
Catrione Lee z,b, Rodrigo Ortega Polo a, Rahat Zaheer a, Gary Van Domselaar c, Athanasios Zovoilis b, and Tim A. McAllister a,*

a Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Government of Canada, 5403 1st Avenue South, Lethbridge, AB, T1J 4B1, Canada
b Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, AB, T3M 2L7, Canada
c National Microbiology Laboratory, Public Health Agency of Canada, Government of Canada, 1015 Arlington Street, Winnipeg, MB, R3E 3R2 Canada

* Correspondence: tim.mcallister@agr.gc.ca
Highlights

- Targeted assembly methods are not suitable for association studies on taxonomy, ARGs, and MGEs using 100 bp Illumina short-reads
- MAG-based analyses into AMR do not provide comprehensive coverage of ARG diversity, relative abundance, or overall prevalence to adequately inform surveillance projects
- MGEs are too repetitive to be accurately detected in quality filtered sequence data (especially for short-reads) and need longer sequences to establish the context of surrounding genes

Abstract

Antimicrobial resistance genes (ARGs) can be transferred between members of a bacterial population by mobile genetic elements (MGE). Understanding the risk of these transfer events is important in monitoring and predicting antimicrobial resistance (AMR), especially in the context of a One Health Continuum. However, there is no universally accepted method for detection of ARGs and MGEs, and especially for determining their linkages. This study used publicly available shotgun metagenomic DNA short-read (Illumina, 100 bp paired-end) sequence data from samples across the One Health Continuum (including beef cattle composite feces from feedlots, catch basin water at feedlots, agricultural soil from feedlot manured surrounding fields, and urban/municipal sewage influent from two municipal wastewater treatment plants) to develop a workflow to identify and associate ARGs and MGEs. ARG- and MGE-based targeted-assemblies with available short-read data were unable to meet this analysis goal. In contrast, de novo assembly of contigs provided enough sequence context to associate ARGs and
MGEs, without compromising discovery rate. However, to estimate the relative abundance of these elements, unassembled sequence data must still be used.

Keywords

metagenomics, antimicrobial resistance, mobile genetic elements, targeted assembly, de novo assembly

Data Summary

The DNA sequence data used in this study was accessed from BioProject IDs PRJNA420682, PRJNA529711, PRJNA507800 and PRJNA482680 [DOI: 10.1186/s12866-019-1548-x (Zaheer et al., 2019)]. All analysis was performed using in-house code or through the National Microbiology Laboratory (Public Health Agency of Canada)’s instance of Galaxy. All parameters for specific tools are described in the text.
1. Introduction

Antimicrobial resistance (AMR) is a global health issue that should be addressed using a One Health approach (McEwen & Collignon, 2018; Shallcross & Davies, 2014). One Health considers the health of three sectors, humans, animals, and the ecosystem as a single continuum (Behravesh, 2019). One Health can be applied to the pathogenicity of microbial infections and the associated risks of antimicrobial resistance causing treatment failure. Antimicrobial resistance (AMR) is a concern to all One Health sectors, but particularly in the human and animal sectors where over-prescription and misuse of antimicrobials are of major concern (Ramachandran et al., 2019; Xiong et al., 2018).

Antimicrobial resistance genes (ARGs) are encoded within the genomes of both commensal and pathogenic bacteria (human, agricultural, and environmental). They can be transferred horizontally (Dröge et al., 1998) as mediated by mobile genetic elements (MGEs) (Frost et al., 2005). Integrative and conjugative elements (ICEs) are among the most common MGEs. They can integrate into the bacterial chromosome, enabling them to be replicated and transferred to daughter cells during cell division. However, ICEs can also excise from the chromosome and transfer to other bacteria through conjugation (Burrus et al., 2002). ICEs have been shown to harbour various cargo genes, including ARGs in bacteria of greatest health concern (Botelho & Schulenburg, 2020; Partridge et al., 2018). These pathogenic bacteria are most frequently members of the Firmicutes and Proteobacteria (Botelho et al., 2020; Botelho & Schulenburg, 2020; Farzand et al., 2019; Partridge et al., 2018). Other MGEs that can play a role in the transfer of ARGs include plasmids, transposons, and Class I integrons. Due to their importance in transferring ARGs from commensal to potentially pathogenic bacteria,
Understanding the linkage of MGEs and ARGs is integral to developing approaches to reduce AMR (Slizovskiy et al., 2020).

