

DATABASES, RESOURCES AND TOOLS  
FOR ANTIMICROBIAL RESEARCH

---

# Comparing and reporting AMR results using hAMRonization

---

Inês Mendes - University of Lisbon  
15th October 2021 - AMR Workshop



Public Health Alliance for  
Genomic Epidemiology



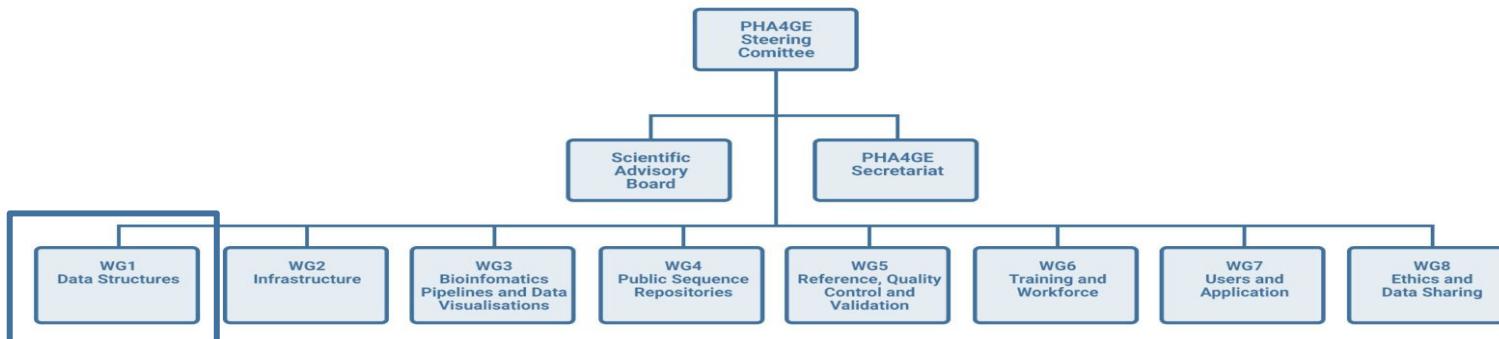
CLIMB  
BIG DATA



## Public Health Alliance for Genomic Epidemiology

<https://pha4ge.org/>

The **Public Health Alliance for Genomic Epidemiology (PHA4GE)** is a global coalition that is actively working to establish consensus standards; document and share best practices; improve the availability of critical bioinformatic tools and resources; and advocate for **greater openness, interoperability, accessibility and reproducibility in public health microbial bioinformatics.**





**Dr Emma Griffiths**  
*Chair*

## Main

## Goal:

Develop, share, promote **data standards and standardized data** models to improve data harmonization, integration, transformation, sharing, and control/security.

### Areas of Interest:

Metadata standards, ontologies and conventions; Contextual data harmonization and sharing; Data inputs/outputs, APIs and interoperability; Result reporting and views; Data Security and Encryption; Identity management for role/resource based access

### Current projects:

- SARS-CoV-2 Contextual Data Specification
- Gene Detection/AMR Output Specification

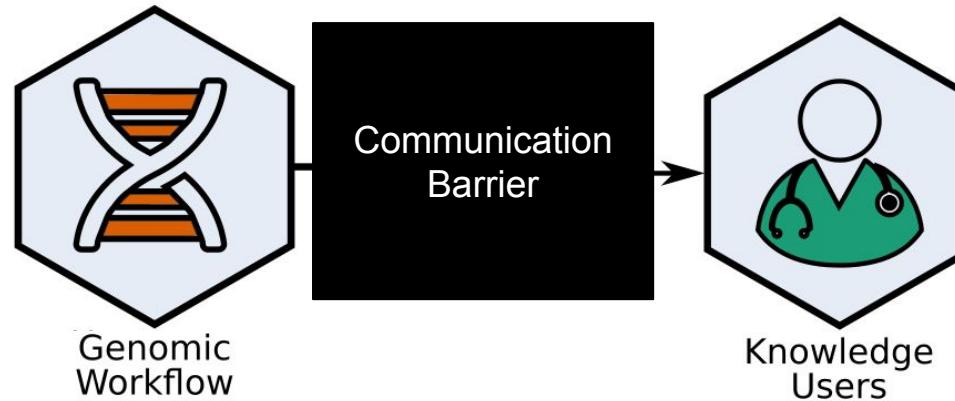
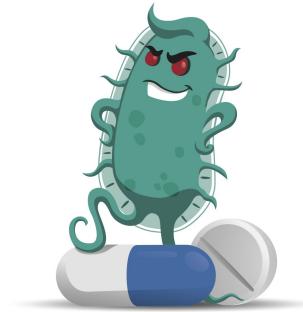


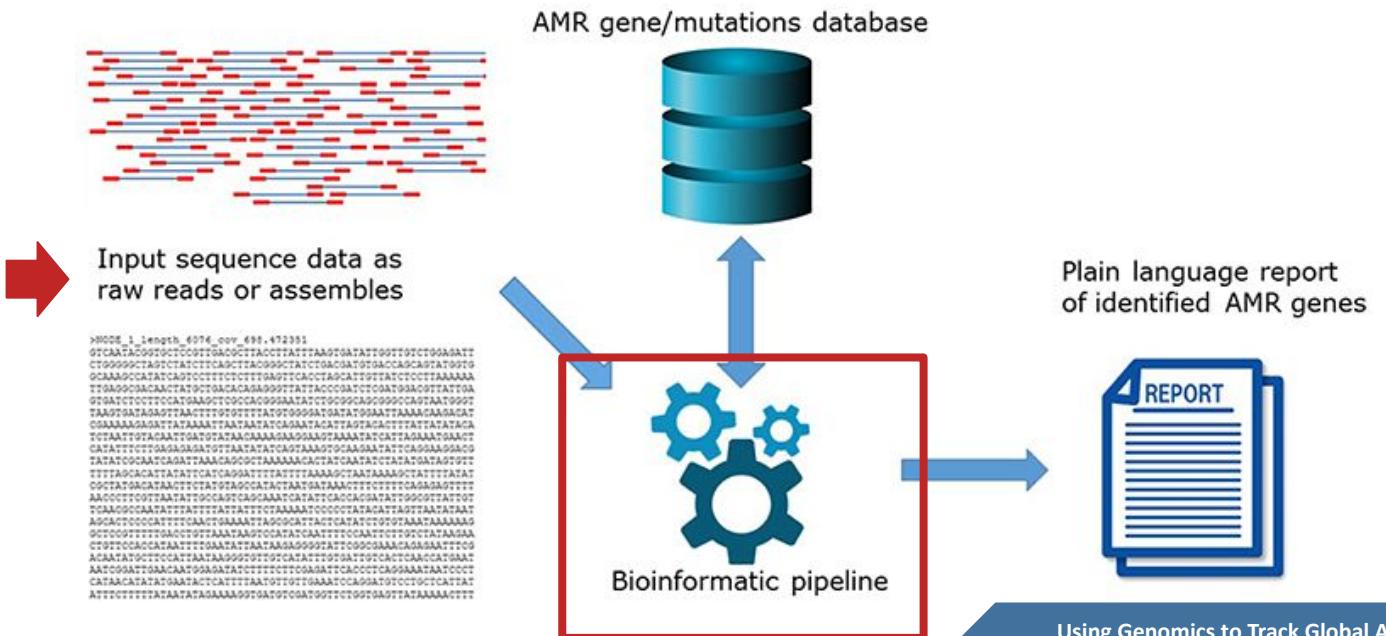
# The Antimicrobial Resistance Data Problem



Antimicrobial resistance (**AMR**) represents a growing public health crisis of global scope and relies on the **detection of AMR genes from genomic data**.

