

Using Metagenomic Approaches to Determine Antimicrobial Resistance Diversity and Prevalence within Four Landfills

by

Isabella Ippolito

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2023

© Isabella Ippolito 2023

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Data contributions

I would like to acknowledge and thank the following contributions to this work. Dr. Aneisha Collins-Fairclough from the University of Technology in Jamaica provided us with the Jamaican metagenomes. Metagenomic analyses associated with the Jamaican metagenomes were carried out by Dr. Collins-Fairclough, Rebecca Co, Dr. Ellis and Dr. Hug. Please see Collins-Fairclough, A. M., Co, R., Ellis, M. C., & Hug, L. A. (2018). Widespread antibiotic, biocide, and metal resistance in microbial communities inhabiting a municipal waste environment and anthropogenically impacted river. *Mosphere*, 3(5), e00346-18.

The Southern Ontario landfill was sampled in 2016 and 2017 collaboratively by Hug lab members. The Jamaican Dump in Kingston, Jamaica was sampled by Dr. Aneisha Collins-Fairclough. The NEUS landfill was sampled by Hug Research Group members Rebecca Co, Nikhil George and Alex Sauk in 2019.

Canadian and US landfill metagenomes samples were binned by Nikhil George. Many public data sets were used in the completion of this work. Metagenomes and MAGs were accessed through the Integrated Microbial Genomes & Microbiomes (IMG/M) database (<https://img.jgi.doe.gov/>) (Markowitz et al., 2012) . These datasets were generated by researchers from around the world. Please see Table 2.3 provided in this thesis to find PI information and more details for each dataset.

Abstract

Antibiotics are one of the many hazardous substances present in municipal solid waste (MSW) landfills. Antibiotics can contaminate the leachate generated in landfills, which poses a threat to human health as leachate can permeate and contaminate the surrounding ground water and surface water systems. The presence of antibiotics in landfills can create selection pressures on the microorganisms present in the landfill, selecting for antibiotic resistance genes (ARGs) and antibiotic resistant bacteria (ARB). This makes landfills important reservoirs of ARGs and ARB. This study used genome resolved metagenomic sequencing to observe ARG diversity and prevalence from four active municipal landfills, their adjacent ground or surface water systems, as well as other environments for comparison. The Resistance Gene Identifier (RGI) and DeepARG were used to predict resistance mechanisms, drug classes, and ARGs. The results from RGI and DeepARG highlighted landfill signatures that were distinct from other environments, and that landfill ARG profiles were more similar to each other than to other environments. PlasClass and Phage and Plasmid Recognizer for Metagenomes (PPR-Meta) were used to predict plasmids within the landfill metagenome datasets. Plasmids encoding ARGs were identified to assess the role of plasmids in ARG mobility in landfills, and to identify enriched ARG types on these mobile elements, such as multidrug resistance and antibiotic inactivation resistance mechanisms. The results of this study clarify target drug classes of interest, such as tetracyclines, aminoglycosides, macrolide-lincosamide-streptogramin (MLS) antibiotics, glycopeptides, and peptide antibiotics, as well as the impact that landfills along with their ground and surface water systems may have on surrounding environments.

Acknowledgements

I would like to express my deepest appreciation to my supervisor Dr. Laura Hug for her encouragement, guidance, patience, enthusiasm, and extensive knowledge. Throughout the duration of this project Laura has continuously provided me with unparalleled support in all of my research and writing of my thesis. I feel extremely lucky to have had Laura as my supervisor for my M.Sc. I would also like to extend my sincere thanks to my committee: Dr. Trevor Charles, and Dr. Brendan McConkey for their valuable advice and suggestions.

I am also grateful to the many members of the Hug lab (past and present) for all of their encouragement and advice. I have learned a lot from everyone and they have made my M.Sc. a very enjoyable experience.

I would like to thank Dr. Aneisha Collins-Fairclough from the University of Technology in Jamaica who provided us with the Riverton dump metagenomes.

Table of Contents

Author's Declaration.....	ii
Statement of Contributions	iii
Abstract	iv
Acknowledgements.....	v
List of Figures	vii
List of Tables.....	ix
List of Abbreviations.....	x
Chapter 1: Introduction and Objectives	1
1.1 Antibiotics and Antibiotic Resistance	1
1.2 Environmental Antibiotic Resistance and Landfills.....	3
1.3 Horizontal Gene Transfer of Antibiotic Resistance Genes and Plasmids	5
1.4 Objectives.....	6
Chapter 2: Antibiotic Resistance Gene (ARG) and Antibiotic Resistant Bacteria (ARB) Diversity and Prevalence	7
2.1 Introduction	7
2.2 Methods.....	8
2.2.1 Landfill Sampling and Metagenome Processing.....	8
2.2.2 Data Processing	9
2.2.3 Comparison to Other Environments.....	10
2.2.4 Antibiotic Resistant Bacteria (ARB) Classification	11
2.3 Results	11
2.3.1 Antibiotic Resistance Gene (ARG) Prevalence in Landfill and Outgroup Metagenomes	11
2.3.2 Resistance Mechanism Distribution and Prevalence Among Landfill and Outgroup Metagenomes.....	23
2.3.3 Target Drug Class Distribution and Prevalence Among Landfill and Outgroup Metagenomes.....	28
2.3.4 Gene Family Distribution and Prevalence Among Landfill and Outgroup Metagenomes.....	34
2.3.5 Classification of Bacteria with Antibiotic Resistance Genes (ARGs) in Landfill Metagenomes.....	41
2.4 Discussion	46

2.4.1 Differences Between the RGI and DeepARG	46
2.4.2 Antibiotic Resistance Gene (ARG) Burden Across Environments	47
2.4.3 Landfill Specific signatures for Resistance Mechanisms, Drug Classes, and Gene Families	51
2.4.4 Antibiotic Resistant Bacteria (ARB) Classification in Landfill Metagenomes.....	62
2.4.5 New Trends Seen in Landfill Metagenomes	64
2.4.6 Conclusions	64
Chapter 3: Assessment of Plasmids and ARG Mobility within Landfill Metagenomes	66
3.1 Introduction	66
3.2 Methods	68
3.2.1 Landfill Sampling	68
3.2.2 Data Processing	68
3.2.3 Determination of Plasmid Load and Antibiotic Resistance Gene (ARG) Mobility	69
3.3 Results	70
3.3.1 Total Plasmid Counts	70
3.3.2 Plasmid Loads	73
3.3.3 Classification of Plasmid-Encoded Antibiotic Resistance Genes (ARGs).....	75
3.4 Discussion	81
3.4.1 Total Plasmid Counts within Landfills	81
3.4.2 Total Antibiotic Resistance Genes (ARGs) encoded on Plasmids and Plasmid Loads	81
3.4.3 Plasmid-Encoded Antibiotic Resistance Genes (ARGs) within Landfills	84
3.4.4 Conclusions	85
Chapter 4: Conclusions and Future Work.....	87
References.....	91
Appendix A: Supplementary Material: Chapter 2	108

List of Figures

Figure 2.1: Heatmap of resistance mechanism relative abundance per 100,00 genes, as predicted by the RGI.....	24
Figure 2.2: Ordination plot of resistance mechanisms using results from the RGI.....	27
Figure 2.3: Heatmap of drug class relative abundance from the RGI	29
Figure 2.4: Heatmap of drug class relative abundance from DeepARG	30
Figure 2.5: Ordination plots for drug class predictions from the RGI (A,B) and DeepARG (C,D)	32
Figure 2.6: Heatmap of gene family relative abundance using results from the RGI	36
Figure 2.7: Heatmap of gene family relative abundance using results from DeepARG	37
Figure 2.8: Ordination plot of gene families using results from the RGI (A,B) and DeepARG (C,D)	39
Figure 2.9: Bar chart of classified MAGs and ARG encoding genes that have been classified using results from the RGI (A) and DeepARG (B)	44
Figure 3.1: Resistance mechanism relative abundance per 100,000 genes using results from the RGI for full metagenome and plasmid-specific predictions	76
Figure 3.2: Proportional bar chart of resistance mechanism relative abundance per 100,000 genes using results from the RGI.....	77
Figure 3.3: Drug class relative abundance per 100,000 genes using results from DeepARG.....	79
Figure 3.4: Proportional bar chart of drug class relative abundance per 100,000 genes using results from DeepARG	80

List of Tables

Table 2.1: Landfill metagenome statistics	13
Table 2.2: Landfill metagenome statistics by site.....	16
Table 2.3: Outgroup environment metagenome statistics.....	17
Table 2.4: Landfill and outgroup environment metagenome statistics by sample type.....	21
Table 2.5: Average normalized abundances of ARGs per 100,00 genes and ranges by landfill and outgroup environment sample type.....	22
Table 3.1: Predicted plasmids and phages by sample type.....	71
Table 3.2: Predicted plasmids and phages by site.....	72
Table 3.3: Plasmid-encoded ARGs and plasmid load by sample type	74
Table 3.4: Plasmid-encoded ARGs and plasmid load by site	75
A.1 Supplementary Data 2.1: Percent variance explained by the first ten axes for landfill metagenome – based ordinations	108
A.2 Supplementary Data 2.2: Percent variance explained by the first ten axes for landfill and outgroup environment based ordinations.....	108

List of Abbreviations

AG	Agriculture
ARB	Antibiotic Resistant Bacteria
ARDB	Antibiotic Resistance Genes Database
ARG	Antibiotic Resistance Gene
BLAST	Basic Local Alignment Search Tool
CA	California
CARD	Comprehensive Antibiotic Resistance Database
CLC	Composite Leachate Cistern
DW	Drinking Water
GI	Genomic Island
GW	Groundwater
HGT	Horizontal Gene Transfer
IMG/M	Integrated Microbial Genomes & Microbiomes
JA	Jamaica
LW	Leachate Well
M	Mining
MAG	Metagenome Assembled Genome
MFS	Major Facilitator Superfamily
MGE	Mobile Genetic Element
MLS	Macrolide; Lincosamide; Streptogramin
NEUS	Northeastern United States
NLLA	Non-Leachate Landfill Associated
ORF	Open Reading Frame
PLDSB	Plasmid Database
PPR-Meta	Phage and Plasmid Recognizer for Metagenomes
RGI	Resistance Gene Identifier
RND	Resistance-Nodulation-Cell Division
S	Sewage
SO	Southern Ontario
UNIPROT	Universal Protein Resource
WHO	World Health Organization
WWTP	Wastewater Treatment Plant

Chapter 1: Introduction and Objectives

1.1 Antibiotics and Antibiotic Resistance

The World Health Organization (WHO) recognizes antimicrobial resistance as a major threat to public health (Public Health Agency of Canada, 2022). Antibiotics are drugs that kill or inhibit the growth of bacteria, specifically, where antimicrobials are a class of compounds affecting a broader group of microbes (e.g., including fungi and viruses). Antibiotics play a critical role in modern medicine, as they have lowered disease burden, reduced childhood mortality and increased life expectancy (Blair et al., 2015). Antibiotics can have either broad spectrum activity, meaning that they target both Gram-positive and Gram-negative bacteria, or narrow spectrum activity, meaning that they target either Gram-positive or Gram-negative bacteria (Singh et al., 2017). Through evolution, bacteria can develop resistance mechanisms against antibiotics, meaning that antibiotic effectiveness decreases over time (Singh et al., 2017). The overuse and misuse of antibiotics increases selection pressure for the evolution of resistance mechanisms (Huemer et al., 2020; Singh et al., 2017). There are limited biochemical modifications that can rescue the effectiveness of current antibiotics. As a result, the creation of new antibiotics has not kept pace with the loss of effective antibiotics, posing an increasing threat to health care systems (Huemer et al., 2020).

Bacteria can develop resistance to antibiotics through different mechanisms, including target modification, antibiotic inactivation, decreased antibiotic entry, and expression of efflux pumps (Blair et al., 2015; Huemer et al., 2020; Singh et al., 2017). In target modification, the antibiotic target molecule changes structure through spontaneous mutations or post-translational modifications, rendering the antibiotic ineffective as it is unable to bind to the target. In antibiotic inactivation, antibiotics are inactivated by bacteria through direct destruction by hydrolysis or through the transfer of a chemical group that prevents the antibiotic from binding

to the target molecule. Bacteria can decrease antibiotic entry by reducing the number of porins or by mutating porins in their cell membrane. Finally, bacteria can also overexpress antibiotic-targeting efflux pumps in the membrane. Notably, both decreasing entry and increasing the efflux of antibiotics lead to decreased intracellular concentrations of antibiotic within the bacterial cell (Blair et al., 2015; Huemer et al., 2020).

Bacteria that have evolved resistance to certain antibiotic classes are called antibiotic resistant bacteria (ARB). Bacteria can have intrinsic resistance or acquired resistance. Intrinsic resistance refers to resistance to an antibiotic due to the inherent structural or functional characteristics of the bacteria (Blair et al., 2015; Jian et al., 2021). Some examples of intrinsic resistance include the absence of the antibiotic's target molecule or protein or the inability of some antibiotic compounds to cross that organism's membrane due to an absence of appropriate transporters (Blair et al., 2015). Resistance can also be in the form of antibiotic resistance genes (ARGs), providing the bacteria that carries the ARG on its genome with resistance (Jian et al., 2021). In acquired resistance, resistance mechanisms are gained through horizontal gene transfer of ARGs (Blair et al., 2015; Davies & Davies, 2010; Jian et al., 2021; San Millan, 2018). ARGs therefore support the proliferation and development of ARB, either through intrinsic resistance or acquired resistance.

Mitigating the rapid rise of antibiotic resistance requires an approach that considers connections between human health, animal health, and the environment, otherwise known as the “one-health perspective”. The “one-health perspective” is necessary because ARB and ARGs can impact any of these three axes, with implications for the other two (Larsson et al., 2018). ARB and ARGs are not solely a public health issue as many ARB and ARGs originate from the environment. The repertoire of ARGs present in the environment is, in turn, impacted by human

activities, which can influence the resistance genes that are acquired by clinically relevant pathogens. ARG abundance has increased since World War II because of the industrialization of antibiotic production. This led to an increase in antibiotic residues entering the environment through pharmaceutical disposal, wastewater effluents, manure, crop treatment, and aquaculture (Finley et al., 2013). Ultimately, to address the issue of antibiotic resistance, the contribution of ARGs and ARB from the environment needs to be considered.

1.2 Environmental Antibiotic Resistance and Landfills

Studying environmental antibiotic resistance is important as it can address critical knowledge gaps such as how human-impacted environments are contributing to antibiotic resistance prevalence and movement (Larsson et al., 2018). Environments act as reservoirs of ARB and ARGs (Ben et al., 2019), but the differences in ARG reservoirs between environment types and the relationship to human impacts has not been examined.

Landfills are engineered environments that act as disposal sites for municipal solid wastes, which can contain a variety of materials such as mechanical waste, detergents, pharmaceuticals, general household wastes, and industrial by-products (Lakhout & Alsulami, 2020; Ozbay et al., 2021). Landfills are designed to keep waste, leachate, and gas separate from the surrounding environment (Hilger & Barlaz, 2007). Leachate formation is one of the most common problems associated with landfill sequestration (Ozbay et al., 2021). Leachate is the liquid that develops as rainwater and water from organic decomposition moves through the waste in the landfill. The waste present in the landfill contaminates leachate as it develops (Bandala et al., 2021; Ozbay et al., 2021). Landfills contain drainage layers to move the contaminated leachate to a leachate collection system at the bottom of the landfill (Hilger & Barlaz, 2007; Ozbay et al., 2021). Waste is deposited above this collection system, with the waste covered daily to prevent windblown

refuse, contamination from storm water runoff, and attraction of disease vectors like rats and other animals. The final cover of low permeability soil is applied once waste has reached the maximum elevation and this cover minimizes storm water infiltration and supports vegetative growth (Hilger & Barlaz, 2007). Despite measures taken to contain leachate within the landfill, leachate can percolate into and contaminate the surrounding groundwater and surface water systems (Bandala et al., 2021; Ozbay et al., 2021; Parvin & Tareq, 2021). The eventual contamination of surrounding groundwater and surface water systems by leachate poses risks to human health.

In recent years pharmaceuticals such as antibiotics have been identified as emerging contaminants within landfills (Bandala et al., 2021). Antibiotics can end up in landfills through the use of veterinary drugs, household waste, illegal clinical waste, and activated sludge (Bai et al., 2022; Y. Wang et al., 2015). As leachate develops in landfills it can become contaminated by the antibiotics that are present (Bai et al., 2022). Development of novel ARGs and ARB can occur within landfills as ARGs evolve during refuse decomposition under strong selection pressures (Liu et al., 2018). ARGs and ARB present within landfills can enter the surrounding environment as leachate seeps out of the landfill, polluting the surrounding ground and surface water systems (Bai et al., 2022; Liu et al., 2018; Y. Wang et al., 2015). ARGs and ARB are seen as emerging contaminants as they are increasing in prevalence and threatening human health through environmental to human transmission (Ben et al., 2019). Some studies have profiled ARGs and ARB in landfills using PCR (Y. Wang et al., 2015, 2022; S. Xu et al., 2021), though these studies are limited by the diversity and breadth of ARGs targeted by the primers that are available (Zhang et al., 2022). One study on a landfill in Chongqing, China investigated ARG and ARB abundance and diversity using metagenomics based on Illumina high throughput

sequencing. The authors of this study focused solely on eight antibiotics and their resistance mechanisms, rather than reporting the total ARG diversity present. Main findings included the abundance of multidrug resistance genes increasing during refuse decomposition, where antibiotic resistance was influenced by the abundance of *Pseudomonas*, as well as bacterial community diversity (Liu et al., 2018). Two studies on landfill leachate from China profiled ARGs with qPCR, but were also limited to the primers the authors chose to study (Y. Wang et al., 2015, 2022). Another study done on a landfill in China using high throughput PCR was able to detect a larger diversity of ARGs as they used a total of 296 primers, however their detection of ARGs was limited to those targeted by the selected primers (S. Xu et al., 2021). Full shotgun sequencing for ARG detection from landfills has not previously been described.

1.3 Horizontal Gene Transfer of Antibiotic Resistance Genes and Plasmids

Antibiotic resistance has become increasingly widespread across bacterial lineages, in large part due to the acquisition of ARGs via horizontal gene transfer (HGT). Dissemination of ARGs occurs via mobile genetic elements (MGEs). MGEs allow for intracellular and intercellular DNA mobility (Partridge et al., 2018). MGEs are acquired by HGT (Davies & Davies, 2010; San Millan, 2018) via transformation, transduction, or conjugation (Brown-Jaque et al., 2015; Davies & Davies, 2010; Li & Wang, 2021). Plasmids, phages, and genomic islands (GIs) are all vectors for HGT (Smillie et al., 2010), and are all capable of carrying ARGs and integrating new ARGs into a host's genomic repertoire (Brown-Jaque et al., 2015; Li & Wang, 2021; San Millan, 2018; Smillie et al., 2010).

Plasmids are widespread throughout the bacterial Domain and are important for bacterial evolution (San Millan & MacLean, 2017) as they are one of the most important drivers of HGT (Davies & Davies, 2010). Plasmids are circular pieces of extrachromosomal DNA that replicate

independently from the chromosome (San Millan, 2018). ARGs are carried on plasmids, making plasmids important routes for acquired antibiotic resistance (Partridge et al., 2018; Smillie et al., 2010). Conjugation, where the transfer of genetic material between bacteria requires cell-to-cell contact, is the most important form of HGT when considering ARG transfer, as there is the potential for some plasmids to conjugate between distantly related organisms (San Millan, 2018; Smillie et al., 2010). In stressful conditions, plasmids provide novel genetic information, which can allow rapid adaptation, leading to bacterial evolution (Smillie et al., 2010). ARGs present on plasmids provide bacteria with a selective advantage in environments with high amounts of antibiotics (Eberhard, 1990).

1.4 Objectives

The purpose of this research was to determine the prevalence and potential for mobility of ARGs and the diversity of ARB in landfills, as key environments selecting for ARG evolution and ARB development. Chapter 2 explores the ARG diversity and prevalence within four landfills. To assess if landfills possess specific ARG signatures, results from landfills were compared to other contaminated and pristine environments. This chapter hypothesizes that landfills will have distinct ARG signatures that are enriched in diverse ARG types compared to other environments. The work presented in Chapter 3 includes reconstruction of plasmids from metagenomic datasets to clarify the mobility of ARGs in the landfill microbial communities. This chapter hypothesizes that ARG diversity within the four landfills would correlate to ARG mobility, indicating strong selection pressures for antibiotic resistance at a given site. Furthermore, the role of plasmids as an important mobile reservoir of ARGs in landfills was explored.

Chapter 2: Antibiotic Resistant Gene (ARG) and Antibiotic Resistant Bacteria (ARB) Diversity and Prevalence

2.1 Introduction

Antibiotic resistant bacteria (ARB) develop when bacteria acquire genes that code for antibiotic resistance, also known as antibiotic resistance genes (ARGs) (Jian et al., 2021). There are different tools and pipelines available for identifying putative ARGs, each of which has distinct mechanisms of searching public or proprietary databases of known ARGs. The Resistance Gene Identifier (RGI) is a bioinformatic tool associated with the Comprehensive Antibiotic Resistance Database (CARD) that analyzes submitted sequences to predict resistance mechanism, targeted drug classes, and ARGs (McArthur et al., 2013). Annotation using the RGI follows a perfect/strict/loose paradigm. A perfect hit is an exact match to a known ARG sequence in the CARD. A strict hit has some variation from a sequence in the CARD and is useful for determining potential mutations. A loose hit has lower similarity to a sequence in the CARD and is useful for determining new potential ARGs and novel antimicrobial threats (Alcock et al., 2020). The RGI uses a best hit approach, which has been seen to have a high false negative rate, meaning that some ARGs are not predicted as ARGs (Arango-Argoty et al., 2018). A second tool for identifying putative ARGs is DeepARG, a deep learning approach that predicts ARGs and drug classes based on the DeepARG-DB database. This database is manually updated and comprised of information from the CARD, the Antibiotic Resistance Genes Database (ARDB), and the Universal Protein Resource (UNIPROT) database (Arango-Argoty et al., 2018). DeepARG has two models: DeepARG-SS and DeepARG-LS, which use either short reads or full-length gene sequences, respectively. DeepARG considers the similarity distribution instead of only relying on the best hit approach. Additionally, DeepARG does not rely on human intervention for feature discrimination. However, DeepARG can only identify genes predicted to

belong to the 30 ARG categories it was trained on and cannot *de novo* predict new categories. The quality of the results from DeepARG rely on the quality of the original database (Arango-Argoty et al., 2018).

In this chapter, I make use of existing metagenomic datasets from four geographically distant landfills to assess the prevalence and diversity of ARGs and their encoding ARBs within landfills using The RGI and DeepARG. I compare the landfill ARG profiles with metagenomes from four freshwater samples associated with agriculture samples, four drinking water samples, four groundwater samples, two mining waste water samples, two sewage samples, and fourteen wastewater treatment plant samples to assess whether landfills carry a unique ARG burden and whether landfills represent a specific contribution to ARG profiles in the environment.

2.2 Methods

2.2.1 Landfill Sampling and Metagenome Processing

Previous members of the Hug Research Group sampled four landfills located in Southern Ontario (SO), Southern California (CA), the northeastern coast of the US (NEUS), and Jamaica (JA) (site anonymity by site owners' requests). Total community DNA was extracted, sequenced, assembled and binned into high quality MAGs for assessments of total microbial diversity (Collins-Fairclough et al., 2018; Sauk & Hug, 2022), virus-host interactions (George & Hug, 2023), and determination of methane cycling population dynamics across time (Grégoire et al., 2023). No prior investigation has explored ARG annotation or distribution within these datasets, and the CA landfill samples have not previously been reported in the literature.

In brief, landfill biomass was collected from leachate wells, leachate ponds, leachate collection systems, leachate treatment plants, and adjacent water systems including groundwater, stormwater catchment systems, and a river abutting the JA landfill. Biomass was harvested either through centrifugation (JA), filtration onto a 0.2 μm sterivex filter (CA, NEUS), or onto 0.1 μm

polyethersulfone membrane filters (SO) (Collins-Fairclough et al., 2018; George & Hug, 2023; Grégoire et al., 2023; Sauk & Hug, 2022). A full list of sampling locations and their respective metagenome sequences is detailed in the Results and in Table 2.1.

For the CA landfill, four samples were collected including from one leachate well (CA_LW1), one from the collected leachate at the treatment plant influent (CA_TPin), and two samples from the treatment plant bioreactor (CA_Filter & CA_Sludge). For all four samples, DNA was extracted with a Qiagen PowerSoil kit, with one modification, where sterivex filters were sliced into small fragments as input in place of soil. CA metagenomes were sequenced by the Center for Applied Genomics (TCAG, Toronto, Ontario) on an Illumina HiSeq machine using 2x150 bp paired-end libraries.

Metagenome processing was conducted by Nikhil George and Laura Hug. Metagenomes were assembled using spades3 using the -meta flag, and reads mapped to the assembled scaffolds using bowtie (Bankevich et al., 2012; Langmead & Salzberg, 2012). Scaffolds longer than 2.5 kb were binned using a combination of CONCOCT, MaxBin2, and MetaBAT2, with consensus bins generated by DAS Tool (Alneberg et al., 2014; Kang et al., 2019; Sieber et al., 2018; Wu et al., 2015). Metagenome assembled genomes (MAGs) were kept for further analysis if they were above 70% complete and less than 10% contaminated according to CheckM (Parks et al., 2015). Taxonomic affiliations for MAGs were determined using GTDB-tk (Chaumeil et al., 2019).

2.2.2 Data Processing

To ensure all metagenomes had been processed consistently for each dataset, pullseq was used to extract all sequences from metagenome fasta files that were greater than 1000bp (<https://github.com/bcthomas/pullseq>). Prodigal was used to annotate predicted genes from the 1000 bp+ scaffolds, generating nucleotide and amino acid files of open reading frames (Hyatt et

al., 2010). Amino acid files generated from Prodigal were used as input to the Resistance Gene Identifier (RGI) (McArthur et al., 2013). The “rgi main” command was used with “-t” set to protein, “-n” set to 8, and the “—clean” flag included. Nucleotide files generated from prodigal were used as input into DeepARG with the “—model” flag set to LS, and the “—type” flag set to nucl.

For each metagenome, outputs from The RGI and DeepARG were summarized using pivot tables in Excel. Counts of open reading frames (ORFs) for predicted ARGs classified to resistance mechanism (RGI only), drug class, and gene family were collated and compared. Comparison tables were normalized to # predicted ARGs per 100,000 ORFs per metagenome to address differences in metagenome sizes.