Integrating ARGs and their association with MGEs into AMR surveillance and risk assessment requires assessing the co-localization of the mobilome and resistome. This process is tedious if done through clonal isolation and whole-genome sequencing (WGS). Alternatively, shotgun metagenomics can facilitate high-throughput analyses for AMR surveillance (Pillay et al., 2022; Sherry et al., 2023). Short-read sequencing technologies are much more affordable and have lower error rates than long-read approaches, making them popular for genomic surveillance. However, the read lengths achievable by these technologies limits their ability to characterize ARGs in their surrounding genomic context.

A variety of tools and strategies are being developed or proposed that utilize metagenomic data for AMR surveillance (Cheng et al., 2022; Juraschek et al., 2021; Nobrega et al., 2021; Pillay et al., 2022; Sherry et al., 2023; Zankari et al., 2017). Targeted and de novo assembly methods have been proposed, along with binning of metagenome-assembled genomes (MAGs) based on similar relative abundance and sequence composition. Targeted assembly uses an algorithm that constructs contiguous sequences (contigs) from sequenced reads via alignment to a guiding sequence scaffold. De novo metagenomic assembly constructs contigs without a guiding scaffold and provides a more comprehensive metagenome without database biases, but at the cost of reduced coverage. MAG binning is an additional construction step that generates hypothetical genomes to closely approximate the biological genomes in a metagenomic sample (Bowers et al., 2017). This study aims to use short-read sequences to determine which of these methods, assembled contigs (targeted or de novo) or binned MAGs is most conducive to determining the context of ARGs to MGE.
2. Materials and Methods

2.1. The Dataset

A previously generated dataset of sequenced metagenomic samples was selected for this study (Zaheer et al., 2019). This dataset originated from four different sample types: pen fecal composites (FC) collected at cattle feedlots, catch basin water (CB) at feedlots, soil from surrounding fields that had been fertilized with feedlot manure, and urban/municipal sewage influent from two municipal wastewater treatment plants. Samples were sequenced using Illumina HiSeq2000 to generate short, 100 bp paired end reads (Zaheer et al., 2019). Sequence Read Archive (SRA) accession numbers for paired-end short-read sequenced metagenomic samples (n=43, including FC n=20, CB n=13, soil=4, municipal sewage influent n=6) were accessed from the National Center for Biotechnology Information (NCBI) SRA database from the BioProject IDs PRJNA420682, PRJNA529711, PRJNA507800 and PRJNA482680. For trial-and-error optimization of targeted assembly, de novo assembly as well as AMR and MGE co-localization, a single fecal composite sample (SRR6512893) was selected based on its high abundance of Proteobacteria and Firmicutes as these phyla tend to have a high abundance of MGEs (Botelho et al., 2020; Botelho & Schulenburg, 2020; Farzand et al., 2019; Partridge et al., 2018). Therefore, selecting a sample with a high number of MGEs with adequate coverage increased the likelihood of identifying co-localized MGEs and ARGs. The criteria for a successful method were: 1) number of assemblies was comparable to the literature, 2) co-localization was confirmed, and 3) automatable and replicable workflows to ensure
reproducibility for large datasets. All samples were used to compare the overall microbiota composition and quality distribution of MAGs.