A large number of **different bioinformatic tools** have been developed to perform this task but **differ in supported inputs**, search algorithm, parameterisation, underlying reference databases and **output formats**.





## Using Genomics to Track Global Antimicrobial Resistance

Hendriksen et al, Front. Public Health, 2019



abricate

tseemann/abricate

\$ head abricate/report.tsv

#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE_MAP	GAPS	%COVERAGE	%IDENTITY	DATABASE	ACCESSION	PRODUCT	RESISTANCE
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039729.1	656118	657620	+	eat(A)	1-1503/1503	=====	0/0	100.00	100.00	ncbi	NG_047762.1	ABC-F type	
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039729.1	2163439	2163987	-	aacA-ENT1	1-549/549	=====	0/0	100.00	100.00	ncbi	NG_052371.1		
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039729.1	2564038	2565516	-	msr(C)	1-1479/1479	=====	0/0	98.92	ncbi	NG_048003.1	ABC-F		
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039730.1	43619	44197	-	vanX-M	1-579/609	=====	0/0	95.07	100.00	ncbi	NG_048495.1	D-Ala-D	
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039730.1	44203	45234	-	vanM	1-1032/1032	=====	0/0	100.00	100.00	ncbi	NG_048396.1	D-alanine	
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039730.1	45227	46207	-	vanH-M	1-981/981	=====	0/0	100.00	100.00	ncbi	NG_048388.1	D-lactate dehydrogenase VanH-M	VANCOMYCIN
GCF_010120755.1_ASM1012075v1_genomic.fna	NZ_CP039730.1	47306	48418	-	vanS-M	1-1113/1113	=====	0/0	100.00	100.00	ncbi	NG_048450.1	VanM-type vancomycin resistance histidine kinase VanS	VANCOMYCIN

\$ head amrfinder/report.nucleotide.tsv

Protein identifier	Contig id	Start	Stop	Strand	Gene symbol	Sequence name	Scope	Element type	Element subtype	Class	Subclass	Method	Target length	Reference sequence length	% Coverage of reference sequence	%		
Identity to reference sequence	Alignment length	Accession of closest sequence	Name of closest sequence						HMM id	HMM description								
NA	NZ_CP039729.1	6208	8676	+	gyrA_S83I	Enterococcus faecium quinolone resistant Gyra	core	AMR	AMR	PLEUROMUTILIN	PLEUROMUTILIN	EXACTX	500	500	100.00	100.00	500	
Gyra	NA	NA													WP_137277832.1	DNA gyrase subunit A		
NA	NZ_CP039729.1	656118	657617	+	eat(A)	ABC-F type ribosomal protection protein Eat(A)	core	AMR	AMR						WP_002296175.1	ABC-F type ribosomal protect		
ion protein Eat(A)	NA	NA																
NA	NZ_CP039729.1	1072195	1074642	+	parC_S80I	Enterococcus faecium quinolone resistant ParC	core	AMR	AMR	POINT	QUINOLONE	QUINOLONE	POINTX	816	816	100.00	100.00	100
subunit A ParC NA	NA																	
NA	NZ_CP039729.1	2163442	2163987	-	aac(6')-I	aminoglycoside 6'-N-acetyltransferase	core	AMR	AMR	AMINOGLYCOside	AMINOGLYCOside	EXACTX	182	182	100.00	100.00	100	
ansferase	NA	NA																
NA	NZ_CP039729.1	2564041	2565516	-	msr(C)	ABC-F type ribosomal protection protein Msr(C)	core	AMR	AMR	MACROLIDE	MACROLIDE	BLASTX	492	492	100.00	100.00	100	
ion protein Msr(C)	NA	NA																
NA	NZ_CP039730.1	43529	44197	-	vanX-M	D-Ala-D-Ala dipeptidase VanX-M	core	AMR	AMR	GLYCOPEPTIDE	VANCOMYCIN	INTERNAL_STOP	202	202	100.00	97.03	202	
A	NA																	
NA	NZ_CP039730.1	44206	45234	-	vanM	D-alanine--(R)-lactate ligase VanM	core	AMR	AMR	GLYCOPEPTIDE	VANCOMYCIN	EXACTX	343	343	100.00	100.00	343	
A	NA														WP_063856777.1	D-Ala-D-Ala dipeptidase VanX-M NA		
NA	NZ_CP039730.1	45230	46207	-	vanH-M	D-lactate dehydrogenase VanH-M	core	AMR	AMR	GLYCOPEPTIDE	VANCOMYCIN	EXACTX	326	326	100.00	100.00	100	
NA	NZ_CP039730.1	47309	48418	-	vanS	VanM-type vancomycin resistance histidine kinase VanS	core	AMR	AMR	GLYCOPEPTIDE	VANCOMYCIN	EXACTX	370	370	100.00	100.00	370	
resistance histidine kinase VanS	NA	NA													WP_063856748.1	VanM-type vancomycin		

