Resistance Gene Identifier (RGI) - Prediction of antimicrobial resistance genes and mutations for genomic and metagenomic sequencing data

Amogelang R. Raphenya[1], Tammy T. Y. Lau[2], Brian Alcock[1], Kara K. Tsang[1], Finlay Maguire[3], Fiona S. Brinkman[4], Robert G. Beiko[3], Andrew G. McArthur[1]

[1]The Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada

[2]B.C. Cancer Agency, Vancouver, British Columbia, Canada

[3]Faculty of Computer Science, Dalhousie University, Halifax, Canada

[4]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

Introduction

Antimicrobial resistance (AMR) is a global health concern. Improved sequencing technology offers a solution, but maximizing its utility requires the Comprehensive Antibiotic Resistance Database[a] (CARD; card.mcmaster.ca) and the Resistance Gene Identifier software (RGI; card.mcmaster.ca/analyze/rgi or github.com/arpcard/rgi)

Methods

CARD is organized based on the Antibiotic Resistance Ontology and its primary classes are: AMR Gene Family, Drug Class, and Resistance Mechanism. RGI was used to mine Genbank, and the results were used to produce a catalog of predicted AMR alleles called *in silico* variants.
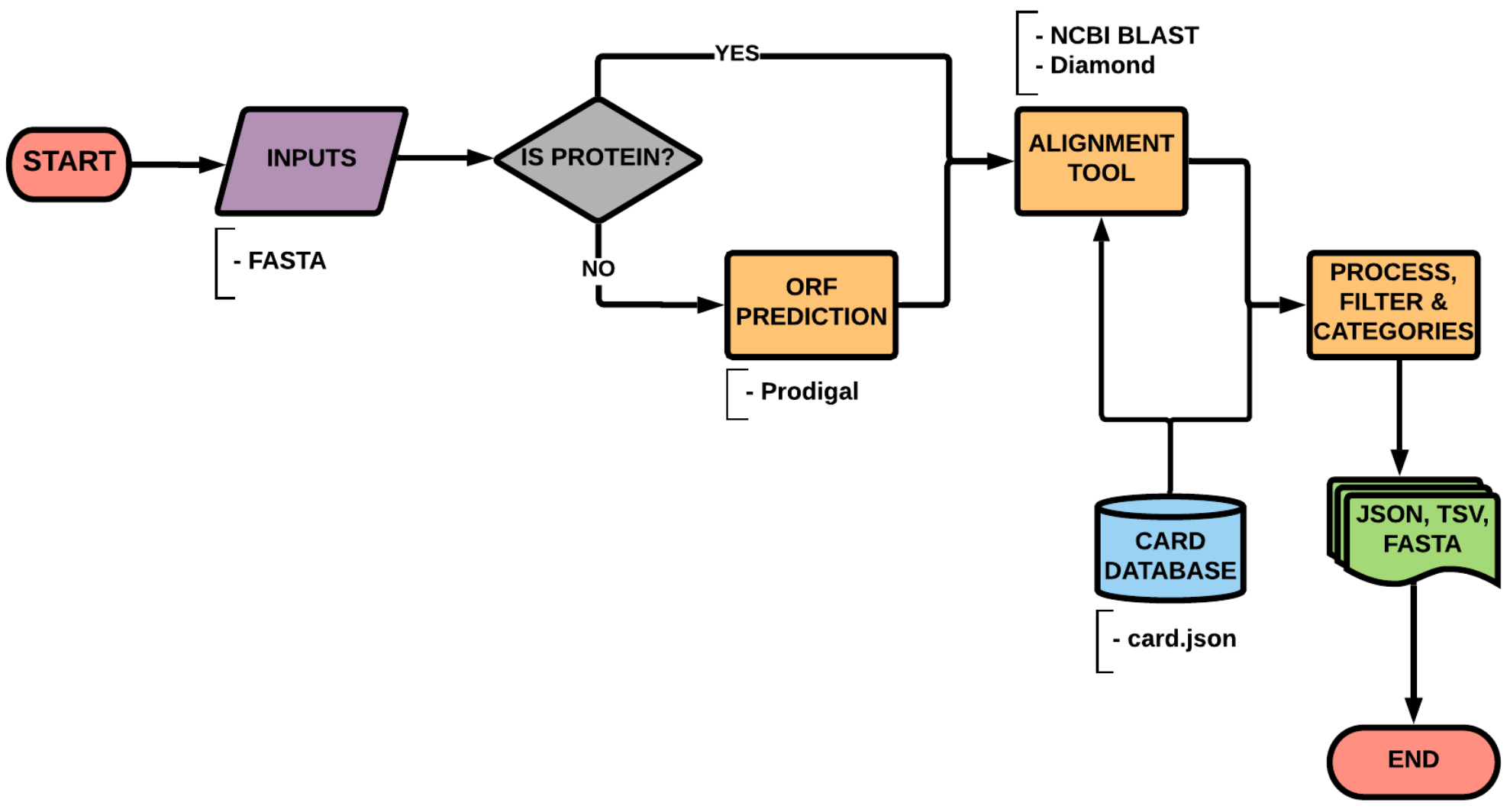

