#O588

Effective use of antibiotic resistance gene databases for whole genome and metagenomic sequencing data

Anupam J Das, Basil Britto Xavier, Guy Cochrane, Sandra de Ganck,
Samir Kumar-Singh, Frank Moller Aarestrup, Herman Goossens, Surbhi Malhotra-Kumar

Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerpen, Belgium

Laboratory of Medical Microbiology
Vaccine & Infectious Disease Institute
University of Antwerp

Horizon 2020 COMPARE (http://www.compare-europe.eu/)

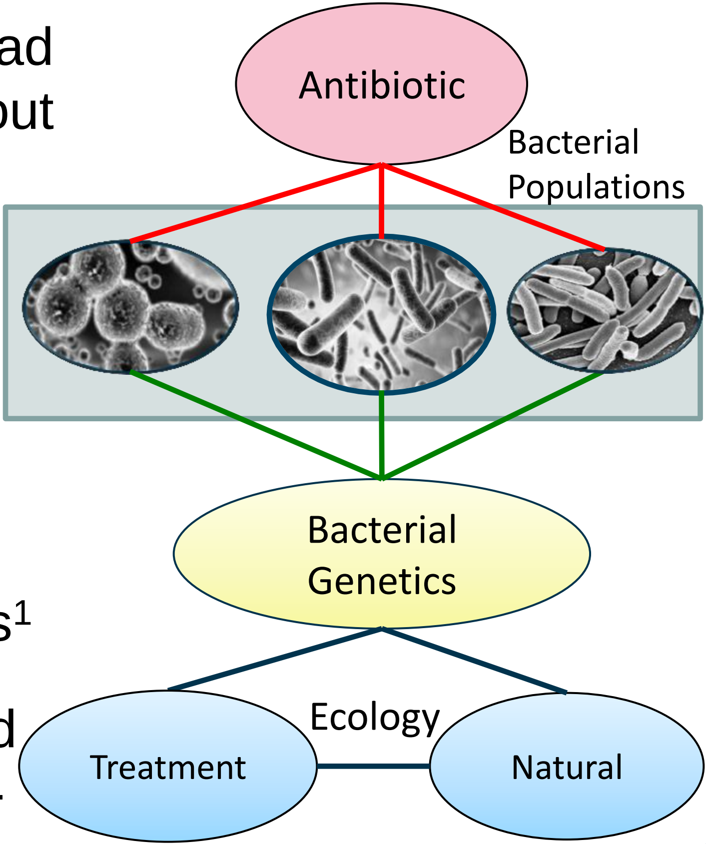
Evolution of Resistance to Antibiotics

- physical and biological forces cause widespread dissemination of resistance genes throughout many environments

- resistance genes exist naturally in the environment owing to a range of selective pressures in nature

- first β-lactamase was identified in *E coli* prior to the release of penicillin-use in medical practices[1]

- overuse of antibiotics has led to rampant spread of these genes, resulting in emergence of multi-drug resistant species



---

[1] Abraham and Chain, 1940

Antibiotic Resistance Gene Databases

- provide a centralized compendium of information on antibiotic resistance, manually curated

- facilitate consistent annotation of resistance information in newly sequenced organisms