2.2. ARG and ICE Detection via Sequence Targeted Assembly

Sequence-targeted assembly of the metagenomic samples was accomplished with MetaCherchant v0.1.0 (Olekhnovich et al., 2018) with ICE and ARG sequences in the FC sample SRR6512893 assembled using the ICEberg 2.0 (Liu et al., 2019) and MEGARes 2.0 (Lakin et al., 2017) databases, respectively. Antimicrobial resistance that was linked to genes with a single-nucleotide polymorphism (SNP) and genes involved in the regulation of ARGs were excluded from the analysis.

2.3. De novo Assembly and Binning of MAGs

De novo assembly of all metagenomic samples was accomplished using the nf-core/mag pipeline v1.0.0 (Krakau et al., 2022), which also performs binning of MAGs. The nf-core pipeline (Ewels et al., 2020) is based on the Nextflow workflow manager (Di Tommaso et al., 2017), which uses two assemblers, MEGAHIT (Li et al., 2015; Li et al., 2016) and metaSPAdes (Nurk et al., 2017) to perform parallel assemblies and subsequent binning by metaBAT2 (Kang et al., 2019) (Figure 1). The two assemblers were used to compare assembly quality, particularly the number of quality filtered contigs and the completeness of subsequent MAGs. High-quality was defined as >90% completeness with <5% contamination; medium-quality was defined as >50% completeness with <10% contamination, with parameters below this considered to be of low-quality (Bowers et al., 2017).
The following parameters were supplied to nf-core/mag: a mean quality of 20 and trimming quality of 20 for fastp (Chen et al., 2018), Kraken 2 (Wood et al., 2019) database limited to 8 GB (constructed on 2019-04-01), the Contig Annotation Tool (CAT) (von Meijenfeldt et al., 2019) (generated on 2020-06-18), and a minimum contig size of 1500 bp for both MEGAHIT and metaSPAdes. BUSCO (Simão et al., 2015) was used to identify contigs containing bacterial genes along with bacterial reference database “bacteria_odb9” which contained 148 single-copy bacterial genes from 3,663 species. All other tools included in nf-core/mag [Bowtie2 (Langmead & Salzberg, 2012), FastQC, PRODIGAL (Hyatt et al., 2010), QUAST (Gurevich et al., 2013), BUSCO (Simão et al., 2015), CheckM (Parks et al., 2015), MultiQC (Ewels et al., 2016)] were run using default parameters.

**Figure 1.** Workflow of the nf-core/mag pipeline depicting steps from DNA extraction to the point of assembly of metagenomics assembled genomes (MAGs).
2.4. ARG and ICE Detection via de novo Assembly

Metagenome-assembled genomes (MAGs) binned from the random FC sample SRR6512893 were locally aligned with ABRicate (Seemann, 2020) using default parameters to the MEGARes 2.0 (Doster et al., 2020) database to detect ARGs. These same MAGs were aligned with BLASTn (Johnson et al., 2008) to the ICEberg 2.0 database to detect ICEs. The MAGs containing ARGs were taxonomically classified with the Bin Annotation Tool (BAT) (von Meijenfeldt et al., 2019). The MAG annotations containing both ARGs and ICEs were visualized using Geneiousv10.2.5 to determine synteny. To compare BLASTn-based methods to another method such as ABRicate, the same MAGs from FC sample SRR6512893 were passed through Staramr (Bharat et al., 2022) with default parameters. Staramr was run using ResFinder (Florensa et al., 2022) and PlasmidFinder (Carattoli et al., 2014) databases. ARGs with the tag ‘SNP-confirmation required’ were filtered out as no SNP analysis was performed to identify point mutations in housekeeping genes. Similarly, ARGs with the ‘regulator’ tag were excluded because while regulatory genes modulate AMR expression, they do not result in resistance if ARGs are absent.

