AMD TRAINING LEAD *and* BIOINFORMATICS REGIONAL RESOURCE

Demystifying Series: AMR Detection Using NGS

Joel R. Sevinsky, Ph.D.
WY PHL AMD Training Lead
MA DPH AMD Bioinformatics Resource
26 February 2020

Demystifying Series Objectives:

- Discuss a specific application of NGS data
  - Often the application is constantly evolving/improving
- Create more interactive content than static content
  - Google sheets, websites, Tableau views
- Record sessions for review
  - Unlisted youtube content available by link on www.statph.org
- Provide slides and materials on www.staphb.org as a resource
  - Free to download slides as pdf file
- Since it is not possible to cover all aspects of a proposed topic in a single webinar, these webinars serve as a beginning to your investigation and education on the topic.
- Encourage additional discussions on Slack (contact Noah Hull, Kelly Oakeson, Erin Young, or Joel Sevinsky to be connected).

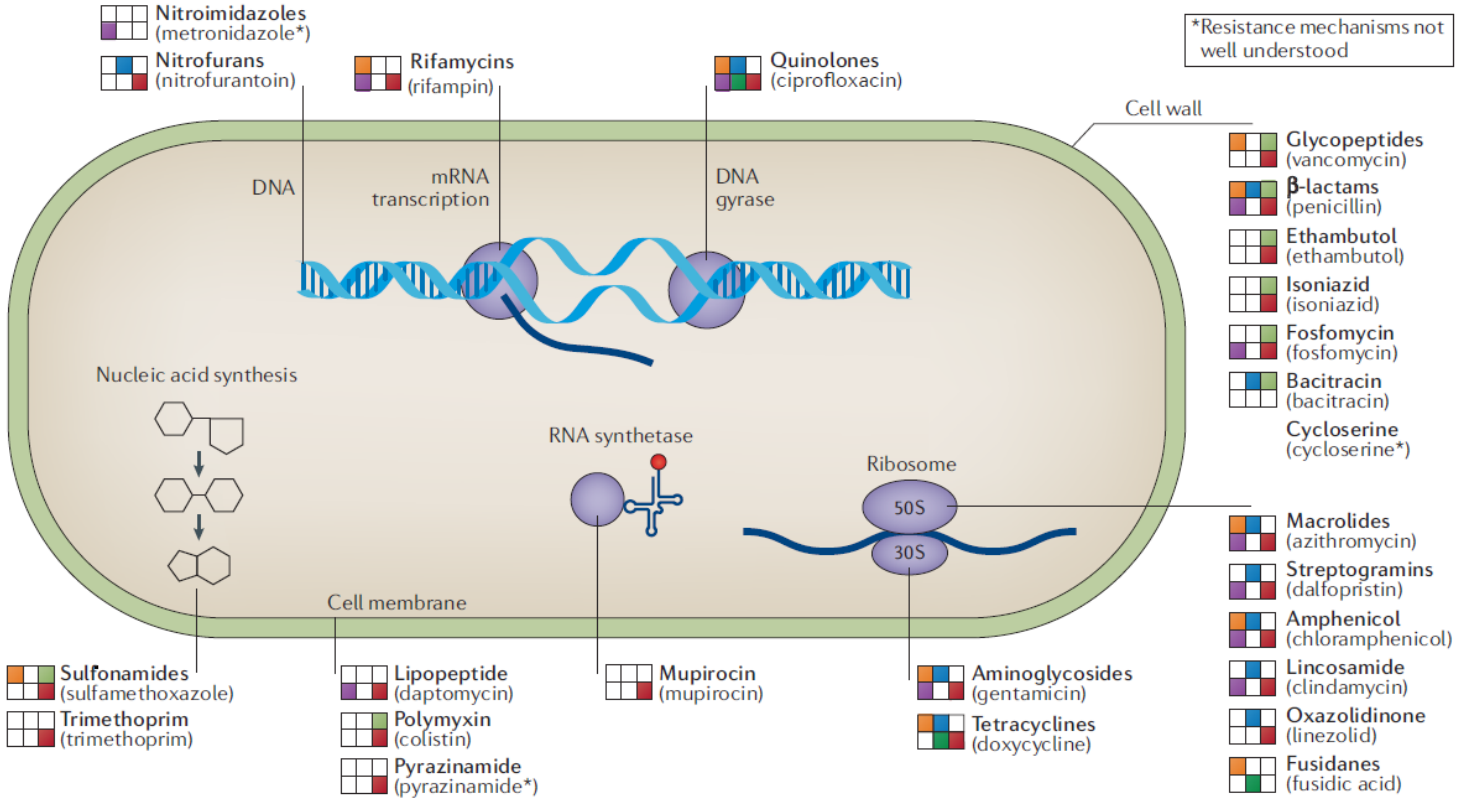
Outline for today's webinar

- Brief background on antimicrobial resistance (AMR)
- How genotype (NGS data) can help predict phenotype (resistance)
- Two popular algorithms for AMR detection
- Mutational resistance

- Disclaimer: I borrow extensively from other resources, including journal articles and online resource. These are referenced in the slides and used for educational purposes only. All references can be found in the Paperpile library (https://paperpile.com/shared/BbQJT2).