\$ head ariba/report.tsv

#ariba_ref_name	ref_name	gene	var_only	flag	reads	cluster	ref_len	ref_base_assembled	pc_ident	ctg	ctg_len	ctg_cov	known_var	var_type	var_seq_type	known_var_change	has_known_var	ref_ctg_chan	
ge	ctg_eff	ref_start	ref_end	ref_nt	ctg_start	ctg_end	ctg_nt	smtls_total_depth	smtls_nts	smtls_nts	smtls_nts	smtls_nts	var_description	free_text					
AAC_3_IId	3004263_EU022314.1_0..861..5344	AAC_3_IId	3004263_EU022314.1_0..861..5344	1	0	27	256	AAC_3_IId	861	861	861	861	99.88	AAC_3_IId..1..115.c4.ctg.1	1547	37.8	.	.	
.	.	.	.	.	b'AAC(3)-IId	is a plasmid-encoded aminoglycoside acetyltransferase in E. coli';AAC(3)-IId													
AAC_6_Ib4	3002577_AF445082_2788..3343.401	AAC_6_Ib4	3002577_AF445082_2788..3343.401	1	0	667	510	AAC_6_Ib4..	555	555	555	555	99.64	AAC_6_Ib4..1..115.c30.ctg.1	619	128.1	0	p .. 0 W	
87R	NONSYN	259	261	TGG	264	266	CGG	86;89;90	C,T,A;G;6	30;28;1;59;63	.	.	AAC(6')-Ib4;b'AAC(6')-Ib4	is an aminoglycoside acetyltransferase in Serratia spp.'					
AAC_6_Ib4	3002577_AF445082_2788..3343.401	AAC_6_Ib4	3002577_AF445082_2788..3343.401	1	0	667	510	AAC_6_Ib4..	555	555	555	555	99.64	AAC_6_Ib4..1..115.c30.ctg.1	619	128.1	0	p .. 0 S	
102L	NONSYN	304	306	TCA	309	311	TTA	92;92;92	T;T;A	70;70;71	.	.	AAC(6')-Ib4;b'AAC(6')-Ib4	is an aminoglycoside acetyltransferase in Serratia spp.'					
AAC_6_Ib4	3002577_AF445082_2788..3343.401	AAC_6_Ib4	3002577_AF445082_2788..3343.401	1	0	667	510	AAC_6_Ib4..	555	555	555	555	99.64	AAC_6_Ib4..1..115.c30.ctg.1	619	128.1	0	p .. 0 S	
490T	.	490	490	G	495	495	G	139	G,T	57;46	.	.	AAC(6')-Ib4;b'AAC(6')-Ib4	is an aminoglycoside acetyltransferase in Serratia spp.'					
BRP_MBL_3001205_KC503911.1..9883..10249..5446	BRP_MBL_3001205_KC503911.1..9883..10249..5446	BRP_MBL_3001205_KC503911.1..9883..10249..5446	1	0	27	132	BRP_MBL	366	366	366	366	100.0	BRP_MBL..1..16.c4.ctg.1	BRP(MBL);A novel bleomycin resistance protein encoded by a metallo-beta-lactamase-associated ble gene. Express					
n-like antibiotics	in Enterobacteriaceae and Acinetobacter, where it is co-expressed with an MBL and controlled by the same promoter region.'																		
CRP_3000518_AP009048.1..4153663..4154296..4752	CRP_3000518_AP009048.1..4153663..4154296..4752	CRP_3000518_AP009048.1..4153663..4154296..4752	1	0	528	58	CRP	633	146	93.84	CRP..1..15.c4.ctg.1	477	24.1	0	p .. 0 K29T	NON5			
YN	85	87	AAAG	414	416	ACG	14;14;13	C;G;6	10;10;10	.	.	.	CRP is a global regulator that represses MdtF multidrug efflux pump expression.';CRP						
CTX_M_15_3001878_AY044436_1435_2311_1430	CTX_M_15_3001878_AY044436_1435_2311_1430	CTX_M_15_3001878_AY044436_1435_2311_1430	1	0	27	274	CTX_M_-	876	876	100.0	CTX_M_-..1..15.c4.ctg.1	1793	36.0	.	.	.	.		
DHA_7_3002136_HQ456945_0..1140..1865	DHA_7_3002136_HQ456945_0..1140..1865	DHA_7_3002136_HQ456945_0..1140..1865	1	0	16	170	DHA_-	1140	796	100.0	DHA_..1..15.c4.ctg.1	1350	29.8	.	.	.	.		
E	Escherichia_coli_23S_3004173_AE014075..1..237159..240063..4164	Escherichia_coli_23S_3004173_AE014075..1..237159..240063..4164	0	0	659	8302	Escherichia_coli_23S	2904	2904	97.1	Escherichia_coli_23S.l15.c17.ctg.1	2822							
611.4	0	0	T72A	SNP	72	72	T	86	86	A	506	125	.	Escherichia coli 23S rRNA with mutation conferring resistance to oxazolidinone antibiotics					

\$ head srst2/SAMN13064234\_srst2\_report.tsv

Sample	DB	gene	allele	coverage	depth	diffs	uncertainty	divergence	length	maxMAF	clusterid	seqid	annotation					
SRR10313716	ResFinder	cat	catA1	100.0	75.852	15sp	0.152	660	0.037	470	1995	catA1_1..V00622..V00622;	chloramphenicol					
SRR10313716	ResFinder	dfrA	dfrA32	92.616	25.914	42spn1indel34holes	9.545	474	0.027	593	138	dfrA32_1..G0067642..G0067642;	trimethoprim					
SRR10313716	ResFinder	strA	strA4	100.0	74.021	0.0	804	0.029	325	1142	strA_4..AF321551..AF321551;	aminoglycosides						
SRR10313716	ResFinder	aac(6')-I	aac(6')-Iy	100.0	39.246	10spn	2.283	438	0.036	621	1001							
SRR10313716	ResFinder	dfrA	dfrA17	99.578	33.08	40spn2indel	8.439	474	0.167	589	902	dfrA17_1..FJ460238..FJ460238;	trimethoprim					
SRR10313716	ResFinder	su1	su1_11	92.857	62.65	63holes	edge0.0	0.0	882	0.034	227	1889	su1_11..DQ914960..DQ914960;	sulphonamides				
SRR10313716	ResFinder	su1	su1_2	100.0	68.151	trun	0.0	927	0.029	183	1044	su1_2..CP002151..CP002151;	sulphonamides					
SRR10313716	ResFinder	blaTEM	blaTEM-1_1	100.0	117.002	0.0	861	0.032	258	1394	blaTEM-1_1..JF910132..JF910132;	beta lactamase						
SRR10313716	ResFinder	Qnr-S	QnrS2	97.869	79.482	58spn2indel13holes	9.006	657	0.1	476	1997	Qnr-S_2_1..JF261185..JF261185;	quinolone					



srst2

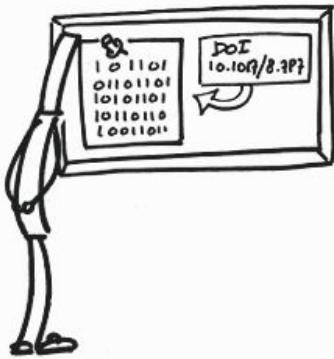
katholt/srst2

## FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE

HOW DO YOU  
OPEN A .X3D FILE?



INTEROPERABLE

HERE!



REUSABLE

The FAIR Guiding Principles for scientific data management and stewardship  
Wilkinson et al, Scientific Data, 2016

# 1.

The **lack of standardization** in the reporting of **AMR gene detection** greatly hinders the comparison of results across the public health sector. The myriad of options available for this purpose highlights **a grave interoperability problem**.

# Standardized Output Specification

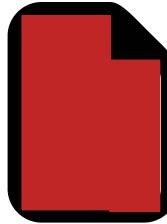
Data **specification** for the detection of AMR genes based on **harmonisable, publicly available, community standards** implementable via a **report template** with a **standardized list of labels and terms** and respective **ontology identifiers**.

