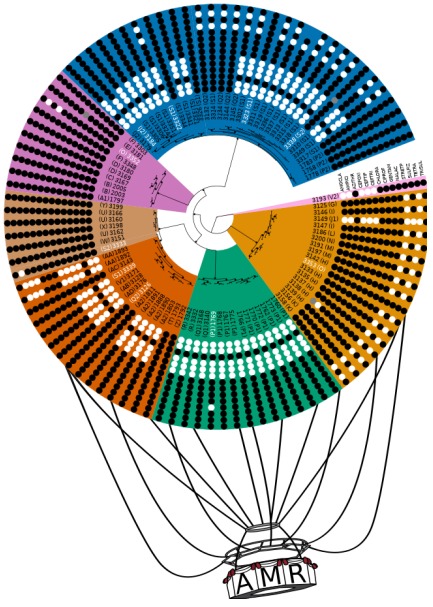
Around the resistome in 80 ways:

an empirical evaluation of antimicrobial resistance gene detection methods



Finlay Maguire
finlaymaguire@gmail.com

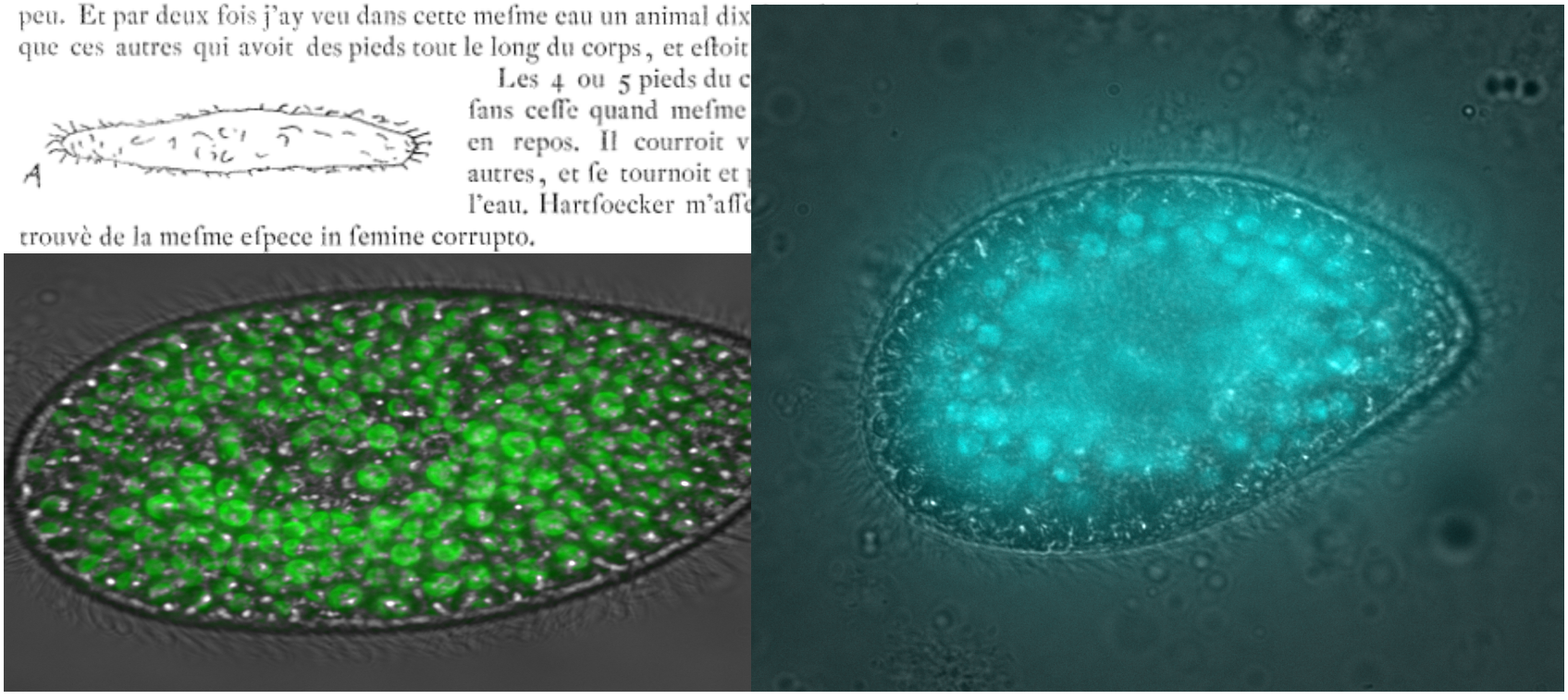
December 2, 2019

Faculty of Computer Science, Dalhousie University

Table of contents

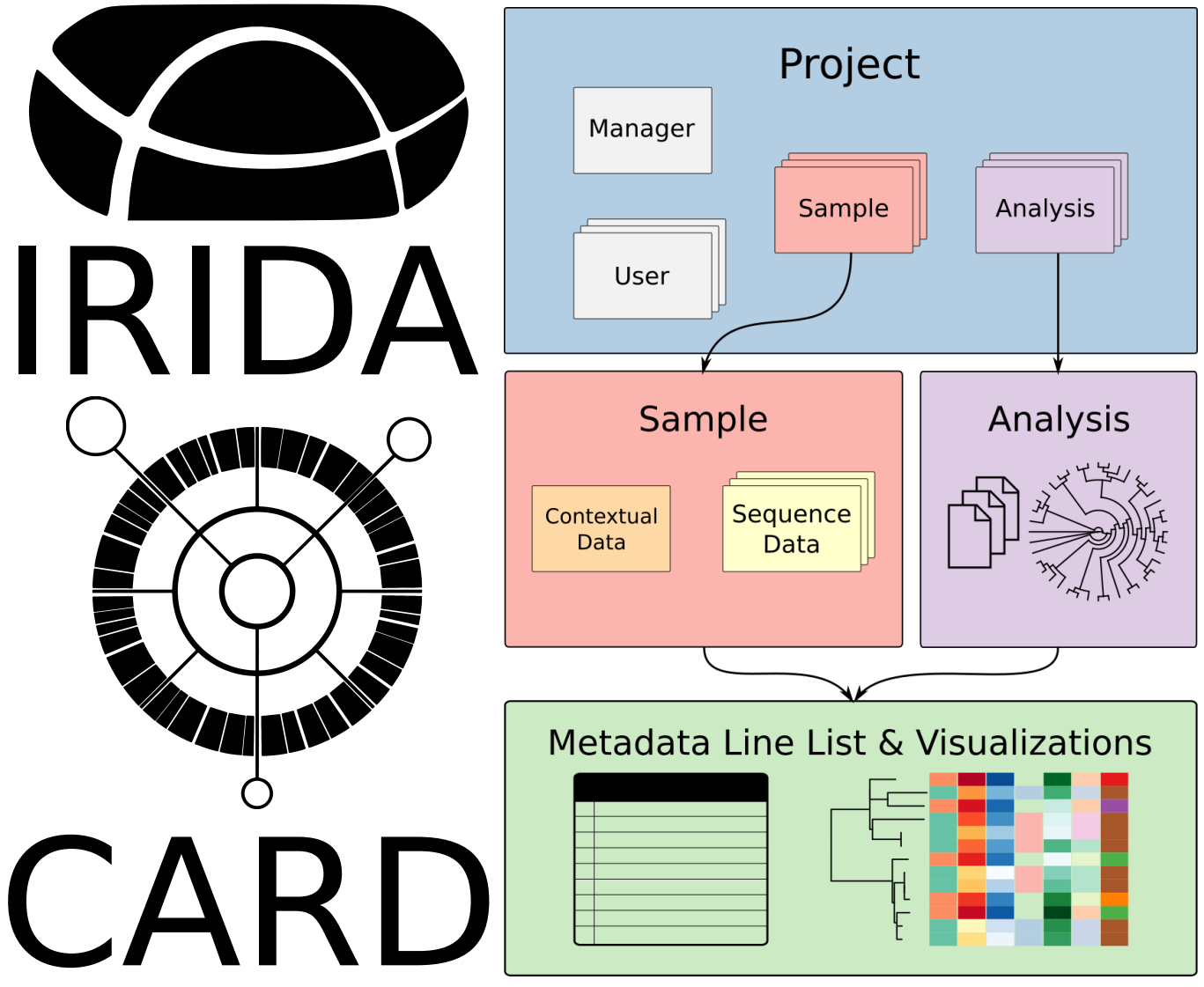
Background

Evolution of Eukaryotic Endosymbioses



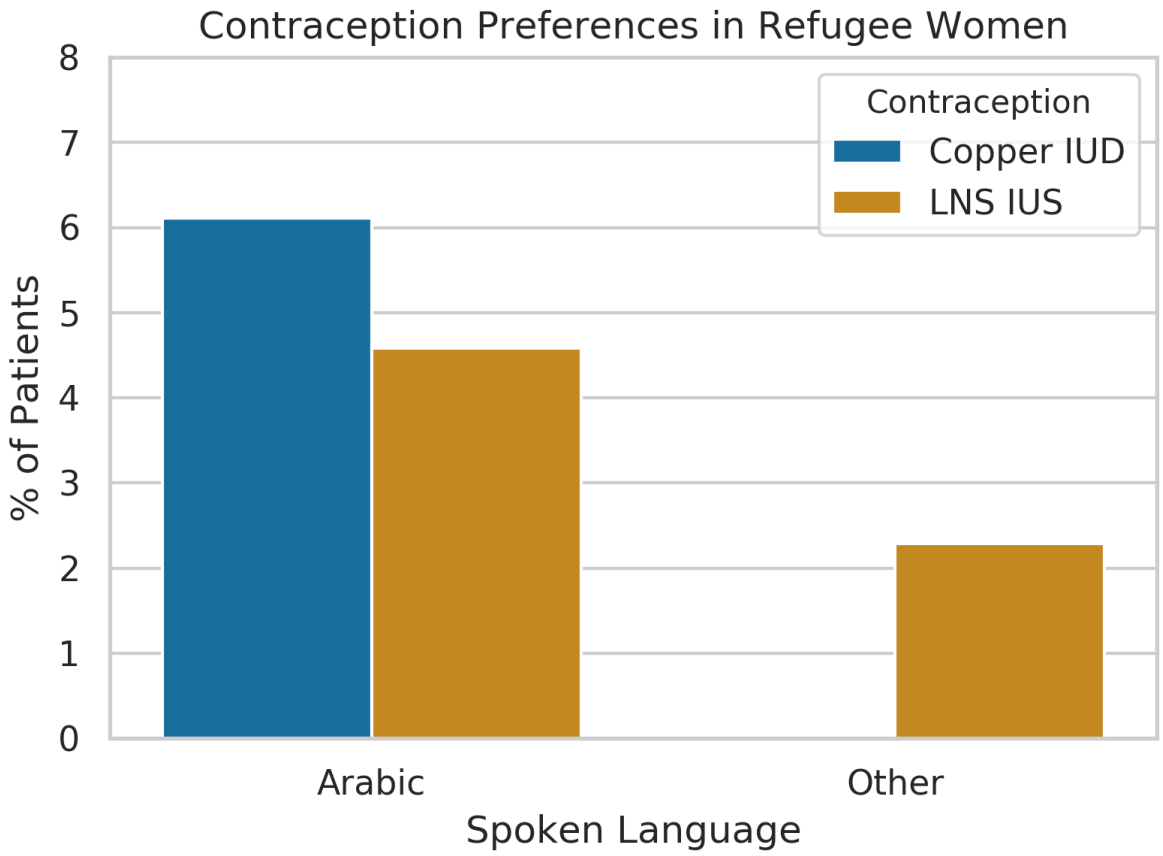
(Maguire, 2016)

Antimicrobial Resistance



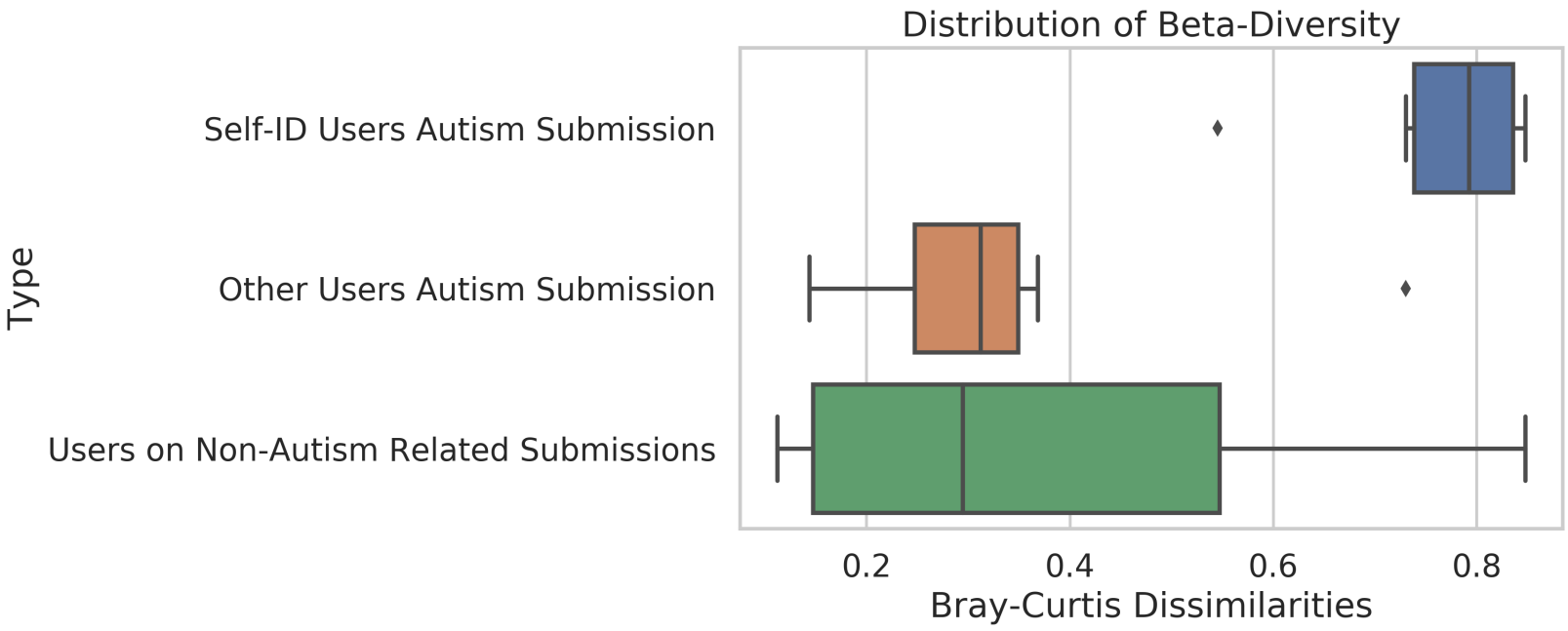
(Matthews et al., 2018)

Epidemiology



(Stairs et al., 2019)

Sociology?


Distribution of Beta-Diversity
Type
Self-ID Users Autism Submission
Other Users Autism Submission
Users on Non-Autism Related Submissions
0.2
0.4
0.6
0.8
Bray-Curtis Dissimilarities

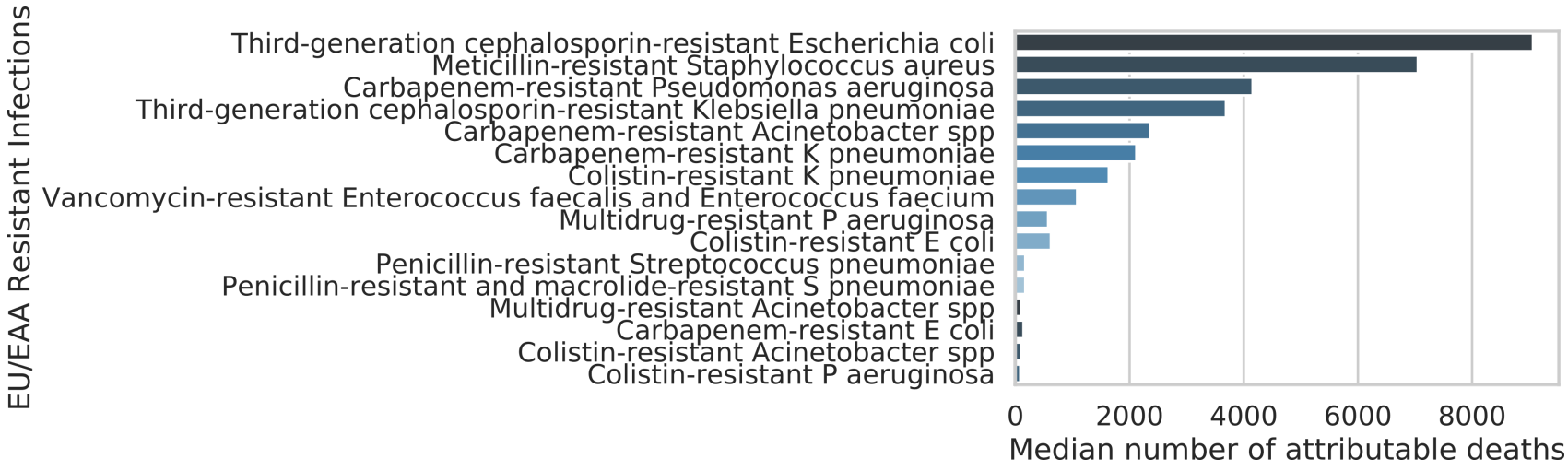
Sociology?

Congratulations, your application to the SSHRC Explore Grants competition has been awarded.

Project Title: NEETs, Incels, and Wizards: The Experiences of Socially Isolated Men

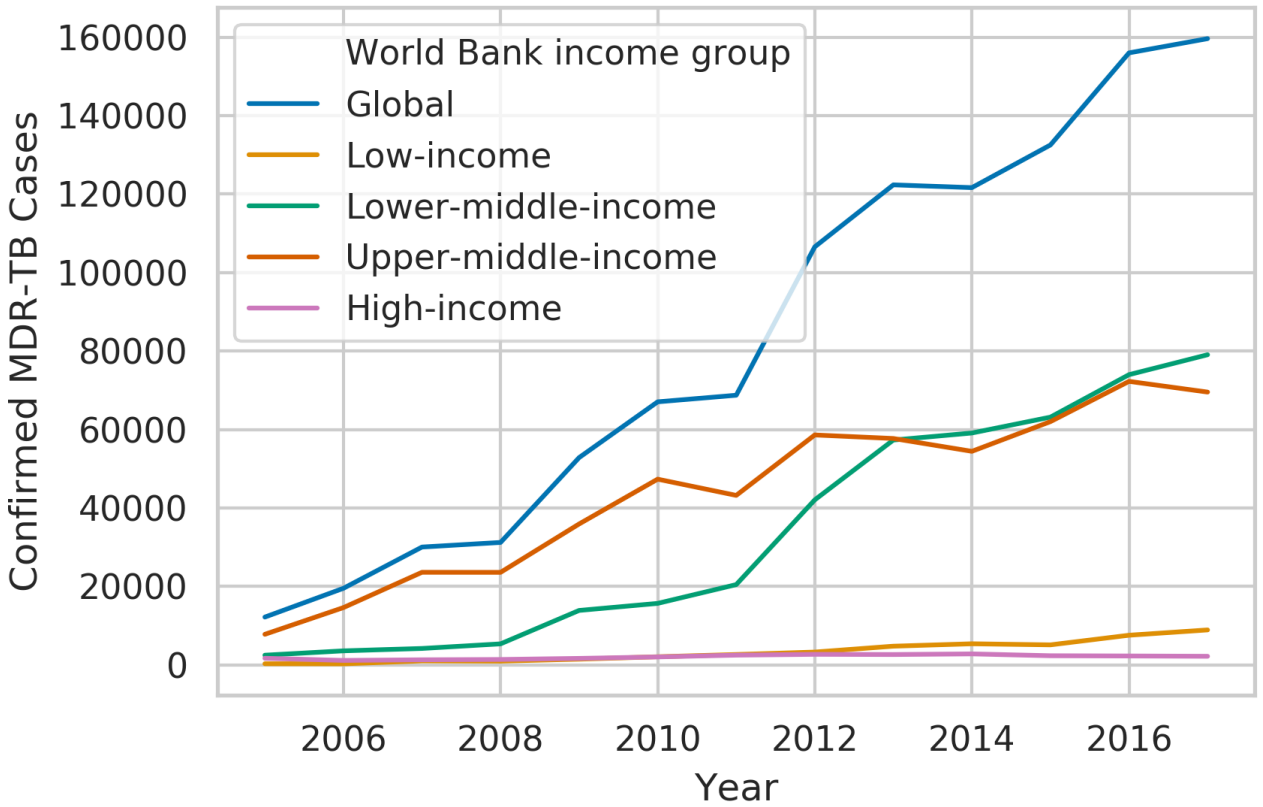
Why do we care about AMR?

AMR is currently a problem



2015 EU/EEA: 33,110 deaths, Data from (Cassini et al., 2019).

AMR is growing



WHO Global Health Observatory Data Repository.

What can we do about it?

Improve surveillance

- Locally: information would help improve patient health.

Improve surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.