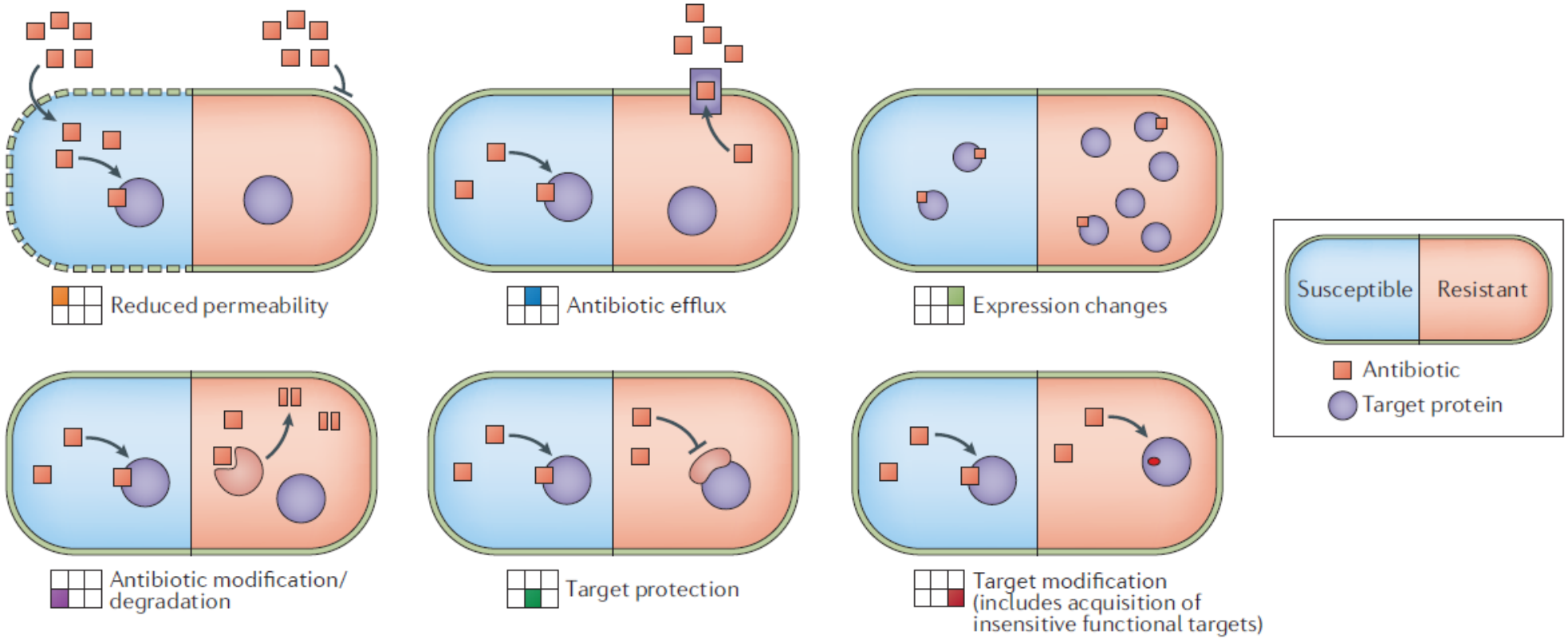
This webinar will be recorded and will be available at http://www.staphb.org/training/amr

Antibiotic Targets in Bacterial Cells



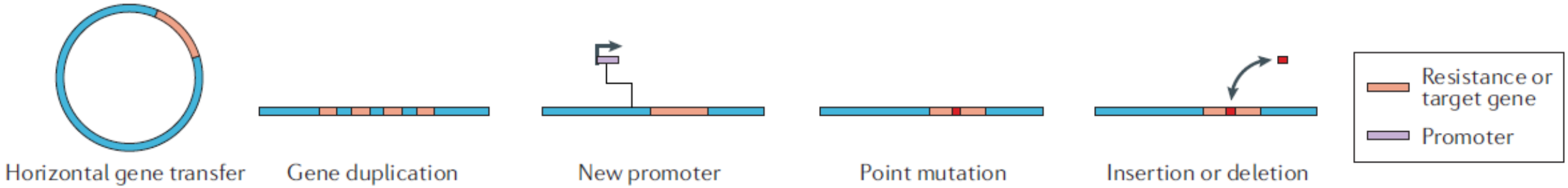
Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

Mechanisms of Antibiotic Resistance



Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

Genetic Determinants of Resistance



Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

Big Picture of AMR Detection Using NGS

- Generally speaking, detection requires well curated databases of AMR genes or mutations.
  - There are many different databases, most general, some species specific.
  - Most of these genetic determinants have been correlated with phenotype and/or experimentally determined.
- There are different tools available for searching these databases.
  - Tools vary on how they search the databases, and sometimes are part of the database implementation.
    - Input: contigs vs reads, nucleotide vs protein
    - Algorithm: Blast, k-mer, HMM, and others