Heatmaps were generated from the normalized data in R. To better visualize changes in relative abundance the data was log-transformed, with zero values converted to -2 to allow inclusion in heatmaps. Using the ComplexHeatmap package in R the columns (metagenomes) were clustered by row data (normalized ARG count) and rows arranged in order of decreasing relative abundance.

Ordinations were generated from the normalized data in R. A distance matrix using the Euclidian distance was created. Using the ggplot2 package in R an ordination plot was created using the geom_point function. The points were coloured by sample type.

2.2.3 Comparison to Other Environments

Environments selected to compare with landfills were: freshwater samples associated with agriculture (AG), drinking water samples (DW), groundwater samples (GW), mining wastewater samples (M), sewage wastewater samples (S), and wastewater treatment plant samples (WWTP). Metagenomes for these environments were downloaded from the Joint Genome Institute’s

Integrated Microbial Genomes (IMG) database. From IMG, the environments' taxon object IDs were used to find representative metagenomes with a minimum size of 2271 kb assembled data (average assembled size: 227.0 Mbp per sample, compared to average size of 1578.8 Mbp for landfill datasets). A full list of outgroup sample types and their metagenome sequence metadata is detailed in the Results and in Table 2.3. The outgroup metagenomes' assembled scaffold files were downloaded via globus. All outgroup metagenomes were analyzed in the same manner as described above, from initial scaffold curation using pullseq onwards.

Normalized ARG counts per 100,000 genes for individual metagenomes are presented in Supplementary file 1, which is available as an excel spreadsheet hosted by the Open Science Framework at https://osf.io/myg8q/?view_only=4a44d89e2044411dba79761a3f3261a5.

2.2.4 Antibiotic Resistant Bacteria (ARB) Classification

Scaffolds containing ARGs identified by the RGI and DeepARG were compared to quality filtered bins to identify the ARG-encoding MAGs using grep. GTDB classification files were referenced to determine taxonomic classifications for ARB carrying ARGs.

2.3 Results

2.3.1 Antibiotic Resistance Gene (ARG) Prevalence in Landfill and Outgroup Metagenomes

A total of 29 landfill metagenomes were used in this study from four sites: Southern Ontario (SO), Southern California (CA), a landfill on the northeastern coast of the USA (NEUS), and a landfill in Jamaica (JA) (Table 2.1). The SO landfill houses residential and commercial waste and was sampled in 2016 and 2017 for a total of 14 metagenomes. Samples for the SO landfill come from leachate wells (2016: LW1 through LW3; 2017: LW1 through LW4), a composite leachate cistern (July 14, 2016: CLC_T1; July 20, 2016: CLC_T2; October 2017: CLC_T3), and

groundwater or stormwater catchments (2016: GW1; 2017: GW1, GW3, and SWC). The CA landfill houses high levels of industrial waste. It was sampled in June 2019, and contributes four metagenomes: two from the bioreactor within the leachate treatment plant (CA_Filter and CA_Sludge), one metagenome from the influent leachate into the treatment plant (CA_TPin), and one from a leachate well (CA_LW1). There are nine metagenomes from the NEUS landfill which houses residential solid waste, municipal solid waste, construction debris, asbestos, white goods, tires, and household yard waste. Of those, eight metagenomes are from leachate wells (March 2019: NEUS_A, B, C, D1, D2, E, F1, F2), and one metagenome is from a composite leachate cistern (March 2019: NEUS_CSWMC). The JA landfill houses municipal solid waste and provides two metagenomes to the analysis, one from a leachate pond sample (2015: JA_Leachate) and the other from a surface water sample of an adjacent river (2015: JA_River). For comparisons, we have grouped leachate wells (LW: SO_2016 LW1 through LW3; SO_2017 LW1 through LW4; CA_LW1; JA_Leachate; NEUS_A, B, C, D1, D2, E, F1, F2), composite leachate cisterns (CLC: SO_2016 CLC T1, CLC T2; SO_2017 CLC T3; CA_TPin; NEUS_CSWMC), treatment plants (TP: CA_Filter; CA_Sludge), and non-leachate landfill-associated samples (NLLA: SO_2016 GW1; SO_2017 GW1, GW3, SWC; JA_River).

Table 2.1: Landfill metagenome statistics. The number of genes detected as ARGs were compiled from the RGI and DeepARG. The number of genes was obtained from the fasta files created from prodigal. The size of the metagenomes was determined using stats.sh on the min1000 bp fasta files. The percentage of genes that are ARGs were determined by dividing the number of ARGs by the total number of genes and multiplying by 100. Landfill locations are Southern California (CA), Jamaica (JA), North Eastern Coast of the US (NEUS), Southern Ontario (SO). Sample types are leachate wells (LW), composite leachate cisterns (CLC), leachate treatment plant influent (TPIn) and bioreactors (Filter, Sludge), groundwater wells (GW), and storm water catchments (SWC).

Landfill sample	Predicted ARGs (The RGI)	Predicted ARGs (DeepARG)	Number of genes	Size (Mb)	% ARG-encoding genes (The RGI)	% ARG-encoding genes (DeepARG)
JA Leachate	181	70	284,945	216.1	0.064	0.025
JA River	98	75	171,281	132.1	0.057	0.044
SO 2016 CLC T1	610	394	1,245,132	982.8	0.049	0.032
SO 2016 CLC T2	756	415	1,464,855	1,153.5	0.052	0.028
SO 2016 GW1	680	286	1,237,322	1,038.4	0.055	0.023
SO 2016 LW1	834	162	1,537,267	1,056.2	0.054	0.011
SO 2016 LW2	822	781	1,473,271	1,209.9	0.056	0.053
SO 2016 LW3	835	428	1,487,888	1,089.5	0.056	0.029
SO 2017 CLC T3	588	408	1,123,044	936.1	0.052	0.036
SO 2017 GW1	1435	955	2,756,479	2,295.9	0.052	0.035
SO 2017 GW3	143	60	314,296	227.1	0.045	0.019
SO 2017 LW1	517	103	1,000,071	685.2	0.052	0.010
SO 2017 LW2	660	751	1,241,054	1,057.5	0.053	0.061
SO 2017 LW3	1448	384	2,954,114	2,086.3	0.049	0.013
SO 2017 LW4	654	276	1,425,800	1,137.2	0.046	0.019
SO 2017 SWC	251	245	486,641	423.1	0.052	0.050
CA LW1	868	1185	1,566,695	1,325.5	0.055	0.076
CA Filter	1306	1246	2,689,267	2,399.	0.049	0.046
CA Sludge	1171	1131	2,471,200	2,204.1	0.047	0.046
CA TPIn	1637	1595	3,396,366	2,275.4	0.048	0.047
NEUS A	1631	832	3,080,447	2,187.5	0.053	0.027
NEUS B	1743	710	3,464,221	2,563.3	0.050	0.020
NEUS C	1443	1239	2,712,470	2,133.3	0.053	0.046
NEUS CSWMC	1450	1456	2,880,015	2,326.1	0.050	0.051
NEUS D1	1962	778	4,243,122	3,165.3	0.046	0.018
NEUS D2	1661	1665	3,368,931	2,733.0	0.049	0.049
NEUS E	1249	946	2,431,772	1,945.8	0.051	0.039
NEUS F1	1535	1148	3,097,280	2,509.3	0.050	0.037
NEUS F2	1411	1408	2,829,227	2,288.6	0.050	0.050

NEUS landfill had the largest amount of ARGs identified, with all NEUS metagenomes combined, with 14,085 ARGs identified from the RGI and 10,182 identified from DeepARG

(Table 2.2). The JA landfill had the least amount of ARGs with all JA metagenomes combined with only 279 ARGs identified from the RGI and 145 identified from DeepARG (Table 2.2). The SO 2016, SO 2017, and CA landfills had 4,537, 2,466, 5,696 ARGs identified from the RGI and 3,182, 4,982, and 5,157 ARGs identified by DeepARG, respectively. The NEUS D1 metagenome had the highest overall count of ARGs from the RGI with 1,962 ARGs identified. The NEUS D2 metagenome had the highest overall count of ARGs from DeepARG with 1,665 ARGs identified (Table 2.1). The metagenome with the fewest ARGs identified by the RGI was the JA River metagenome with 98 ARGs, where the JA leachate metagenome had the fewest identified by DeepARG with 70 ARGs. However, using the RGI results the CA landfill had the lowest proportion of ARG-encoding genes with 0.049% of genes identified as ARGs, the SO 2016; SO 2017; and NEUS metagenomes had similar percentages with 0.054%, 0.050%, and 0.050% respectively (Table 2.2). From the RGI the JA metagenomes had the highest proportion of ARG encoding genes with 0.061% of genes identified as ARGs. Comparatively, from the DeepARG results the CA landfill had the highest proportion of genes identified as ARGs with 0.051% and the SO 2017 metagenomes had the lowest proportion of genes identified as ARGs with 0.028%. The SO 2016 metagenomes were close to the SO 2017 metagenomes with 0.029% of genes identified as ARG encoding. The JA and NEUS metagenomes were in between with 0.032% and 0.036% respectively. Using the RGI results, the metagenome with the highest proportion of genes that were ARG encoding was the JA Leachate metagenome with 0.064% of genes encoding ARGs, and the metagenome with the lowest proportion of genes that were ARG encoding was the SO 2017 GW3 metagenome with 0.045% of genes encoding (Table 2.1). Using results from DeepARG the CA LW1 metagenome had the highest proportion of genes that were

ARGs with 0.076%, and the SO 2016 LW1 metagenome had the lowest proportion of genes that were ARGs with 0.011%.

Table 2.2: Landfill metagenome statistics by site. The number of genes detected as ARGs were compiled from the RGI and DeepARG. The number of genes was obtained from the fasta files created from prodigal. The size of the metagenomes was determined using stats.sh on the min1000 bp fasta files. The percentage of genes that are ARGs were determined by dividing the number of ARGs by the total number of genes and multiplying by 100. Landfill locations are Southern California (CA), Jamaica (JA), North Eastern coast of the US (NEUS), Southern Ontario (SO).

Landfill	Predicted ARGs (the RGI)	Predicted ARGs (DeepARG)	Number of Genes	% ARG- encoding genes (The RGI)	% ARG- encoding genes (DeepARG)
JA	279	145	456,226	0.061	0.032
SO 2016	4,537	2,466	8,445,725	0.054	0.029
SO 2017	5,696	3,182	11,301,499	0.050	0.028
CA	4,982	5,157	10,123,528	0.049	0.051
NEUS	14,085	10,182	28,107,485	0.050	0.036

Table 2.3: Outgroup environment metagenome statistics. The number of genes detected as ARGs were compiled from the RGI and DeepARG. The number of genes was obtained from the fasta files created from prodigal. The size of the metagenomes was determined using stats.sh on the min1000 bp fasta files. The percentage of genes that are ARGs were determined by dividing the number of ARGs by the total number of genes and multiplying by 100. Outgroup sample types are abbreviated as follows: agriculture (AG), drinking water (DW), groundwater (GW), mining (M), sewage (S), and wastewater treatment plants (WWTP).

Outgroup sample ID	JGI Accession Number or Citation	Number of ARGs identified by the RGI	Number of ARGs identified by DeepARG	Number of Genes	Size of Metagenome (Mb)	% ARG-encoding genes (The RGI)	% ARG-encoding genes (DeepARG)
AG1	104888	49	14	79,785	58.8	0.061	0.018
AG2	105220	54	6	80,296	54.8	0.067	0.0075
AG3	105224	52	3	97,725	68.8	0.053	0.0031
AG4	132480	31	12	67,996	50.4	0.046	0.018
DW1	34882	423	619	840,418	784.4	0.050	0.074
DW2	77582	2	5	2,028	2.3	0.099	0.25
DW3	184124	152	49	339,844	270.4	0.045	0.014
DW4	198618	27	5	46,348	30.0	0.058	0.011
GW1	84326	349	120	676,477	517.0	0.052	0.018
GW2	123537	339	99	686,808	586.9	0.049	0.014
GW3	184128	46	42	83,516	63.5	0.055	0.050
GW4	187715	13	34	13,137	13.0	0.099	0.26
M1	10507	6	5	9,974	6.8	0.060	0.050
M2	Ga0394085	44	73	76,266	64.3	0.058	0.10
S1	192479	339	305	707,806	642.6	0.048	0.043
S2	Ga0168608	9	13	4,927	3.7	0.18	0.26
WWTP 1	(Spasov et al., 2020)	164	150	280,145	248.5	0.059	0.054
WWTP 2	(Spasov et al., 2020)	178	149	376,047	329.1	0.047	0.040
WWTP 3	(Spasov et al., 2020)	131	158	293,527	260.1	0.045	0.054
WWTP 4	(Spasov et al., 2020)	126	106	241,186	216.5	0.052	0.044

Outgroup sample ID	JGI Accession Number or Citation	Number of ARGs identified by the RGI	Number of ARGs identified by DeepARG	Number of Genes	Size of Metagenome (Mb)	% ARG-encoding genes (The RGI)	% ARG-encoding genes (DeepARG)
WWTP 5	(Spasov et al., 2020)	168	102	343,408	300.6	0.049	0.030
WWTP 6	(Spasov et al., 2020)	174	170	378,490	333.1	0.046	0.045
WWTP 7	(Spasov et al., 2020)	144	139	275,825	249.1	0.052	0.050
WWTP 8	(Spasov et al., 2020)	74	68	161,278	138.5	0.046	0.042
WWTP 9	(Spasov et al., 2020)	177	184	357,525	321.7	0.050	0.051
WWTP 10	(Spasov et al., 2020)	75	71	145,119	125.1	0.052	0.049
WWTP 11	(Spasov et al., 2020)	171	164	321,362	287.4	0.053	0.051
WWTP 12	(Spasov et al., 2020)	133	122	241,191	215.8	0.055	0.051
WWTP 13	(Spasov et al., 2020)	174	158	338,918	299.1	0.051	0.047
WWTP 14	(Spasov et al., 2020)	146	143	300,370	269.4	0.049	0.048

A total of 30 outgroup environment metagenomes were used (Table 2.3). In the outgroup environment metagenomes, the sample type with the largest number of ARGs identified were the WWTP samples, with a total of 2,035 and 1,884 ARGs identified by the RGI and DeepARG, respectively (Table 2.4). The M samples had the lowest count of ARGs identified, with 50 and 78 ARGs identified by the RGI and DeepARG, respectively. The single outgroup environment sample with the highest count of ARGs was the DW1 metagenome, with 423 and 619 ARGs identified by the RGI and DeepARG, respectively (Table 2.3). Conversely, the metagenome with the lowest count of ARGs was DW2 with only 2 and 5 ARGs identified by the RGI and DeepARG, respectively. Using the RGI results the outgroup sample type with the highest percentage of genes that were ARG encoding was AG sample type with 0.057% of genes encoding ARGs (Table 2.4). The outgroup S and DW sample types had the lowest percentage of genes that were ARGs as they both had 0.049% of genes encoding ARGs. Comparatively, using DeepARG results the outgroup sample type with the highest percentage of genes encoding ARGs was the DW sample type with 0.055%, and the lowest percentage of genes encoding ARGs was the AG sample type with 0.011%. From the RGI results, the metagenome with the highest percentage of genes encoding ARGs was the S2 metagenome with 0.18% of genes encoding ARGs (Table 2.3). The metagenomes with the lowest percentage of genes encoding ARGs from the RGI were the DW4 metagenome and WWTP 3 metagenome with 0.045%. Using DeepARG results the metagenomes with the highest percentage of genes encoding ARGs were the GW4 and S2 metagenomes with 0.26%. Similar to the RGI, the metagenome with the lowest percentage of genes encoding ARGs using DeepARG results was the DW4 metagenome.

Landfill sample types did not have the highest average normalized abundance of ARGs, with 52.21, 50.30, and 52.26 ARGs per 100,000 genes for LW, CLC, and NLLA samples types

from the RGI, respectively, and 34.26, 32.22, and 34.20 ARGs per 100,000 genes for LW, CLC, and NLLA from DeepARG, respectively (see Table 2.5 for ranges). These values are lower than those for S, DW, GW, and M environments, and similar to AG and WWTP samples. The largest average normalized abundance belonged to the S metagenomes with an average normalized abundance of 115.28 ARGs per 100,000 genes from the RGI and 153.47 ARGs per 100,000 genes from DeepARG (Table 2.5). The range for the S metagenomes were 47.89 to 182.67 ARGs per 100,000 genes and 43.09 to 263.85 ARGs per 100,000 genes from the RGI and DeepARG respectively. The smallest average normalized abundance belonged to the TP metagenomes using the RGI results and the AG metagenomes using DeepARG results with average normalized abundances of 47.97 ARGs per 100,000 genes and 11.43 ARGs per 100,000 genes respectively. The range for the TP metagenomes was 47.39 to 48.56 ARGs per 100,000 genes and the range for the AG metagenomes was 3.07 to 17.65 ARGs per 100,000 genes. Using the RGI results the remaining average normalized abundances for AG, LW, CLC, WWTP, DW, GW, NLLA, M, and TP ranged from 50.30 to 63.75 ARGs per 100,000 genes. From the RGI the LW, CLC, NLLA, and M metagenomes had smaller ranges compared to the AG, WWTP, DW, and GW metagenomes. From the DeepARG results the LW, CLC, WWTP, DW, GW, M, and TP metagenomes had large ranges.

These numbers are difficult to compare given the wide range of metagenome sizes and differences in assemblies between the samples considered (Tables 2.1, 2.3), so normalized abundances were calculated based on ARG frequency per 100,000 genes. All further analyses rely on these normalized values.

Table 2.4: Landfill and outgroup environment metagenome statistics by sample type. The number of genes detected as ARGs were compiled from the RGI and DeepARG. The number of genes was obtained from the fasta files created from prodigal. The size of the metagenomes was determined using stats.sh on the min1000 bp fasta files. The percentage of genes that are ARGs were determined by dividing the number of ARGs by the total number of genes and multiplying by 100. Sample types are abbreviated as follows: agriculture (AG), leachate well (LW), composite leachate cistern (CLC), wastewater treatment plant (WWTP), sewage (S), drinking water (DW), groundwater (GW), mining (M), leachate treatment plant (TP), and non-leachate landfill-associated samples (NLLA).

Sample Type	Number of ARGs identified by the RGI	Number of ARGs identified by DeepARG	Number of Genes	% ARG- encoding genes (the RGI)	% ARG- encoding genes (DeepARG)
AG	186	35	325,802	0.057	0.011
LW	19,454	12,866	38,198,575	0.051	0.034
CLC	5,041	4,268	10,109,412	0.050	0.042
WWTP	2,035	1,884	4,054,391	0.050	0.046
S	348	318	712,733	0.049	0.045
DW	604	678	1,228,638	0.049	0.055
GW	747	295	1,459,938	0.051	0.020
M	50	78	6,039,870	0.052	0.032
TP	2,477	2,377	86,240	0.058	0.090
NLLA	3,119	1,940	5,160,467	0.048	0.046

Table 2.5: Average normalized abundances of ARGs per 100,000 genes and ranges by landfill and outgroup environment sample type. Averages were calculated for each sample type of the landfill and outgroup environment by summing the normalized abundances and dividing by the number of metagenomes per sample type. The range was determined for each sample type. Both the average and range were reported for the RGI and DeepARG separately. Sample types are abbreviated as follows: agriculture (AG), leachate well (LW), composite leachate cistern (CLC), wastewater treatment plant (WWTP), sewage (S), drinking water (DW), groundwater (GW), mining (M), leachate treatment plant (TP), and non-leachate landfill-associated samples (NLLA).

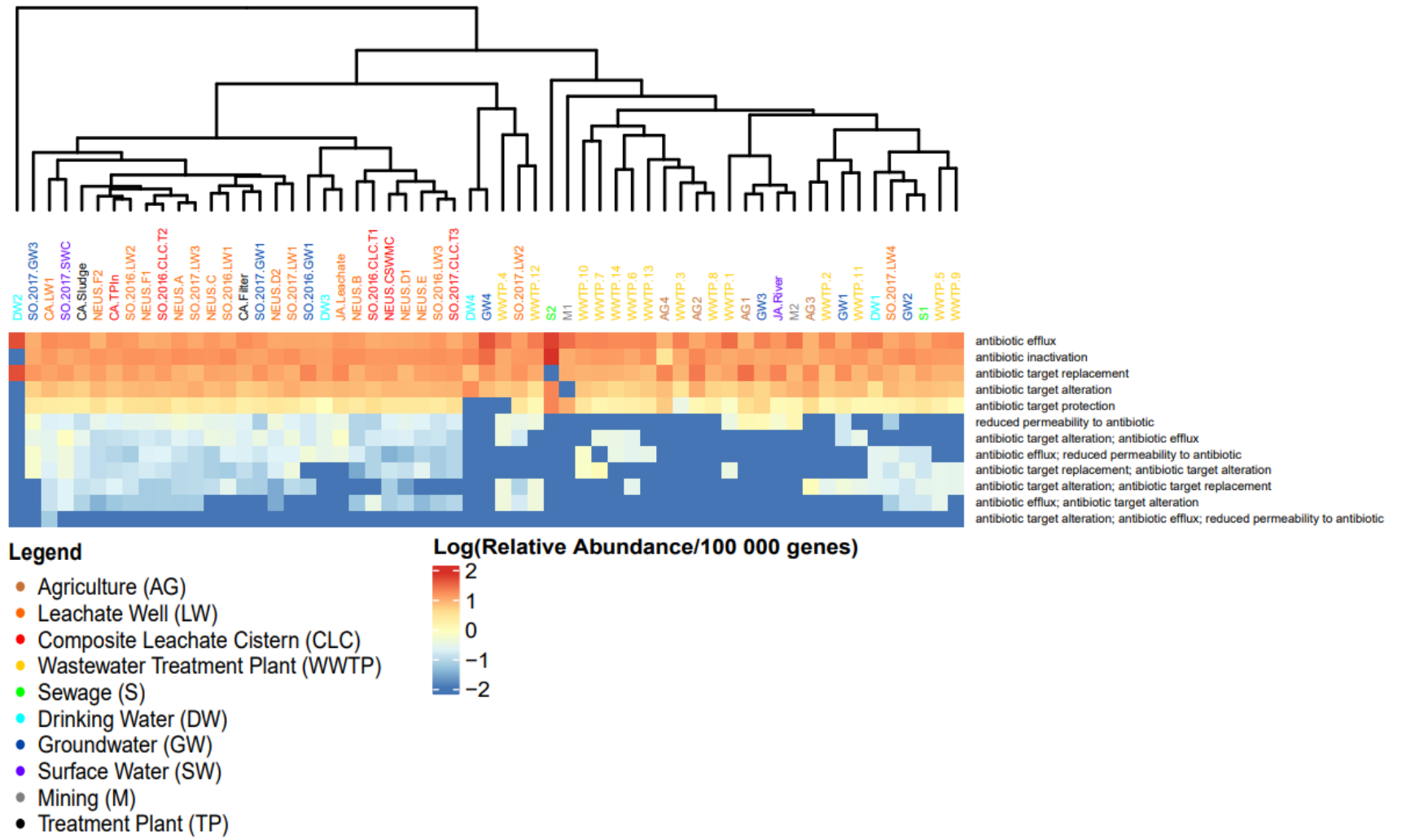
	AG	LW	CLC	WWTP	S	DW	GW	M	TP	NLLA
RGI Average	56.87	52.21	50.30	50.37	115.28	62.98	63.75	58.92	47.97	52.26
DeepARG Average	11.43	34.26	32.22	46.74	153.47	86.35	85.31	72.92	46.05	34.20
RGI Range	45.59- 67.25	45.87- 63.52	48.20- 52.36	44.63- 58.54	47.89- 182.67	44.73- 98.62	49.36- 98.96	57.69- 60.16	47.39- 48.56	45.50- 57.22
DeepARG Range	3.07- 17.65	10.30- 75.64	3.63- 50.56	29.70- 53.83	43.09- 263.85	10.79- 246.55	14.41- 258.81	50.13- 95.72	45.77- 46.33	19.09- 50.35

2.3.2 Resistance Mechanism Distribution and Prevalence Among Landfill and Outgroup Metagenomes

Landfills have distinct resistance mechanism signatures when compared to other environments (Figure 2.1), including a more diverse suite of antibiotic resistance mechanisms.

From the RGI predictions, the most abundant resistance mechanisms across all environments were single drug resistance mechanisms: antibiotic efflux, antibiotic inactivation, antibiotic target replacement, and antibiotic target protection. The remaining single type of resistance mechanism, reduced permeability to antibiotics, showed a substantial decrease in prevalence and was primarily seen in landfill metagenomes (Figure 2.1).

Figure 2.1: Heatmap of resistance mechanism relative abundance per 100,000 genes, as predicted by the RGI. All relative abundances were log scaled. Zero values were set to -2 to allow visualization. The ComplexHeatmap package in R was used to order rows by decreasing abundance and cluster by columns to show similarities across sites. Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA).



When observing the landfill data on an ordination plot, there is little clustering seen between the CA, JA, NEUS, and SO sites (Figure 2.2A). Additionally, there is little clustering between LW, CLC, GW, and TP sample types (Figure 2.2A). However, the TP samples do cluster together (Figure 2.2A). When comparing the landfill samples with other environments via ordination, resistance mechanism does not appear to be influenced by sample type as most metagenomes appear to cluster on the right side of the ordination plot excluding one S and one DW metagenome (Figure 2.2B).