To detect colocalized ARGs and MGEs, all MAGs binned from the random FC sample SRR6512893 were supplied to the tool MOB-recon from the package MOB-suite (Robertson & Nash, 2018). The MOB-recon tool generated plasmid sequences along with information on predicted transferability, replicon family, and relaxase type. Reconstructed plasmid fasta files containing contigs were supplied to Staramr which included ResFinder and PlasmidFinder using default parameters. ResFinder identified ARGs that were colocalized on the previously identified plasmids.
2.5. Microbiota Comparison between Methods

Quality-filtered sequence data were taxonomically classified using Kraken 2 (Wood et al., 2019) to evaluate contig taxonomic classifications. The contig annotation tool (CAT) (von Meijenfeldt et al., 2019) was used to classify assembled contigs from MEGAHIT and metaSPAdes. Taxon counts were generated using phyloseq (McMurdie & Holmes, 2013) in the R v4.0.1 statistical programming language. Non-metric multidimensional scaling (NMDS) was used to visualize ordination between all classification methods across all samples. Non-metric multidimensional scaling was done at the phylum level to capture the majority of the diversity of the taxa present within the dataset.

3. Results

3.1. Targeted Assembly

MetaCherchant successfully identified 333 ARGs from the random single FC sample SRR6512893 (Figure 2). The most prevalent classes were tetracyclines (43.8%), MLS (macrolide, lincosamide, streptogramin; 18.3%), sulfonamides (17.4%), beta-lactams (5.7%) and aminoglycosides (4.8%). Additionally, a total of 173 putative ICE were found to be associated with at least one ARG within this sample (Figure 3).
Figure 2. Distribution of antimicrobial resistant gene (ARG) classes. A total of 333 ARGs were detected in fecal composite sample SRR6512893 (n=1).
Figure 3. Distribution of 173 integrative conjugative element families associated with an antimicrobial resistant gene detected in fecal composite sample SRR6512893 (n=1).

3.2. Taxonomic Comparison between Reads and Contigs

Non-metric multidimensional scaling (NMDS) of all fecal composite (Figure 4A), catch basin (Figure 4B) and soil samples (Figure 4C) revealed three distinct clusters with all three classification methods. In municipal sewage influent (Figure 4D), there was no distinction between the contig assembly methods, with Kraken separating out.
Figure 4. Non-metric multidimensional scaling of the phylum composition of each of the four environments’ [A) fecal composite, n=20; B) catch basin, n=13; C) soil, n=4; and D) municipal sewage influent, n=6] prokaryotic communities compared by taxonomic classification method. Depicting Kraken 2’s classification of short-reads (circles), CAT’s classification of MEGAHIT assembled contigs (triangle), and CAT’s classification of metaSPAdes assembled contigs (square).

3.3. De novo Assembly and MAG Binning

We assessed the binning quality from all environments using all samples. MEGAHIT produced 327 high-quality and 1848 medium-quality MAGs, while metaSPAdes produced 267 high-quality and 1925 medium-quality MAGs (Figure 5). Across all environments and assemblers, there were more medium-quality than high-quality MAGs per sample. The fecal
composite and catch basin samples had the highest number of binned MAGs per sample. Fecal samples averaged 52.3 mean high-quality and 58.4 mean medium-quality MAGs, and CB samples averaged 47.2 mean high-quality and 49.6 mean medium-quality MAGs. Soil samples had the lowest mean number of recovered high-quality and medium-quality MAGs.

![Mean distribution of high and medium-quality MAGs in fecal composite, catch basin, soil, and municipal sewage influent. Error bars represent standard deviation.](image)

**Figure 5.** Mean distribution of high and medium-quality MAGs in fecal composite, catch basin, soil, and municipal sewage influent. Error bars represent standard deviation.

### 3.4. ARG and MGE Co-localization from MAGs

The same random FC sample, SRR6512893, used in section 3.1 was the subset of MAGs selected for further co-localization analyses to provide consistent comparison across all optimization methods. MEGAHIT assembled ten high-quality and 64 medium-quality MAGs. In contrast, metaSPAdes assembled five high-quality and 69 medium-quality MAGs. Alignment of MAGs with BLASTn to the MEGARes database identified one high-quality and four medium-quality MAGs that contained at least one ARG (Table 1). Notably, only the medium-quality MAGs contained ICEs, regardless of their association with an ARG. One medium-quality MAG
derived from MEGAHIT and one medium-quality MAG derived using metaSPAdes, both possessed tetZ and sul1 and were classified as Tessaracoccus sp. One of the MAGs possessed 23 different ARGs. This MAG could not be classified beyond the Bacterial domain. The counterpart for this MAG was not found in the MEGAHIT assemblies.

