

Metagenome Assembled Genome Binning Methods With Short Reads Disproportionately Fail For Plasmids and Genomic Islands

Baofeng Jia^{1†}, Finlay Maguire^{2,3†}, Kristen L. Gray¹, Wing Yin Venus Lau¹, Robert G. Beiko², and Fiona S. L. Brinkman¹

[†] Authors contributed equally. 1. Department of Molecular Biology and Biochemistry, Simon Fraser University, Canada. 2. Faculty of Computer Science, Dalhousie University, Halifax, Canada. 3. Department of Community Health & Epidemiology, Faculty of Medicine, Dalhousie University

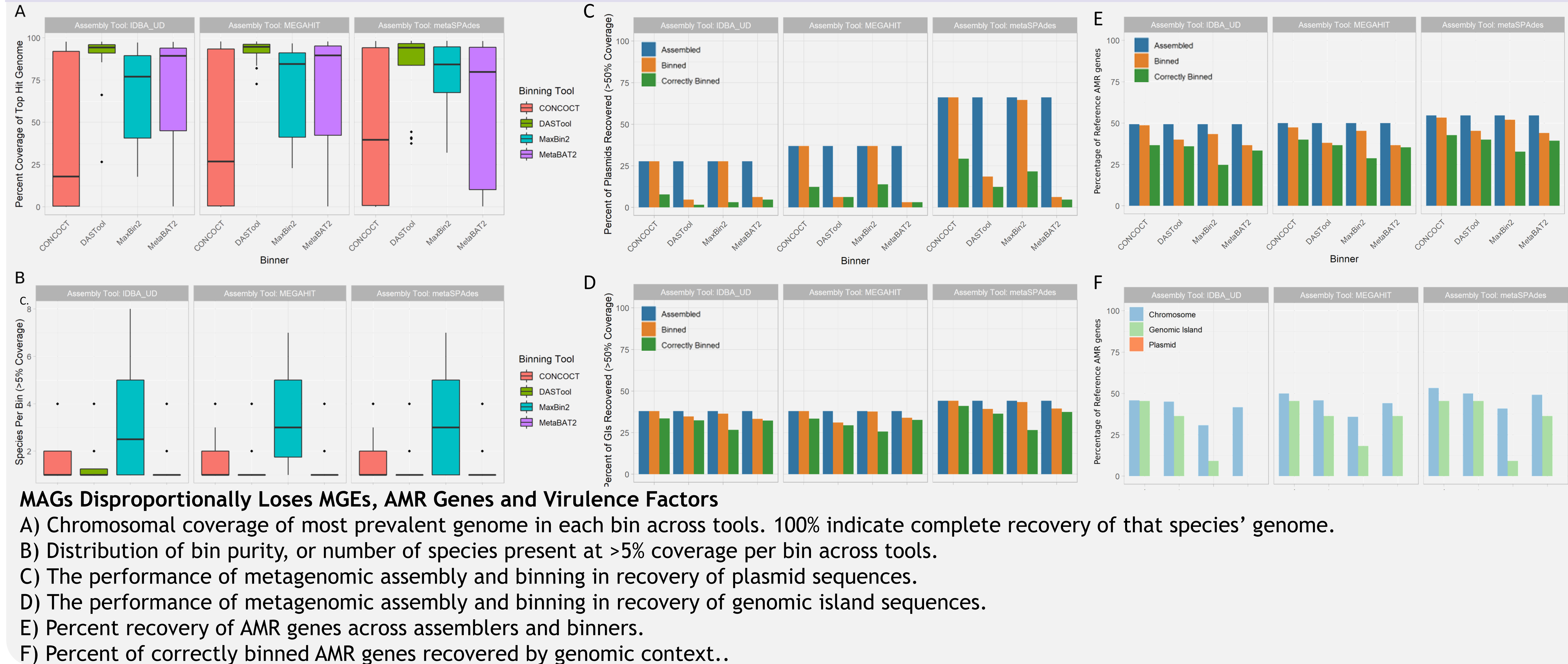
INTRODUCTION

- Metagenomic methods enable the simultaneous characterization of microbial communities without time-consuming and bias-inducing culturing.
- Metagenome-assembled genome (MAG) binning methods aim to reassemble individual genomes from this data.
- Recovery of mobile genetic elements (MGEs), such as plasmids and genomic islands (GIs), by binning has not been well characterized.
- GIs and plasmids have proven particularly difficult to assemble from short-read sequencing data¹.
- Antimicrobial resistance (AMR) genes and virulence factors (VF) often associate with MGEs, studying their transmission is a public-health priority^{2,3}.

