

Exploring microbial diversity across a Southern Ontario landfill

by

Alexandra Sauk

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, 2019

©Alexandra Sauk 2019

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this 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.

Abstract

Sanitary landfills are highly engineered environments that receive a heterogeneous mixture of organic waste, metals, and plastics. Global waste production continues to grow every year and waste management is an increasing environmental and financial concern for municipalities. Over the last 50 years, many municipalities have improved recycling efforts and hazardous waste disposal to limit inputs to landfills; however, landfills still contain and receive a number of hard to degrade and/or dangerous materials, including heavy metals and volatile compounds. Conventional landfills are designed to entomb municipal solid waste and prevent its degradation by microorganisms. Despite this engineering goal, waste degradation in landfills via aerobic and anaerobic decomposition by microorganisms actively reshapes the municipal solid waste over time and must be accounted for in landfill design. Our depth of knowledge on the diversity of landfill microorganisms and how this microbial diversity changes across and between landfills is limited. Much of the current research into landfill microbial diversity has investigated specific groups with known functions, such as cellulose degraders, methanotrophs, and methanogens. Recently, research groups have taken a community-based approach to studying landfill microbes, relating community composition to environmental and chemical parameters. Many of these previous studies rely solely on 16S rRNA gene amplicon sequencing for taxonomic identification, which limits functional predictions to what is already known about related groups. Understanding landfill microbial communities and their functions is important for informing waste management practices, and has the potential to reveal novel degradation metabolisms that can be used in waste remediation.

This research combines 16S rRNA gene amplicon and metagenomic sequencing to examine the taxonomic and functional diversity of a Southern Ontario landfill and to relate this microbial diversity with site geochemistry. Eight samples were collected from a municipal landfill in 2016 via filtration of the liquid from three leachate wells, two temporal samples of a leachate-collecting cistern (one size-filtered in two fractions), and two groundwater wells from the adjacent aquifer. The DNA for all eight samples underwent 16S rRNA gene amplicon sequencing by the U.S. Department of Energy's Joint Genome Institute. In parallel, total community DNA was extracted and then shotgun sequenced by the U.S. Department of Energy's Joint Genome Institute for six samples, generating assembled and annotated metagenomes. Maximum likelihood phylogenies were inferred from the metagenome-derived 16S rRNA genes and a concatenated alignment of 16 syntenic ribosomal proteins co-located on single scaffolds. Based on a comparison between taxonomy derived from the two phylogenies,

high quality metagenome assembled genomes (MAGs), and the 16S rRNA gene amplicon data, 22 bacterial and 2 archaeal phyla were present at >1% relative abundance within at least one landfill sample. The Patescibacteria, Bacteroidota, Firmicutes, and Proteobacteria had the highest relative abundances, whereas most other phyla were present at low abundance, with some fluctuations among sites. Below the phylum level, very few microorganisms were identified across multiple sites, with only 37 of 2989 populations (represented by 16S rRNA exact sequence variants) present in two or more sites. This indicates that, although phylum-level signatures are conserved, individual landfill populations vary widely both spatially and temporally.

Three leachate and two groundwater sites had partial or complete volatile and non-volatile compound concentration data, allowing for limited correlation of geochemical conditions and contaminant concentrations to microbial diversity patterns across the landfill. Significant differences in geochemistry exist across the sites, with calcium, iron, magnesium, boron, meta and para xylenes, ortho xylenes, and ethylbenzene concentrations contributing most strongly to site differences. The genera *Sulfurovum*, *Proteiniphilum*, and *Ferritrophicum* are relatively abundant in the amplicon data, and have predicted roles in nutrient and contaminant cycling in the landfill related to benzene, proteins, and iron. Access to more sites with complete geochemistry data is required to allow direct connections to be drawn between microbial populations and site geochemistry.

A phylogenetic and metabolic analysis was conducted examining the Tenericutes and Erysipelotrichia, which are related radiations dominated by parasitic and host-associated organisms. Landfill metagenomes allowed reconstruction of 12 Tenericute MAGs distributed within four orders, as well as five MAGs within the order RFN20 that are sister to the Erysipelotrichales. The landfill Tenericutes exhibited small genome sizes below 2 Mb, which is expected for this group, and encode a range of biosynthetic capabilities. Genome features suggest potential lifestyles: free-living, commensal, or parasitic, with the potentially parasitic bacteria showing a greater loss of amino acid pathways. The RFN20 MAGs' genome sizes are below 3 Mb, closer to the Tenericute than the Erysipelotrichia reference genomes. One RFN20 MAG showed similar metabolic capacities to *Dielma fastidious*, an Erysipelotrichia species, whereas the other RFN20 MAGs are predicted to be metabolically more similar to the Tenericutes. The reconstruction and analysis of these Tenericute and RFN20 genomes expands our current knowledge of these abundant groups' potential lifestyles in the landfill, and expands genomic representation of these understudied orders.

Acknowledgements

I would like to thank everyone who has supported me throughout this degree. Thank you to my supervisor Laura Hug for your support and guidance, and for introducing a wildlife biologist to wonderful world of microbes. Thank you to my committee members Barb Butler and Josh Neufeld for your insight and advice.

Thank you to the members of the Hug Lab for all your help and support. I especially want to thank Rebecca Co and Veronica Viljakainen for all your help and answering my endless questions about wet lab work.

Thank you to Joanna Blaszcak at Flathead Lake Biological Station at the University of Montana for your advice and sharing your code for principle component analysis. It made figuring out how to conduct my analysis much easier.

Thank you to Curtis Preiss for your support and motivational chats. Thank you for reading my drafts and lending me your map making skills and knowledge.

Thank you to my dog, Teddy, for forcing me to play and go for walks. My mental and physical health would not be the same without you.

Thank you to my dad, Andreas Sauk, for all your support and encouragement. I would not have got this far without you.

Table of Contents

AUTHOR'S DECLARATION.....	ii
Abstract	iii
Acknowledgements	v
List of Figures	1
List of Tables.....	3
List of Abbreviations.....	4
Chapter 1 Introduction	6
1.1 Microbial diversity	6
1.1.1 Technological advances identify the uncultured majority	6
1.2 Microbes in contaminated sites	9
1.3 Municipal Solid Waste Disposal.....	10
1.3.1 Municipal solid waste.....	10
1.3.2 Landfill disposal options for MSW	11
1.4 Landfill microbial communities	12
1.5 A case study on Tenericutes.....	15
1.6 Study site.....	16
1.7 Scope of research and research objectives	16
Chapter 2 Microbial diversity in a Southern Ontario Landfill.....	18
2.1 Introduction	18
2.2 Methods.....	21
2.2.1 Sample collection	21
2.2.2 Non-volatile and volatile compound measurements	21
2.2.3 Metagenomic sequencing.....	21
2.2.4 16S rRNA gene amplicon sequencing	22
2.2.5 QIIME2 analysis	22
2.2.6 Phylogenetics	23
2.2.7 Binning	24
2.2.8 Diversity analysis	24
2.2.9 Chemical data analysis.....	25
2.3 Results and discussion.....	26
2.3.1 Phylum level diversity.....	26

2.3.2 Alpha diversity.....	30
2.3.3 Beta diversity	33
2.3.4 Diversity of exact sequence variants (ESVs).....	35
2.3.5 Temporal diversity	39
2.3.6 Microbial diversity of the groundwater wells.....	41
2.3.7 Analysis of geochemical parameters	43
2.4 Conclusions.....	46
Chapter 3 A phylogenetic and metabolic study of the group Tenericutes.....	48
3.1 Introduction.....	48
3.1.1 A brief history of the Tenericutes and Erysipelotrichia.....	48
3.1.2 The Tenericute cell	52
3.1.3 Two Tenericutes clades	53
3.2 Methods	55
3.2.1 Data collection	55
3.2.2 Metagenome assembled genomes.....	55
3.2.3 Tenericute and Erysipelotrichales phylogenetic tree.....	55
3.2.4 Genome size vs. gene number	56
3.2.5 Metabolic analyses.....	57
3.2.6 Pangenome.....	57
3.2.7 Cultivation trials	58
3.2.8 Sample preparation for 16S rRNA gene sequencing	60
3.2.9 Identification of isolates from 16S rRNA gene sequences	61
3.3 Results and discussion	61
3.3.1 Metagenome assembled genomes.....	61
3.3.2 Phylogeny	63
3.3.3 Genome size vs gene number	66
3.3.4 Metabolic analyses.....	70
3.3.5 Pangenome.....	74
3.3.6 Cultivation trials	77
3.4 Conclusions.....	79
Chapter 4 Conclusions and Future Directions	81
Bibliography	85

Appendix A Synthetic Leachate.....97

List of Figures

Figure 2.1: Map of landfill sampling locations at the Southern Ontario landfill.	19
Figure 2.2: Relative abundance of bacterial and archaeal phyla present at greater than 1% abundance in at least one sample site.	27
Figure 2.3: Maximum likelihood phylogenetic trees derived from metagenome sequences.	29
Figure 2.4: Observed diversity and Shannon index for the eight samples, grouped by sample type on the x axis.	31
Figure 2.5: Alpha diversity analyses by sample type for Faith’s Phylogenetic Diversity and Pielou’s Evenness.	32
Figure 2.6: PCoA using unweighted (A) and weighted (B) UniFrac distances based on 16S rRNA amplicon ESVs for all sites.	34
Figure 2.7: Prevalence of 16S rRNA amplicon ESVs across sampling sites.	36
Figure 2.8: Abundance and prevalence of 16S rRNA gene amplicon ESVs present in two or more sampling sites.	37
Figure 2.9: Top 10% most abundant 16S rRNA gene amplicon ESVs by count.	38
Figure 2.10: Principal component analysis for 16S rRNA gene amplicon ESVs present in two or more sites. ESV count data was Hellinger transformed.	39
Figure 2.11: Composite Leachate Cistern relative abundance of bacterial and archaeal phyla present at greater than 1% abundance based on 16S rRNA gene amplicons.	40
Figure 2.12: Environmental variation between landfill and aquifer sites for non-volatiles (mg/L) and volatiles (µg/L).	42
Figure 3.1: Maximum-likelihood phylogenies of the Erysipelotrichia and Tenericutes using 16S (left) and 23S (middle) rRNA genes and 34 universal ribosomal proteins (right), taken from Davis et al. (2013).	50
Figure 3.2: Annotree (Mendler et al., 2018) view of class Bacilli based on the GTDB taxonomy (Parks et al., 2018).	51
Figure 3.3: Maximum likelihood phylogeny of the Tenericutes using 45 conserved ribosomal proteins, taken from Gupta et al. (2018).	52
Figure 3.4: Maximum likelihood phylogenetic tree inferred from a concatenated alignment of 16 syntenic ribosomal proteins for the landfill MAGs and references within the Tenericutes and Erysipelotrichales.	65
Figure 3.5: Gene number versus genome size for all landfill MAGs and reference Tenericutes and Erysipelotrichia (A) and for genome sizes less than 3 Mb (B), coloured by lifestyle.	68

Figure 3.6: Gene number versus genome size for all landfill MAGs and reference <i>Tenericutes</i> and <i>Erysipelotrichia</i> (A) and for genome sizes less than 3 Mb (B), coloured by taxonomy.	69
Figure 3.7: Heatmap of metabolism pathways for the landfill MAGs and representative reference genomes for the orders of interest.	73
Figure 3.8: Pangenome analysis of the 17 landfill metagenome assembled genomes showing gene cluster presence/absence and shared gene cluster profiles.	76

List of Tables

Table 2.1: Environmental variable measurements averaged from April and October 2016 values for the two groundwater wells and the three leachate wells.....	20
Table 2.2: The number of metagenome-derived scaffolds per metagenome present on the concatenated ribosomal protein (rp16) and the 16S rRNA gene phylogenetic trees.....	30
Table 2.3: Kruskal-Wallis test results for Faith’s phylogenetic diversity (PD) and Pielou’s evenness (E) diversity metrics.	33
Table 2.4: ANOVA results for all volatile compound concentrations from October and April of 2016 for the three leachate wells and the two groundwater wells.	44
Table 2.5: ANOVA results for non-volatile compound concentrations from October and April of 2016 for the three leachate wells and the unimpacted groundwater well.	44
Table 3.1: Growth media treatments for <i>Tenericutes</i> cultivation trials.	59
Table 3.2: Genome statistics and assigned taxonomy for <i>Tenericutes</i> and <i>Erysipelotrichales</i> MAGs from the landfill metagenomes.	63
Table 3.3: Linear regression statistics for genome size vs. gene number as depicted in Figures 3.5 and 3.6.	67
Table 3.4: Cultivation trial isolate taxonomic affiliations based on 16S rRNA gene comparisons to the SILVA database (release 132) and their morphologies.	78
Table 3.5: Isolates with representation in the metagenome sequences, based on BLAST of isolate 16S rRNA genes to assembled metagenomes. A threshold of 1×10^{-20} was used to filter BLAST hits.....	79
Table 3.6: Isolates with representation in the 16S rRNA gene amplicon exact sequence variants.	79

List of Abbreviations

%	percent
°C	degrees Celsius
16rps	concatenation of 16 syntenic ribosomal proteins
ANOVA	analysis of variance
BLAST	Basic Local Alignment Search Tool
bp	base pair
CLC	composite leachate cistern
CLC_T1	composite leachate cistern, time point 1
CLC_T1_0.1	composite leachate cistern, time point 1, 0.1 μm filter size
CLC_T1_0.2	composite leachate cistern, time point 1, 0.2 μm filter size
CLC_T2	composite leachate cistern, time point 2
CLC_T2_0.1	composite leachate cistern, time point 2, 0.1 μm filter size
df	degrees of freedom
dH ₂ O	deionized water
DNA	deoxyribonucleic acid
DOE-JGI	US Department of Energy Joint Genome Institute
e	e-value
ESV	exact sequence variant
FA	formaldehyde assimilation
FAB	fatty acid biosynthesis
g	gram
GC	guanine-cytosine content
GTDB	Genome Taxonomy Database
GW1	groundwater well 1, impacted
GW2	groundwater well 2, unimpacted
HCl	hydrochloric acid
IB	isoprenoid biosynthesis
JGI	Joint Genome Institute
JGI IMG	Joint Genome Institute Integrated Microbial Genomes and Microbiomes
kb	kilobase pair
L	liter
LW1	leachate well 1
LW2	leachate well 2
LW3	leachate well 3
m	meter

MAG	Metagenome assembled genome
Mb	megabase pair
mg	milligram
mL	milliliter
MSW	municipal solid waste
m- & p-xylenes	meta and para xylenes
NADH	nicotinamide adenine dinucleotide hydride
NCBI	National Centre for Biotechnology Information
nm	nanometer
NSB	nucleotide sugar biosynthesis
OTU	operational taxonomic unit
o- xylenes	ortho xylenes
PCA	Principle component analysis
PCoA	Principle coordinate analysis
PCR	polymerase chain reaction
PE	phosphatidylethanolamine
pH	power of hydrogen
PRPP	5-phospho- α -D-ribosyl 1-pyrophosphate
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RRPC	reductive pentose phosphate cycle
SCR 1	Semi Core Region 1
SCR 2	Semi Core Region 2
str.	Strain
TAE	tris-acetate EDTA
<i>Taq</i>	<i>Thermus aquaticus</i>
TCA	tricarboxylic acid
TCAG	The Centre for Applied Genomics
TDS	total dissolved solids
TKD	total Kiejdahl nitrogen
v.	version
w/v	weight/volume
μ g	microgram
μ L	microliter
μ M	micromolar
μ m	micrometer

Chapter 1

Introduction

1.1 Microbial diversity

Over time, and through increased and improved sampling efforts, we have come to understand the ubiquity of microbial life on Earth. Microbes have evolved to survive in every niche on the planet over billions of years of evolution (Gibbons and Gilbert, 2015). Against a background of on-going debates over the taxonomic organization of microbes, the number of described and approved species continues to increase, with estimates for the number of microbial species reaching as high as one trillion (Locey and Lennon, 2016; Parks *et al.*, 2018; Vitorino and Bessa, 2018). High throughput sequencing has expanded our ability to investigate different habitats and has led to the discovery of microbes in areas once thought inhospitable to life (Savage *et al.*, 2016). Microorganisms live in and on humans and other animals (Kowarsky *et al.*, 2017; Reese *et al.*, 2018; Ross *et al.*, 2017); in soil (Terrat *et al.*, 2017); in bodies of water, including freshwater, marine, saline, and hot springs (Fuhrman *et al.*, 2015; Henson *et al.*, 2018; Schuler *et al.*, 2017; Vavourakis *et al.*, 2016); in glaciers (Lutz *et al.*, 2017); and even in the atmosphere over Antarctica (Pearce *et al.*, 2009). Microbes also live in built and engineered environments, such as building interiors, roads, and human and animal waste treatment sites (King, 2014), as well as environments that have been contaminated by human activities, such as mine tailings (Baker and Banfield, 2003) and ocean water impacted by oil spills (Hazen *et al.*, 2010). To survive in these diverse habitats, microbes use diverse metabolisms. Bacteria and archaea produce and consume a variety of metabolites, from carbohydrates to proteins and lipids using a wide variety of metabolic pathways (Gibbons and Gilbert, 2015). This leads to complex microbial communities, with cross feeding microorganisms able to take advantage of all available niches (Gibbons and Gilbert, 2015). Although much has been discovered about the diversity of microbial life taxonomically, geographically, and functionally, new phyla on the tree of life and environments supporting microbial life continue to be identified, leading to new questions about the ecology and evolution of microbes (Hug *et al.*, 2016; Koskella and Vos, 2015; Rappé and Giovannoni, 2003).

1.1.1 Technological advances identify the uncultured majority

We interact with microorganisms throughout our daily lives, whether they are commensal to us, potential pathogens in our environment, or even probiotics in our yogurt. However, we have not

always been so aware of our microbial neighbours. Antonie van Leeuwenhoek first described bacteria nearly 300 years ago, but, since that time, much of the microbial world has remained inaccessible and underestimated (Gibbons and Gilbert, 2015; Keller and Zengler, 2004; Vitorino and Bessa, 2018). Our knowledge and understanding of microorganisms were frequently limited to what we could see and what we could culture (Keller and Zengler, 2004). The difficulty or even inability to culture some microorganisms in laboratories resulted in cultured bacteria representing less than 0.01% of total microbial life, and it is estimated that less than 1% of microbes are culturable even using current techniques (Leung and Lee, 2016; Vitorino and Bessa, 2018). We do not fully understand the metabolic requirements and environmental conditions necessary for many microbes, such that we cannot replicate their optimal conditions for growth. This is particularly true for microorganisms that are restricted geographically, like cyanobacteria in hot springs (Papke *et al.*, 2003), and those with unique or extreme characteristics (Koskella and Voss, 2015; Vitorino and Bessa, 2018). Because we cannot culture the vast majority of bacteria and archaea, alternative approaches are required to better capture the diversity of microbial life in the environment.

Major technological advancements came in the form of DNA sequencing and, later, metagenomics. In the last 30 years, continued progress in sequencing has allowed for an exponential increase in our understanding of the diversity and ecology of bacteria, archaea, fungi, and viruses (Leung and Lee, 2016; Scholz *et al.*, 2012; Tyson *et al.*, 2004). The use of marker genes allowed for the identification and classification of organisms independent of morphology (Janda and Abbott, 2007; Srinivasan *et al.*, 2015). The 16S rRNA gene was chosen as a marker gene because of its universal presence in all bacteria and archaea, and its highly conserved regions (Janda and Abbott, 2007; Srinivasan *et al.*, 2015). The conserved regions of the 16S rRNA gene allow for the development of primers suitable for Bacteria and Archaea, whereas the variable regions of the gene allow for classification of different organisms (Větrovský and Baldrian, 2013). Indeed, 16S rRNA gene amplicon sequencing is a valuable tool for examining the diversity of complex environmental communities, including soil bacterial communities (Will *et al.*, 2010) and the rare biosphere of the deep sea (Lynch and Neufeld, 2015; Sogin *et al.*, 2006). Although 16S rRNA gene sequencing is a powerful tool for studying microbial communities, it is limited to taxonomy and provides limited insight into the functional potential of the microorganisms (Bareither *et al.*, 2013). Further, the variation in 16S rRNA gene copy number between different taxa and the sequence diversity between those copies can lead to imperfect relative abundances and taxonomic assignments at certain ranks (Větrovský and Baldrian, 2013). Thus, although 16S rRNA gene amplicon sequencing is an important

and useful tool for exploring taxonomic diversity of microbial communities, more phylogenetic information and functional insight can be gained by using metagenomic approaches.

Metagenomics, a relatively recent technological advance in DNA sequencing applications, offers an opportunity to deepen our understanding of microbial community structure across a wide range of ecosystems (Cardinali-Rezende *et al.*, 2016; Leung and Lee, 2016). Metagenomics involves shotgun sequencing of community DNA, which is a more direct and less biased technique for sampling microbial communities, compared to 16S rRNA gene amplicon sequencing (von Mering *et al.*, 2007). In metagenomics, there is no community manipulation or PCR amplification step, removing the chance of primer mismatches and increasing detection of deeper-branching organisms, such as the Patescibacteria (Brown *et al.*, 2015; von Mering *et al.*, 2007). If shotgun sequencing is performed to a sufficient level of depth, metagenomics offers access to the total microbial community (Hedlund *et al.*, 2014). Metagenomic datasets require substantial bioinformatics processing to annotate the large number of genomic contigs or scaffolds produced during sequencing and assembly, and to interpret the phylogeny of those sequences (Hedlund *et al.*, 2014). Assembled scaffolds can be binned into metagenome assembled genomes (MAGs) representing a single population within the original environment, with MAG reconstruction quality defined by the level of completeness (the presence of core bacterial or archaeal genes) and redundancy (multiple copies of core genes) (Eren *et al.*, 2015; Hedlund *et al.*, 2014; Venter *et al.*, 2004). MAGs are reconstructed by clustering contigs or scaffolds based on nucleotide composition (including GC content, codon usage, tetranucleotide frequency), read depth, read depth across samples, and taxonomy/homology (Venter *et al.*, 2004; Hedlund *et al.*, 2014). These MAGs can provide both taxonomic and functional potential information about the organisms, as the reconstructed genomes include phylogenetic marker genes as well as the suite of functional genes encoded by that population's genetic material.

Environmental samples interrogated with metagenomics can greatly expand the genetic information available for uncultivated organisms and, coupled with bioinformatic tools, allow us to determine the functional potential of the organisms present in the original environment. Metagenomics has generated a remarkable expansion of known bacterial lineages over the last decade, including identification of the uncultivated members of the candidate phyla radiation/Patescibacteria, which may make up >15% of the bacterial domain (Brown *et al.*, 2015; Parks *et al.*, 2018). The ability to reconstruct bacterial genomes with metagenomics allows us to uncover who is present in the environment and hypothesize potential microbial community

interactions. This has clarified interactions driving geochemical cycles, for instance such as those present in aquifer sediments and groundwater where the microbial consortia rely on metabolic handoffs to survive (Anantharaman *et al.*, 2016). Metagenomics has also expanded our understanding of the Archaea, with the description of previously unknown members of the Asgard superphylum and their functions, including the Helarchaeota, who are capable of anaerobic hydrocarbon cycling (Seitz *et al.*, 2019). The advent of metagenomics technology has allowed us to make great strides in discovering the uncultured majority for both bacteria and archaea, and, as we continue to sample more environments may yet reveal more lineages on the tree of life (Hug *et al.*, 2016).

1.2 Microbes in contaminated sites

Many environments have been affected by complex mixtures of organic pollutants and heavy metals, but we do not fully understand the effects of these compounds on microbial communities (Epelde *et al.*, 2015; Konopka *et al.*, 1999; Venterino *et al.*, 2018). Pollutants can come from urban, industrial, and agricultural activities as well as illegal dumping (Konopka *et al.*, 1999). Microbial bioremediation is a promising option for some contaminants, but if the pollutants are also toxic to the native microbial community, the process of contaminant transformation is slowed or prevented (Konopka *et al.*, 1999). Environmental stressors - heavy metals in particular - can reduce microbial diversity and bioactivity for years after the initial contamination event, resulting in a shift in microbial populations (Konopka *et al.*, 1999; Venterino *et al.*, 2018). These changes or losses to microbial biodiversity can be compensated for by migration of bacteria into the site, local adaptation, and horizontal transfer of resistance mechanisms (Epelde *et al.*, 2015). We can detect changes in microbial community structure and the functional potential of the microbes, including whether any are capable of degrading pollutants or if the geochemistry of the site affects community structure (Epelde *et al.*, 2015; Konopka *et al.*, 1999). A greater understanding of the effects of pollutants on microbial communities would allow us to assess the health and recovery of contaminated sites and identify bacteria or archaea that can transform contaminants and may be used for bioremediation purposes.