Table 1. Antimicrobial resistance genes, mobile genetic elements\ and bin classifications of high-quality and medium-quality metagenome-assembled genomes binned from the representative fecal composite sample, SRR6512893.

<table>
<thead>
<tr>
<th>Assembler</th>
<th>MAG Quality</th>
<th>MAG ID</th>
<th>ABRicate ARGs</th>
<th>MAG Taxonomy</th>
<th>ICE Aligns w/ ARGs</th>
<th>Staramr ARGs</th>
<th>MOB-recon Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEGAHIT</td>
<td>High</td>
<td>29</td>
<td>mphN</td>
<td>Clostridiales</td>
<td>0 (no ARG align)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>72</td>
<td>tetW</td>
<td>Firmicutes</td>
<td>1 tetZ; 3 sul1</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>30</td>
<td>tetZ, sul1</td>
<td>Tessaracoccus sp.</td>
<td>1 (no ARG align)</td>
<td>tetZ, sul1</td>
<td>-</td>
</tr>
<tr>
<td>metaSPAdes</td>
<td>Medium</td>
<td>74</td>
<td>tetZ, sul1</td>
<td>Tessaracoccus sp.</td>
<td>1 (no ARG align)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>149</td>
<td>23 ARGs – MDR</td>
<td>Bacteria</td>
<td>2 (no ARG align)</td>
<td>-</td>
<td>pCM959, IncFIA</td>
</tr>
</tbody>
</table>

1MDR (multi-drug resistance; defined as resistance to >3 classes of antimicrobials).

Analysis of the MAGs with Staramr and ABRicate yielded different results (Table 1 and 2). Only MAG MEGAHIT-30 came back with an AMR-positive profile using Staramr. The same ARGs, tetZ and sul1, were detected with both methods. Staramr’s output from PlasmidFinder (Carattoli et al., 2014) showed that only metaSPAdes-149 contained the plasmid IncFIA. This finding was supported by the three ‘non-mobilizable’ plasmids reconstructed with MOB-recon, all of which were present in metaSPAdes-149, one of which was also IncFIA. No AMR-positive contigs were identified by Staramr, suggesting that the reconstructed plasmids did not contain ARGs. A combination of ABRicate and BLASTn yielded four more ARG- and MGE-positive MAGs than Staramr and MOB-recon.
4. Discussion

Targeted assembly methods to detect ARGs and MGEs like MetaCherchant (Olekhnovich et al., 2018) could be more computationally efficient than de novo assembly approaches such as metaSPAdes or IDBA-UD (Calderón-Franco et al., 2022; Slizovskiy et al., 2020). Leveraging the efficient, focused approach of targeted assembly could simplify and streamline downstream data analysis. Targeted assemblers such as MetaCherchant (Olekhnovich et al., 2018) rely on a reference sequence database to generate assemblies. Targeted assembly methods can also provide context of flanking sequences in relation to reference sequences, which can be useful for defining the association ARG to MGEs. Comparing the resistome from targeted assemblies to the original read-based resistome of fecal samples revealed a remarkable similarity in the composition of ARGs (Zaheer et al., 2019). This comparison identified tetracyclines ARGs as being predominant, followed by those associated with MLS, betalactams, and aminoglycosides. Sulfonamide ARGs were the only class to be detected in the targeted assembly, but not in the read-based resistome.

