



DATABASES, RESOURCES AND TOOLS
FOR ANTIMICROBIAL RESEARCH

Comparing and reporting AMR results using hAMRonization

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



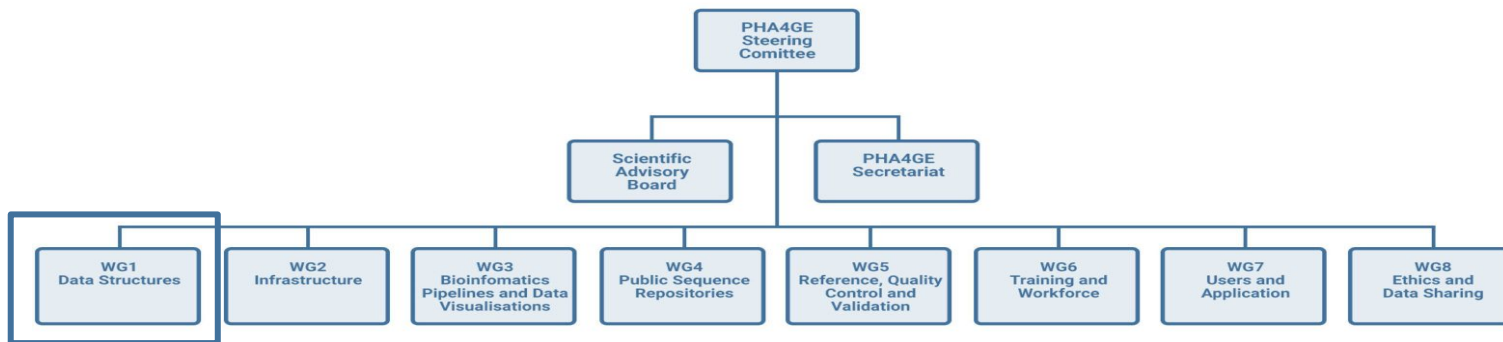
PHA4GE - Data Structures Workgroup



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

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



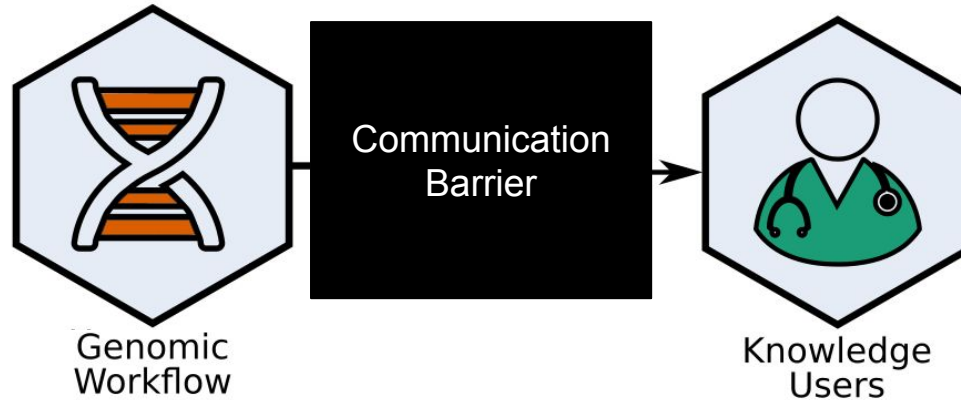
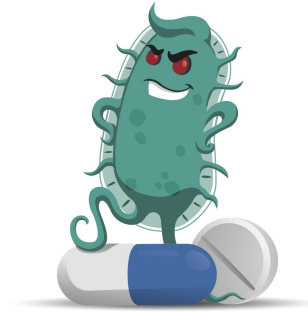
Goal:

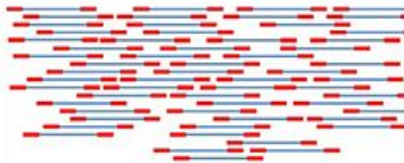
The Antimicrobial Resistance Data Problem



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**.