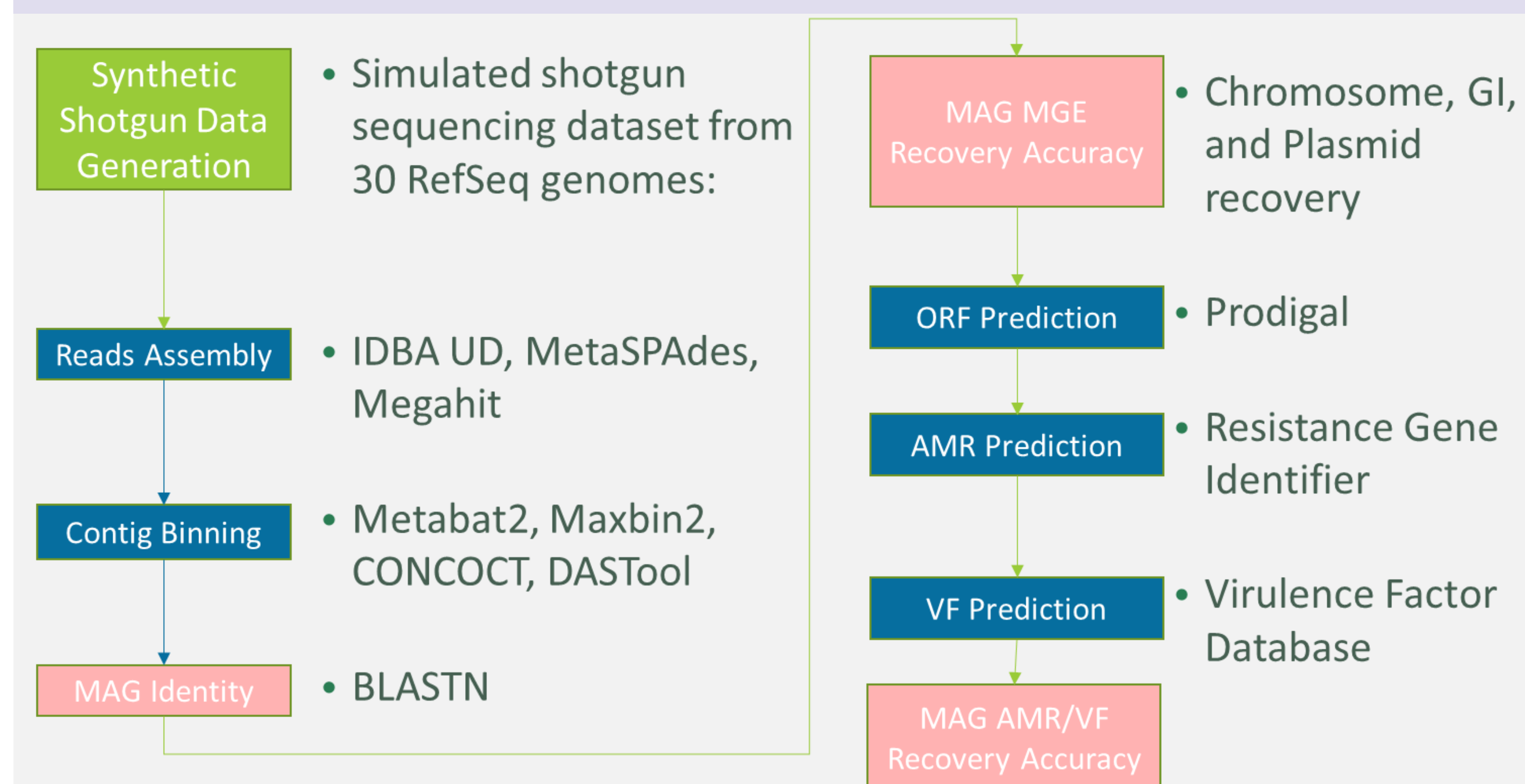
Hypothesis:

The variable copy number and sequence composition of MGEs makes them problematic for MAG binning methods and thus will not be correctly recovered, resulting in a disproportionate loss of AMR genes and VFs.