The reference-based assembly approach employed in this study identified 173 putative ICEs, but their linkage with ARGs could not be confirmed, due to the lack of overlapping reads between ARG and ICE. Thus, this method failed to achieve the goal of co-localization of ARGs and MGEs in metagenomes, but larger sequences as obtained from de novo assembly and/or MAG binning were able to establish this linkage. Others have also been able to use MAGs derived from full-scale granular sludge along with human and cattle fecal samples to establish linkages between ICE and ARGs (Calderón-Franco et al., 2022; Slizovskiy et al., 2020).
Overall taxonomic classifications differed depending on the classifier tool used. The largest differences were seen between Kraken 2 and both CAT classifications (from MEGAHIT and metaSPAdes), as opposed to assembler type approaches (MEGAHIT vs metaSPAdes). Kraken 2 uses an exact alignment of $k$-mers to assign reads to the lowest common ancestor (Wood et al., 2019). In contrast, CAT and BAT use an open reading frame (ORF) algorithm to identify translated protein homologs and assign classification based on the NCBI non-redundant protein database (von Meijenfeldt et al., 2019). Another explanation for these differences could arise from the fact that Kraken 2 (MiniKraken) and CAT/BAT (NCBI-nr protein database) utilize different taxa classification databases. Our finding of the different classifications between the read-based and assembly-based approaches is not surprising, given that the coverage of taxa abundance differed between reads and contigs, with reads having higher coverage. A prior study found that MEGAHIT only assembled 55.8% of available reads (Li et al., 2015), whereas CAT and other contig-based classifiers had higher precision than the read-based classifier Kaiju [0.6-0.7 precision in contig-based versus 0.1-0.25 in read-based (von Meijenfeldt et al., 2019)]. This finding indicates that while assembly-based methods used in combination with contig classifiers may have higher precision than short-read classifiers, not all diversity is captured due to the substantial number of unincorporated reads. We observed this to be the case in our NMDS analysis that showed clear clustering of read-based versus assembly-based taxonomic methods across all samples (Figure 4).

The recovery rate of MAGs that passed quality benchmarking (Bowers et al., 2017) was as expected (8.4 and 7.0 average MAGs per sample for MEGAHIT and metaSPAdes, respectively) and was comparable to another study that recovered 1,150 high-quality dereplicated MAGs at rate of 6.4 MAGs per sample from 179 metagenomic samples (Holman et al., 2022).
Our study selected only MAGs with a cut-off minimum length of 1,000 bp, to filter out low-quality MAGs. This level of stringency should have allowed for the detection and association of even highly repetitive MGEs, ranging from small miniature inverted repeat transposable elements (50–800 bp) (Crescente et al., 2018) to prophages and transposons (6–358 kbp) (Gao et al., 2020). However, we found little to no co-localization between ARGs and MGEs, despite ARGs having sequences ranging from 237 bp (dfrB1) to 1,992 bp (otr(A)) (van Hoek et al., 2011). This finding could be due to the comparatively low recovery of MGE, whether co-localized with an ARG or not. While ARGs with the potential for mobilization were defined as being on the same contig as a MGE, those that were on the same MAG but on different contigs could also be potentially mobile as some MGEs exceeded 300 kbp, a size much larger than the average contig in this study.

We found a range of taxonomic classification levels within FC sample SRR6512893, due to the variability in the taxonomic specificity of the assembled contigs. This variable level of taxonomic assignment is comparable to other studies that looked at swine feces (Holman et al., 2022) and activated sludge (Holman et al., 2022; Zhao et al., 2020). The MAG metaSPAdes-149 were unique from the other assembled MAGs as it contained 23 different ARGs and had three non-mobilizable plasmids that did not contain any ARGs. ARGs tend to be more commonly associated with MGEs and less so with the chromosome (Partridge et al., 2018; Yi et al., 2022), but we found all 23 ARGs to be associated with the chromosome, possibly within ICEs (Burrus et al., 2002). Some studies have found MDR MAGs with over 20 different ARGs (Yi et al., 2022), but the norm is closer to 1–6 ARGs, where a single ARG may confer resistance to multiple antimicrobials as is the case for multi-target efflux pumps (Holman et al., 2022; Zhang et al., 2022; Zhao et al., 2020). Given these observations, it is possible that MAG metaSPAdes-
149 is an artifact of assembly and/or binning due to the repetitive elements that are within MGEs. If several MGEs contained numerous ARGs, then it is possible they might have been assembled/binned together into a single MAG.