Databases

Database	Organisms	Description	Link	Status	Reference
CARD	General	- Ontology-based database that provides comprehensive information of AR genes and their resistance mechanisms - Currently contains >2,200 protein homologues and includes a curated set of resistance-conferring chromosomal mutations in protein-coding genes	https://card.mcmaster.ca/	Active; launched in 2013; updated monthly	Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45, D566–D573 (2017).
ResFinder	General	Collation of AR genes involved in HGT events	https://cge.cbs.dtu.dk/services/ResFinder/	Active; started in 2012; update regularly ; last update in February 2019	Zankari, E. et al. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644 (2012). This article describes Resfinder, a widely used tool for the identification of acquired antimicrobial resistance genes in whole-genome data.
ResFinderFG	General	Collection of resistance gene variants identified in multiple functional metagenomics studies	https://cge.cbs.dtu.dk/services/ResFinderFG/	Active; last update in November 2016	Munk, P. et al. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. Nat. Microbiol. 3, 898–908 (2018).
Resfams	General	A profile HMM-based curated database confirmed for AR function	http://www.dantaslab.org/resfams/	Active; last update in January 2015	Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 9, 207–216 (2015).
ARDB	General	- First centralized resource of AR gene information - Manually curated; contains >4,500 AR sequences	https://ardb.cbcb.umd.edu/	Archived; last updated in 2009	Liu, B. & Pop, M. ARDB—antibiotic resistance genes database. Nucleic Acids Res 37, D443–D447 (2009).
MEGARes	General	- Collation of multiple databases (CARD, ARG-ANNOT and ResFinder) to avoid redundancy between entries - For high-throughput screening and statistical analysis	https://megares.meglab.org/	Active; last update in December 2016	Lakin, S. M. et al. MEGARes: an antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res. 45, D574–D580 (2017).
NDARO	General	- Collated and curated data from multiple databases (CARD, Lahey , Pasteur Institute β-Lactamases and ResFinder) - Contains 4,500 AR sequences	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047	Active; started in 2016	
ARG-ANNOT	General	- Repository of >1,800 AR sequences collated from scientific literature	Not available	Archived; last update in	Gupta, S. K. et al. ARG-ANNOT, a new bioinformatic tool to discover

.....

- Go to http://www.staphb.org/training/amr/

Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

Tools

Name	Input	Description	Accessibility	Year	Link	Status	Reference
ResFinder	Assembly/Reads	Tool for detecting acquired AR genes from sequenced or partially sequenced bacterial isolates	Web and/or standalone	2012	https://cge.cbs.dtu.dk/services/ResFinder/	Active	Zankari, E. et al. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644 (2012). This article describes Resfinder, a widely used tool for the identification of acquired antimicrobial resistance genes in whole-genome data.
ARG-ANNOT	Assembly	Tool for pairwise comparison of query sequence with ARG-ANNOT database	Standalone	2014	NA	Archived	Gupta, S. K. et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob. Agents Chemother. 58, 212–220 (2014).
RGI	Assembly	- Pairwise comparison of query sequence with the CARD - Uses curated AR detection models to predict intrinsic resistance genes, dedicated resistance genes and acquired resistance from mutations in drug targets	Web and/or standalone	2015	https://card.mcmaster.ca/analyze/rgi	Active	Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45, D566–D573 (2017).
ARGs-OAP (v2)	Assembly	- Online analysis pipeline for AR genes - Detection from metagenomic data using an integrated structured database of AR sequences	Web and/or standalone	2016	https://galaxyproject.org/use/args-oap/	Active	Yin, X. et al. ARGs-OAP v2.0 with an expanded SARG database and hidden Markov models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics 34, 2263–2270 (2018).
ARIBA	Reads	Tool for rapid AR genotyping directly from sequencing reads using curated public databases	Standalone	2017	https://github.com/sanger-pathogens/ariba	Active	Hunt, M. et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb. Genom. 3, e000131 (2017).
PointFinder	Assembly	Web tool for WGS-based detection of AR associated with chromosomal point mutations in bacterial pathogens	Web and/or standalone	2018	https://cge.cbs.dtu.dk/services/ResFinder/	Active	Zankari, E. et al. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72, 2764–2768 (2017).
NCBI-AMRFinderPlus	Assembly	Tool for identification of acquired resistance genes using NCBI's curated AR database and curated collection of HMMs	Standalone	2018	https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/	Active	
SRST2	Reads	Tool for direct mapping of reads to curated AR databases	Standalone	2014	https://github.com/katholt/srst2	Active	Inouye, M. et al. SRST2: rapid genomic surveillance

.....