## Mandatory terms:

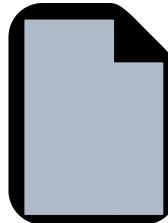
- Input File Name
- Gene Symbol
- Gene Name
- Sequence Identity
- **Reference Database ID**
- **Reference Database Version**
- Reference Accession
- Reference Sequence Length
- Target Sequence Length
- **Analysis Software Name**
- **Analysis Software Version**



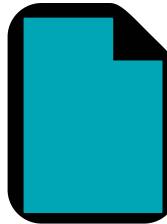
# Standardized Output Specification



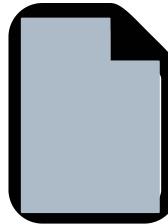
Report of  
tool X



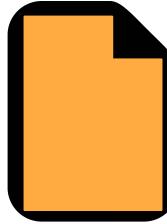
Standard Report  
of tool X



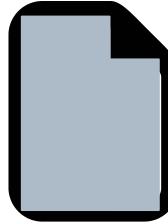
Report of  
tool Y



Standard Report  
of tool Y



Report of  
tool Z



Standard Report  
of tool Z

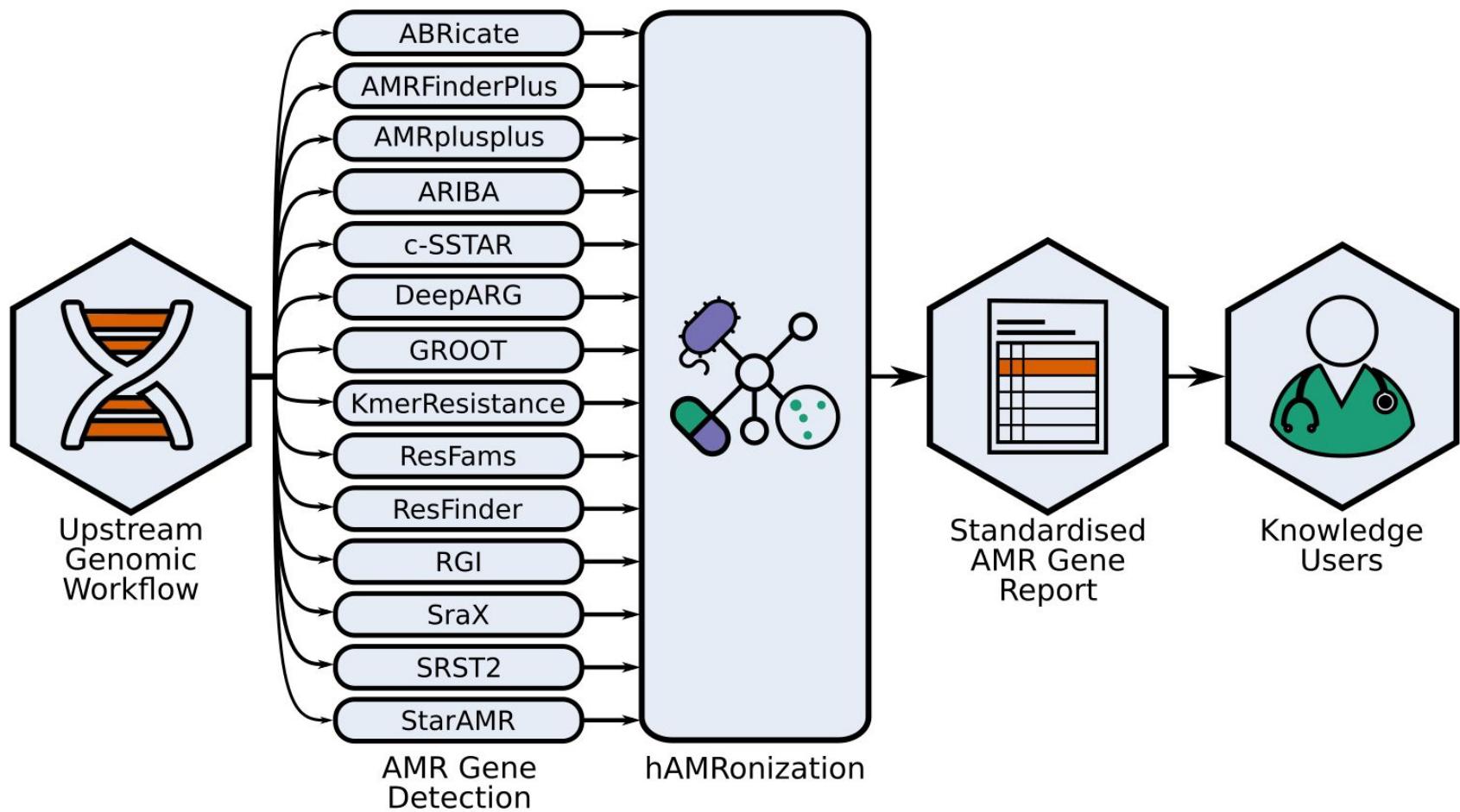


# Standardized AMR Gene Detection Output



AMR Prediction Tools	hAMRonization Workflow	hAMRonization Package	Standardized Output Specification
<i>Lots of software with different output formats.</i>	<i>Run any dataset through the set of tools.</i>	<i>Take tool-specific output and reformat to standardized output.</i>	<i>Enables consistent comparison of AMR genes detected by different tools.</i>

**The hAMRonization package automates conversion to a standardized output where tools can be built on.**



# 2.

We developed a **standardized data specification** to improve data harmonization and interoperability, implemented in a **Biopython-compatible parser** and **command-line utility**.

# The hAMRonization Workflow



[https://github.com/pha4ge/hAMRonization\\_workflow](https://github.com/pha4ge/hAMRonization_workflow)

**Snakemake workflow** to run **14 different species-agnostic AMR gene detection tools** selected tools on a dataset, installing **fixed versions** of the tools from **conda** on execution.

## Ensures:

- Reproducibility
- Scalability

## Tool inclusion criteria:

- Specificity for AMR gene detection
- Open-source



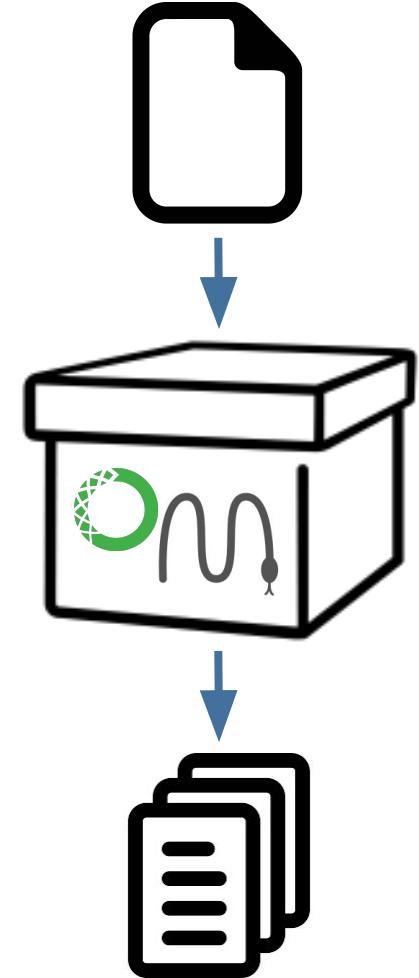
Dr Finlay Maguire



Dr Adam Witney



Dr Simon Tausch



# The hAMRonization Package



<https://github.com/pha4ge/hAMRonization>

**Biopython-compatible parser** and **command-line utility** automatically transformation of reports from 14 different species-agnostic AMR gene detection tools into “hAMRonization”-compatible reports.