MAGs Provide Contextual Information but Disproportionally Lose MGEs, AMR Genes and Virulence Factors



METHOD

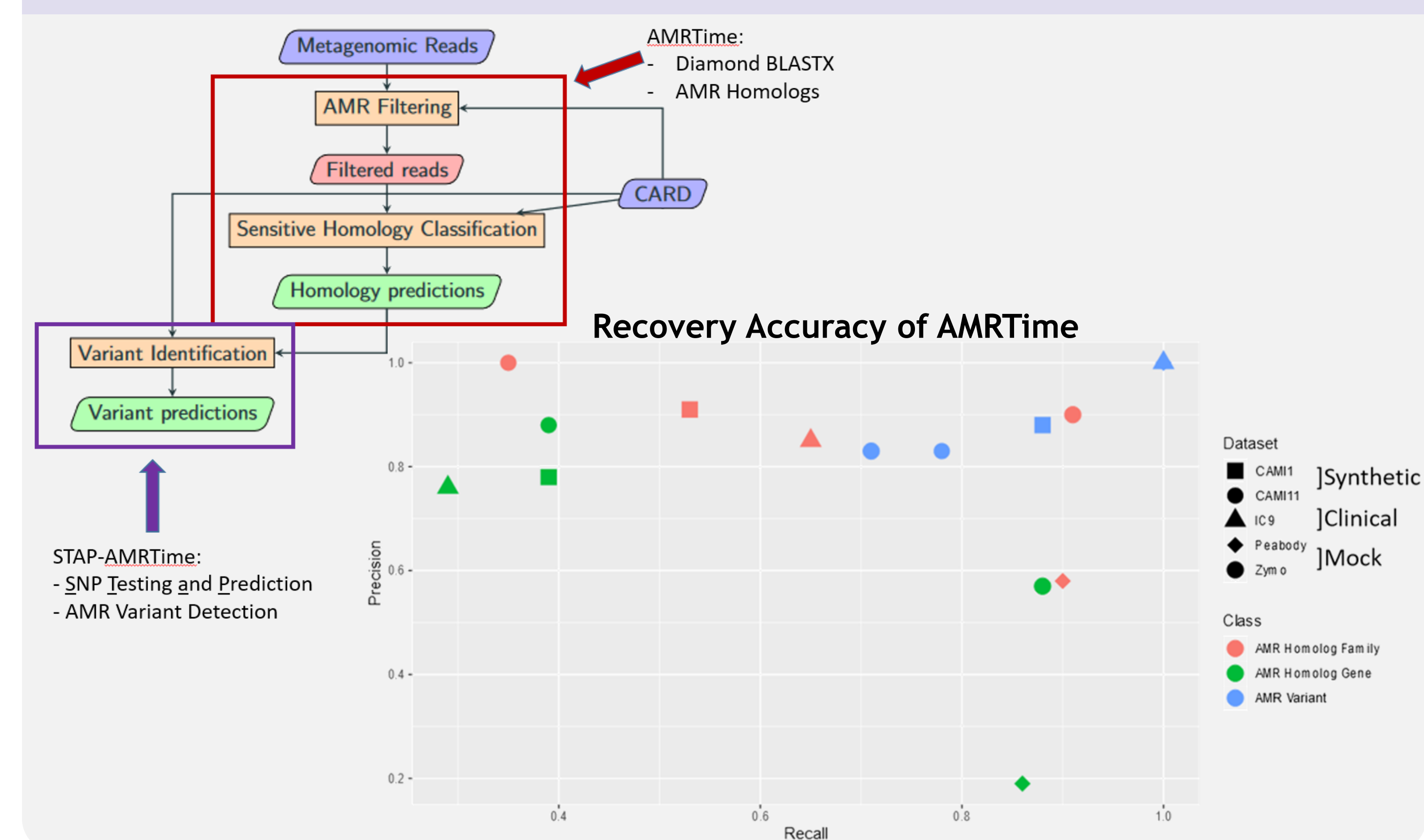


Schematic of the Evaluation MAG Binning Methods

SUMMARY & FUTURE DIRECTIONS

- Short-read MAG-binning approaches provide a useful tool to study a bacterial species' core chromosomal elements. However, they have severe limitations in the recovery of MGEs.
- The majority of these MGEs will either fail to be assembled or be incorrectly binned.
- There is a disproportionate loss of key MGE-associated VF and AMR genes in clinically relevant pathogens.
- It is vital we utilize MAGs in conjunction with other methods (e.g. long-read sequencing, read-based sequence homology) before drawing most biological or epidemiological conclusions.

AMRTime: Read-Based AMR Gene Predictor



Funding & Acknowledgements



References

- Williams KP. Nucleic Acids Res. 2002;30:866–875. doi: 10.1093/nar
- Ho Sui SJ, et al. PLoS One. 2009;4:e8094. doi: 10.1371/journal.pone.0008094
- Von Wintersdorff CJH, et al. Front Microbiol. 2016;7:173. fmicb.201
- Maguire F. and Jia B. et al. Microb Genom. 2020 Oct; 6(10): mgen000436



Scan For Copy