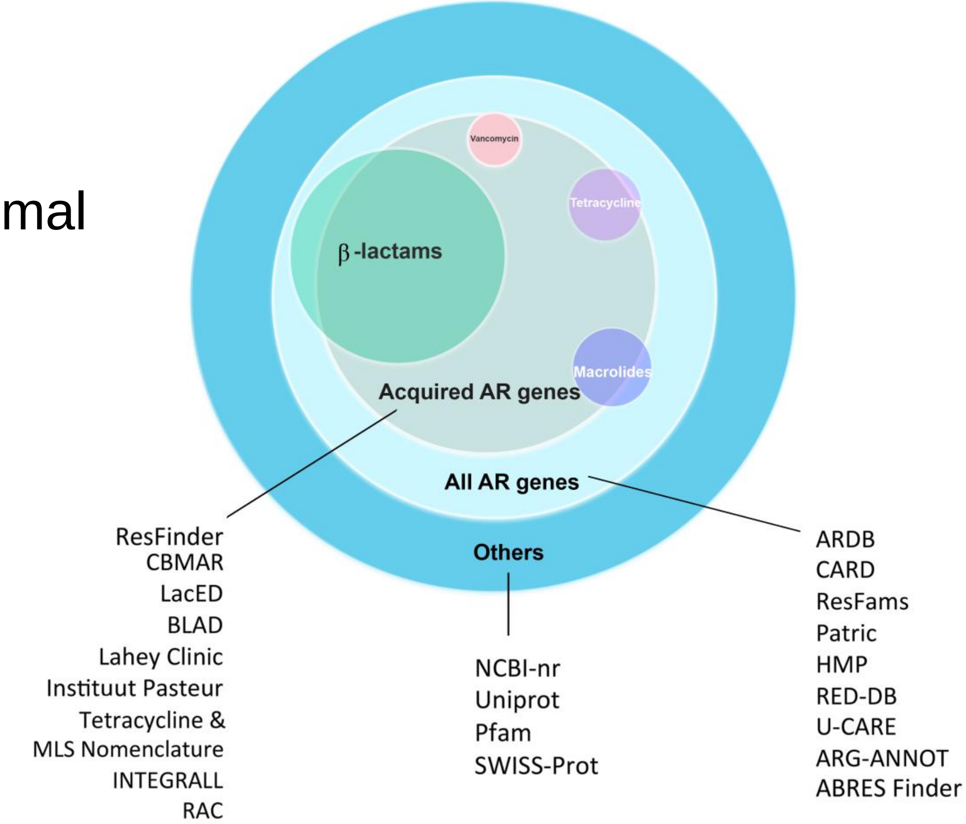
- facilitate identification and characterization of new genes

- platform to integrate informatics tool to bridge antibiotic resistance concerns in health care, agriculture, and the environment

Subset of AR Gene Data Resources

Include all possible AR genes

- both acquired and chromosomal gene databases
- major antibiotic classes
- virulence factors
- Reliability of results?



ResFinder
CBMAR
LacED
BLAD
Lahey Clinic
Instituut Pasteur
Tetracycline & MLS Nomenclature
INTEGRALL
RAC

NCBI-nr
Uniprot
Pfam
SWISS-Prot

ARDB
CARD
ResFams
Patric
HMP
RED-DB
U-CARE
ARG-ANNOT
ABRES Finder

Assessment and Validation of Databases

- for the benefit of the microbiology research community, assessment was performed on the AR data resources to identify their limitations, and to propose potential solutions