Figure 1. RGI analyzes genomes under three paradigms: **Perfect**, **Strict**, and **Loose**. **Perfect** hits match the curated reference sequences in CARD. The **Strict** algorithm identifies previously unknown variants of known AMR genes, and it conducts a secondary screen for mutations using detection models with CARD's curated bitscore cut-offs. The **Loose** hits fall below the detection model cut-offs. The Loose algorithm detects putative emergent AMR genes.
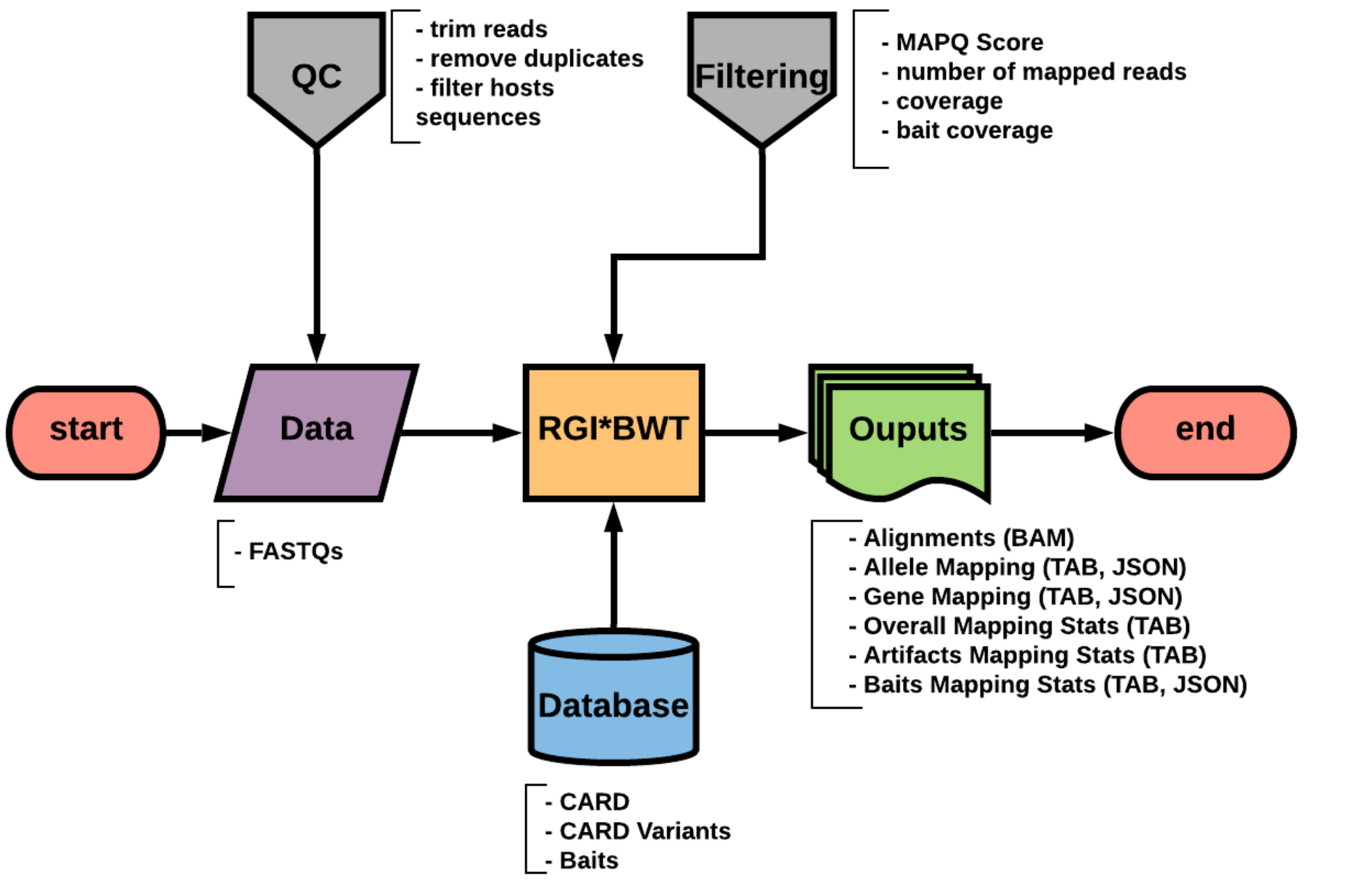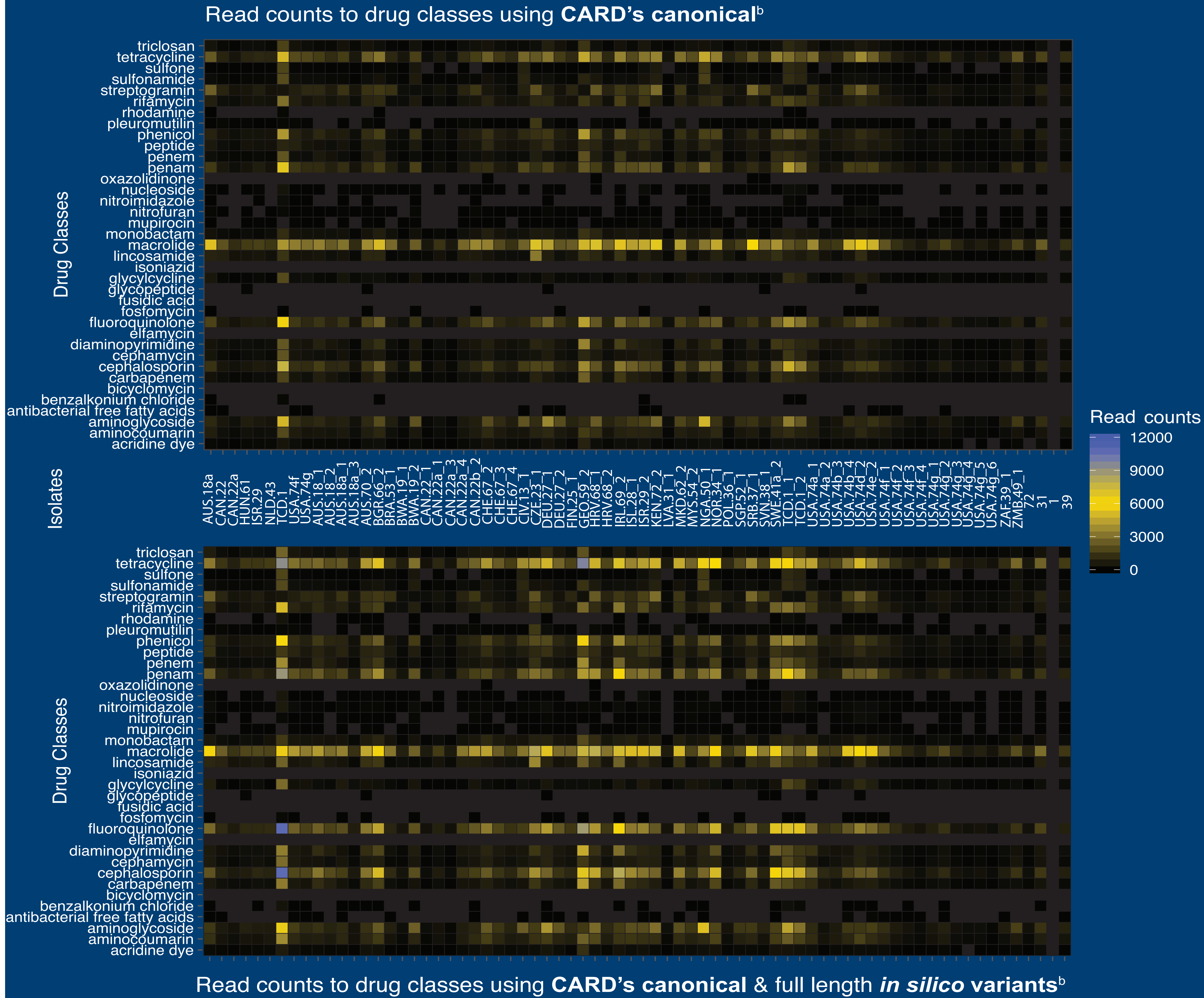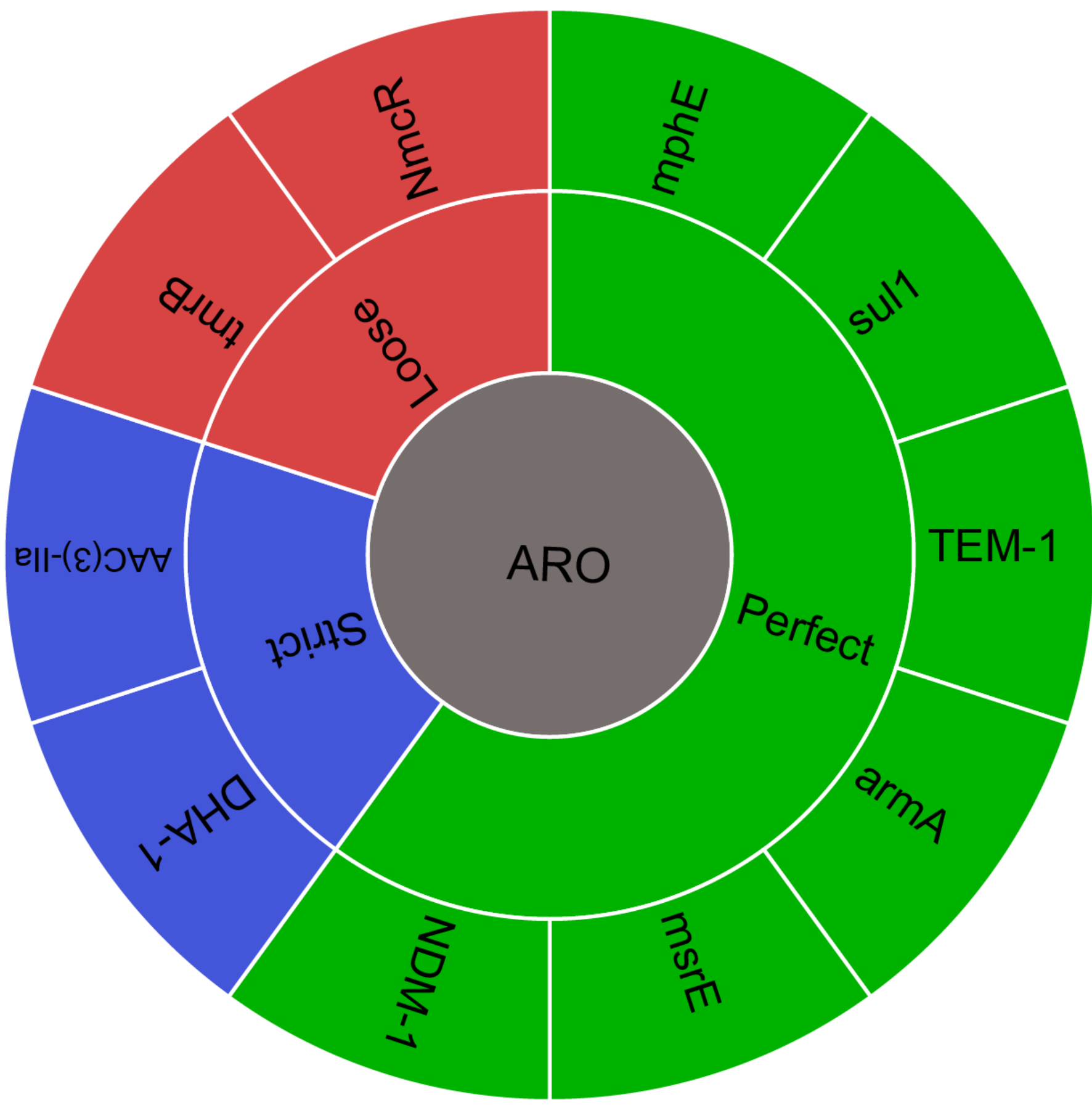

Figure 2. In metagenomic analyses, RGI uses both **CARD's canonical** and *in silico* variants for resistome prediction.

RGI predicts AMR genes in genomes with high confidence.

In silico variants are needed to predict AMR genes in metagenomes.

Read counts to drug classes using CARD's canonical[b]

Read counts to drug classes using CARD's canonical & full length *in silico* variants[b]

Results

[a]**Figure 3**. Predicted AMR genes in *Escherichia coli* strain HK-01 plasmid pNDM-HK, HQ451074.1 using RGI for genomes. All hits are confirmed from NCBI annotations.

Future directions

- Add **Perfect**, **Strict** and **Loose** paradigms for RGI metagenomic analyses
- Include a screen to detect AMR single nucleotide polymorphisms (SNPs)
- Develop a feature to distinguish Gram-positive and Gram-negative AMR genes

Acknowledgements

References

[a]Jia, B., et al. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database.

[b]Hendriksen, R.S., et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage.

Take a picture to See more details at https://github.com/arpcard/rgi