**Validation** and programmatic use of the **specification** was facilitated via the development of **JSON** and **SALAD** schemata. **Validation** of the parsing utilities ensured by unit testing.



Dr Finlay Maguire



Inês Mendes



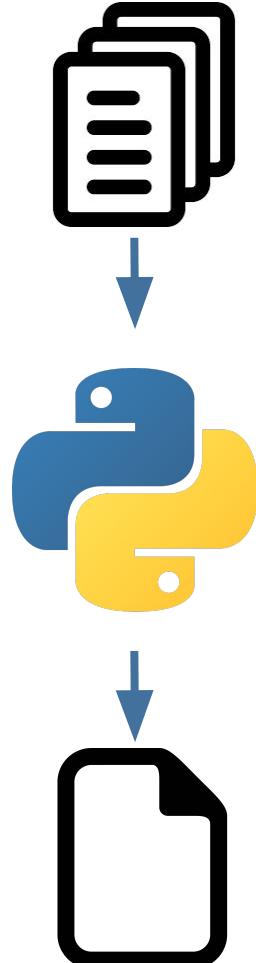
Alex Manuele



Amos Raphenya



And Counting!



# The hAMRonization Package



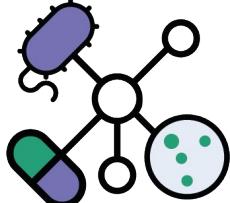
<https://github.com/pha4ge/hAMRonization>

An example from abricate:

```
> hamronize abricate <report> --reference_database_version <version> \  
--analysis_software_version <version> --format json  
  
Applies hAMRonization specification to output from abricate (OUTPUT.tsv)
```

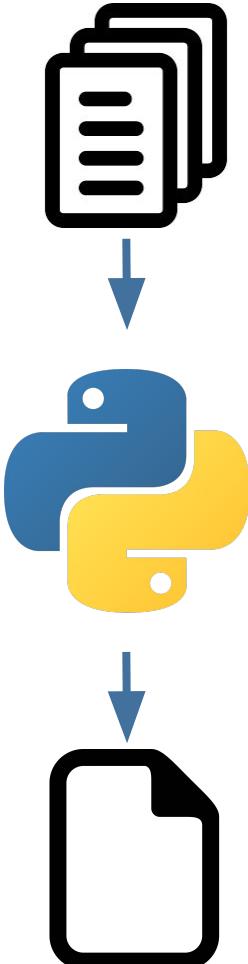
Combining all the reports:

```
> hamronize summarize --summary_type interactive <list of reports>
```



## Summary output formats:

- TSV
- Json
- Interactive HTML



# The hAMRonization Package



<https://github.com/pha4ge/hAMRonization>

Public Health Alliance for Genomic Epidemiology

Search Show Only Genomes With Hits Restore Results

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873306	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873307	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873308	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873309	16 hits	14 hits	14 hits	6 hits	11 hits
ERR873310	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873311	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873312	11 hits	10 hits	10 hits	6 hits	9 hits

Search Results

Total hits: 0

Genomes with hits: 0

Tools with hits: 0

Differential results: 0

Selected

Compare Clear

Public Health Alliance for Genomic Epidemiology

Search

Show Only Genomes With Hits

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873306	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873307	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873308	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873309	16 hits	14 hits	14 hits	6 hits	11 hits
ERR873310	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873311	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873312	11 hits	10 hits	10 hits	6 hits	9 hits

Public Health Alliance for Genomic Epidemiology

fosfomycin

Show Only Genomes With Hits

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
aa(6')-29a	aa(6')-29a	aa(6')-29a	sac(6')-29a	aa(6')-29a	
aph(3')-lb	aph(3')-lb	aph(3')-lb	blaVIM-2	aph(3')-lb	
blaOXA-395	blaOXA-395	blaOXA-395	catB7	blaOXA-50	
blaPDC-158	blaPDC-158	blaPDC-3	crpP	blaPAO	
blaPDC-55	blaPDC-55	blaVIM-2	fosA	blaVIM-2	
blaVIM-2	blaVIM-2	carB7	suI	carB7	
catB7	catB7	cpnP	cpnP	cpnP	
cmrB1	cmrB1	fosA	fosA	fosA	
crpP	crpP	qacEdelta1	suI	suI	
fosA-354827590	fosA-354827590	suI			
ERR873306					

<https://hamronization-demo.herokuapp.com>

17

# The hAMRonization Package



<https://github.com/pha4ge/hAMRonization>

Public Health Alliance for Genomic Epidemiology

Search Show Only Genomes With Hits Restore Results

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873306	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873307	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873308	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873309	16 hits	14 hits	14 hits	6 hits	11 hits
ERR873310	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873311	11 hits	10 hits	10 hits	6 hits	9 hits
ERR873312	11 hits	10 hits	10 hits	6 hits	9 hits

Search Results

Total hits: 0

Genomes with hits: 0

Tools with hits: 0

Differential results: 0

Selected

Compare Clear

Public Health Alliance for Genomic Epidemiology

fosfomycin Show Only Genomes With Hits

	abricate: config 0	amrfinderplus: config 0	csstar: config 0	resfinder.py: config 0	staramr: config 0
ERR873305	11 hits	10 hits	10 hits	6 hits	9 hits
	aac(6')-29a	aac(6')-29a	aac(6')	aac(6')-29a	aac(6')-29a
	aph(3')-Ib	aph(3')-Ib	aph(3')	aph(3')-Ib	aph(3')-Ib
	blaOXA-395	blaOXA-395	blaVIM-2	blaOXA-50	blaOXA-50
	blaPDC-158	blaPDC-3	bcr1	blaPAO	blaPAO
	blaPDC-55	blaVIM-2	blaPAO	blaVIM-2	blaVIM-2
	blaVIM-2	casB7	casB7	sul1	casB7
	catB7	crpP	crpP	fosA	crpP
	cmB1	fosA	fosA	sul1	fosA
	crpP	qacEdelta1	qacEdelta1	fosATR	qacEdelta1
	fosA-354827590	sul1	sul1	sul1	sul1

ERR873306

Public Health Alliance for Genomic Epidemiology

fosfomycin Show Only Genomes With Hits

	fosA-354827590	fosA	fosATR
ERR873305	abricate: config 0	ERR873305	ERR873305
	amrfinderplus: config 0	amrfinderplus: config 0	amrfinderplus: config 0
	csstar: config 0	csstar: config 0	csstar: config 0
input_file_name	ERR873305	ERR873305	ERR873305
gene_symbol	fosA-354827590	fosA	fosATR
gene_name	FosA family fosfomycin resistance glutathione transferase	FosA family fosfomycin resistance glutathione transferase	FosA family fosfomycin resistance glutathione transferase
reference_database_id	ncbi	NCBI Reference Gene Database	ResGANNOT
reference_database_version	2020-Apr-19	2020-05-20.1	2020-Nov-05
reference_accession	NG_047883.1	WP_003082280.1	fosATR
analysis_software_name	abricate	amrfinderplus	csstar
analysis_software_version	1.0.1	3.6.10	2.0.0
sequence_identity	98.53	98.52	98.529
contig_id	gnl JUGS ERR873305_36	gnl JUGS ERR873305_36	gnl JUGS ERR873305_36
query_start_aa			
query_stop_aa			
query_start_nt	8371	8374	