- Go to http://www.staphb.org/training/amr/

Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

ABRicate (Blast-based nucleotide contig searching)

- ABRicate website
  - https://github.com/tseemann/abricate
- ResFinder Databases
  - https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/

ABRicate - Results

- ABRicate will output one file for each database searched:

```
-rw-r--r--  1 ubuntu ubuntu  2658 Dec 23 03:53 SRR9988086_argannot.tab
-rw-r--r--  1 ubuntu ubuntu 25407 Dec 23 03:57 SRR9988086_card.tab
-rw-r--r--  1 ubuntu ubuntu   295 Dec 23 04:00 SRR9988086_ecoh.tab
-rw-r--r--  1 ubuntu ubuntu 15364 Dec 23 04:03 SRR9988086_ecoli_vf.tab
-rw-r--r--  1 ubuntu ubuntu  1223 Dec 23 04:09 SRR9988086_ncbi.tab
-rw-r--r--  1 ubuntu ubuntu   488 Dec 23 04:14 SRR9988086_plasmidfinder.tab
-rw-r--r--  1 ubuntu ubuntu  1511 Dec 23 04:17 SRR9988086_resfinder.tab
-rw-r--r--  1 ubuntu ubuntu 42044 Dec 23 04:21 SRR9988086_vfdb.tab
```

ABRicate - Results

- Each file is a tab delimited file with the Blast results:

```
(base) ubuntu@ip-172-31-7-175:~/workspace/fourth_test/abricate$ more SRR9988086_resfinder.tab
#FILE   SEQUENCE        START   END     STRAND  GENE    COVERAGE        COVERAGE_MAP    GAPS    %COVERAGE       %I
DENTITY       DATABASE        ACCESSION       PRODUCT RESISTANCE
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00001     569447  569884  +       aa
c(6')-Iaa_1   1-438/438       ================ 0/0     100.00  100.00  resfinder       NC_003197 aac(6')-Iaa      A
mikacin;Tobramycin
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00013     64484   65702   +       md
f(A)_1        1-1219/1233     ================ 0/0     98.86   79.41   resfinder       Y08743  mdf
(A)
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00029     10268   11473   +       te
t(B)_2        1-1206/1206     ================ 0/0     100.00  100.00  resfinder       AF326777  tet(B)   Doxycycli
ne;Tetracycline;Minocycline
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00034     136     972     -       ap
h(6)-Id_1     1-837/837       ================ 0/0     100.00  100.00  resfinder       M28829  aph
(6)-Id  Streptomycin
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00034     972     1775    -       ap
h(3'')-Ib_5   1-804/804       ================ 0/0     100.00  100.00  resfinder       AF321551  aph(3'')-Ib      S
treptomycin
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00034     1836    2651    -       su
l2_3  1-816/816       ================ 0/0     100.00  100.00  resfinder       HQ840942        sul
2       Sulfamethoxazole
/home/ubuntu/workspace/fourth_test/shovill/SRR9988086/SRR9988086.fa     contig00043     649     1509    +       bl
aTEM-1B_1     1-861/861       ================ 0/0     100.00  100.00  resfinder       AY458016  blaTEM-1B        A
moxicillin;Ampicillin;Cephalothin;Piperacillin;Ticarcillin
```