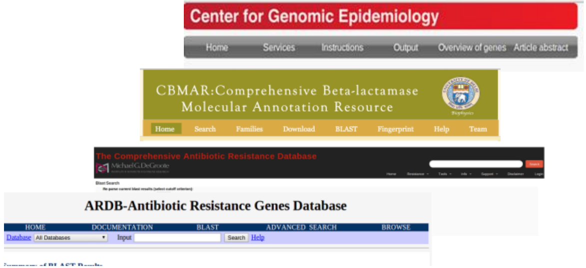
Assessment

  - availability of resources

  - consistency and accuracy of information variations across different databases

Popular Databases

Validation and assessment of the databases

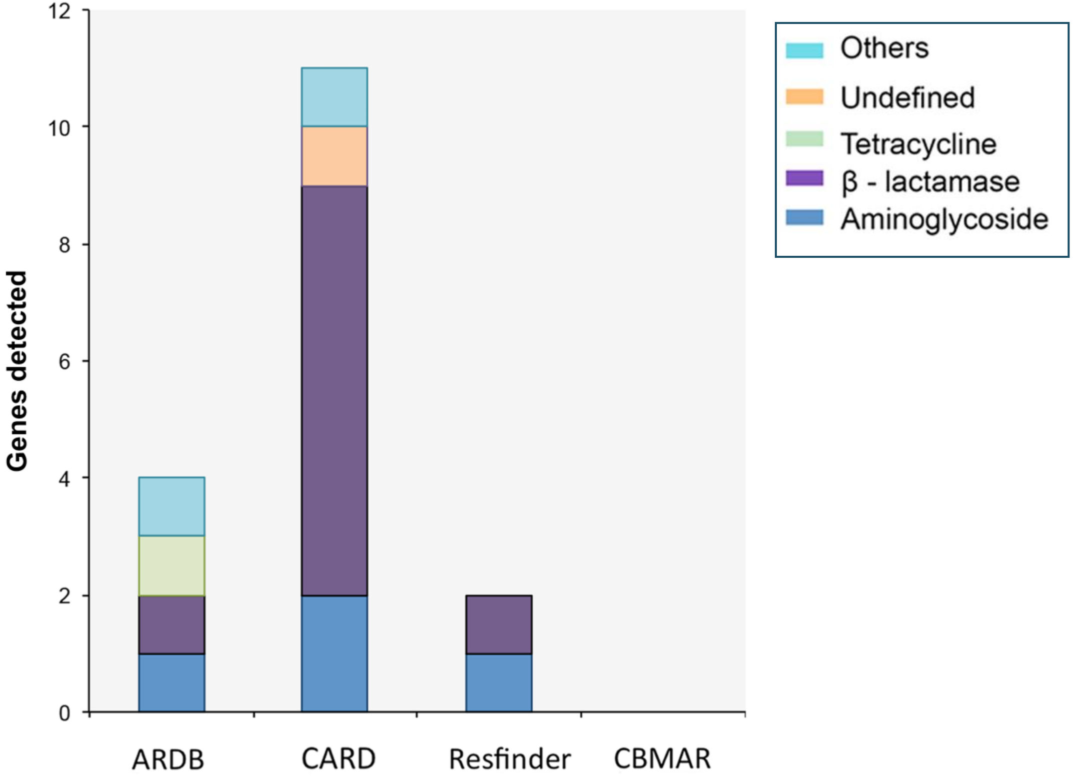


Resfinder	https://cge.cbs.dtu.dk/services/ResFinder/
CARD	https://card.mcmaster.ca/analyze/rgi
ARDB	http://ardb.cbcb.umd.edu/
CBMAR	http://14.139.227.92/mkumar/lactamasedb/
Resfams	http://www.dantaslab.org/resfams/

Data used for the assessment

Sequence Data	Background	DDBJ/EMBL/GenBank databases - Accession numbers	Reference
Metagenomic	fecal samples - SDD treatment	PRJEB3977	Buelow et al, 2014
Whole genome	MRSA strains clinically relevant	EMRSA -15 (CP007659) JDK6008 (CP002120) UAS391(CP007690)	Sabirova et al, 2014 Howden et al, 2010 Sabirova et al, 2014
Single gene	Verona imipenemase metallo-beta-lactamase *bla*$_{\text{VIM}}$ New Delhi metallo-beta-lactamase *bla*$_{\text{NDM}}$	*bla*$_{\text{VIM-1}}$ (KT124311) *bla*$_{\text{VIM-2}}$ (KR337992.1) *bla*$_{\text{VIM-4}}$ (AJ585042.1), *bla*$_{\text{VIM-19}}$ (KT124310) *bla*$_{\text{IM-35}}$ (JX982634.1) *bla*$_{\text{NDM-1}}$ (KP770030.1), *bla*$_{\text{NDM-2}}$ (JF703135.1) *bla*$_{\text{NDM-4}}$ (KP772213), *bla*$_{\text{NDM-6}}$ (KJ872581.1) *bla*$_{\text{NDM-8}}$ (NG_036906.1)	
Latest updates	Colistin resistance gene	*mcr-1*	Liu et al, 2015