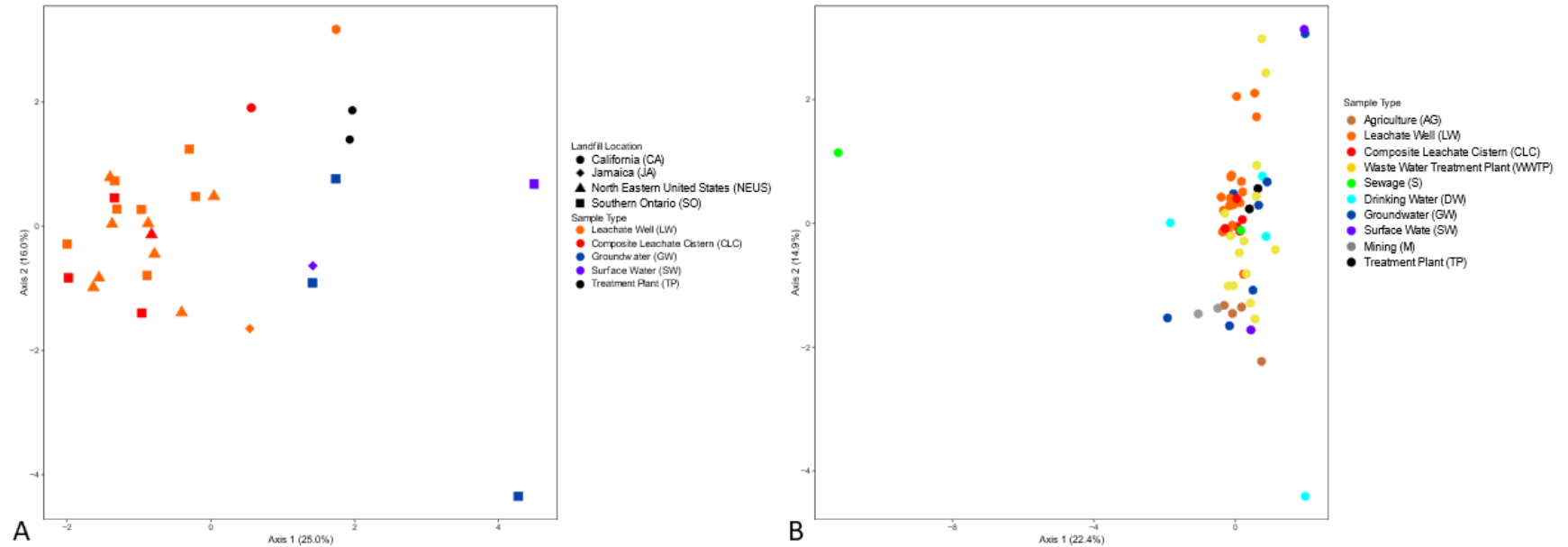


Figure 2.2: Ordination plot of resistance mechanisms using results from the RGI. The ggplot2 package in R was used to create ordination plots to summarize the community data from the resistance mechanism relative abundances. **(A)** Landfill locations are represented by shape; Southern California (CA) = circle, Jamaica (JA) = diamond, North Eastern Coast of the US (NEUS) = triangle, Southern Ontario (SO) = square. **(B)** Sample types are represented by colour; Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. See Table A.1 Supplementary Data 2.1 for scree plot information on percent variation per axis.

2.3.3 Target Drug Class Distribution and Prevalence Among Landfill and Outgroup Metagenomes

The top five target drug classes detected by the RGI were 1) diaminopyrimidine; 2) tetracycline & fluoroquinolone; 3) aminoglycoside; 4) tetracycline; and 5) carbapenem & penam & cephalosporin (Figure 2.3). The top five target drug class detected by DeepARG were 1) multidrug; 2) aminoglycoside; 3) macrolides & lincosamides & streptogramins (MLS); 4) fosmidomycin; and 5) unclassified (Figure 2.4). Antibiotic resistance to target drug classes that were identified at high abundance by both programs were tetracycline, aminoglycoside, MLS, glycopeptide, and peptide antibiotics resistance (Figure 2.3; Figure 2.4). Compared to other environments, landfills possess the widest diversity of resistance to different drug classes (Table 2.2; Table 2.4). The similarity in landfill-associated metagenomes' ARG complements can be observed through clustering on heatmaps (Figure 2.3; Figure 2.4). There are some exceptions – a small number of metagenomes from other environments group together with the landfill sample types on the heatmaps, from S, GW, and DW (Figure 2.3, Figure 2.4). The results from DeepARG show the landfill TP samples clustering with the WWTP samples, unlike the RGI results, which group the landfill TP samples with other landfill samples.

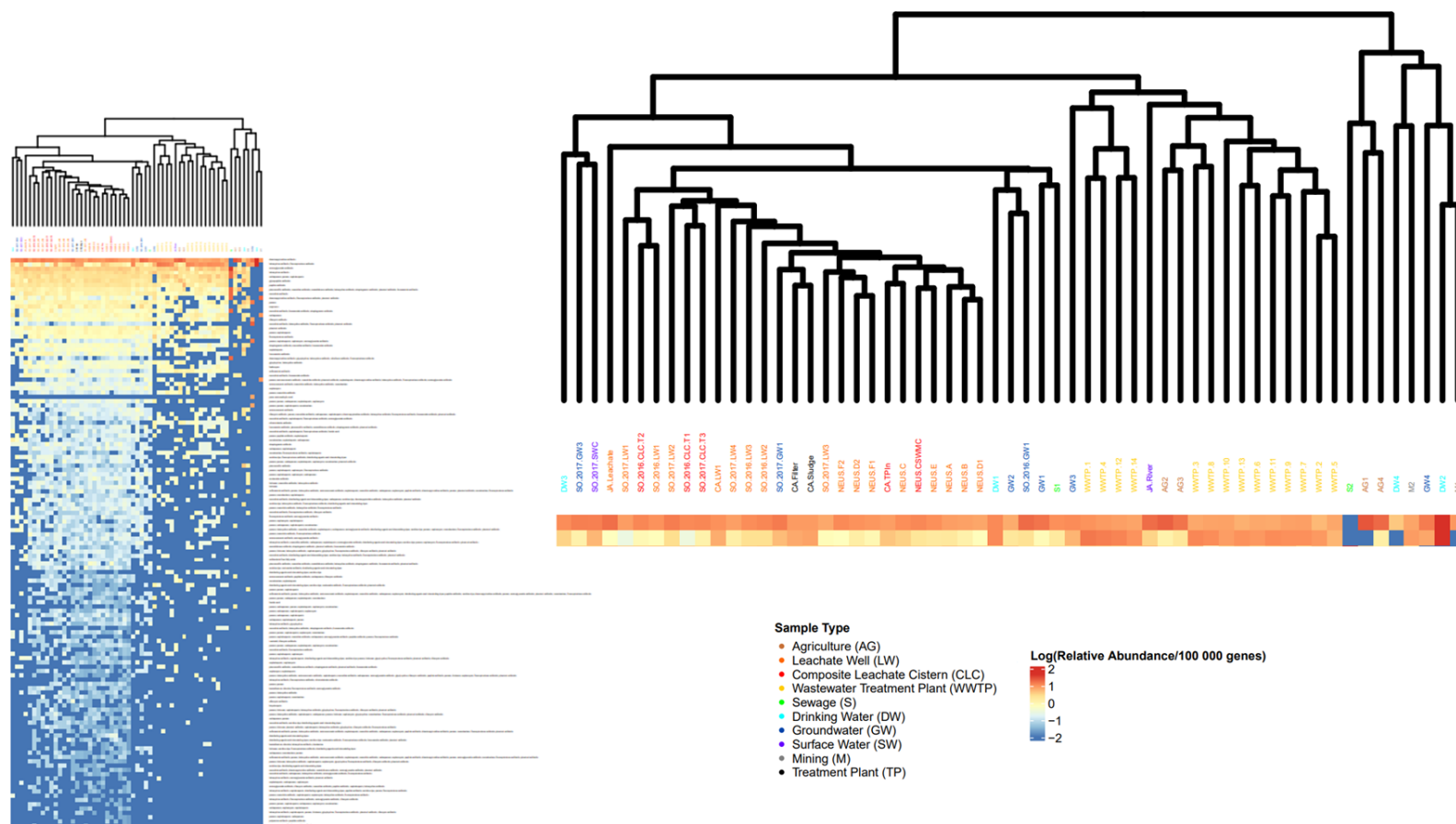


Figure 2.3: Heatmap of drug class relative abundance from the RGI. All relative abundances were log scaled. Zero values were set to -2 to allow visualization. The ComplexHeatmap package in R was used to order rows by decreasing abundance and cluster by columns to show similarities across sites. Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA). Full size heatmaps are presented in Supplementary file 1, which is available as a pdf file hosted by the Open Science Framework at https://osf.io/myg8q/?view_only=4a44d89e2044411dba79761a3f3261a5.

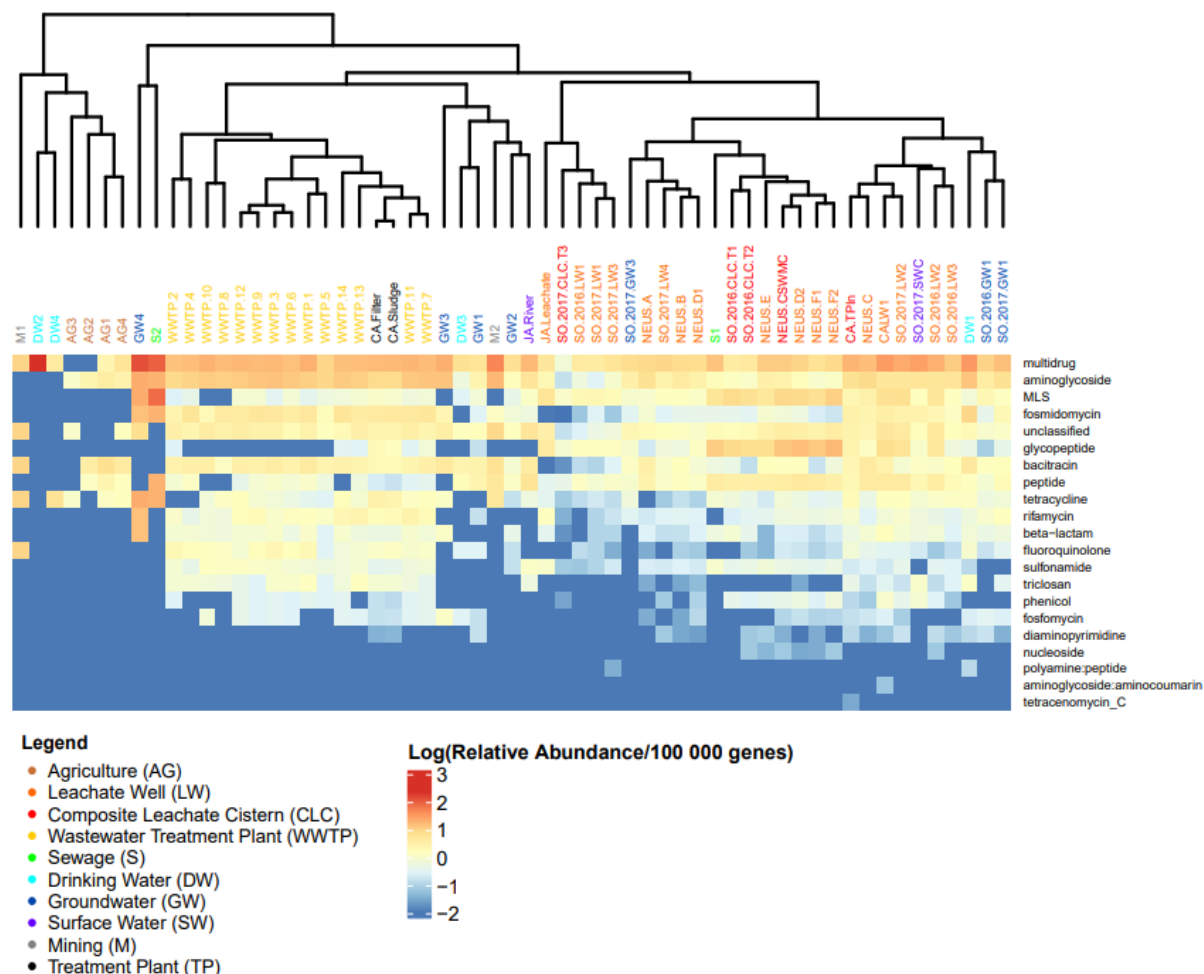
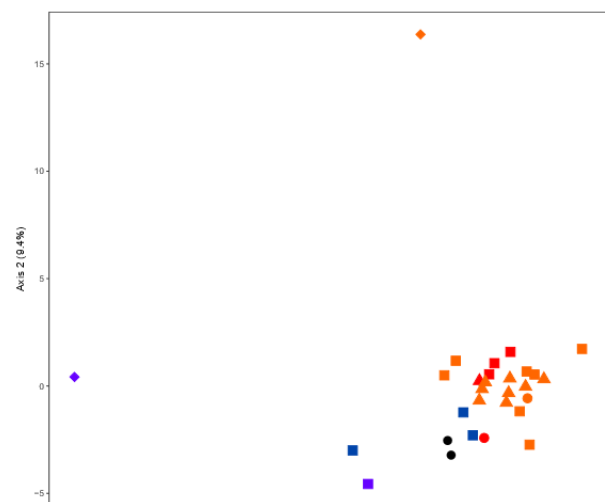


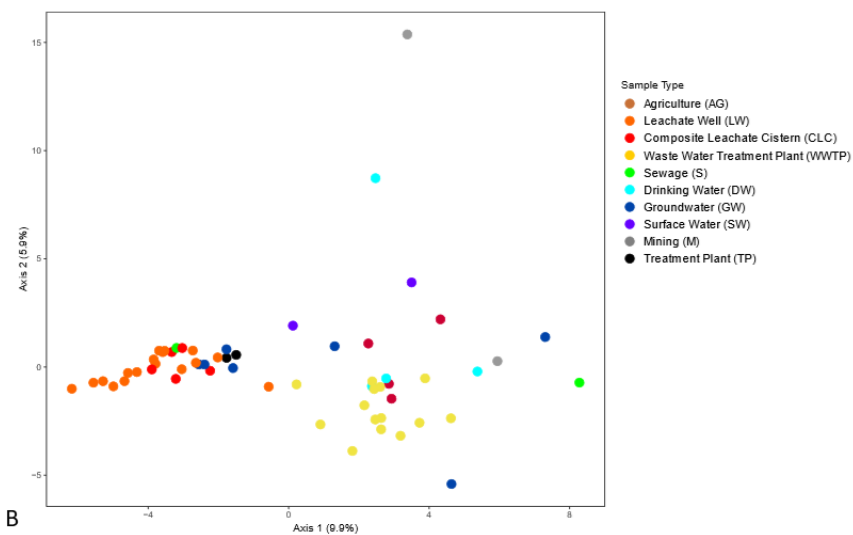
Figure 2.4: Heatmap of drug class relative abundance from DeepARG. All relative abundances were set to a log scale. Zero values were set to -2 to allow visualization. The ComplexHeatmap package in R was used to order rows by decreasing abundance and cluster by columns to show similarities across sites. Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA).

The ordination plot of antibiotic resistance to target drug classes within the landfill samples from the RGI does not show clustering by site, but does show clustering by LW, CLC, TP and some GW samples (Figure 2.5A). Similarly, the ordination plot of antibiotic resistance to target drug classes from DeepARG, for landfill-only samples, does not show clustering by site, though some clustering by LW, CLC, and TPs is seen (Figure 2.5B). There is less clustering seen when using the DeepARG predictions when compared to RGI result-based ordinations (Figure 2.5A; Figure 2.5B). When considering patterns of antibiotic resistance to target drug classes between landfill metagenomes and outgroup environment metagenomes, using RGI data, the sample types from landfills cluster together, with the LWs, CLCs, and TPs towards the left of the plot (Figure 2.5C). A S metagenome also clusters with the LW and CLC samples. The other outgroup environment metagenomes do not cluster closely with their own sample types, nor with other sample types (Figure 2.5C). Similarly, on ordination plots of antibiotic resistance to target drug classes using DeepARG results from landfill metagenomes and outgroup environment metagenomes, the landfill samples cluster together (Figure 2.5D). Some other environmental metagenomes cluster with the landfill samples, namely 3 DW samples and the same S metagenome. Comparatively, using the DeepARG results, the outgroup environment samples cluster closer to their own sample types, and are more closely associated with other outgroup categories as well (Figure 2.5D).

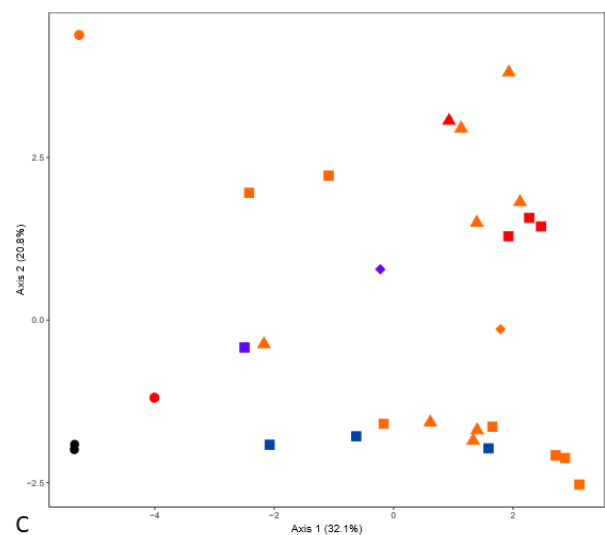
Figure 2.5: Ordination plots for drug class predictions from the RGI (A, B) and DeepARG (C, D). The ggplot2 package in R was used to create ordination plots to summarize the community data from the drug class relative abundances. (A and C) Landfill locations are represented by shape: Southern California (CA) = circle, Jamaica (JA) = diamond, North Eastern Coast of the US (NEUS) = triangle, Southern Ontario (SO) = square. (A - D) Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. See Tables A.1 Supplementary Data 2.1 and A.2 Supplementary Data 2.2 for scree plot information on percent variation per axis.



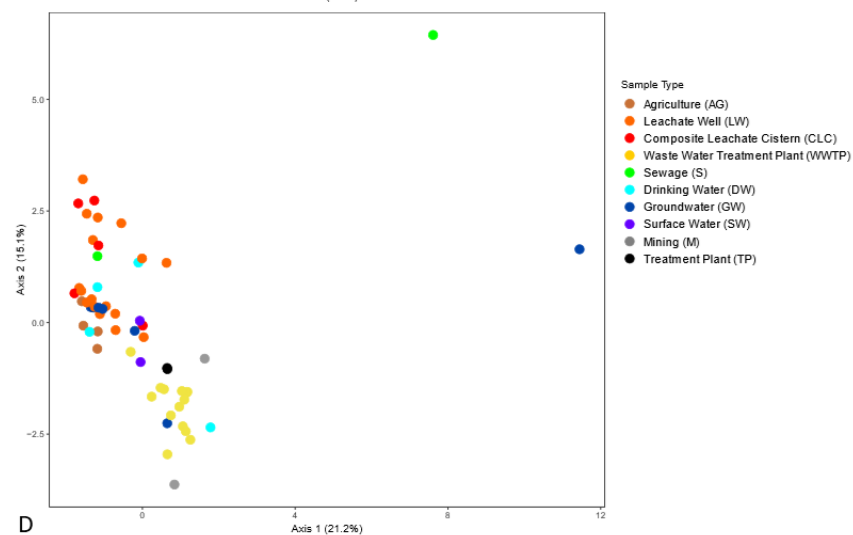
A



B



C



D

2.3.4 Gene Family Distribution and Prevalence Among Landfill and Outgroup Metagenomes

The top five gene families detected by the RGI were 1) resistance – nodulation – cell division (RND) antibiotic efflux pump; 2) trimethoprim resistance dihydrofolate reductase; 3) major facilitator superfamily (MFS) antibiotic efflux pump; 4) OXA beta – lactamase; and 5) ABC – F ATP – binding cassette ribosomal protection protein (Figure 2.6). The top five gene families detected by DeepARG were 1) *ompR* (an outer membrane porin component); 2) *kdpE* (part of a two component regulatory system and member of the *kdpDE* gene family in the CARD); 3) *bacA* (recycles undecaprenyl pyrophosphate during cell wall biosynthesis and is a member of the undecaprenyl pyrophosphate related proteins gene family in the CARD); 4) *rpoB2* (the rifamycin – resistant beta – subunit of RNA polymerase); and 5) *oprM* (an outer membrane protein that is part of the resistance – nodulation – cell division (RND) antibiotic efflux pump gene family in the CARD) (Figure 2.7). For DeepARG, the most abundant gene families for the landfills specifically differ from the overall DeepARG output, with *rpoB2*, *oprM*, *vanU* (a transcriptional regulator that is a *vanG* variant in the *vanG* gene cluster), *ugd* (part of the *pmr* phosphoethanolamine transferase in the CARD), and truncated response regulator *arcR* as the highest detected gene families in landfill samples. Gene families that were present in both the RGI and DeepARG results were *ompR*, *rpoB2*, *vanU*, and *ugd* (Figure 2.6; Figure 2.7). However, *rpoB2*, *vanU*, and *ugd* were at lower relative abundance in the RGI output, with *vanU* being the least abundant of all gene families, and only present in a few of the landfill samples. Landfills have a larger diversity of gene families when compared to other environments (Table 2.2; Table 2.4; Figure 2.6; Figure 2.7). In the heat map of gene families using results from the RGI, landfill samples group together, with the LWs, CLC, TPs, and two GW samples clustering together (Figure 2.6). Within these samples, there is also a S sample and two non-landfill-associated GW

samples. Similarly, the heatmap of gene families using results from DeepARG has landfill samples grouping together, with LW, CLCs, TPs, GW, and SW samples are grouping together (Figure 2.7). However, in this analysis, the landfill samples form two separate groups rather than one big cluster. The same S sample, and some GW and DW samples are interleaved with the landfill samples.

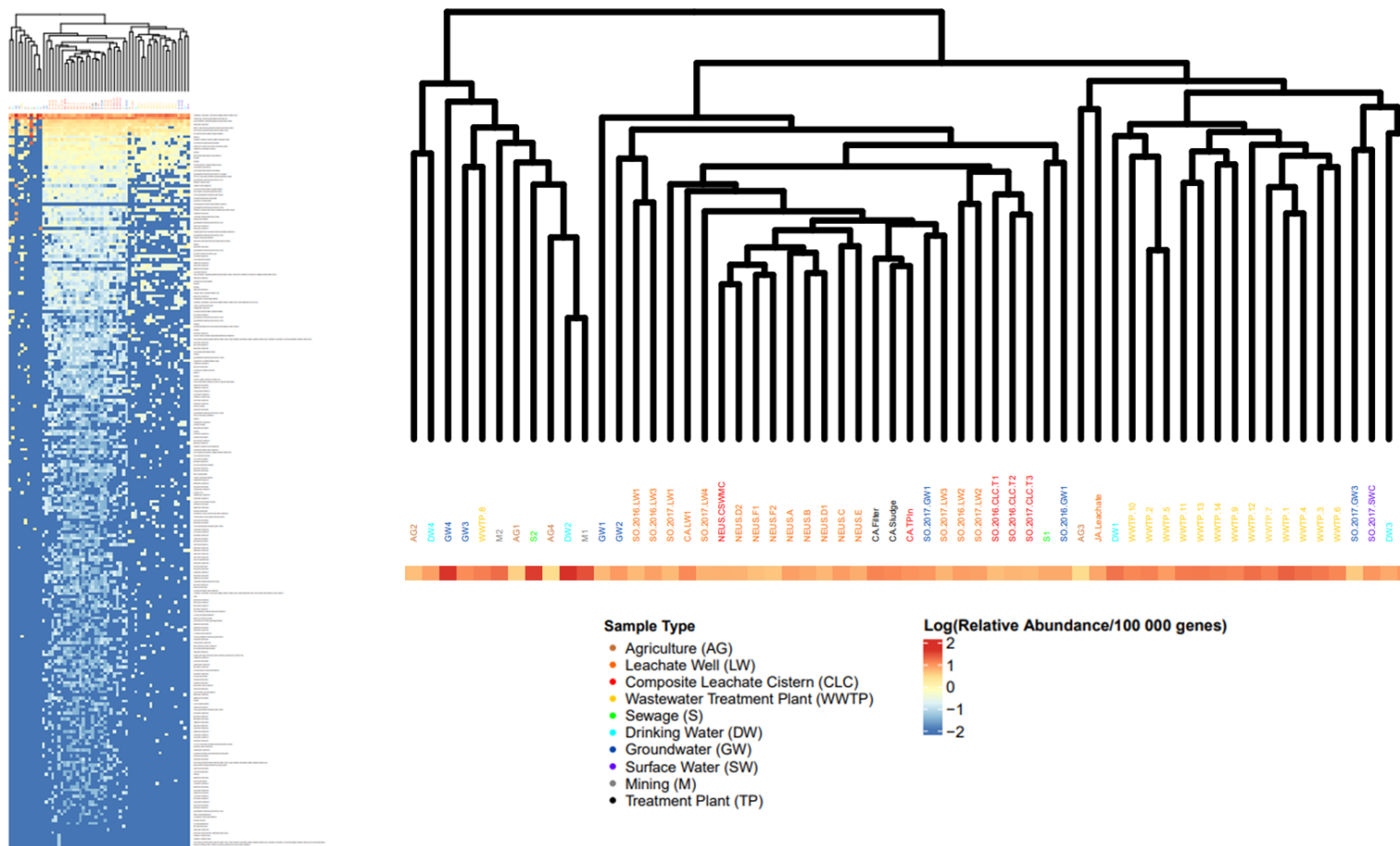


Figure 2.6: Heatmap of gene family relative abundance using results from the RGI. All relative abundances were set to a log scale. Zero values were set to -2 to allow visualization. The ComplexHeatmap package in R was used to order rows by decreasing abundance and cluster by columns to show similarities across sites. Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA). Full size heatmaps are presented in Supplementary file 1, which is available as a pdf file hosted by the Open Science Framework at https://osf.io/myg8q/?view_only=4a44d89e2044411dba79761a3f3261a5.