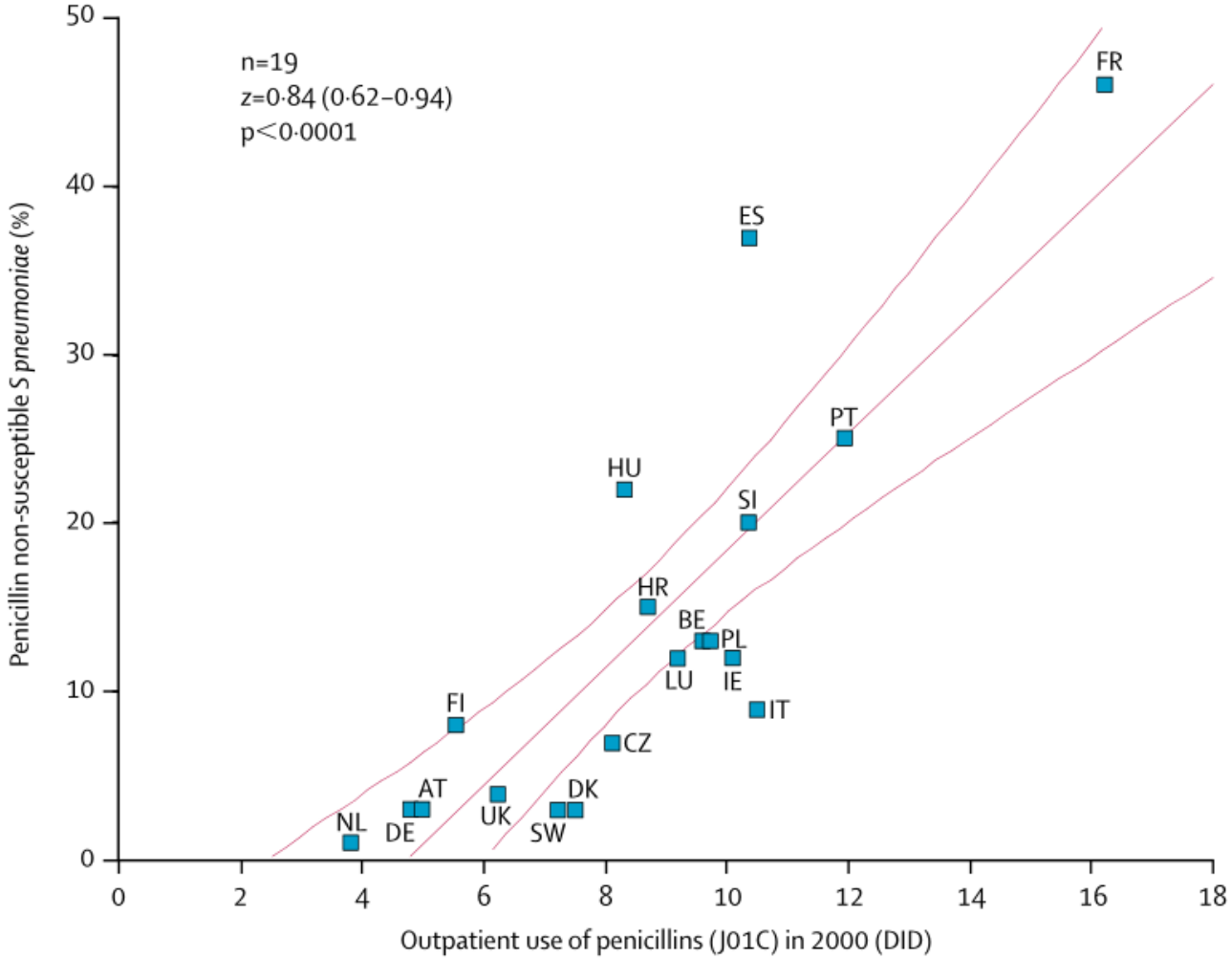
Improve surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.
- Globally: emerging threats and long–term trends.

Improve surveillance

- Locally: information would help improve patient health.
- Nationally: health policies and responses to emergencies.
- Globally: emerging threats and long–term trends.
- Scientifically: better understanding of underlying biology.

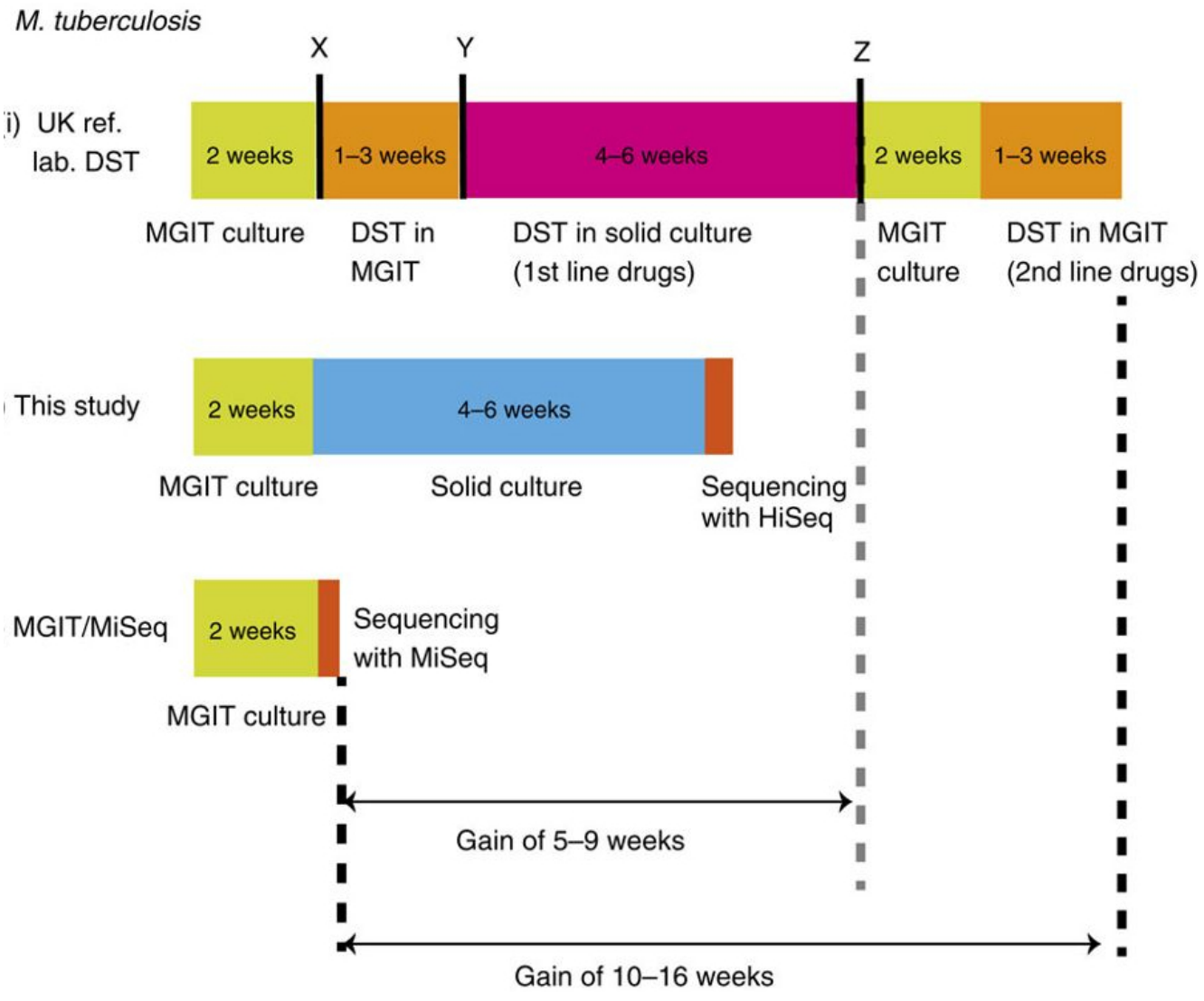
Improve diagnostics



(Goossens et al., 2005)

How do we do this?

Phenotypically?



(Bradley et al., 2015)

DNA sequencing

- DNA is relatively tractable and stable
- Sequencing technology is mature
- Represents the substrate of evolution

Downside of DNA: capacity not expression

E. coli gene regulatory networks are inconsistent with gene expression data

Simon J Larsen ✉, Richard Röttger, Harald H H W Schmidt, Jan Baumbach

Nucleic Acids Research, Volume 47, Issue 1, 10 January 2019, Pages 85–92,

Downside of DNA: capacity not expression

Random sequences rapidly evolve into de novo promoters

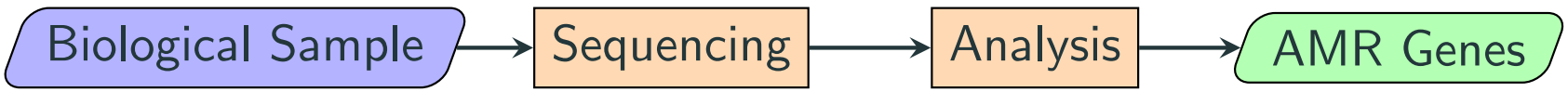
Avihu H. Yona, Eric J. Alm & Jeff Gore

Nature Communications **9**, Article number: 1530 (2018) | Cite this article

- 10% of random sequences can serve as active promoters
- 60% of random sequences can modulate expression with only one mutation

Which DNA sequencing method?

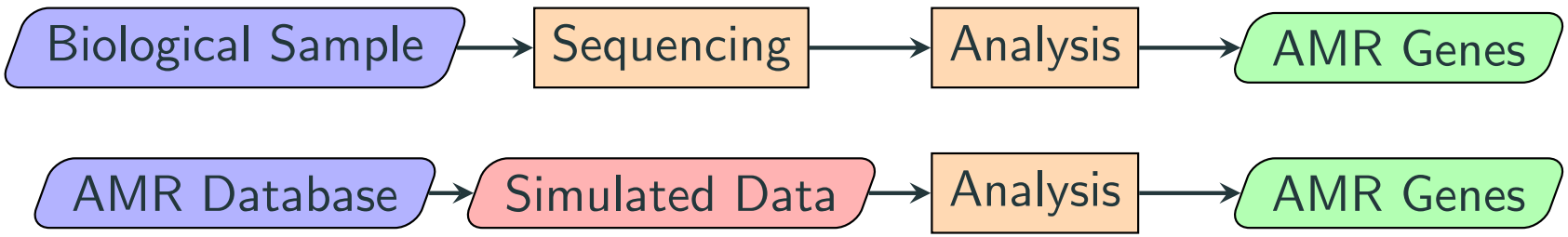
Choosing a method



Additional factors:

- Does method provide other information?
- Cost/experimental considerations

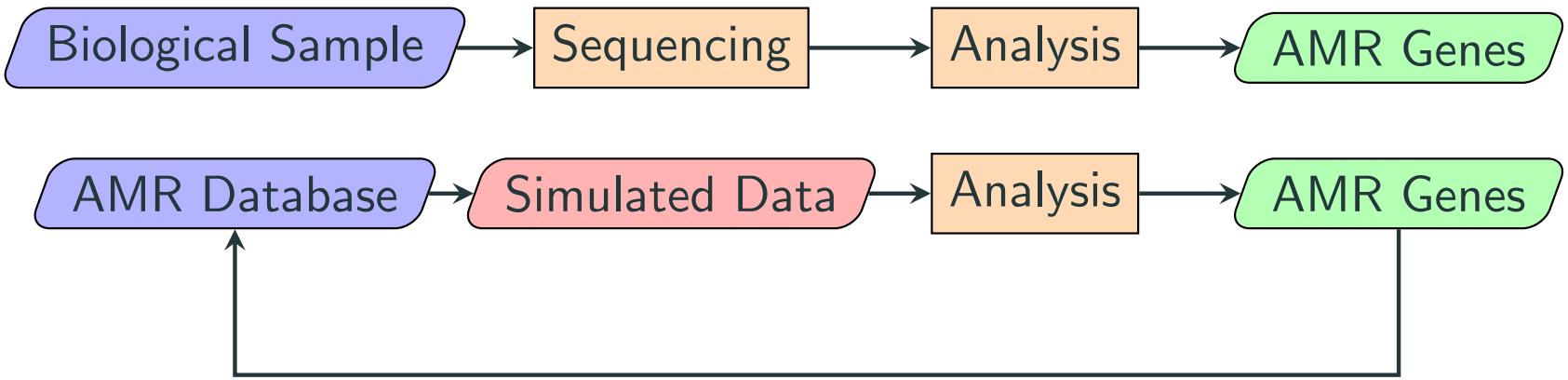
Choosing a method



Additional factors:

- Does method provide other information?
- Cost/experimental considerations

Choosing a method

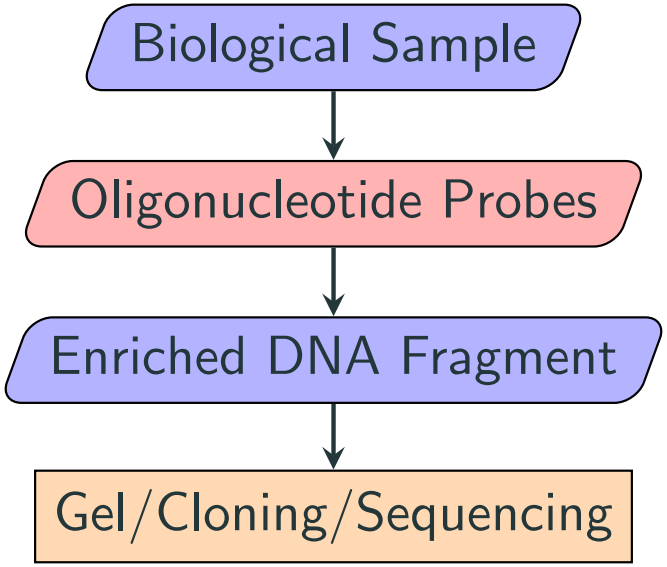


Additional factors:

- Does method provide other information?
- Cost/experimental considerations

Targeted sequencing

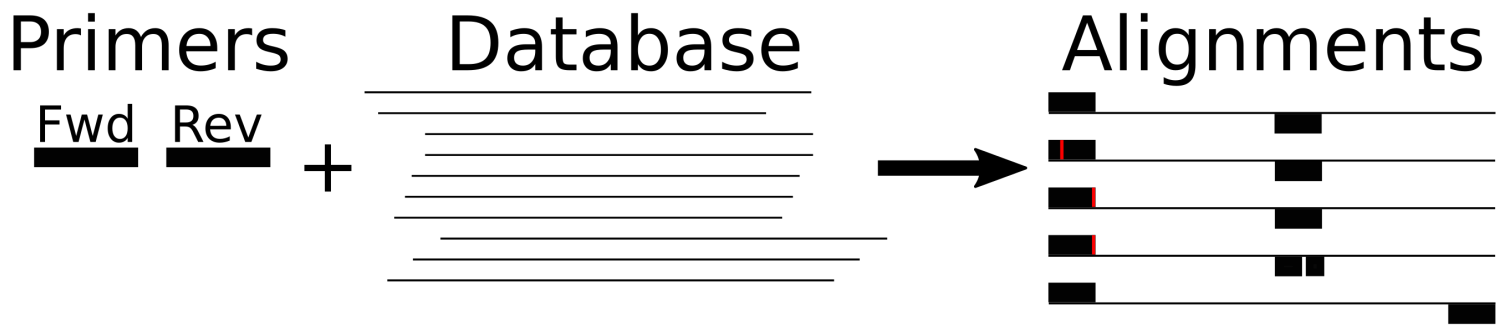
Targeted sequencing



- Cheap/simple infrastructure
- Multiple sample types
- Low input requirements

Choosing and evaluating primers

Testing primers computationally



`github.com/mwhall/VAware`: Mike Hall

Needleman-Wunsch alignments:

- Perfect: no mismatches, insert < 1500
- Intermediate: (1-2 minor mismatches)
- Low: (2-4 minor; 0-1 major - terminal, gaps)
- Missed: (> 4 minor; > 1 major)

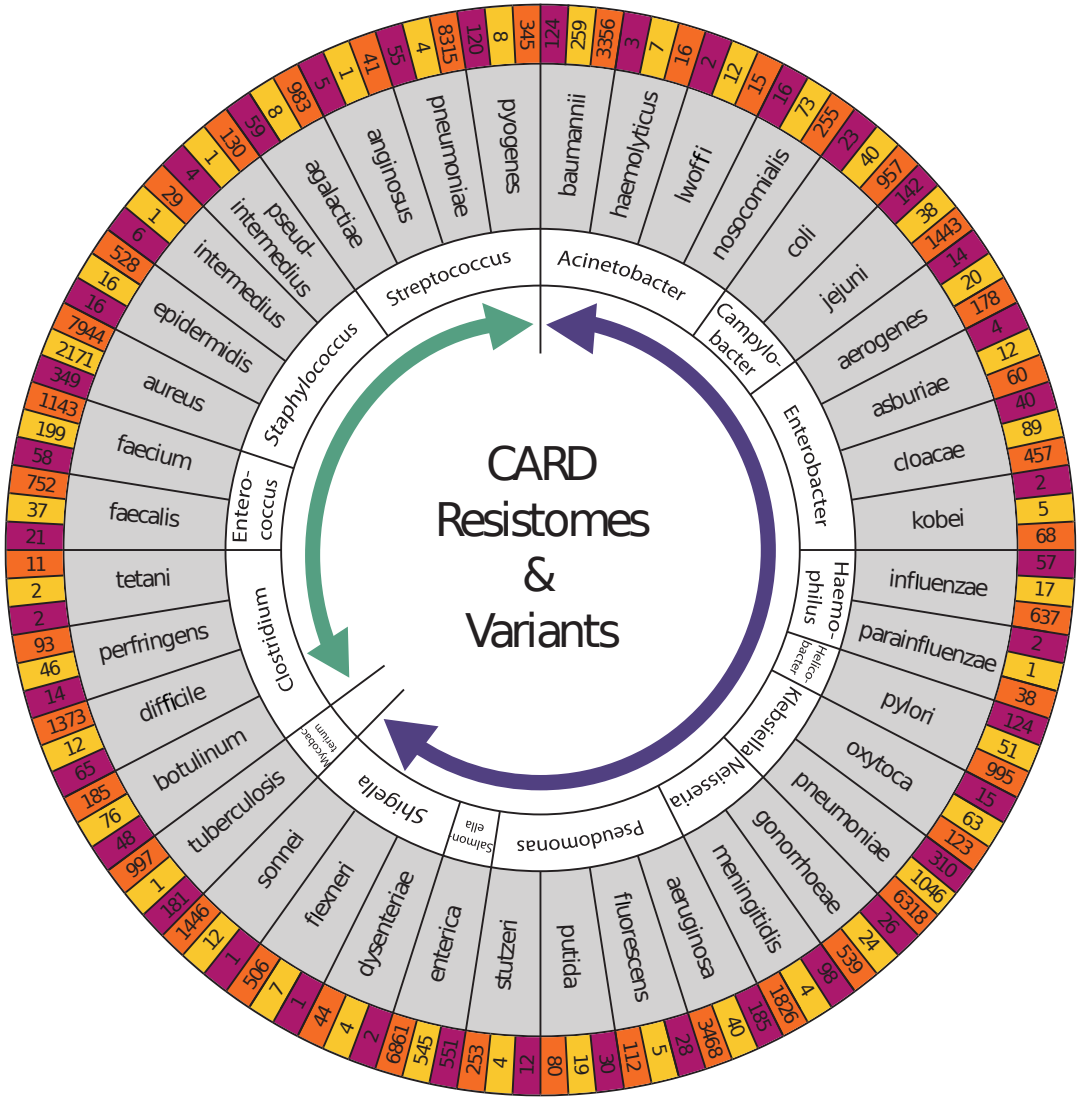
Which primers?

List of primers for detection of antimicrobial resistance genes

Antimicrobial	Group	Gene/Codon no.	Primer name	Internal number	Sequence	Temperature (°C)	Reference
Betalactams	TEM	All	TEM front P1	Primer 757	5'-GCGGAACCCCTATTTG-3'	55	Olesen, I., H. Hasman, and F. M. Aarestrup. 2004. Prevalence of beta-lactamases among ampicillin-resistant Escherichia coli and Salmonella isolated from food animals in Denmark. Microb.Drug Resist. 10:334-340.Moodley, A. and Guardabassi, L.Transmission of IncN Plasmids Carrying blaCTX-M-1 between Commensal Escherichia coli in Pigs and Farm Workers. Antimicrobial Agents and Chemotherapy. 2009. 53:1709-1711.
			TEM-C-R-ny	Primer 686	5'-ACC AAT GCT TAA TCA GTG AG-3'		
	CTX	M-All	ctx M U1	Primer 1354	5'-ATGTGCAGYACCAGTAARGTKATGGC-3'	60	Hasman, H., D. Mevius, K. Veldman, I. Olesen and F. M. Aarestrup. 2006. Beta-lactamases among Extended spectrum Beta-lactamase resistant (ESBL) Salmonella from poultry, poultry products and human patients in The Netherlands. J. Antimicrob. Chemother. 56:115-121.
			CTX-M-U-2new	Primer 1580	5'-TGGGTRAARTARGTSACCAGAAYSAGCGG-3'		Hendriksen R. S., Mikoleit M., Kornschober C., Rickert R.L., Duyne S.V., Kjelsø C., Hasman H., Cormican M., Mevius D., Threlfall J., Angulo F. J., Aarestrup F. M. 2009. Emergence of Multidrug-Resistant Salmonella Concord Infections in Europe and the United States in Children Adopted From Ethiopia, 2003-2007. Pediatr Infect Dis J. 28:814-818.
		CTX-M1 group	ctx-M-15 front P1	Primer 1537	5'-CCATGGTTAAAAAATCACTGCG-3'		Moodley, A. and Guardabassi, L.Transmission of IncN Plasmids Carrying blaCTX-M-1 between Commensal Escherichia coli in Pigs and Farm Workers. Antimicrobial Agents and Chemotherapy. 2009. 53:1709-1711.

European Committee on Antimicrobial Susceptibility Testing: 78 PCR Primer Sets

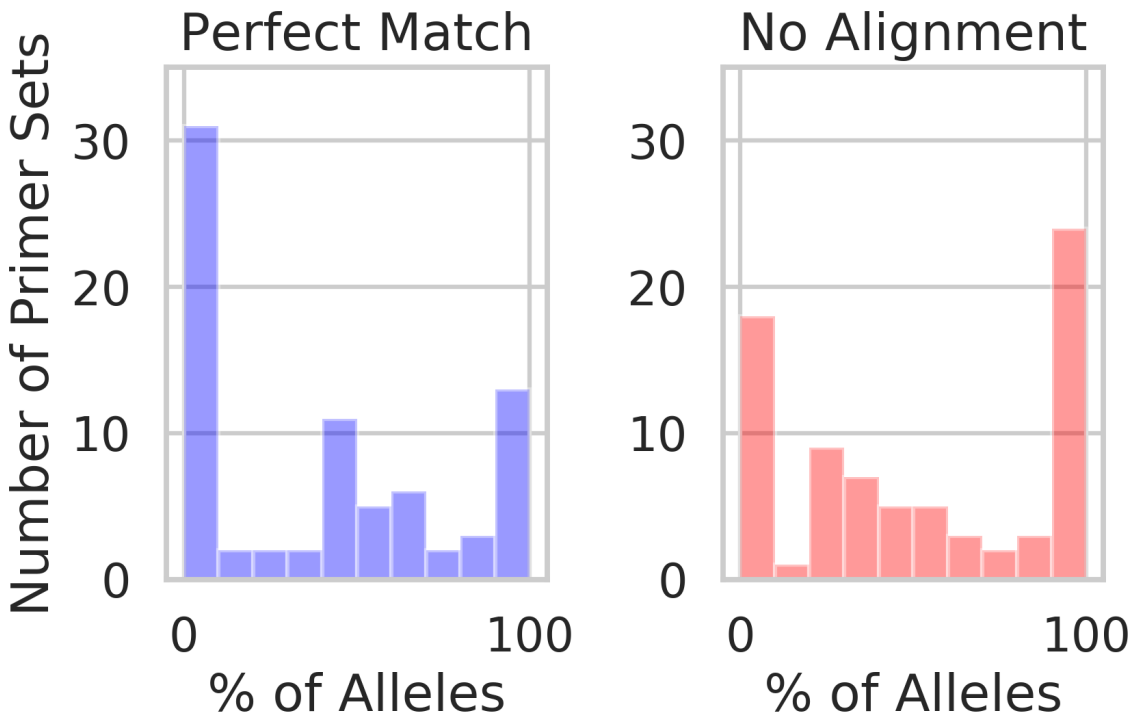
Which AMR genes?