The number ARGs detected in the quality-filtered sequence read data of SRR6512893 were three orders of magnitude greater than that in MAGs, reflecting a loss in the coverage of ARGs. For this reason, metagenomic genomic surveillance of AMR using short-reads must still rely upon read-based methods as opposed to assembly-based approaches. However, to identify associations of ARGs and MGEs, assembly-based approaches are necessary, even though this approach generates an incomplete picture of relative abundance or the quantity of these elements.

There are numerous ARG databases available, many with their own specific annotation tools. For example, CARD (Comprehensive Antibiotic Resistance Database) uses the Resistance Gene Identifier (Alcock et al., 2023), whereas Staramr (Bharat et al., 2022) uses ResFinder (Florensa et al., 2022). ABRicate (Seemann, 2020) identifies ARGs using a variety of reference sources, including MEGARes 2.0 (Doster et al., 2020) and ResFinder for the resistome. PlasmidFinder for the mobilome, and the Virulence Factors Database (VFDB) for virulence determinants (Liu et al., 2022). The AMR++ pipeline uses the MEGARes database for resistome discovery (Doster et al., 2020), but is unique in that it is a combination of ResFinder, CARD, and NCBI’s Bacterial Antimicrobial Resistance Reference Gene Dataset (Feldgarden et al., 2019). These datasets encompass resistance genes and determinants to conventional antibiotic drugs, biocides, metals, and multi-class resistances. In this study, the MEGARes 2.0 database was selected due to its characteristic of using multiple databases, in addition to being implemented in the AMR++ pipeline. However, all current AMR databases have a human-centric bias to them as they are primarily constructed from commensal or pathogenic bacteria isolated from humans,
with an underrepresentation of ARGs from agricultural and natural environments. A One Health-focused AMR database has yet to be developed, and until one is constructed, future studies will continue to have a human bias.

Detecting MGEs in metagenomes is challenging, as in addition to the absence of a comprehensive MGE database, no single tool encompasses all of algorithms needed to detect all types of MGEs (Pillay et al., 2022). The method employed in this study was to treat MAGs as incomplete genomes and screen them for MGEs using MOB-suite. This method failed to associate ARG with MGE within MAGs. While this may be a consequence of the low recovery of binned MAGs, it may also be due to a failure to detect ARGs. Methods for the detection of MGEs often involve BLAST alignment to a collection of MGE databases (Calderón-Franco et al., 2022; Slizovskiy et al., 2020), including ICEberg 2.0 for ICEs (Liu et al., 2019); PlasmidFinder for plasmids (Carattoli et al., 2014); ACLAME (A CLAssification of Mobile genetic Elements) for phage genomes, plasmids, and transposons (Leplae et al., 2004); IS finder for insertion sequences (Siguier et al., 2006); and INTEGRALL for integrons, integrases, and gene cassettes (Moura et al., 2009). Another method proposed by Pillay et al. (2022) aligns to MGE databases using a mapping tool, such as BWA-MEM (Li & Durbin, 2009) or Bowtie2 (Langmead & Salzberg, 2012). This method of aligning to multiple databases could be more successful in the recovery of MGE in metagenomes as compared to aligning to a single database or using a single tool designed for whole-genome sequenced isolates such as MOB-suite.