Microbes have been found thriving in environments that are contaminated by human standards. Some microbes have evolved the ability to use pollutants as carbon sources whereas others have evolved resistance mechanisms for toxic metals, such as permeability barriers and enzymatic detoxification pathways (Epelde *et al.*, 2015; Venterino *et al.*, 2018). This is exemplified by microbes in mining sites who are able to survive in the acid mine drainage produced by other species catalyzing

iron and oxidizing sulfur (Baker and Banfield, 2003; Ferris *et al.*, 1989). Similarly, although arsenic is toxic to humans, microorganisms have been found thriving in arsenic rich aquifers and using the arsenic as part of their metabolisms (Oremland and Stolz, 2005). Oil spills are generally regarded as ecological disasters for both marine and coastal species, but after the Deepwater Horizon blowout, petroleum-degrading proteobacteria were stimulated in the environment (Hazen *et al.*, 2010). These discoveries show that some microbes are able to capitalize on contamination and use pollutants to further their survival. Still, contaminated sites typically show lower biodiversity than uncontaminated sites. Groundwater wells that have been impacted by leachate leaks from landfills show lower diversity closer to the leak site (Lu *et al.*, 2012). As well, soils contaminated with heavy metals, such as mercury and cadmium, show a decrease in microbial community diversity (Müller *et al.*, 2002; Xie *et al.*, 2016). Just as microbes have evolved adaptations to enable them to live in a variety of environments, some microbes have also evolved mechanisms for thriving in contaminated sites that remain inhospitable to other life forms, including members of the native microbial community.

1.3 Municipal Solid Waste Disposal

1.3.1 Municipal solid waste

Municipal solid waste (MSW) is a growing concern for municipalities and countries across the globe, both in terms of its environmental consequences and the monetary costs of its storage and management. MSW includes organics such as food and yard waste, paper, plastics, and metals (Hoornweg and Bhada-Tata, 2012). The exact composition and quantity of the waste varies temporally and geographically, within and between countries, and becomes more complex as countries develop (Idris *et al.*, 2004). Financially, MSW is often one of the greatest expenses for municipalities, and regional abilities to process the increasing volumes and the ever-changing types of waste varies greatly (Hoornweg *et al.*, 2013; Idris *et al.*, 2004). MSW generation has increased exponentially with rising populations, increased development, and urbanization (Hoornweg *et al.*, 2013; Idris *et al.*, 2004). It is estimated that by 2025 the global annual production of waste will reach 2.2 billion tons, and is not predicted to peak in this century, reaching 4 billion tons annually in 2100 (Hoornweg and Bhada-Tata, 2012; Hoornweg *et al.*, 2013). These increases in waste production are not sustainable for municipalities or the environment. Environmentally, landfills contributed 12% of annual global methane emissions, making them a focus area for mitigating climate change (Broun and Sattler, 2016; Hoornweg and Bhada-Tata, 2012).

1.3.2 Landfill disposal options for MSW

The waste that humans generate must be managed, and methods of disposal vary substantially. Although landfills are the most common end point for MSW in many countries, including Canada, the United States, and China, disposal sites for MSW range from open dumping to sanitary landfills (Bareither *et al.*, 2013; Idris *et al.*, 2004; Köchling *et al.*, 2015; Xu *et al.*, 2017). Sanitary landfills are superior to open dumping for their reduced ecological impact and increased sustainability (Köchling *et al.*, 2015). Open dumps do not have safeguards, such as cover soil and in-ground liners, to control waste degradation and environmental contamination (Indris *et al.*, 2004). The exposed nature of open dumps also puts them at risk of above-ground fires, further increasing air pollution and safety concerns (Idris *et al.*, 2004). Sanitary landfills, on the other hand, are designed to prevent environmental contamination through the use of cover soils, waste containment in cells, and collection and treatment of leachate, which is liquid that accumulates in landfills (Idris *et al.*, 2004; Köchling *et al.*, 2015). The disposal options ranging from open dumps to sanitary landfills can have various levels of environmental protection and waste containment and can aim to prevent or promote aerobic or anaerobic microbial metabolisms (Indris *et al.*, 2004). Within the available range of waste disposal site engineering, sanitary landfills are the most environmentally friendly.

Sanitary landfills are highly engineered built environments that can be designed as conventional landfills or bioreactor landfills (Broun and Sattler, 2016). Conventional landfills minimize the amount of moisture entering the waste through the use of cover soils and drainage systems to reduce the production of leachate (Broun and Sattler, 2016). In comparison, bioreactor landfills increase the rate of solid waste decomposition by adding supplemental water or recirculating the leachate (Broun and Sattler, 2016). The distinct design principles for conventional and bioreactor landfills present different challenges for the microbial communities present in and degrading the waste (Broun and Sattler, 2016; Gilbert and Stephens, 2018). Bioreactor landfills pose an elegant solution for the need to efficiently store increasing volumes of MSW and offer an opportunity to harness methane emissions as an alternative fuel source. The comparatively rapid waste decomposition in bioreactor landfills is more economical for the production of biogas, a combination of methane and carbon dioxide, as an alternative fuel source. Broun and Sattler (2016) found that the production of methane in bioreactor landfills was 135 kilograms of carbon dioxide equivalents per one ton of MSW above that of conventional landfills. Bioreactor options for landfills provide *in situ* treatment and stabilization of solid waste and leachate (Bareither *et al.*, 2013). Microbial degradation reduces the organic waste fraction, which is the largest component of MSW, providing the potential to recover

space within the existing landfill for additional waste (Broun and Sattler, 2016; Cardinali-Rezende *et al.*, 2016; Hoornweg and Bhada-Tata, 2012). Bioreactor landfills also exhibit lower production of air pollutants, including nitrous oxide, carbon monoxide, and sulfur dioxide gases. Bioreactor landfills are a promising alternative to conventional landfills to better harness the energetic potential of waste and to more efficiently use the land.

The performance of waste degradation is dependent on the microbial communities within the landfill (Cardinali-Rezende *et al.*, 2016). Microbial waste degradation occurs in both conventional and bioreactor landfills, although conventional landfills are designed to reduce microbial processes (Broun and Sattler, 2016). The first three steps of the waste decomposition process are reliant on bacteria: hydrolysis; acidogenesis, both fermentation and beta oxidation; and acetogenesis (Cardinali-Rezende *et al.*, 2016). The last step, methanogenesis, is dependent on methanogenic archaea (Cardinali-Rezende *et al.*, 2016). Landfill deposits are diverse, both chemically and physically, which can make these microbial degradation processes difficult (Stamps *et al.*, 2016). Different bacteria and archaea in the waste degradation process can be affected by stress-inducing waste conditions that reduce their metabolic efficiencies, such as high ammonia concentrations that inhibit saccharolytic bacteria, important for the hydrolysis of cellulose materials (Cardinali-Rezende *et al.*, 2016). Understanding the ecology and diversity of the bacterial and archaeal community structure in bioreactor landfills, and landfills in general, will lead to a better understanding of the decomposition process, allowing for better control of methane production for energy conversion and more efficient waste management strategies.

1.4 Landfill microbial communities

Landfills are understudied built environments, and much is still unknown about the microbial communities inhabiting landfills and their associated functions. This is in part because of the challenge posed by the heterogeneity of landfills (Hilger and Barlaz, 2006; Stamps *et al.*, 2016). The range of inputs to built environment microbial communities like landfills and the resulting unique selection pressures lead to microbial communities that can vary greatly compared to more natural communities. High heterogeneity of MSW deposits makes it difficult to gather a representative sample of the microbial community. Additionally, some researchers are limited to sampling leachate from capped landfills (Hilger and Barlaz, 2006; McDonald *et al.*, 2010). Scientists have developed methods to circumvent these problems in order to meet research objectives, including taking multiple replicate samples from full scale landfills and running laboratory simulations with shredded waste

(Hilger and Barlaz, 2006). Leachate samples are also used for gathering biomass for cultivation assays and DNA extractions; however, the leachate community may differ from that associated with the solid waste portion of the landfill (McDonald *et al.*, 2010). Despite this potential disconnect, the leachate samples provide insights into the microbial community composition and the degradation potential of the community (Stamps *et al.*, 2016).

With on-going development of research methods for sampling landfills and growing interest in characterizing uncultivated microbes from unique environments, the number of studies investigating landfill microbial diversity has been increasing. Most of landfill microbiology research has focused on specific aspects of waste degradation in landfills and the microbes responsible, with special interest in methane cycling and cellulose degradation. For example, Lin *et al.* (2009) used 16S rRNA gene sequencing to examine the diversity and abundance of type I and type II methanotrophs in the microbial community of cover soils in a landfill in Ontario, Canada. Also using 16S rRNA gene sequencing, Burrell *et al.* (2004) determined that the microbes responsible for cellulose degradation in a landfill leachate bioreactor were affiliated with two groups of *Clostridium* (phylum Firmicutes). Similarly, McDonald *et al.*, (2010) used a community approach using 16S and 18S rRNA gene probes to target bacterial, archaeal, and fungal groups associated with cellulose degradation. Although this function-focused research provides important insights into the diversity and abundance of these specific activities in landfills, only a subset of the total microbial community was identified or characterized, without investigating interactions between these cellulose degraders or methanotrophs and the wider community.

Several research groups have examined overall landfill microbial diversity and community composition, typically using high throughput sequencing techniques like 16S rRNA gene amplicon sequencing. The Firmicutes and Bacteroidetes are consistently the most abundant phyla across studies (Cardinali-Rezende *et al.*, 2016; Song *et al.*, 2015; Stamps *et al.*, 2016; Xu *et al.*, 2017). Campylobacterota (formerly Epsilonproteobacteria (Parks *et al.*, 2018)) and Proteobacteria are also often highly abundant, particularly the Gammaproteobacteria (Song *et al.*, 2015; Stamps *et al.*, 2016) but are more likely to be enriched in the landfill leachate than in the solid fraction of the waste (Xu *et al.*, 2017). Actinobacteria, candidate division OP3, and Tenericutes were also among the abundant phyla across different landfills (Cardinali-Rezende *et al.*, 2016; Song *et al.*, 2015; Stamps *et al.*, 2016). The high throughput nature of 16S rRNA amplicon sequencing also allow examination of the rare or novel biosphere in landfills. Cardinali-Rezende *et al.*, (2016) identified that very low

abundance organisms (singletons and doubletons in 16S rRNA gene amplicon data) accounted for 62% of the operational taxonomic units. Additionally, Stamps *et al.*, (2016) showed that more than 10% of their operational taxonomic units were unclassified. The phylum level signature of the microbial communities was consistent across landfills overall and all studies show that there are many rare and uncharacterized microorganisms present in landfills.

Although these 16S rRNA gene studies have expanded our understanding of landfill microbial diversity, such approaches are limited in their ability to determine the metabolic functions of the microorganisms. Song *et al.*, (2015) inferred functions about known groups that they found in the landfills, including for members of the genus *Sulfurimonas*, which were predicted to be involved in sulfur cycling. Similarly, Cardinali-Rezende *et al.*, (2016) found a predominance of known proteolytic bacteria, including *Proteiniphilum* and *Gallicola* species. These inferences are useful in predicting the potential functional capabilities of the microbial community, but researchers would need to use techniques like metagenomics to confirm the presence of relevant functional genes, and expression assays like transcriptomics to suggest that the identified microbes are active in the inferred processes. As well, the 16S rRNA gene does not provide sufficient functional information for unclassified microorganisms that are uncultivated or newly discovered. Thus, 16S rRNA gene studies are most useful for studying the taxonomic diversity of microbial communities in landfills, but fall short when considering the functional diversity.

In terms of landfill heterogeneity, previous studies identified several environmental or geochemical factors that influenced microbial community composition. Both Song *et al.*, (2015) and Stamps *et al.*, (2016) studied multiple landfills within their respective study countries, China and the USA, and determined that age of the landfill or the age of the waste correlated with the microbial community composition. This was also seen by Cardinali-Rezende *et al.*, (2016), as their bioreactor communities changed over the two year period from start up, with shifts in the dominant phyla as the community established. These findings suggest that landfill microbial community composition changes over time, but more long term studies are needed to determine if the change in community structure is predictable or if it is dependent on the landfill and its waste components. Community composition was also correlated with moisture (Song *et al.*, 2015; Xu *et al.*, 2017) and ammonium concentration (Cardinali-Rezende *et al.*, 2016; Xu *et al.*, 2017). Other chemicals that showed a link to microbial composition included barium, chloride, sulfate, and copper (Stamps *et al.*, 2016; Xu *et al.*, 2017). The correlations between landfill age, ammonium concentration, and moisture content on landfill

microbial communities may be useful as predictors for community composition; however, hypotheses based on the observed correlations would need to be tested. Other chemical factors seem to affect microbial communities in a site-specific manner, and their effects will depend on the types of waste deposited and other geochemical conditions at each site of interest.

1.5 A case study on Tenericutes

Landfill-associated bacteria that warrant further study are those affiliated with the phylum Tenericutes. Both Stamps *et al.*, (2015) and Song *et al.*, (2015) identified Tenericutes within their landfill samples, but the lifestyles and metabolic roles of these landfill-associated Tenericutes have yet to be determined. Life strategies for the Tenericutes are typically parasitic or commensal with eukaryotic hosts (Skenneron *et al.*, 2016), with several being animal or human pathogens (Ludwig *et al.*, 2010). Tenericutes have undergone genome reductions typically associated with host-associated life strategies during which many biosynthetic pathways for amino acids, fatty acids, and other molecules have been lost (Joblin and Naylor, 2002; Skenneron *et al.*, 2016). Most characterized Tenericutes lack the tricarboxylic acid cycle, relying on anaerobic fermentation of carbohydrates and small organic acids for energy production whereas the *Ureaplasma* couple urea hydrolysis with ATP production (Razin, 2006; Skenneron *et al.*, 2016; Smith *et al.*, 1993). This anaerobic metabolism is consistent with the low oxygen environments where Tenericutes have been found, such as landfill leachate (Song *et al.*, 2015; Stamps *et al.*, 2016) and the gastrointestinal tracts of humans (Trosvik and de Muinck, 2015) and other animals, including the Chilean octopus, *Octopus mimus* (Iehata *et al.*, 2015), and the scorpion *Androctonus australis* (Elmnasri *et al.*, 2018). Although most known or cultivated Tenericutes are host-associated, Tenericutes have also been associated with free-living lifestyles based on their environmental location. For example, Skenneron *et al.*, (2016) cultured Tenericutes from the new order Izimaplasma from an ocean methane seep, and determined these Tenericutes were not directly associated with eukaryotic hosts. Much of the currently described diversity of Tenericutes focuses on medically or agriculturally important organisms. The total diversity of Tenericutes is potentially much greater both metabolically and phylogenetically.

The Tenericutes are an interesting group from a phylogenetic perspective, because they are typically nested within the Firmicutes in bacterial phylogenies, making the Tenericute phylum status a point of contention (Davis *et al.*, 2013; Ludwig *et al.*, 2010; Parks *et al.*, 2018). The Tenericutes were elevated to a phylum because they possess a distinctive wall-less cell envelope (Ludwig *et al.*, 2010) and they exhibited phylogenetic separation based on a number of conserved molecular markers,

including elongation factor Tu and RNA polymerase (Ludwig and Schleifer, 2005). In a recent reorganization of microbial taxonomy undertaken by Parks *et al.*, (2018), Tenericutes were removed as a distinct phylum and instead the former Tenericutes orders were placed within the class Bacilli (phylum Firmicutes). The appropriate taxonomy of this group may become clearer with the addition of members from more diverse environments, such as landfills.

1.6 Study site

The study site for my research is a municipal waste landfill in southern Ontario, Canada that opened in 1972. This landfill is a conventional sanitary landfill with onsite waste sorting, compacting, and daily soil covers. The landfill design includes a leachate capture system, including wells for monitoring the composition of leachate and composite cisterns for collecting and mixing the leachate before it is sent to a waste water treatment plant. There are also a number of groundwater wells on the landfill property that access the adjacent aquifer to detect any leachate leaks and to assess groundwater quality. The landfill also has a gas capture system and biogas plant to collect and convert emitted methane to biogas. Three leachate wells in the active landfill section, a composite leachate cistern, and two groundwater wells, one pristine and one impacted by leachate, were sampled in 2016.

1.7 Scope of research and research objectives

Landfills are final deposition sites for household and municipal waste, and are sites housing considerable microbial diversity. Microbes conduct waste degradation in landfills, whether planned (sanitary bioreactor) or unplanned (sanitary conventional), and contribute to two environmental concerns with landfills: 1) landfills emit methane, a greenhouse gas (Hoornweg and Bhada-Tata, 2012); and 2) landfill leachates contain potentially toxic compounds with unknown consequences for environmental and human health (Stamps *et al.*, 2016). The heterogeneity of landfill waste is a main reason for the taxonomic and functional diversity of microbial inhabitants. The microbial diversity in landfills is of specific interest because these microbes may possess functional capabilities that are applicable to industrial applications, bioremediation of contaminated sites, or modulation of methane production, either to stimulate production in bioreactor landfills for an alternative biogas fuel source, or to reduce methane production and emissions for landfills lacking methane capture systems.

My research aims to expand on the current body of knowledge assessing the diversity of landfill microbial communities, microbial relationships to geochemical parameters, and the potential ecological roles and functions within the landfill microbial communities. The objective of my

research in Chapter 2 is to use both metagenomic and 16S rRNA gene amplicon sequencing technologies to investigate the distribution, heterogeneity, and diversity of microbial communities in the Southern Ontario landfill. I additionally leverage 16S rRNA gene amplicon data in combination with sample site chemical data to determine which, if any, geochemical variables are linked to microbial distributions within the landfill. In Chapter 3, my objective is to use metagenomic data to investigate the taxonomy and predicted metabolisms of landfill-associated Tenericutes, and then test those predictions with cultivation trials. The landfill-associated Tenericutes are relatively abundant in my study site and their roles within the landfill are uncertain as current knowledge of the Tenericutes is largely limited to medically or agriculturally important members, and little is known about non-host-associated Tenericutes. Here, I identify Tenericutes from the landfill, reconstruct genomes for key representatives, and predict their ecological roles, life histories, and possible functions within the landfill.

Chapter 2

Microbial diversity in a Southern Ontario Landfill

2.1 Introduction

Global waste production continues to increase, with the majority of waste destined for landfills despite recycling and composting efforts (Hoornweg and Bhada-Tata, 2012; Hoornweg *et al.*, 2013). Landfills are engineered ecosystems unique from other built and natural environments, and are understudied in terms of their microbial communities (Stamps *et al.*, 2016). Landfills receive heterogeneous mixtures of waste that contribute nutrients, toxins, and microorganisms to the ecosystem (Köchling *et al.*, 2015; Stamps *et al.*, 2016). The composition of waste input, local climate, and soil and groundwater geochemistry all affect the establishment of microbial communities in landfills, and this community structure changes over time as populations stabilize and waste composition changes (Cardinali-Rezende *et al.*, 2016; Sawamura *et al.*, 2010; Stamps *et al.*, 2016). Previous studies have examined landfill microbial diversity to the phylum level, but have not adequately addressed whether organismal diversity is similar across landfills or over time at lower levels of classification (Cardinali-Rezende *et al.*, 2016; Köchling *et al.*, 2015).

The sampling site for this research, the Southern Ontario landfill described in Chapter 1, is well-instrumented, with over 100 leachate wells across the site as well as three composite leachate cisterns. The leachate wells are routinely sampled by regional waste management staff to determine the chemical composition of the leachate (Figure 2.1, Table 2.1). There are also groundwater wells bordering the landfill for monitoring the conditions of the adjacent aquifer and any leachate leaks (e.g., there is known leachate infiltration from the area near LW3 into the aquifer near GW1, Figure 2.1). In order to understand waste degradation processes and the transformation and movement of contaminants within the site, it is important to understand how the microbial communities change across the landfill. The objectives of the research presented in this chapter are to characterize the distribution, heterogeneity, and diversity of the microbial communities in the Southern Ontario municipal landfill by investigating community diversity and compositions across the study site. Further, this work investigates how the observed microbial heterogeneity relates to geochemical conditions across the site.

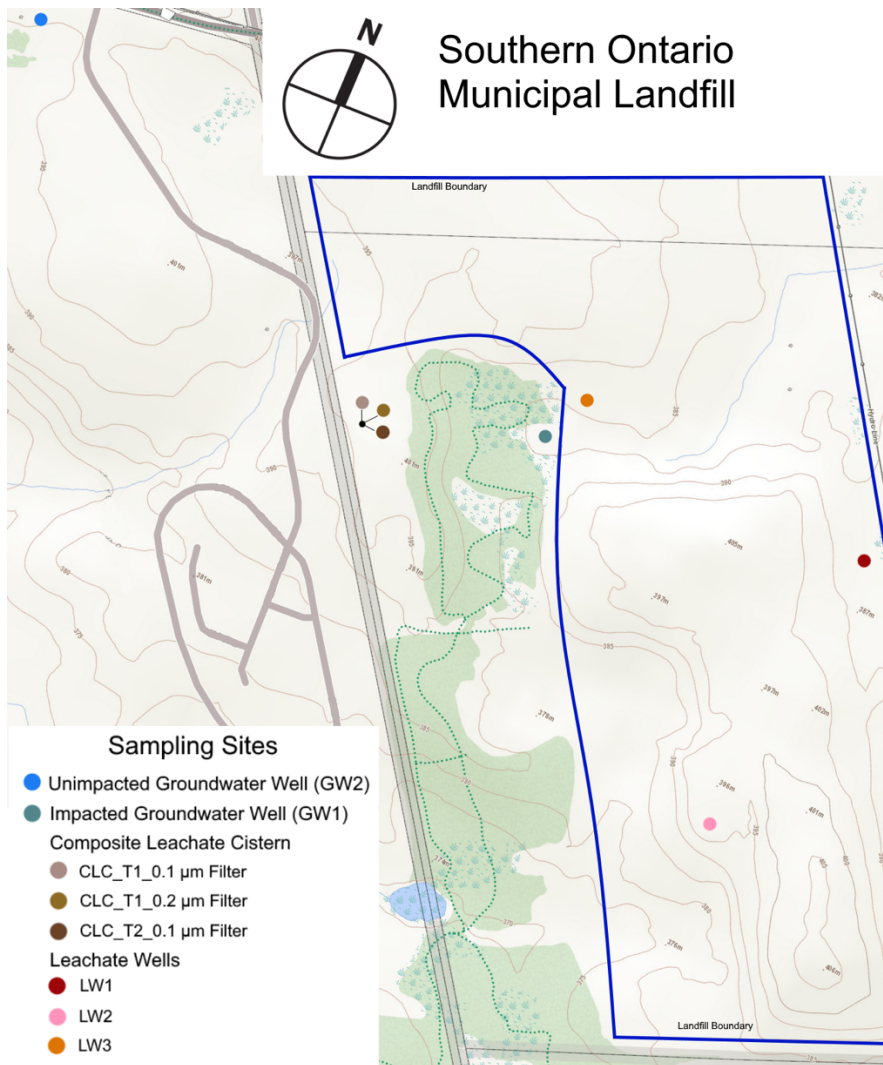


Figure 2.1: Map of landfill sampling locations at the Southern Ontario landfill. Two groundwater wells accessing the adjacent aquifer were sampled. GW1, the impacted groundwater well, is closer to the active landfill and is not as deep as GW2, the unimpacted groundwater well. The three samples from the leachate collecting cistern were all sampled from the same cistern at two time points. Two filter sizes were used for collecting microbial biomass on July 14, 2016 (CLC_T1_0.1 µm filter and CLC_T1_0.2 µm filter) and one filter size was used on July 20, 2016 (CLC_T2_0.1 µm filter). The three leachate wells are located across the active landfill, and leachate from these wells is pumped to the leachate collecting cistern. The catchment area for LW3 has an active leak infiltrating into the groundwater near GW1. The topographic map was modified from maps provided by the Ontario Ministry of Natural Resources and Forestry.

Table 2.1: Environmental variable measurements averaged from April and October 2016 values for the two groundwater wells and the three leachate wells.