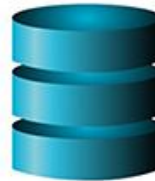




Input sequence data as raw reads or assembles

```
>NODE_1_length_6076_cov_695.472351
GTCAATACGGTCTCCGTTGACCTTAACTTATTAAAGTATATGGTTGTCTGGAGATT
CTGGGGGCTAGTCTATCTTCAAGCTTAAGGGCTATCTGACGATGTGACAGCAGTATGGTG
GGAAAGCCATACAGTCTCTTCTTTGAGTTCACTAGCAGTATGTTACTCTTAAAGAA
TTGAGGCGACACACTATGCTGACAGAGGGTTATACCCGATCTTGGATGGAGGTTATGA
GTATCTCTTCCATGAGCTCCGCCAGGGAATCTGCGCGGAGGGGCGAGTAAATGGGT
TAAGTGAAGAGTTAAATTTGTTTATGTGGGATGATATGGAAATTAAGAACAGACAT
CGAAAGAGAGTATAAAATTAATATACAGAAATACATTAAGTACATTTATATATACA
TCTAATTGTACATGATGATTAAGAAAGAGAGGAGTAAATATGCTATTGAAAGTAACT
CATATTTCTGAGAGGATGTTAAATATACAGTAAAGTGGCAAGATATCGAGAAAGGCG
TATATCCCAATCAGATTAAACAGGCTTAAAAACAGTATCAATATCTATATAGTGT
TTTAAAGCATTATATCATCAGGATTTTATTTAAAGCTAAATAAGCTATTTATAT
GGCTATGACATAATTTATGAGGATCTATGATAAGCTTTCTTTGAGAGGTTT
AAAGCTTGTAAATTTGGCAGTCCAGCAATCATATTCAACAGATATGGGTTATGGT
TCAAAGCCAAATTTATTTATTTTCTAAAAATCCCGCTATACATTAAGTAAATAAAT
AGCAGTCCGATTTCAACTGAAATTAAGGCGATGATCATATCTGTAAATAAAAAAG
GGTGGTTTTGAGCTGTTAAATAGGCGATATGAAATTTCCAGATTTGTCTATAGAA
CTGTCCAGCATAATTTGAAATTAATAGAGGGGTTATCGGCGAAGCAGAGATTCG
ACATATGCTTCATTAAAGGGTGTGTCAATTTGTGATGTCACCTCAACCATGAAT
AATCGAATGACCAATGGAGATCTTTTCTTGGAGATCAAGCTCAGGAAATATATCCG
CATACATATGAAATCTCATTTAGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
ATTTCTTTTAAATATAGAAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG
```

AMR gene/mutations database



Bioinformatic pipeline

Plain language report of identified AMR genes



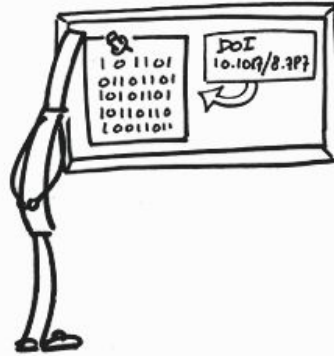
Using Genomics to Track Global Antimicrobial Resistance
Hendriksen et al, Front. Public Health, 2019

FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE



INTEROPERABLE



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

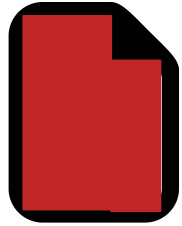
Data **specification** for the detection of AMR genes based on **harmonisable, publicly available, community standards** implementable via a **report template** with a **standardized list of labels and terms** and respective **ontology identifiers**.

Mandatory terms:

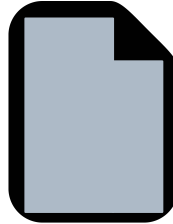
- Input File Name
- Gene Symbol
- Gene Name
- Sequence Identity
- **Reference Database ID**
- **Reference Database Version**
- Reference Accession
- Reference Sequence Length
- Target Sequence Length
- **Analysis Software Name**
- **Analysis Software Version**



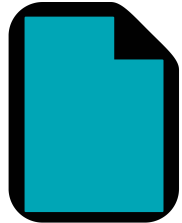
Standardized Output Specification



Report of
tool X



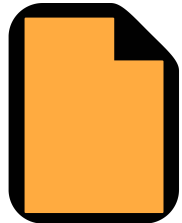
Standard Report
of tool X



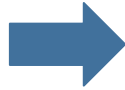
Report of
tool Y



Standard Report
of tool Y



Report of
tool Z



Standard Report
of tool Z



Standardized AMR Gene Detection Output



AMR Prediction Tools

Lots of software with different output formats.

hAMRonization Workflow

Run any dataset through the set of tools.