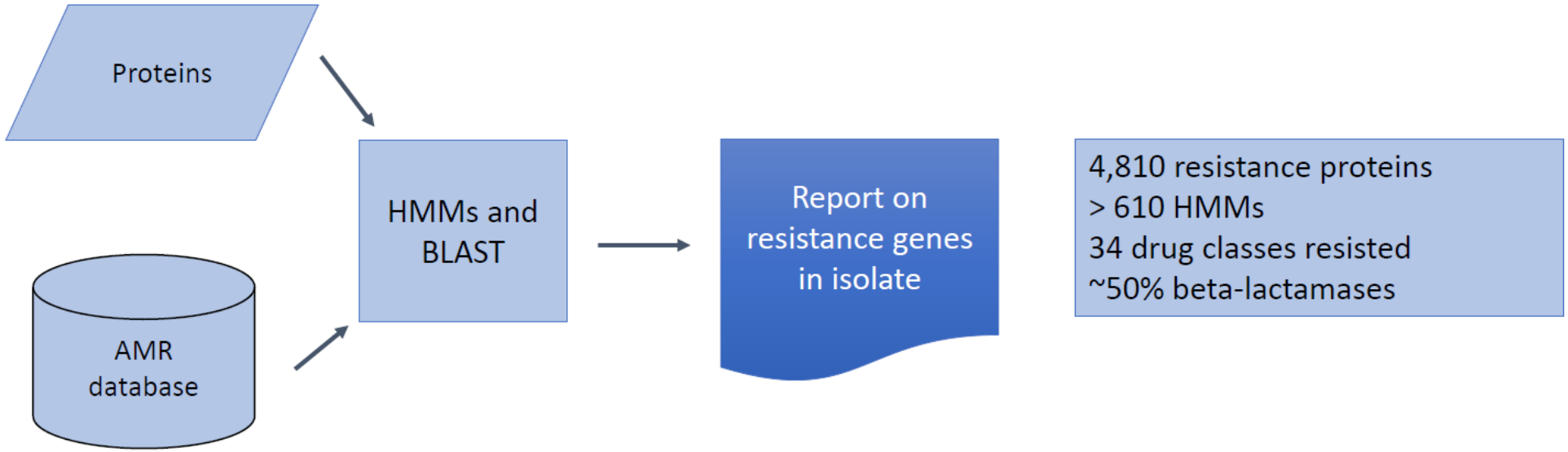
ABRicate - Results

- Let’s look at several results files in formatted view:
  - Google Sheets View

William Klimke
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

AMRFinderPlus

AMRFinder Uses a Curated Database, HMMs and BLAST to Identify AMR genes

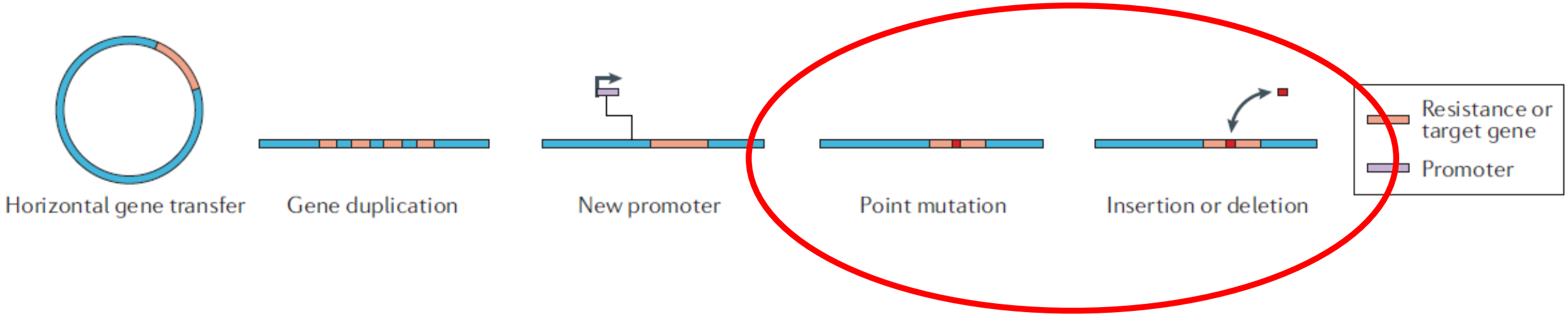


AMRFinderPlus - Results

- https://www.ncbi.nlm.nih.gov/pathogens/isolates/#/search/SRR9988086

Mutational Resistance

- Best example is *Mycobacterium tuberculosis*
  - Drug resistance is predominantly mutational.



Cite this article as: Miotto P, Tessema B, Tagliani E, *et al.* A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J* 2017; 50: 1701354 [https://doi.org/10.1183/13993003.01354-2017].

TB Mutational Resistance

TABLE 1 Overview of the data included in the study

		Collected data			Studies	
	Loci of interest	Total isolates	Isolation time frame years	Countries represented	Screened	Included
Rifampicin (R)	*rpoB*	13424	1999–2014	37	459	95
Isoniazid (H)	*katG*	11847	1992–2014	42	650	127
	inhA-mabA	9407				
	furA	361				
	mshA	288				
Ethionamide and prothionamide (ETO/PTO)	*inhA-mabA*	346				
	ethA	181				
	mshA	117				
Ofloxacin (OFX)	*gyrA*	5911	1991–2013	36	243	75
	gyrB	3078				
Moxifloxacin (MFX)	*gyrA*	1019				
	gyrB	735				
Levofloxacin (LFX)	*gyrA*	449				
	gyrB	218				
Pyrazinamide (Z)	*pncA*	4949	1990–2014	36	378	81
Streptomycin (S)	*rpsL*	3263	1985–2013	43	423	104
	tap	0				
	rrs	2598				
	whiB7	0				
	gidB	812				
Amikacin (AM)	*rrs*	2105				
Capreomycin (CM)	*rrs*	2533				
	tlyA	1854				
Kanamycin (KM)	*rrs*	1727				
	eis	2029				
	whiB7	56				

Data are presented as n. Inclusion and exclusion criteria for individual studies are reported in online supplementary material 2.

Cite this article as: Miotto P, Tessema B, Tagliani E, *et al.* A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J* 2017; 50: 1701354 [https://doi.org/10.1183/13993003.01354-2017].

TB Mutational Resistance

TABLE 3 List of confidence-graded mutations associated with phenotypic drug resistance as determined by best confidence values