<i>Environmental Parameter</i>	<i>units</i>	<i>GW1*</i>	<i>GW2</i>	<i>LW1</i>	<i>LW2</i>	<i>LW3</i>
Arsenic	mg/L	0.02	0	0	0.04	0
Boron	mg/L	0.79	0.01	2.10	7.53	1.45
Cadmium	mg/L	0	0	0.10e ⁻³	0	0
Calcium	mg/L	240	86.50	175	93.25	155
Chromium	mg/L	0	0	0.04	0.40	0.02
Copper	mg/L	0	0	0.01	0.36	0.03
Iron	mg/L	40	0.37	56.50	11.78	10.55
Lead	mg/L	0	0	0.01	0.05	0
Magnesium	mg/L	38	24	145.50	145	58.50
Manganese	mg/L	1.20	0.03	0.32	0.29	0.67
Mercury	mg/L	0	0	0	0.30e ⁻³	0
Nickel	mg/L	0.01	0	0.05	0.36	0.02
Potassium	mg/L	16	1.40	345	917.50	79.50
Sodium	mg/L	76	4.55	615	2525	200
Zinc	mg/L	0	0	0.06	0.28	0.01
Alkalinity total (as CaCO ₃)	mg/L	687	250	2800	7025	1200
Ammonia-N	mg/L	18	0	275	1375	123.50
Chloride	mg/L	180	6.70	655	3050	255
Nitrite	mg/L	0	0	0.01	46.24	0.01
Sulfate	mg/L	17	60.50	0	70	51.50
Total dissolve solids	mg/L	1040	315	3255	9657.50	1345
Total Kjeldahl nitrogen	mg/L	20	0.11	315	1925	135
Un-ionized ammonia	mg/L	0.03	0	2.32	99.88	0
1,4-Dichlorobenzene	ug/L	0	0	0	0	11.50
Benzene	ug/L	0.69	0	13.50	0	5.50
Chlorobenzene	ug/L	0	0	6	0	3.80
Ethylbenzene	ug/L	0.13	0	275	8.25	17.50
m&p-Xylenes	ug/L	0.07	0.06	1040	26.25	200
o-Xylene	ug/L	0.06	0	250	17.05	11.20
Toluene	ug/L	0	0	17	0	0

*All measurements for GW1 except the volatile compounds (1,4-dichlorobenzene to toluene) are from April 2011 and are for comparison only. These values were not included in the PCA analysis.

2.2 Methods

2.2.1 Sample collection

Samples were collected from a Southern Ontario landfill in July 2016. Samples were collected from three leachate wells, two samples from a composite leachate cistern at time points separated by one week, and samples from two groundwater wells adjacent to the landfill (Figure 2.1). In the initial sampling event, a sample was collected from the composite leachate cistern on July 14, 2016 by filtering the leachate through a 0.2 μm poly-ethersulfone filter followed by a 0.1 μm poly-ethersulfone filter in series. Both filters were kept for DNA extractions. On July 20, 2016, a larger-scale sampling was conducted. Leachate and water samples were collected by pumping the liquid through a filter apparatus with a 3 μm glass fiber pre-filter in series with a 0.1 μm poly-ethersulfone filter until filters clogged. The pre-filter was discarded while the 0.1 μm filters carrying the microbial biomass were kept. All filters were frozen on dry ice in the field and transferred to a $-80\text{ }^{\circ}\text{C}$ freezer until processed. DNA was extracted from the cells using the Powersoil DNA extraction kit (MoBio) following the manufacturer's instructions with one modification: filters were sliced into pieces and added to the bead tube in place of a soil sample.

2.2.2 Non-volatile and volatile compound measurements

The landfill employs a monitoring company to assess the non-volatile and volatile compound concentrations across the site. The monitoring company takes measurements throughout the year at a suite of leachate and groundwater wells. The relevant measurements for the leachate and groundwater wells are taken in October and April, and the average values for these two dates in 2016 were used to estimate compound concentrations at the time of microbial biomass sampling (Table 2.1). The impacted groundwater well, GW1, did not have current non-volatile concentration measurements. For this well, measurements from 2011 are included for comparison purposes only (Table 2.1). No geochemical measurements are available for the composite leachate cistern.

2.2.3 Metagenomic sequencing

Six DNA samples were sent to the US Department of Energy's Joint Genome Institute (JGI) for metagenomic sequencing, assembly, and annotation: LW1, LW2, LW3, CLC_T1 (0.2 μm filter), CLC_T2, and GW1. The CLC_T1 and GW2 0.1 μm filters were not sent for metagenomic sequencing due to insufficient DNA recovered from those filters. The JGI sequenced the

metagenomes using the HiSeq platform (Illumina). Metagenomes were annotated using the DOE-JGI Metagenome Annotation Pipeline (MAP v.4) (Huntemann *et al.*, 2016).

2.2.4 16S rRNA gene amplicon sequencing

Eight samples were sent to the JGI for 16S rRNA gene amplicon sequencing: LW1, LW2, LW3, CLC_T1 0.1 μm and 0.2 μm filters, CLC_T2, GW1, and GW2. The JGI amplified the V4-V5 region of the 16S rRNA gene using the forward primer 16S_515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and the reverse primer 16S_926R (5'-CCGYCAATTYMTTTRAGTTT-3') using in-house protocols. Amplicons were sequenced on the MiSeq platform (Illumina) and reads were quality control checked using the iTagger pipeline (Tremblay *et al.*, 2015).

2.2.5 QIIME2 analysis

The 16S rRNA gene amplicons were analyzed using QIIME2 (Bolyen *et al.*, 2019). Forward and reverse reads from the JGI were separated using khmer (Crusoe *et al.*, 2015). Primers were trimmed from the forward and reverse reads using cutadapt in QIIME2 (Martin, 2011). Reads were demultiplexed by the JGI, allowing us to bypass the demultiplexing step for the QIIME2 pipeline. Barcodes had already been removed during read quality processing by the JGI. The forward reads were truncated at 270 base pairs and the reverse reads at 230 base pairs based on the quality score visualization produced by QIIME2 in the demux summary step. Reads were denoised using paired denoising in DADA2 within the QIIME2 platform which also merges the reads (Callahan *et al.*, 2016). Sequence variants were determined using DADA2 and summarized using feature-table summarize in QIIME2. Taxonomic assignment of the 16S rRNA gene amplicons was based on a phylogenetic tree produced by QIIME2 in which the taxonomy classifier was trained with the SILVA 99% taxonomy classification for the 16S rRNA gene from the April 2018 SILVA 132 release (Quast *et al.*, 2012). Phylum names were updated as per the GTDB database taxonomy changes by Parks *et al.* (2018) for diversity comparisons.

2.2.6 Phylogenetics

2.2.6.1 16S rRNA gene tree

All assembled and annotated 16S rRNA genes in the landfill metagenomes were downloaded from the JGI IMG server (IMG Genome IDs: 3300014203, 3300014204, 3300014205, 3300014206, 3300014208, 3300015214). Genes were sorted by length in Geneious 11.0.5 (<https://www.geneious.com>) and curated to a minimum length of 600 bp. We submitted the landfill metagenome-derived 16S rRNA genes as well as a reference set of 16S rRNA genes from known organisms to the SILVA SINA algorithm (Pruesse *et al.*, 2012) for alignment. Unaligned bases at the ends of the genes were removed and sequences below 70% identity to a reference sequence were automatically removed from the dataset by SINA. To curate the SINA alignment, columns containing 97% or more gaps were removed, a region of poor alignment was manually trimmed from the 3' end, and sequences falling below 600 bp post-trimming were removed. A phylogenetic tree was inferred using FastTree in Geneious to check for poorly aligned or divergent sequences. In this processing, 195 sequences were removed that did not meet quality standards. The final 16S rRNA gene alignment included 1,903 reference sequences and 2,306 sequences from the metagenome samples, and had 1,521 positions. A high-quality phylogenetic tree was inferred from the curated final alignment using RAxML-HPC2 8.2.12 (Stamatakis, 2014) on Cipres (Miller *et al.*, 2010) under model GTRCAT, with 100 alternative bootstrap iterations run from 100 starting trees.

2.2.6.2 Ribosomal protein tree

All amino acid sequences for 16 syntenic, universally-present, single copy ribosomal protein genes (RpL2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24 and RpS3, S8, S10, S17, S19) for the landfill metagenomes were downloaded from the JGI IMG server using annotation keyword-based identification (Hug *et al.*, 2013). Ribosomal protein datasets were screened for the Archaeal/Eukaryotic type, which were removed, as were short (<45 aa) sequences. Each individual protein set was aligned with a reference set of genes (Hug *et al.*, 2016) using MAFFT 7.402 (Katoh and Standley, 2013) on Cipres. Alignment columns containing $\geq 95\%$ gaps were removed using Geneious. IMG-derived sequence names were trimmed to 8 digits after the metagenome code (e.g. Ga0172377_100004578 \rightarrow Ga0172377_10000457) to remove gene-specific identifiers and allow for concatenation by scaffold name. The protein gene alignments were concatenated in numeric order (L2 \rightarrow L24, followed by S3 \rightarrow S19). As not all scaffolds contained all of the 16 proteins in either partial

or complete form, concatenated sequences that contained less than 50% of the total expected number of aligned amino acids were removed. The final alignment was 3,452 columns long and contained 2,914 reference organisms and 1,265 scaffolds from the metagenome samples. A phylogenetic tree was inferred using RAxML-HPC Blackbox on Cipres using the following parameters: sequence type - protein; protein substitution matrix – LG; and estimate proportion of invariable sites (GTRGAMMA + I) – yes (Miller *et al.*, 2010; Stamatakis, 2014).

2.2.7 Binning

All scaffolds >2,500 bp were included in the binning process. The binning algorithm CONCOCT (Alneberg *et al.*, 2014) was used in Anvi'o (Eren *et al.*, 2015) to automatically cluster each metagenome's scaffolds using a combination of scaffold tetranucleotide frequencies and read-mapped coverage data from all six metagenomes. Gene annotations were imported from the JGI annotations, overriding the automated annotation pipeline in Anvi'o. The bins were manually refined for the six metagenomes using Anvi'o, focusing on completion and quality metrics to guide bin refinements. High quality bins were considered those with greater than 70% completion and less than 10% redundancy. An in-house pipeline was developed to connect Anvi'o bin information with the scaffold data from the JGI.

2.2.8 Diversity analysis

Metagenome-derived sequences were assigned to the phylum level based on their placement within reference clades on the 16S rRNA and concatenated ribosomal protein phylogenetic trees. Metagenome sequences placing outside of or between phyla were assigned to either “Unclassified Archaea” or “Unclassified Bacteria” as appropriate. Phylum names were updated from the NCBI taxonomy to conform to the GTDB database taxonomy by Parks *et al.* (2018). Bins were identified at the phylum level using the scaffold assignments from the 16S rRNA gene and concatenated ribosomal protein phylogenetic trees. Bin abundances were determined using the average fold coverage data for all scaffolds in the bin. Phylum abundance per sample was calculated by summing the average fold coverage data for each scaffold on the tree assigned to the phylum, where the scaffold acts as a proxy for the underlying microbial population. Microbial diversity comparisons were visualized using stacked bar plots produced using ggplot2 in R (Wickham, 2011).

Additional diversity analyses were conducted using the 16S rRNA gene amplicon exact sequence variants (ESVs) identified by QIIME2 (Bolyen *et al.*, 2019). The alpha diversity metrics

Faith's phylogenetic diversity (Faith and Baker, 2007) and Pielou's evenness (Pielou, 1966) were calculated using QIIME2 for the four sample types: impacted groundwater well, unimpacted groundwater well, leachate well, and composite leachate cistern. A Shannon diversity index analysis with rarified sequence depth of 20,000 was conducted using QIIME2 and visualized using phyloseq (McMurdie and Holmes, 2013) in R. A Chao1 statistic was not calculated, as data processing with QIIME2 and DADA2 removes all singleton ESVs, which the Chao1 statistic requires. For beta diversity measures, full ESV and taxonomy tables were input to unweighted and weighted UniFrac distances principle coordinate analysis, calculated using phyloseq and visualized in R for all samples. The top 10% most abundant ESVs by count across samples were visualized using ggplot2 in R. The prevalence across samples of ESVs with a count of 3 or more and belonging to phyla with relative abundance greater than 1% was determined using phyloseq and visualized in R. The ESVs present in two or more sites were visualized using ggplot2 in R.

A principal component analysis (PCA) for the 16S rRNA gene amplicon ESVs present at two or more sites was conducted using vegan (Oksanen *et al.*, 2018) in R. The ESV count data was Hellinger transformed to reduce the weight of ESVs with low counts and zeros. The leachate wells and the two groundwater well samples were included in the PCA to allow for comparison with environmental parameters, which are available for those sites. Environmental data was not available for the composite leachate cistern site and so CLC samples were excluded from this analysis.

2.2.9 Chemical data analysis

Chemical measurements provided by the Southern Ontario landfill 2016 annual report were used to determine variance of metals and volatile compounds over time for the three leachate wells and two ground water wells. GW1 only has metal measurements for one time point in 2011 and so variance could not be calculated. Non-detects, where a compound, if present, is below the detection limit, were treated as zeros. The measurements were log transformed and visualized in a heatmap using heatmap3 (Zhao *et al.*, 2014) in R. Metal and volatile compounds detected in a majority of samples were used for further analysis. The measurements from April and October of 2016 were averaged to estimate the concentrations at the time of microbial biomass sampling.

PCA for the metals and volatile compounds were conducted using vegan (Oksanen *et al.*, 2018) in R. The metal and volatile compound concentrations were square root transformed to reduce the range of the values as measurements differed by orders of magnitude (Table 2.1). Data for leachate wells and the two groundwater well samples were included for the volatile analysis, but

GW1 was excluded from the metals analysis as no data were available for that site in 2016. A PCA was also conducted using *vegan* in R for the other geochemical parameters measured at the sites that are not characterized as metals or volatile compounds (*e.g.*, total dissolved solids (TDS)).

2.3 Results and discussion

2.3.1 Phylum level diversity

Twenty-five phyla are present at greater than 1% relative abundance in at least one landfill sample (Figure 2.2) and these phyla represent lineages within both Bacteria and Archaea (Figure 2.3). These phyla are present across multiple samples (Figure 2.2, Table 2.2). Phylum level profiles are relatively consistent between the 16S rRNA gene amplicon and metagenomic sequencing (Figure 2.2). A notable exception to this are the Patescibacteria, which make up a reduced fraction in the 16S rRNA gene amplicon results (max relative abundance of 23% in GW1) but exhibit the highest relative abundance in the landfill (mean relative abundance of 34% (based on the 16 ribosomal protein-carrying scaffold calculated abundances) and peak relative abundance of 79% in GW1). Members of the Patescibacteria are often underrepresented in 16S rRNA gene amplicon studies due to primer mismatches and long insertions in the gene (Brown *et al.*, 2015), but can be more robustly identified using metagenomic techniques (Schulz *et al.*, 2017). Previous landfill microbial diversity studies, as described in Chapter 1, have relied on 16S rRNA gene amplicon sequencing, which may have systematically underestimated the abundance of Patescibacteria. The Bacteroidota (mean: 16%, peak: 31.89% in CLC_T1), Firmicutes (mean: 10.19%, peak: 28.74% in CLC_T1), and Proteobacteria (mean: 10%, peak: in 28% LW2) were also highly abundant in the landfill. These findings are consistent with other studies where Bacteroidota, Firmicutes, and Proteobacteria are frequently detected as the most abundant bacterial phyla in landfills (Song *et al.*, 2015; Stamps *et al.*, 2016; Xu *et al.*, 2017).

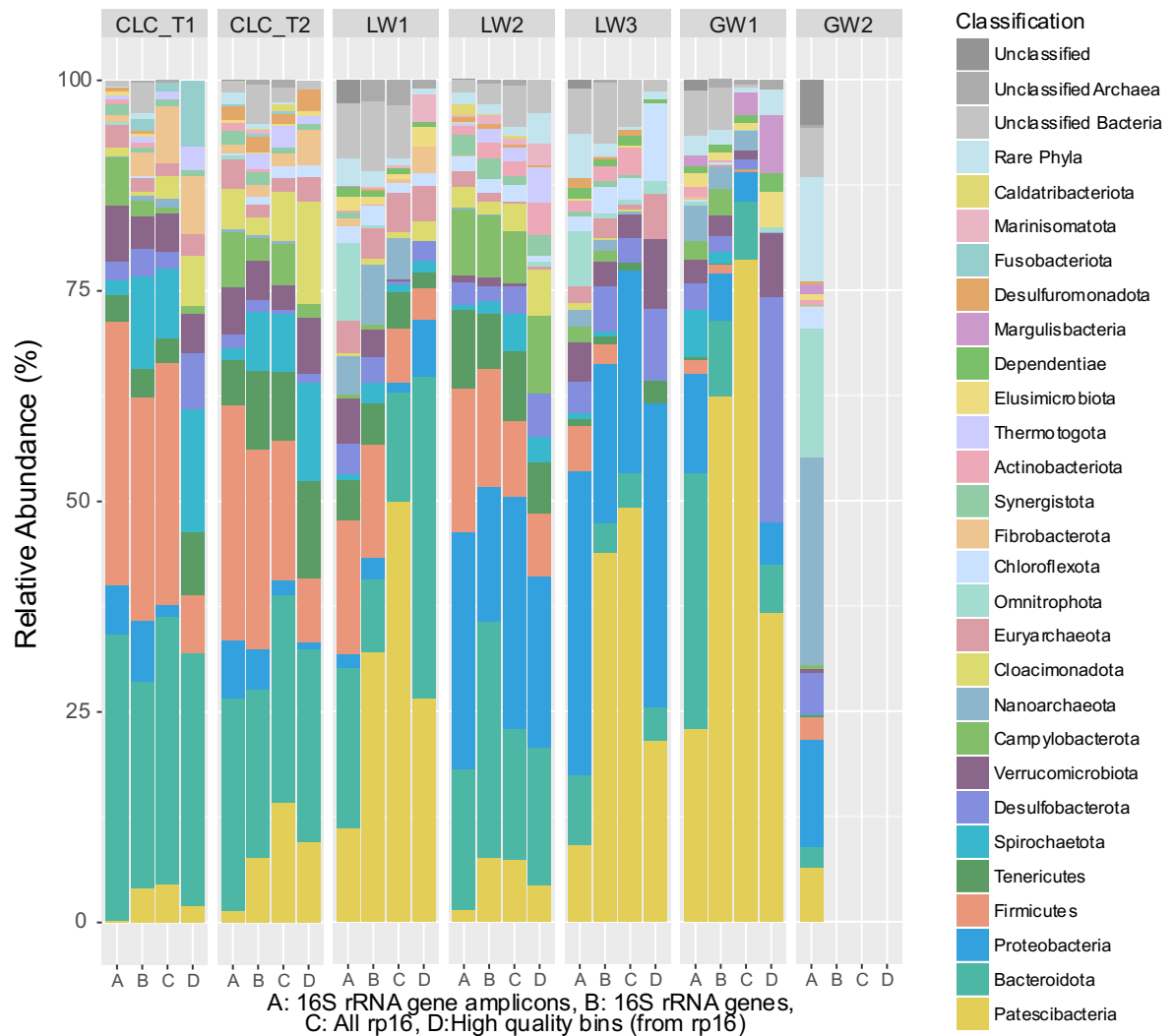
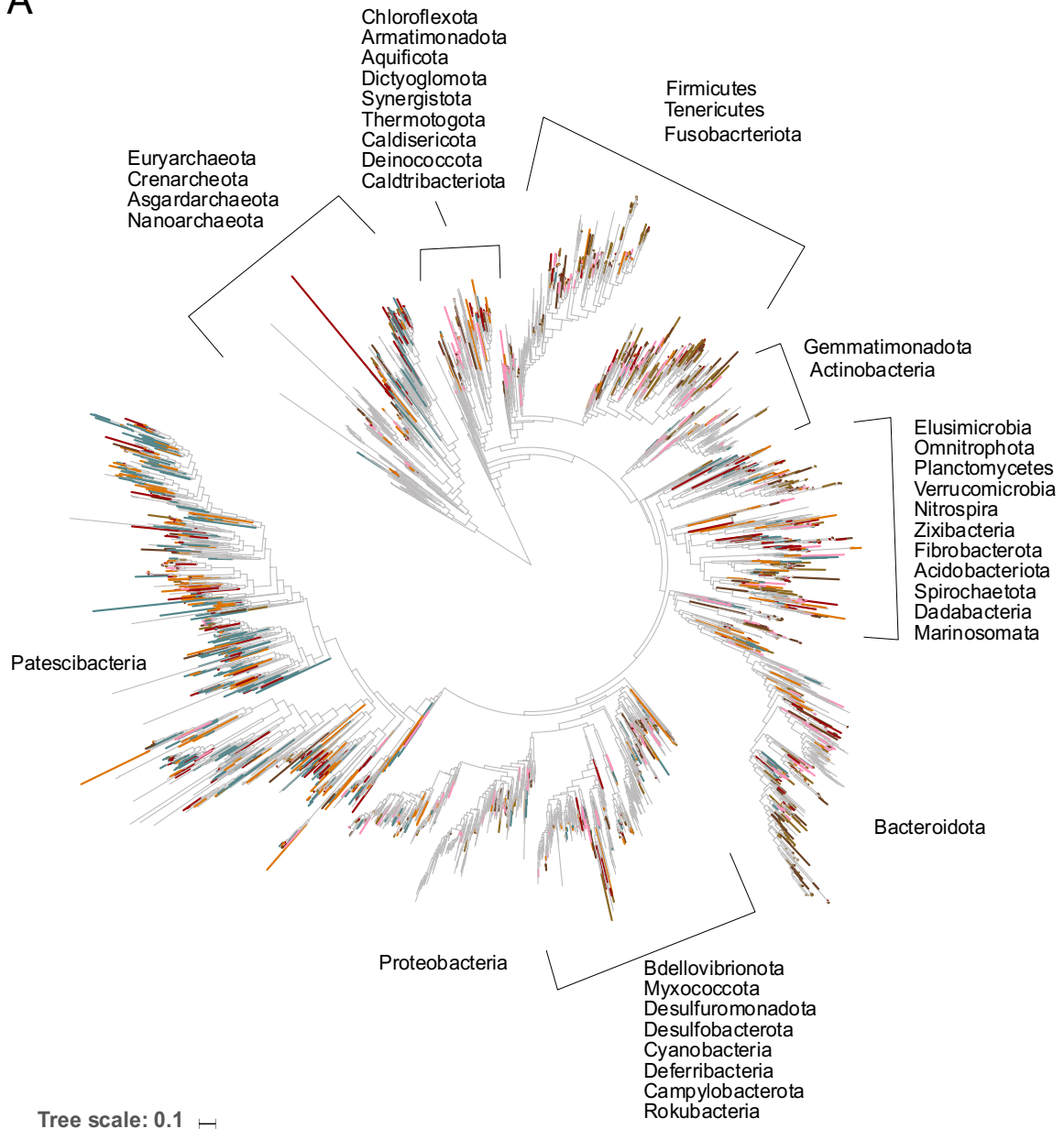


Figure 2.2: Relative abundance of bacterial and archaeal phyla present at greater than 1% abundance in at least one sample site. A) 16S rRNA gene amplicons; B) 16S rRNA genes derived from the assembled metagenomes; C) concatenation of 16 syntenic ribosomal proteins derived from the assembled metagenomes, with scaffold coverage as a proxy for abundance (all rp16); and D) high quality bins (containing the 16 concatenated ribosomal proteins, and with abundances calculated from the ribosomal protein-encoding scaffolds' coverages). The Composite Leachate Cistern at timepoint 1 (CLC_T1) is represented by the 0.2 μm filter data as the 0.1 μm filter showed highly similar results (Figure 2.11). Site GW2, the unimpacted groundwater well, did not yield sufficient biomass for metagenomic sequencing, thus only 16S rRNA gene amplicons were generated and reported here.