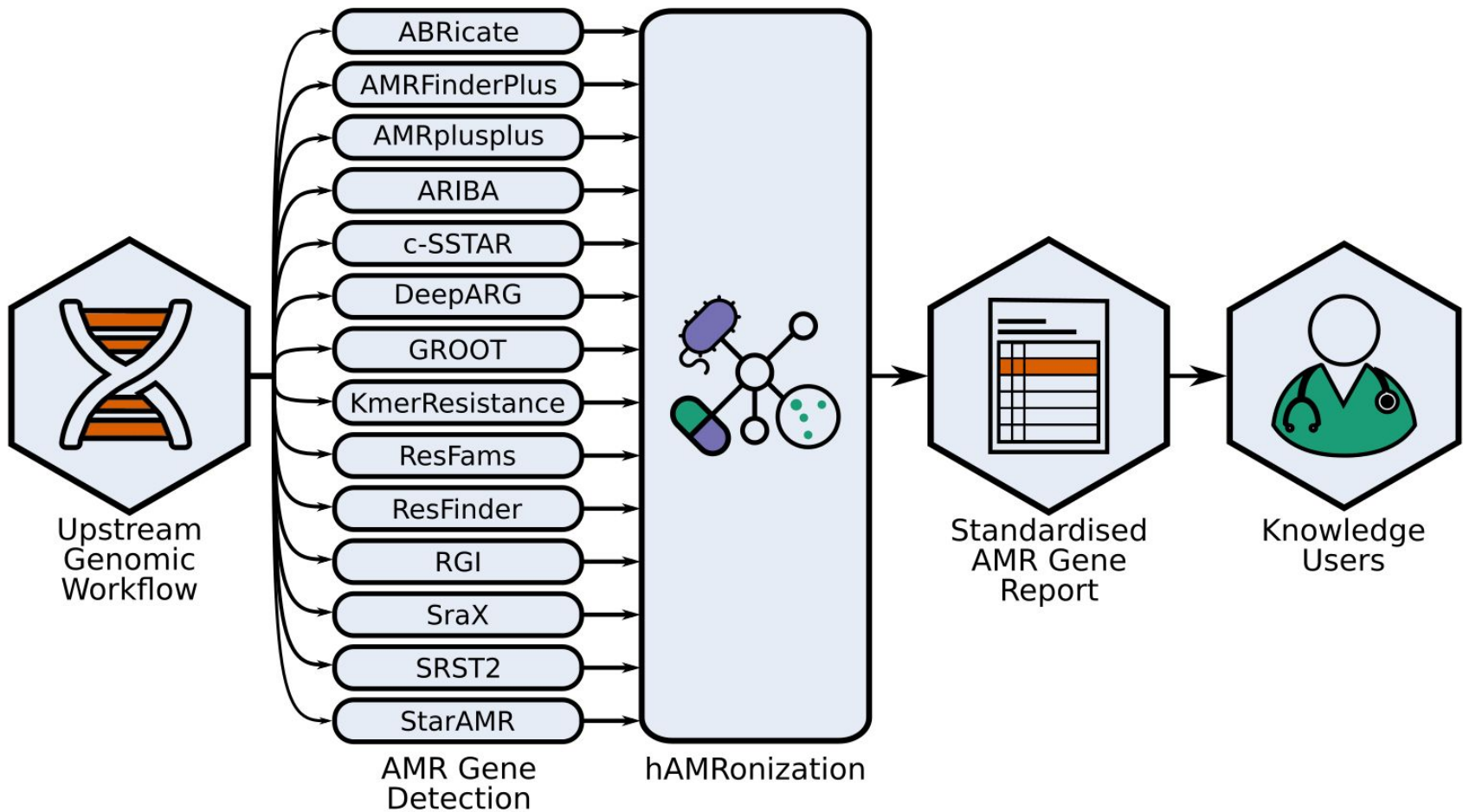
hAMRonization Package

Take tool-specific output and reformat to standardized output.

Standardized Output Specification

Enables consistent comparison of AMR genes detected by different tools.