Drug (phenotypic testing)	Gene	High-confidence mutations	Moderate-confidence mutations	Minimal-confidence mutations	No association with resistance
First-line					
Rifampicin (R)	*rpoB*	F505V+D516Y, S512T, Q513H+L533P, Q513-F514ins, **Q513K**, **Q513L**, **Q513P**, **F514dupl**, M515I+D516Y, **D516A**, **D516F**, **D516G**, D516G+L533P, D516ins, D516N, **D516V**, Del N518, **S522Q**, **H526C**, **H526D**, H526F, **H526G**, **H526L**, **H526R**, **H526Y S531F**, **S531L**, S531Q, **S531W**, S531Y, D626E	**D516Y**, **S522L**, H526P, **L533P**	**L511P**, **H526N**, **I572F**	
Isoniazid (H)	*inhA-mabA*	g-102a[#,¶]	**c-15t**		**g-102a**[#,¶], t-80g, *g-47c*, **T4I**
	katG	S315I, **S315N**, **S315T**, pooled frameshifts and premature stop codons			A110V, **R463L**, L499M
	furA				**L68F**
	mshA		A187V[#,¶]		**N111S**
Second-line (group A)					
Moxifloxacin (MFX)	*gyrA*	G88C, **A90V**, **S91P**, **D94A**, **D94G**, **D94N**, **D94Y**			E21Q, **S95T**, G247S, G668D, V712L
Ofloxacin (OFX)/ levofloxacin (LFX)	*gyrA*	**G88A**, **G88C**, **S91P**, **A90V**, **D94A**, **D94G**, **D94H**, **D94N**, **D94Y**	D89N		E21Q, **T80A**, **S95T**, G247S, G668D, V712L
	gyrB	E459K, **A504V**			
Second-line (group B)					
Amikacin (AM)	*rrs*	**a1401g**, **g1484t**			
Kanamycin (KM)	*eis*	**c-14t**, **g-10a**		g-37t, c-12t	**a1338c**
	rrs	a514c[#], **a1401g**, c1402t, g1484t			
	rrs+eis	*rrs* c517t[#] + *eis* g-37t			
Capreomycin (CM)	*rrs*	**a1401g**, **c1402t**, **g1484t**			**c517t**
	tlyA	**N236K**, **pooled frameshifts and premature stop codons**			D149H
Streptomycin (S)	*rpsL*	**K43R**, K43T, **K88Q**, **K88R**, T40I			
	rrs	a1401g[#], **a514c**, a514t, c462t, c513t, **c517t**			
	gidB		**E92D**[#,¶]		**L16R**, **V110G**, pooled frameshifts and premature stop codons
Second-line (group C)					
Ethionamide and prothionamide (ETO/PTO)	*inhA*	c-15t+I194T, c-15t+S49A	**c-15t**		
	ethA				Q347Stop
Second-line (group D)					
Pyrazinamide (Z)	*pncA*	t-12c, **a-11g**, t-7c, **A3E**, L4S, **I6T**, V7G, D8E, **D8G**, D8N, **Q10P**, **D12A**, **D12N**, **C14R**, G17D, L19P, G24D, **Y34D**, **A46V**, **K48T**, **D49G**, D49N, **H51Q**, H51R, P54S, **H57D**[¶], H57P, **H57R**, **H57Y**, S59P, P62L, P62Q, **D63G**, S66P, **S67P**, W68C, **W68R**, **H71D**, H71Q, **H71Y**, **C72R**, **T76P**, **H82R**, **L85P**, **L85R**, **F94L**, **F94S**, K96N, K96R, G97C, **G97D**, **G97S**, **Y103H**, **S104R**, **G108R**, L116P, **L116R**, **L120P**, R123P, V125F, **V125G**, **V128G**, G132A, G132D, G132S, A134V, **T135N**, **T135P**, H137P, C138Y, **V139G**, **V139L**, **Q141P**, **T142A**, **T142K**, **T142M**, **indel - R148ins (inframe)**, L151S, **V155G**, L159P, T160P, G162D, T168P, **L172P**, M175T, **M175V**, V180F, V180G, **Pooled frameshifts and premature stop codons**	**V7G**, Q10R, **P54L**, **W68G**, K96E, K96T, A171E, M175I	**D12G**, F58L, H71R, I133T, **V139A**	**indel - c-125del**, **I31T**, **L35R**, **T47A**, **I6L**, K48T, T114M

The table includes all the mutations graded according to the proposed standardised approach for providing confidence levels to their association with phenotypic drug resistance. Standard type represents associations based on nominal p-values (putative); bold type represents associations based on corrected p-values. The rationale for pooling insertions/deletions and nonsense mutations can be found in online supplementary material 5. Tables 1 and 2 provide the details of the data included in the grading system and the definitions for the confidence categories. Indeterminate mutations were not included in the table and can be found in online supplementary material 8. Drugs were classified based on the updated guidelines for short and individualised regimens [4]. [#]: six associations were not considered for further analysis as there was probably no causative relationship between these genetic changes and the resistance to the antibiotic in question; [¶]: genotype-specific mutation.

TB Mutational Resistance

ARTICLE
DOI: 10.1038/s41467-017-00721-2 OPEN

A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide

Adam N. Yadon[1,2], Kashmeel Maharaj[2], John H. Adamson[2], Yi-Pin Lai[3], James C. Sacchettini[4], Thomas R. Ioerger[3], Eric J. Rubin[1] & Alexander S. Pym[2]