<https://hamronization-demo.herokuapp.com>

18

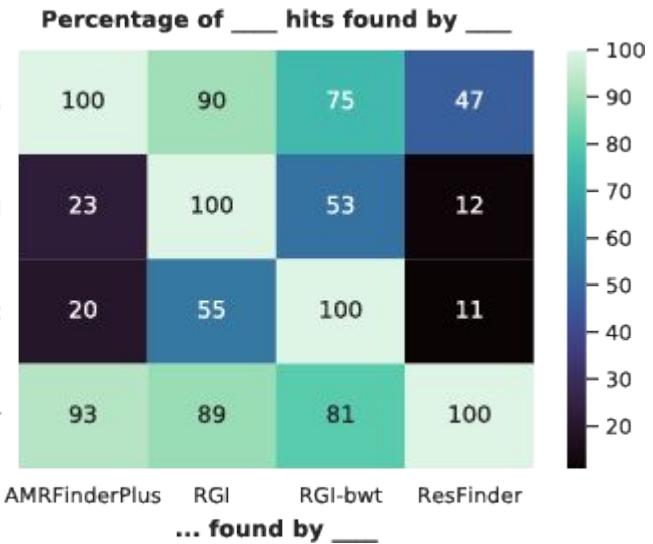
# 3.

The hAMRomization allows the **dissemination of results to stakeholders in a single consistent format**, allowing not only the comparison of tools and databases, but the validation of results through multiple detection algorithms.

# 89 *Klebsiella pneumoniae* sequence data



Percentage of hits ::



**RGI:** BLASTp + CARD

**RGI bwt (beta):** Bowtie2 + CARD

**AMRFinderPlus:** BLASTx + BARRG DB

**ResFinder:** BLASTn + ResFinder DB

	RGI			
	RGI	bwt	AMR	Res
Total # unique gene names	140	223	70	46



Elizabeth Culp



# 87 *Pseudomonas aeruginosa* sequence data

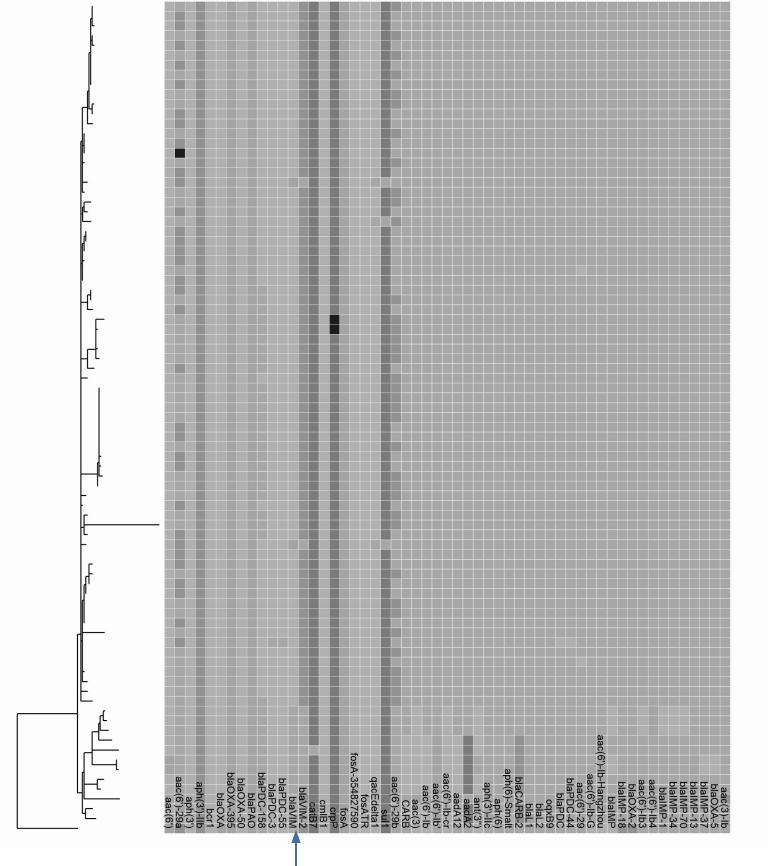


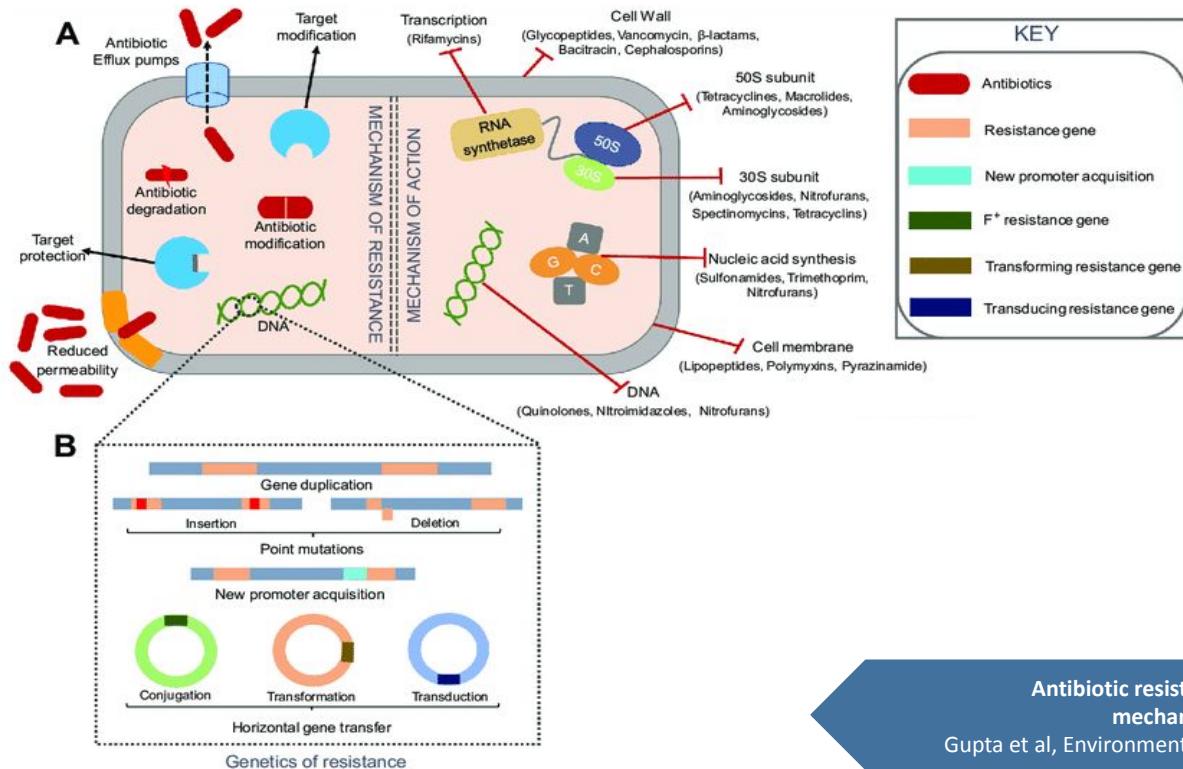
## Isolates (ST=111) associated with Metallo-Carbapenemases resistance outbreak in the United Kingdom