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

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



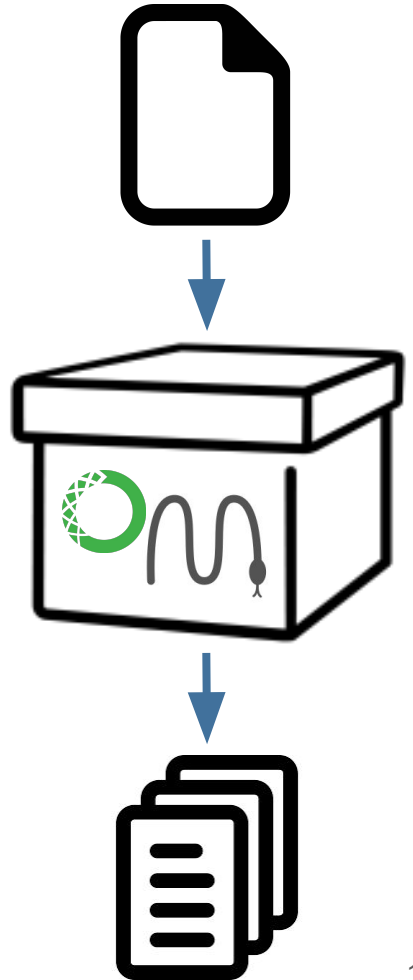
Dr Finlay Maguire



Dr Adam Witney



Dr Simon Tausch



The hAMRonization Package



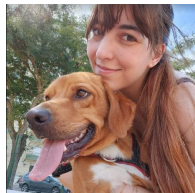
<https://github.com/pha4ge/hAMRonization>