A



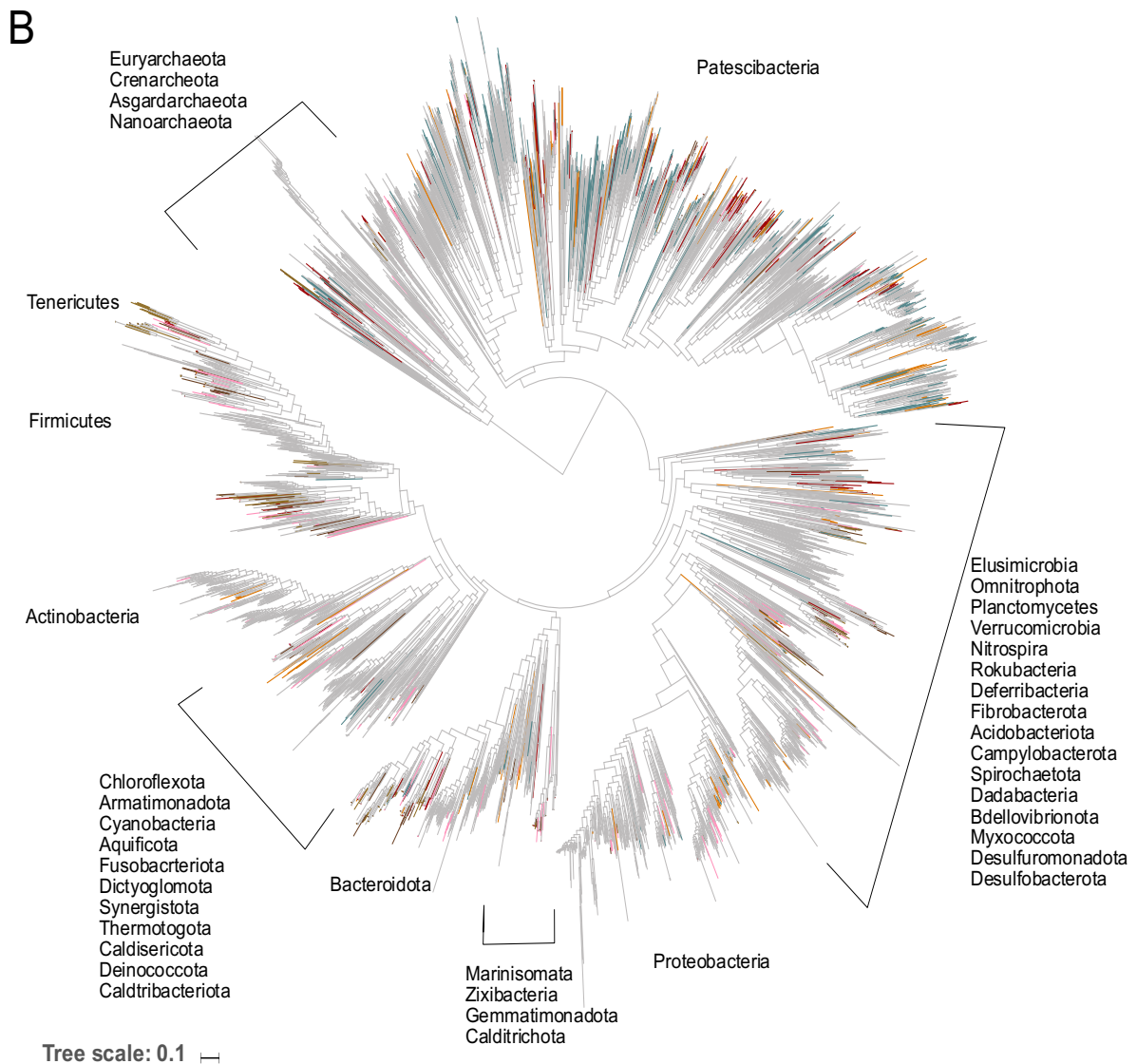


Figure 2.3: Maximum likelihood phylogenetic trees derived from metagenome sequences. A) A maximum likelihood phylogenetic tree inferred from 16S rRNA genes of 600 bp or longer identified in the assembled metagenomes. The tree includes 2,306 metagenome-derived 16S rRNA genes and 1,903 reference sequences. B) A maximum likelihood phylogenetic tree inferred from a concatenation of 16 syntenic ribosomal proteins. The tree includes 1,265 metagenome-derived scaffolds and 2,914 reference genome sequences. Coloured branches correspond to the landfill sampling site for metagenome-derived scaffolds: GW1 in teal, CLC_T1 in light brown, CLC_T2 in dark brown, LW1 in red, LW2 in pink, and LW3 in orange. Number of scaffolds included per sample are listed in Table 2.2. Trees were visualized using iTOL (Letunic and Bork, 2019).

Table 2.2: The number of metagenome-derived scaffolds per metagenome present on the concatenated ribosomal protein (rp16) and the 16S rRNA gene phylogenetic trees.

<i>Sampling Site</i>	<i>rp16</i>	<i>16S rRNA genes</i>
<i>GW1</i>	357	448
<i>LW1</i>	187	333
<i>LW2</i>	208	358
<i>LW3</i>	173	371
<i>CLC_T1</i>	137	353
<i>CLC_T2</i>	203	443

2.3.2 Alpha diversity

All of the landfill samples have a Shannon index above 5.0 for 16S rRNA gene amplicon data, suggesting that these sites have relatively high levels of microbial richness and evenness (Figure 2.4). The eight samples from our study also exhibit high Pielou's evenness (>0.9) without significant differences between sample types (Figure 2.5, Table 2.3). There is also no significant difference between the sample types when considering Faith's phylogenetic diversity (Figure 2.5, Table 2.3). Although there are no significant differences among samples, GW1 shows lower richness and overall phylogenetic distance than the other samples, but similar evenness with GW2. The landfill leachate well and composite leachate cistern sample diversities are consistent with the findings of Stamps *et al.* (2016) who showed high richness and evenness across the 19 landfills in their study. The implication of these high alpha diversity values is that the landfill microbial communities consist of phylogenetically diverse microorganisms that have relatively equal abundances at the species level regardless of the presence of dominant phyla (Figure 2.2).

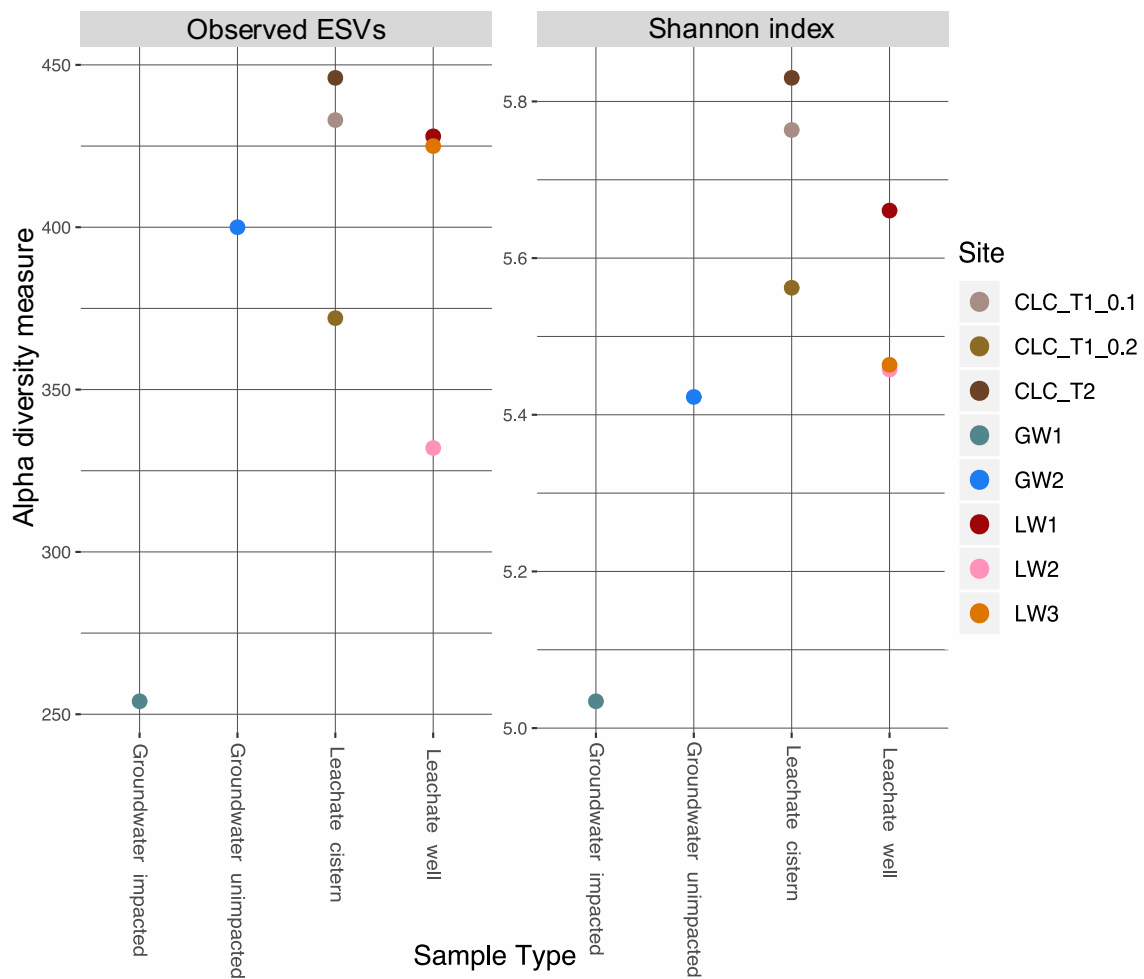


Figure 2.4: Observed diversity and Shannon index for the eight samples, grouped by sample type on the x axis. Observed ESVs fall between 250-450 total ESVs for each sample, indicating similar levels of microbial community heterogeneity across sites. All sites show a high level of diversity, with Shannon's indices greater than 5.

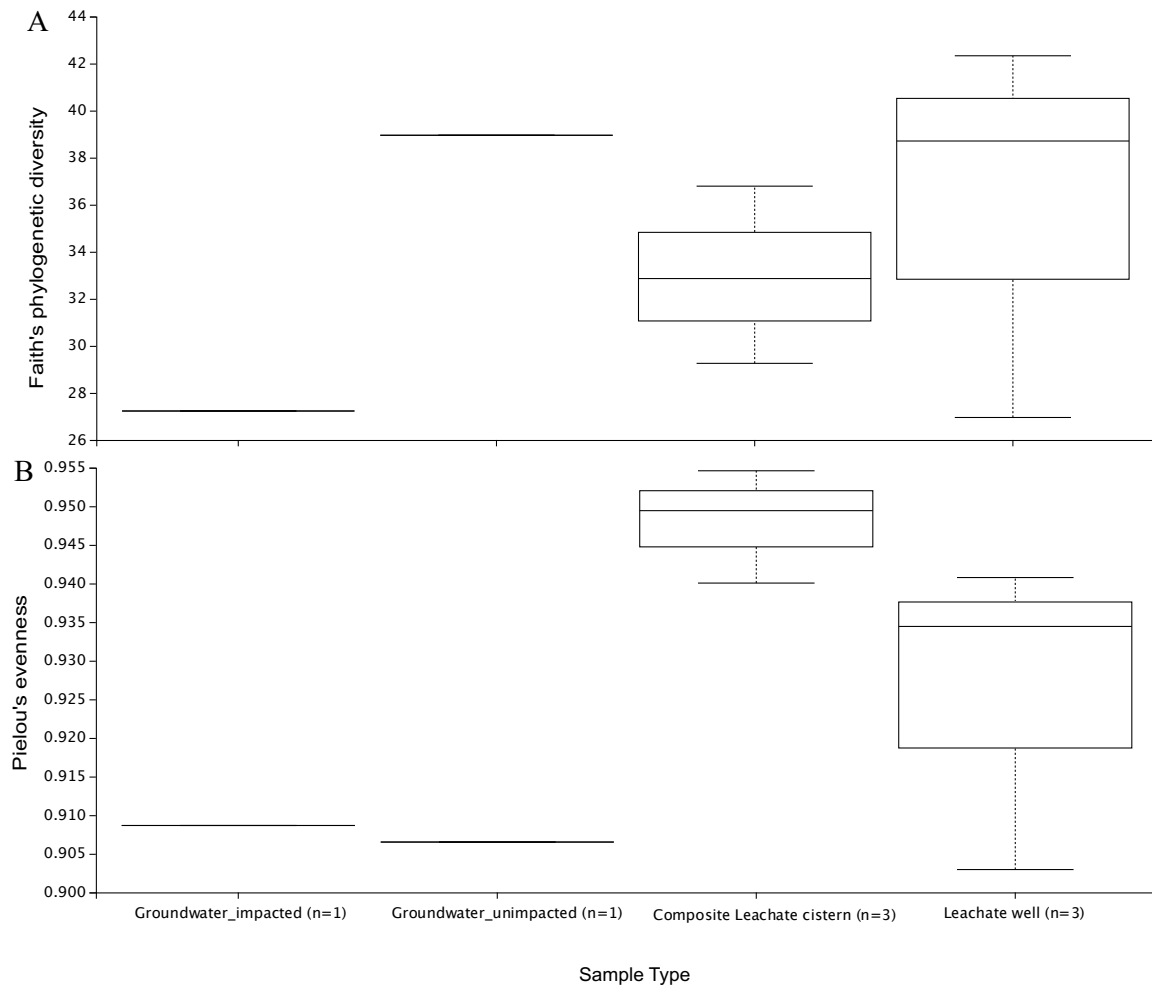


Figure 2.5: Alpha diversity analyses by sample type for Faith's Phylogenetic Diversity and Pielou's Evenness. A) Faith's phylogenetic diversity. The impacted groundwater sample shows lower feature diversity than the other samples. The leachate well samples show a larger range of feature diversity than the composite leachate cistern samples. A Kruskal-Wallis test for all groups is not significant ($H=2.33$ and $p=0.51$). Kruskal-Wallis pairwise tests are also not significant for all pairs ($p \geq 0.18$). B) Pielou's evenness. All sample types show high levels of evenness among their respective species ($J' > 0.9$). A Kruskal-Wallis test for all groups is not significant ($H=4.11$ and $p=0.25$). Kruskal-Wallis pairwise tests are also not significant for all pairs ($p \geq 0.13$).

Table 2.3: Kruskal-Wallis test results for Faith’s phylogenetic diversity (PD) and Pielou’s evenness (E) diversity metrics.

<i>Test</i>	<i>Groups</i>	<i>H</i>	<i>p-value</i>	<i>q-value</i>
<i>Faith’s</i>	All groups	2.33	0.51	NA
<i>PD</i>	Groundwater impacted Groundwater unimpacted	1	0.32	0.63
	Groundwater impacted Leachate cistern	1.80	0.18	0.54
	Groundwater impacted Leachate well	0.20	0.65	0.65
	Groundwater unimpacted Leachate cistern	1.80	0.18	0.54
	Groundwater unimpacted Leachate well	0.20	0.65	0.65
	Leachate cistern Leachate well	0.43	0.51	0.65
<i>Pielou’s</i>	All groups	4.11	0.25	NA
<i>E</i>	Groundwater impacted Groundwater unimpacted	1	0.32	0.48
	Groundwater impacted Leachate cistern	1.80	0.18	0.36
	Groundwater impacted Leachate well	0.20	0.65	0.65
	Groundwater unimpacted Leachate cistern	1.80	0.18	0.36
	Groundwater unimpacted Leachate well	0.20	0.65	0.65
	Leachate cistern Leachate well	2.33	0.13	0.36

None of the tests were significant at a *p* value of 0.05.

2.3.3 Beta diversity

The PCoA plots using weighted and unweighted UniFrac distances based on 16S rRNA gene amplicon ESVs show separation of the samples by type (Figure 2.6). The samples for the composite leachate cistern group together in both PCoA analyses, which agrees with their shared similarity at the phylum level of diversity (Figure 2.2) but does not match the prevalence data for the ESVs, as there are no shared ESVs between CLC_T1 and CLC_T2. The distance matrix values for the composite leachate cistern samples less distant to each other than to the leachate and groundwater samples, which contributes to their grouping in the PCoA. The unimpacted groundwater well (GW2) is separated from the impacted groundwater well (GW1) and the leachate wells (LW) in both PCoA analyses (Figure 2.6). In the unweighted UniFrac plot, the impacted groundwater well is close to LW3, the leachate well closest to GW1. This relationship is not seen in the weighted UniFrac, when ESV abundances are included in the analysis. In the unweighted UniFrac plot, axes 1 and 2 explain

27.2% and 18.2% of differences in sampled communities respectively. More of the variation between samples is explained in the weighted UniFrac plot, with axes 1 and 2 explaining 50.2% and 17% of differences, respectively. As a result, the inclusion of abundance data in the weighted UniFrac analysis increases the explained diversity on axes 1 and 2 by a combined 21.6%, suggesting that presence/absence and phylogenetic distance data in the unweighted UniFrac are not sufficient to resolve the differences in beta diversity between sites in two dimensions. Notably, three of the top five most abundant ESVs are present in only one sample (ESV_1 in LW3, ESV_2 in GW2, and ESV_5 in GW1) while ESV_4 is present in the three leachate wells and CLC_T2 and ESV_3 is present in both samples for CLC_T1. The differences in abundance and overlap of ESVs between sites contributes to the separation of the samples in the PCoA plots.

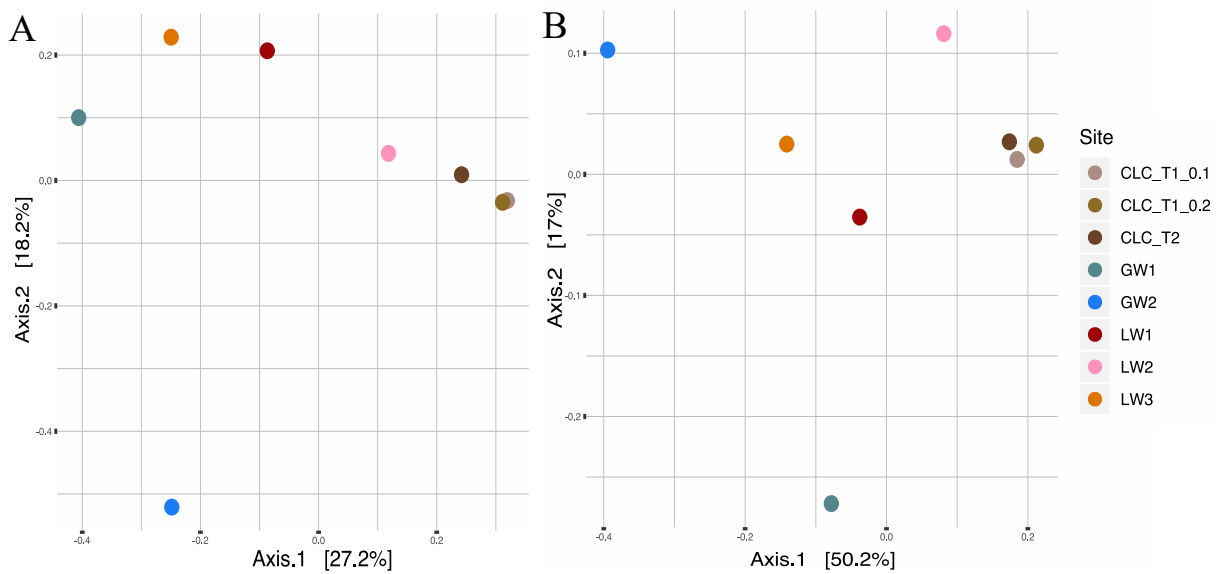


Figure 2.6: PCoA using unweighted (A) and weighted (B) UniFrac distances based on 16S rRNA amplicon ESVs for all sites.

2.3.4 Diversity of exact sequence variants (ESVs)

Although phylum level diversity is relatively consistent across the composite leachate cistern, leachate well, and GW1 samples, the diversity at the ESV level varies. The overwhelming majority of ESVs identified to the top 25 phyla are present in only a single sample (Figure 2.7) with only 37 ESVs shared across two or more samples (Figures 2.7 and 2.8). Only two of the ESVs, ESV_4 (Phylum: Bacteroidota, Genus: DMER64) and ESV_18 (Phylum: Cloacimonadota, Genus: LNR_A2-18) are present across four samples: LW1, LW2, LW3, and CLC_T2 (Figures 2.7 and 2.8). None of the ESVs from the two CLC_T1 samples are shared with any of the other sites. Of the ESVs present across multiple sites, only 9 are present within the top 10% most abundant ESVs for the study site (Figure 2.9). Twenty of the 37 multi-sample ESVs are shared between LW2 and CLC_T2, suggesting that the chemical composition of the leachate at these sites may be similar or that LW2 may contribute a larger proportion of the leachate to the composite leachate cistern. Interestingly, GW2, the unimpacted well, shares more ESVs in common with the leachate wells and CLC_T2 than GW1, the impacted well, with eight shared ESVs to one ESV, respectively. This is in contrast to the phylum-level differences seen for GW2 compared to landfill samples, and suggests there is some interconnectivity even between the unimpacted groundwater and the leachate – possibly with groundwater infiltrating the landfill. The one shared ESV between LW3 and GW1 is ESV_8 (Phylum: Proteobacteria, Genus: *Shewanella*) and its shared presence drives the position of GW1 in the principal component analysis (Figure 2.12). Similarly, the abundance of ESV_4 (Phylum: Bacteroidota, Genus: DMER64) is driving the separation of LW1 from the others (Figure 2.). The shared ESVs between GW2 and LW2 contribute to the close clustering of these samples in the PCA plot (Figure 2.10). The high number of ESVs present at only one site may be akin to the rare operational taxonomic units described by Köchling *et al.* (2015) and Cardinali-Rezende *et al.* (2016), where rare or non-prevalent organisms were hypothesized to act as seeder or starter communities during environmental changes or disturbances. Although there are some shared ESVs between the study samples, the vast majority (98.8%) of the ESVs are limited to one sample, suggesting that each site within the landfill hosts its own unique community. In addition, the CLC comparison across one week, with no shared ESVs between time points, suggests a rapid turnover of these populations within the landfill, despite the observed phylum-level similarities among sites and over time.



Figure 2.7: Prevalence of 16S rRNA amplicon ESVs across sampling sites. Only ESVs with total counts greater than three were included in this graph. Only ESVs identified to the top 25 most abundant phyla as per Figure 2.2 are included, to allow a comparison to the phylum level diversity. Data from CLC_T1 0.1 and 0.2 μm filters were combined, because these samples represent the same biomass from the same site and sample date. The majority of ESVs occur in only one sample. Only two ESVs are present in four samples, ESV_4 (Phylum: Bacteroidota, Genus: DMER64) and ESV_18 (Phylum: Cloacimonadota, Genus: LNR_A2-18).

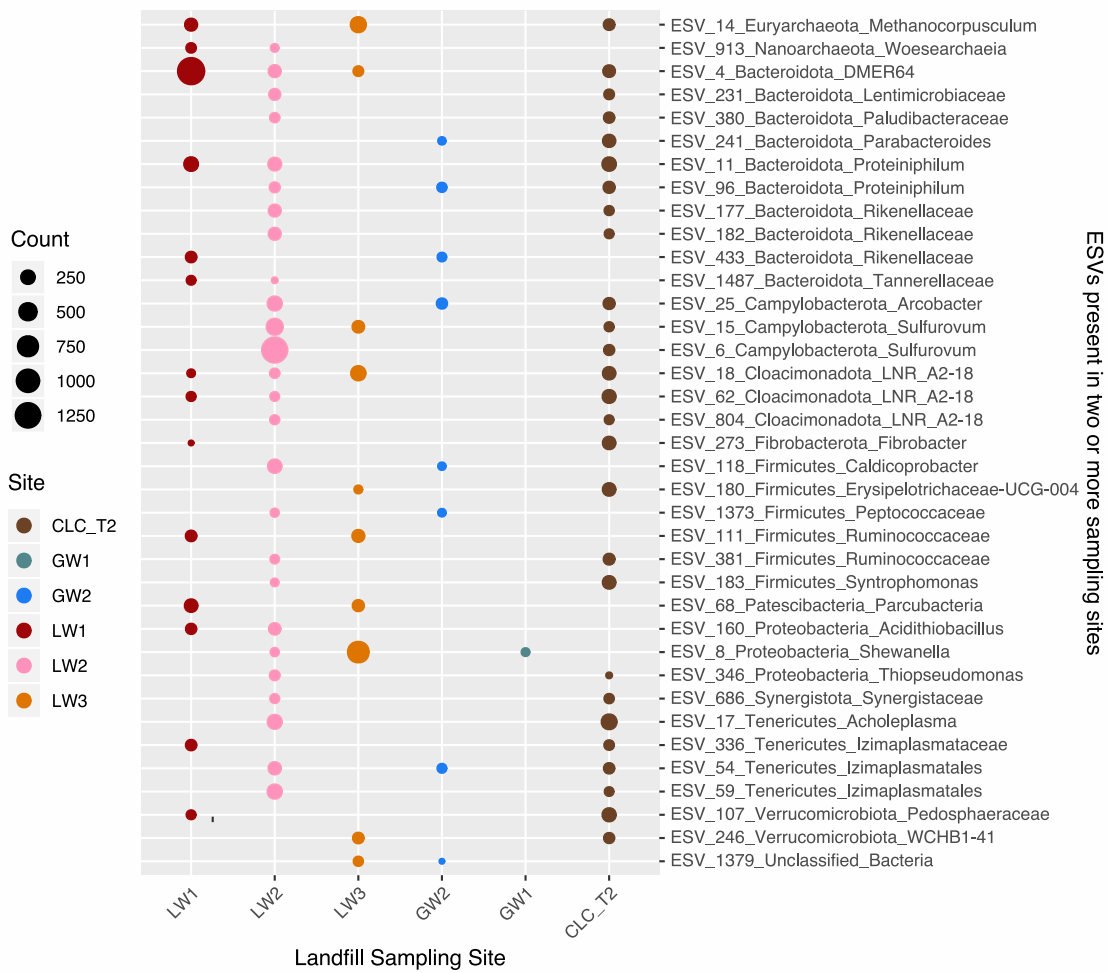


Figure 2.8: Abundance and prevalence of 16S rRNA gene amplicon ESVs present in two or more sampling sites. The CLC_T1 0.1 and 0.2 μ m filter samples were excluded as they do not share any ESVs with other sites and show numerous overlaps with each other, having come from the same filtered material. ESV numbers correspond to count abundances with ESV_4_Bacteroidota_DMER64 having the highest abundance based on counts for the profiled ESVs (#4 in count-based abundance for all ESVs).