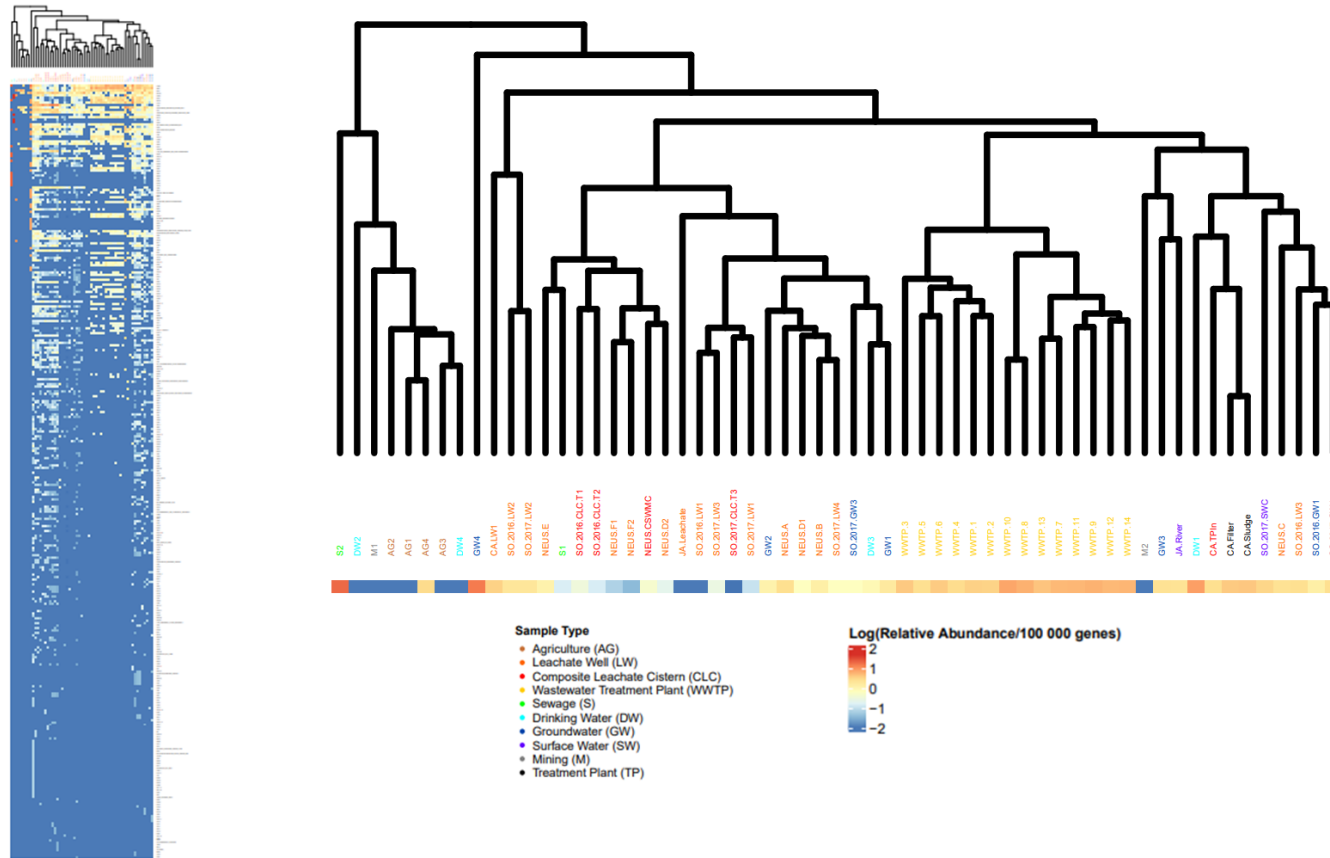
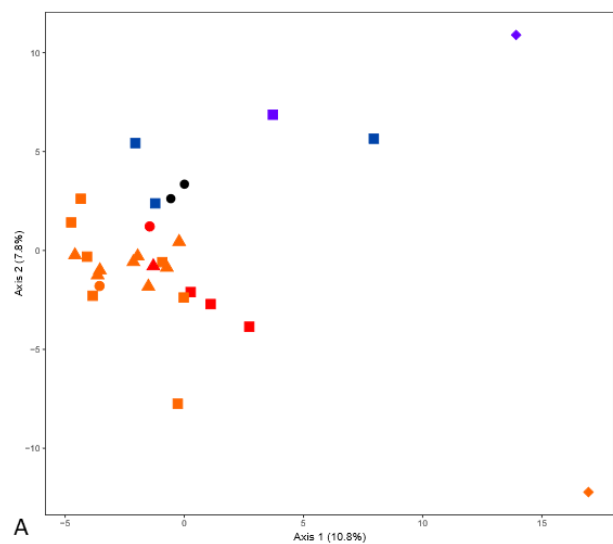


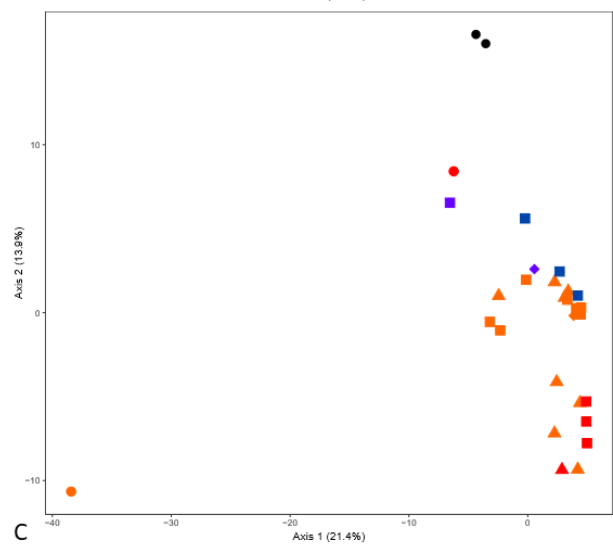
Figure 2.7: Heatmap of gene family relative abundance using results from DeepARG. All relative abundances were set to a log scale. Zero values were set to -2 to allow visualization. The ComplexHeatmap package in R was used to order by rows and cluster by columns to show decreasing abundance and similarities across sites. Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA). Full size heatmaps are presented in Supplementary file 1, which is available as a pdf file hosted by the Open Science Framework at https://osf.io/myg8q/?view_only=4a44d89e2044411dba79761a3f3261a5.

Similar to drug class categorization, ordination plots of gene families within landfills do not show clustering by site, but do show clustering by sample type as LWs, CLCs, and TPs cluster together, for both the RGI and DeepARG data (Figure 2.8A,C). Ordination plots of gene families using data from both landfills and outgroup environments using the RGI results show LWs, CLCs, TPs, and some GW samples clustering to the right of the ordination plot (Figure 2.8B), with one S sample clustering with these samples as well. The other GW samples, and samples from WTP, AG, some GW, DW, and SW do not cluster as tightly and are on the left side of the ordination plot. Conversely, the ordination plot of gene families within landfills and outgroup environments using DeepARG results shows less clustering between LWs and CLCs than the RGI results (Figure 2.8D). Unlike the RGI results, the DeepARG results show stronger clustering between WWTPs, DW, GW, TPs, and AG at the bottom left corner of the heatmap.

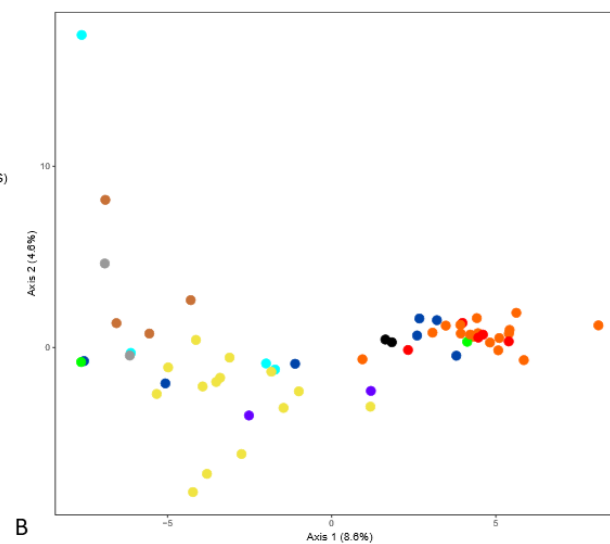
Figure 2.8: Ordination plot of gene families using results from the RGI (A, B) and DeepARG (C, D). The ggplot2 package in R was used to create ordination plots to summarize the community data from the gene family relative abundances. (A and C) Landfill locations are represented by shape: Southern California (CA) = circle, Jamaica (JA) = diamond, North Eastern Coast of the US (NEUS) = triangle, Southern Ontario (SO) = square. (A - D) Sample types are represented by colour: Agriculture (AG) = brown, Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Wastewater Treatment Plant (WWTP) = yellow, Sewage Waste (S) = green, Drinking Water (DW) = bright blue, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Mining Waste (M) = grey, Leachate Treatment Plant (TP) = black. See Tables A.1 Supplementary Data 2.1 and A.2 Supplementary Data 2.2 for scree plot information on percent variation per axis.



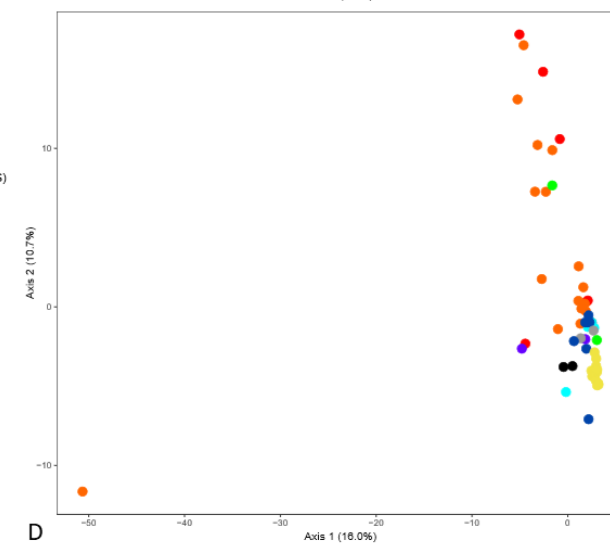
A



C



B



D

2.3.5 Classification of Bacteria with Antibiotic Resistance Genes (ARGs) in Landfill Metagenomes

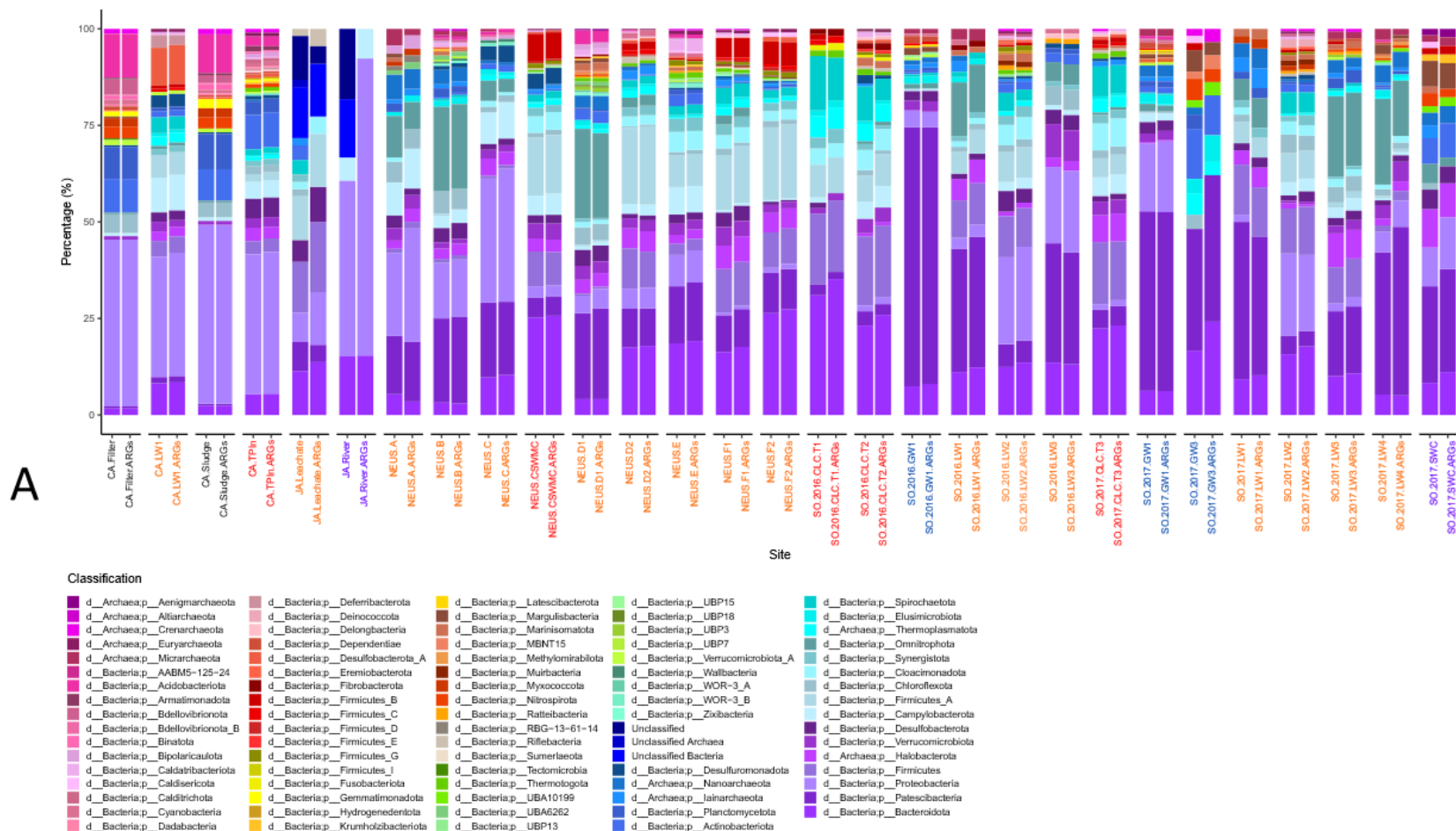
The taxonomic affiliations of the ARB identified differ between the four landfills (Figure 2.9). Observing the baseline community of all the CA landfills sites (Filter, LW1, Sludge, and TPIIn) there was a large amount of Proteobacteria classified. When looking at the hosts within the CA landfills that were classified as carrying ARGs Proteobacteria remains as the most abundant among all CA landfills sites with slight increases in abundance. According to the RGI both the baseline and hosts carrying ARG CA filter/sludge communities were abundant in Planctomycetota, Actinobacteriota, and Acidobacteriota (Figure 2.9A). DeepARG results also displayed Acidobacteriota at increased abundance in both communities for CA filter and CA sludge sites (Figure 2.9B). Omnitrophota and Spirochaetota were also at increased abundance in both baseline and non-baseline CA filter and sludge communities from DeepARG results. For the CA LW1 site Campylobacterota (Figure 2.9A), Chloroflexota (Figure 2.9B), and Desulfobacterota A (Figure 2.9) were at increased abundance for the baseline and non-baseline communities. The baseline JA leachate community was dominated by Bacteroidota, Patescibacteria, Proteobacteria, and Firmicutes. A large amount of the baseline JA leachate community was unclassified. The baseline JA leachate community had a large abundance of Firmicutes A (Figure 2.9A) and Campylobacterota (Figure 2.9B). In the non-baseline JA leachate community the abundance of unclassified bacteria decreased (Figure 2.9) and the abundance of Proteobacteria, Bacteroidota (Figure 2.9A), and Campylobacterota (Figure 2.9B) increased. The baseline JA river community was dominated by Proteobacteria and had a large abundance of unclassified bacteria (Figure 2.9). The non-baseline JA river community had an increase in the abundance of Proteobacteria and the disappearance of unclassified bacteria. From the NEUS A, B, and C sites there is an increase in Proteobacteria in the non-baseline community

compared to the baseline community (Figure 2.9). From the NEUS CSWMC, D2, E, F1, and F2 sites Firmicutes A increases in the non-baseline communities compared to the baseline communities. Compared to the other NEUS sites the NEUS D1 baseline and non-baseline communities have a higher abundance of Omitrophota and Patescibacteria. The SO 2016 CLC T1 and SO 2016 CLC T2 sites have an increase in the abundance of Bacteroidota in the non-baseline community compared to the baseline community in the RGI results (Figure 2.9A). The DeepARG results show a decrease in Bacteroidota for the non-baseline community from the SO 2016 CLC T1 site and no change for the SO 2016 CLC T2 site (Figure 2.9B). Both sites show an increase in Firmicutes A in the non-baseline community from the DeepARG results. The SO 2016 GW1 site has a large abundance of Patescibacteria in both the baseline and non-baseline communities (Figure 2.9). However, the abundance of Patescibacteria decreases in the non-baseline community. For the SO 2016 LW1 and LW3 sites there is a large abundance of Patescibacteria in both the baseline and non-baseline communities. The SO 2016 LW2 and LW3 sites have an increase in the abundance of Proteobacteria in the non-baseline community compared to the baseline community. The ARG scaffolds that were classified show high abundance of Patescibacteria in the SO 2016 GW1, SO 2016 LW1, and SO 2016 LW3 metagenomes. The SO 2017 CLC T3 site had a large abundance of Bacteroidota and Firmicutes in the baseline and non-baseline communities. The SO 2017 GW1 site had a large amount of Patescibacteria in the baseline and non-baseline communities as well as an increase in Firmicutes in the non-baseline communities. From the RGI results the SO 2017 GW3 site increased in Bacteroidota in the non-baseline communities (Figure 2.9A) whereas Bacteroidota decreased in the non-baseline communities from the DeepARG results (Figure 2.B). The DeepARG results showed an increase in Margulisbacteria and Actinobacteriota in the non-baseline communities.

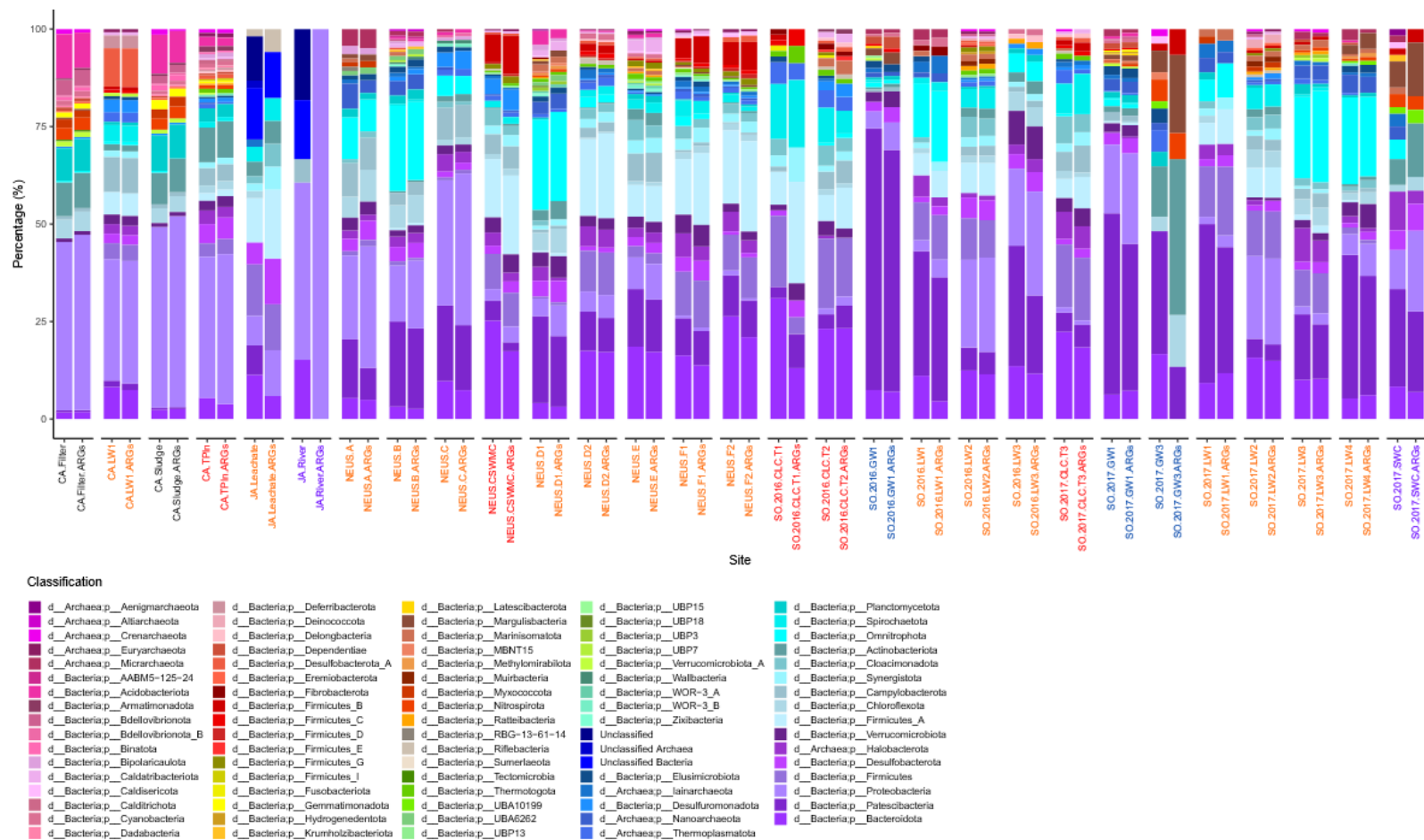
Patescibacteria was seen at high abundance in both baseline and non-baseline communities from the SO 2017 LW1, LW3, LW4, and SWC sites, however the RGI results showed an increase in Patescibacteria in non-baseline communities (Figure 2.9A) whereas DeepARG results showed a decrease in Patescibacteria in non-baseline communities (Figure 2.9B). SO 2017 LW3 and LW4 have a high abundance of Omnitrophota in the baseline and non-baseline communities (Figure 2.9). From the RGI results Omnitrophota decreases in abundance within the SO 2017 LW4 non-baseline community (Figure 2.9A) and from the DeepARG results Omnitrophota increases in abundance in the SO 2017 LW3 non-baseline community but decreases in abundance in the SO 2017 LW4 non-baseline community (Figure 2.9B). The SO 2017 LW2 site has a high abundance of Bacteroidota, Proteobacteria, and Firmicutes in both the baseline and non-baseline communities (Figure 2.9). Firmicutes slightly increases in the non-baseline community in the RGI and DeepARG results. In the RGI results Proteobacteria decreases in abundance in the non-baseline community whereas Bacteroidota increases in abundance in the non-baseline community (Figure 2.9A). Comparatively, in the DeepARG results Proteobacteria increases in abundance in the non-baseline community whereas Bacteroidota decreases in abundance in the non-baseline community (Figure 2.9A).

Figure 2.9: Bar chart of classified MAGs and scaffolds with ARGs that have been classified using results from the RGI (A) and DeepARG (B)

The ARG encoding genes within landfills from the RGI and DeepARG were compared to GTDB-TK files for each landfill. The ggplot package and geom_col function in R were used to create bar charts of the taxonomic affiliations of the total MAGs (left column) and predicted ARB-encoding MAGs (right column) for each sample. MAGs are summarized by phylum. The landfill sites used were in Southern Ontario (SO), North Eastern US (NEUS), California (CA), and Jamaica (JA).



B



2.4 Discussion

2.4.1 Differences Between the RGI and DeepARG

In this study, results from both the RGI and DeepARG were used and compared to each other rather than using the union of the results from both tools. By comparing the results from the two tools, the limitations of one tool can be overcome by using results from the other tool. For both the total number of ARGs and the percentage of ARG encoding genes that were identified as ARGs for the landfill metagenomes, the RGI nearly always identified a greater number and proportion of ARGs than DeepARG (Table 2.1; Table 2.2). Similarly, when looking at the total number of ARGs that were identified for the outgroup environment metagenomes, the RGI nearly always identified a greater number of ARGs than DeepARG (Table 2.3; Table 2.4). Unlike the percentage of genes that were identified as ARGs for the landfill metagenomes, neither the RGI and DeepARG were consistently higher or lower for the outgroup environmental metagenomes. DeepARG was trained across 30 ARG categories from the CARD, ARDB, and UNIPROT databases, and will only identify a putative ARG if the gene places within one of the 30 categories the model was trained on. DeepARG may have a lower total number of ARGs identified and percentage of ARG encoding genes as some submitted genes may belong to a new ARG category, meaning they may not match the training categories (Arango-Argoty et al., 2018). The RGI can account for genes that may belong to new ARG categories with its perfect, strict, or loose algorithms. More total ARGs and thus a higher percentage of ARG encoding genes were determined by the RGI, likely because it allows for variation between the reference sequences in the CARD and the metagenome-derived ARGs (Alcock et al., 2020).

The drug classes and gene families predicted by the RGI and DeepARG differed. One of the main differences between the target drug classes predicted by the RGI and DeepARG was that DeepARG predicted a high number of multidrug antibiotic resistance genes as well as

putative ARGs without a specific classification (“unclassified” on Figure 2.4). Within the DeepARG database’s 30 antibiotic resistance categories there are 2,149 groups and 14,933 reference sequences, of which 877 sequences belong to the multidrug category. If the submitted gene has determinants that can confer resistance to multiple antibiotic resistance groups it will be classified as multidrug by DeepARG, but DeepARG will not specify which antibiotic categories that gene is resistant to (Arango-Argoty et al., 2018). The RGI relies on the CARD, which is made up of 458 gene families, 64 drug classes, and 8 resistance mechanisms. For every resistance determinant identified, the gene family, resistance mechanism, and target drug class are reported (Alcock et al., 2023). The CARD does not have a specific multidrug category like the DeepARG database. Instead, if a gene has an antibiotic resistant determinant that can belong to multiple gene families, target drug classes, or resistance mechanisms, all categories are listed. This means that if a gene has multidrug resistance, the results from the RGI will specify what the multiple types of drug classes a gene is resistant to, unlike for the DeepARG output (Alcock et al., 2023; Arango-Argoty et al., 2018). Furthermore, DeepARG had a high relative abundance of unclassified target drug classes (Figure 2.4). These unclassified target drug classes may be genes that have antibiotic resistance determinants, but don’t match any of the antibiotic categories so they are labelled as unclassified (Arango-Argoty et al., 2018). The RGI may face less of a problem with divergent sequences, as it can classify genes that do not match the reference sequences with the use of the loose algorithm (Alcock et al., 2020).

2.4.2 Antibiotic Resistance Gene (ARG) Burden Across Environments

Looking at the total amount of ARGs across different environments, the landfill metagenomes have more ARGs classified compared to the outgroup environment metagenomes. Landfill metagenomes had 29,579 and 21,132 ARGs classified by the RGI and DeepARG, respectively,

whereas the outgroup environment metagenomes had 3,980 and 3,315 ARGs classified by the RGI and DeepARG respectively (Table 2.1; Table 2.3). However, the landfill (LW, CLC, and TP) metagenomes and NLLA metagenomes did not have the highest average normalized abundance per 100,000 ORFs – in fact, contrary to our expectations, landfills and associated samples had some of the lowest normalized abundances of ARGs (Table 2.5). Comparatively, the S metagenomes had the highest average normalized abundance per 100,000 ORFs, though with the highest range of variance between samples. The DW metagenomes showed high variability, with the highest range for drug classes and gene families for both the RGI and DeepARG. Despite the outgroup environment metagenomes having higher average relative abundance of ARGs per 100,000 ORFs, since there was a larger range for the outgroup environment metagenomes this means the average is less representative of the data compared to the landfill and non-leachate landfill associated metagenomes, which had smaller ranges.