Results - Metagenomic Data

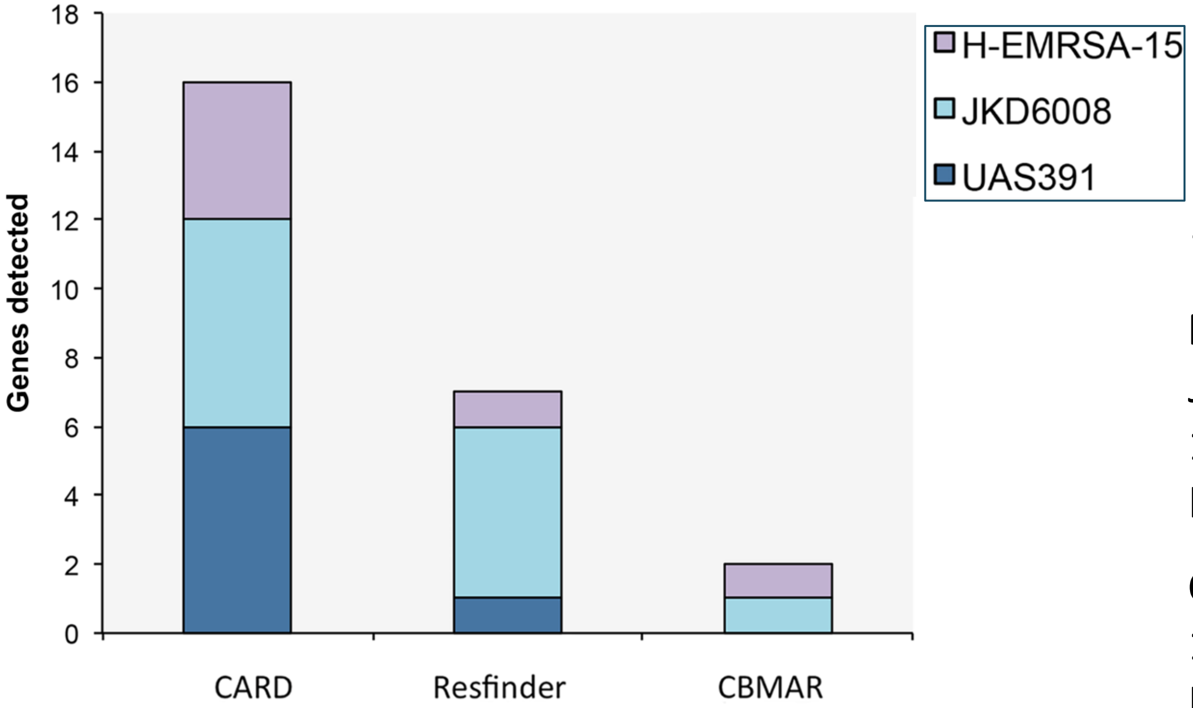


ARDB
1 each from aminoglycoside
β-lactamase
tetracycline and others

CARD
2 aminoglycoside resistance genes
7 β-lactamases
2 either undefined or other genes