CARD-prevalence: 85 pathogens, 116,914 resistomes (chromosome, plasmid, and WGS assembly). Brian Alcock/McArthur Lab

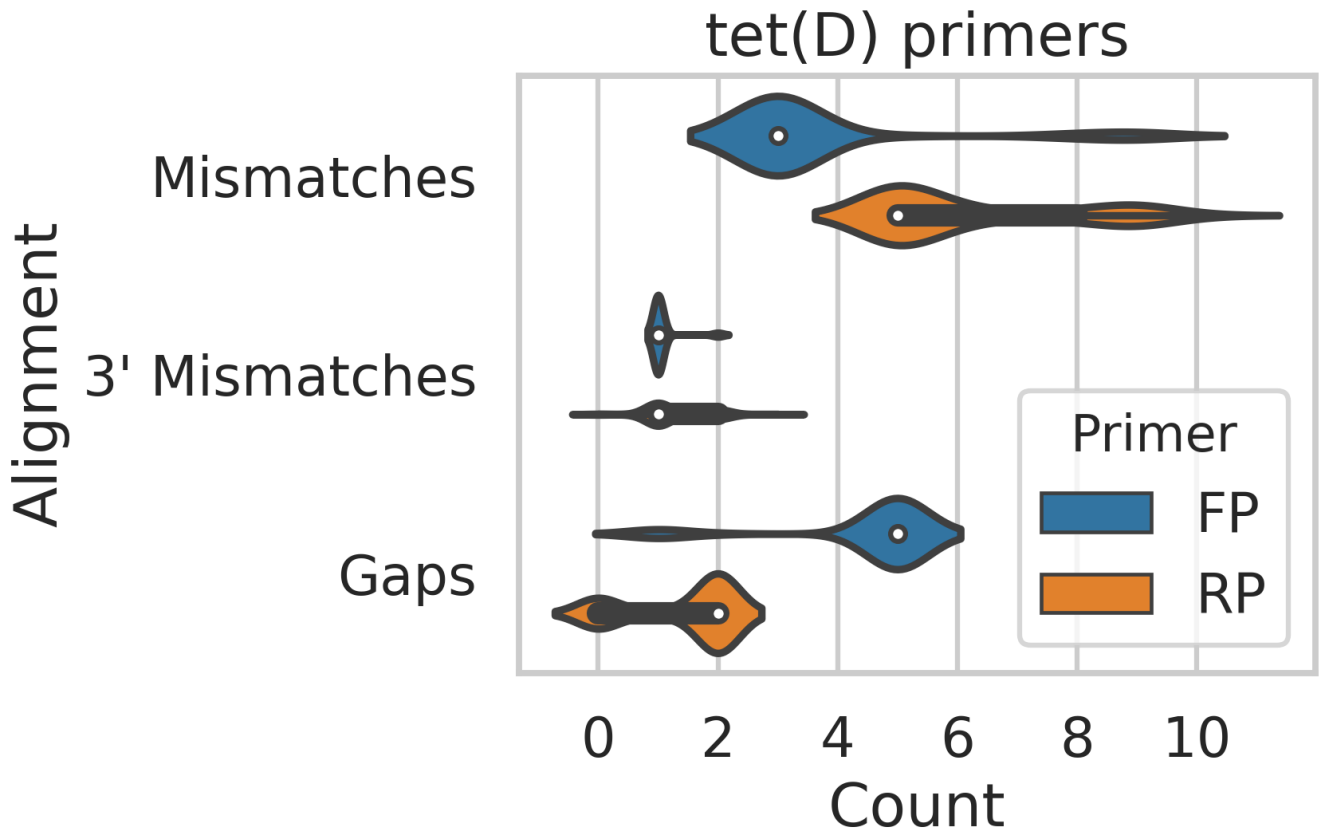
How well do these primers work?

Surprisingly poorly



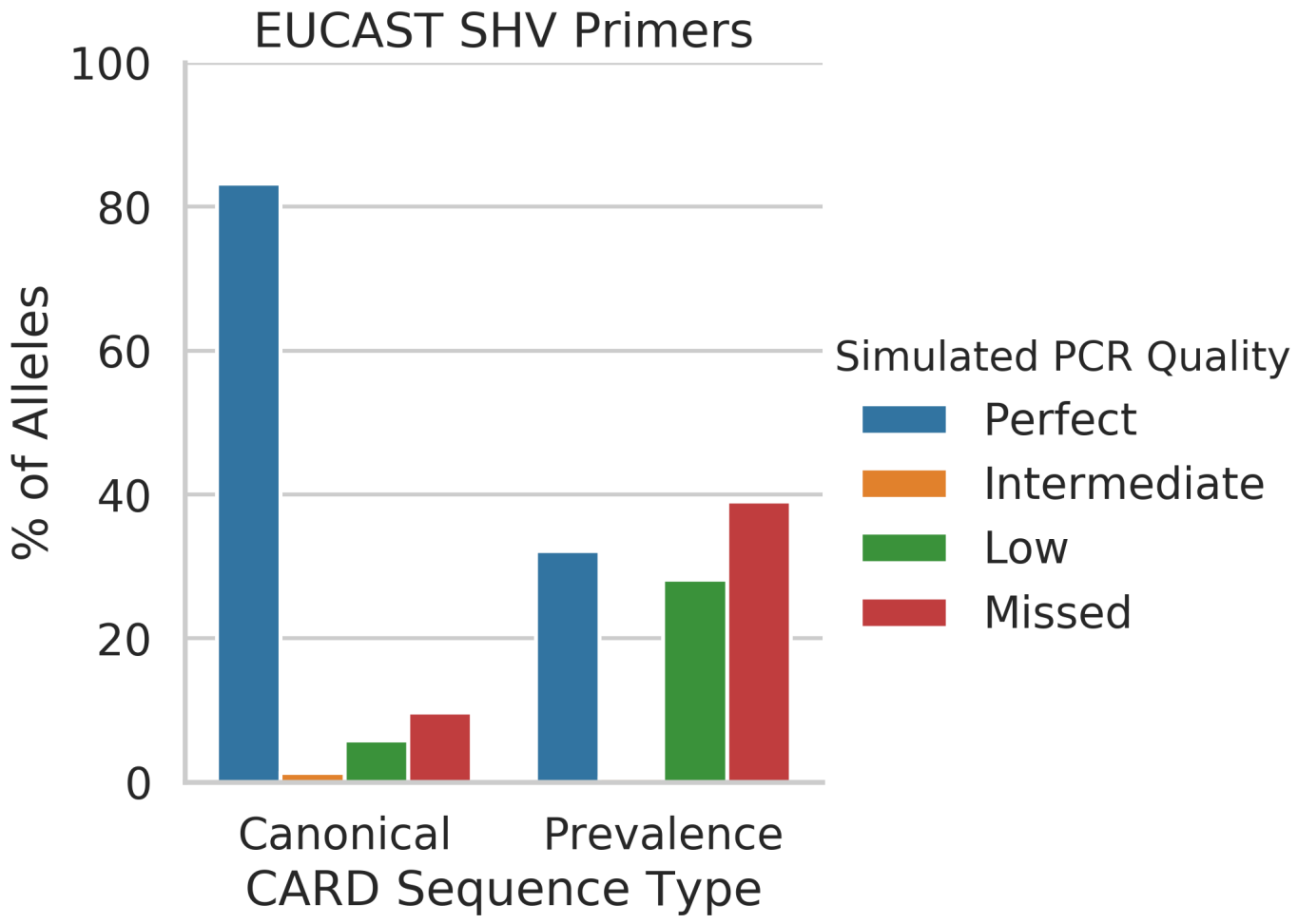
- Many aminoglycosides and tetracycline resistance genes totally missed
- Caveat: needs experimental validation

Lots of serious mismatches



No primer alignment in 27.58% of *tetD* alleles

Stagnation of primers



off-target hits (1 mismatch in RP) to *LEN-3*, *LEN-4*

Can we improve on this?

Designing probes with up-to-date AMR allele diversity

Antimicrobial Agents and Chemotherapy

Mechanisms of Resistance

Capturing the Resistome: A targeted capture method to reveal antibiotic resistance determinants in metagenomes

Allison K. Guitor, Amogelang R. Raphenya, Jennifer Klunk, Melanie Kuch, Brian Alcock, Michael G. Surette, Andrew G. McArthur, Hendrik N. Poinar, Gerard D. Wright

DOI: 10.1128/AAC.01324-19

(Guitor et al., 2019)

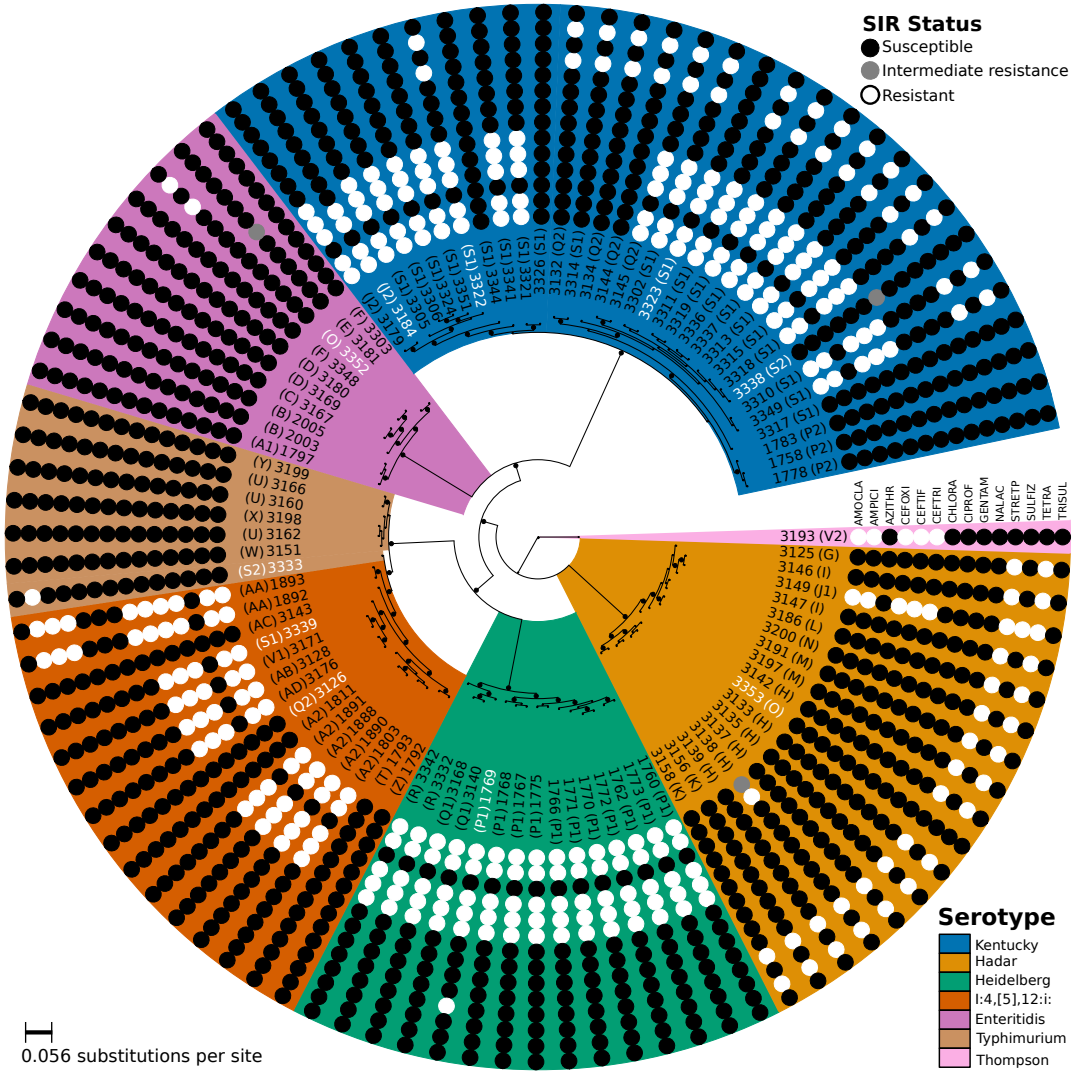
Downsides of targeted-approaches

- *a priori* target decisions
- Need constantly updated
- No easy genomic context
- No easy source-genome attribution

Why do we care about context?

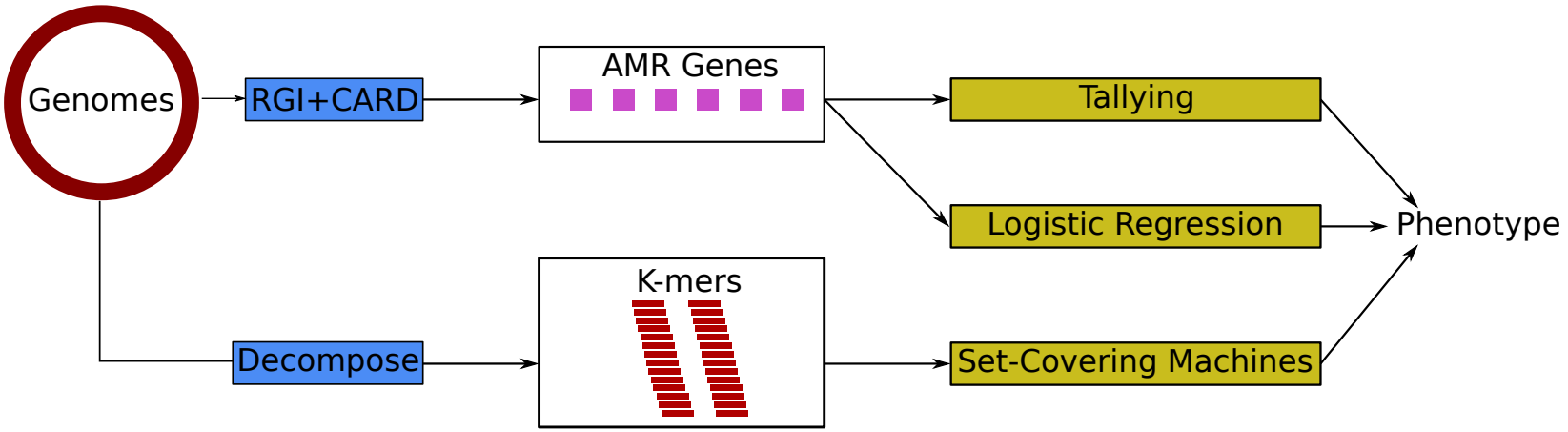
Genomics

Case-study on strengths of genomics



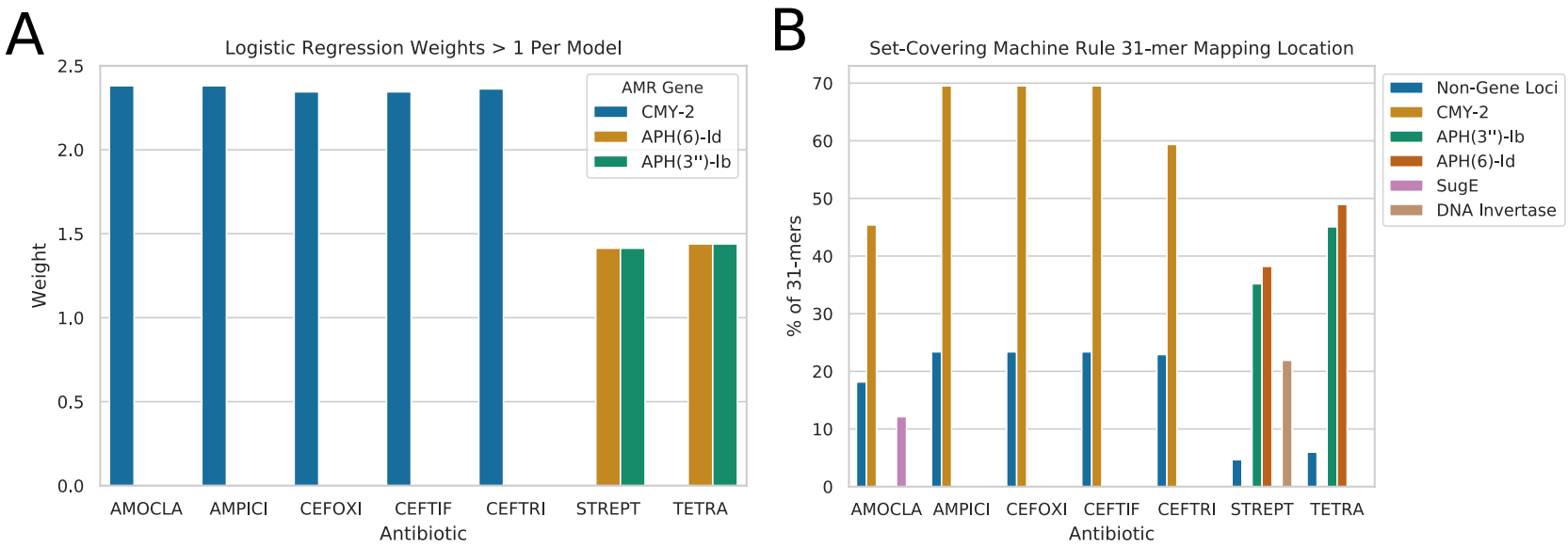
(Maguire et al., 2019)

Phenotype prediction modelling



(Maguire et al., 2019)

Genomes allow gene-free models



(Maguire et al., 2019)

Generate co-selection hypotheses

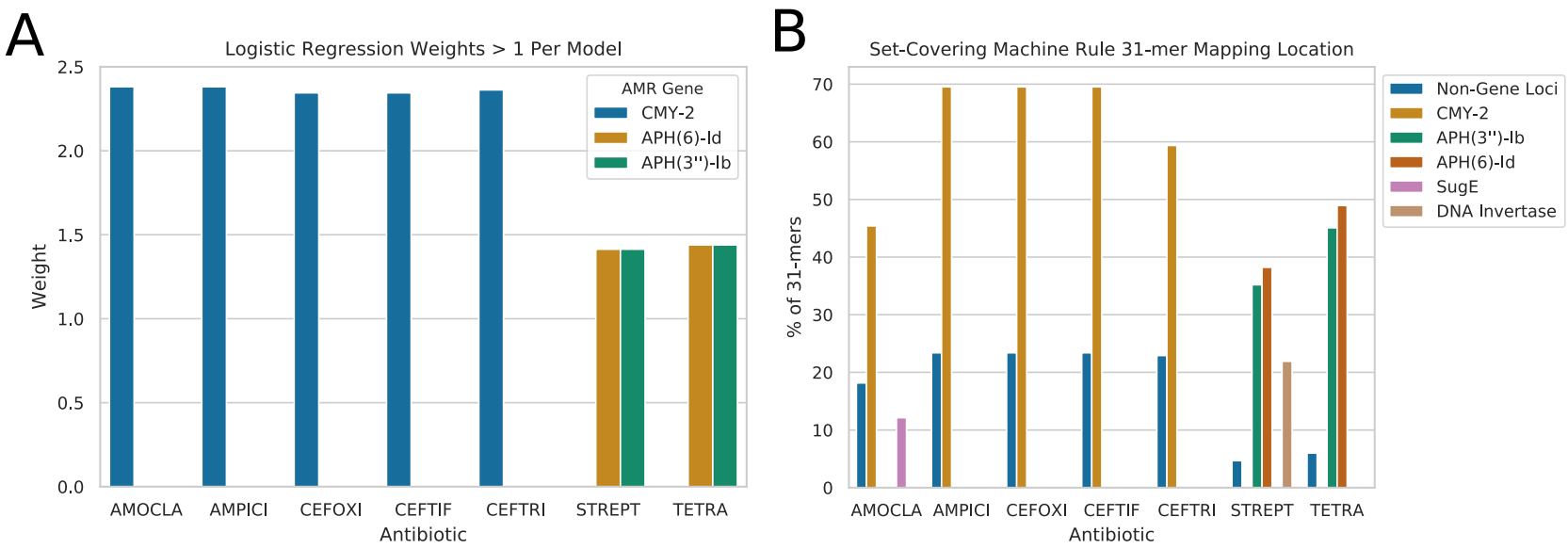
Appl Environ Microbiol. 2011 Jul;77(13):4486-93. doi: 10.1128/AEM.02788-10. Epub 2011 May 20.

Selection pressure required for long-term persistence of blaCMY-2-positive IncA/C plasmids.

Subbiah M[1], Top EM, Shah DH, Call DR.