The distribution of ARGs is influenced by several factors. The antibiotics present within an environment are a direct selection pressure for the ARGs that accumulate in that environment. Additionally, heavy metals can also impact the ARGs within an environment as metal ions promote HGT, and increased heavy metal concentration has been shown to be correlated with ARG enrichment (Lin et al., 2021). Exposure to non-antibiotic antimicrobial chemicals can also impact the ARGs present within an environment, as these chemicals still create conditions that can select for types of antibiotic resistance. The diversity of the microbial community present within an environment can also impact the ARGs within that same environment. The physical and chemical properties of an environment can also impact ARG presence and abundance as these properties can interact with the microbial communities that carry and transfer the ARGs (Lin et al., 2021). The abundance of ARGs may differ between the landfill metagenomes and

outgroup environment metagenomes as they may have different factors that are impacting ARG abundances. Landfills typically contain many pollutants such as heavy metals, pathogenic organisms bearing ARGs, and pharmaceuticals, all of which can impact the abundance of ARGs present in landfills (Bandala et al., 2021; Lin et al., 2021). Landfill-specific signatures are discussed in the next section. A study conducted on Lake Honghu, a lake in China disturbed by agriculture run off, saw that the microbial communities present in the lake made the strongest contribution to ARG profiles (Z. Wang et al., 2020). This study also observed increasing ARG concentration with increasing N and P (which indicate eutrophication) within lakes. Eutrophic waters enable the growth of ARB which in turn increases the amount of ARG (Z. Wang et al., 2020). The AG outgroup metagenomes used in this study were all freshwater samples associated with agriculture, exposure that may have impacted the ARB that could grow in the water samples, in turn impacting the ARGs present. AG metagenomes had mid-range levels of ARGs per 100,000 genes from the RGI predictions (average 56.87), but the lowest of any environment from DeepARG predictions (average 11.43). A study done on filtered water, chlorine disinfected water, and tap water in China observed an increase in ARGs present in drinking water after chlorination (Shi et al., 2013). Two of the outgroup environment metagenomes were unchlorinated drinking water from treatment plants, and the other two DW metagenomes' metadata did not specify if they were from chlorinated samples or not. DW samples may have had a higher average relative abundance of ARGs per 100,000 genes as the DW1 and DW4 samples may have been chlorinated, decreasing the microbial diversity, but enriching the microbes carrying ARGs, thus increasing the abundance of ARGs. Mining contributes to the presence of heavy metals in the environment close to the mining area. The presence of heavy metals in surrounding water systems could co-select for antibiotic resistance. One study done on

a mine in China assessed the impact that heavy metals from the mine had on the ARG abundance in the surrounding water systems (Zou et al., 2021). The authors saw that ARG abundance was correlated with the abundance of MGEs and heavy metals (Zou et al., 2021). Heavy metals create reactive oxygen species, which increase cell permeability, ultimately promoting the transfer of MGEs and any ARGs present on those MGEs (Lin et al., 2021). The mining outgroup metagenomes in this study had relatively low counts of ARGs, possibly due to low sequence coverage missing low abundance ARGs, low community diversity or a low abundance of MGEs. Given the average ARGs/100,000 genes was mid-range (56.87 for the RGI; 72.92 for DeepARG) it is likely a sequence depth issue impacting ARG detection. In sewage WWTPs, fecal pollution can explain ARG abundance rather than selection or HGT. Treated sewage has also shown a lower abundance of ARGs compared to raw sewage as fecal pollution is decreased (Karkman et al., 2019). The S outgroup metagenomes have the highest ARG proportion per 100,000 genes by a large amount (115.28 for the RGI; 153.47 for DeepARG) – indicating direct entry of ARGs to a system, even a high-flow transient system, can contribute high ARG prevalence. Groundwater environments have distinct microbial communities as a result of limited nutrient availability, which decreases the degradation rate of antibiotics, allowing for longer selection periods for ARGs and ARB (Huang et al., 2023). The ARGs present in groundwater are influenced by the surrounding river and spring water resistomes, as well as surrounding anthropogenic activities (Huang et al., 2023). The GW outgroup metagenomes came from four different locations. Three in the USA (New York, Nevada, and Colorado) and one from Sweden. The GW outgroup metagenomes may have had higher average relative abundance of ARGs because of potential limited nutrient availability decreasing the degradation rate of any antibiotics that are present, this would result in selection for ARGs and ARB. This would depend on the sampling location

as the surrounding soil composition, and anthropogenic activities would impact the ARG relative abundance as well (Huang et al., 2023). The GW metagenome from New York specifies that it is contaminated, but not by what it is contaminated with. Depending on what it is contaminated with it can impact the ARG relative abundance within that sample. In WWTPs, the ARG abundance has been seen to be impacted by the presence of heavy metals and antibiotics, meaning the removal of these materials can impact the ARGs that proliferate (Mao et al., 2015). The WWTP outgroup metagenomes had very similar ARG prevalences to landfill leachate – possibly due to their shared feature of high antibiotic load.

2.4.3 Landfill Specific Signatures for Resistance Mechanisms, Drug Classes, and Gene Families

Compared to other environments, landfills have specific signatures for resistance mechanisms, target drug classes, and gene families identified. Landfills and other environments had a high abundance of single drug types of resistance mechanisms (Figure 2.1). These resistance mechanisms are antibiotic efflux, antibiotic inactivation, antibiotic target replacement, antibiotic target alteration, and antibiotic target protection. One single drug type of resistance mechanism that was not seen across most outgroup environment metagenomes but is seen across all landfill metagenomes, though at lower abundance, is reduced permeability to antibiotics. Unlike the outgroup environment metagenomes, the landfill metagenomes have ARGs with multiple types of resistance mechanisms, again at lower abundances. Examples of these ARGs with multiple types of resistance mechanisms are antibiotic target alteration; antibiotic efflux, and antibiotic efflux; reduced permeability to antibiotic.

Under both the RGI and DeepARG analyses, resistance to the target drug classes tetracycline, aminoglycoside, MLS, glycopeptide, and peptide antibiotics were found in the landfill samples (Figure 2.3; Figure 2.4). These drug classes were not exclusive to the landfill

datasets and were detected in WWTP metagenomes as well by both the RGI and DeepARG. Using the RGI, tetracyclines and aminoglycosides were detected in all metagenomes except for GW4, DW2, and M1 (Figure 2.3). MLS antibiotics were absent from AG1, AG2, DW2, DW4, M1, M2, and GW4. Glycopeptides were absent from M1, DW2, GW3, GW4, and S2. Peptide antibiotics were absent from M1, DW2, GW4, and S2 metagenomes. Common target drug classes for which resistance has been detected within landfills at high abundance are macrolides, β -lactams, fluoroquinolones, sulfonamides, tetracyclines, aminoglycosides, MLS, multidrug, and vancomycin (Czatkowska et al., 2022; J.-Y. Wang et al., 2020; Zhao et al., 2018). Tetracyclines, aminoglycosides, MLS, β -lactams, fluoroquinolones, sulfonamides, and glycopeptides are all target drug classes seen in landfill studies previously that were detected by both the RGI and DeepARG in the landfill datasets.

Resistance to tetracyclines and aminoglycosides was detected by both the RGI and DeepARG. The RGI detected resistance to both at high abundance whereas DeepARG detected resistance to aminoglycosides at higher abundance than tetracyclines. Using the RGI results resistance to tetracyclines was absent from M1, DW4, AG1, and S2 metagenomes and resistance to aminoglycosides was absent from M1, DW2, and GW4 metagenomes. Using DeepARG results resistance to tetracyclines was absent from NEUS A, GW1, DW2-3, AG2, WWTP2 & 4 metagenomes and resistance to aminoglycosides was absent from M1, DW2, and GW4 metagenomes. Tetracyclines and aminoglycosides are broad spectrum antibiotics that inhibit bacterial protein synthesis (Singh et al., 2017). Tetracyclines are some of the earliest identified antibiotics, initially discovered from an actinomycete culture, and have been widely used for clinical treatment and animal agriculture (Bradford & Jones, 2012; Schnappinger & Hillen, 1996). Tetracyclines bind to the 16S rRNA of the 30S ribosomal subunit which prevents protein

synthesis (Singh et al., 2017). Some resistance mechanisms related to tetracyclines are 1) protection of the antibiotic target through binding site mutations as well as tetracycline specific ribosomal protection proteins; 2) reduction of intracellular concentrations of tetracycline with the use of MFS efflux transporters; and 3) inactivation of the antibiotic by enzymatic additions of hydroxyl groups to the core of tetracycline (Grossman, 2016; Schnappinger & Hillen, 1996). Comparatively, aminoglycosides are only used in clinical settings for the treatment of serious infections (Armstrong et al., 2012). Similar to tetracyclines, aminoglycosides prevent protein synthesis by binding the 16S rRNA of the 30S ribosome (Singh et al., 2017). Some resistance mechanisms related to aminoglycosides are 1) aminoglycoside modifying enzymes (AMEs); 2) target modification by mutations or methyltransferases; 3) change of uptake or efflux by the RND and MFS family; and 4) membrane proteases (Becker & Cooper, 2013). Macrolides were first discovered in actinomycetes, and are important in clinical settings for treating respiratory tract infections (Singh et al., 2017; Z.-Q. Xu et al., 2012). The RGI detected resistance to macrolides at high abundance, meanwhile resistance to this drug class was not detected by DeepARG. Macrolides stimulate the dissociation of the peptidyl tRNA from the ribosome which inhibits bacterial protein synthesis (Singh et al., 2017). Macrolides bind the 50S ribosomal subunit in the same way as the lincosamides and streptogramin B antibiotic classes. Resistance against these three drug classes also occurs in the same way, which is why these three drug classes are referred to as the MLS complex (Roberts et al., 1999). Both the RGI and DeepARG detected resistance to MLS antibiotics at high abundance. Resistance to MLS antibiotics was absent from M1, M2, DW2, DW4, AG1, and AG3 metagenomes in both the RGI and DeepARG results. The metagenomes GW4, WWTP 12, WWTP 14, and SO 2017 GW3 did not have resistance to MLS antibiotics according to the RGI. Additionally, the metagenomes GW3,

WWTP 10, WWTP 8, and AG 2-3 did not have resistance to MLS antibiotics according to DeepARG. Some resistance mechanisms for MLS antibiotics are 1) rRNA methylation of an adenine; 2) efflux from efflux pumps with homology to the MFS family or members of the ABC transporter superfamily; and 3) ribosomal mutation of the drug binding domain (Roberts et al., 1999; Z.-Q. Xu et al., 2012). Both the RGI and DeepARG detected resistance to β -lactam antibiotics in the landfill samples. The RGI detected resistance to β -lactams at higher abundance than DeepARG, categorized as carbapenem & penam & cephalosporin by the RGI. β -lactams were not detected in AG4, DW2, and M1 metagenomes in both the RGI and DeepARG results. The metagenomes GW3 and GW4 did not have resistance to β -lactams using the RGI results. Metagenomes that did not have resistance to β -lactams using DeepARG results were SO 2017 CLC T3, JA River, M2, GW1, DW3, S2, and AG1-3. β -lactams bind penicillin-binding enzymes that are responsible for synthesis and maturation of the bacterial cell wall, thus inhibiting cell wall synthesis (Page, 2012). There are many classes of β -lactam antibiotics that are used in clinical settings. β -lactam resistance can occur through the activation of or obtaining β -lactamases, which inactivate β -lactams through hydrolysis, which in turn prevents them from binding to penicillin-binding proteins, ultimately stopping the antibiotics from inhibiting cell wall synthesis (Singh et al., 2017). The RGI and DeepARG both detected the presence of resistance to fluoroquinolones in landfill metagenomes. Resistance to fluoroquinolones was detected at a slightly higher abundance by the RGI when compared to DeepARG. Resistance to fluoroquinolones was absent from GW4, DW2 & 4, AG1, and JA leachate metagenomes in both the RGI and DeepARG results. From the RGI results resistance to fluoroquinolones was absent from M1, S1, and WWTP 14 metagenomes. The metagenomes missing resistance to fluoroquinolones from DeepARG results were S1 & 2, SO 2016 CLC T1, SO 2017 GW3 & CLC

T3, JA River, M2, GW3, and AG2-3. Fluoroquinolones are part of the quinolone family, which are clinically used to treat bacterial infections (Singh et al., 2017). DNA gyrase and DNA topoisomerase IV are targeted by the quinolone family, which bind to these enzymes and prevent cell growth. Quinolone resistance has arisen through mutation of the target enzymes (Jacoby & Hooper, 2012; Singh et al., 2017). The RGI and DeepARG detected resistance to sulfonamide antibiotics at lower abundance. Resistance to sulfonamides was absent from SO 2017 GW1, SO 2017 GW3, SO 2017 SWC, M2, GW3 & 4, S2, DW2-4, and AG1 & 4 metagenomes in both the RGI and DeepARG results. The metagenomes GW1 and AG 2&3 were also missing sulfonamide resistance in the RGI results. Using DeepARG results the metagenomes DW1 and NEUS D1 were missing sulfonamide resistance. Sulfonamides have been used in clinical, veterinary, and agricultural settings (Deng et al., 2018; Zhou et al., 2021). Sulfonamides work by inhibiting cell growth by binding to dihydropteroate synthase (DHPS), which interferes with folate synthesis (Deng et al., 2018). Through either gene mutation of the *folP* gene or gene substitution with *sul* genes, bacteria can develop resistance to sulfonamides (Deng et al., 2018; Zhou et al., 2021). Resistance to glycopeptide antibiotics was detected by both the RGI and DeepARG. The RGI detected resistance to glycopeptides at a higher abundance than DeepARG. Glycopeptides inhibit bacterial cell wall synthesis in Gram-positive bacteria. Vancomycin is a well-known and frequently used glycopeptide that was naturally derived and originally used to treat MRSA infections that has become less effective due to increasing resistance (Arhin et al., 2012; Singh et al., 2017). Resistance to vancomycin has been seen in other landfills, and despite not being specifically detected within the landfill datasets by the RGI and DeepARG, related glycopeptide resistance was identified (J.-Y. Wang et al., 2020). In the CARD vancomycin is classified by the glycopeptide antibiotic drug class and don't have their own category. Since the RGI and

DeepARG use information from the CARD the results may be including vancomycin in glycopeptide antibiotics category. Resistance to glycopeptides was absent from M1, DW2, GW3 & 4, and S2 metagenomes in the RGI and DeepARG results. Metagenomes that were missing glycopeptide resistance exclusively in the DeepARG results were JA River, GW2, M2, DW3 & 4, AG 1-4, and WWTP 1; 3-6; 8-10; 12-14. With glycopeptide antibiotics, bacterial cell wall synthesis is prevented as glycopeptides bind to the terminal dipeptide, which is the precursor of the peptidoglycan chains, decreasing the integrity of the cell envelope and leading to lysis (Arhin et al., 2012; Singh et al., 2017; Stogios & Savchenko, 2020). Resistance mechanisms for glycopeptides relate to the production of modified peptidoglycan precursors, preventing glycopeptides from binding (Arhin et al., 2012). Peptide antibiotics are small, short peptides that have broad spectrum activity and have existed in nature for many years (Lai et al., 2023; Rima et al., 2021). Resistance to peptide antibiotics was detected by the RGI at higher abundance and across more metagenomes when compared to the DeepARG results. Metagenomes that were missing resistance to peptide antibiotics in the RGI and DeepARG results were M1, DW2, and GW4 metagenomes. Metagenomes missing resistance to peptide antibiotics that were exclusive to the RGI were AG4 and S2 metagenomes. The metagenomes that did not have resistance to peptide antibiotics in the DeepARG results were GW3, AG3, DW4, and WWTP8; 10; 14. Peptide antibiotics currently have little drug resistance, but there have been cases of reported resistance to peptide antibiotics through modification of the PhoP/PhoQ system, which decreases sensitivity to cationic peptide antibiotics (Xuan et al., 2023). Peptide antibiotics can be broken down into two classes: nonribosomally synthesized peptides and ribosomally synthesized peptides (Hancock & Chapple, 1999).

Common ARGs found within landfills are *aac6*, *aadA1*, *aadA*, *ampC*, *bacA*, *bla_{CTX-M}*, *bla_{OXA10}*, *bla_{OXY}*, *bla_{SHV}*, *bla_{TEM}*, *cat*, *dfrA*, *ermB*, *ermF*, *mefA*, *mexF*, *penA*, *qnrA*, *qnrB*, *qnrD*, *qnrS*, *sul1*, *sul2*, *tetO*, *tetM*, *tetQ*, *tetW*, and *tetX* (Czatkowska et al., 2022; Zhang et al., 2022). Gene families that were common between our landfill dataset and ARGs previously found within other landfills were *aac6*, *aadA*, *ampC*, *bacA*, *bla_{CTX-M}*, *bla_{OXA10}*, *bla_{OXY}*, *bla_{SHV}*, *bla_{TEM}*, *dfrA*, *ermB*, *ermF*, *mefA*, *mexF*, *qnrA*, *sul1*, *sul2*, *tetO*, *tetM*, *tetW*, and *tetX*. The *aac6* gene was detected by the RGI and DeepARG with a range of normalized abundance of 0.029-1.04 for the RGI and 0.0090-0.95 for DeepARG. The *aac6* gene was not the most abundant gene within the landfills, the *mexF* gene had the highest abundance across all landfills according to the RGI. The *aac6* gene was detected by the RGI in all landfill samples except for the JA metagenomes (Figure 2.6). With DeepARG, the *aac6* was absent from the JA metagenomes, as well as the NEUS B and NEUS D1 metagenomes. The *aac6* genes code for one of four classes of aminoglycoside acetyltransferases (AACs), some of the most common aminoglycoside modifying enzymes, and genes that are often found on MGE (Becker & Cooper, 2013). These AACs acetylate the 6' amino group in aminoglycoside antibiotics, an important feature for rRNA binding, providing resistance as the acetylated antibiotic can no longer bind the 16S rRNA of the 30S ribosome (Becker & Cooper, 2013; Singh et al., 2017). The *aadA* gene was detected by DeepARG at low abundance in the CA, NEUS CSWMC, NEUS E, NEUS F2, SO 2016 CLC T1, SO 2016 CLC T2, SO 2016 LW2, SO 2017 CLC T3, and SO 2017 LW2 metagenomes. Across all landfills the range of the normalized abundance of the *aadA* gene was 0.0090-2.62. *aadA* codes for an aminoglycoside adenylyltransferase which is a member of the aminoglycoside nucleotidyltransferase (ANT) (3'') class (Becker & Cooper, 2013; Stern et al., 2018). This gene is frequently found on plasmids (Stern et al., 2018). The *ampC* gene family was detected by the

RGI in all of the landfill metagenomes, with the exception of JA, at low abundance. Across all landfills the range of the normalized abundance for the *ampC* gene was 0.068-0.32. The *ampC* gene codes for AmpC β -lactamase, which breaks down β -lactams through hydrolysis. Since 1989, *ampC* have been detected on plasmids from both nosocomial and non-nosocomial samples (Jacoby, 2009). The *bacA* gene was detected by DeepARG in most of the landfills, except for the JA leachate metagenome, at varying abundance. Across all landfills the range of normalized abundance for the *bacA* gene was 0.018-4.53. The BacA protein has undecaprenyl pyrophosphate phosphatase activity, which decreases the cell integrity and causes the cell to lyse. Peptide antibiotics such as bacitracin are the targets of BacA proteins (El Ghachi et al., 2004). The RGI did detect low amounts of undecaprenyl pyrophosphate proteins in the CA LW1, JA leachate, NEUS A, NEUS B, NEUS C, NEUS D1, NEUS D2, NEUS E, NEUS F1, NEUS F2, SO 2016 CLC T1, SO 2017 GW1, SO 2017 LW 3, SO 2017 LW4, and SO 2017 SWC metagenomes, which is the gene family for *bacA* in the CARD. The range of normalized abundances of undecaprenyl pyrophosphate across all landfills was 0.024-0.35. The RGI detected *bla*_{CTX-M} in low amounts in most of the landfills except for the NEUS A, SO 2016 GW1, and SO 2017 SWC metagenome (0.060-0.58); *bla*_{OXA} at high amounts in all landfill metagenomes (0.95-3.50); *bla*_{OXY} at low amounts in 18 of the 29 landfill metagenomes (0.024-0.24) (excluding the JA metagenomes, NEUS C, SO 2016 CLC T1, SO 2016 CLC T2, SO 2016 GW1, SO 2016 LW1, SO 2017 CLC T3, SO 2017 GW 3, SO 2017 LW1, and SO 2017 SWC metagenomes); *bla*_{SHV} in low amounts in most of the landfills except for the JA metagenomes, NEUS D1, and the SO 2016 CLC T1 metagenome (0.040-0.32); and *bla*_{TEM} in low amounts in most of the landfills except for the JA metagenomes and the SO 2017 LW2 and SWC metagenomes (0.047-0.64). DeepARG detected *bla*_{OXA} at low amounts in all the CA

metagenomes; the JA leachate; NEUS A, C, CSWMC, D1, D2, E; SO 2016 CLC T1, GW1, LW1, LW2; and SO 2017 GW1, LW2, and LW3 metagenomes (0.032-0.44); and *bla*_{TEM} in low amounts in the NEUS A, B, C, CSWMC, and D1 metagenomes (0.024-0.037). The genes *bla*_{CTX-M}, *bla*_{OXA}, *bla*_{SHV}, and *bla*_{TEM} code for CTX-M, OXA, SHV, and TEM β -lactamases, which are all members of the extended spectrum β -lactamases and can be present on plasmids or chromosomes (Bush & Jacoby, 2010; Gundran et al., 2019). The *bla*_{OXY} gene belongs to *Klebsiella oxytoca* and encodes an OXY β -lactamase providing resistance against β -lactams through hydrolysis (Yang et al., 2022). The gene *cat* was detected by the RGI at high abundance among the landfill metagenomes (0.21-0.78) and at lower abundance among some of the landfills using DeepARG (0.040-0.38). The *cat* gene encodes chloramphenicol acetyltransferase (CAT), which inhibits ribosome binding and protein synthesis by transferring an acetyl group from acetyl-CoA to chloramphenicol molecules (Potrykus & Wegrzyn, 2001). The RGI detected *dfrA* in high amounts across all landfill metagenomes (7.45-19.72). On the other hand, DeepARG detected *dfrA* genes in very low amounts in only the CA TPin; NEUS B, CSWMC, D1, E; SO 2016 CLC T2, LW2, LW3; and SO 2017 LW2 metagenomes (0.024-0.16). The *dfrA* gene encodes a mutant dihydrofolate reductase that provides resistance to trimethoprim by making it less sensitive to trimethoprim (Ambrose & Hall, 2021; J. C. Lee et al., 2001). These genes are frequently found on mobile genetic elements, including being originally identified on plasmids (Sánchez-Osuna et al., 2020; Totake et al., 1998). The RGI did detect Erm 23S ribosomal RNA methyltransferase (*erm*) in all landfills at high amounts (0.32-3.16). DeepARG detected *ermB* at very low amounts in the NEUS C, CSWMC, D1, D2, E; SO 2016 LW 2; and SO 2017 LW2 metagenomes (0.024-0.081), and *ermF* at low amounts in 20 of 29 metagenomes, except CA Filter and Sludge; JA River; SO 2016 GW1, LW3; SO 2017 GW1, GW3, LW4, and SWC

(0.029-1.05). The *erm* genes are carried on plasmids and encode an adenosine-N⁶ methyl transferase which methylates the adenine residue in the peptidyl transferase center of the 23S rRNA. This disrupts hydrogen bonding, resulting in a conformational change in the 50S ribosomal subunit and resistance to MLS antibiotics (Z.-Q. Xu et al., 2012). The *mefA* gene was found by DeepARG in 15 of 29 landfill samples in low abundance except for the CA Filter and Sludge; JA; NEUS A; SO 2016 GW1, LW1, LW3; SO 2017 GW1, GW3, LW1, LW4, and SWC metagenomes (0.024-0.85). In the CARD the *mef* genes are classified as the MFS antibiotic efflux pump gene family. From the RGI results the MFS antibiotic efflux pump was seen across all landfills with a range of normalized abundance from 2.10-4.72. The *mefA* gene encodes an efflux protein with homology similar to the MFS efflux proteins and provides macrolide efflux (Roberts et al., 1999). The RGI detected the *mexF* gene at the highest abundance across all of the landfills (4.14, 23.20), whereas DeepARG detected the *mexF* gene at lower abundance across some of the landfill metagenomes and did not detect it from the JA Leachate; NEUS CSWMC, D2; SO 2016 CLC T1, CLC T2, LW1; SO 2017 CLC T3, GW3, LW1, and LW2 metagenomes (0.032-3.32). The *mex* genes encode RND efflux pumps. The *mexF* gene is part of the *mexEF-oprN* efflux operon which provides resistance to quinolones, chloramphenicol, and trimethoprim (Köhler et al., 1999). The *mexEF-oprN* operon becomes overexpressed as a result of mutational events (Aires et al., 2002). The RGI detected *qnr* at low abundance across most of the landfills except for the JA; SO 2016 LW2; SO 2017 CLC T3 and GW3 metagenomes (0.070-0.39). The six families of *qnr* genes are *qnrA*, B, C, D, S, and VC, and they provide resistance to quinolones (Flach et al., 2013). The *qnr* genes are plasmid-encoded genes that provide resistance by protecting type II topoisomerases, DNA gyrase and DNA topoisomerase IV, from quinolone inhibition (Flach et al., 2013; Hooper & Jacoby, 2015; Jacoby & Hooper, 2012). The RGI

detected *sul* genes at low abundance across most landfills except for SO 2017 GW1, GW3, and SWC (0.080-1.75). DeepARG detected *sul1* (0.0090-0.58) and *sul2* (0.027-0.70) across most landfills, except for SO 2016 GW1, LW1; SO 2017 GW1, GW3, LW1, LW4, and SWC metagenomes, at low abundance as well. The *sul* genes provide resistance to sulfonamides through gene substitution, as these genes replace the *folP* which codes for dihydrophenolic acid synthase, thus replacing the antibiotic target (Zhou et al., 2021). The RGI detected *tetX* at high abundance across all landfill sites (0.21-1.39). DeepARG also detected *tetX*, however it was across only 10 landfill sites (NEUS CSWMC, D1, E, F1, F2; SO 2016 CLC T1, CLC T2, LW2; SO 2017 CLC T3, and LW2) and at very low abundances (0.0090-0.14). DeepARG also detected *tetO*, *tetM*, and *tetW* across 10 landfill sites (CA LW1; NEUS C, CSWMC, D2, E, F1, F2; SO 2016 LW1; SO 2017 CLC T3, and LW2), and at low abundance. The *tetO* (0.030-0.035), *tetM* (0.027-0.11), and *tetW* (0.0090-0.16) genes all encode ribosomal protection proteins which provide resistance to tetracycline as they bind to the ribosome, causing tetracycline to be released and allowing protein synthesis to continue (Warburton et al., 2016). The *tetX* gene encodes tetracycline modifying enzymes that inactivate tetracyclines by adding hydroxyl groups to the tetracycline core (Grossman, 2016).