- Most isolates (73) carried VIM-2, conferring the resistance phenotype
  - Dataset run on **abricate**, **amrfinderplus**, **csstar**, **resfinder** and **staramr**
  - Multiple tools finding the same genes increases evidence, which increases confidence



Dr Adam Witney





### Antibiotic resistance dissemination mechanisms and pathways

Gupta et al, Environment International, 2020



CREATIVE  
Safety Supply

creativesafetysupply.com

866-777-1360 #FS72927-24

# Standardized Output Specification

## The hAMRonization Package



<https://github.com/pha4ge/hAMRonization>

### Gene Detection Standard



### Mutation Detection Standard

#### Additional terms:

- **Genetic Variation Type**
- Predicted Phenotype
- Predicted Phenotype Confidence Level
- Nucleotide mutation
- Nucleotide mutation interpretation
- Protein mutation
- Protein mutation interpretation
- Frequency of variant

# Standardized Output Specification

## The hAMRonization Package



<https://github.com/pha4ge/hAMRonization>



**Sequence variant nomenclature** system implemented for genetic changes and denoting sequence coordinates.

Due to the vast types of genetic changes that might occur, nomenclature is not easily understood. Programmatic interface to provide an interpretation is required.

<https://varnomen.hgvs.org/recommendations/general/>  
<https://github.com/conmeehan/laymansHGVS>

# The TB-Profiler case study



**TBProfiler database** (tbdb) incorporates **1541 individual mutation/resistance** associations across 16 anti-tuberculosis drugs



<https://github.com/jodyphelan/tbdb/>

- **Genetic Variation Type:** protein\_mutation
- Nucleotide mutation: c.1349C>T
- Nucleotide mutation interpretation:

*This is a subst found in rpoB at position 1349 where the reference has a C and the sample has a T*

- Protein mutation: p.Ser450Leu
- Protein mutation interpretation:  
*This is an amino acid subst found in rpoB at position 450 where the reference has a Serine and the sample has a Leucine*
- Frequency of variant: 0.57



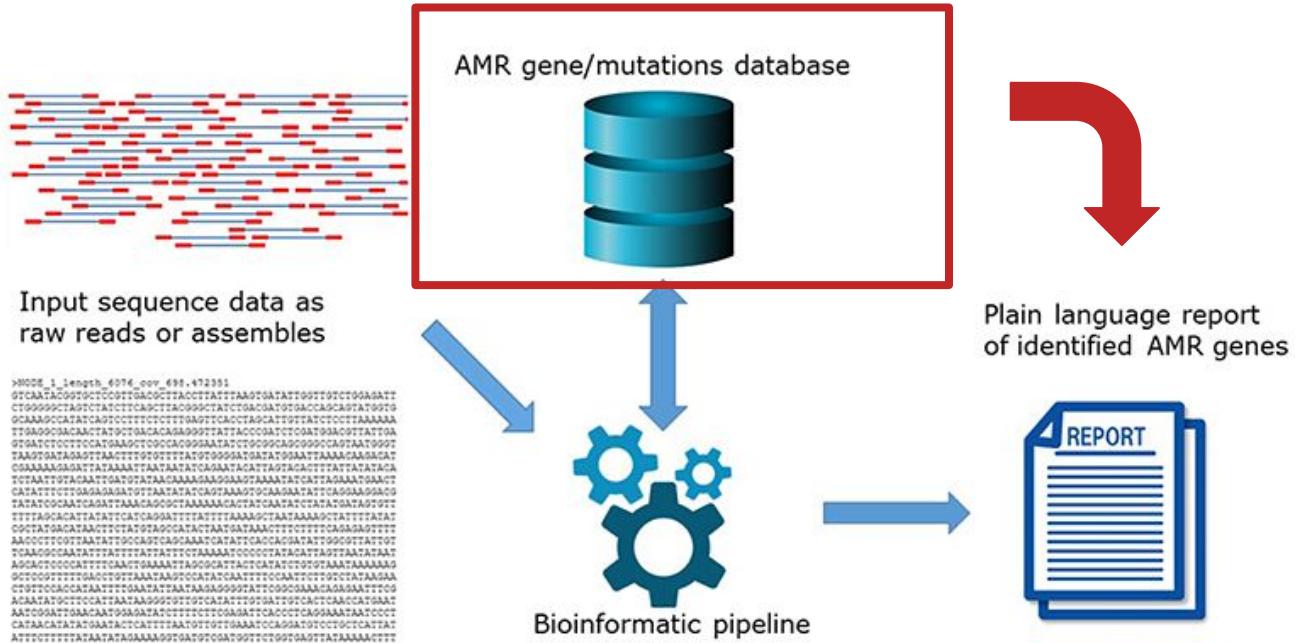
Jody Phelan



Conor Meehan

# 4.

Resistance phenotypes can be not only plasmid-mediated, or due to the presence of whole genes, but also chromosomal mutations. The ability to detect **not just gene absence/presence**, but **more granular changes** is of uttermost importance for different public health analyses.



## Using Genomics to Track Global Antimicrobial Resistance

# AMR Database Harmonization

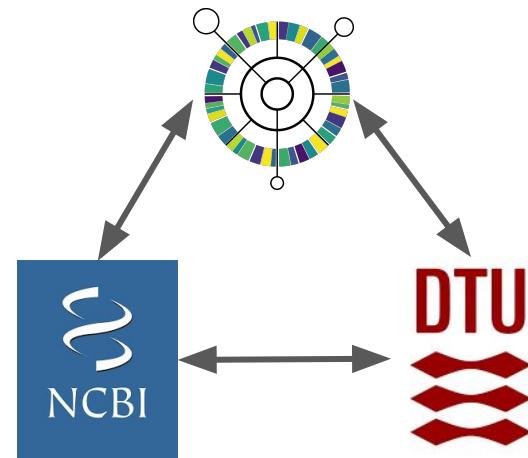


## Charm 😍 D<sub>b</sub>

(pronounced 'charmed' /tʃɑ:(r)mɪd/ )