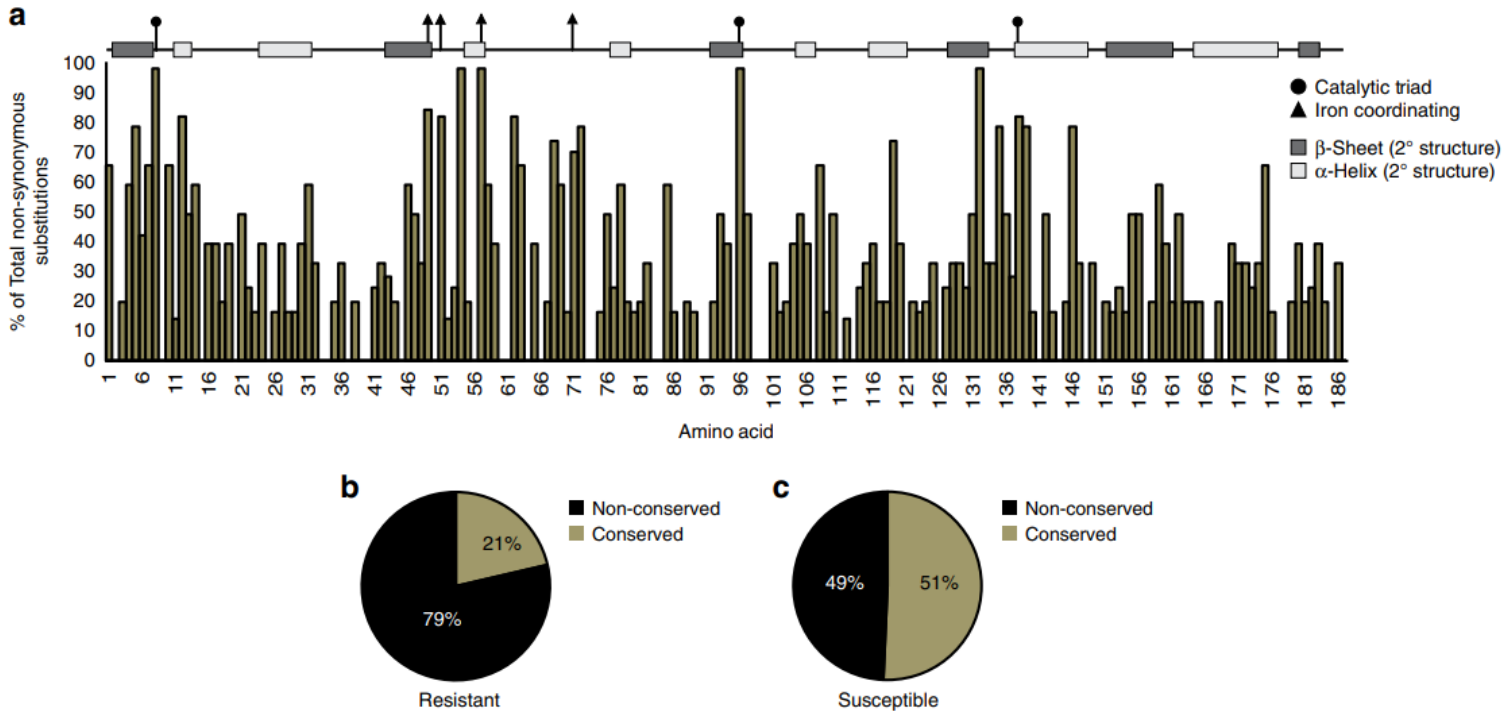


Fig. 4 Pyrazinamide resistant amino acid substitutions occur throughout PncA and are enriched for non-conserved amino acid substitutions. **a** The proportion (%) of non-synonymous amino acid substitutions represented in the *pncA* library that were pyrazinamide resistant at each PncA amino acid. Amino acids corresponding to the catalytic triad are marked with a *circle*. Amino acids responsible for iron coordination are marked with a *triangle*. **b** Percentage (%) of non-conserved (*black*) and conserved (*brown*) pyrazinamide resistant amino acid substitutions. **c** Percentage (%) of non-conserved (*black*) and conserved (*brown*) pyrazinamide susceptible amino acid substitutions

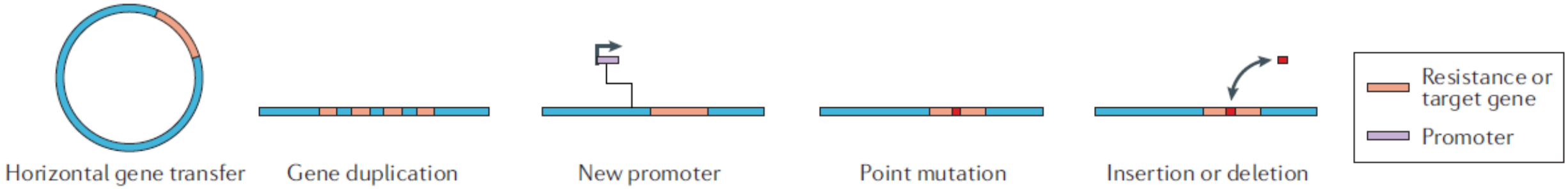
TB Mutational Resistance

- Very active area of research
- Databases currently being curated
- Potential for regional differences in curated databases
- Often data will be used for patient care, thus there is a lot of recalcitrance in sharing workflows until there is greater confidence in results