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>

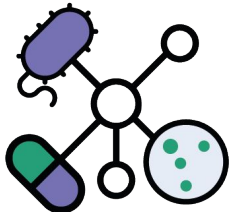
An example from abricate:

```
> 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

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

fofosoycoil 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

Search Results

Total hits: 130

Genomes with hits: 43

Tools with hits: 1

Differential results: 43

Selected

Compare Clear

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 aac(6)-29a aph(3)-Ib blaOXA-395 blaPDC-158 blaPDC-55 blaVIM-2 blaVIM-2 catB7 cmlB1 crpP fosA-354827590 sul1	10 hits aac(6)-29a aph(3)-Ib blaOXA-395 blaPDC-3 blaVIM-2 catB7 crpP fosA qacEdelta1 sul1	10 hits aac(6) aph(3) bcrl blaOXA blaPAO blaVIM catB7 crpP sul1	6 hits aac(6)-29a blaVIM-2 catB7 crpP fosA sul1	9 hits aac(6)-29a aph(3)-Ib blaOXA-50 blaPAO blaVIM-2 catB7 crpP fosA sul1

Search Results

Total hits: 44

Genomes with hits: 41

Tools with hits: 4

Differential results: 41

Selected

fosA-354827590 from genome 0, tool 0

fosA from genome 0, tool 1

fosATR from genome 0, tool 2

Compare Clear

Public Health Alliance for Genomic Epidemiology

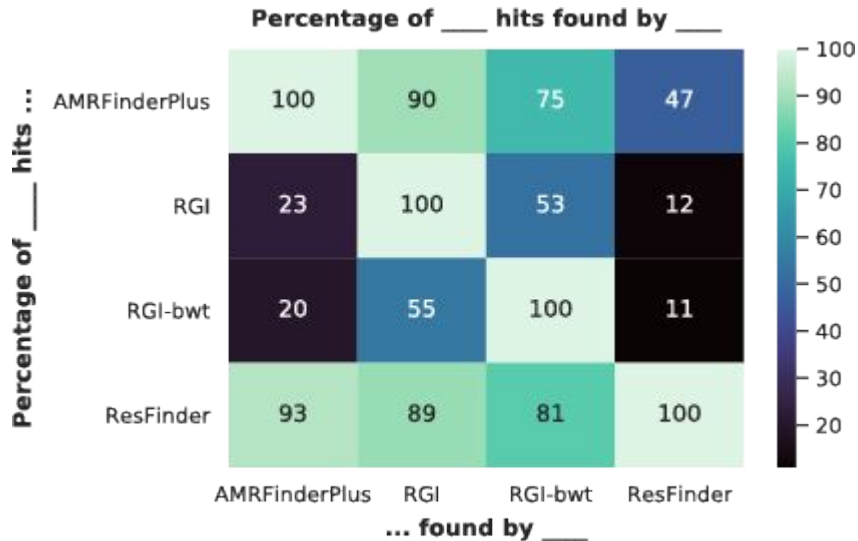
fosfomycin Show Only Genomes With Hits