<https://gitlab.com/antunderwood/chamredb>



Anthony  
Underwood



Inês Mendes



Alex Manuele



Dr Adam Witney

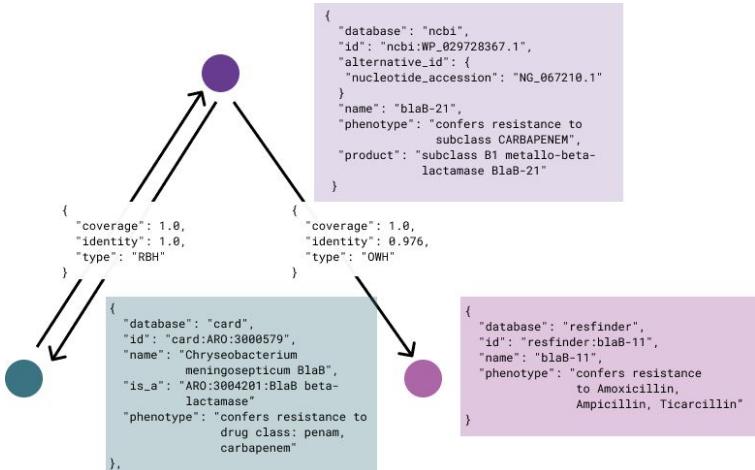
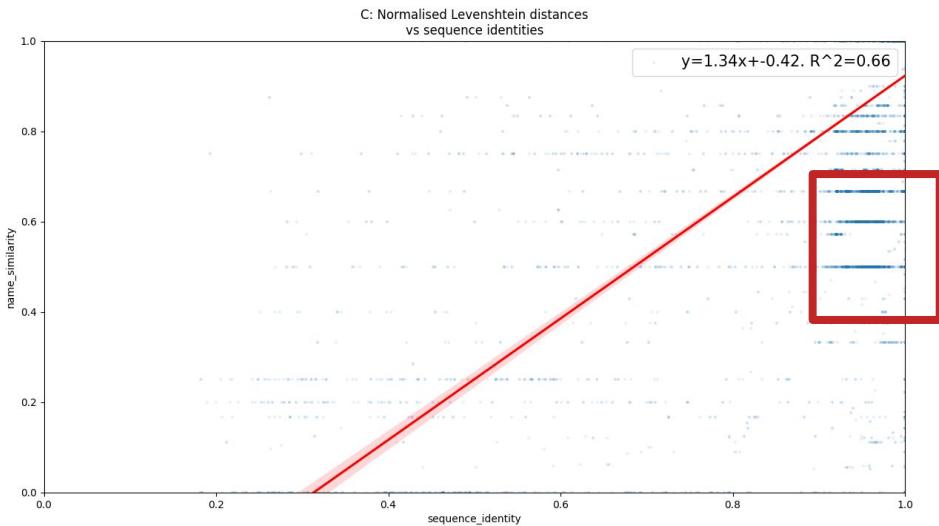


Trestan Pillonel



Varun Shamanna

# AMR Database Harmonization



# AMR Database Harmonization



```
$ chamredb query -d ncbi -i qnrB2
```

```
qnrB2 (qnrB2)
  phenotype: confers resistance to subclass QUINOLONE
  product: quinolone resistance pentapeptide repeat protein QnrB2
  card
    ↪ ARO3002735 (QnrB20)
      ⚡ coverage: 1.0
      ⚡ identity: 0.995
      ⚡ type: RBH
      PMID: ['18993034']
      is_a: ARO:3000419:quinolone resistance protein (qnr)
      phenotype: [['confers resistance to antibiotic: gatifloxacin,sparfloxacin,levofloxacin,nalidixic acid,moxifloxacin,ciprofloxacin,norfloxacin', 'confers resistance to drug class: fluoroquinolone antibiotic']]
    ↪ resfinder
      ↪ qnrB2 (qnrB2)
        ⚡ coverage: 1.0
        ⚡ identity: 1.0
        ⚡ type: RBH
        phenotype: confers resistance to Ciprofloxacin
=====
```

# AMR Database Harmonization



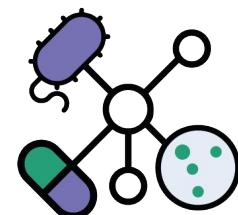
```
$ hamronize abricate report.tsv --reference_database_version db_v_1 \
--analysis_software_version tool_v_1 --format json
```

...

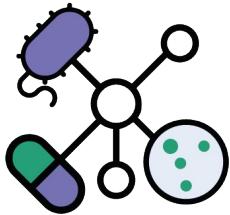
```
$ hamronize summarize -o hamronize_summary.json -t json *.json
```

```
$ chmreldb query -j hamronize_summary.json -o hamronize_summary.tsv
```

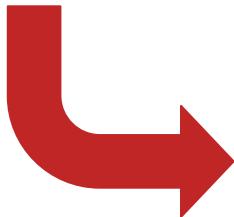
Charm 😍 D<sub>b</sub>



# AMR Database Harmonization



Charm 😍 D<sub>b</sub>



Public Health Alliance for  
Genomic Epidemiology

Search Show Subgraph

sample	id	database	name	metadata	AI
G18002568	NG_047244.1	ncbi	aac(3')-Ile	phenotype: confers resistance to subclass GENTAMICIN, product: aminoglycoside N-acetyltransferase AAC(3')-Ile	AI
G18002568	NG_054648.1	ncbi	ant(3')-Ila	phenotype: confers resistance to subclass STREPTOMYCIN/SPECTINOMYCIN, product: aminoglycoside nucleotidyltransferase ANT(3')-Ila	AI
G18002569	NG_049444.1	ncbi	blaOXA-146	phenotype: confers resistance to subclass CARBAPENEM, product: OXA-23 family carbapenem-hydrolyzing class D beta-lactamase OXA-146	AI
G18002569	NG_051852.1	ncbi	sul2	phenotype: confers resistance to subclass SULFONAMIDE, product: sulfonamide-resistant dihydropteroate synthase Sul2	AI
G18002570	NG_047282.1	ncbi	aac(6')-Ian	phenotype: confers resistance to subclass AMIKACIN/KANAMYCIN/TOBRAMYCIN, product: aminoglycoside N-acetyltransferase AAC(6')-Ian	AI
G18002570	NG_049326.1	ncbi	blaNDM-1	phenotype: confers resistance to subclass CARBAPENEM, product:	AI

# 5.

A unified global picture requires not only a **common ground** for the comparison of results from different **tools**, but a way to compare the **databases** that are used to generate those results.



# 15th October 2021 AMR Workshop

Inês Mendes

[cimendes@medicina.ulisboa.pt](mailto:cimendes@medicina.ulisboa.pt)

Twitter: @ines\_cim

PHA4GE Data Structures Workgroup

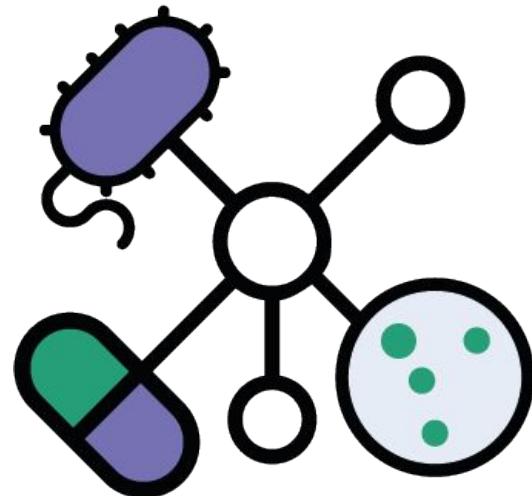
[datastructures@pha4ge.org](mailto:datastructures@pha4ge.org)

PHA4GE Consortium

[help@pha4ge.org](mailto:help@pha4ge.org)

<https://pha4ge.org/>

Twitter: @pha4ge



Public Health Alliance for  
Genomic Epidemiology