Antibiotic efflux and antibiotic inactivation were the most abundant resistance mechanisms as many of the top targeted drug classes are predicted to use those resistance mechanisms. Resistance mechanisms for tetracyclines and aminoglycosides are antibiotic efflux and antibiotic inactivation, where for macrolides it is largely antibiotic efflux (Becker & Cooper, 2013; Grossman, 2016; Roberts et al., 1999). A gene that was detected that encodes antibiotic efflux was *mefA* as it encodes a protein similar to MFS efflux pumps (Roberts et al., 1999). Genes that were detected that encode antibiotic inactivation were *tetX*, *aac6*, and *aadA* as they

encode tetracycline modifying enzymes, aminoglycoside acetyltransferases, and aminoglycoside adenylyl transferases respectively (Becker & Cooper, 2013; Grossman, 2016). Lower abundance resistance mechanisms included antibiotic target protection, a type of resistance mechanism against tetracyclines and quinolones (Flach et al., 2013; Warburton et al., 2016); antibiotic permeability, a type of resistance mechanism to aminoglycosides (Becker & Cooper, 2013); and antibiotic target modification, a type of resistance mechanism to aminoglycosides and MLS antibiotics (Becker & Cooper, 2013; Z.-Q. Xu et al., 2012). Genes that encode antibiotic target protection that were detected include *tetO*, *tetM*, *tetW*, and *qnr* as the *tet* genes encode ribosomal protection proteins, and the *qnr* gene protects type II topoisomerases, DNA gyrase and DNA topoisomerase IV, from quinolone inhibition (Flach et al., 2013; Hooper & Jacoby, 2015; Jacoby & Hooper, 2012; Warburton et al., 2016). A gene that was detected that encodes antibiotic target modification was the *erm* gene which provides resistance to MLS antibiotics by encoding an adenosine-N⁶ methyl transferase which methylates the adenine residue in the peptidyl transferase center of the 23S rRNA, disrupting hydrogen bonding, causing a conformational change in the target 50S ribosomal subunit (Z.-Q. Xu et al., 2012).

2.4.4 Antibiotic Resistant Bacteria (ARB) Classification in Landfill Metagenomes

The most numerous MAGs associated with ARGs were associated with the phyla Proteobacteria, Firmicutes, Firmicutes A, Patescibacteria, Bacteroidota, and Omnitrophota (Figure 2.9).

Proteobacteria, Firmicutes, and Bacteroidota have previously been identified as dominant hosts of ARGs (Rice et al., 2020; Y. Wang et al., 2022). Proteobacteria and Bacteroidota can impact human health, as they have the ability to produce or metabolize antibiotics as well as transfer environmental ARB and ARGs to humans (Y. Wang et al., 2022). Proteobacteria can carry *sul1* resistance genes, and Bacteroidota and Firmicutes can carry *tetO* and *tetW* resistance genes

(Y. Wang et al., 2022). Proteobacteria have been seen to be enriched in leachate (Y. Wang et al., 2022). In this study Proteobacteria was seen at higher abundance in leachate wells and treatment plants (Figure 2.9). Proteobacteria abundance did increase in the non-baseline communities indicating that they may be hosting ARGs, giving them a selective advantage. Bacteroidota and Firmicutes are primarily seen in anaerobic environments (Y. Wang et al., 2022). Bacteroidota and Firmicutes were seen at higher abundance in younger LW samples, as well as CLCs (Figure 2.9). From the RGI results they did tend to increase in abundance in the non-baseline samples which may indicate that they are hosting ARGs, giving them a selective advantage against antibiotic stressors. Patescibacteria were seen at highest abundance in the groundwater samples. Patescibacteria have previously been seen to be prevalent in groundwater environments (Tian et al., 2020). The Patescibacteria may be decreasing in abundance when looking at hosts that carry ARGs as Patescibacteria have a reduced genome size with functions for growth and reproduction (Tian et al., 2020). The presence of antibiotic selection pressures may be selecting for Patescibacteria that have acquired antibiotic resistance through HGT, thus decreasing those that don't have ARGs. Omnitrophota have been seen in oligotrophic water systems (Perez-Molphe-Montoya et al., 2022). In the landfill metagenomes Omnitrophota were primarily seen in leachate well samples (Figure 2.9). Common traits that have been reported for Omnitrophota are type IV pili, type II secretion systems, glycogen storage, and glycolysis (Perez-Molphe-Montoya et al., 2022). The slight increase or no change in Omnitrophota abundance between the baseline and non-baseline communities may indicate the presence of ARGs, aiding in selection of Omnitrophota against antibiotic selection pressures (Figure 2.9).

2.4.5 New Trends Seen in Landfill Metagenomes

Using the RGI and DeepARG, gene families detected in the landfills were *ompR*, *rpoB2*, *vanU*, and *ugd*. These gene families have previously not been reported in landfills before. The *ompR* gene codes for the OmpR protein which is a response regulator in the EnzV/OmpR regulatory system. The OmpR protein regulates outer membrane porins in response to β – lactams. Increases in β – lactams in the environment increases the expression of *ompR*, thus enhancing resistance to β – lactams (Ko & Choi, 2022). The *rpoB2* gene is a RNA polymerase β subunit gene that provides resistance to rifampin resistance. Rifampin is used in clinical settings to treat infections such as tuberculosis (Ishikawa et al., 2006). The *vanU* gene is part of the *vanG* regulatory operon and it encodes a transcriptional regulator. The *vanG* operon provides resistance to vancomycin (Depardieu et al., 2015). The *ugd* gene encodes a UDP-D-glucose dehydrogenase and provides resistance to polymyxin B (a peptide antibiotic) by synthesizing 4 – aminoarabinose which modifies the lipopolysaccharide of the organism (Mouslim & Groisman, 2003).

Peptide antibiotics have been reported to not have high amounts of resistance, but they are part of the top 5 target drug classes in this study. One mechanism of resistance to peptide antibiotics is through the modification of the lipopolysaccharides using the PhoP/PhoQ system (Xuan et al., 2023). One of the ways that the *ugd* gene alters an organisms lipopolysaccharide is through a PhoP/PhoQ two component system (Mouslim & Groisman, 2003).

2.4.6 Conclusions

Ultimately, the RGI and DeepARG were able to characterize the resistance mechanisms, target drug classes, and ARGs that are present within landfills and other environments. Antibiotic efflux, antibiotic inactivation, antibiotic target replacement, and antibiotic target protection were

the most abundant resistance mechanisms detected by the RGI. Overall, the RGI and DeepARG detected tetracyclines, aminoglycosides, MLS antibiotics, glycopeptides, and peptides antibiotics as the most abundant target drug classes. The ARGs that were the most abundant detected by the RGI and DeepARG were *ompR*, *rpoB2*, *vanU*, and *ugd*. Consequently, landfills harbour a large diversity of resistance mechanisms, target drug classes, and ARGs when compared to other environments, despite not having the highest average normalized abundance of ARGs per 100,000 genes, highlighting the specific ways in which landfills are important reservoirs for a variety of ARGs.

Chapter 3: Assessment of Plasmids and ARG Mobility Within Landfill Metagenomes

3.1 Introduction

Plasmids are circular DNA molecules that can replicate independently from the chromosome in bacteria and archaea (San Millan, 2018). Plasmids are mobile genetic elements that are acquired through horizontal gene transfer, predominantly via conjugation. For conjugation to occur, cell to cell contact between bacteria is required (San Millan, 2018). Plasmids can conjugate between organisms that are distantly related, allowing transfer of divergent genetic material to new hosts. Plasmids can be categorized as conjugative or mobilizable (Smillie et al., 2010). Conjugative plasmids encode their own transfer system, which is made up of mobility genes and genes encoding a membrane-associated mating pair formation complex (Francia et al., 2004; Smillie et al., 2010). Mobilizable plasmids rely on genes encoded elsewhere within a cell (e.g., chromosome or other plasmid) to construct a membrane-associated mating pair formation complex (Smillie et al., 2010). Plasmid acquisition aids in adaptation to new or stressful environments as plasmids can provide new genetic information to organisms, including novel metabolic capacity or other traits to benefit survival. Overall, plasmids have an important role in providing new genetic information through horizontal gene transfer of DNA.

Plasmids often encode genes for detoxification, virulence, and antibiotic resistance (Smillie et al., 2010). As plasmids can carry genes for antibiotic resistance, they are implicated in the spread of antibiotic resistance genes (ARGs) to pathogens (San Millan, 2018). Under a selective antibiotic pressure, horizontal gene transfer of ARGs between bacteria can increase fitness for the recipient, as the genes necessary for survival are now part of their genetic repertoire (Lerminiaux & Cameron, 2019).

Several tools have been developed to identify plasmids from sequence datasets. PlasClass is a bioinformatic tool for plasmid sequence classification that uses four logistic regression models to classify sequences (Pellow et al., 2020). The regression models were trained on sequences from the PLDSB for plasmid references (a plasmid database) and complete bacterial genome assemblies from the NCBI for bacterial chromosome references (Pellow et al., 2020). When classifying assembled metagenomic sequences, sequences are initially categorized based on their length, placed into categories of closest length of either 1 kb, 10 kb, 100 kb, or 500 kb. The length-appropriate logistic regression model is then applied to each category of sequences, and a probability of a sequence being a plasmid is assigned (Pellow et al., 2020). Typically, any assigned probability greater than 0.5 sufficient to predict a sequence is a plasmid, but more stringent thresholds can be applied. Another bioinformatic tool for plasmid sequence classification is Phage and Plasmid Recognizer for Metagenomes (PPR-Meta, (Fang et al., 2019). PPR-Meta classifies metagenomic sequences as either phages, plasmids, or chromosomes using a deep learning approach that was trained on sequences downloaded from the NCBI (Fang et al., 2019). A base one-hot matrix and codon one-hot matrix are used to represent the sequences, and the transformed sequences are then used as input for the base path and codon path, respectively. At the end of the paths the data are combined and output into a table of sequences along with three scores that add to one (Fang et al., 2019). These three scores represent the probability of a sequence being a phage, plasmid, or chromosome, where the category with the highest score is what that sequence is predicted to be (Fang et al., 2019).

In this chapter, I identify plasmids from the 29 metagenomic datasets associated with four landfills described in Chapter 2, to assess plasmid prevalence and ARG mobility using state-of-the-art plasmid detection software programs.

3.2 Methods

3.2.1 Landfill Sampling

Landfill sampling and metagenome processing were as described in Chapter 2.3.1 “Landfill sampling and metagenome processing”. Plasmid prevalence and ARG mobility has not been previously explored, and the CA landfill samples have not been previously reported in the literature.

3.2.2 Data Processing

The 1000bp+ scaffold files previously generated by pullseq (Chapter 2, 2.3.2) were used for plasmid classification by PlasClass and PPR-Meta (Fang et al., 2019; Pellow et al., 2020). For PlasClass, the “classify_fasta.py” command was used with default parameters. Any scaffolds with a probability greater than 0.79 were kept for consideration as a potential plasmid, a more stringent threshold than the default 0.5 (Pellow et al., 2020). The “./PPR_Meta” command was used with default parameters, and any scaffolds annotated as “plasmid” were kept for consideration as a potential plasmid from PPR-Meta (Fang et al., 2019).

For each landfill metagenome, outputs from PlasClass and PPR-Meta were compared to identify scaffolds that were considered plasmids by both plasmid classifiers. The intersection of the two prediction algorithms, all scaffolds that were detected by both classifiers, was kept, creating a “clean” plasmid list.

Virsorter and CheckV were used on the 1000bp+ scaffold files to detect potential phages within the scaffold files (Guo et al., 2021; Nayfach et al., 2021). The “virsorter” command was used with default settings. The VirSorter results were checked for quality with CheckV by using the “checkv” command under default settings. Scaffolds that were listed as “High-quality” phage elements from CheckV were compiled into a list for comparison to the plasmid list. Scaffolds that were detected as phages by VirSorter and CheckV and as plasmids by PlasClass and PPR-

Meta were removed from the plasmid lists and further analyses. The percent scaffolds that were plasmids was calculated by dividing the number of predicted plasmid scaffolds by the total number of scaffolds and multiplying by 100.

3.2.3 Determination of Plasmid Load and Antibiotic Resistance Gene (ARG) Mobility

Total plasmid and phage counts were determined for each metagenome. The total counts were used to calculate the percent scaffolds that are plasmids and phages. ARG counts from the Resistance Gene Identifier (RGI) and DeepARG were combined to get total ARG counts. Duplicate ARGs that were detected by both programs were dereplicated to ensure they were only included once in the total ARG count. Scaffolds that were determined to be plasmids from the clean plasmid lists were compared to ARG-encoding scaffolds to identify ARG-encoding plasmids. The plasmid load was calculated as the count of plasmids encoding ARGs divided by the total number of ARG-containing scaffolds in the sample (Kraemer et al., 2022).

Plasmid-encoded ARGs were summarized using their classifications from the RGI and DeepARG into resistance mechanism (RGI only), drug class, and gene family categories, for comparison between samples. ARG counts were normalized to the number of predicted ARGs per 100,000 genes per metagenome to normalize for differences in the metagenome sizes. The plasmid-encoded ARGs' classifications were compared to the full, normalized ARG datasets generated by the RGI and DeepARG, which were described in Chapter 2 (2.3.2). Bar charts were created for resistance mechanism (RGI only), drug class, and gene family to compare the distribution of classification categories between the plasmid-encoded ARGs and the full ARG datasets.

Bar charts comparing ARG prevalence and classifications between plasmid-encoded ARGs and the full datasets were made using the normalized datasets in R using the ggplot package and the geom_col function (<https://github.com/tidyverse/ggplot2>).

3.3 Results

The 29 landfill metagenomes previously described in Chapter 2 were used as the datasets in this chapter. Overall, leachate wells (LWs) and other sample sites from the North Eastern United Stated (NEUS) had the highest total plasmid counts. Conversely, the treatment plants (TP) and other Californian (CA) sample types had the highest proportion of scaffolds that were plasmids. Leachate wells (LWs) and samples from the Southern Ontario (SO) site in 2017 had the highest count of scaffolds detected as phages, as well as the highest proportion of scaffolds that were phages. The plasmid load was largest for TPs and the CA landfill samples. The count of ARGs associated with resistance mechanisms and drug classes detected on plasmids was markedly lower than the ARG count within the entire datasets (Figure 3.1). The proportion of resistance mechanisms and drug classes on plasmids was generally consistent with the proportion seen in the full landfill metagenomic datasets, with an increase in antibiotic efflux and multidrug (Figure 3.3).

3.3.1 Total Plasmid Counts

LWs were the sample type with the largest total plasmid count (500,939) followed by composite leachate cisterns (CLCs) (158,068), and then TPs (107,108; Table 3.1). Groundwater (GW) and surface water (SW) samples had the smallest total plasmid counts (Table 3.1). In all cases, the size of the combined metagenome datasets roughly scaled with number of plasmids detected. The data was normalized to percentages to allow direct comparisons. TP samples had the largest

percentage of predicted plasmid scaffolds by a substantial margin, with 8.39% (Table 3.1). SW samples had the second highest percent of predicted plasmid scaffolds at 5.81% (Table 3.1), despite relatively smaller metagenome sizes. GW, CLC, and LW samples had similar percentages (4.12-4.72%), with LW samples having the lowest percentage (Table 3.1).

Table 3.1: Predicted plasmids and phages by sample type. The number of plasmids represents the intersection of predictions from PlasClass and PPR-Meta. Scaffolds that were also predicted as phages were removed from the plasmid counts.

Sample Type	Predicted Plasmids	Total Scaffolds	Percent Scaffolds that are Predicted Plasmids	Predicted Phages	Percent Scaffolds that are Predicted Phages
Leachate Well (LW)	500,939	12,153,908	4.12	52,932	0.44
Composite Leachate Cistern (CLC)	158,068	3,345,771	4.72	7,324	0.22
Groundwater (GW)	47,667	1,105,570	4.31	3,135	0.28
Surface Water (SW)	9,957	171,274	5.81	285	0.17
Treatment Plant (TP)	107,108	1,276,618	8.39	1,163	0.13

When comparing across site locations, the CA samples and samples from the NEUS landfill had a higher total plasmid count compared to the other sites, with the NEUS sites having the greatest total plasmid count (388,977; Table 3.2). The Jamaican (JA) samples had the lowest total plasmid counts (7,755; Table 3.2). When normalized to percentages, the CA samples had the highest percent scaffolds that are predicted plasmids (7.43%) followed by the JA samples (5.21%; Table 3.2). The NEUS and SO samples from 2017 had similar percentages for predicted plasmid scaffolds, at 4.08 and 4.01%, respectively. The lowest percent scaffolds that are plasmids were the Southern Ontario samples from 2016 (3.77%; Table 3.2).

Table 3.2: Predicted plasmids and phages by site. The number of plasmids represents the intersection of predictions from PlasClass and PPR-Meta. Potential phages were removed from the plasmid counts. Scaffolds that were also predicted as phages were removed from the plasmid counts. Landfill locations are Southern California (CA), Jamaica (JA), North Eastern Coast of the US (NEUS), Southern Ontario (SO).

Site	Predicted Plasmids	Total Scaffolds	Percent Scaffolds that are Predicted Plasmids	Predicted Phages	Percent Scaffolds that are Predicted Phages
SO 2016	96,052	2,548,561	3.77	15,444	0.61
SO 2017	118,568	2,954,692	4.01	23,910	0.81
JA	7,755	148,833	5.21	705	0.47
NEUS	388,977	9,543,940	4.08	21,275	0.22
CA	212,387	2,857,115	7.43	4,005	0.14

When looking at the total number of scaffolds that were detected as phages, the LW samples had the largest amount of phages detected (52,932) followed by CLC samples (7,324; Table 3.1). GW and TP samples contained 3,135 and 1,663 predicted phages respectively. The surface water samples had the least amount of phages detected (285). The sample site with the largest count of phages detected were the SO samples from 2017 (23,910), followed by the samples from the NEUS samples (21,275; Table 3.2). The SO samples from 2016 had a higher count of phages detected (15,444) than the CA (4,005). The JA samples had the lowest number of phages detected (705).

Once normalized to percent scaffolds that are phages, the LW samples had the highest percentage (0.44%; Table 3.1) and the TP samples had the lowest percentage (0.13%). The GW, CLC, and SW samples had similar percentages (0.17 – 0.28%). When looking at the data by sample site, the SO samples from 2017 had the highest percentage (0.81%; Table 3.2) and the CA samples had the lowest percentage (0.14%).

3.3.2 Plasmid Loads

LWs were the sample type seen to have the largest number of total ARGs detected (31,162), followed by CLCs (8,876; Table 3.3). The SW samples had the lowest count of ARGs detected (628). Similarly, the LW samples and CLCs had the largest number of ARGs characterized as plasmid encoded (1,687 and 615, respectively). The SW samples also had the lowest count of ARGs characterized as plasmid encoded (45). However, the plasmid load, which is the proportion of plasmid-encoded ARGs out of all detected ARGs, was highest for the TP samples (0.12) and lowest for the LW samples (0.054; Table 3.3). When looking at the proportion of plasmid-encoded ARGs as a function of the total number of plasmids, the TP samples had the highest percentage (0.49%) and the LW samples had the lowest percentage (0.34%). The CLC, GW, and SW samples had similar percentages of plasmids encoding ARGs (0.39-0.45%).

Table 3.3: Plasmid-encoded ARGs and plasmid load by sample type. The number of ARGs were compiled from the intersection of predictions from RGI and DeepARG. The plasmid load was calculated by dividing the number of ARGs that were characterized as plasmid-encoded by the total number of ARGs in a sample type.

Sample Type	Number of ARGs	Number of Plasmid-Encoded ARGs	Plasmid Load	Percent Plasmids Carrying ARGs
Leachate Well (LW)	31,162	1,687	0.054	0.34
Composite Leachate Cistern (CLC)	8,876	615	0.069	0.39
Groundwater (GW)	3,364	209	0.062	0.44
Surface Water (SW)	628	45	0.072	0.45
Treatment Plant (TP)	4,468	522	0.12	0.49

The sample site with the highest count of ARGs detected were the samples from the NEUS landfill (23,454), followed by the CA landfill samples (9,422; Table 3.4). The JA samples had the least amount of ARGs detected (411). Similarly, the samples from the NEUS had the highest number of ARGs characterized as plasmid encoded (1,198), followed by the CA samples (989), and the JA samples had the smallest number (21). However, the CA samples had the largest plasmid load (0.105) and the lowest plasmid loads were seen in the JA and NEUS samples (0.051). The SO 2016 and SO 2017 samples had similar plasmid loads (0.055 and 0.059). When looking at the percentage of plasmids encoding ARGs out of all plasmids, the CA samples had the highest percentage (0.47%) and the Jamaican samples had the lowest percentage (0.27%). The SO 2016, SO 2017, and NEUS samples had percent plasmids encoding ARGs ranging from 0.31% to 0.42%.

Table 3.4: Plasmid-encoded ARGs and plasmid load by site. The number of ARGs were compiled from the intersection of predictions from RGI and DeepARG. The plasmid load was calculated by dividing the number of ARGs that were characterized as plasmid encoded by the total number of ARGs for each site. Landfill locations are Southern California (CA), Jamaica (JA), North Eastern Coast of the US (NEUS), Southern Ontario (SO).

Site	Number of ARGs	Number Plasmid-Encoded ARGs	Plasmid Load	Percent Plasmids Carrying ARGs
SO 2016	6,752	368	0.055	0.38
SO 2017	8,459	502	0.059	0.42
JA	411	21	0.051	0.27
NEUS	23,454	1,198	0.051	0.31
CA	9,422	989	0.11	0.47

3.3.3 Classification of Plasmid-Encoded Antibiotic Resistance Genes (ARGs)

Plasmid-encoded ARGs represent a relatively small fraction of the total ARGs within a given landfill metagenome dataset, but there are enrichments of specific resistance mechanisms and drug classes within plasmid-encoded ARGs (Figures 3.1-3.4). Antibiotic inactivation, antibiotic efflux, antibiotic target replacement, and antibiotic target alteration are resistance mechanisms that were at high abundance on plasmids compared to other resistance mechanisms on plasmids (Figure 3.1; Figure 3.2).

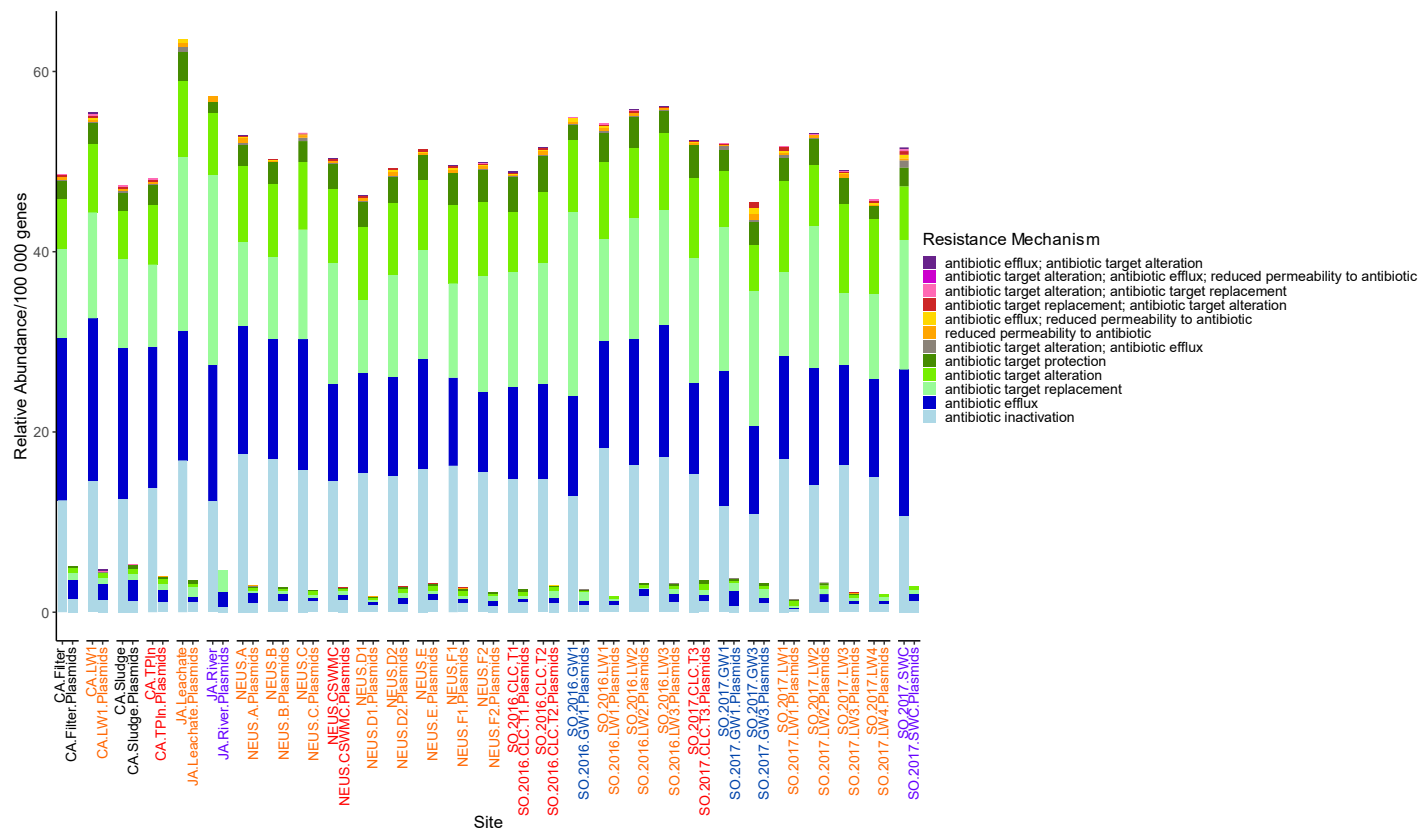
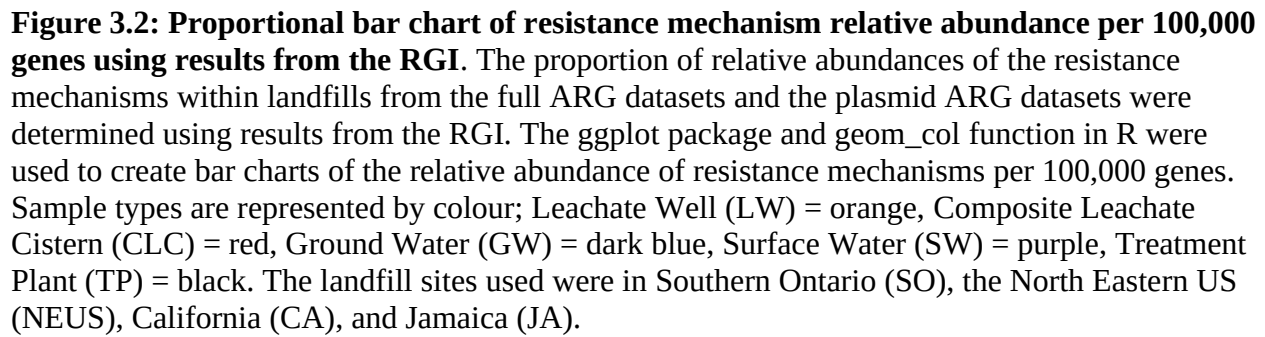


Figure 3.1: Resistance mechanism relative abundance per 100,000 genes using results from the RGI for full metagenome and plasmid-specific predictions. The relative abundances of the resistance mechanisms within landfills from the full ARG datasets and the plasmid ARG datasets were determined using results from the RGI. The ggplot package and geom_col function in R were used to create bar charts of the relative abundance of resistance mechanisms per 100,000 genes. Sample types are represented by colour; Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Treatment Plant (TP) = black. The landfill sites used were in Southern Ontario (SO), the North Eastern US (NEUS), California (CA), and Jamaica (JA).