Despite these numerous methods, MGEs such as plasmids and genomic islands have been shown to be disproportionally absent as compared to core chromosomal regions in MAGs generated using short-read assembled metagenomes (Maguire et al., 2020). This calls for the use of long-read assembly technologies that can span entire repeat regions that make the
identification of MGEs difficult. Due to the high error rate (10-12%) associated with long-read sequencing technologies (Morisse et al., 2021), it is preferable to couple the greater length of long-read sequencing with the higher accuracy of short-reads through hybrid assemblies (Zhang et al., 2020). Different hybrid assemblers have been outlined and compared, with Unicycler being more accurate and producing more contiguous assemblies (Chen et al., 2020). The authors of this study noted the lack of MGEs, particularly plasmids. Undertaking long-read sequencing and applying hybrid assembly to this data would increase the recovery of MGEs and allow for more accurate co-localization analysis.

In conclusion, targeted assembly methods such as MetaCherchant are not suitable for association studies on taxonomy, ARGs, and MGEs using 100 bp Illumina short-reads. De novo assembly approaches are likely needed for metagenomics when investigating association experimental questions, such as how the use of antimicrobials affects the prevalence of mobile ARGs in a given population as opposed to determining the specific relative abundance of elements. Second, MAG-based analyses into AMR do not provide comprehensive coverage of ARG diversity, relative abundance, or overall prevalence to adequately inform surveillance studies. Sequence read data is still required to investigate these attributes in metagenomic sequenced environmental samples. Increased sequencing depth, sequence read length and/or hybrid assemblies could overcome this limitation. Third, MGEs are too repetitive to be accurately detected in quality filtered sequence data (especially for short-reads) and need longer sequences gained from assembly to establish the surrounding context of other genes. Long-read technologies coupled with hybrid assembly could help circumvent this issue as well, though not to the same coverage as would be possible from read-based relative abundances. There are two possible recommended approaches to associate the resistome and mobilomes: a) use quality
filtered read data (short or long-read) to get relative abundances and MAGs for a more comprehensive accounting of associations; or b) if time is a limiting factor, only investigate contigs for the presence of co-localized ARGs and MGEs. The findings presented here help inform future projects that strive to use metagenomics to characterize AMR in a variety of environments.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding Information

The authors are grateful to the Genomics Research and Development Initiative (GRDI) of the Government of Canada, and the Major Innovation Fund of the Government of Alberta in conjunction with the University of Calgary AMR One Health Consortium for financial support.

Consent for Publication

All authors have read and agree to the published version of the manuscript.

Author Contributions

Study was conceptualized by R.O.P, R.Z., and T.A.M.; Provision, management, and access of bioinformatics tools and computing environment was facilitated by R.O.P. and G.V.D. Metagenomic assemblies were performed by C.L. and R.O.P. Data was analyzed by C.L.
Interpretation of data was performed by C.L., R.O.P., and R.Z. Manuscript draft prepared by C.L. Manuscript reviewed and edited by all authors. Funding acquisition by R.Z., A.Z., and T.A.M. All authors have read and agree to the published version of the manuscript.

References


Florensa, A. F., Kaas, R. S., Clausen, P., Ayta-Aktug, D., & Aarestrup, F. M. (2022). ResFinder - an open online resource for identification of antimicrobial resistance genes in next-


Florensa, A. F., Kaas, R. S., Clausen, P., Ayta-Aktug, D., & Aarestrup, F. M. (2022). ResFinder - an open online resource for identification of antimicrobial resistance genes in next-
generation sequencing data and prediction of phenotypes from genotypes. Microb Genom, 8(1). https://doi.org/10.1099/mgen.0.000748


advanced methodologies and community practices. *Methods*, 102, 3-11.

https://doi.org/10.1016/j.ymeth.2016.02.020


https://doi.org/10.1093/bioinformatics/btp324


https://doi.org/10.1093/nar/gkab1107


https://doi.org/10.1093/mgen.000436


https://doi.org/10.1128/microbiolspec.arba-0009-2017


https://doi.org/10.1371/journal.pone.0061217


https://doi.org/10.1101/gr.213959.116


https://doi.org/10.1093/bioinformatics/btx681


This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4512659