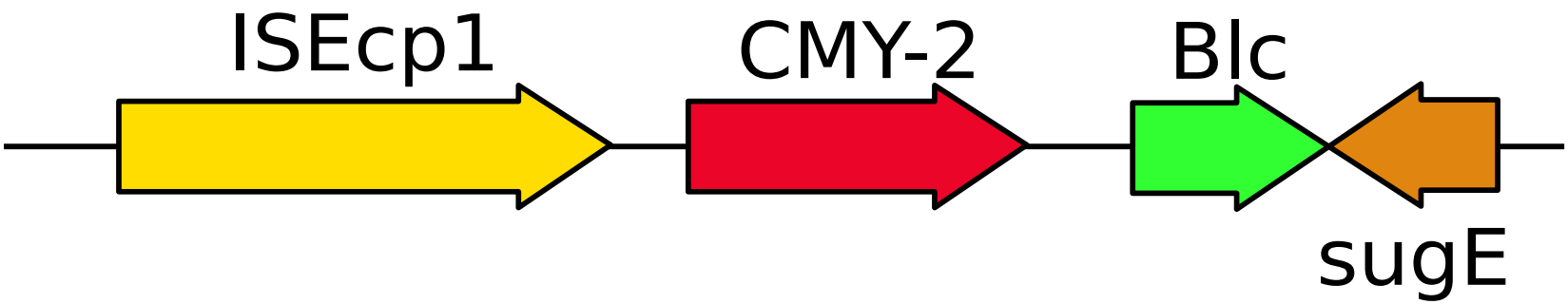
(Maguire et al., 2019)

Generate co-selection hypotheses



(Maguire et al., 2019)

Generate co-selection hypotheses



(Maguire et al., 2019)

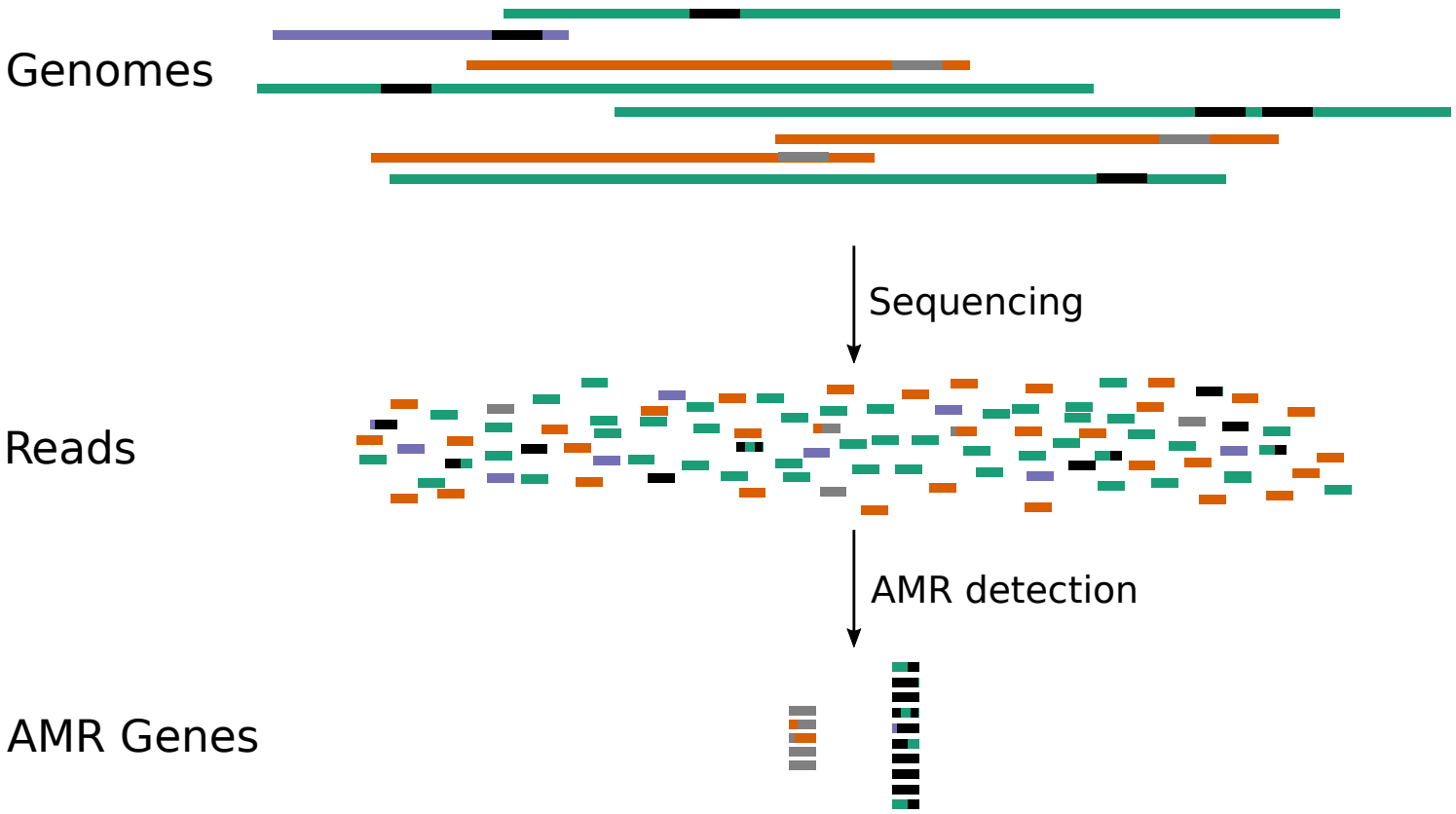
Downsides of genomics

We need genomes to identify previously unknown factors, but:

- Culturing is expensive, time-consuming, and difficult
- Single cell methods are noisy and analytically complex
- Only profile 'one' genome per sample

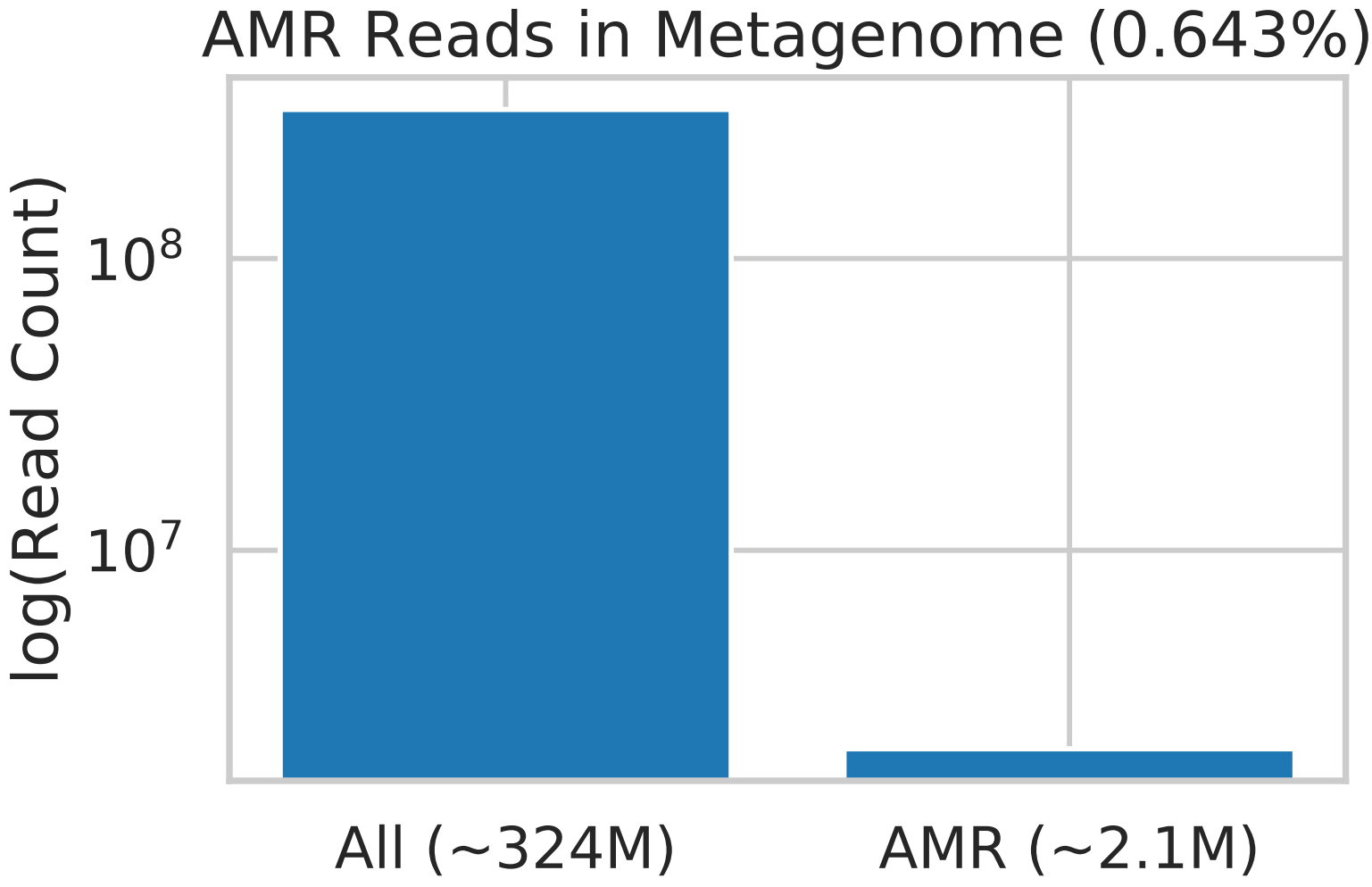
Metagenomics

Read-based AMR Metagenomics



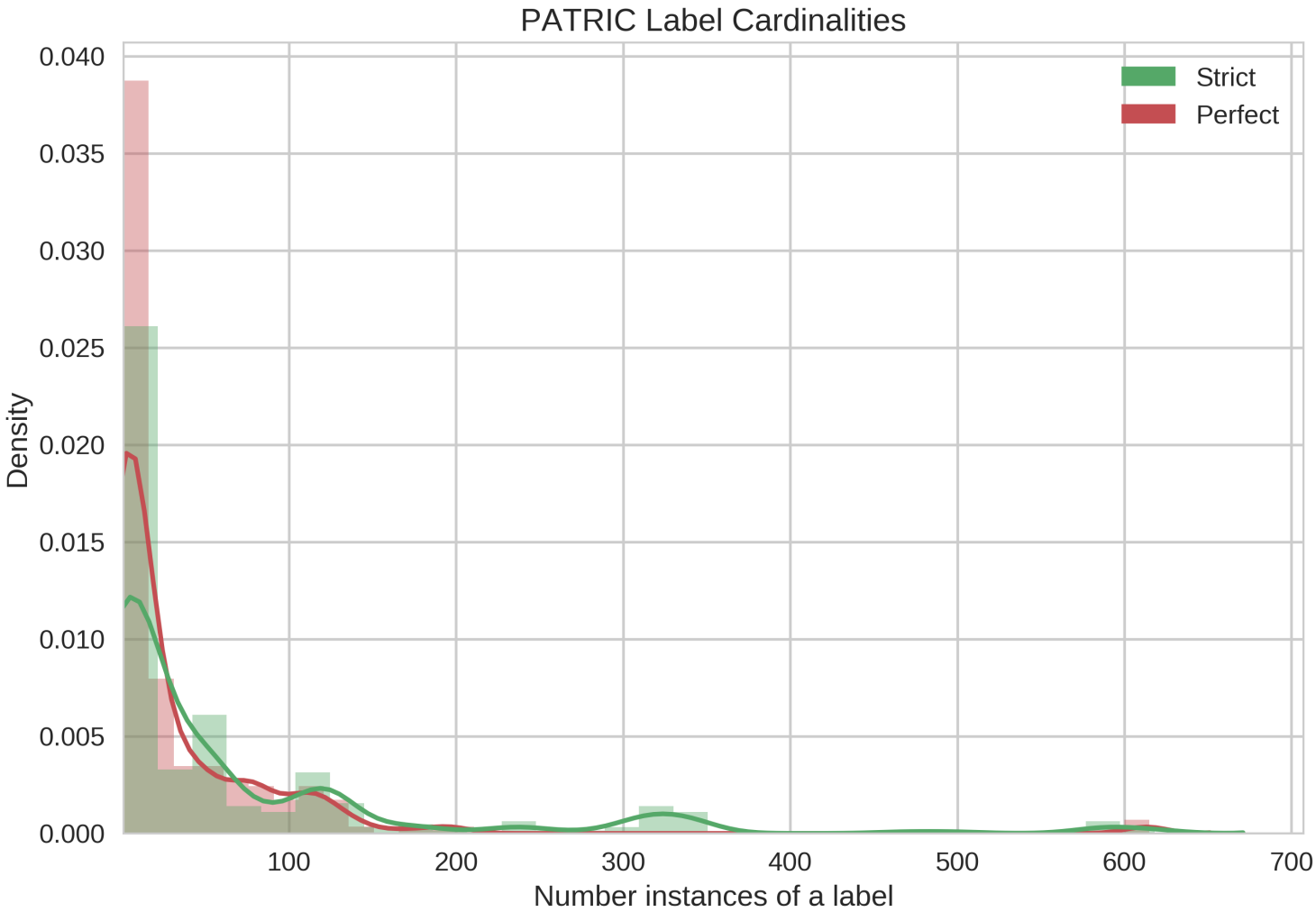
Difficulties of metagenomics

AMR genes are rare genomically



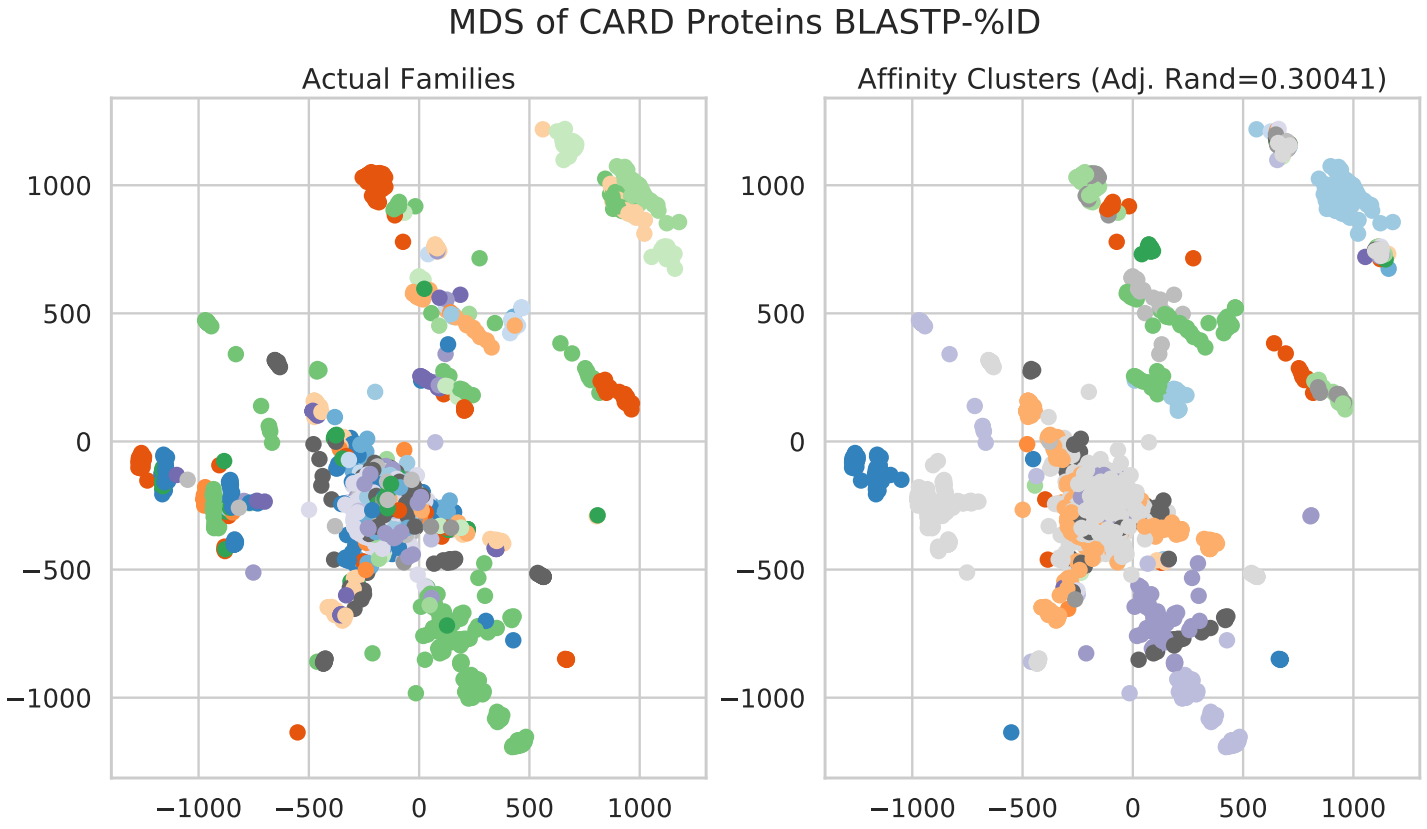
2184 CARD-prevalence genomes at 1-10X abundance