Multidrug, MLS, aminoglycosides, and glycopeptides are target drug classes that are seen at higher relative abundances on plasmids compared to other drug classes on plasmids (Figure 3.3; Figure 3.4). The distribution of resistance mechanisms on plasmids is consistent with the distribution in the full landfill datasets (Figure 3.2), with the exception that antibiotic inactivation is slightly enriched on plasmids. At the drug class level, the distribution of drug classes on plasmids is fairly consistent with the distribution in the full landfill datasets excluding a few metagenomes (JA Leachate; SO 2016 GW1, LW1, LW2; SO 2017 GW3, LW1, SWC) (Figure 3.4), and with the exception that the multidrug target drug class is slightly enriched on plasmids.

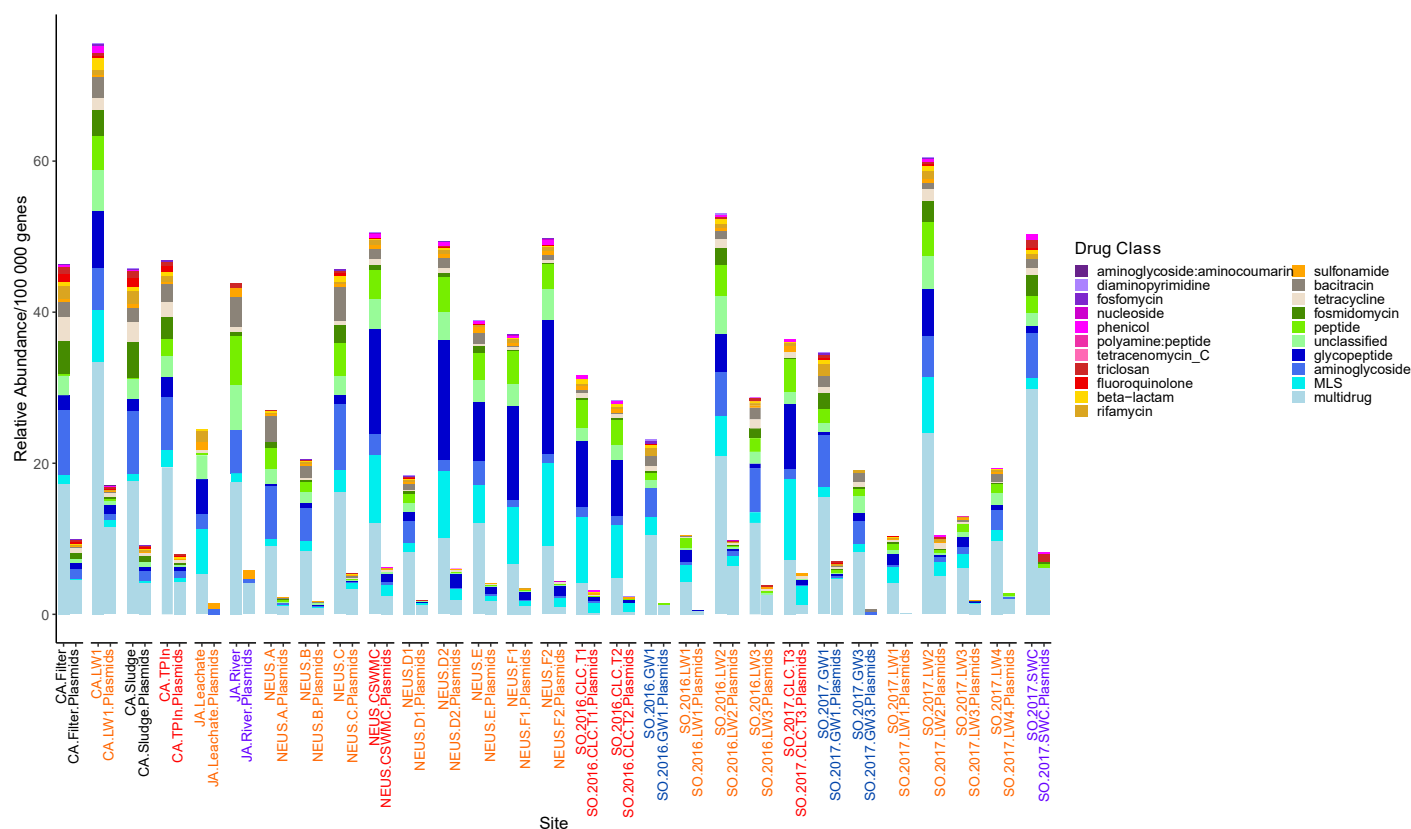


Figure 3.3: Drug class relative abundance per 100,000 genes using results from DeepARG.

The relative abundances of the drug classes within landfills from the full ARG datasets and the plasmid ARG datasets were determined using results from DeepARG. The ggplot package and geom_col function in R were used to create bar charts of the relative abundance of resistance mechanisms per 100,000 genes. Sample types are represented by colour: Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Treatment Plant = black. The landfill sites used were in Southern Ontario (SO), the North Eastern US (NEUS), California (CA), and Jamaica (JA).

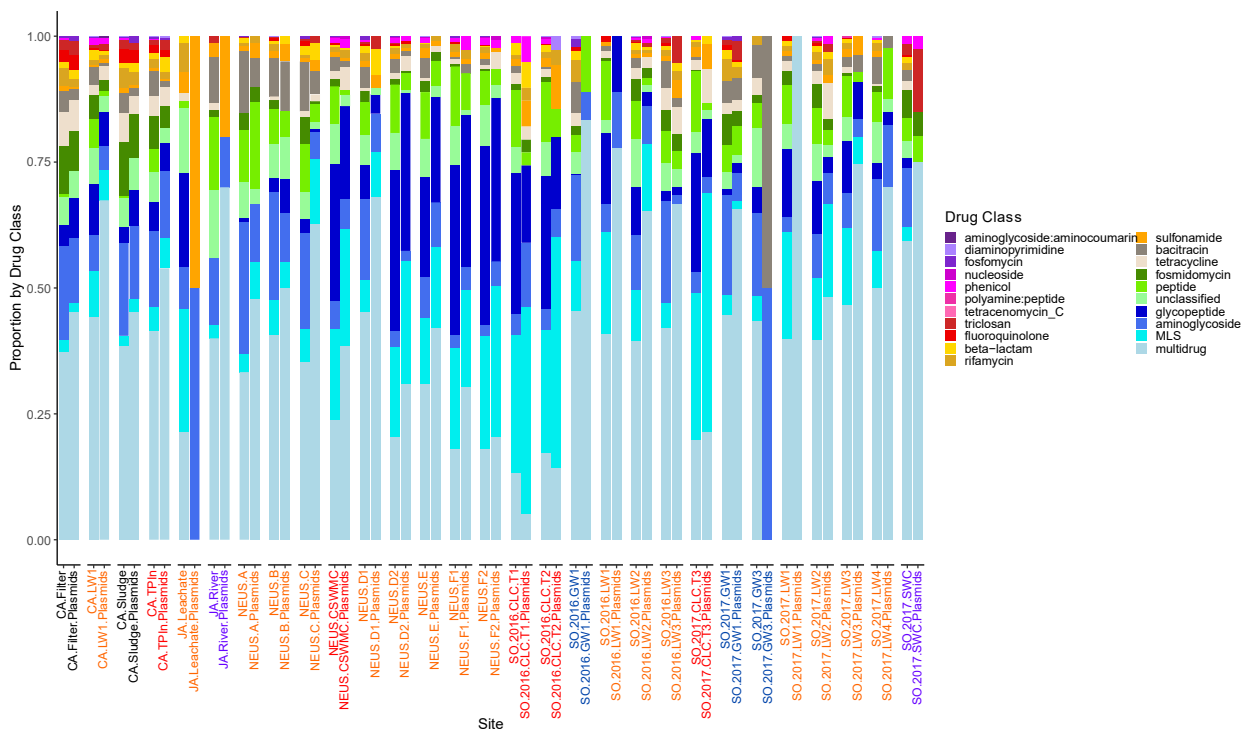


Figure 3.4: Proportional bar chart of drug class relative abundance per 100,000 genes using results from DeepARG

The proportion of relative abundances of the drug classes within landfills from the full ARG datasets and the plasmid ARG datasets were determined using results from DeepARG. The ggplot package and geom_col function in R were used to create bar charts of the relative abundance of resistance mechanisms per 100,000 genes. Sample types are represented by colour: Leachate Well (LW) = orange, Composite Leachate Cistern (CLC) = red, Ground Water (GW) = dark blue, Surface Water (SW) = purple, Treatment Plant = black. The landfill sites used were in Southern Ontario (SO), the North Eastern US (NEUS), California (CA), and Jamaica (JA).

3.4 Discussion

3.4.1 Total Plasmid Counts within Landfills

Sample types present at a landfill help determine the plasmid dynamics within that landfill (Table 3.1; Table 3.2). For example, compared to the rest of the other sample types, TP had the highest prevalence of predicted plasmids. This was reflected at the site level, as the CA site also had the highest prevalence of predicted plasmids. TP samples make up half of the CA landfill samples. Plasmid prevalence was not equal across sample types (4.12 – 8.39%) or sites (3.77 – 7.43%). The abundance of plasmids within a microbial community can be impacted by many biotic and abiotic factors (Smalla et al., 2015). Plasmids rely on host cells, meaning that the diversity of the host cells in a target community influences the presence and abundance of the plasmids within that community (Medaney et al., 2016). The ability of plasmids to persist in a community also increases as the microbial diversity increases in a community (Alonso-del Valle et al., 2021). For this reason, the TP samples, which contain leachate from across the full breadth of the CA landfill site, may have a higher percentage of scaffolds that are plasmids, as these samples encompass a higher diversity microbial community.

3.4.2 Total Antibiotic Resistance Genes (ARGs) Encoded on Plasmids and Plasmid Loads

When looking at the amount of ARGs detected, sample type also influences the ARG abundance within the associated landfills (Table 3.3; Table 3.4). As mentioned in Chapter 2, section 2.5.2, the abundance of ARGs may differ due to numerous factors, including the antibiotics present within an environment, the presence of heavy metals, the exposure of non-antibiotic antimicrobial chemicals, the diversity of the microbial community, and the physical and chemical properties of an environment (Lin et al., 2021). The LWs, CLCs, and TPs are landfill samples associated with leachate and may have higher ARG abundance due to the leachate being

contaminated by heavy metals, pathogenic organisms, and/or pharmaceuticals, which are known to impact ARG abundance (Bandala et al., 2021; Lin et al., 2021; Ozbay et al., 2021). The leachate well samples had the highest count of plasmid-encoded ARGs (Table 3.3). The NEUS landfill had the largest number of ARGs that were characterized as plasmid encoded, a further example of sample type influencing the site characteristics, as eight of the nine NEUS samples are from leachate wells (Table 3.4). The distribution of ARGs that were characterized as plasmid encoded was not equal between all sample types and site locations. Despite LWs and CLCs having higher numbers of ARGs that were characterized as plasmid encoded, these sample types did not have the highest plasmid load (Table 3.4), indicating a disproportionately larger pool of ARGs at these sites, outstripping the higher counts of ARGs associated with plasmids. Similarly, despite the NEUS having the highest number of ARGs that were characterized as plasmid encoded, this site did not have the largest plasmid load (Table 3.4). The TPs samples, and, relatedly, the CA site had both the highest plasmid loads and percentage of plasmids that are carrying ARGs (Table 3.3; Table 3.4). Plasmids typically carry a fitness cost to the host as plasmid genes are often encoding systems for their transmission and maintenance, which may not benefit the host. However, plasmids can carry traits that facilitate bacterial evolution, ultimately increasing the fitness of the host by increasing the number of environments that the host can survive in (Rodríguez-Beltrán et al., 2021). By carrying genes for antibiotic resistance, the plasmid can provide a selective advantage in an environment where antibiotics are an active stressor (San Millan, 2018; Smillie et al., 2010). As different environments may have differing magnitudes of selection pressures, the distribution of ARGs that are characterized as plasmid encoded within a sample will vary. The plasmid load, which measures the mobile proportion of ARGs within an environment, was highest for the CA treatment plant samples, despite these

samples not having the highest total ARG counts or highest number of ARGs found on plasmids. The high plasmid load in the CA treatment plant samples may indicate higher selection pressures enabling the maintenance of ARG-encoding plasmids in these environments. TPs collect all leachate from within the site, meaning that they have multiple inputs carrying different selective pressures. Having a larger proportion of plasmids carrying ARGs may be a result of increased antibiotic input in TPs from all leachate sites, or a reflection of different antibiotic pressures across the site captured in this composite sample. With more plasmid-encoded ARGs within the community, a higher diversity of bacteria would be expected to survive antibiotic stressors (San Millan, 2018; Smillie et al., 2010). The GW and SW samples are non-leachate landfill-associated (NLLA) samples and were expected to experience lower antibiotic pressure. Converse to expectations, these sample types did not have the lowest plasmid load or percentage of plasmids carrying ARGs (Table 3.3). Landfills have physical barriers to prevent the migration of leachate to surrounding water systems (Bandala et al., 2021; Hilger & Barlaz, 2007; Ozbay et al., 2021; Parvin & Tareq, 2021). The higher proportion of plasmid-encoded ARGs in GW and SW samples may indicate an increase in antibiotic residues acting as selective pressures from landfill leachate contamination. However, ground and surface water resistomes may also be impacted by other surrounding anthropogenic activities. Groundwater specifically has long selection periods for ARGs as a result of limited nutrient availability decreasing the degradation rate of antibiotics, meaning that there is a longer period of exposure to selection pressures which can increase the proportion of plasmid-encoded ARGs (Huang et al., 2023; San Millan, 2018; Smillie et al., 2010).

3.4.3 Plasmid-Encoded Antibiotic Resistance Genes (ARGs) within Landfills

Compared to the full metagenomic datasets, the plasmid-encoded ARGs represented a small fraction of ARGs within landfills (Figure 3.1; Figure 3.3). Plasmid-encoded ARGs may be present at a lower prevalence as the presence of a plasmid can introduce a fitness cost on the host, as well as the plasmid may be lost during bacterial cell division. Alternatively, beneficial genes that are introduced from the plasmid may move to the chromosome (San Millan & MacLean, 2017). If positive selection for ARGs is occurring, selection would favour the integration of ARGs into the host chromosome, resulting in the loss of the plasmid (Hall et al., 2016). The resistance mechanisms that were encoded on plasmids at high abundances were antibiotic inactivation, antibiotic efflux, antibiotic target replacement, antibiotic target alteration, and antibiotic target protection (Figure 3.1; Figure 3.2). The plasmid-encoded resistance mechanisms are proportionally similar in relative abundance to the full ARG datasets (Figure 3.2). Resistance mechanisms for antibiotic inactivation and antibiotic efflux are slightly enriched on the plasmids. The plasmid-encoded target drug classes that were seen at high relative abundances were multidrug, MLS, aminoglycoside, and glycopeptides (Figure 3.3 and 3.4). The plasmid-encoded target drug classes were not proportionally similar across all landfill sites (Figure 3.4). As mentioned in Chapter 2, section 2.5.3, common resistance mechanisms to MLS antibiotics are; 1) rRNA methylation of adenine, 2) efflux from efflux pumps with homology to the MFS family or members of the ABC transporter superfamily, and 3) ribosomal mutation of the drug binding domain (Roberts et al., 1999; Z.-Q. Xu et al., 2012). All three resistance mechanisms are seen in the plasmid-encoded resistance mechanisms, as antibiotic target protection, antibiotic efflux, and antibiotic target protection. *Erm* genes, which were seen in the landfill metagenomes, do provide resistance to MLS antibiotics and are carried on plasmids (Z.-Q. Xu et al., 2012). *Erm* genes were seen within the plasmid ARG datasets, supporting the

dissemination of this gene within landfills. Common resistance mechanism to aminoglycosides are 1) aminoglycoside modifying enzymes (AMEs); 2) target modification by mutations or methyltransferases; 3) change of uptake or efflux by the RND and MFS family; and 4) membrane proteases (Becker & Cooper, 2013). Three of those resistance mechanisms were seen in the plasmid-encoded resistance mechanisms as antibiotic inactivation, antibiotic target alteration, and antibiotic efflux. *Aac* genes, which were seen in the full landfill datasets, provide resistance to aminoglycoside antibiotics and are often on MGEs (Becker & Cooper, 2013). *Aac* genes were also seen in the plasmid ARG datasets, supporting dissemination of this gene within landfills. A resistance mechanism to glycopeptide antibiotics is the production of modified peptidoglycan precursors, which is an example of antibiotic target alteration (Arhin et al., 2012). Antibiotic target alteration was one of the plasmid-encoded resistance mechanisms that was present at high abundance. Vancomycin, which is a type of glycopeptide, has demonstrated dissemination through horizontal gene transfer (Singh et al., 2017; Stogios & Savchenko, 2020). Within the plasmid ARG datasets resistance genes for vancomycin resistance were identified, indicating dissemination of this gene within landfills. Overall, distributions of plasmid-encoded resistance mechanisms and target drug classes aligned with distributions of landfill resistance mechanism and target drug classes. The target drug classes for multidrug and MLS are slightly enriched on plasmids.

3.4.4 Conclusions

Ultimately, the PlasClass and PPR-Meta were able to reconstruct plasmids from the landfill metagenomic datasets. Plasmid dynamics within a landfill were determined by the sample types within that landfill. The TP samples, and by association the CA site, had the highest plasmid load and the highest percentage of plasmids that were carrying ARGs indicating higher selection

pressures enabling the maintenance of ARG-encoding plasmids in these environments.

Interestingly, plasmid-encoded ARGs only made up a small fraction of the ARGs within the landfills. When looking at proportions, the plasmid-encoded resistance mechanisms are proportionally similar in relative abundance to the full ARG datasets where as the plasmid encoded drug classes were not proportionally similar. Antibiotic inactivation, antibiotic efflux, antibiotic target replacement, antibiotic target alteration, and antibiotic target protection were encoded on plasmids at high abundances. Multidrug, MLS, aminoglycoside, and glycopeptides were the plasmid-encoded target drug classes that were seen at high abundances. It is important to note that the plasmids identified in this analysis were not able to be linked to hosts, meaning the taxonomic distribution of the plasmid encoded ARGs and the populations that are benefitting from the ARGs were unable to be determined. This would be an interesting avenue of research in the future.

Chapter 4: Conclusions and Future Work

Chapter 2 hypothesized that landfills would have distinct antibiotic resistance gene (ARG) signatures that were enriched in diverse ARG types compared to other environments and we saw that landfills did have ARG signatures consisting of a larger diversity of ARG types compared to other environments. Chapter 3 hypothesized that the ARG diversity within the four landfills would correlate to ARG mobility, indicating strong selection pressures for antibiotic resistance at a given site and we saw that plasmid-encoded ARGs made up a small fraction of the ARGs within the landfills, however proportions of resistance mechanisms and drug classes were similar indicating strong selection pressures from certain classes of antibiotics.

Overall, landfills have distinct resistance mechanism, drug class, and gene family signatures when compared to other environments, with higher diversity of categories across all levels of detection. Antibiotic efflux, antibiotic inactivation, antibiotic target replacement and antibiotic target protection were the most abundant resistance mechanisms detected. Target drug classes that were the most abundant as detected by the RGI were diaminopyrimidine; tetracycline; fluoroquinolone; aminoglycoside; tetracycline; and carbapenem, penam, and cephalosporin. The target drug classes that were the most abundant as detected by DeepARG were multidrug; aminoglycoside; macrolides, lincosamides, streptogramins (MLS); fosmidomycin; and unclassified. Tetracycline; aminoglycoside; MLS; glycopeptide; and peptide antibiotics were detected at high abundance by both programs. The gene families that were the most abundant as detected by the RGI were resistance – nodulation – cell division (RND) antibiotic efflux pump, trimethoprim resistance dihydrofolate reductase, major facilitator superfamily (MFS) antibiotic efflux pump, OXA beta – lactamase, and ABC – F ATP – binding cassette ribosomal protection protein. DeepARG detected *rpoB2*, *oprM*, *vanU*, *ugd*, and

truncated response regulator *arcR* as the most abundant gene families. The gene families *ompR*, *rpoB2*, *vanU*, and *ugd* were detected by both programs. Landfills had a larger diversity of gene families when compared to other environments and distinct gene family signatures.

The treatment plant (TP) and California (CA) samples had the largest number of predicted plasmids. The abundance of plasmids was influenced by the sample types that made up a certain landfill location. Leachate wells (LWs) and samples from the North Eastern Coast of the United States (NEUS) had the largest number of ARGs that were characterized as plasmid originating. However, the TP and CA samples had the largest plasmid load as well as the largest percentage of plasmids that encoded ARGs. The plasmid-encoded ARGs made up a small fraction of the ARGs within the landfill datasets. Plasmids were seen to carry antibiotic inactivation, antibiotic efflux, antibiotic target replacement, antibiotic target alteration, and antibiotic target protection resistance mechanisms. Antibiotic inactivation and antibiotic efflux were enriched on plasmids. Plasmids were also seen to carry resistance to multidrug, MLS, aminoglycoside, and glycopeptide drug classes. Multidrug resistance was enriched on plasmids.

The ARG data, plasmid load, and percentage of plasmids carrying ARGs differed between sample types and landfill locations, likely due to differing environmental inputs for each sample type and site creating different selection pressures. The resistance mechanisms and drug classes present on plasmids were also seen at similar proportions in the full landfill ARG datasets. The *erm* and *aac* genes, which were detected in the full landfill datasets, are commonly seen on plasmids. These genes code for resistance to MLS antibiotics and aminoglycosides which were two types of target drug classes found in the plasmid data. The *erm* and *aac* genes were also detected within the plasmid-encoded ARG datasets.

Our work included using the RGI's 'Loose' algorithm that may detect new ARGs, but phenotypic screening is required to determine if any potential new ARGs are functional (Alcock et al., 2020). Moreover, metal resistance genes and ARGs are often co-selected as a result of similar resistance mechanisms (Vats et al., 2022). In future, it would be interesting to investigate the metal resistance genes present within the landfills or on plasmids to determine if their abundance correlates to the ARGs that they are often co-selected with within these environments.

The research presented in this thesis used metagenomic data for determining ARG diversity and prevalence. Metagenomics can investigate abundance of plasmids and the ARGs that are carried by plasmids. However, the metagenomic techniques employed here did not allow identification of the host that a plasmid is associated with. One technique that can be used to overcome this is using Hi-C sequencing (Smalla et al., 2015). Hi-C sequencing can link the plasmid DNA to parts of chromosomal DNA by cross-linking molecules that are near each other prior to DNA extraction. This method can identify both plasmid and chromosomal DNA, allowing connection of plasmid and host in the downstream analyses (Smalla et al., 2015).

Additionally, this study looked at antibiotic resistance in landfills. The RGI and DeepARG look for antibiotic resistance using metagenomic data based on databases containing bacteria and archaea (Alcock et al., 2023; Arango-Argoty et al., 2018). However, fungal infections contribute to infectious disease-related deaths, and pathogenic fungi are becoming resistance to antifungal treatments, a development which negatively impacts human health (Y. Lee et al., 2021). One interesting question would be if landfills have unique antifungal resistance signatures, could potentially contribute to antifungal resistance spread, or if there is no association. Antifungal resistance in landfills as well as antifungal resistance in general is an

understudied area. To date there are currently no tools to study antifungal resistance using metagenomics, and no study examining antifungal resistance in landfills has been published.

Ultimately, landfill environments are an important hot spot for ARGs. Landfill-specific signatures were observed for resistance mechanisms, drug classes, and gene families. Plasmid-encoded ARGs made up a small fraction of the full landfill ARG datasets, but had similar proportions at the resistance mechanism and drug class levels. Future work could investigate connecting plasmids carrying ARGs to their associated hosts as well as antifungal and metal resistance within landfills.

References

- Aires, J. R., Pechère, J.-C., Van Delden, C., & Köhler, T. (2002). Amino acid residues essential for function of the MexF efflux pump protein of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 46(7), 2169–2173. <https://doi.org/10.1128/AAC.46.7.2169-2173.2002>
- Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., ... McArthur, A. G. (2023). CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 51(D1), D690–D699. <https://doi.org/10.1093/nar/gkac920>
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517–D525. <https://doi.org/10.1093/nar/gkz935>
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11), 1144–1146. <https://doi.org/10.1038/nmeth.3103>
- Alonso-del Valle, A., León-Sampedro, R., Rodríguez-Beltrán, J., DelaFuente, J., Hernández-García, M., Ruiz-Garbajosa, P., Cantón, R., Peña-Miller, R., & San Millán, A. (2021).

- Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. *Nature Communications*, 12(1), 2653. <https://doi.org/10.1038/s41467-021-22849-y>
- Ambrose, S. J., & Hall, R. M. (2021). dfrA trimethoprim resistance genes found in Gram-negative bacteria: Compilation and unambiguous numbering. *The Journal of Antimicrobial Chemotherapy*, 76(11), 2748–2756. <https://doi.org/10.1093/jac/dkab212>
- Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6(1), 23. <https://doi.org/10.1186/s40168-018-0401-z>
- Arhin, F. F., Belley, A., Rafai Far, A., Lehoux, D., Moeck, G., & Parr Jr., T. R. (2012). Glycopeptides and lipoglycopeptides. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 301–346). Springer.
- Armstrong, E. S., Kostrub, C. F., Cass, R. T., Moser, H. E., Serio, A. W., & Miller, G. H. (2012). Aminoglycosides. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 229–269). Springer.
- Bai, L., Tan, Z., Gong, H., Xu, M., Li, Z., Yue, J., Liu, L., Yang, D., & Li, R. (2022). Study on antibiotics, antibiotic resistance genes, bacterial community characteristics and their correlation in the landfill leachates. *Journal of Applied Microbiology*, 132(1), 445–458. <https://doi.org/10.1111/jam.15229>
- Bandala, E. R., Liu, A., Wijesiri, B., Zeidman, A. B., & Goonetilleke, A. (2021). Emerging materials and technologies for landfill leachate treatment: A critical review. *Environmental Pollution*, 291, 118133. <https://doi.org/10.1016/j.envpol.2021.118133>

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Becker, B., & Cooper, M. A. (2013). Aminoglycoside antibiotics in the 21st century. *ACS Chemical Biology*, 8(1), 105–115. <https://doi.org/10.1021/cb3005116>
- Ben, Y., Fu, C., Hu, M., Liu, L., Wong, M. H., & Zheng, C. (2019). Human health risk assessment of antibiotic resistance associated with antibiotic residues in the environment: A review. *Environmental Research*, 169, 483–493. <https://doi.org/10.1016/j.envres.2018.11.040>
- Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13(1), 42–51. <https://doi.org/10.1038/nrmicro3380>
- Bradford, P. A., & Jones, C. H. (2012). Tetracyclines. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 147–179). Springer.
- Brown-Jaque, M., Calero-Cáceres, W., & Muniesa, M. (2015). Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid*, 79, 1–7. <https://doi.org/10.1016/j.plasmid.2015.01.001>
- Bush, K., & Jacoby, G. A. (2010). Updated functional classification of beta-lactamases. *Antimicrobial Agents and Chemotherapy*, 54(3), 969–976. <https://doi.org/10.1128/AAC.01009-09>

- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
- Collins-Fairclough, A. M., Co, R., Ellis, M. C., & Hug, L. A. (2018). Widespread antibiotic, biocide, and metal resistance in microbial communities inhabiting a municipal waste environment and anthropogenically impacted river. *MSphere*, 3(5), 10.1128/msphere.00346-18. <https://doi.org/10.1128/mSphere.00346-18>
- Czatzkowska, M., Wolak, I., Harnisz, M., & Korzeniewska, E. (2022). Impact of anthropogenic activities on the dissemination of ARGs in the environment-a review. *International Journal of Environmental Research and Public Health*, 19(19). <https://doi.org/10.3390/ijerph191912853>
- Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews : MMBR*, 74(3), 417–433. <https://doi.org/10.1128/MMBR.00016-10>
- Deng, Y., Li, B., & Zhang, T. (2018). Bacteria that make a meal of sulfonamide antibiotics: Blind spots and emerging opportunities. *Environmental Science & Technology*, 52(7), 3854–3868. <https://doi.org/10.1021/acs.est.7b06026>
- Depardieu, F., Mejean, V., & Courvalin, P. (2015). Competition between VanUG repressor and VanRG activator leads to rheostatic control of *vanG* vancomycin resistance operon expression. *PLOS Genetics*, 11(4), 1–21. <https://doi.org/10.1371/journal.pgen.1005170>
- Eberhard, W. G. (1990). Evolution in bacterial plasmids and levels of selection. *The Quarterly Review of Biology*, 65(1), 3–22. <https://doi.org/10.1086/416582>

- El Ghachi, M., Bouhss, A., Blanot, D., & Mengin-Lecreulx, D. (2004). The *bacA* gene of *Escherichia coli* encodes an undecaprenyl pyrophosphate phosphatase activity. *The Journal of Biological Chemistry*, 279(29), 30106–30113.
<https://doi.org/10.1074/jbc.M401701200>
- Fang, Z., Tan, J., Wu, S., Li, M., Xu, C., Xie, Z., & Zhu, H. (2019). PPR-Meta: A tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience*, 8(6). <https://doi.org/10.1093/gigascience/giz066>
- Finley, R. L., Collignon, P., Larsson, D. G. J., McEwen, S. A., Li, X.-Z., Gaze, W. H., Reid-Smith, R., Timinouni, M., Graham, D. W., & Topp, E. (2013). The scourge of antibiotic resistance: The important role of the environment. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 57(5), 704–710.
<https://doi.org/10.1093/cid/cit355>
- Flach, C.-F., Boulund, F., Kristiansson, E., & Larsson, D. J. (2013). Functional verification of computationally predicted qnr genes. *Annals of Clinical Microbiology and Antimicrobials*, 12, 34. <https://doi.org/10.1186/1476-0711-12-34>
- Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., & de la Cruz, F. (2004). A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiology Reviews*, 28(1), 79–100. <https://doi.org/10.1016/j.femsre.2003.09.001>
- George, N. A., & Hug, L. A. (2023). CRISPR-resolved virus-host interactions in a municipal landfill include non-specific viruses, hyper-targeted viral populations, and interviral conflicts. *Scientific Reports*, 13(1), 5611. <https://doi.org/10.1038/s41598-023-32078-6>
- Grégoire, D. S., George, N. A., & Hug, L. A. (2023). *Microbial methane cycling in a landfill on a decadal time scale*. <https://doi.org/10.1101/2023.01.20.524919>

- Grossman, T. H. (2016). Tetracycline antibiotics and resistance. *Cold Spring Harbor Perspectives in Medicine*, 6(4), a025387. <https://doi.org/10.1101/cshperspect.a025387>
- Gundran, R. S., Cardenio, P. A., Villanueva, M. A., Sison, F. B., Benigno, C. C., Kreausukon, K., Pichpol, D., & Punyapornwithaya, V. (2019). Prevalence and distribution of *bla*CTX-M, *bla*SHV, *bla*TEM genes in extended- spectrum β - lactamase- producing *E. coli* isolates from broiler farms in the Philippines. *BMC Veterinary Research*, 15(1), 227. <https://doi.org/10.1186/s12917-019-1975-9>
- Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, 9(1), 37. <https://doi.org/10.1186/s40168-020-00990-y>
- Hall, J. P. J., Wood, A. J., Harrison, E., & Brockhurst, M. A. (2016). Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8260–8265. <https://doi.org/10.1073/pnas.1600974113>
- Hancock, R. E., & Chapple, D. S. (1999). Peptide antibiotics. *Antimicrobial Agents and Chemotherapy*, 43(6), 1317–1323. <https://doi.org/10.1128/AAC.43.6.1317>
- Hilger, H. H., & Barlaz, M. A. (2007). Anaerobic decomposition of refuse in landfills and methane oxidation in landfill covers. In *Manual of Environmental Microbiology* (pp. 818–842). John Wiley & Sons, Ltd. <https://doi.org/10.1128/9781555815882.ch67>
- Hooper, D. C., & Jacoby, G. A. (2015). Mechanisms of drug resistance: Quinolone resistance. *Annals of the New York Academy of Sciences*, 1354(1), 12–31. <https://doi.org/10.1111/nyas.12830>

- Huang, F.-Y., Zhao, Y., Neilson, R., Zhou, X.-Y., Li, H., Ding, L., Zhou, S.-Y.-D., & Su, J.-Q. (2023). Antibiotic resistome in groundwater and its association with mountain springs and river. *Ecotoxicology and Environmental Safety*, 252, 114603. <https://doi.org/10.1016/j.ecoenv.2023.114603>
- Huemer, M., Mairpady Shambat, S., Brugger, S. D., & Zinkernagel, A. S. (2020). Antibiotic resistance and persistence-implications for human health and treatment perspectives. *EMBO Reports*, 21(12), e51034. <https://doi.org/10.15252/embr.202051034>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- Ishikawa, J., Chiba, K., Kurita, H., & Satoh, H. (2006). Contribution of *rpoB2* RNA polymerase β subunit gene to rifampin resistance in *Nocardia* species. *Antimicrobial Agents and Chemotherapy*, 50(4), 1342–1346. <https://doi.org/10.1128/aac.50.4.1342-1346.2006>
- Jacoby, G. A. (2009). AmpC beta-lactamases. *Clinical Microbiology Reviews*, 22(1), 161–182, Table of Contents. <https://doi.org/10.1128/CMR.00036-08>
- Jacoby, G. A., & Hooper, D. C. (2012). Review of the quinolone family. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 119–146). Springer.
- Jian, Z., Zeng, L., Xu, T., Sun, S., Yan, S., Yang, L., Huang, Y., Jia, J., & Dou, T. (2021). Antibiotic resistance genes in bacteria: Occurrence, spread, and control. *Journal of Basic Microbiology*, 61(12), 1049–1070. <https://doi.org/10.1002/jobm.202100201>
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. <https://doi.org/10.7717/peerj.7359>

- Karkman, A., Pärnänen, K., & Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nature Communications*, 10(1), 80. <https://doi.org/10.1038/s41467-018-07992-3>
- Ko, D., & Choi, S. H. (2022). Mechanistic understanding of antibiotic resistance mediated by EnvZ/OmpR two-component system in *Salmonella enterica* serovar Enteritidis. *Journal of Antimicrobial Chemotherapy*, 77(9), 2419–2428. <https://doi.org/10.1093/jac/dkac223>
- Köhler, T., Epp, S. F., Curty, L. K., & Pechère, J. C. (1999). Characterization of MexT, the regulator of the MexE-MexF-OprN multidrug efflux system of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 181(20), 6300–6305. <https://doi.org/10.1128/JB.181.20.6300-6305.1999>
- Kraemer, S. A., Barbosa da Costa, N., Oliva, A., Huot, Y., & Walsh, D. A. (2022). A resistome survey across hundreds of freshwater bacterial communities reveals the impacts of veterinary and human antibiotics use. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.995418>
- Lai, S., Zhang, Q., & Jin, L. (2023). Natural and man-made cyclic peptide-based antibiotics. *Antibiotics*, 12(1). <https://doi.org/10.3390/antibiotics12010042>
- Lakhout, A., & Alsulami, B. T. (2020). Evaluation of risk assessment of landfill emissions and their impacts on human health. *Arabian Journal of Geosciences*, 13(22), 1185. <https://doi.org/10.1007/s12517-020-06218-5>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Larsson, D. G. J., Andremont, A., Bengtsson-Palme, J., Brandt, K. K., de Roda Husman, A. M., Fagerstedt, P., Fick, J., Flach, C.-F., Gaze, W. H., Kuroda, M., Kvint, K., Laxminarayan,

- R., Manaia, C. M., Nielsen, K. M., Plant, L., Ploy, M.-C., Segovia, C., Simonet, P., Smalla, K., ... Wernersson, A.-S. (2018). Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environment International*, 117, 132–138. <https://doi.org/10.1016/j.envint.2018.04.041>
- Lee, J. C., Oh, J. Y., Cho, J. W., Park, J. C., Kim, J. M., Seol, S. Y., & Cho, D. T. (2001). The prevalence of trimethoprim-resistance-conferring dihydrofolate reductase genes in urinary isolates of *Escherichia coli* in Korea. *Journal of Antimicrobial Chemotherapy*, 47(5), 599–604. <https://doi.org/10.1093/jac/47.5.599>
- Lee, Y., Puumala, E., Robbins, N., & Cowen, L. E. (2021). Antifungal drug resistance: Molecular mechanisms in *Candida albicans* and beyond. *Chemical Reviews*, 121(6), 3390–3411. <https://doi.org/10.1021/acs.chemrev.0c00199>
- Lerminiaux, N. A., & Cameron, A. D. S. (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian Journal of Microbiology*, 65(1), 34–44. <https://doi.org/10.1139/cjm-2018-0275>
- Li, W., & Wang, A. (2021). Genomic islands mediate environmental adaptation and the spread of antibiotic resistance in multiresistant *Enterococci*—Evidence from genomic sequences. *BMC Microbiology*, 21(1), 55. <https://doi.org/10.1186/s12866-021-02114-4>
- Lin, Z., Yuan, T., Zhou, L., Cheng, S., Qu, X., Lu, P., & Feng, Q. (2021). Impact factors of the accumulation, migration and spread of antibiotic resistance in the environment. *Environmental Geochemistry and Health*, 43(5), 1741–1758. <https://doi.org/10.1007/s10653-020-00759-0>

- Liu, X., Yang, S., Wang, Y., Zhao, H.-P., & Song, L. (2018). Metagenomic analysis of antibiotic resistance genes (ARGs) during refuse decomposition. *Science of The Total Environment*, 634, 1231–1237. <https://doi.org/10.1016/j.scitotenv.2018.04.048>
- Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., Hou, J., Mu, Q., & Alvarez, P. J. J. (2015). Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water Research*, 85, 458–466. <https://doi.org/10.1016/j.watres.2015.09.010>
- Markowitz, V. M., Chen, I.-M. A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., Ratner, A., Jacob, B., Pati, A., Huntemann, M., Liolios, K., Pagani, I., Anderson, I., Mavromatis, K., Ivanova, N. N., & Kyrpides, N. C. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research*, 40(Database issue), D123-129. <https://doi.org/10.1093/nar/gkr975>
- McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva, K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J. V., Spanogiannopoulos, P., ... Wright, G. D. (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*, 57(7), 3348–3357. <https://doi.org/10.1128/AAC.00419-13>
- Medaney, F., Ellis, R. J., & Raymond, B. (2016). Ecological and genetic determinants of plasmid distribution in *Escherichia coli*. *Environmental Microbiology*, 18(11), 4230–4239. <https://doi.org/10.1111/1462-2920.13552>

- Mouslim, C., & Groisman, E. A. (2003). Control of the *Salmonella ugd* gene by three two-component regulatory systems. *Molecular Microbiology*, 47(2), 335–344.
<https://doi.org/10.1046/j.1365-2958.2003.03318.x>
- Nayfach, S., Camargo, A. P., Schulz, F., Eloie-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, 39(5), 578–585. <https://doi.org/10.1038/s41587-020-00774-7>
- Ozbay, G., Jones, M., Gadde, M., Isah, S., & Attarwala, T. (2021). Design and operation of effective landfills with minimal effects on the environment and human health. *Journal of Environmental and Public Health*, 2021, 6921607. <https://doi.org/10.1155/2021/6921607>
- Page, M. G. P. (2012). Beta-lactam antibiotics. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 79–117). Springer.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055.
<https://doi.org/10.1101/gr.186072.114>
- Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4), e00088-17. <https://doi.org/10.1128/CMR.00088-17>
- Parvin, F., & Tareq, S. M. (2021). Impact of landfill leachate contamination on surface and groundwater of Bangladesh: A systematic review and possible public health risks assessment. *Applied Water Science*, 11(6), 100. <https://doi.org/10.1007/s13201-021-01431-3>

- Pellow, D., Mizrahi, I., & Shamir, R. (2020). PlasClass improves plasmid sequence classification. *PLOS Computational Biology*, 16(4), 1–9.
<https://doi.org/10.1371/journal.pcbi.1007781>
- Perez-Molphe-Montoya, E., Küsel, K., & Overholt, W. A. (2022). Redefining the phylogenetic and metabolic diversity of phylum Omnitrophota. *Environmental Microbiology*, 24(11), 5437–5449. <https://doi.org/10.1111/1462-2920.16170>
- Potrykus, J., & Wegrzyn, G. (2001). Chloramphenicol-sensitive *Escherichia coli* strain expressing the chloramphenicol acetyltransferase (<i>cat</i>) gene. *Antimicrobial Agents and Chemotherapy*, 45(12), 3610–3612. <https://doi.org/10.1128/AAC.45.12.3610-3612.2001>
- Public Health Agency of Canada. (2022). *Canadian antimicrobial resistance surveillance report*.
- Rice, E. W., Wang, P., Smith, A. L., & Stadler, L. B. (2020). Determining hosts of antibiotic resistance genes: A review of methodological advances. *Environmental Science & Technology Letters*, 7(5), 282–291. <https://doi.org/10.1021/acs.estlett.0c00202>
- Rima, M., Rima, M., Fajloun, Z., Sabatier, J.-M., Bechinger, B., & Naas, T. (2021). Antimicrobial peptides: A potent alternative to antibiotics. *Antibiotics (Basel, Switzerland)*, 10(9). <https://doi.org/10.3390/antibiotics10091095>
- Roberts, M. C., Sutcliffe, J., Courvalin, P., Jensen, L. B., Rood, J., & Seppala, H. (1999). Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. *Antimicrobial Agents and Chemotherapy*, 43(12), 2823–2830.
<https://doi.org/10.1128/AAC.43.12.2823>
- Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C., & San Millán, Á. (2021). Beyond horizontal gene transfer: The role of plasmids in bacterial evolution.

- Nature Reviews Microbiology*, 19(6), 347–359. <https://doi.org/10.1038/s41579-020-00497-1>
- San Millan, A. (2018). Evolution of plasmid-mediated antibiotic resistance in the clinical context. *Trends in Microbiology*, 26(12), 978–985.
<https://doi.org/10.1016/j.tim.2018.06.007>
- San Millan, A., & MacLean, R. C. (2017). Fitness costs of plasmids: A limit to plasmid transmission. *Microbiology Spectrum*, 5(5).
<https://doi.org/10.1128/microbiolspec.MTBP-0016-2017>
- Sánchez-Osuna, M., Cortés, P., Llagostera, M., Barbé, J., & Erill, I. (2020). Exploration into the origins and mobilization of di-hydrofolate reductase genes and the emergence of clinical resistance to trimethoprim. *Microbial Genomics*, 6(11).
<https://doi.org/10.1099/mgen.0.000440>
- Sauk, A. H., & Hug, L. A. (2022). Substrate-restricted methanogenesis and limited volatile organic compound degradation in highly diverse and heterogeneous municipal landfill microbial communities. *ISME Communications*, 2(1), 58. <https://doi.org/10.1038/s43705-022-00141-4>
- Schnappinger, D., & Hillen, W. (1996). Tetracyclines: Antibiotic action, uptake, and resistance mechanisms. *Archives of Microbiology*, 165(6), 359–369.
<https://doi.org/10.1007/s002030050339>
- Shi, P., Jia, S., Zhang, X.-X., Zhang, T., Cheng, S., & Li, A. (2013). Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water Research*, 47(1), 111–120. <https://doi.org/10.1016/j.watres.2012.09.046>

- Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843.
<https://doi.org/10.1038/s41564-018-0171-1>
- Singh, S. B., Young, K., & Silver, L. L. (2017). What is an “ideal” antibiotic? Discovery challenges and path forward. *Biochemical Pharmacology*, 133, 63–73.
<https://doi.org/10.1016/j.bcp.2017.01.003>
- Smalla, K., Jechalke, S., & Top, E. M. (2015). Plasmid detection, characterization, and ecology. *Microbiology Spectrum*, 3(1), PLAS-0038-2014.
<https://doi.org/10.1128/microbiolspec.PLAS-0038-2014>
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of plasmids. *Microbiology and Molecular Biology Reviews : MMBR*, 74(3), 434–452. <https://doi.org/10.1128/MMBR.00020-10>
- Spasov, E., Tsuji, J. M., Hug, L. A., Doxey, A. C., Sauder, L. A., Parker, W. J., & Neufeld, J. D. (2020). High functional diversity among *Nitrospira* populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant. *The ISME Journal*, 14(7), 1857–1872. <https://doi.org/10.1038/s41396-020-0650-2>
- Stern, A. L., Van der Verren, S. E., Kanchugal P, S., Näsvall, J., Gutiérrez-de-Terán, H., & Selmer, M. (2018). Structural mechanism of AadA, a dual-specificity aminoglycoside adenylyltransferase from *Salmonella enterica*. *The Journal of Biological Chemistry*, 293(29), 11481–11490. <https://doi.org/10.1074/jbc.RA118.003989>

- Stogios, P. J., & Savchenko, A. (2020). Molecular mechanisms of vancomycin resistance. *Protein Science : A Publication of the Protein Society*, 29(3), 654–669.
<https://doi.org/10.1002/pro.3819>
- Tian, R., Ning, D., He, Z., Zhang, P., Spencer, S. J., Gao, S., Shi, W., Wu, L., Zhang, Y., Yang, Y., Adams, B. G., Rocha, A. M., Detienne, B. L., Lowe, K. A., Joyner, D. C., Klingeman, D. M., Arkin, A. P., Fields, M. W., Hazen, T. C., ... Zhou, J. (2020). Small and mighty: Adaptation of superphylum *Patescibacteria* to groundwater environment drives their genome simplicity. *Microbiome*, 8(1), 51. <https://doi.org/10.1186/s40168-020-00825-w>
- Totake, K., Kobayashi, N., & Odaka, T. (1998). Trimethoprim resistance and susceptibility genes in *Staphylococcus epidermidis*. *Microbiology and Immunology*, 42(7), 497–501.
<https://doi.org/10.1111/j.1348-0421.1998.tb02315.x>
- Vats, P., Kaur, U. J., & Rishi, P. (2022). Heavy metal-induced selection and proliferation of antibiotic resistance: A review. *Journal of Applied Microbiology*, 132(6), 4058–4076.
<https://doi.org/10.1111/jam.15492>
- Wang, J.-Y., An, X.-L., Huang, F.-Y., & Su, J.-Q. (2020). Antibiotic resistome in a landfill leachate treatment plant and effluent-receiving river. *Chemosphere*, 242, 125207.
<https://doi.org/10.1016/j.chemosphere.2019.125207>
- Wang, Y., Tang, W., Qiao, J., & Song, L. (2015). Occurrence and prevalence of antibiotic resistance in landfill leachate. *Environmental Science and Pollution Research International*, 22(16), 12525–12533. <https://doi.org/10.1007/s11356-015-4514-7>
- Wang, Y., Zhang, R., Lei, Y., & Song, L. (2022). Antibiotic resistance genes in landfill leachates from seven municipal solid waste landfills: Seasonal variations, hosts, and risk

- assessment. *The Science of the Total Environment*, 853, 158677.
<https://doi.org/10.1016/j.scitotenv.2022.158677>
- Wang, Z., Han, M., Li, E., Liu, X., Wei, H., Yang, C., Lu, S., & Ning, K. (2020). Distribution of antibiotic resistance genes in an agriculturally disturbed lake in China: Their links with microbial communities, antibiotics, and water quality. *Journal of Hazardous Materials*, 393, 122426. <https://doi.org/10.1016/j.jhazmat.2020.122426>
- Warburton, P. J., Amodeo, N., & Roberts, A. P. (2016). Mosaic tetracycline resistance genes encoding ribosomal protection proteins. *The Journal of Antimicrobial Chemotherapy*, 71(12), 3333–3339. <https://doi.org/10.1093/jac/dkw304>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2015). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Xu, S., Qasim, M. Z., Zhang, T., Wang, R., Li, C., & Ge, S. (2021). Diversity, abundance and expression of the antibiotic resistance genes in a Chinese landfill: Effect of deposit age. *Journal of Hazardous Materials*, 417, 126027.
<https://doi.org/10.1016/j.jhazmat.2021.126027>
- Xu, Z.-Q., Flavin, M. T., & Eiznhamer, D. A. (2012). Macrolides and ketolides. In T. J. Dougherty & M. J. Pucci (Eds.), *Antibiotic Discovery and Development* (pp. 181–228). Springer.
- Xuan, J., Feng, W., Wang, J., Wang, R., Zhang, B., Bo, L., Chen, Z.-S., Yang, H., & Sun, L. (2023). Antimicrobial peptides for combating drug-resistant bacterial infections. *Drug Resistance Updates*, 68, 100954. <https://doi.org/10.1016/j.drug.2023.100954>

- Yang, J., Long, H., Hu, Y., Feng, Y., McNally, A., & Zong, Z. (2022). *Klebsiella oxytoca* complex: Update on taxonomy, antimicrobial resistance, and virulence. *Clinical Microbiology Reviews*, 35(1), e0000621. <https://doi.org/10.1128/CMR.00006-21>
- Zhang, R., Yang, S., An, Y., Wang, Y., Lei, Y., & Song, L. (2022). Antibiotics and antibiotic resistance genes in landfills: A review. *The Science of the Total Environment*, 806(Pt 2), 150647. <https://doi.org/10.1016/j.scitotenv.2021.150647>
- Zhao, R., Feng, J., Yin, X., Liu, J., Fu, W., Berendonk, T. U., Zhang, T., Li, X., & Li, B. (2018). Antibiotic resistome in landfill leachate from different cities of China deciphered by metagenomic analysis. *Water Research*, 134, 126–139. <https://doi.org/10.1016/j.watres.2018.01.063>
- Zhou, Y., Fang, J., Davood, Z., Han, J., & Qu, D. (2021). Fitness cost and compensation mechanism of sulfonamide resistance genes (*sul1*, *sul2*, and *sul3*) in *Escherichia coli*. *Environmental Microbiology*, 23(12), 7538–7549. <https://doi.org/10.1111/1462-2920.15783>
- Zou, H.-Y., He, L.-Y., Gao, F.-Z., Zhang, M., Chen, S., Wu, D.-L., Liu, Y.-S., He, L.-X., Bai, H., & Ying, G.-G. (2021). Antibiotic resistance genes in surface water and groundwater from mining affected environments. *The Science of the Total Environment*, 772, 145516. <https://doi.org/10.1016/j.scitotenv.2021.145516>

Appendix A: Supplementary Material: Chapter 2

A.1 Supplementary Data 2.1: Percent variance explained by the first ten axes for landfill metagenome-based ordinations presented in Figures 2.2, 2.5, & 2.8. The percent variance was determined using data from scree plots created using ggplot and the tibble function in R.

Axis	Resistance mechanism RGI	Drug class RGI	Drug class DeepARG	Gene family RGI	Gene family DeepARG
1	25.0	13.8	32.1	10.8	21.4
2	16.0	9.40	20.8	7.80	13.9
3	13.2	8.01	9.18	6.67	9.50
4	11.3	6.53	8.55	6.41	8.66
5	8.60	6.25	6.33	5.75	6.65
6	7.88	5.84	4.71	5.55	4.81
7	6.12	4.89	4.09	5.20	3.98
8	5.05	4.74	3.75	4.77	3.42
9	3.23	4.19	3.36	4.27	3.10
10	2.09	3.66	1.98	3.89	2.78

A.2 Supplementary Data 2.2: Percent variance explained by the first ten axes for landfill and outgroup environment metagenome-based ordinations presented in Figures 2.2, 2.5, & 2.8. The percent variance was determined using data from scree plots created using ggplot and the tibble function in R.

Axis	Resistance mechanism RGI	Drug class RGI	Drug class DeepARG	Gene family RGI	Gene family DeepARG
1	22.4	9.91	21.2	8.57	16.0
2	14.9	5.90	15.1	4.62	10.7
3	12.2	5.81	11.3	3.88	7.88
4	9.54	5.01	8.39	3.53	5.57
5	8.69	4.54	7.66	3.36	5.19
6	7.52	4.00	6.96	3.26	4.74
7	6.86	3.83	5.67	3.16	3.83
8	5.15	3.53	4.93	3.09	3.64
9	4.47	3.08	4.39	3.04	3.40
10	4.17	2.88	3.18	2.93	3.03

Normalized ARG counts per 100,000 genes for individual metagenomes and full size heatmaps are presented as Supplementary files, which are available on the Open Science Framework at https://osf.io/myg8q/?view_only=4a44d89e2044411dba79761a3f3261a5.