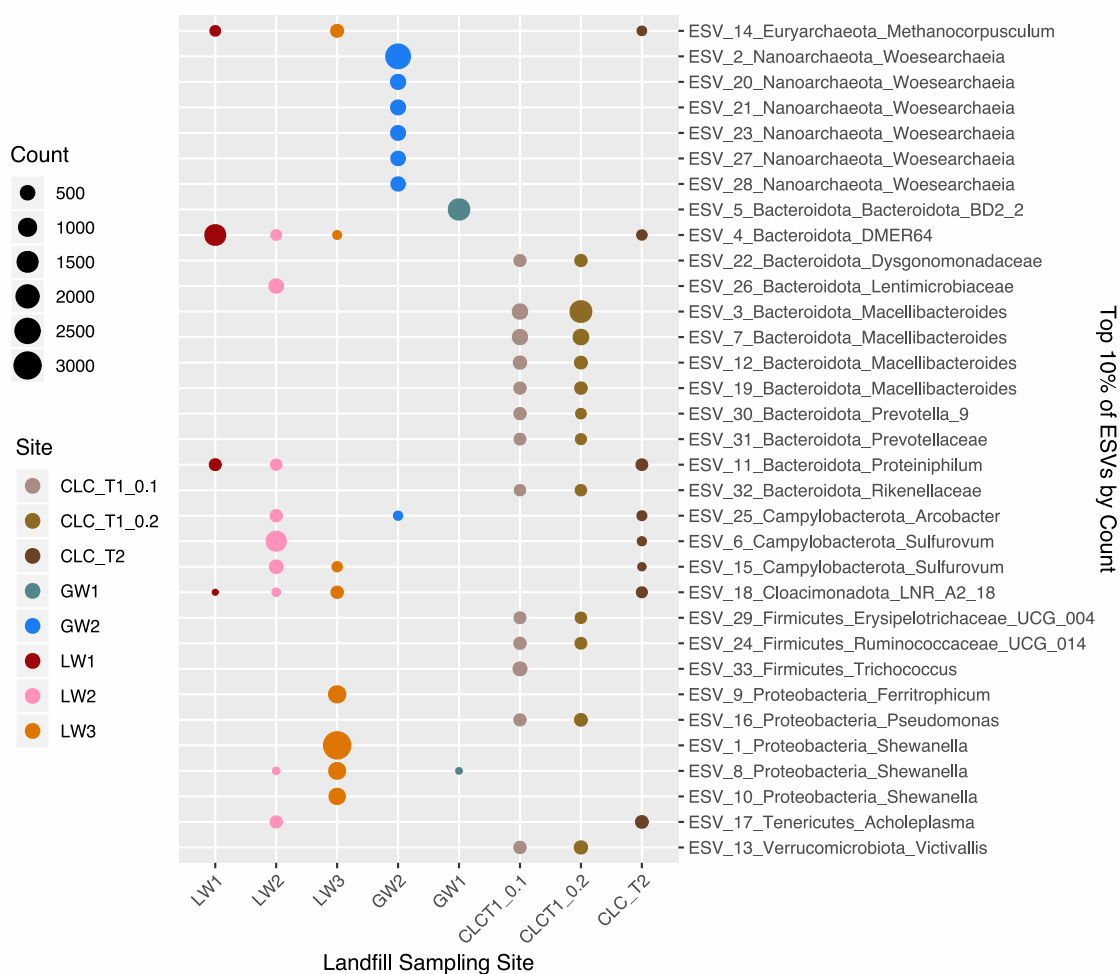


Figure 2.9: Top 10% most abundant 16S rRNA gene amplicon ESVs by count. The 33 ESVs are affiliated with nine phyla, all of which are included in the 25 phyla with greater than 1% relative abundance (Figure 2.2). ESV ID numbers correspond to their overall abundance, with ESV_1_Proteobacteria_Shewanella having the highest count when all samples' counts are summed.

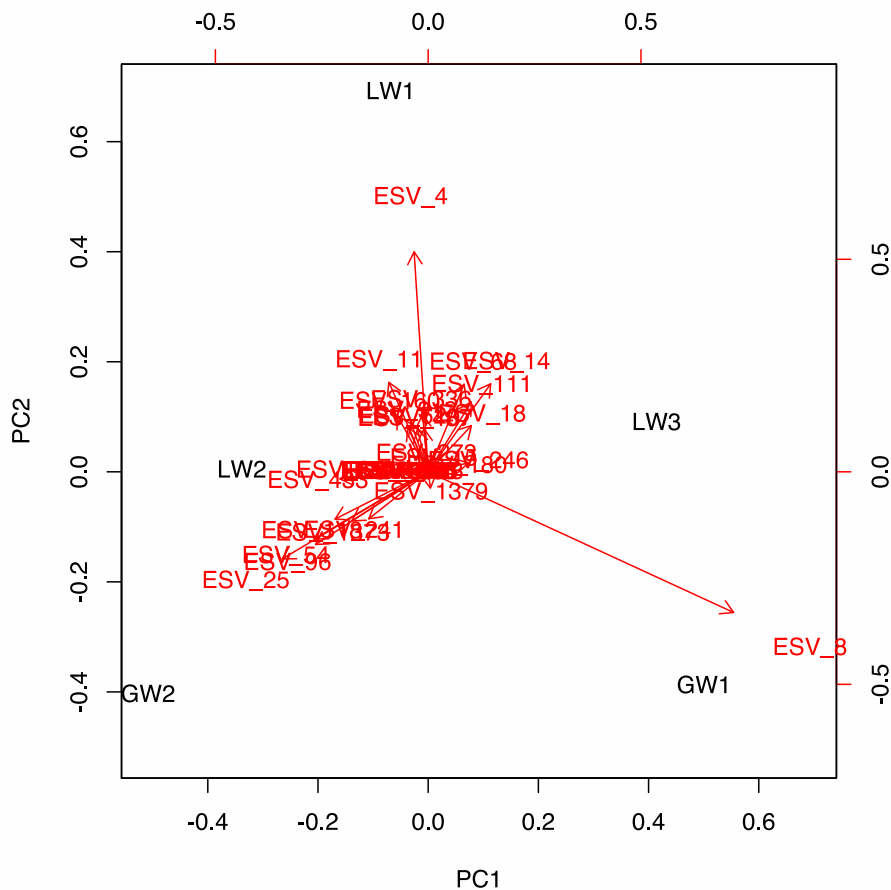
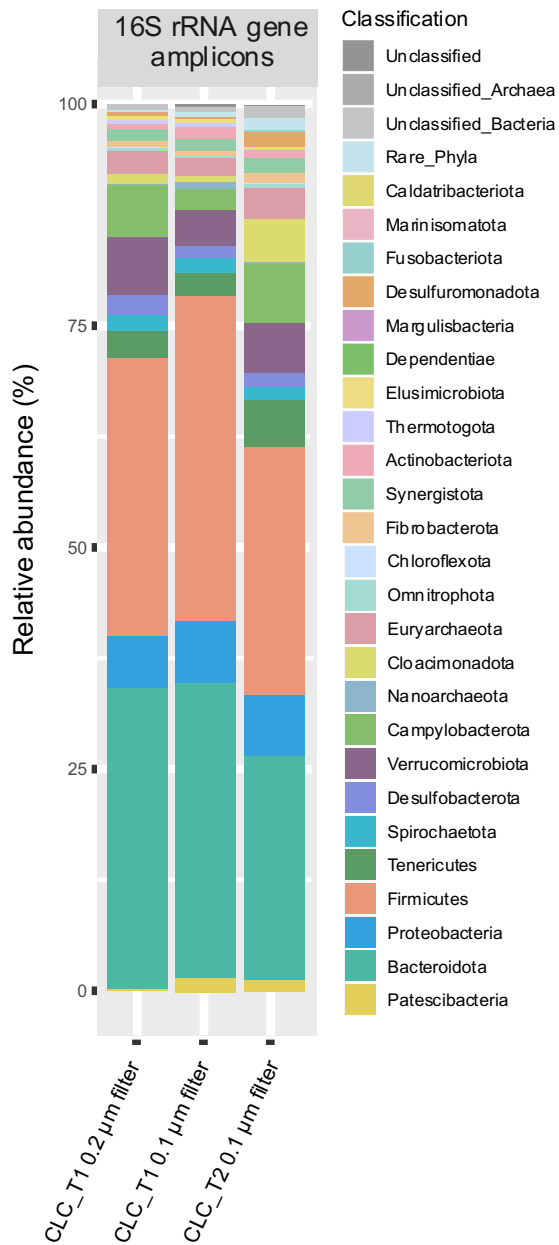


Figure 2.10: Principal component analysis for 16S rRNA gene amplicon ESVs present in two or more sites. ESV count data was Hellinger transformed. Environmental data were not available for the composite leachate cistern site and so these samples were excluded from this analysis to allow comparison between PCAs. PC1 explains 41.9% of the variation and PC2 explains 30.1% of the variation.

2.3.5 Temporal diversity

We sequenced three samples from the composite leachate cistern: two from July 14 (CLC_T1_0.1 and CLC_T1_0.2) and one from July 20 (CLC_T2), from which we examined the change in microbial diversity over time. Although these three samples show similar phylum level profiles (Figure 2.11), there is no overlap in ESVs between the CLC_T1 and CLC_T2 samples. This demonstrates that there was complete or near-complete turnover of the leachate cistern microbial community within a week,

and implies a similar level of turn-over within the leachate microbial community that is included in the catchment area for this composite leachate cistern. Alternatively, this turn-over of ESVs could imply rapid evolution within these microbial communities with a sufficient level of mutations between generations at time point 1 and time point 2 for the ESVs to differ. This is an unprecedented finding, as most landfill studies sampled their sites at only one time point (Sawamura *et al.*, 2010; Stamps *et al.*, 2016) or sampled landfill sites of different ages (Köchling *et al.*, 2015; Song *et al.*,



2015). The studies that include multiple landfills of different ages, such as Song *et al.* (2015), show differences in community structure correlating to the age of waste. Cardinali-Rezende *et al.* (2016) showed that microbial community composition changes in a landfill bioreactor as it progresses from start up to a steady state, but these stages were separated by two years. Prior surveys focusing on age differences in years will not resolve fine scale temporal changes in landfill microbial communities. From our results, the consistency of phylum level diversity across time and across landfills may be less meaningful than previously thought given our observation of rapid and near-complete changes in the microbial communities.

Figure 2.11: Composite Leachate Cistern relative abundance of bacterial and archaeal phyla present at greater than 1% abundance based on 16S rRNA gene amplicons. Both time point 1 (0.1 and 0.2 µm filters) and time point 2 (0.1 µm filter) are included. The phylum level diversity signature is consistent between the three samples with shifts in percent relative abundance for some phyla.

2.3.6 Microbial diversity of the groundwater wells

The two groundwater wells allow for comparison between a natural groundwater environment and a leachate contaminated environment. GW1 is closer to the active landfill and leachate from the region around LW3 is leaking into the groundwater near GW1 (Figure 2.1). GW2, on the hand, is further from the active landfill and is embedded in a region of the aquifer that shows no evidence of contamination from the landfill leachate (Figure 2.1, top left corner). Landfill leachate solubilizes a number of potentially harmful chemicals (Stamps *et al.*, 2016) and a number of metals and volatile compounds are detected at GW1 but not GW2 (Figure 2.12). The higher concentration of 1,4-dioxane, vinyl chloride, and various chloroethane compounds in GW1 in comparison to the leachate wells (Figure 2.12) may be due to microorganisms capable of degrading those compounds in the leachate wells not surviving to colonize the aquifer near GW1. Lu *et al.* (2012) showed that when landfill leachate contaminates groundwater, the landfill microbes are unable to survive in the more dilute groundwater, and the addition of chemicals from the leachate negatively impacts the native groundwater microorganism diversity.

Differences in the groundwater microbial community are evident throughout the data for GW1 and GW2. GW1 has a high abundance of Patescibacteria while also sharing a more similar phylum-level profile to the leachate wells than to GW2 (Figure 2.2). The sample from GW2 had insufficient microbial biomass for metagenomic sequencing, but we can see from the 16S rRNA gene amplicon sequencing that GW2 has a distinct microbial community compared to all other sites (Figure 2.2). GW2 has a higher relative abundance of Nanoarchaeaota (25%) and Omnitrophota (15%) (Figure 2.2). The difference in microbial community composition between GW1 and GW2 is also reflected in their alpha diversity metrics. GW1 has the lowest Shannon index (Figure 2.4) of the eight samples as well as the lowest Faith's phylogenetic diversity (Figure 2.5). A lower richness and evenness are expected in GW1 in light of the Lu *et al.* (2012) findings, as the mixing of leachate and groundwater creates a suboptimal environment for microorganisms adapted to either environment.

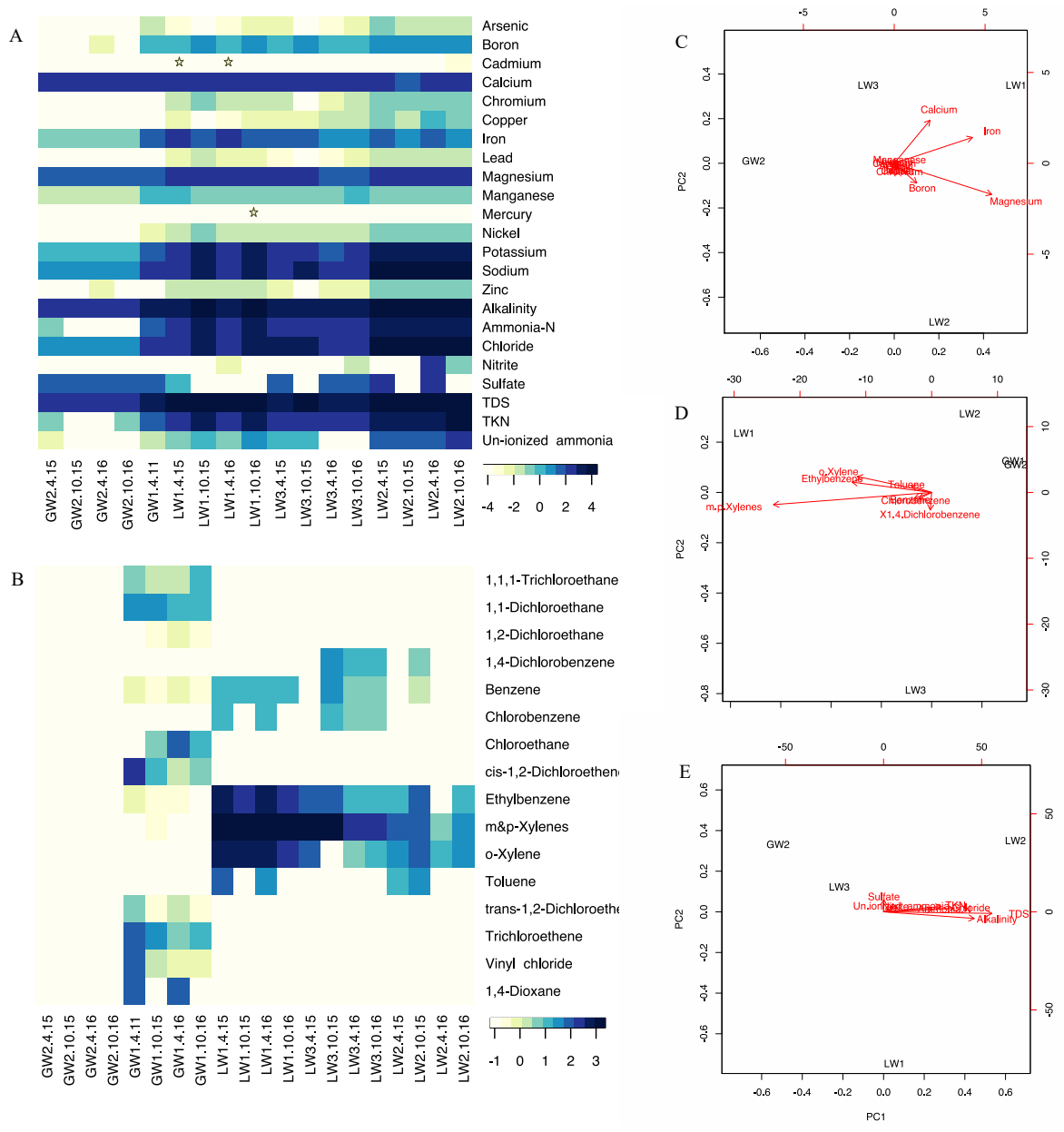


Figure 2.12: Environmental variation between landfill and aquifer sites for non-volatiles (mg/L) and volatiles (µg/L). A) Heat map of non-volatile compound concentrations and other site parameters (log10 transformed) across four dates for the three leachate wells and the unimpacted groundwater well. Only one date (April 2011) was available for the impacted groundwater well (GW1). Alkalinity and total dissolved solids (TDS) are high across sites, although lower in GW2. Nitrogen compounds (ammonia-N, total Kiejdahl nitrogen (TKD), and un-ionized ammonia) are higher in the leachate wells. B) Heat map of volatile compound concentrations (log10 transformed)

across four dates for the leachate and groundwater wells. C) Principal component analysis of metal concentrations at the three leachate wells and the unimpacted groundwater well. Metal concentration data was square root transformed. Sodium and potassium concentrations were considerably higher in LW2, so these compounds were removed to allow visualization of separation by other variables. PC1 explains 79.6% of variation and PC2 explains 18.4% of variation. D) Principal component analysis of volatile compound concentrations at the three leachate wells and the unimpacted groundwater well. Volatile compound concentration data were square root transformed. PC1 explains 97.4% of variation and PC2 explains 2.4% of variation. E) Principal component analysis for other geochemical parameters of the site, derived from measurements at the three leachate wells and the unimpacted groundwater well. Measurements (mg/L and $\mu\text{g/L}$) were square root transformed. PC1 explains 99% of the variation. Stars indicate low levels of cadmium and mercury not visible on the heatmap that may be biologically relevant.

2.3.7 Analysis of geochemical parameters

The statistical power available for analysis of geochemical parameters in the landfill was limited by the availability of data. Non-volatile compound concentrations were only available for four sites and volatile compound concentrations for five sites (Table 2.1). Non-volatile and volatile compound concentrations did vary significantly between sites ($p < 9.14 \times 10^{-14}$ and $p < 2 \times 10^{-16}$, respectively) (Tables 2.4 and 2.5) with a large range between sites for several non-volatile and volatile compounds (Figure 2.12). The date of measurement was not significant for either volatiles or non-volatiles ($p = 0.56$ and $p = 0.73$, respectively) (Tables 2.4 and 2.5). The April and October 2016 measurements for the PCA analysis were averaged for the non-volatiles and volatiles. Sodium and potassium were treated as outliers because they were at much higher concentrations in LW2 (Table 2.1) and their variation masked any differences in the other compounds in the analyses. In the PCA, calcium, iron, magnesium, and to a lesser degree, boron contributed to the difference between the leachate wells and GW2, pulling the leachate wells away from GW2 (Figure 2.12C). For the volatile compounds, nearly all of the observed variation is explained by PC1 (97.4%), largely due to the punctuated presence of m- & p- xylenes in LW1 and LW3, and o. xylenes and ethylbenzene in LW1 (Figure 2.12B and D). GW1 contains a number of volatile compounds not present in the leachate wells, potentially due to differences in the microbial functional capacity for degradation of these compounds between the leachate communities and the groundwater community (Figure 2.12B). Other measured metrics, such as total dissolved solids and chloride, are present at higher levels in LW1 and LW2, contributing to

those samples segregating from GW2 and LW3 and from each other (Table 2.1 and Figure 2.12). LW1 is also the only well out of these four without detectable sulfate, which also contributes to its segregation. From this, we can see significant variations exist between the groundwater and leachate wells' chemical compositions, however, data from additional sites is needed to determine if any of these chemical parameters are driving overall microbial heterogeneity.

Table 2.4: ANOVA results for all volatile compound concentrations from October and April of 2016 for the three leachate wells and the two groundwater wells.

<i>Variables</i>	<i>P value</i>
Date	0.56
Site	< 9.14e-14*
Volatiles	< 2e-16 *
Date : Site	0.06
Date : Volatiles	0.48
Site : Volatiles	< 4.9e-16 *

Concentrations were log transformed for the calculations. * indicates statistical significance at a p-value of 0.05.

Table 2.5: ANOVA results for non-volatile compound concentrations from October and April of 2016 for the three leachate wells and the unimpacted groundwater well.

<i>Variables</i>	<i>P value</i>
Date	0.73
Site	< 2e-16 *
Non-volatiles	< 2e-16 *
Date : Site	0.01*
Date : Non-volatiles	0.19
Site : Non-volatiles	< 2e-16 *

Concentrations were log transformed for the calculations. * indicates statistical significance at a p-value of 0.05.

Although the available geochemical data limit our ability to draw conclusions about the effects of geochemical parameters on the overall microbial heterogeneity of the landfill, some associations can be made for particular ESVs. Approximately 41.7% of the ESVs were classified to the genus level, either with an accepted genus name (27.5%) or other genus level grouping (14.2%). For the ESVs identified to known genera, some interactions with the site geochemistry can be inferred based on current knowledge of those genera and their metabolisms. Three genera with either high abundance and/or prevalence that are known to interact with particular compounds are discussed below as examples of bacteria that are likely affected by the site geochemical heterogeneity.

2.3.7.1 Genus *Sulfurovum*

One genus that is abundant in the landfill and has potential contaminant degradation capabilities is *Sulfurovum*. Two abundant ESVs, ESV_6 and ESV_15, were classified to the *Sulfurovum* genus and are present in LW2, LW3, and CLC_T2 (Figure 2.9). Three other ESVs were also classified as *Sulfurovum* at lower abundances bringing the total abundance for the genus to 1.01%. Characterized *Sulfurovum* bacteria are sulfur oxidizing bacteria and have been previously found in contaminated groundwater where it appears to participate in the communal anoxic degradation of benzene (Rakoczy et al., 2013; Vogt et al., 2011). Benzene is present in LW3 and has historically been present in LW2 (Table 2.1, Figure 2.12). The abundance of *Sulfurovum* at impacted or historically impacted sites may suggest that this community is capable of degrading benzene but genome analyses would be needed to determine if any community members possess the metabolic capabilities to degrade benzene.

2.3.7.2 Genus *Proteiniphilum*

One genus that is abundant and diverse in the landfill and associated with the degradation of organic waste is *Proteiniphilum*. Both ESV_11 and ESV_96 were classified to the genus *Proteiniphilum*, each present at three sites (Figure 2.8). An additional 25 ESVs were classified as *Proteiniphilum* at lower abundances bring the total abundance for the genus to 1.53%. Bacteria in this genus are proteolytic (Chen and Dong, 2005) and their prevalence and abundance suggests a potential enrichment of organic material at these sites. Both ESV_11 and ESV_96 are present in LW2, which has higher ammonia-N and total Kjeldahl nitrogen (Table 2.1) than all other sites. ESV_11 is also present in LW1, which does not have the same level of nitrogen species as LW2, but which does have elevated nitrogen in comparison to the groundwater wells and LW3. This genus was also identified at high abundance in the organic waste bioreactors studied by Cardinali-Rezende *et al.* (2016), which also

exhibited high levels of ammonia. There appears to be a correlation between high levels of nitrogen and abundant *Proteiniphilum* in landfill systems. It is likely that the populations associated with these ESVs are contributing to protein degradation in the landfill, but this would need to be confirmed with further analyses.

