752	-	mcr-1.1	phosphoetha	core	AMR	AMR	COLISTIN	COLISTIN

AMRFinderPlus

Method	Target length	Reference seq	% Coverage c	% Identity to	Alignment len	Accession of closest sequence	Name of closest sequence	HMM id	HMM
POINTP	575	575	100	99.65	575	WP_000623140.1	ptsI	NA	NA
POINTP	463	463	100	99.78	463	WP_000879194.1	uhpT	NA	NA
BLASTP	377	377	100	98.94	377	WP_001443153.1	cephalosporin-hydrolyzing class C beta-lactamase EC-5	NF000185.2	BlaEC
POINTP	875	875	100	99.77	875	WP_001281242.1	gyrA	NA	NA
ALLELEP	286	286	100	100	286	WP_000027057.1	class A broad-spectrum beta-lactamase TEM-1	NF000531.2	TEM f
ALLELEP	286	286	100	100	286	WP_000557454.1	aminoglycoside N-acetyltransferase AAC(3)-IId	NF033080.0	AAC(3
EXACTP	157	157	100	100	157	WP_004201280.1	trimethoprim-resistant dihydrofolate reductase DfrA14	NF000330.1	trimet
EXACTP	301	301	100	100	301	WP_000219391.1	Mph(A) family macrolide 2'-phosphotransferase	NF000254.1	Mph(/
ALLELEP	286	286	100	100	286	WP_000027057.1	class A broad-spectrum beta-lactamase TEM-1	NF000531.2	TEM f
EXACTP	263	263	100	100	263	WP_001261740.1	ANT(3'')-Ia family aminoglycoside nucleotidyltransferase AadA2	NF012157.0	ANT(3
BLASTP	263	263	100	99.62	263	WP_000034420.1	sulfonamide-resistant dihydropteroate synthase Sul3	NF000296.1	sulfon
EXACTP	110	110	100	100	110	WP_000800531.1	quaternary ammonium compound efflux SMR transporter QacL	NF000067.2	QacF/
EXACTP	263	263	100	100	263	WP_001206316.1	ANT(3'')-Ia family aminoglycoside nucleotidyltransferase AadA1	NF033126.1	ANT(3
PARTIALX	333	419	79.47	100	333	WP_000095725.1	chloramphenicol efflux MFS transporter CmlA1	NF000509.1	CmlA
EXACTP	271	271	100	100	271	WP_000018329.1	aminoglycoside O-phosphotransferase APH(3'')-Ia	NF033059.2	APH(3
EXACTP	267	267	100	100	267	WP_001082319.1	aminoglycoside O-phosphotransferase APH(3'')-Ib	NF032895.1	aminc
EXACTP	278	278	100	100	278	WP_000480968.1	aminoglycoside O-phosphotransferase APH(6)-Id	NF012171.0	APH(6
BLASTP	404	404	100	99.5	404	WP_000214119.1	chloramphenicol/florfenicol efflux MFS transporter FloR	NF000219.1	chlora
EXACTP	399	399	100	100	399	WP_000804064.1	tetracycline efflux MFS transporter Tet(A)	NF012193.0	tetrac
ALLELEP	541	541	100	100	541	WP_049589868.1	phosphoethanolamine--lipid A transferase MCR-1.1	NF000465.1	MCR-

Conclusions

- Curation underpins quality automation and standardization
- Still outstanding nomenclature problems that can inhibit communication and research
- Prediction in NARMS food safety-related isolates was very accurate
- Hierarchical structure of gene organization limits overspecification errors
- HMMs versus arbitrary identity thresholds
 - ‘Beta-lactamases’ versus beta-lactamases
 - **If it’s new, investigate it!**

Future Directions and Coming Attractions

- AMRFinderPlus (*just released*):
 - New release of AMRFinder has point mutations for *Campylobacter* sp., *E. coli*, and *Salmonella* (29 genes covered)
 - AMRFinderPlus output describes what drug or drug classes could be affected by that gene or mutation
- AMRFinderPlus optionally can identify: biocide, stress, heat, virulence
- Pathogen Detection Reference 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 Zasytkin
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



Ask Us
for a full schedule
or posters & slides

The NCBI team would like to learn more about workflows and **user needs for large scale analyses of microbial genomes** to support our users better.

We would be especially interested in anyone interested in doing **large-scale data analyses using cloud-based resources**.

If you are interested in participating and would be willing to **share your email address, contact us at info@ncbi.nlm.nih.gov**.

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:

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

Isolate Browser:

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

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

movie