AMR genes have wildly different abundances



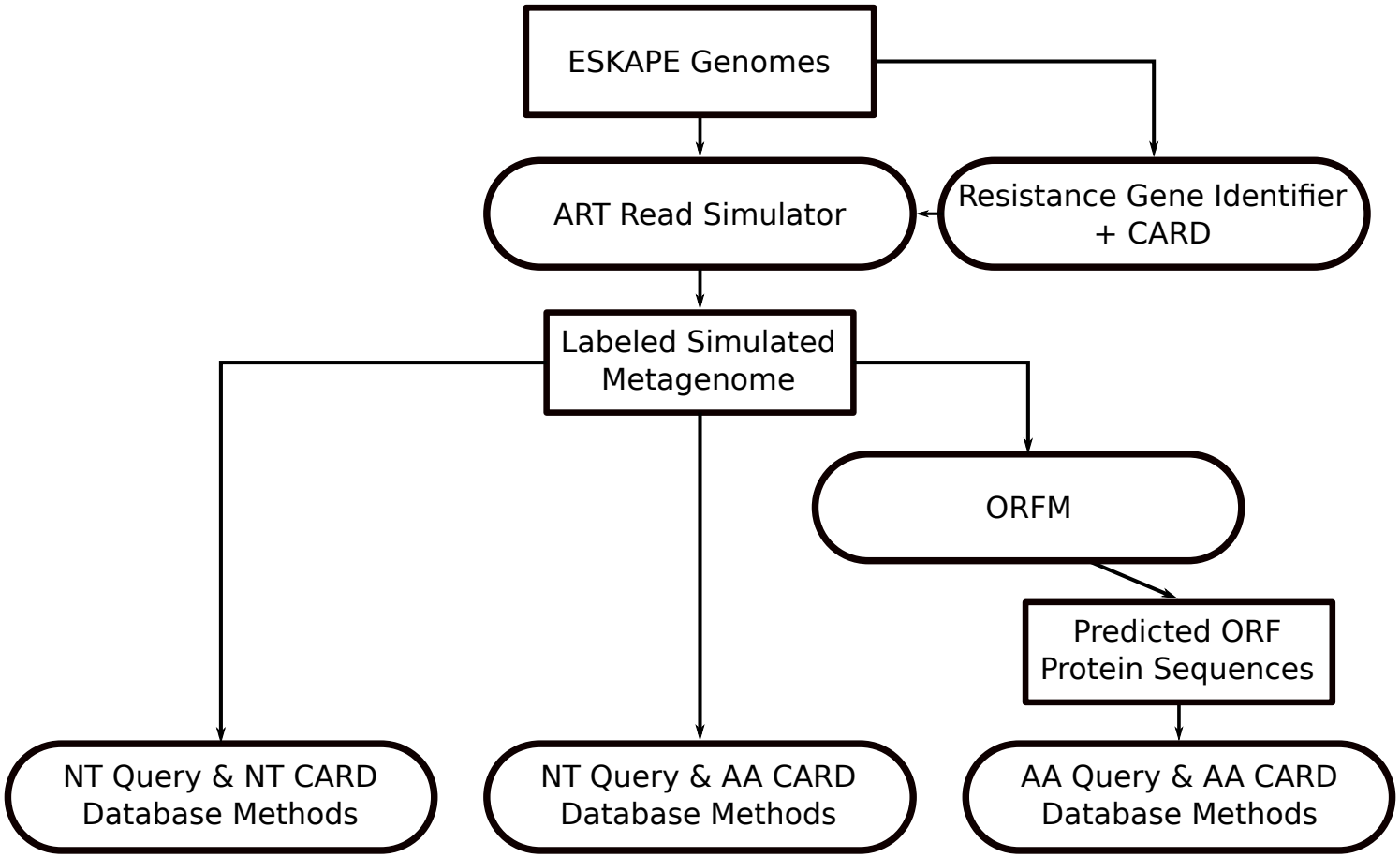
1236 AMR PATRIC genomes

AMR sequence space overlaps

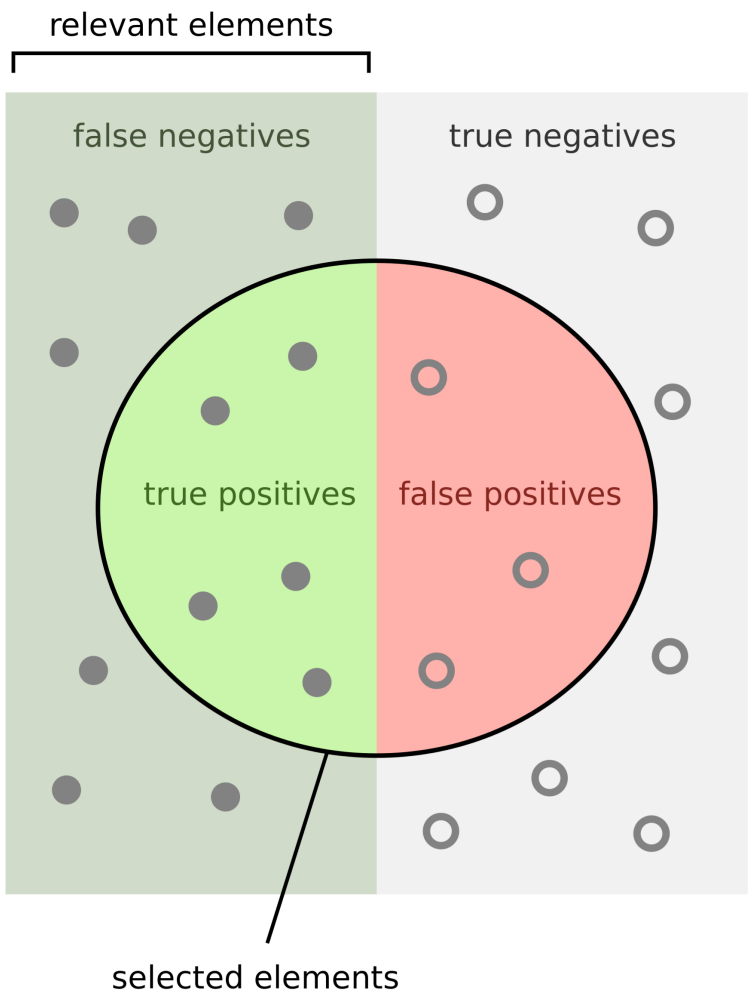


Choosing an analysis approach

Simulate data and compare tools



Terminology refresher



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

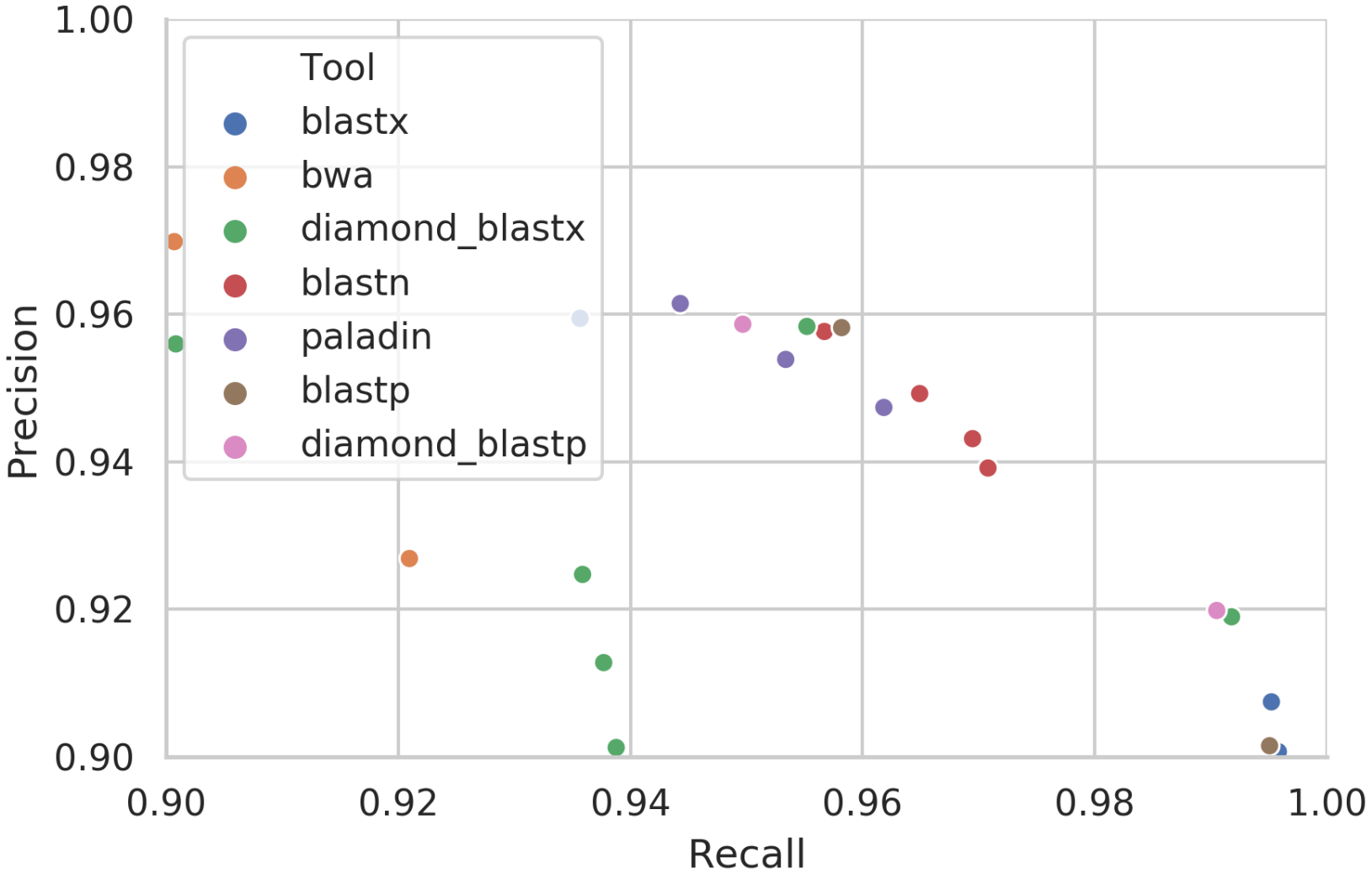
How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

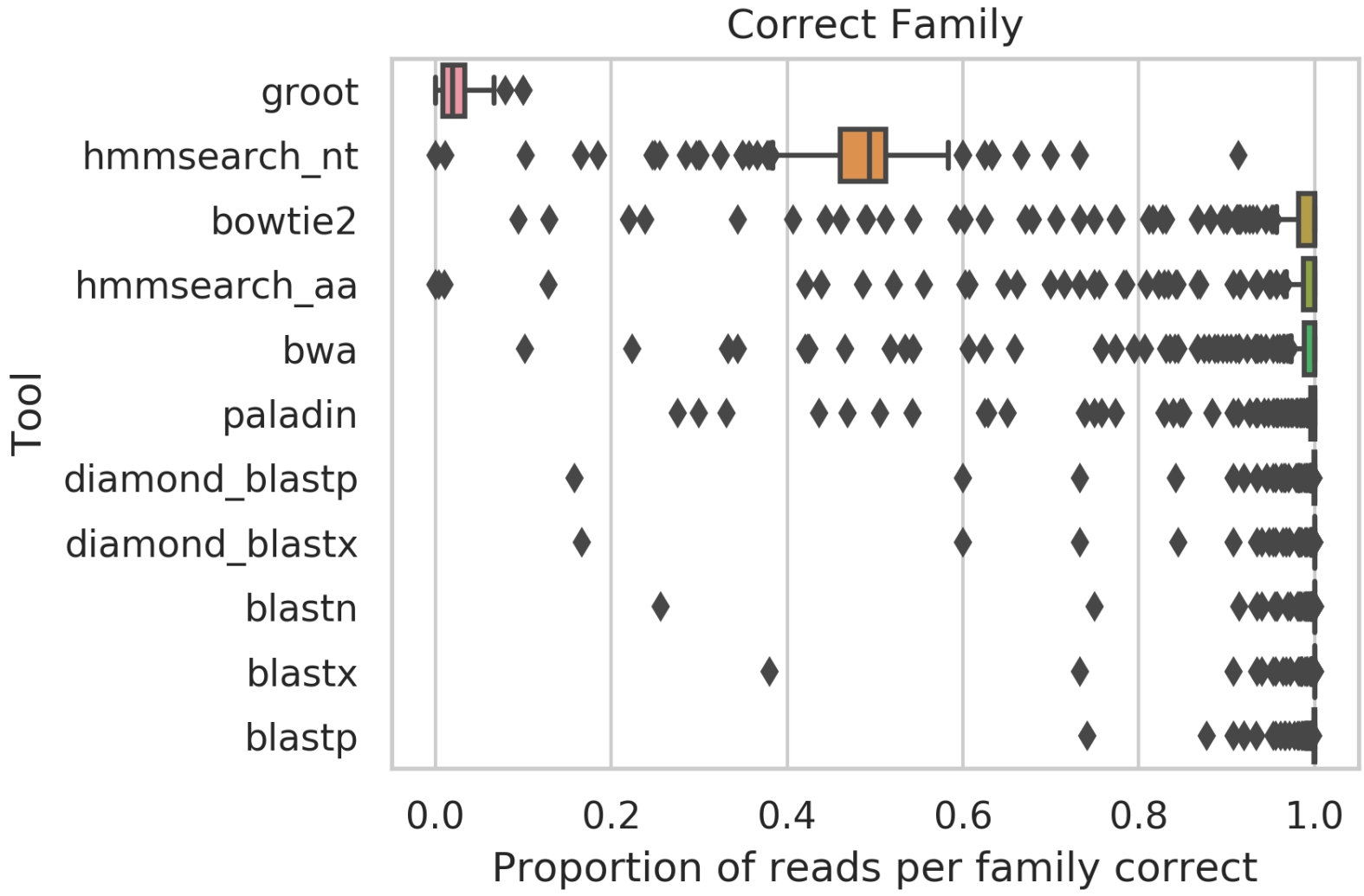
bit.ly/2pZzxJU

How well do different methods do?

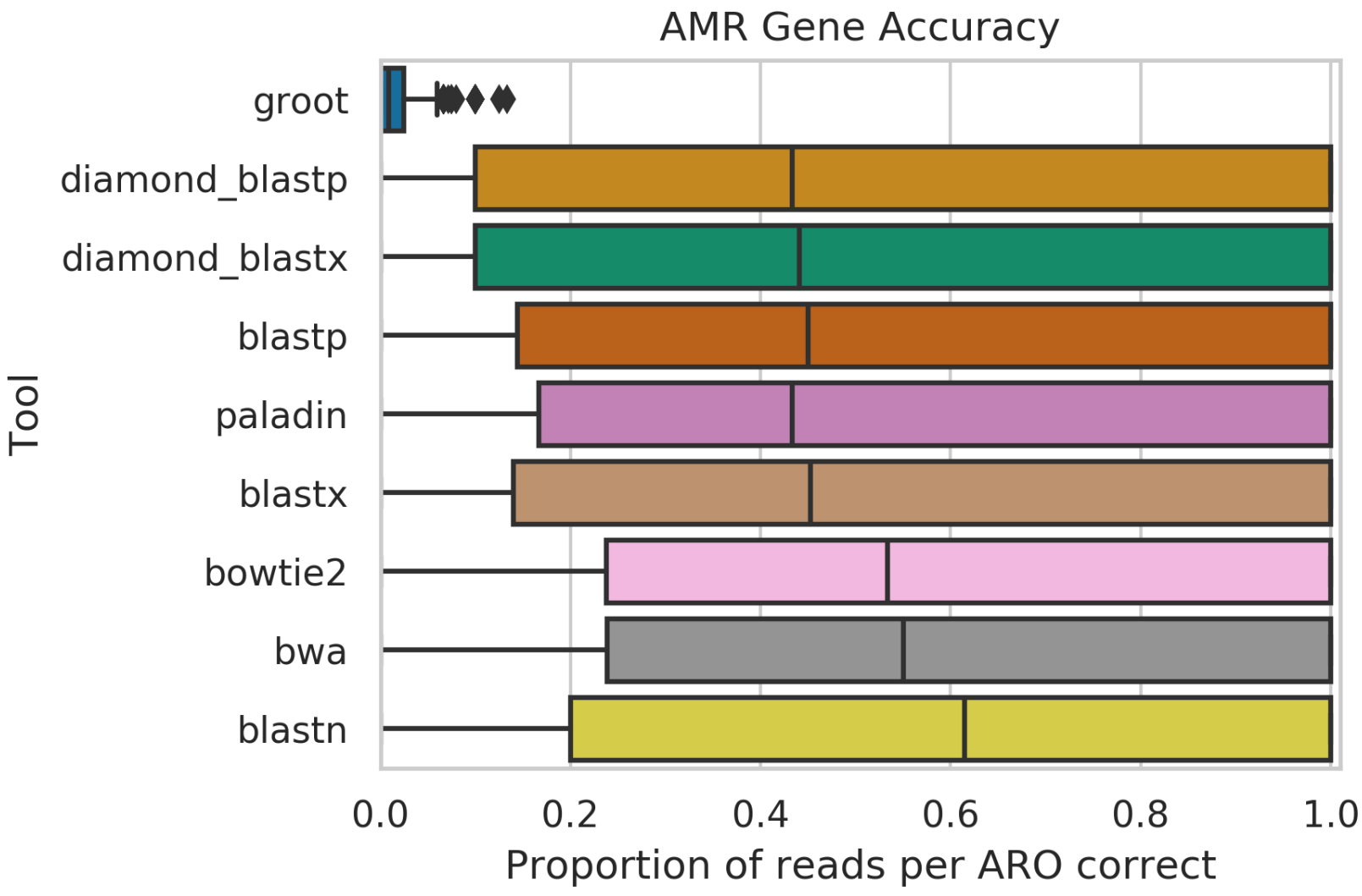
We can find reads from AMR genes



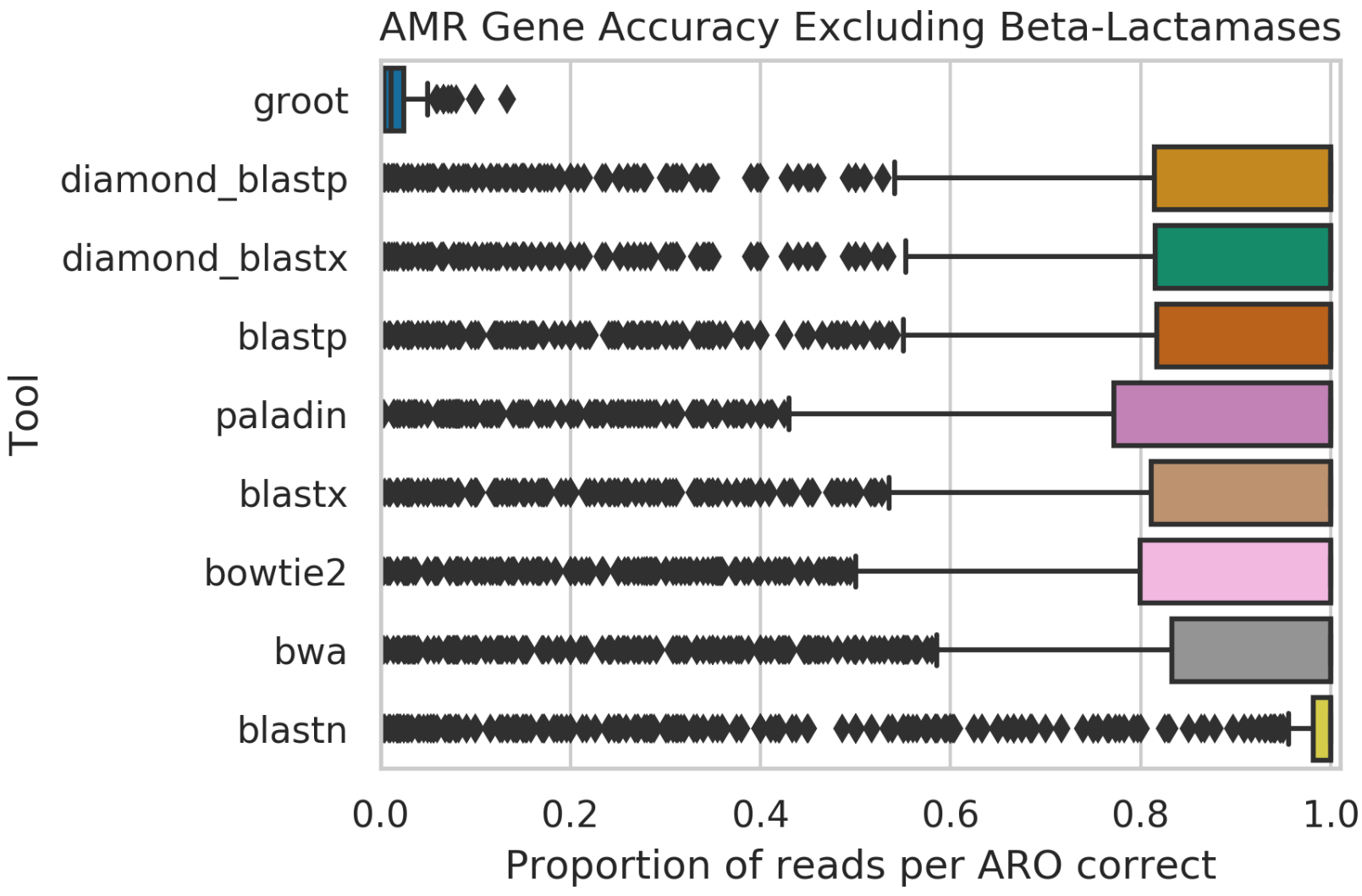
We can mostly identify which family



We cannot identify which specific gene



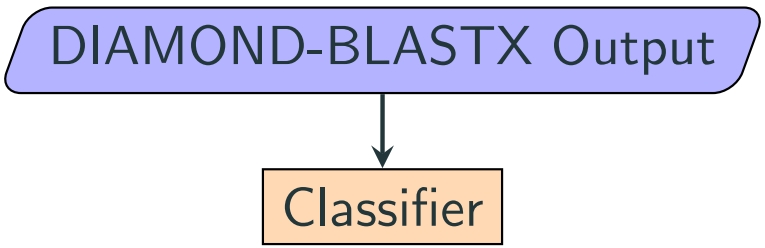
Highly similar families to blame



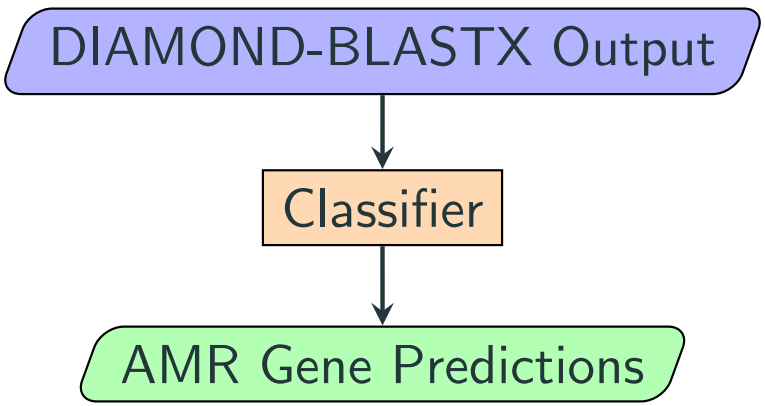
Is there any way to improve this?

Statistical/Machine-Learning Correction

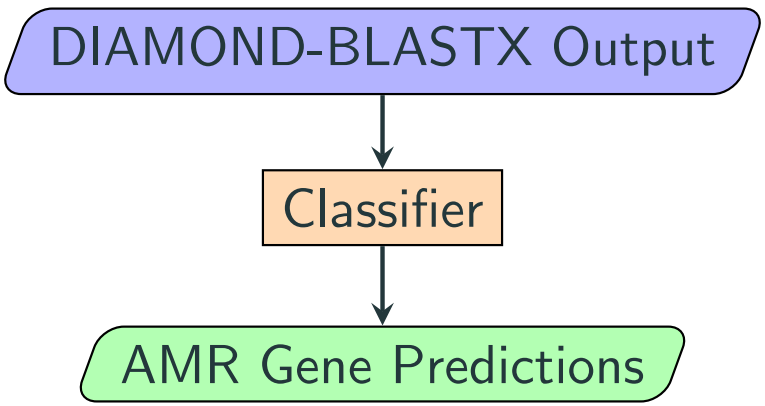
Statistical/Machine-Learning Correction



Statistical/Machine-Learning Correction

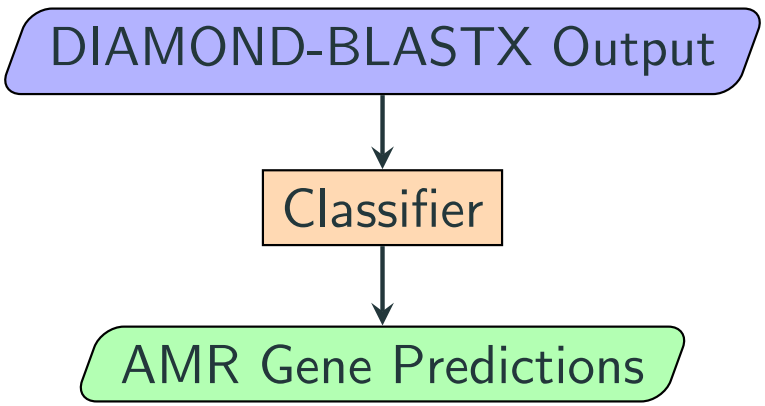


Statistical/Machine-Learning Correction



Average Precision: 0.63

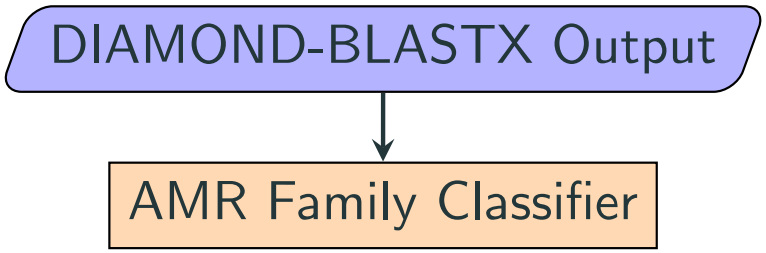
Statistical/Machine-Learning Correction



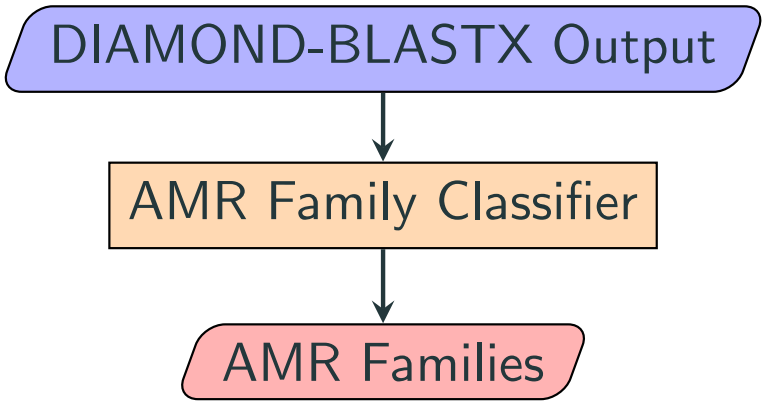
Average Precision: 0.63 %

Revised classifier structure: exploiting the ARO

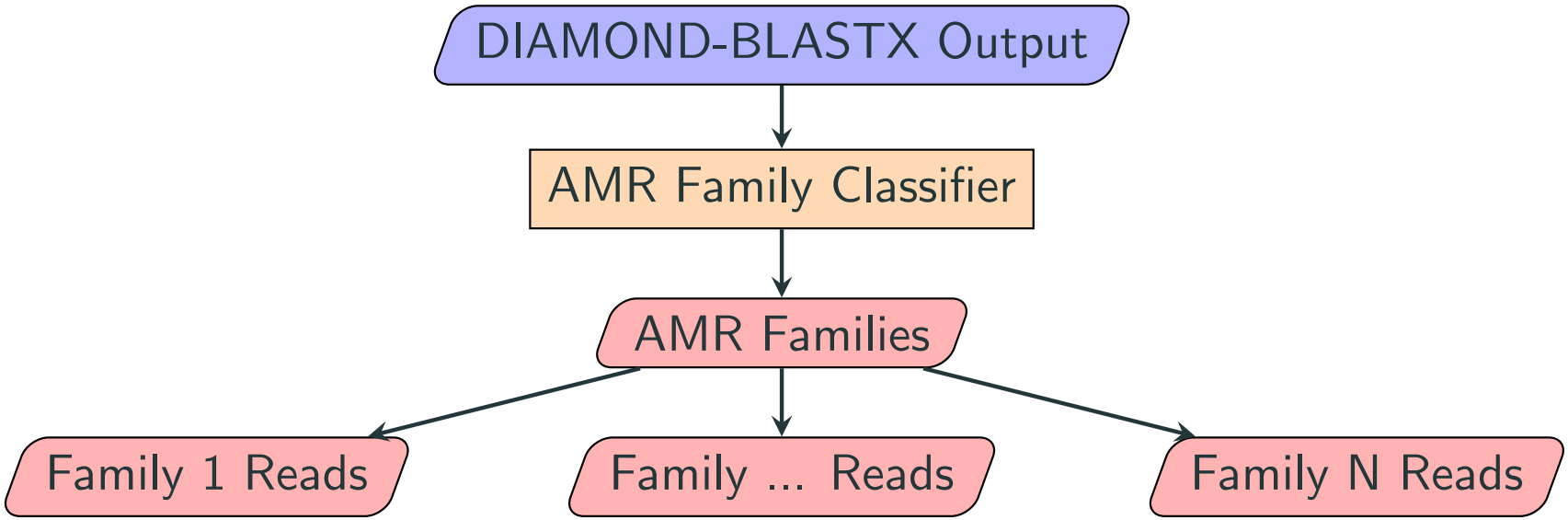
Revised classifier structure: exploiting the ARO