	ERR873305 abricate: config 0	ERR873305 amrfinderplus: config 0	ERR873305 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	fosATR
reference_database_id	ncbi	NCBI Reference Gene Database	RefGANNOT
reference_database_version	2020-Apr-19	2020-03-20.1	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
config_id	griBUGSERR873305_36	griBUGSERR873305_36	griBUGSERR873305_36
query_start_aa			
query_stop_aa			
query_start_nt	8371	8374	

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



RGI: BLASTp + CARD

RGI bwt (beta): Bowtie2 + CARD

AMRFinderPlus: BLASTx + BARRG DB

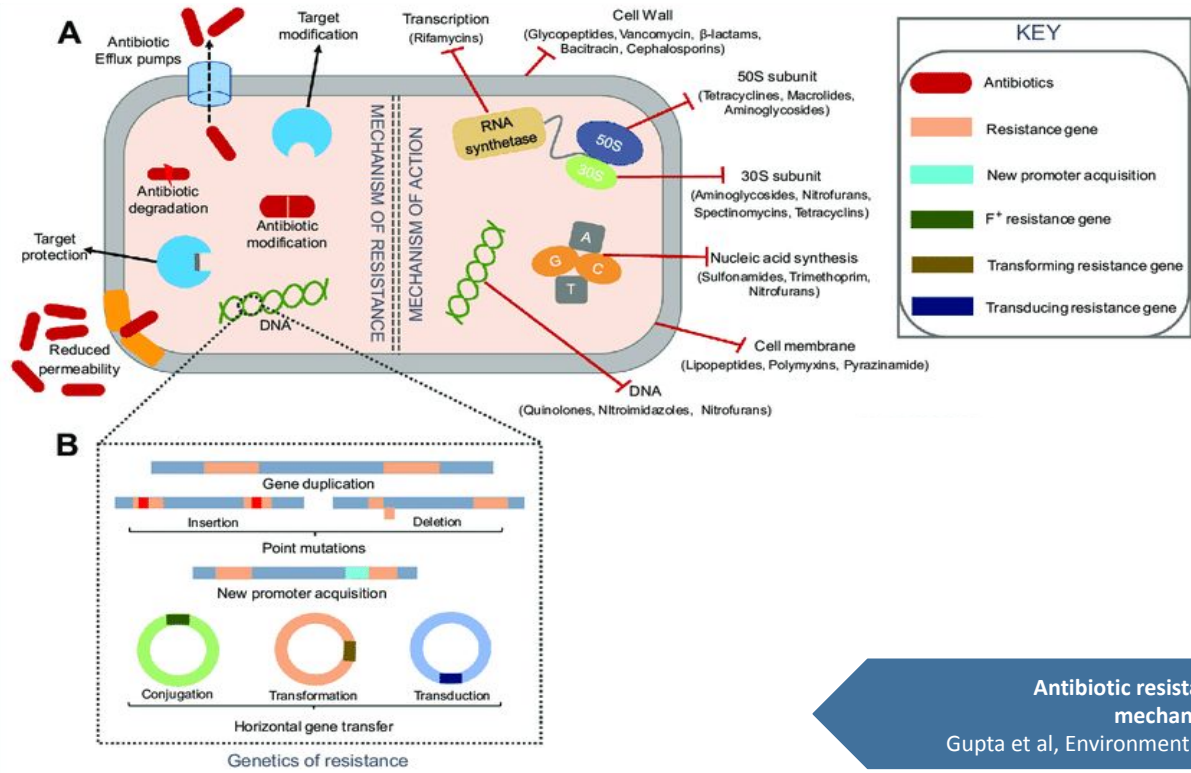
ResFinder: BLASTn + ResFinder DB



Elizabeth Culp



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



Antibiotic resistance dissemination mechanisms and pathways
 Gupta et al, Environment International, 2020



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



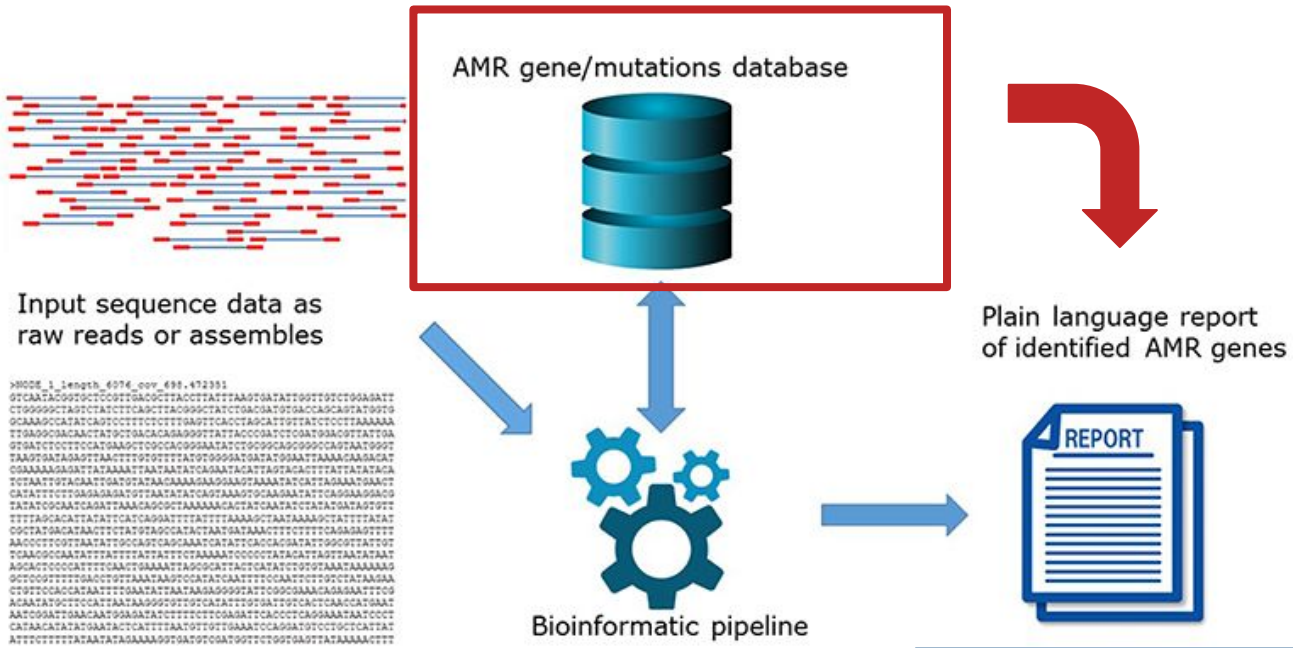
Jody Phelan



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
 Hendriksen et al, Front. Public Health, 2019

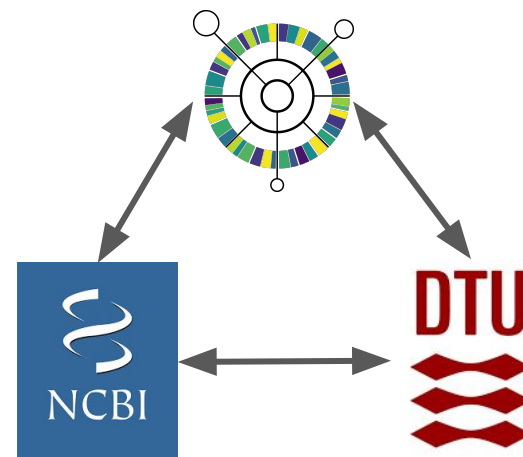


Charm Db

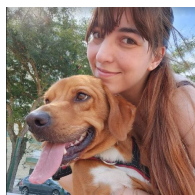
(pronounced 'charmed' /tʃɑ:(r)md/)



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



Anthony Underwood



Inês Mendes



Alex Manuele



Dr Adam Witney

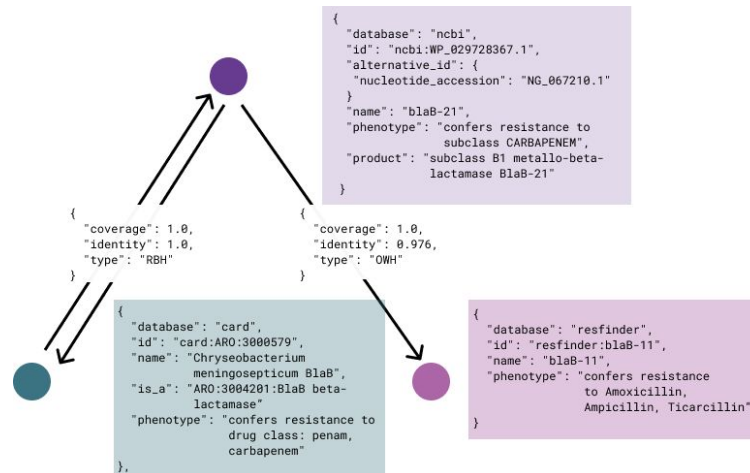