2.3.7.3 Genus *Ferritrophicum*

Another genus that was relatively abundant in the landfill and associated with contaminant degradation is *Ferritrophicum*. ESV_9 was classified to the genus *Ferritrophicum*, and was detected at high abundance in LW3 (Figure 2.). Three other ESVs in LW3 were also identified as *Ferritrophicum*, further increasing the total abundance of this genus (0.45%). The only other recorded observation of *Ferritrophicum* species was by Weiss *et al.* (2007) who isolated the Fe(II)-oxidizing bacteria from the rhizosphere of wetland plants. There is iron present in each of the landfill leachate wells likely in the form of Fe(II), but it is at the lowest concentration in LW3 (Table 2.1). These *Ferritrophicum* bacteria may be important for iron cycling in LW3 and pose an interesting avenue for future study to understand iron cycling in the landfill and expand our knowledge of this rare genus.

2.4 Conclusions

The phylum level profiles for the composite leachate cistern, leachate wells, and GW1 are consistent with previous landfill microbial community studies, with Bacteroidota, Firmicutes, and Proteobacteria among the most abundant phyla. We uncovered an abundance of Patescibacteria in the landfill with the use of metagenomics sequencing, a group that may have been missed in previous studies relying on 16S rRNA gene amplicon sequencing. Richness and evenness are high for all of the studied samples, with the impacted groundwater well, GW1, being notably, but not significantly, lower than the other samples. The samples separate by sample type for beta diversity on the PCoA plots, which, along with the prevalence patterns of the ESVs, suggests that the microbial heterogeneity is largely driven by lower order taxonomic differences between samples. Additionally, the absence of shared ESVs between CLC_T1 and CLC_T2, with only a week between sampling, suggests that there is a rapid turnover of microbial communities at the population level within the landfill.

There are differences in the chemical composition of the leachate and groundwater between sites that could potentially influence microbial heterogeneity. Possible compounds of interest include sodium, potassium, calcium, iron, magnesium, and boron for the non-volatile compounds, and m- & p- xylenes, o. xylenes, ethylbenzene, vinyl chloride and 1,4-dioxane for the volatile compounds. The

genera *Sulfurovum*, *Proteiniphilum*, and *Ferritrophicum* are relatively abundant in the amplicon data, and have predicted roles in nutrient and contaminant cycling in the landfill but this would need to be confirmed with genomic analyses. A greater sample size of sites with paired genetic data and chemical data is required to further elucidate the effects of leachate chemical composition on landfill microbial community structure and composition.

Chapter 3

A phylogenetic and metabolic study of the group Tenericutes

3.1 Introduction

3.1.1 A brief history of the Tenericutes and Erysipelotrichia

Tenericutes, also called Mollicutes, are a group of bacteria that have historically confounded researchers in regard to their classification and phylogeny. Organisms within the Mycoplasma group were first discovered in 1898 and were classed with other cell wall-less prokaryotes (Freundt, 1975). Edward and Freundt (1956) assigned the known species of mycoplasmas to the family Mycoplasmataceae and the genus *Mycoplasma*. In 1957, the group was placed as the order Mycoplasmatales within the now defunct Schizomycetes class (Edward, 1971; Freundt, 1975). Early cultivated Mollicutes, including mycoplasmas, required sterols for growth and, in combination with their lower GC content, this led to their designation as a new class within the Firmicutes in 1967 (Edward, 1971; Freundt, 1975). Murry (1984) proposed the phylum Tenericutes for the Mollicutes, separating them from the Firmicutes on the basis of the Mollicutes' distinctive wall-less cell envelope and their inability to synthesize cell wall precursors, including muramic and diaminopimelic acids (Ludwig et al., 2010; Mollicutes, 1995). This designation was also supported by phylogenies using several conserved molecular markers, including elongation factor Tu and RNA polymerase (Ludwig and Schleifer, 2005). Thus, the Tenericutes were designated as a phylum, though this taxonomy remained controversial.

After the Mollicutes were elevated to the phylum Tenericutes, the class Erysipelotrichia, containing the order Erysipelotrichales and family Erysipelotrichaceae, was created within the phylum Firmicutes (De Vos *et al.*, 2009). Erysipelotrichaceae was classified as a family by Verburg *et al.* (2004), with *Erysipelothrix inopinata* as the type species. The Erysipelotrichia are similar to other Firmicutes phenotypically, including possession of a cell wall and low GC content, but have low rRNA identity with the rest of the Firmicutes (Davis *et al.*, 2013; De Vos *et al.*, 2009). Some members of the Erysipelotrichia are free-living, environmental microbes, such as *Catenisphaera adipataaccumulans*, which was isolated from an organic waste bioreactor (Kanno *et al.*, 2015). Others are animal pathogens, such as the swine pathogen *Erysipelothrix rhusiopathiae* (Ogawa *et al.*, 2011). Several Erysipelotrichia species have undergone genome reductions independently, and have lost metabolic genes, including those for synthesis of purines and pyrimidines, fatty acids, arginine, and

tryptophan, which is a similar metabolic profile to many of the Tenericutes (Davis *et al.*, 2013). Early phylogenetic work placed some members of the Erysipelotrichaceae closer to the Tenericutes, based on 16S and 23S rRNA phylogenies, but the limited sequencing of members of the Erysipelotrichia prevented the use of protein phylogenies (De Vos *et al.*, 2009).

In recent years, a number of Erysipelotrichia species have had their genomes sequenced, allowing for more in-depth phylogenetic comparisons between the Tenericutes and the Erysipelotrichia (Davis *et al.*, 2013; Ogawa *et al.*, 2011; Ramasamy *et al.*, 2013). Davis *et al.* (2013) analyzed phylogenies for the Tenericutes and Erysipelotrichales based on 16S rRNA genes, 23S rRNA genes, ribosomal proteins, and aminoacyl-tRNA synthetase proteins. The Erysipelotrichia placed outside of the Tenericutes when looking at the 16S rRNA gene tree, but placed within the Tenericutes between the “Acholeplasma” group and the other three groups in the 23S rRNA gene and the ribosomal protein trees (Figure 3.1). Most recently, the microbial taxonomy revisions suggested by Parks *et al.* (2018) as part of the Genome Taxonomy Database (GTDB) group the Tenericutes and the Erysipelotrichia within class Bacilli, in the phylum Firmicutes (Figure 3.2). The Tenericutes orders Acholeplasmatales and Izimaplasmatales are maintained in the GTDB taxonomy, whereas the Entomoplasmatales has been included within the Mycoplasmatales. The Erysipelotrichia order Erysipelotrichales has been maintained. Three currently unnamed order-level groups also place phylogenetically with the Tenericute orders and the Erysipelotrichales. Classification of the Tenericutes and Erysipelotrichia has a controversial history based on phenotypic and molecular techniques for classification. As more genomic data becomes available for these groups, the underlying relationships and phylogeny become clearer. For the purpose of this thesis, the term “Tenericutes” is referring collectively to the “Hominis”, “Pneumoniae”, “Spiroplasma”, and “Acholeplasma” groups (Figure 3.3), separate from the Erysipelotrichales, as we attempt to clarify the relationships between these groups.

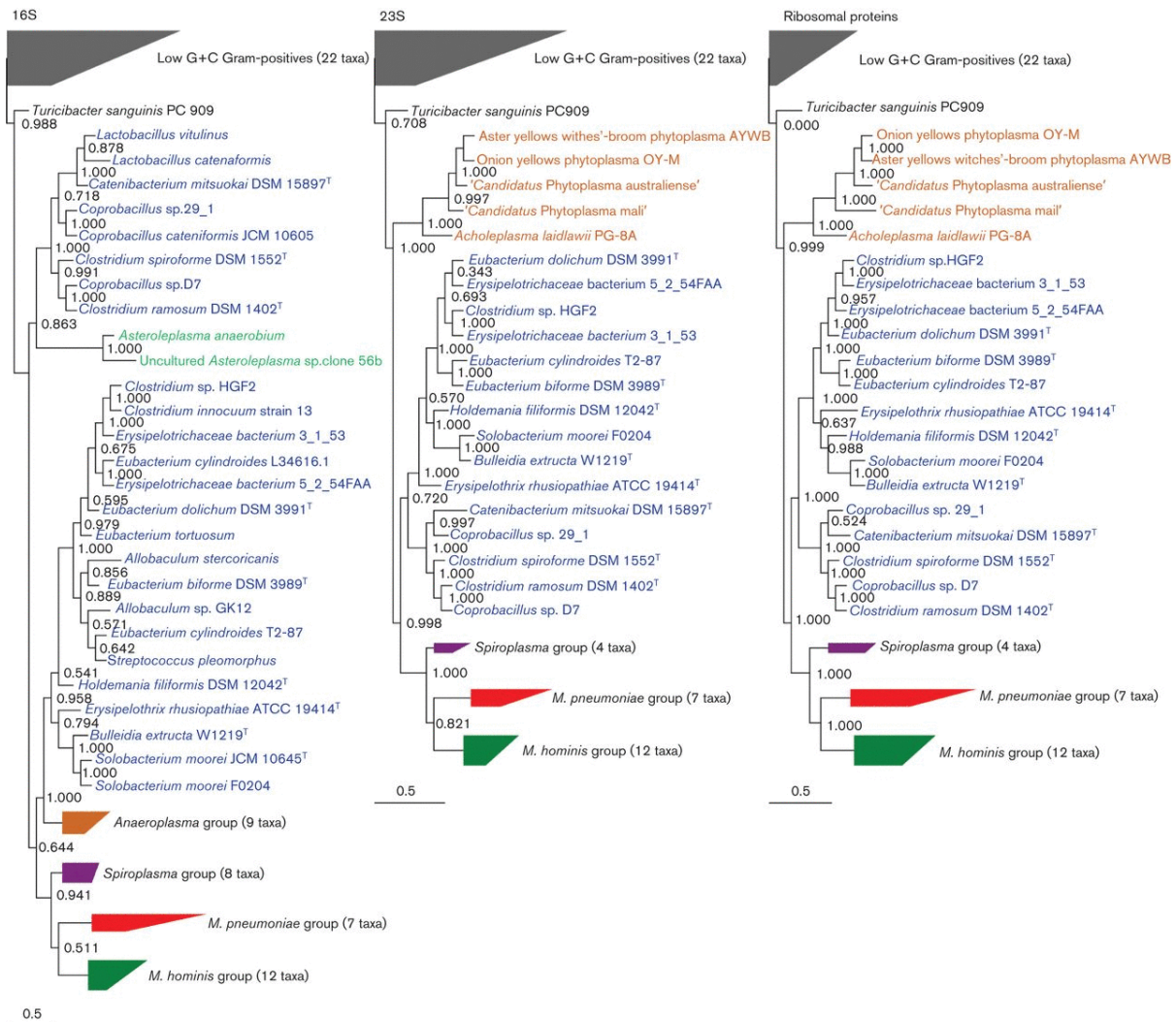


Figure 3.1: Maximum-likelihood phylogenies of the Erysipelotrichia and Tenericutes using 16S (left) and 23S (middle) rRNA genes and 34 universal ribosomal proteins (right), taken from Davis et al. (2013). “Anaeroplasma” group, also called the “Acholeplasma” group, is in orange, the Erysipelotrichia group is expanded in blue, the “Spiroplasma”, “Pneumoniae”, and “Hominis” groups are in purple, red, and green respectively. The outgroup of low GC Gram-positive bacteria is in grey. Groups that have been collapsed are depicted by wedges with the top and bottom edges describing longest and shortest branches lengths within the group, respectively. Bootstrap values are proportional support out of 1000 replicates.

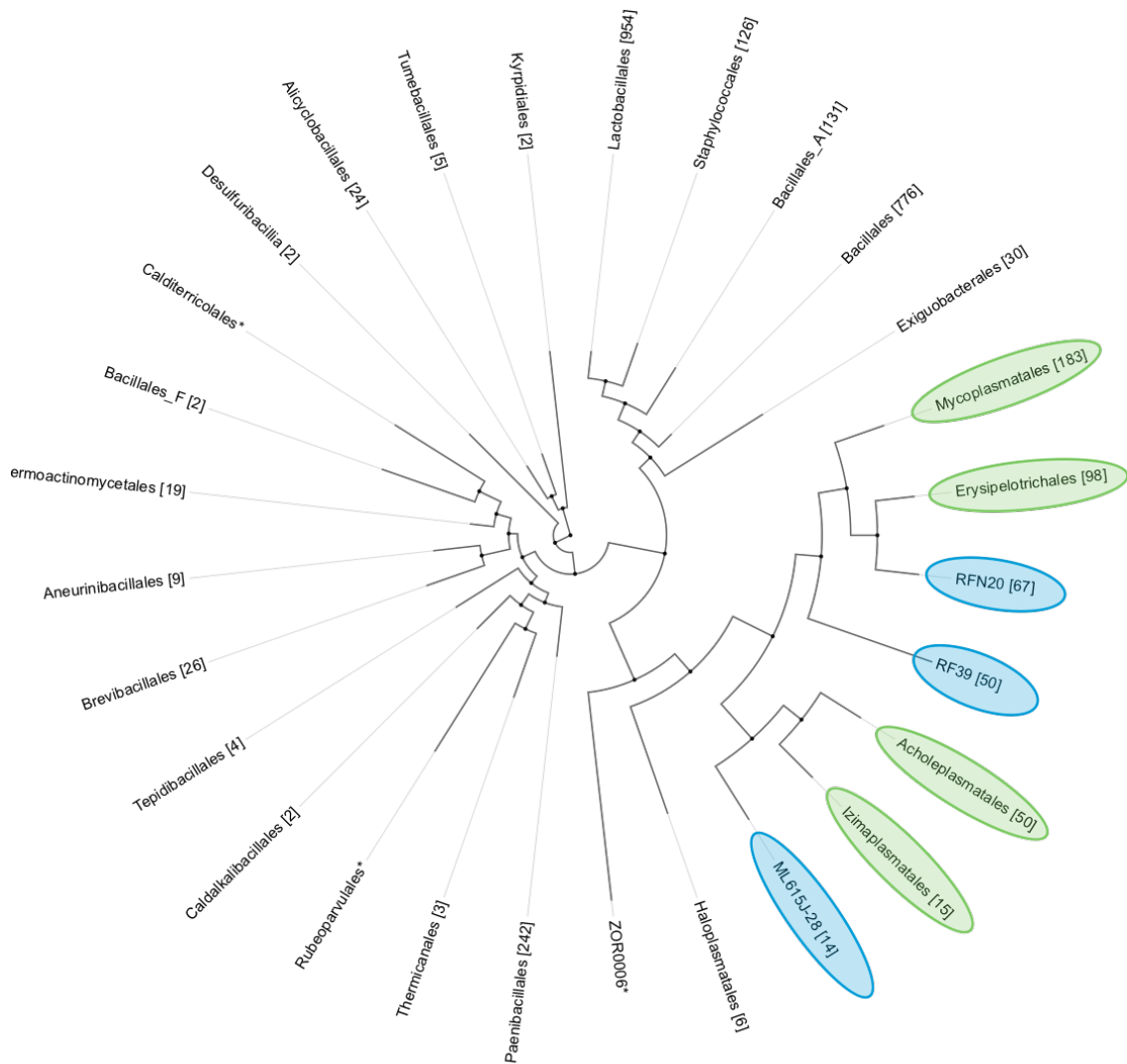


Figure 3.2: Annotree (Mendler et al., 2018) view of class Bacilli based on the GTDB taxonomy (Parks et al., 2018). The Tenericute orders Mycoplasmatales, Acholeplasmatales, and Izimaplasmatales, as well as the Erysipelotrichales are circled in green and the currently unclassified groups related to the Tenericutes and included in the ribosomal protein phylogeny for this research are circled in blue.

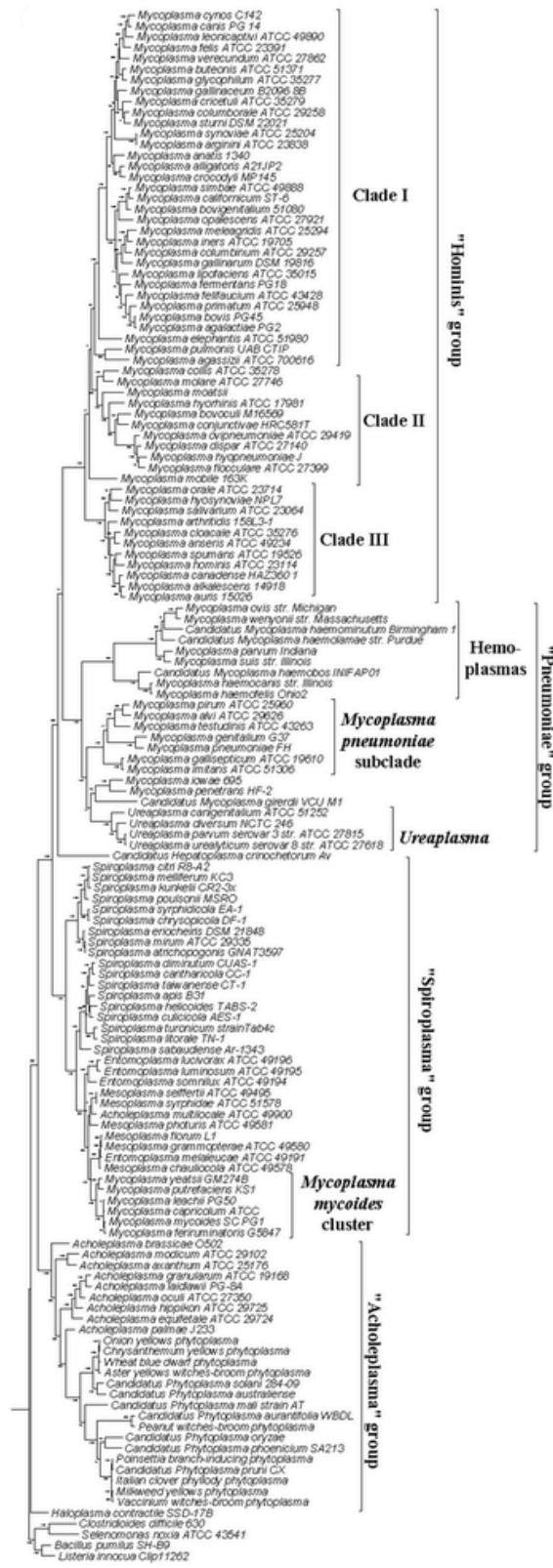


Figure 3.3: Maximum likelihood phylogeny of the Tenericutes using 45 conserved ribosomal proteins, taken from Gupta et al. (2018). The four main groups within the Tenericutes are labelled the “Hominis”, “Pneumoniae”, “Spiroplasma”, and “Acholeplasma” groups. The tree was rooted with members of the Firmicutes. The Erysipelotrichia were not included in this analysis.

3.1.2 The Tenericute cell

The Tenericutes are small, cell wall-less bacteria with diameters in the range of 300 nm for spherical cells, with helical filaments as small as 200 nm in diameter (ICSB Subcommittee on the Taxonomy of Mollicutes, 1995). The relatively small cell size is matched by their genomes, which are among the smallest for autonomously replicating organisms (Lazarev *et al.*, 2011). Without a cell wall, Tenericutes cells come in many shapes, including spherical, coccoid, filamentous, and ring and dumbbell forms, with some Tenericutes being pleomorphic (Joblin and Naylor, 2002; Mollicutes, 1995). Tenericutes cells can be motile, as seen in members of the *Spiroplasma*, or non-motile, as seen in members of the *Anaeroplasma* (Joblin and Naylor, 2002; Lazarev et al., 2011).

Tenericutes are capable of forming colonies

on solid media, with colony appearances varying between motile and non-motile species (ICSB Subcommittee on the Taxonomy of Mollicutes, 1995). Most non-motile species form fried-egg-like colonies whereas motile species will form diffuse colonies with a number of satellite colonies (ICSB Subcommittee on the Taxonomy of Mollicutes, 1995). These cell and colony characteristics helped direct cultivation trials aimed at isolating landfill-associated Tenericutes.

3.1.3 Two Tenericutes clades

There are two clades within the Tenericutes that resolve phylogenetically as two polyphyletic groupings around the Erysipelotricales (Figure 3.1) (Davis *et al.*, 2013), and whose members share distinct characteristics. The first clade is the “Acholeplasma” or “Anaeroplasma” group, which contains the *Anaeroplasma*, *Acholeplasma*, *Candidatus Phytoplasma*, and for the purpose of this thesis, *Candidatus Izimaplasma* (Figures 3.1 and 3.3) (Davis *et al.*, 2013; Gupta *et al.*, 2018). The second clade is the Mycoplasmatales order, as per Parks *et al.* (2018), which includes all of the *Mycoplasma*, *Mesoplasma*, *Spiroplasma*, *Entomoplasma*, and *Ureaplasma* (Figure 3.3).

The Tenericutes are predicted to have diverged approximately 600 million years ago from the low GC content Gram-positive bacteria (Firmicutes; Davis *et al.*, 2013). The *Anaeroplasma* then diverged from the other Tenericutes approximately 490 million years ago, around the same time that the Erysipelotrichales radiation emerged (Davis *et al.*, 2013). Since that time, the Tenericutes have been evolving rapidly and dynamically (Joblin and Naylor, 2002). Members of the “Acholeplasma” group have larger genome sizes, a greater range of metabolic capacities, and use the universal genetic code (Joblin and Naylor, 2002; Siewert *et al.*, 2014; Skennerton *et al.*, 2016). Many of the Acholeplasmatales are saprophytic or commensal with plant or animal hosts (Atobe *et al.*, 1983; Siewert *et al.*, 2014) while the Izimaplasmatales appear to be free-living without a eukaryotic host (Skennerton *et al.*, 2016). The *Anaeroplasma* are obligate anaerobes that often reside in the digestive systems or rumens of animals (Joblin and Naylor, 2002). The “Acholeplasma” group also contains *Candidatus Phytoplasma*, which are parasites and/or pathogens of plants, and have the characteristic reduced genomes and metabolic capacities of parasitic bacteria (Mollicutes, 1995; Siewert *et al.*, 2014). Thus, the “Acholeplasma” group of Tenericutes includes members with a diverse range of functions and lifestyles.

The second clade, the Mycoplasmatales, is made up of three smaller groups: the “Spiroplasma”, “Pneumoniae”, and “Hominis” groups, which are based on the genetics of the groups – genera names are often polyphyletic, particularly for the *Mycoplasma* (Gupta *et al.*, 2018) (Figures 3.1 and 3.3).

This clade contains host-associated organisms, and shares the reduced genomes and metabolic capacities seen in the *Phytoplasma* (Skenneron *et al.*, 2016). Mycoplasmatales members have complex nutritional needs, and frequently rely on their host for many amino acids, fatty acids, sterols, and vitamins for survival (Joblin and Naylor, 2002). The Mycoplasmatales includes commensal organisms, such as *Candidatus* Hepatoplasma crinochetorum in terrestrial isopods (Leclercq *et al.*, 2014) as well as primary and opportunistic parasites and pathogens, such as the bee parasite *Spiroplasma melliferum* (Lo *et al.*, 2013; Mollicutes, 1995). The *Mycoplasma* and *Ureaplasma* genera are of medical and agricultural importance, as many are human and animal pathogens (Ludwig *et al.*, 2010; Skenneron *et al.*, 2016). Members from the Mycoplasmatales use a variation of the genetic code that differs from other bacteria (Joblin and Naylor, 2002). The Mycoplasmatales use UGA, typically a stop codon, to code for tryptophan, and use this substitution at a frequency 10 times greater than the canonical tryptophan codon, UGG (Joblin and Naylor, 2002). This change in genetic code usage occurred after the separation of the Mycoplasmatales from the “Acholeplasma” group (Joblin and Naylor, 2002).