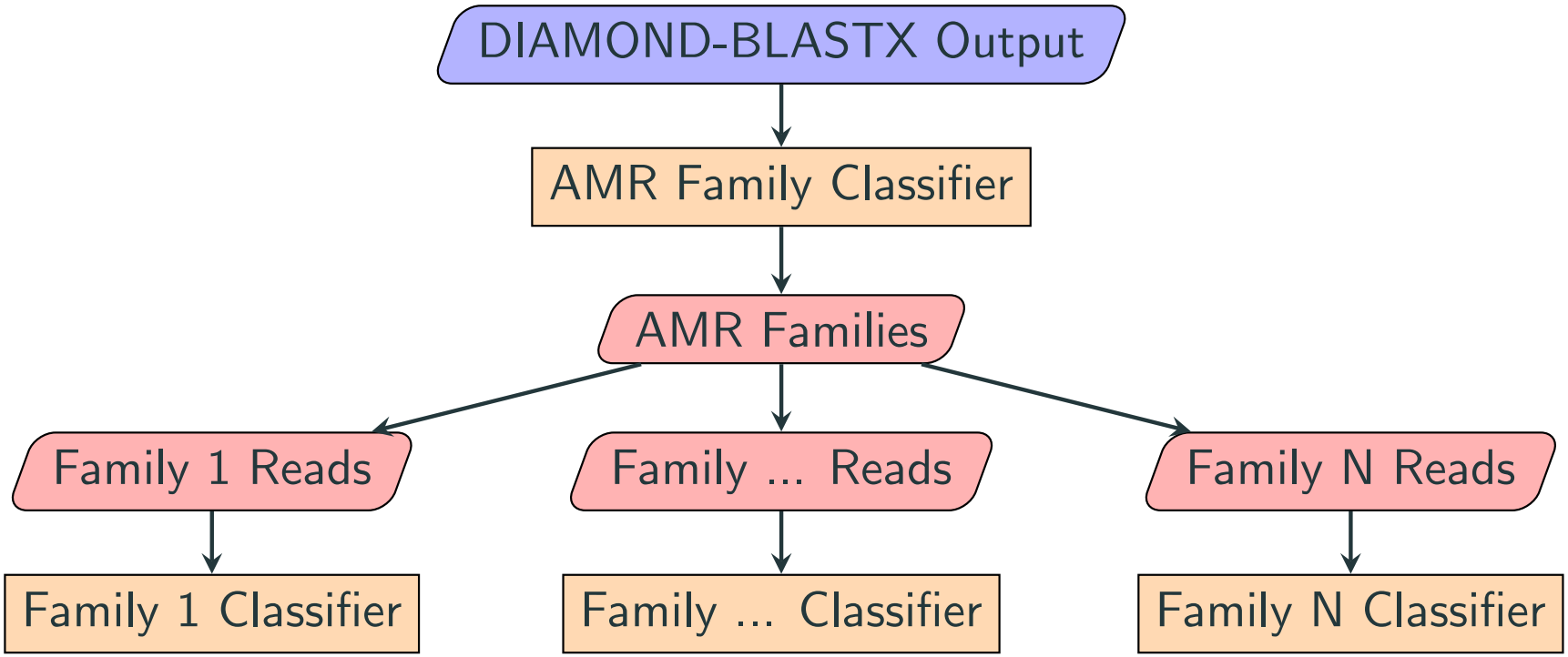
Revised classifier structure: exploiting the ARO



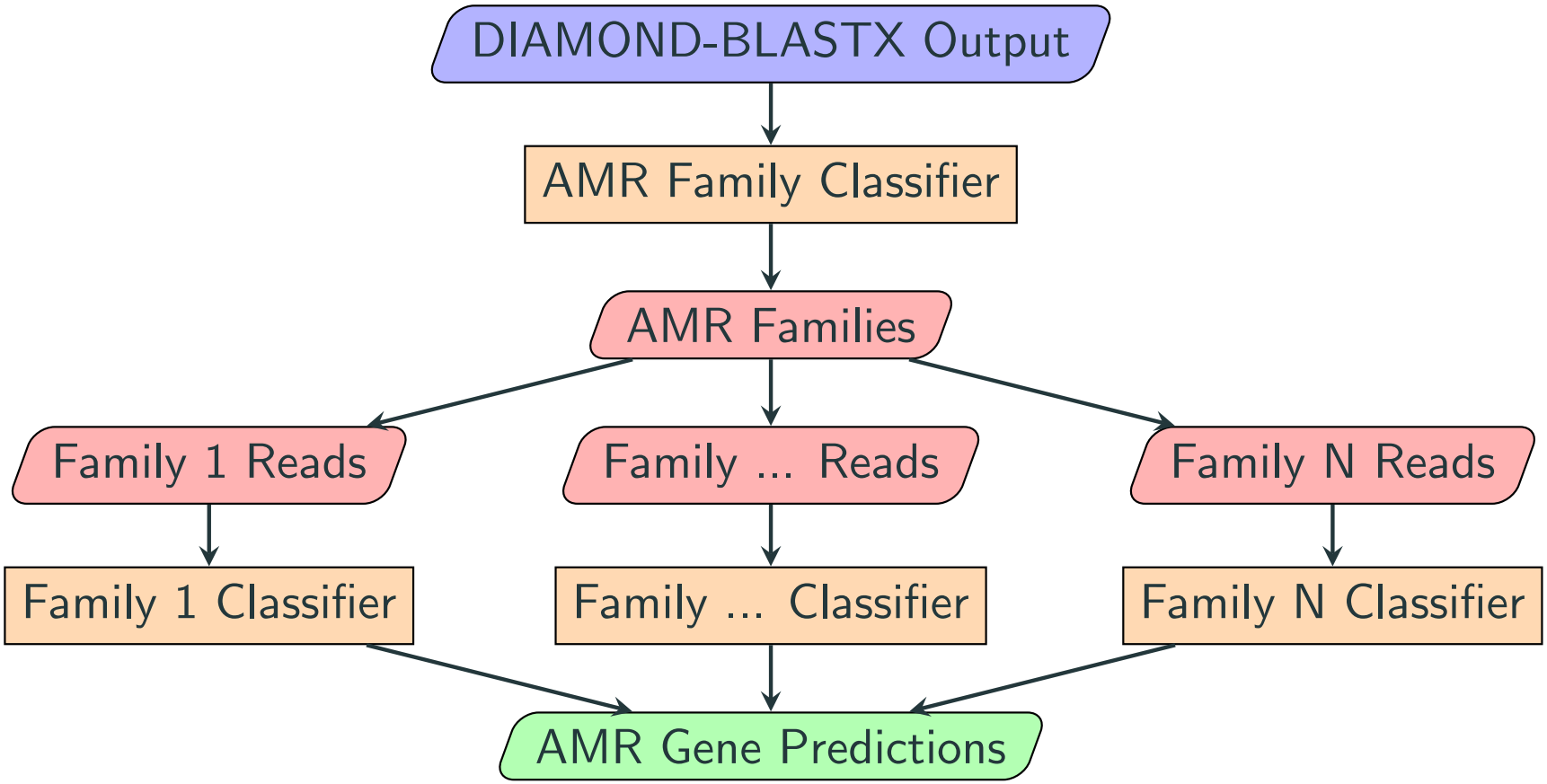
Revised classifier structure: exploiting the ARO



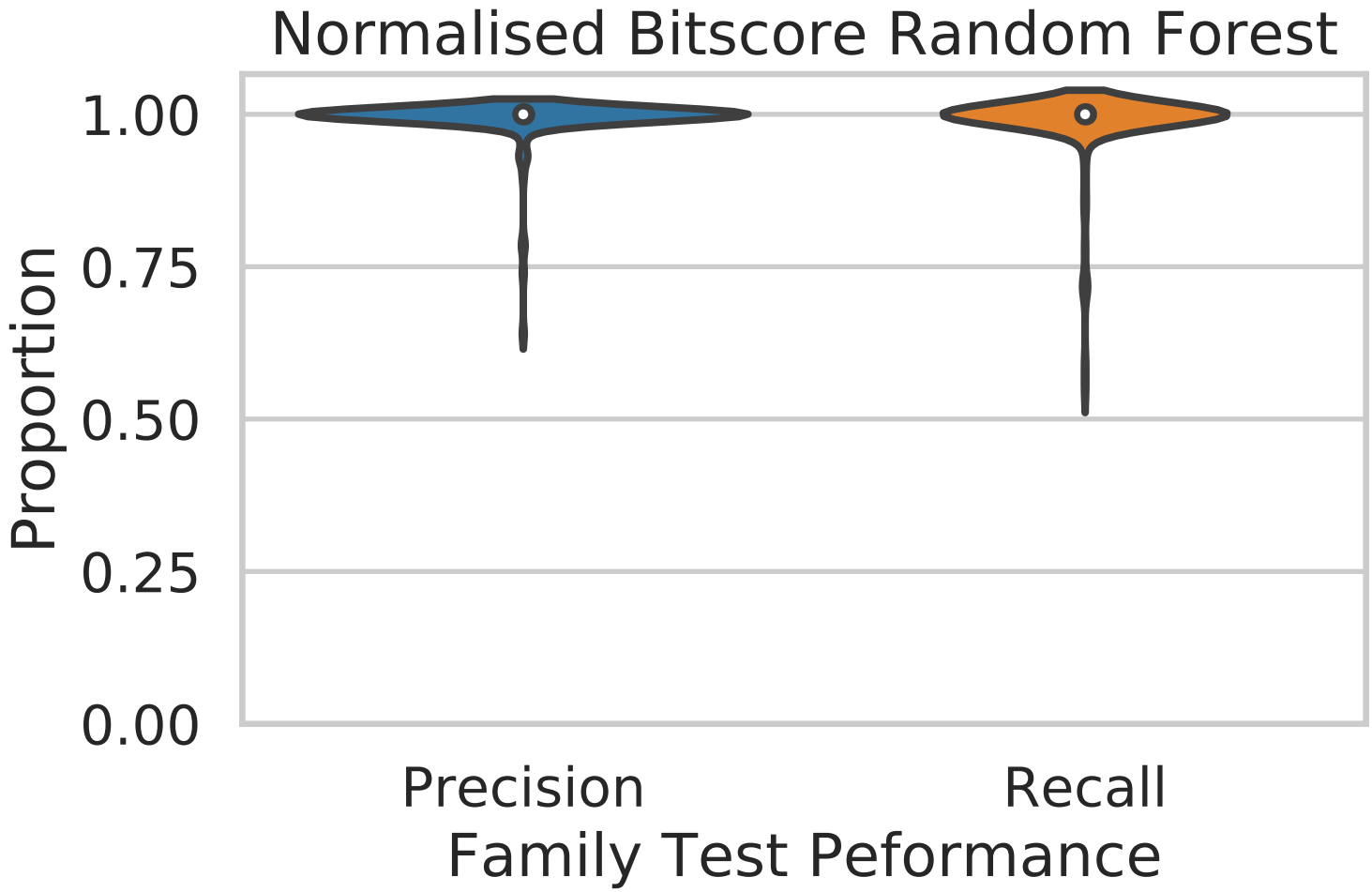
Revised classifier structure: exploiting the ARO