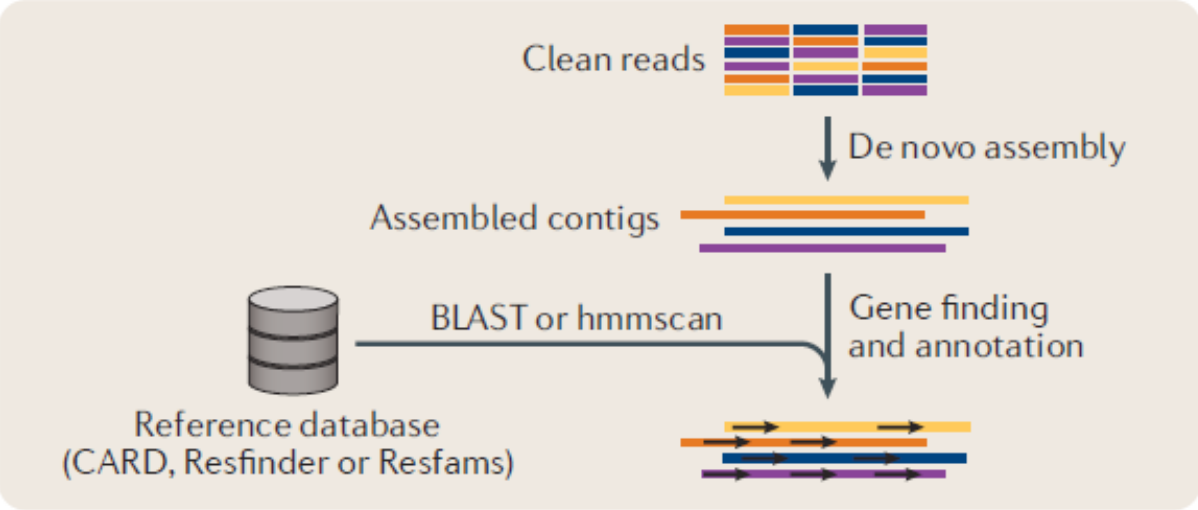
Take Away Messages

- Well curated databases of both genes and mutations are essential for effective AMR detection.
- Mutational resistance is often species specific.
- Resources are constantly being updated as new genomic and phenotypic data is curated.
- Collaboration between bioinformaticists, microbiologists, and epidemiologists are essential for fully understanding AMR.
- Although not discussed in this webinar, specificity and sensitivity of AMR detection using NGS data is often very high, comparable or better than many other molecular diagnostics for AMR detection (see references).

Questions?


Horizontal gene transfer
Gene duplication
New promoter
Point mutation
Insertion or deletion
Resistance or target gene
Promoter

Reads vs Contigs

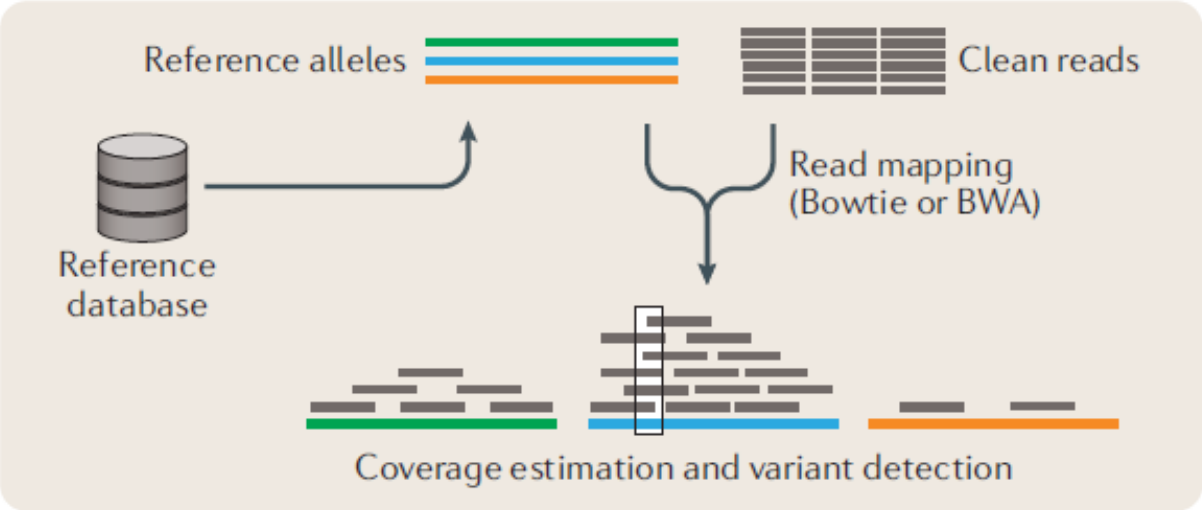


Assembly-based approach:

- Computationally expensive and time consuming, particularly in resistome profiling of large complex communities
- Identification of both known and novel resistance genes that share low similarity with reference database; however, requires high genome coverage
- Genomic context such as regulatory and mobile element sequences can be captured

Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4

Reads vs Contigs



Read-based approach:

- Fast and less computationally demanding, enabling resistome analysis of large data sets
- Identification of resistance genes is dependent on completeness of the reference database of organisms under analysis
- Nearby genes and genomic context are missing; may lead to false positives due to spurious mapping

Boolchandani, M., D'Souza, A.W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet* **20,** 356–370 (2019). https://doi.org/10.1038/s41576-019-0108-4