The vast majority of known Tenericutes have commensal or parasitic relationships with eukaryotic hosts, many of which are medically or agriculturally important. It is uncertain what lifestyles and roles Tenericutes may have in landfills. There is the strong potential for landfill-associated Tenericutes to be free-living which, before Skenneron *et al.* (2016) described the Izimaplasmatales, was unprecedented for this group. The objectives of the research presented in this chapter were to determine the phylogeny, metabolic capabilities, and possible lifestyles of the Tenericutes detected at the Southern Ontario landfill site. Members of both the Tenericutes and Erysipelotrichales were present on the 16 concatenated ribosomal protein and 16S rRNA gene trees (Figure 2.3), and high-quality bins were available for both groups, allowing for metabolic profiling. Tenericutes have previously been detected in non-host associated environments, including landfill leachate (Song *et al.*, 2015), sediments and groundwater from an aquifer (Anantharaman *et al.*, 2016), and the discovery of *Izimaplasma* in methane seeps by Skenneron *et al.* (2016). From the phylum level diversity analysis in Chapter 2, Tenericutes were present in each leachate sample a relative abundance of 8% in both the CLC_T2 and LW2 samples suggesting that this group is abundant within this landfill. As well, Tenericutes were detected at <10% and 30.6% abundance in landfills by Stamps *et al.* (2016) and Song *et al.* (2015), respectively, indicating Tenericutes may be frequently or stably associated with landfill environments. Determining where these landfill Tenericutes place phylogenetically within the Tenericutes groups and predicting their metabolic potential from metagenome-assembled genomes

(MAGs) will clarify possible lifestyles (parasitic, host-associated, or free-living), and expand our understanding of the phylogenetic and metabolic diversity within this curious clade.

3.2 Methods

3.2.1 Data collection

Phylogenetic and metabolic analyses on the Tenericutes were completed using the same metagenomic datasets as described for the diversity analyses presented in Chapter 2. Landfill sample collection, DNA extractions, metagenomic and 16S rRNA amplicon sequencing steps are detailed in Chapter 2 Section 2.2. Publicly available Tenericutes genomes were downloaded from the US Department of Energy's Joint Genome Institute's Integrated Microbial Genomes and Metagenomes (JGI-IMG) and NCBI Genome databases.

3.2.2 Metagenome assembled genomes

Metagenome assembled genomes (MAGs) affiliated with the Tenericutes and Erysipelotrichales were identified from the binned landfill and aquifer metagenomes. MAGs were reconstructed via binning using CONCOCT (Alneberg *et al.*, 2014) in Anvi'o (Eren *et al.*, 2015), as described in Chapter 2. Scaffolds that placed as either Tenericutes or Erysipelotrichales on the concatenated ribosomal protein tree (Figure 2.3) were connected to the bin database for the 2016 landfill samples. Scaffolds encoding the biomarker set of 16 syntenic ribosomal proteins that were assigned to a MAG with >80% completion and <5% redundancy were subset from the complete microbial community dataset and used to produce a phylogenetic tree focused on the Tenericutes and Erysipelotrichales. The associated high quality MAGs were also used for assessing landfill-associated Tenericutes metabolic potential.

3.2.3 Tenericute and Erysipelotrichales phylogenetic tree

To determine the phylogenetic placement of the landfill-associated Tenericutes MAGs within the Tenericutes group, reference sequences were downloaded for all genomes identified as Tenericutes, with the exception of the *Mycoplasma* genus, from the Joint Genome Institute IMG database. There are 254 genomes identified as belonging to the genus *Mycoplasma*, which was higher representation for a low-diversity group than was necessary for this phylogeny. A subset of 22 *Mycoplasma* spp. were included in the phylogeny, ensuring a minimum of four from the "Hominis" and "Pneumoniae" groups as classified by Gupta *et al.* (2018). Additional reference sequences were also downloaded

from the NCBI protein database, including sequences from the *Izimaplasma* genomes and the newly proposed groups within the Tenericutes orders in the GTDB taxonomy, as visualized in AnnoTree (“orders” RFN20, RF39, and ML615J-28) (Figure 3.2; Mendler *et al.*, 2019; Parks *et al.*, 2018). Sixteen Tenericutes MAGs downloaded from the JGI IMG as environmentally derived reference sequences were de-replicated using dRep (Olm *et al.*, 2017), reducing them to two unique genomes. Eight Erysipelotrichales were included to anchor the landfill Erysipelotrichales MAGs and eight Bacilli were included as an outgroup to the Tenericutes and Erysipelotrichales. The Erysipelotrichales and Bacilli reference sequences were downloaded from the JGI.

To construct the tree, a dataset of amino acid sequences for 16 syntenic, universally present, single copy ribosomal protein genes (RpL2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24 and RpS3, S8, S10, S17, S19) was developed for the landfill Tenericutes and Erysipelotrichales MAGs and the reference genomes from the JGI IMG database. The ribosomal protein genes for the references from *Izimaplasma*, RFN20, RF39, and ML615J-28 lineages were downloaded from the NCBI server using annotation keywords. Ribosomal protein datasets were screened for the A/E type, which were removed, as were short (<45 aa) sequences. Sequence sets for each protein were aligned using MUSCLE (Edgar, 2004) on Cipres (Miller *et al.*, 2010). Alignment columns containing $\geq 95\%$ gaps were removed using Geneious 11.0.5 (<https://www.geneious.com>). The sequence names were trimmed to 8 digits after the metagenome code (e.g. Ga0172377_100004578 → Ga0172377_10000457) to remove gene-specific identifiers and allow for concatenation by scaffold name. The protein gene alignments were concatenated in numeric order (L2 → L24, followed by S3 → S19). As not all scaffolds contained all of the 16 proteins in either partial or complete form, removed concatenated sequences that contained less than 50% of the total expected number of aligned amino acids were removed. The final alignment was 2,946 columns long and contained sequences from 16 MAGs and 181 references. A phylogenetic tree was inferred from the curated final alignment using RAxML-HPC Blackbox on Cipres using the following parameters: sequence type - protein; protein substitution matrix - LG; and estimate proportion of invariable sites (GTRGAMMA + I) - yes (Stamatakis, 2014).

3.2.4 Genome size vs. gene number

Some members of the Tenericutes have undergone extreme genome reductions as they evolved to be commensals and parasites, to the point that mycoplasmas have the smallest genomes of any self-replicating cell (Davis *et al.*, 2013; Joblin and Naylor, 2002). Certain members of the

Erysipelotrichales have also undergone genome reductions, but this is less widespread (Davis *et al.*, 2013). Small genomes and an associated reduced number of genes is considered indicative of a parasitic lifestyle (Davis *et al.*, 2013). In order to determine where the landfill *Tenericutes* and *Erysipelotrichia* MAGs fall within the gradient of free-living to parasitic for genome size vs gene number, the genome size was estimated for the MAGs from their completion estimate from Anvi'o and the gene number was summed from the identified genes for each MAG. The genome size and gene number were extracted for the *Tenericutes* and *Erysipelotrichales* references from the JGI. The *Tenericutes* references for this analysis are the same as were used for the phylogeny, whereas 34 additional *Erysipelotrichales* references were included to include the varying genome sizes within the *Erysipelotrichales* (Davis *et al.*, 2013). The genome size and gene number values were then plotted in R using ggplot2 (Wickham, 2011). Lifestyles were inferred for the reference genomes where possible, based on published literature or metadata associated with the genome sequences. A linear regression was performed in R to verify the relationship between genome size and gene number.

3.2.5 Metabolic analyses

Metabolic analyses were completed to determine whether the landfill *Tenericutes* exhibited characteristics of a free-living lifestyle, a host-associated lifestyle, or a parasitic lifestyle, like the *Izimaplasma*, *Acholeplasma*, and *Mycoplasma*, respectively. Landfill-associated *Tenericutes* and *Erysipelotrichales* MAGs as well as reference genomes were submitted to MAPLE-2.3.0 (Anwar *et al.*, 2017) to map genes to functional pathways and assign KO numbers. MAPLE-2.3.0 results were qualitatively analyzed for patterns of function presence/absence. A subset of the most complete references was selected to maintain representatives of each genera for analysis. Pathways not present in any of the MAGs or the reference subset were removed from consideration, as well as pathways that did not appear in the MAGs and were not complete in any of the references. The remaining pathways were visualized using heatmap3 (Zhao *et al.*, 2014) in R, showing their completeness from 0 to 100%.

3.2.6 Pangenome

The *Tenericutes* pangenome was analyzed using Anvi'o v. 5.5 for the 12 *Tenericutes* MAGs with greater than 80% completion and less than 5% redundancy (Delmont and Eren, 2018; Eren *et al.*, 2015). The five *Erysipelotrichales* MAGs meeting these criteria were also included for comparison. From the binning described in Chapter 2, the Anvi'o profile and contig databases were upgraded to

versions 31 and 12, respectively, to enable the pangenome analysis. The MAGs had originally been defined using the external gene calls provided by the JGI so the gene caller flag (--gene-caller JGI-pipeline) was used to maintain that change from the default Anvi'o pipeline parameters. For the pangenome analysis, the minbit flag was set to 0.3 instead of the default of 0.5, reducing the minimum bitscore required for a shared gene to be identified, and allowing for order-level comparisons between MAGs instead of the more common strain- or species-level comparisons.

3.2.7 Cultivation trials

3.2.7.1 Growth media

A total of four different growth media were assayed in an effort to cultivate landfill-associated Tenericutes (Table 3.1). All media used a baseline recipe of artificial leachate, developed by Co (2019) to mimic geochemical conditions of the Southern Ontario landfill (Appendix A Table A1). Tenericutes lack a cell wall and thus are unaffected by penicillin (Whitecomb *et al.*, 1995). This allowed us to add 5 UI/ml penicillin to each growth medium to act as a selective pressure against Gram-positive bacteria in the suspension (Polak-Vogelzang *et al.*, 1987). Thallium acetate was also added to Tenericutes growth medium #1 at a concentration of 0.05% w/v as per Polak-Vogelzang *et al.* (1987) to act as an additional biocide for non-Tenericute bacteria (Table 3.1). A yeast extract solution was made from 62.5 mg of yeast extract in 250 ml of water and was added to the growth media solution at 10% of the final solution to provide amino acids for the bacteria (Rose *et al.*, 1980). Sterols are required by members of the Mycoplasmatales for growth and, although not essential for growth of Acholeplasmatales, the availability of sterols in growth media allow Acholeplasmatales to produce their fried egg colony morphology (Edward, 1971; Martini *et al.*, 2014). Cholesterol was dissolved in ethanol and added to Tenericutes growth medium #1-3 to provide for the sterol requirement at a concentration of 5 µg/mL (Edward 1971) (Table 3.1). Cholesterol was not added to Tenericutes growth medium #4 to test for non-sterol requiring Tenericutes (Table 3.1). Fatty acids were added to Tenericutes growth media #1, #2, and #4 at a concentration of 10 µg/ml (Edward, 1971; Rose *et al.*, 1980) (Table 3.1).

Table 3.1: Growth media treatments for *Tenericutes* cultivation trials.

<i>Media</i>	<i>Penicillin</i>	<i>Thallium acetate</i>	<i>Yeast</i>		
			<i>extract</i>	<i>Cholesterol</i>	<i>Fatty acids</i>
#1	yes	no	yes	yes	yes
#2	yes	no	yes	yes	yes
#3	yes	no	yes	yes	no
#4	yes	yes	yes	no	yes

The synthetic leachate (Appendix A) with agar and yeast extract was autoclaved for sterilization. Hydrochloric acid was added to the synthetic leachate to adjust to a pH of ~7. Penicillin, thallium acetate, cholesterol, and fatty acid solutions were filter sterilized to prevent denaturing during autoclaving. Growth media solutions were mixed once the synthetic leachate agar had cooled to 55°C to prevent heat-denaturation of the penicillin, and then poured into petri plates.

3.2.7.2 Plating and growth of cultures

Biomass from the CLC_T2 sample was plated on four different growth medium compositions in an attempt to cultivate environmental *Tenericutes* from the landfill. One half of a 0.1 µm filter from the CLC_T2 sample after storage at -80°C was cut into small pieces and vortexed in 14 mL of synthetic leachate medium to resuspend the microorganisms. The suspension was plated on the cooled growth media plates at volumes of 100 µL and 1 mL as spread plates, with three replicates per growth medium. The plates were incubated at room temperature ($\cong 23^\circ\text{C}$) for two days until colonies formed. Two plates of each growth medium were then stored at 4°C while the third plate was incubated for an additional three days to allow for growth of any slower organisms prior to refrigeration. No growth was observed on *Tenericutes* growth medium #1 containing thallium acetate at this stage, so these plates were incubated at room temperature for an additional two weeks.

Colonies from growth media #2-4 were streaked on new plates to isolate colonies. Three colonies were chosen for each observed colony morphology on each kind of medium, which were picked and streaked onto new plates of the same growth medium. Streak plates were incubated at room temperature until isolated colonies formed on the plates, at which point they were stored at 4°C. Some of the colonies were streaked a second time to ensure only one colony morphology was present on the final streak plate.

3.2.8 Sample preparation for 16S rRNA gene sequencing

Colony PCR was performed on *Tenericutes* cultivation trial isolated microbes to amplify 16S rRNA genes for identification. The PCR reaction mixture contained PCR grade water (12 μ L), bovine serum albumin (10 mg/mL, 1.0 μ L), the forward primer 27F (10 μ M, 0.5 μ L; ACAGTTGATCMTGGCTCAG), and the reverse primer 1492R (10 μ M, 0.5 μ L; GGTTACCTTGTTACGACTT) (Sigma). 14 μ L of the prepared PCR mixture and 10 μ L of 5Prime HotMasterMix (Quantabio) was added to each reaction. DNA template was added to the mixture by touching a sterile pipette tip to a colony and swirling it in the PCR mixture. Samples were amplified in a C1000 Touch thermal cycler (Bio-Rad). An initial denaturing step occurred at 94°C for 5 minutes. A total of 30 cycles of denaturation at 94°C for 45 seconds, annealing at 50°C for 60 seconds, and extension at 72°C for 90 seconds followed. A final extension at 72 °C for 10 minutes ensures an A-overhang was added to the amplified products. A quality control gel was run to confirm amplification using a 1% agarose and TAE gel with SYBER Safe dye (Invitrogen). PCR products were purified using the ThermoScientific GeneJET PCR Purification Kit following the kit instructions. Purified PCR products' DNA concentrations were assayed using a NanoDrop spectrophotometer (Thermo Scientific). The *Taq* polymerase was added to 5 μ L of the purified PCR product and incubated for 8 minutes in the C1000 Touch thermal cycler (Bio-Rad) at 72°C. This additional step ensured the 3' A-overhangs were present post-PCR product clean-up, following troubleshooting suggestions in the manual for the TOPO TA Cloning Kit (Invitrogen) for Sequencing. Cloning kit protocols were followed for transformation of the 16S rRNA gene with one deviation from the instructions – the ligation reaction was incubated for 10 minutes to increase the likelihood of transformation with the PCR product. 2 μ L of the ligated plasmid was added to the Transform One Shot competent cells (Invitrogen) and incubated on ice for 20 minutes. Cells were then heat shocked for 30 seconds at 42°C. A total of 250 mL of room temperature S.O.C. medium was added to the competent cells. Tubes of cells were shaken horizontally for 1 hour at 37°C. Two volumes of cell suspension, 10 μ L and 50 μ L, were added to prepared, prewarmed LB agar plates with an ampicillin concentration of 50 μ g/mL. The LB plates were incubated at 37°C overnight to allow for growth of clone colonies. Six colonies from each reaction were chosen for colony PCR amplification to check for the 16S rRNA gene insert. The same PCR amplification protocol was used with substitution of the plasmid-binding M13 forward (GTAAAACGACGGCCAG) and reverse (CAGGAAACAGCTATGAC) primers. A quality control gel was run (as above) to check for amplification of the plasmid inserts. Colonies positive for an insert of the correct size were inoculated

into 5 mL of LB broth with an ampicillin concentration of 50 mg/mL and incubated overnight. Plasmids were extracted from 3 mL of the overnight cultures using the Thermo Scientific GeneJET Plasmid Miniprep kit. Purified plasmid products were quality and quantity checked using a NanoDrop spectrophotometer (Thermo Scientific). Twenty plasmid samples, two for each of ten isolates, were prepared for sequencing as per the The Center for Applied Genomics (TCAG; Toronto, ON) guidelines. All selected 16S rRNA gene clones were sequenced using the M13 forward and reverse primers by TCAG.

3.2.9 Identification of isolates from 16S rRNA gene sequences

The 16S rRNA gene sequences for the isolates were downloaded from the TCAG platform and trimmed using their quality values. Forward and reverse reads were *de novo* assembled in Geneious. Any areas of poor quality at the ends of the sequences were removed and a consensus sequence created in Geneious. The consensus sequences were submitted to the SILVA SINA algorithm (Pruesse *et al.*, 2012) for alignment and classification. Unaligned bases at the ends of the genes were removed and sequences below 70% identity to a reference sequence were automatically removed. The isolate 16S rRNA gene sequences were searched with BLAST against the six landfill metagenomes via the JGI IMG platform. BLAST results with 99% identity over 90% of the metagenome 16S rRNA gene sequence length were considered matches. Metagenome sequences were required to have a minimum length of 600 bp. The isolate 16S rRNA gene sequences were also searched with BLAST against the 16S rRNA gene amplicon exact sequence variant (ESV) representative sequences from QIIME2 (Bolyen *et al.*, 2019; see Chapter 2 for full ESV descriptions) using Geneious 11.0.5. The BLAST results with 99% identity across the length of an ESV were considered matches.

3.3 Results and discussion

3.3.1 Metagenome assembled genomes

There were 12 Tenericutes MAGs and 5 Erysipelotrichales MAGs that met the greater than 80% completion and less than 5% redundancy requirements (Table 3.2). The MAGs Bin_16_2, an Achleplasmatales from LW2, and Bin_72_1, a RFN20 from LW2, had the highest mean coverage values of 132.7 and 160.0, respectively (Table 3.2). The LW2 sample had the highest abundance of MAGs with a total mean coverage of 335.1, followed by CLC_T2 with 147.2, CLC_T1 with 88.1, and LW1 with 22.5. The CLC_T2 sample had the highest number of MAGs with 8 with at least one from each order (Table 3.2). No high quality Tenericutes or Erysipelotrichia MAGs were identified

from LW3 or GW1. The Tenericutes MAGs range in estimated size from 0.99 Mb for Bin_16_5 to 1.59 Mb for Bin_35_5. The Erysipelotrichales MAGs range in estimated size from 0.74 Mb for Bin_58_4 to 1.62 Mb for Bin_30_6. GC content for the MAGs ranges from 26.1% for Tenericutes Bin_22_4 to 47.6% for Tenericutes Bin_62_2. The Tenericutes and Erysipelotrichales are hypothesized to have a common ancestor with the low GC content Gram-positive bacteria (Davis *et al.*, 2013). As a result, Tenericutes are expected to have low GC content in the 20-30% range, with a few species reaching up to 40% (Joblin and Naylor, 2002; Mollicutes, 1995). The lower GC contents of the landfill-associated Tenericute and RFN20 MAGs are expected based on this previous work; however, the higher GC contents (>40%), seen in the two *Izimaplasmatales* bins and two of the ML615J-28 bins (Table 3.2) are higher than is expected for Tenericutes. For comparison, the *Izimaplasma* genomes from Skennerton *et al.* (2016) have GC contents of 31.3% (HR1) and 29.2% (HR2). With expansion of these groups from environmental sequencing, we observe a wider range of GC contents for both *Izimaplasma* genomes and ML615J-28 genomes from the GTDB/Annotree database, with GC contents reaching into the 50% range for some genomes (Mendler *et al.*, 2019). Lower GC contents are typical for parasitic and endosymbiotic organisms with small genomes, while free-living microbes with larger genomes are more GC rich (Dutta and Paul, 2012). As most of the older work on the Tenericutes was concentrated on parasitic and host-associated species, it appears that the trend for lower GC content is more consistent for organism with those lifestyles, whereas free-living Tenericutes have a greater range of GC content. Merhej *et al.* (2009), through their comparative analysis of 317 bacterial genomes, found that lower GC content correlates strongly with genome size reduction in obligate intracellular bacteria, including parasites and endosymbionts such as *Ca. Zinderia insecticola* and *Ca. Carsonella ruddii* (Proteobacteria) with GC contents of 13.5% and 16.6%, respectively (McCutcheon and Moran, 2012), whereas free-living bacteria have higher GC contents and larger genome sizes. Thus, variation in GC content for the Tenericute and RFN20 MAGs may provide an indication of their potential lifestyles.

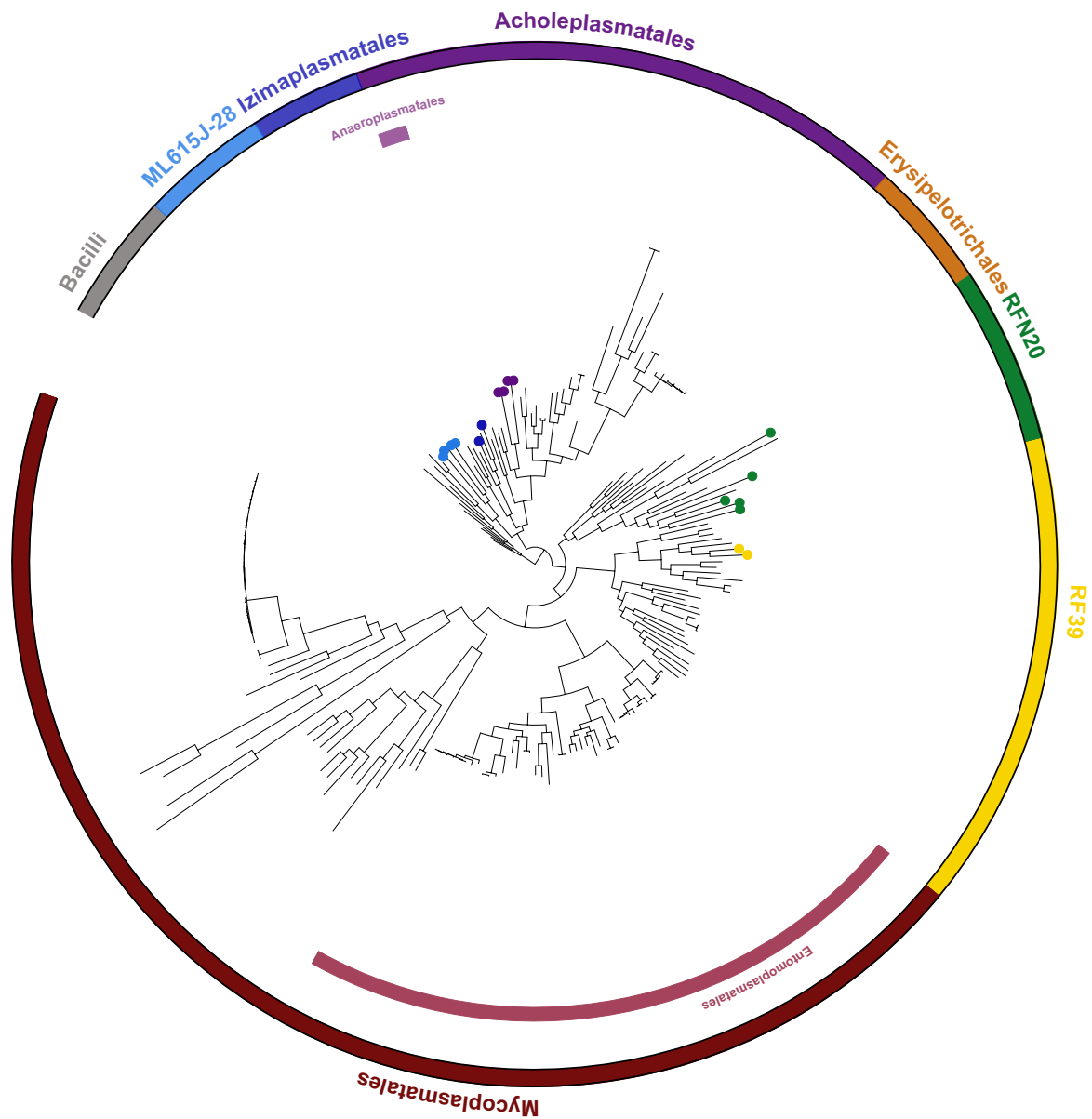
Table 3.2: Genome statistics and assigned taxonomy for *Tenericutes* and *Erysipelotrichales* MAGs from the landfill metagenomes.