Revised classifier structure: exploiting the ARO

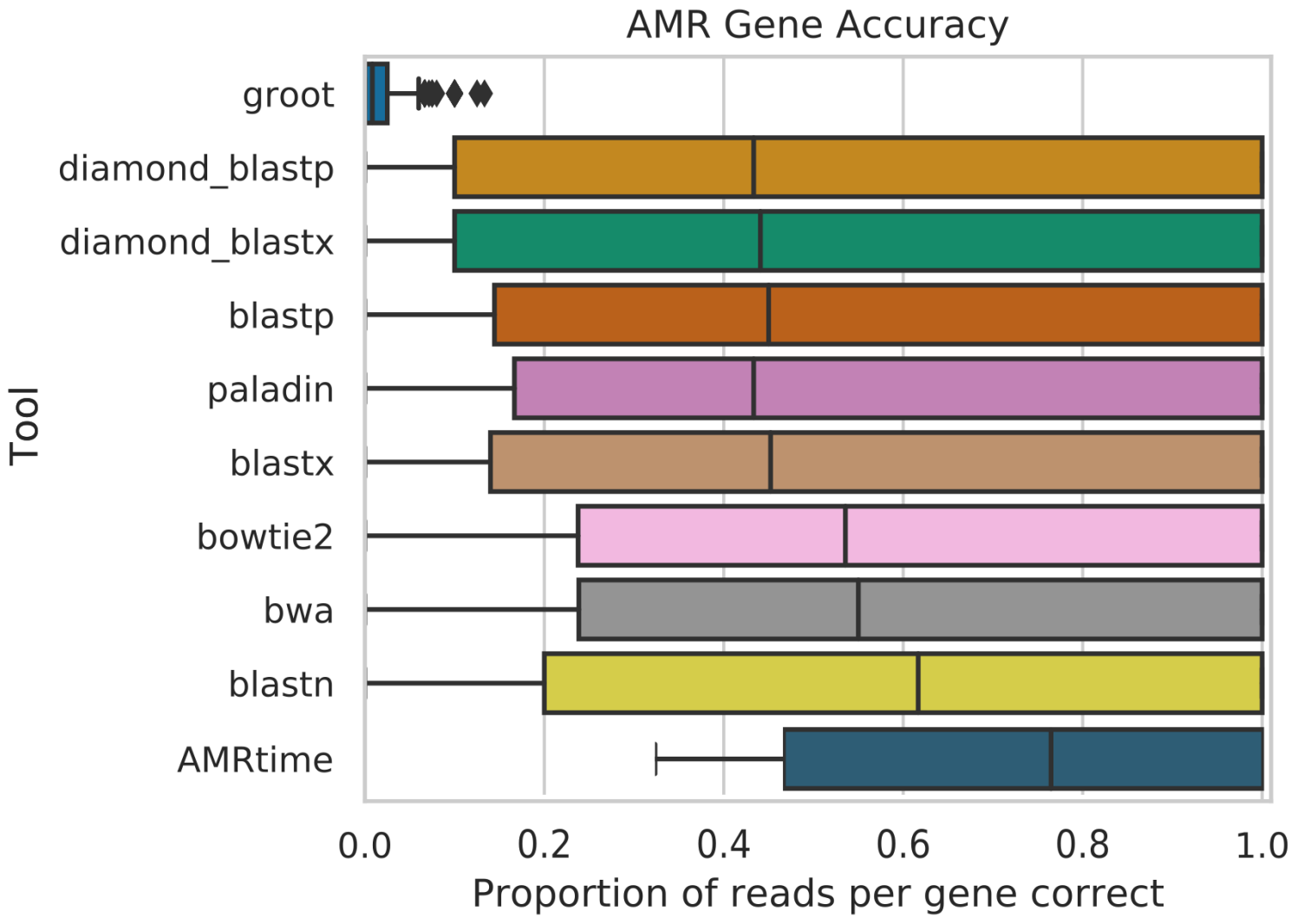


Slightly improved family performance

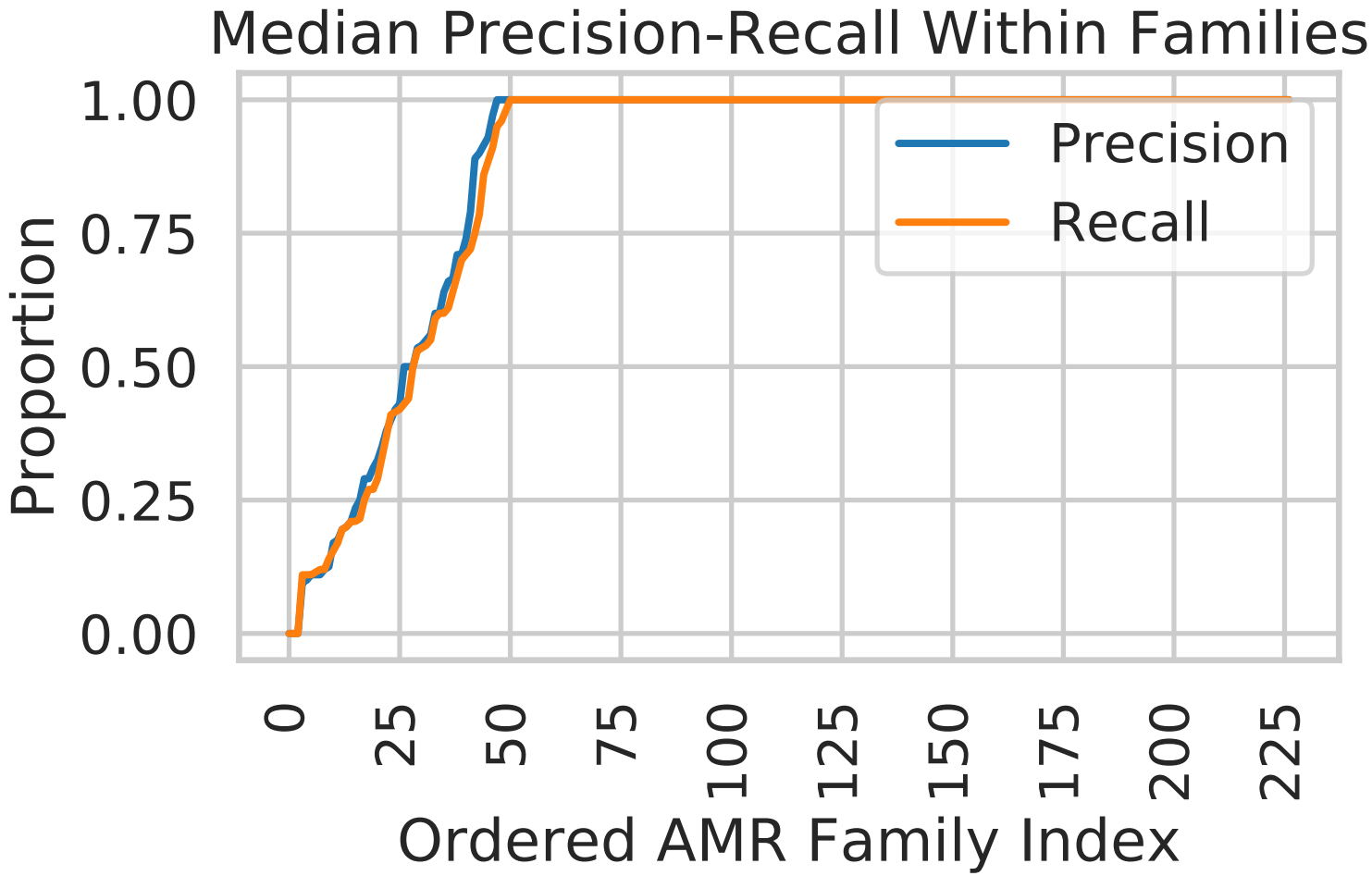


Mean Precision: 0.995, Mean Recall: 0.985

Greatly improved gene performance

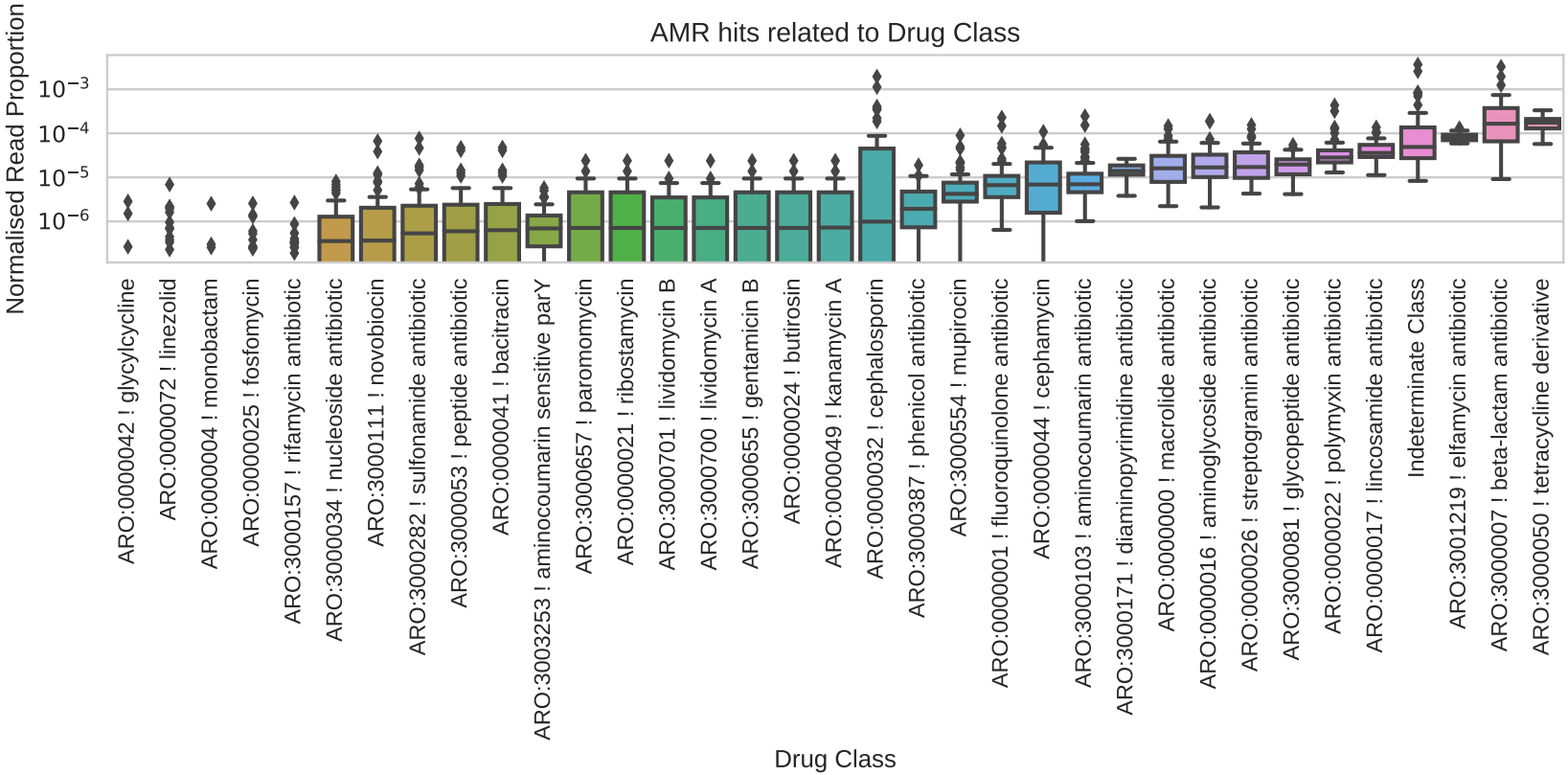


Gains not evenly distributed



- Not enough signal in read so output compatible set
- Some fixed bugs

Metagenomic resistome profile



47 human gut metagenome profiles

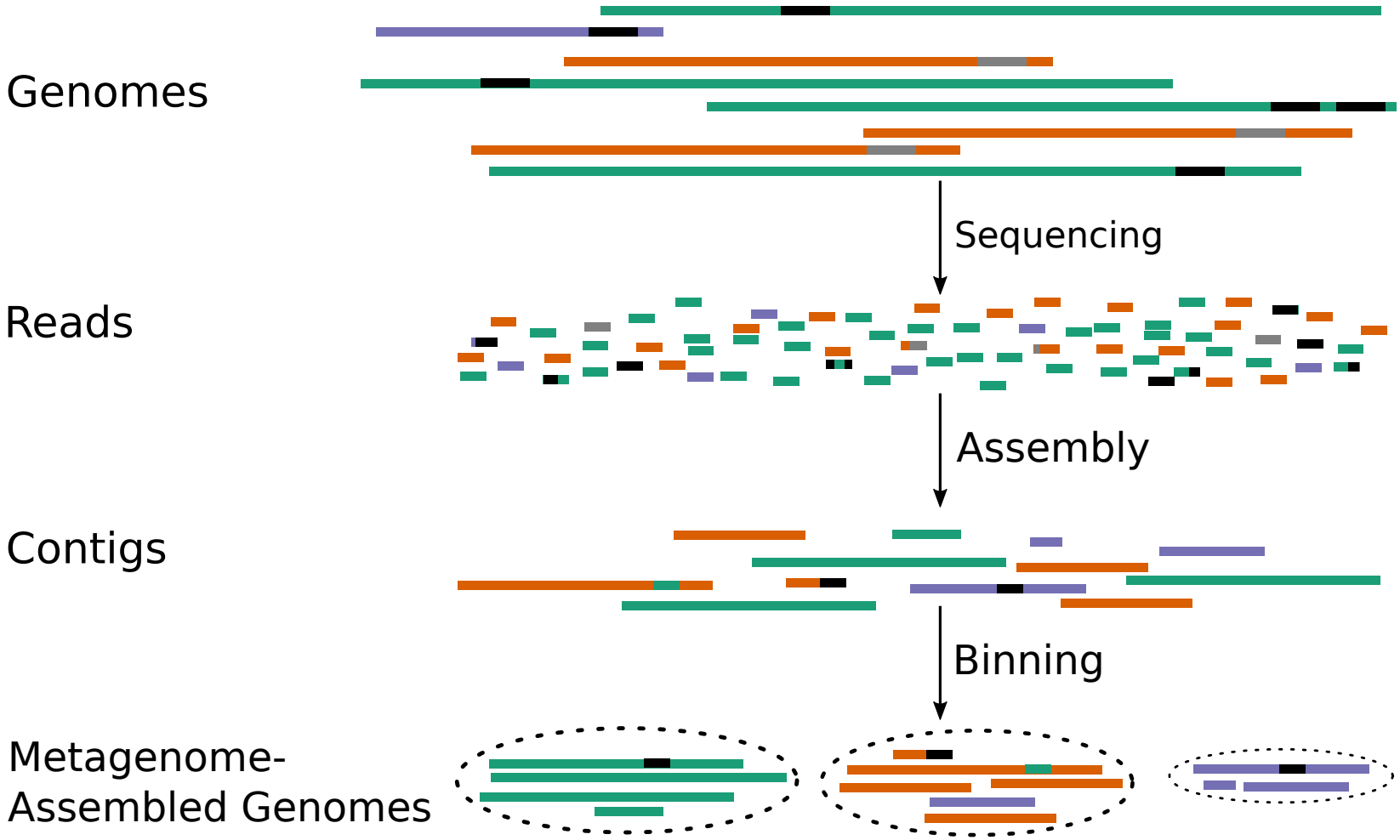
Great, but...

- Known AMR genes
- Is one organism resistant to everything?
- Are many organisms each resistant to one thing?
- Have AMR genes been laterally transferred?

Can we get the best of metagenomics and genomics?

Metagenomic-Assembled Genomes

MAG binning


Genomes
Sequencing
Reads
Assembly
Contigs
Binning
Metagenome-
Assembled Genomes

MAGs are popular

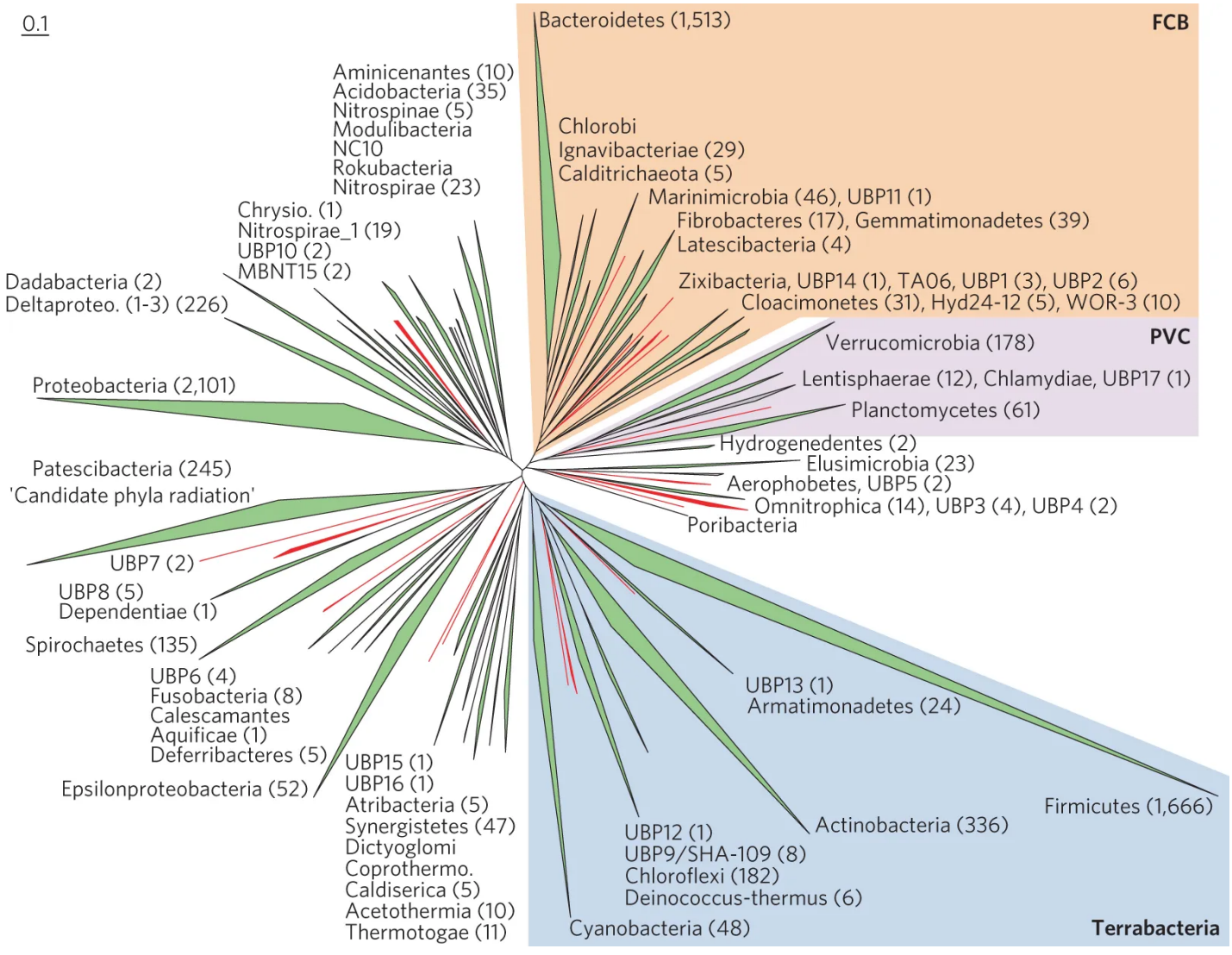


Figure from (Parks et al., 2017)

What about plasmids?

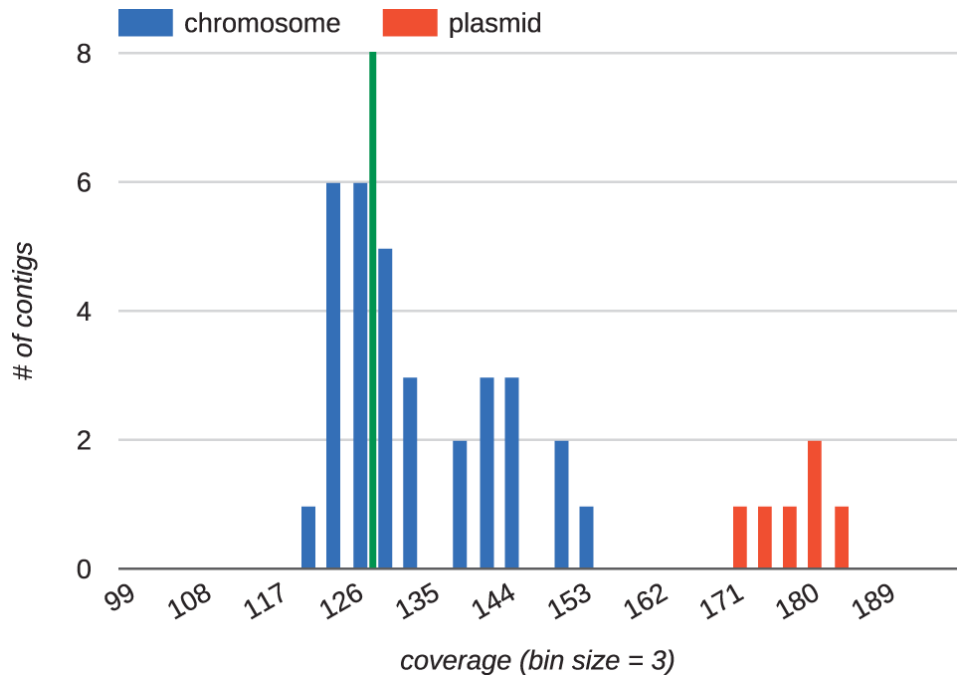
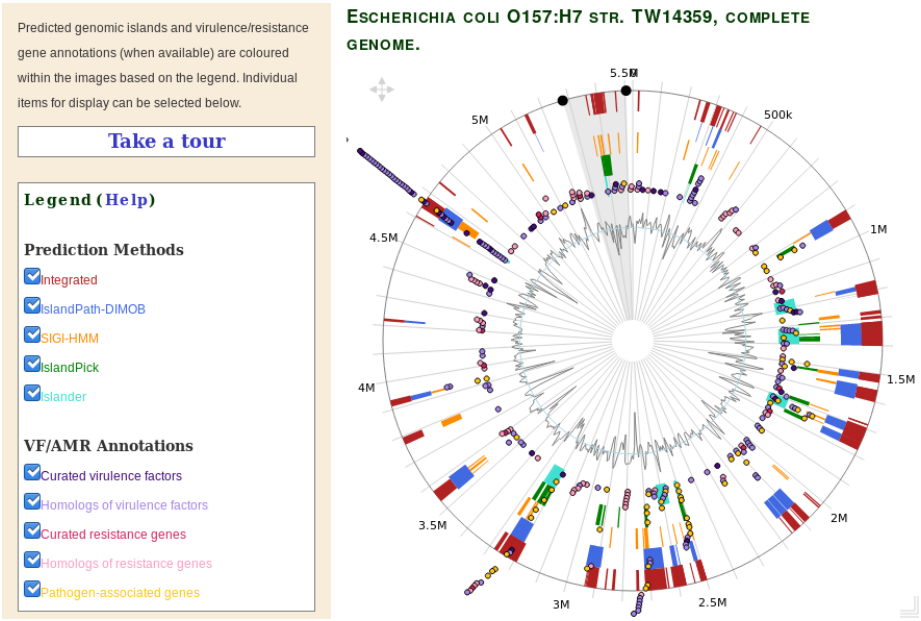


Figure from (Antipov et al., 2016)

- Circular or linear extrachromosomal self-replicating DNA.
- Dissemination of AMR genes.
- Repetitive, variable copy number, different sequence composition.

Or genomic islands



`www.pathogenomics.sfu.ca/islandviewer`

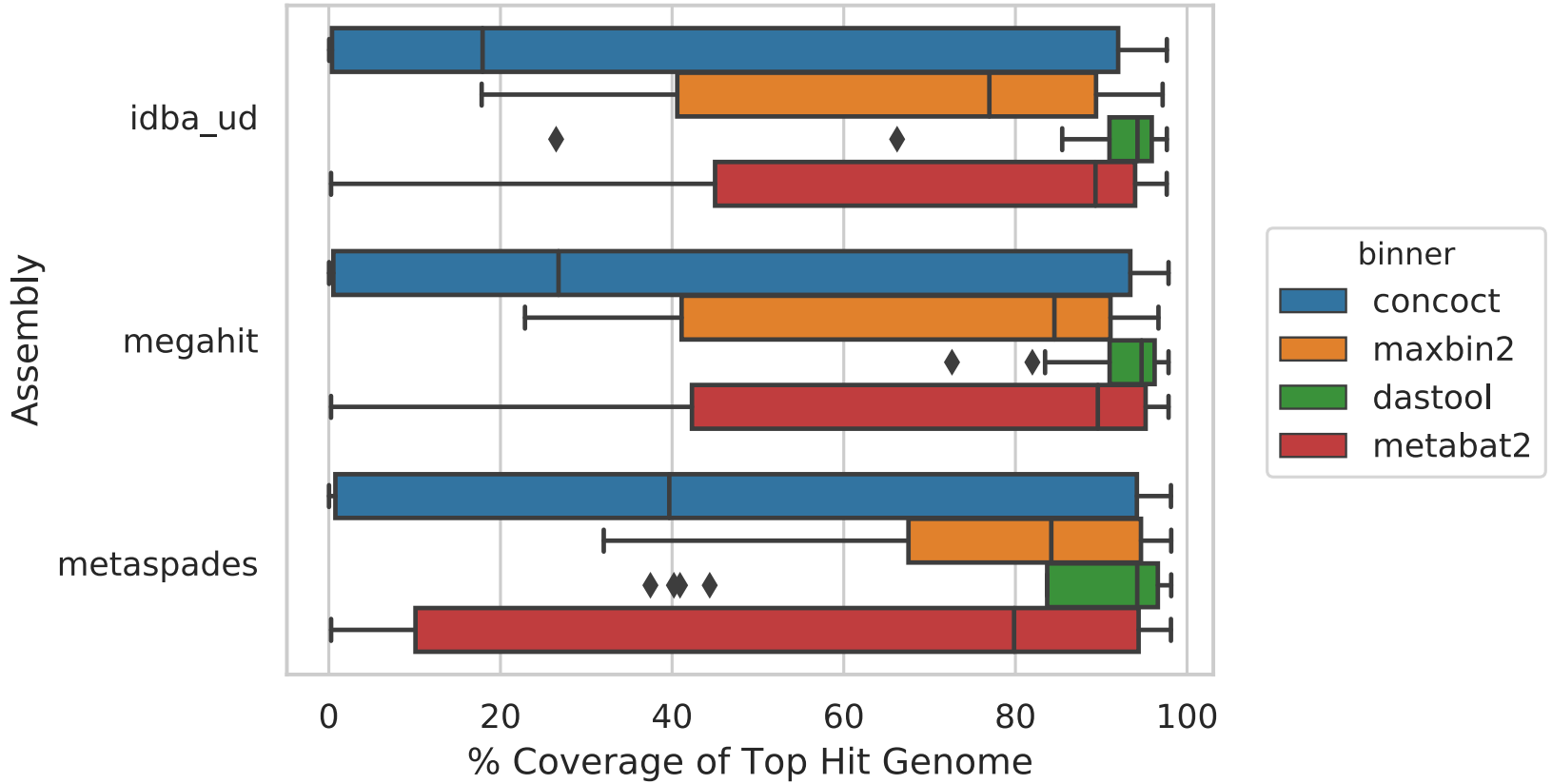
- Clusters of genes acquired through LGT
- Integrons, transposons, integrative and conjugative elements (ICEs) and prophages
- Variable copy number and composition (used by SIGI-HMM, IslandPath-DIMOB)

How well do MAGs recover these sequences?

Time to start simulating again

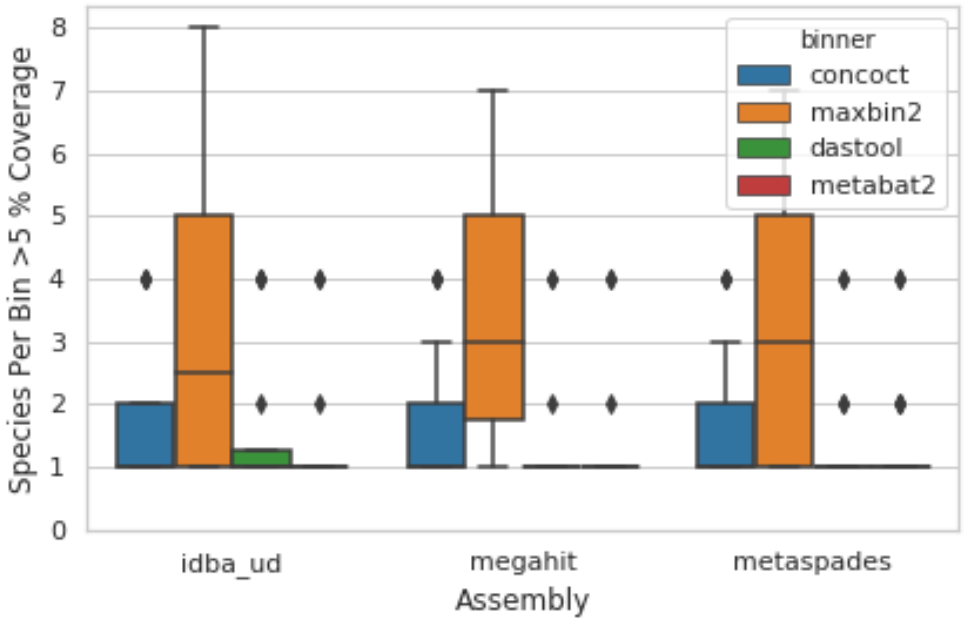
- Simulate some metagenomes (lognormal abundance distribution) from difficult genomes
  - 10 genomes: lots of plasmids
  - 10 genomes: high % of genomic islands (compositional)
  - 10 genomes: low % of genomic islands
- Assembly using 3 alternative methods: IDBA_UD, MetaSPAdes, Megahit
- Bin contigs using 4 different tools: metabat2, maxbin2, concoct, dastool

Chromosomes fairly well binned



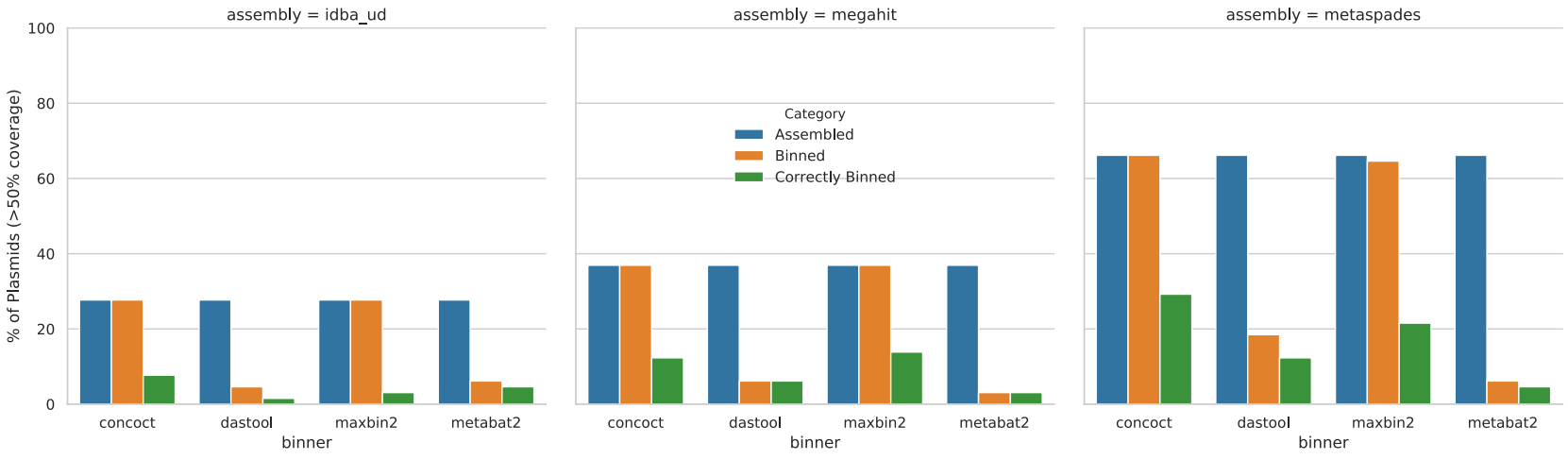
26-94.3% median chromosomal coverage (Pre-print draft `github.com/fmaguire/mag_sim_paper`)