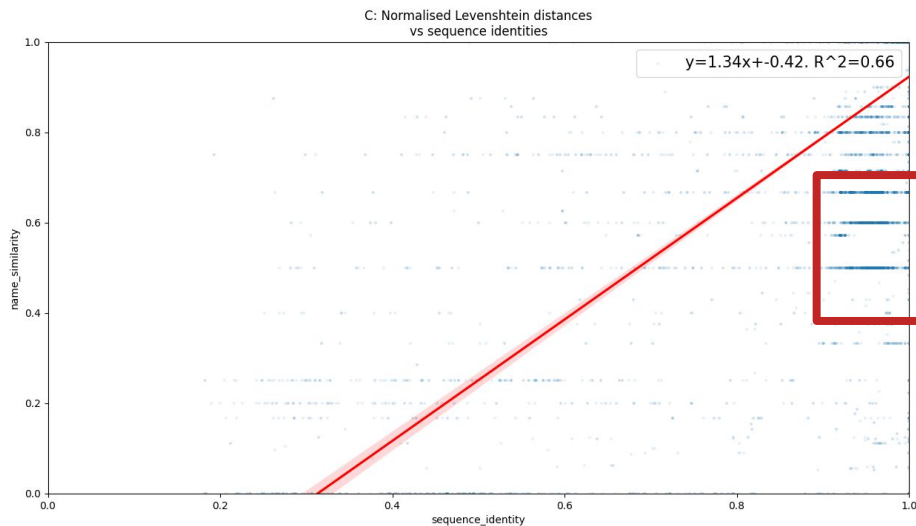


Trestan Pillonel



Varun Shamanna

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



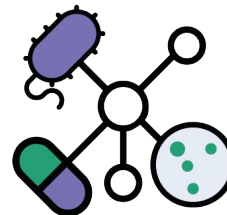
```
$ hamronize abricate report.tsv --reference_database_version db_v_1 \  
--analysis_software_version tool_v_1 --format json
```

...

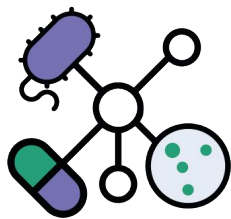
```
$ hamronize summarize -o hamronize_summary.json -t json *.json
```

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

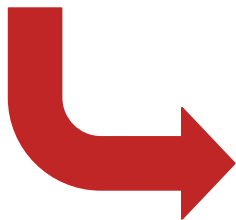
Charm  Db



AMR Database Harmonization



Charm  Db



Public Health Alliance for Genomic Epidemiology

Search [Show Subgraph](#)

sample	id	database	name	metadata	
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'')-IIa	phenotype: confers resistance to subclass STREPTOMYCIN/SPECTINOMYCIN, product: aminoglycoside nucleotidyltransferase ANT(3'')-IIa	Ni
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)-Ilan	phenotype: confers resistance to subclass AMIKACIN/KANAMYCIN/TOBRAMYCIN, product: aminoglycoside N-acetyltransferase AAC(6)-Ilan	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

Twitter: @ines_cim

PHA4GE Data Structures Workgroup

datastructures@pha4ge.org

PHA4GE Consortium

help@pha4ge.org

<https://pha4ge.org/>

Twitter: @pha4ge