Resfinder
1 aminoglycoside
1 β -lactamase

Results - Whole Genome Sequencing Data

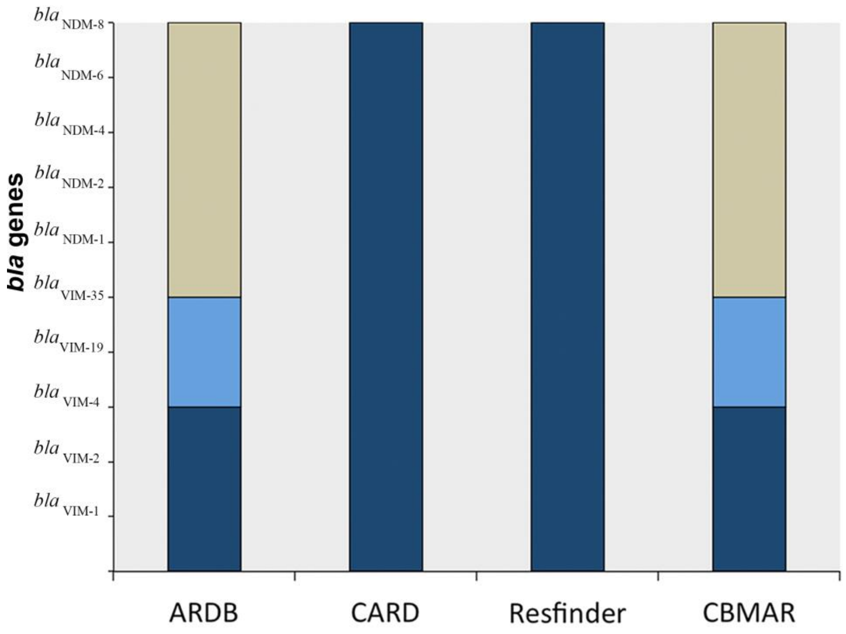


CARD
6 for UAS391 and JKD6008 strains
4 for the H-EMRSA-15

Resfinder
JKD6008
1 each for UAS391 and
H-EMRSA-15

CBMAR
1 each for JKD6008
H-EMRSA-15
no hits for UAS391

Results - Single gene sequences



- 3 (*bla*$_{\text{VIM-1}}$, *bla*$_{\text{VIM-2}}$, and *bla*$_{\text{VIM-4}}$) correctly identified by all of the 4 databases

- ARDB and CBMAR returned several non-specific hits

- *bla*$_{\text{VIM-19}}$ and *bla*$_{\text{VIM-35}}$ were incorrectly detected by ARDB and CBMAR

- *bla*$_{\text{NDM}}$ genes - ARDB and CBMAR returned no hits

- CARD and Resfinder found to consistently return correct hits

Overview of results

Database	Last update*	Metagenomic Sequences	Whole genome sequences	Single gene Sequences
ARDB	07/2009	4	-NA-	3
CARD	03/2016	11	16	10
RESFINDER	03/2016	2	8	10
CBMAR	09/2014	0	2	3

* As of 30 March, 2016

Summary

Database	Merits	Demerits
ARDB	1. First database introduced	1. Limited scope due to lack of regular updates
CARD	1. Detects all AR genes 2. Up to date (*mcr-1* gene records) 3. Accurate predictions 4. No limitations on size or type of data	
RESFINDER	1. Detects acquired mutations 2. Up to date (mcr-1 gene records) 3. Accurate predictions 4. No limitations on size or type of data	1. Ignores chromosomal mutations
CBMAR	1. Repository updated with all carbapenamase genes	1. Webtools need to be updated 2. Latest information unavailable (*mcr-1* records missing)

Main findings

While we have noted the value of the AR related data resources, we have also noted a number of limitations

- gaps

- inconsistent results of searches against different resources with the same query data

- lack of up-to-date reference data

Two step approach

Approach 1
- Simpler to lay out in practical terms
- we do not seek to fill gaps where they exist in AR data resource services, nor to benchmark precision and reliability, but rather seek to create a landscape of transparent and tractable elements that can contribute to many different current and future analytical infrastructures
- Establishing best practice around the open sharing of richly and systematically described reference data - sequences, annotations, alignments and models
- Will reduce redundant effort in discovering source data for analysis and curation in specialist resources and will maximise opportunities to fill gaps
- Systematic descriptions of computational methods and query services offered by specialist AR data resources will aid in users' selection of appropriate tools for their analyses and minimise risk of misinterpretation

Approach 2
- Requiring significant conceptual planning before practical work
- Broad community engagement required
- Both generalist and AR specialist, should move forward to fill gaps in coverage, to provide consistency between query tools that are intended to serve the same function
- To remove redundant data processing, curation and software development steps to maximise overall productivity to guide consumers in making informed analyses

Collaborators

Guy Cochrane

EMBL-EBI, UK

Frank Moller Aarestrup

DTU, Denmark

Horizon 2020 COMPARE (http://www.compare-europe.eu/)

Laboratory of Medical Microbiology
Vaccine & Infectious Disease Institute
University of Antwerp

https://www.uantwerpen.be/en/rg/lab-of-medical-microbiology/

LMM@uantwerpen.be