Chromosomes fairly well binned



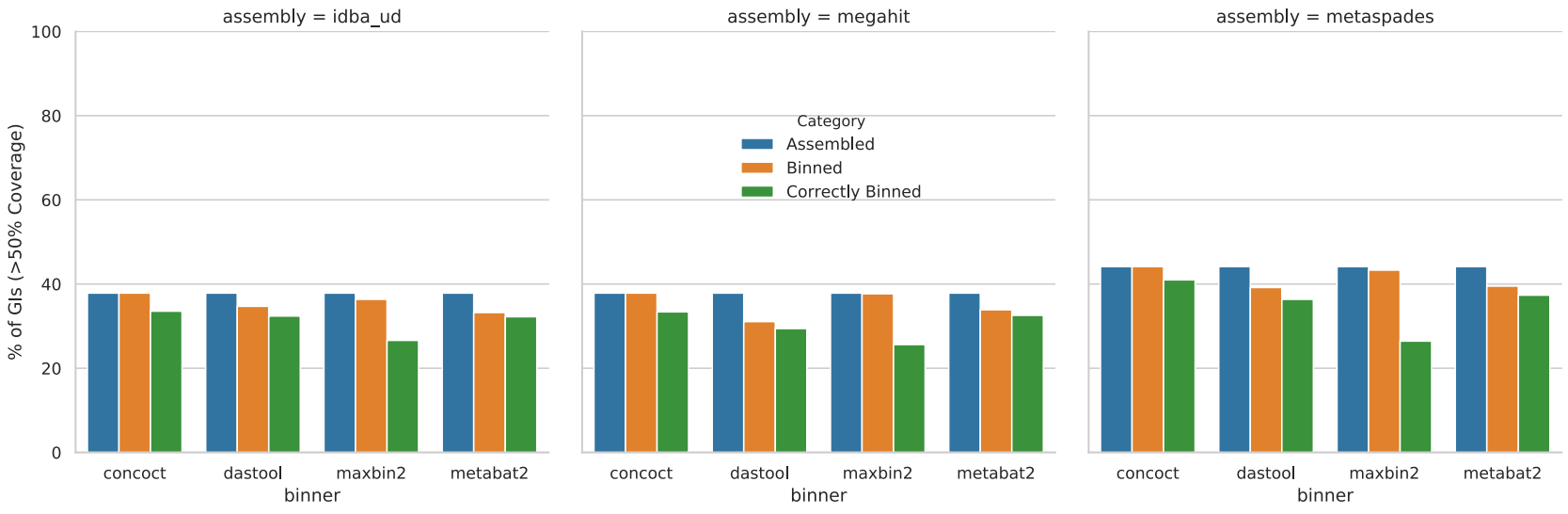
26-94.3% median chromosomal coverage (Pre-print draft `github.com/fmaguire/mag_sim_paper`)

Plasmids are not



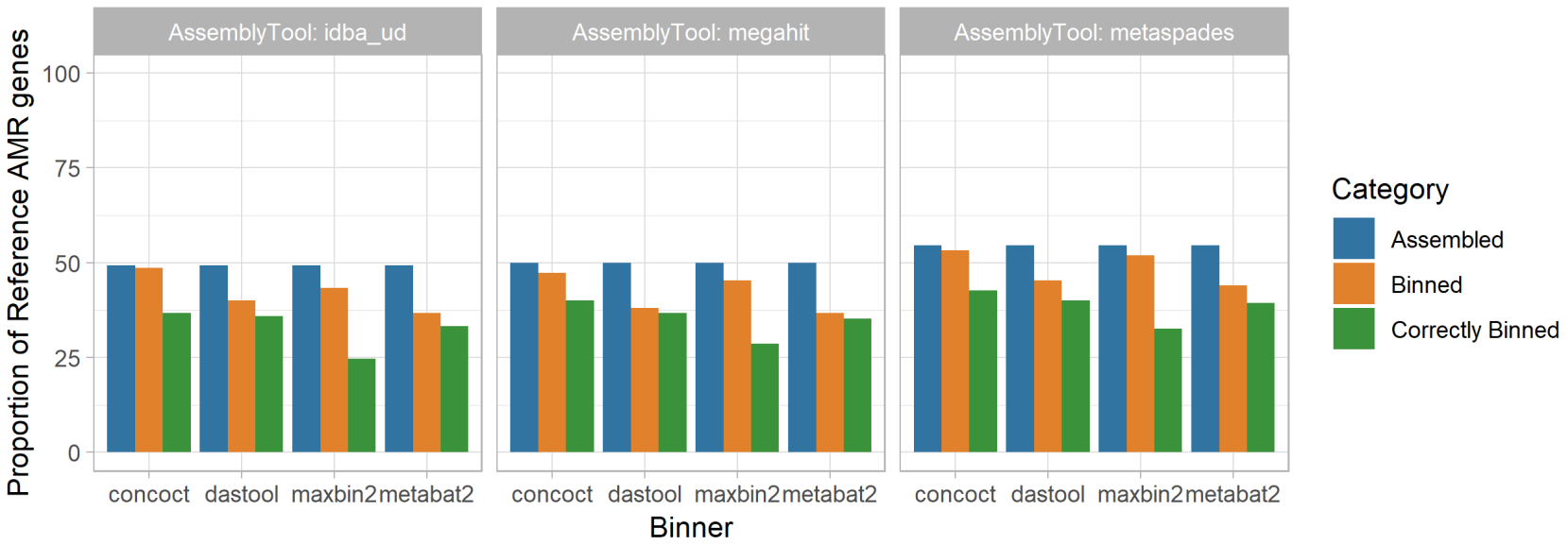
1.5-29.2% plasmids binned

Genomic islands are better but bad



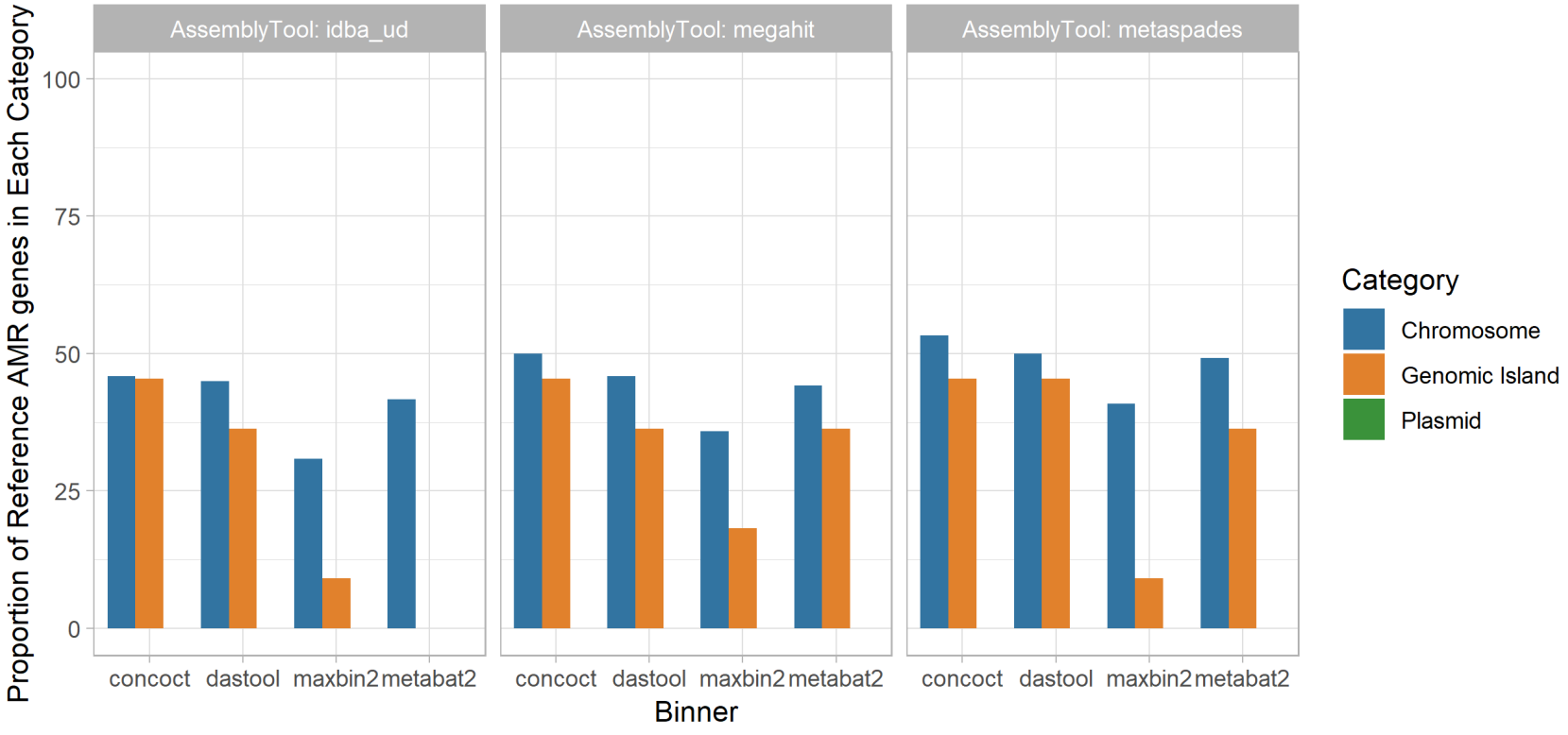
28-42% GIs binned

What about AMR genes?



24-43% AMR genes binned

Which AMR genes are lost?



- 30-53% chromosomal AMR genes (n=120)
- 0-45% genomic island AMR genes (n=11)
- 0% of plasmid AMR genes (n=20)

Be cautious with MAGs

- Regain some context but with biased data loss
- Disproportionate loss of AMR genes
- Mobile Genetic Elements poorly recovered
- Cautionary tale: more processing = more data loss

Conclusions

Conclusions

Method	Strengths	Weaknesses

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation
Genomics	Context, moderate analysis	Isolation, throughput

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation
Genomics	Context, moderate analysis	Isolation, throughput
Metagenomics	Many genomes at once	Fragmented, no context, difficult analysis

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation
Genomics	Context, moderate analysis	Isolation, throughput
Metagenomics	Many genomes at once	Fragmented, no context, difficult analysis
Metagenomic-Assembed Genomes	Context for many genomes	Lose key data, complex analysis

- Simulation fundamental to evaluating approaches

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation
Genomics	Context, moderate analysis	Isolation, throughput
Metagenomics	Many genomes at once	Fragmented, no context, difficult analysis
Metagenomic-Assembed Genomes	Context for many genomes	Lose key data, complex analysis

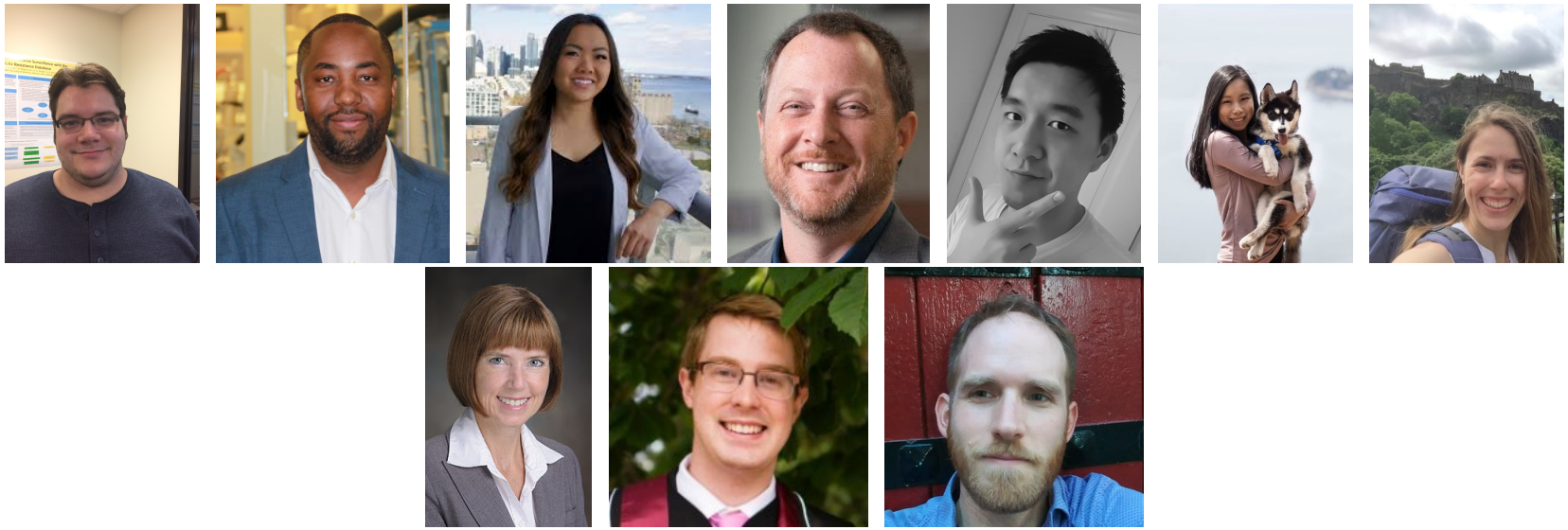
- Simulation fundamental to evaluating approaches
- Characterisation necessary to mitigate weaknesses and promote strengths

Conclusions

Method	Strengths	Weaknesses
Targeted	Cheap, easy analysis	*a priori*, stagnation
Genomics	Context, moderate analysis	Isolation, throughput
Metagenomics	Many genomes at once	Fragmented, no context, difficult analysis
Metagenomic-Assembed Genomes	Context for many genomes	Lose key data, complex analysis

- Simulation fundamental to evaluating approaches
- Characterisation necessary to mitigate weaknesses and promote strengths
- Machine-Learning represents useful tools for this (e.g. AMRtime, gene-free AST prediction models)

Acknowledgements



- McMaster University: Brian Alcock, Amos Raphenya, Kara Tsang, Andrew McArthur
- Simon Fraser University: Justin Jia, Kristen Gray, Venus Lau, Fiona Brinkman
- Dalhousie University: Mike Hall, Robert Beiko
- Funding: Donald Hill Family Fellowship; Genome Canada.

Questions?

References

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. (2016). plasmidspades: assembling plasmids from whole genome sequencing data. *bioRxiv*, page 048942.

Bradley, P., Gordon, N. C., Walker, T. M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L. J., Anson, L., De Cesare, M., et al. (2015). Rapid antibiotic-resistance predictions from genome sequence data for staphylococcus aureus and mycobacterium tuberculosis. *Nature communications*, 6:10063.

Cassini, A., Hogberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., Colomb-Cotinat, M., Kretzschmar, M. E., Devleesschauwer, B., Cecchini, M., et al. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the eu and the european economic area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, 19(1):56–66.

de Kraker, M. E., Stewardson, A. J., and Harbarth, S. (2016). Will 10 million people die a year due to antimicrobial resistance by 2050? *PLoS medicine*, 13(11):e1002184.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., Group, E. P., et al. (2005). Outpatient antibiotic use in europe and association with resistance: a cross-national database study. *The Lancet*, 365(9459):579–587.

Guitor, A. K., Raphenya, A. R., Klunk, J., Kuch, M., Alcock, B., Surette, M. G., McArthur, A. G., Poinar, H. N., and Wright, G. D. (2019). Capturing the resistome: A targeted capture method to reveal antibiotic resistance determinants in metagenomes. *Antimicrobial agents and chemotherapy*, pages AAC–01324.

Maguire, F. (2016). A multi-omic analysis of the photosynthetic endosymbioses of paramecium bursaria. *PhD Thesis*.

Maguire, F., Rehman, M. A., Carrillo, C., Diarra, M. S., and Beiko, R. G. (2019). Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal salmonella enterica serovars by using machine learning. *MSystems*, 4(4):e00211–19.

Matthews, T. C., Bristow, F. R., Griffiths, E. J., Petkau, A., Adam, J., Dooley, D., Kruczkiewicz, P., Curatcha, J., Cabral, J., Fornika, D., et al. (2018). The integrated rapid infectious disease analysis (irida) platform. *bioRxiv*, page 381830.

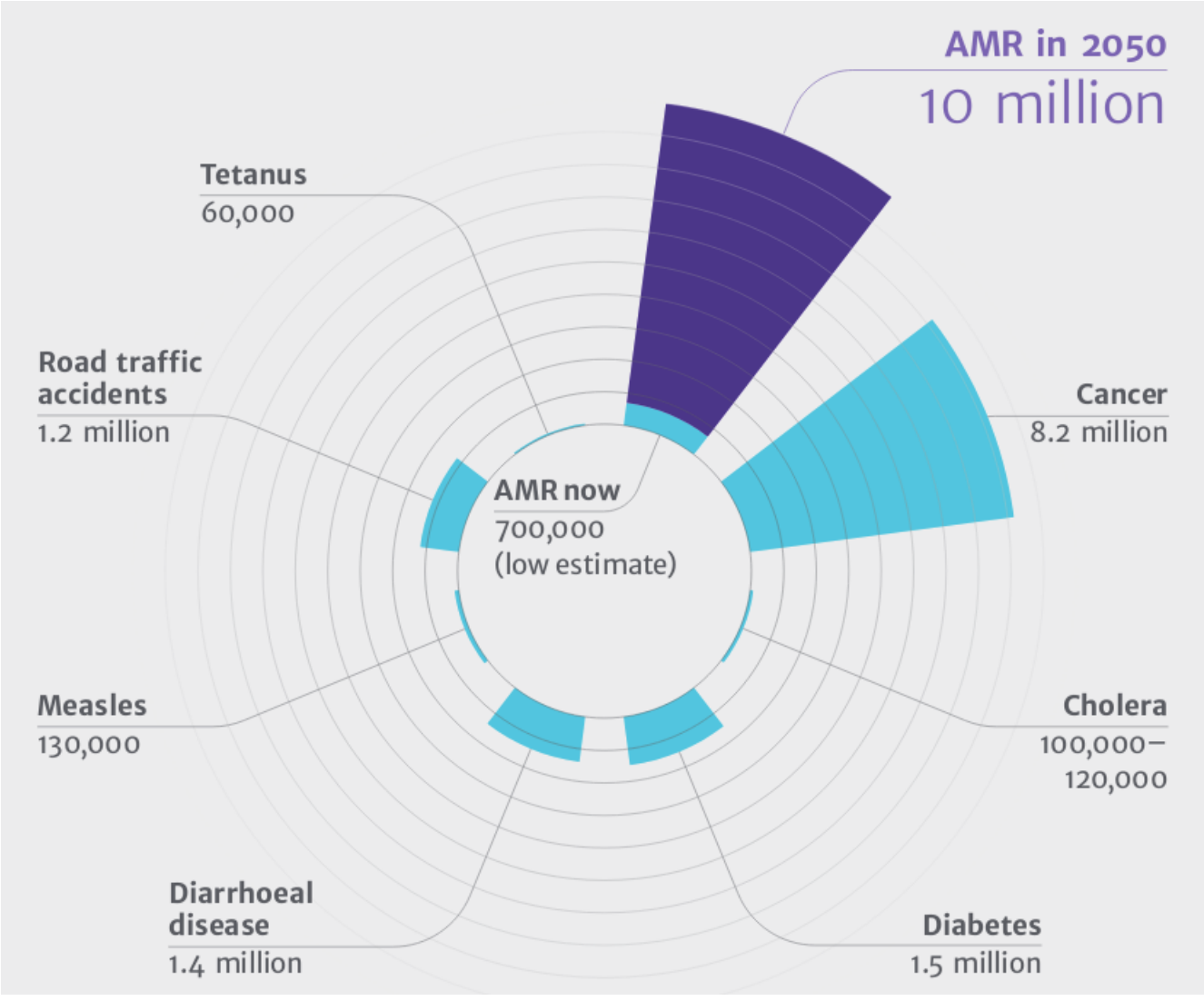
on Antimicrobial Resistance, R. (2016). *Tackling drug-resistant infections globally: final report and recommendations*. Review on antimicrobial resistance.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., and Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature microbiology*, 2(11):1533.

Stairs, J., Bal, N., Maguire, F., and Scott, H. (2019). A resident-led clinic that promotes the health of refugee women through advocacy and partnership. *Canadian Medical Education Journal*.

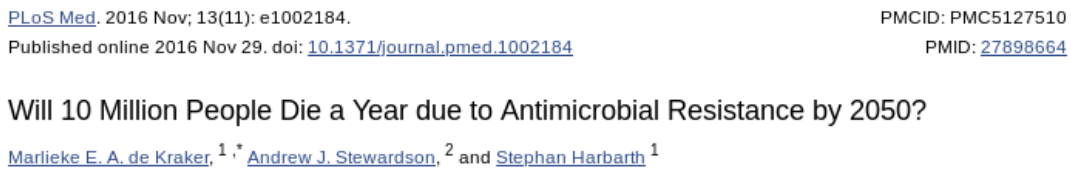
Backup

10 million deaths?



(on Antimicrobial Resistance, 2016), (de Kraker et al., 2016)

10 million deaths?



PLoS Med. 2016 Nov; 13(11): e1002184. PMCID: PMC5127510
Published online 2016 Nov 29. doi: 10.1371/journal.pmed.1002184 PMID: 27898664

Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?

Marlieke E. A. de Kraker,[1,*] Andrew J. Stewardson,[2] and Stephan Harbarth[1]

(on Antimicrobial Resistance, 2016), (de Kraker et al., 2016)

Where does 10 million come from?

For 3rd-generation cephalosporin resistant *E. coli*, *K. pneumoniae*, and MRSA:

- Estimate global BSIs (multiply average incidence in tertiary European hospitals by global population).
- Estimate AMR (proportion of resistant blood-cultures per country)
- Extrapolate to other infections sites (via relative incidence to BSI in 2 studies n=16 BSIs)
- Estimate attributable mortality rates from adjusted odds-ratios in an unspecified manner.
- Assume no change in mortality, 40% increase in resistance, and doubled infection rates by 2050.