<i>Bin number</i>	<i>Site</i>	<i>Mean coverage</i>	<i>Taxonomy</i>	<i>Total size (mb)</i>	<i>Number of contigs</i>	<i>GC content (%)</i>	<i>Completion (%)</i>	<i>Redundancy (%)</i>	<i>Estimated size (mb)</i>
<i>Bin_62_2</i>	LW2	10.0	ML615J-28	1.49	188	47.9	95.7	1.4	1.55
<i>Bin_31_2</i>	CLC_T2	7.9	ML615J-28	1.27	175	47.7	89.9	1.4	1.41
<i>Bin_22_2_</i> <i>LW</i>	LW1	22.5	ML615J-28	1.10	110	27.1	95	1.4	1.16
<i>Bin_17_2</i>	CLC_T1	20.2	ML615J-28	1.19	125	27.9	95	2.2	1.25
<i>Bin_35_5</i>	CLC_T2	12.1	Izimaplasmatales	1.52	143	44	95.7	1.4	1.59
<i>Bin_25_6</i>	CLC_T2	10.3	Izimaplasmatales	1.25	157	38.4	86.3	1.4	1.45
<i>Bin_16_2</i>	LW2	132.7	Acholeplasmatales	0.96	120	31.2	94.2	0.7	1.02
<i>Bin_42_3</i>	CLC_T2	65.3	Acholeplasmatales	1.18	108	31.2	96.4	0	1.22
<i>Bin_42_4</i>	CLC_T2	10.9	Acholeplasmatales	1.39	146	31.3	97.1	2.2	1.43
<i>Bin_16_5</i>	LW2	15.6	Acholeplasmatales	0.94	35	31.2	95.7	0	0.99
<i>Bin_33_2</i>	CLC_T2	8.6	RFN20	1.14	141	49.7	84.2	0.7	1.35
<i>Bin_30_6</i>	CLC_T1	21.6	RFN20	1.33	35	28.7	82	3.6	1.62
<i>Bin_58_4</i>	LW2	16.8	RFN20	0.60	35	36.3	82	0.7	0.74
<i>Bin_72_1</i>	LW2	160.0	RFN20	1.59	48	33.8	95.7	0.7	1.59
<i>Bin_44_6</i>	CLC_T1	46.3	RFN20	0.91	100	38	91.4	2.9	1.00
<i>Bin_22_4</i>	CLC_T2	19.8	RF39	0.87	104	26.1	82	2.9	1.06
<i>Bin_22_2_</i> <i>CLC_T2</i>	CLC_T2	12.3	RF39	1.28	117	26.6	85.6	1.4	1.50

3.3.2 Phylogeny

The *Tenericutes* MAGs are associated with four of the GTDB-defined orders: 3 in ML615J-28, 2 in Izimaplasmatales, 4 in Acholeplasmatales, and 2 in RF39 (Figure 3.4). None of the landfill *Tenericutes* were phylogenetically placed within the Mycoplasmatales – the closest were the RF39 MAGs, the sister group to the Mycoplasmatales. The five “*Erysipelotrichales*” bins were originally identified as *Erysipelotrichales* based on annotations to reference genomes, but placed phylogenetically within the RFN20 order. The RFN20 order is sister to the *Erysipelotrichales*, but currently considered a separate order within the GTDB taxonomy structure. The phylogenetic placement of the landfill-associated *Tenericutes* MAGs is consistent with our expectation that they

would not be Mycoplasmatales, given that it would be difficult for animal and plant parasites to persist in the landfill leachate and aquifer environments. The overall phylogenetic placements of the Tenericutes and the Erysipelotrichales in our concatenated ribosomal protein tree are consistent with the Davis *et al.* (2013) and Parks *et al.* (2018) findings that the Tenericutes are polyphyletic, as their most recent common ancestral node also includes the Erysipelotrichales. The inclusion of the Tenericutes orders within class Bacilli, as suggested by Parks *et al.* (2018), is consistent with our molecular evidence. As such the name Tenericutes no longer applies to a specific taxonomic level. Both Tenericutes and Mollicutes have specific definitions and classifications within the context of this group, and neither are inclusive of all the currently known orders. More research needs to be done to better understand the environmental representatives of the group, in particular the currently undescribed orders, and it would be beneficial to establish a name for this clade to refer to the collective.



Tree scale: 1

Figure 3.4: Maximum likelihood phylogenetic tree inferred from a concatenated alignment of 16 syntenic ribosomal proteins for the landfill MAGs and references within the Tenericutes and Erysipelotrichales. The final alignment contained 2,946 positions with sequences from 17 MAGs and 181 reference genomes. The tree is rooted with a Bacilli outgroup in grey. The current orders as per the GTDB taxonomy are included in the outer ring, coloured by order. Two orders that were previously included in the Tenericutes but are not considered order-level lineages per the GTDB

taxonomy are also noted inside the outer ring: Anaeroplasmatales in light purple, which are now included in Acholeplasmatales; and Entomoplasmatales in pink, which are now included in the Mycoplasmatales. The MAG sequences are indicated by coloured circles, matching the order colour, at the tips of the trees. The hash marks on one of the Mycoplasmatales branches indicates that this branch was shortened for the tree representation. The phylogenetic tree was visualized using iTOL v.4 (Letunic and Bork, 2019).

3.3.3 Genome size vs gene number

There is a strong correlation between genome size and gene number for Tenericute and Erysipelotrichales genomes, both for all genomes ($R^2 = 0.97$) and for genomes smaller than 3 Mb ($R^2 = 0.94$) (Table 3.3). The known parasitic organisms, particularly those belonging to *Phytoplasma* of the Acholeplasmatales and *Ureaplasma* of the Mycoplasmatales, have the smallest genome sizes and gene numbers, with a larger range in sizes observed for the host-associated members (Figure 3.5B), and particularly for the Erysipelotrichales (orange, Figure 3.6A). The environmental Tenericutes (purple triangles, Figure 3.6), which were collected from a geyser, also exhibit a larger range in genome size to gene number, including some with genome sizes and gene complements near those of the Izimaplasmatales (dark blue, Figure 3.6B), suggesting that these populations may be free-living. The *Anaeroplasma* reference genomes have larger genome sizes and place with the Izimaplasmatales and environmental Tenericutes (Figures 3.5B and 3.6B), aligning it with known and potentially free-living organisms, although this genus is considered to be host-associated, as members have been isolated from animal rumens (Joblin and Naylor, 2002). The landfill Tenericutes MAGs genome sizes and gene complements fall in the same range as the Entomoplasmatales and the Acholeplasmatales (not including the *Phytoplasma*), suggesting that these organisms may be host-associated or commensal with other microorganisms. Historically, Tenericutes have been described as having genome sizes between 600 and 2,200 kbp (ICSB Subcommittee on the Taxonomy of Mollicutes, 1995), which is consistent with the estimated genome sizes for the landfill-associated MAGs. The RF39 reference genomes are larger than the majority of the Entomoplasmatales and the Mycoplasmatales. The landfill-associated RFN20 MAGs are in line with the reference RFN20x genomes. One, Bin_58_4, has an estimated genome size of 0.74 Mb, which places it with the *Mycoplasma* (Figure 3.6B), suggesting that this organism may be parasitic or similarly heavily reliant on other organisms for its metabolism. Notably, the landfill-associated MAGs fit closely to the regression line, suggesting that they are not currently undergoing a genome reduction.

Table 3.3: Linear regression statistics for genome size vs. gene number as depicted in Figures 3.5 and 3.6.

<i>Statistic</i>	<i>All genomes</i>	<i>Genomes < 3Mb</i>
<i>Intercept</i>	-60.11	-94.68
<i>Intercept Std. Error</i>	19.05	24.50
<i>Intercept t value</i>	-3.16	-3.86
<i>Intercept p value</i>	0	0
<i>Slope</i>	1013.79	1043.06
<i>Slope Std. Error</i>	11.17	17.39
<i>Slope t value</i>	90.75	59.99
<i>Slope p value</i>	<2e-16	<2e-16
<i>Residual Std. Error</i>	146.2 (236 df)	138.8 (220 df)
<i>Multiple R-squared</i>	0.97	0.94
<i>Adjusted R-squared</i>	0.97	0.94
<i>F-statistic</i>	8235 (1 and 236 df)	3598 (1 and 220 df)
<i>F-statistic p value</i>	<2.2e-16	<2.2e-16

Degrees of freedom are denoted as df.

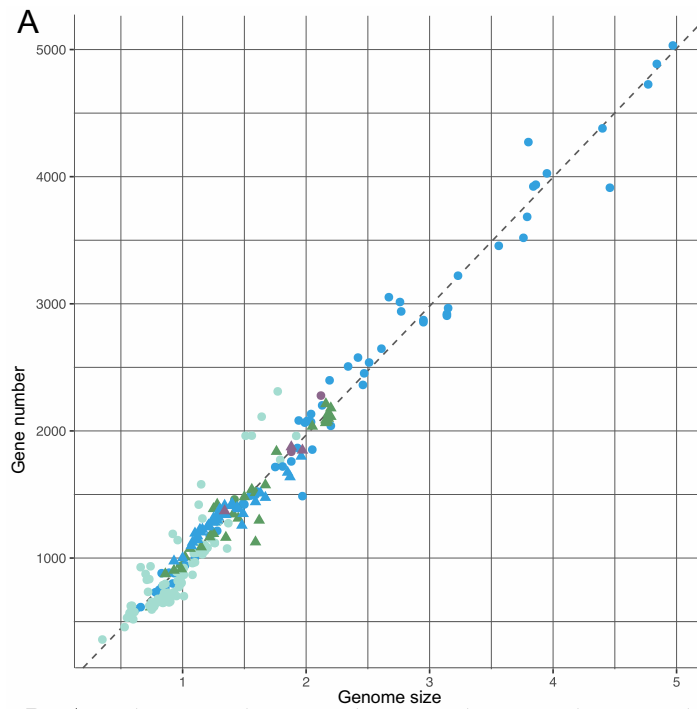
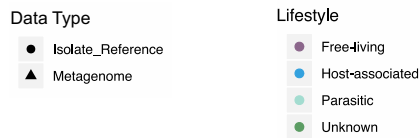
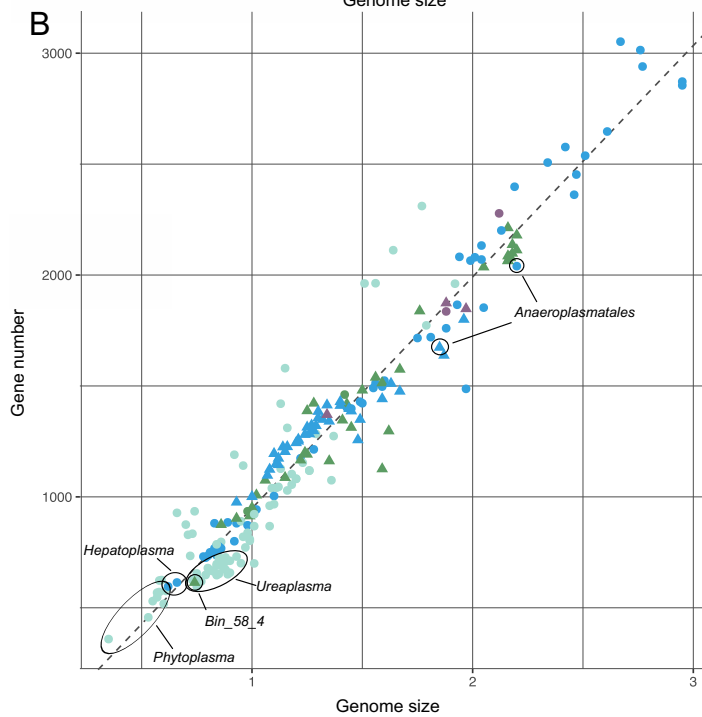
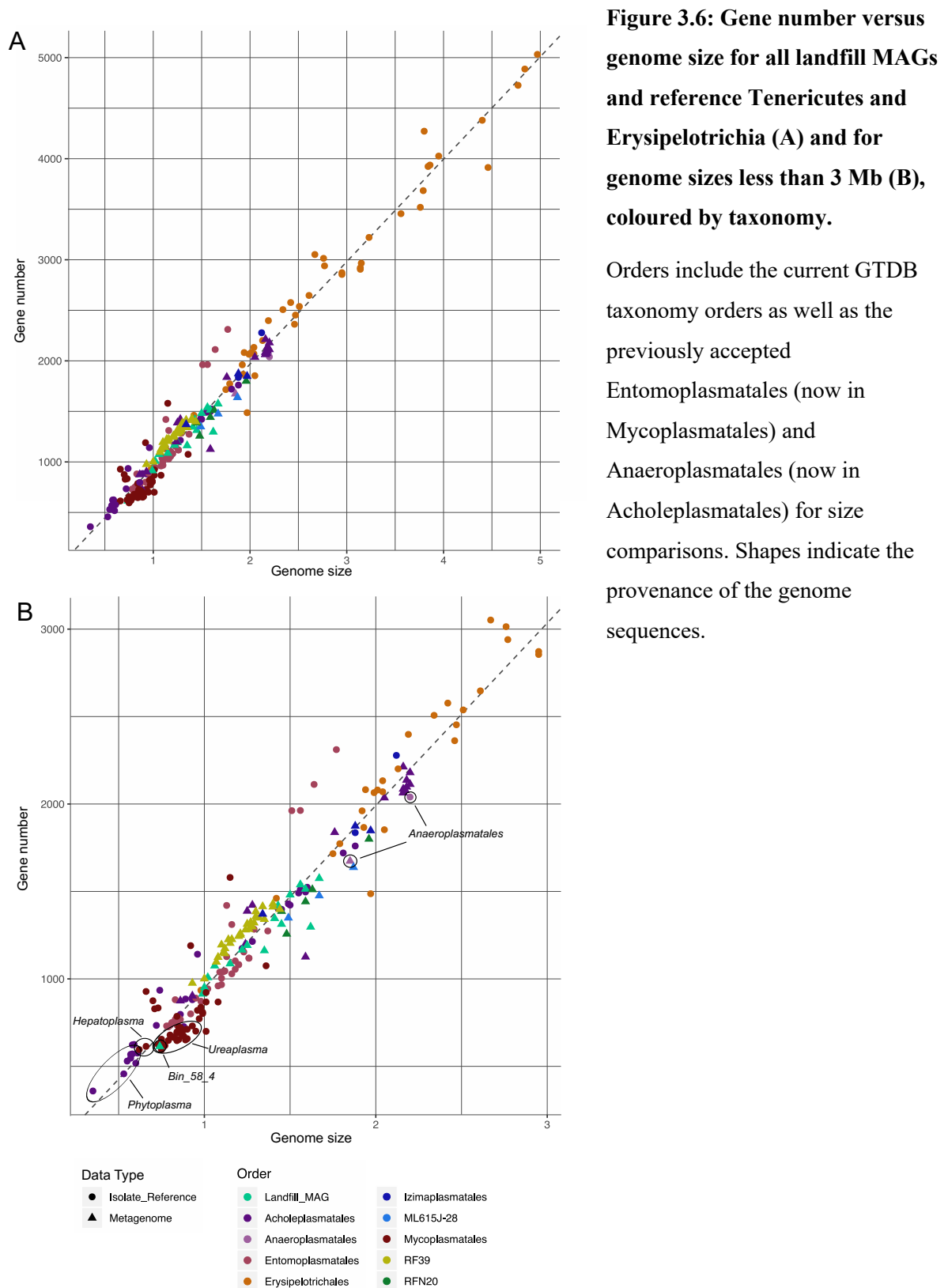


Figure 3.5: Gene number versus genome size for all landfill MAGs and reference Tenericutes and Erysipelotrichia (A) and for genome sizes less than 3 Mb (B), coloured by lifestyle.

Shapes indicate the provenance of the genome sequences.





3.3.4 Metabolic analyses

3.3.4.1 Cell wall

One of the phenotypic characteristics separating the Tenericute orders from the Firmicutes was their lack of cell wall (Freundt, 1975). The landfill-associated and reference Tenericutes genomes do not encode the components of the pathways for peptidoglycan and lipopolysaccharide production, which is consistent with the lack of a cell wall. One MAG, Bin_36_1, which did not meet quality thresholds for inclusion in the in-depth phylogenetic and metabolic analyses, encodes 50% of the peptidoglycan pathway, as does the Erysipelotrichia reference genome for *Coprobacillus* sp. 29_1. This is only weak evidence for a cell wall within the orders of interest, and the ability to produce a cell wall does appear to be universally absent throughout these groups.

3.3.4.2 Sterols

The requirement for sterols was used to separate early Tenericutes into the *Mycoplasma* and *Acholeplasma* genera (Freundt, 1975). The C5 isoprenoid mevalonate and non-mevalonate pathways are used to produce sterol precursors in bacteria (Kuzuyama and Seto, 2012). Complete or partial C5 isoprenoid biosynthesis pathways are present in the genomes for the Izimaplasmatales, ML615J-28, and Acholeplasmatales, except for the *Phytoplasma*, as well as in the Erysipelotrichia and RFN20 genomes (Figure 3.7C). Interestingly, *Spiroplasma taiwanense* possesses a partial non-mevalonate pathway and *Spiroplasma mirum* has a complete non-mevalonate pathway, suggesting that these two species may be capable of producing sterols (Figure 3.7C). With those two exceptions, members of the Mycoplasmatales and RF39 orders do not possess any components for the C5 isoprenoid mevalonate and non-mevalonate biosynthesis pathways.

3.3.4.3 Motility

None of the landfill MAGs or the references possessed pathway components for flagellar synthesis. This suggests that these organisms do not move by flagellar locomotion and thus may depend on twitch motility or external forces for locomotion.

3.3.4.4 Sensory

None of the landfill MAGs or the reference possessed pathway components for phototaxis. Whereas, the *Izimaplasma* HR1 reference genome, *Acholeplasma equifetale*, Bin_22_2_CLC_T2, and one lower quality MAG, Bin_18_5, encode 20% of the pathway for chemotaxis. The Erysipelotrichia

reference genomes from Erysipelotrichaceae bacterium 3_1_53 and Erysipelotrichaceae bacterium 5_2_54 encode 60% and 40% of the chemotaxis pathway, respectively, with nine other Erysipelotrichia reference genomes encoding 20% of the pathway. The partial chemotaxis pathways may allow for some sensory capabilities depending on which parts of the pathways are present but full chemotaxis is not possible for these organisms.

3.3.4.5 Energy metabolism

In general, the reduced genomes of host-associated and parasitic Tenericutes maintain metabolic pathways involved in energy acquisition (Lazarev *et al.*, 2011). None of the Tenericute genomes examined encode an electron transport chain, and so these organisms are hypothesized to be facultatively anaerobic, like most Tenericutes (Joblin and Naylor, 2002). The majority of the Tenericutes genomes, both MAGs and references, encode a complete phosphate acetyltransferase-acetate kinase pathway to convert acetyl-CoA to acetate (Figure 3.7A). The pyruvate pathway to produce acetyl-CoA is complete in the majority of the Tenericutes and Erysipelotrichales MAGs and reference genomes examined, aside from the *Ureaplasma* and *Hepatoplasma* references. Components for the citrate cycle (TCA cycle) and second carbon oxidation are absent in the Mycoplasmatales reference genomes and in the RFN20 and RF39 MAGs. Less than 40% of the TCA cycle and second carbon oxidation components are present in the other profiled reference genomes and MAGs (Figure 3.7A). The first carbon oxidation pathway for the TCA cycle is complete in the reference genomes *Anaeroplasma*, *Dielma*, and *Kandleria*. Glycolysis pathways are complete or partial in all of the MAGs and reference genomes, with the lowest completion values in MAGs Bin_58_4, Bin_33_2, Bin_22_2_CLC_T2, and Bin_22_4, and in the reference genomes for *Mycoplasma pneumoniae* and the two *Ureaplasma* (Figure 3.7A). Other sugar metabolisms, including nucleotide sugars, trehalose, galactose, and D-galactonate have partial or complete pathways in the majority of the Erysipelotrichia, RFN20, and “Acholeplasma” group genomes, except for the *Phytoplasma* references and MAGs Bin_22_2_LW and Bin_17_2. These alternate sugar metabolisms are lacking in the Mycoplasmatales reference genomes and the RF39 MAGs.



^ Figure 3.7: Heatmap of metabolism pathways for the landfill MAGs and representative reference genomes for the orders of interest. Metabolic potential was determined using MAPLE version 2.3.0 and the bi-directional best hit method (Arai et al., 2018). A) Carbon and carbohydrate pathways. B) Amino acid biosynthesis and degradation pathways. C) Cofactor, fatty acid, purine, pyrimidine, and terpenoid biosynthesis pathways. Taxa are coloured by order, with MAGs in bold: ML615J-28 in light blue, Izimaplasmatales in dark blue, Acholeplasmatales in purple, Erysipelotrichales in orange, RFN20 in yellow, RF39 in green, and Mycoplasmatales in red. Heatmap colour scheme is for pathway completion from 0% to 100%. Many pathways involving carbon and carbohydrate metabolism are consistent across groups. The division between metabolic capacities of free-living and commensal organisms compared to parasitic organisms is more evident in the amino acid biosynthesis pathways. MAGs Bin_62_2 and Bin_30_6, along with environmental reference *Tenericutes bacterium* GWC2_34_14, did not return results from MAPLE-2.3.0 and are not included in the heatmap. *Tenericutes bacterium* GWF1_35_14 was substituted for *Tenericutes bacterium* GWC2_34_14 as a reference. Pathway short forms: RRPC= reductive pentose phosphate cycle; FA=formaldehyde assimilation; PRPP= 5-phospho- α -D-ribose 1-pyrophosphate; NSB=nucleotide sugar biosynthesis; FAB= fatty acid biosynthesis; PE= phosphatidylethanolamine; and IB=isoprenoid biosynthesis.

3.3.4.6 Fatty acid and lipid metabolism

The *Izimaplasma* and *Acholeplasma* reference genomes, and *Tenericutes* Bin_31_2 and Bin_35_5 have complete pathways for fatty acid synthesis initiation and elongation, notably different from the 0% or 25% for fatty acid initiation and 0% fatty acid elongation in the other reference genomes and MAGs (exceptions were 33.3% in RFN20 Bin_72_1 and 66.7% in the *Anaeroplasma* reference genome). The *Anaeroplasma*, four of the *Tenericutes* MAGs, and two of the RFN20 bins possess the complete pathway for acetyl-CoA synthesis through beta-oxidation (KEGG pathway M00086), but none of the genomes possess components of the KEGG beta-oxidation pathway (M00087) (Figure 3.7C). Phosphatidylethanolamine biosynthesis and triacylglycerol biosynthesis pathways range from 20-60% completion for all genomes.

3.3.4.7 Amino acid metabolism

Tenericutes typically have reduced biosynthetic capacities as a result of genome reduction (Joblin and Naylor, 2002; Lazarev et al., 2011). This reduced capacity is evident in our results for environmental *Tenericutes* genomes as well, as many amino acid biosynthesis pathways are absent or partial, such as

those for lysine, cysteine, tyrosine, and phenylalanine (Figure 3.7B). *Anaeroplasma bactoclasticum*, *Dielma fastidiosa*, and *Kandleria vitulina* reference genomes, which have larger sizes, possess a number of complete amino acid synthesis pathways that are missing from the other reference genomes and the landfill MAGs, including pathways for proline, tryptophan, and isoleucine synthesis (Figure 3.7B). The *Phytoplasma* and Mycoplasmatales reference genomes, from organisms with parasitic lifestyles, are missing more of the amino acid synthesis pathways than the landfill, environmental, and host-associated genomes. In general, the majority of the landfill-associated MAGs show a larger metabolic potential for biosynthesis of amino acids than the strictly parasitic Tenericutes, but do not have the same capacities as organisms with larger genomes.

3.3.4.8 Purine and pyrimidine synthesis

The majority of the Tenericutes genomes possess complete pathways for purine synthesis but not for pyrimidine synthesis, which is consistent with current knowledge of Tenericute biosynthetic abilities (Lazarev *et al.*, 2011). The *Izimaplasma*, environmental Tenericutes references, *Anaeroplasma*, *Dielma*, *Kandleria*, *Spiroplasma*, *Entomoplasma*, and *Mesoplasma* reference genomes possess 75-100% completeness for adenine and guanine synthesis (Figure 3.7 C). Purine synthesis is at 50% completion in the MAGs and other references, except Bin_58_4, Bin_72_1, Bin_22_2_CLC_T2, CLC_22_4, and the *Phytoplasma* references (Figure 3.7 C). Pyrimidine deoxyribonucleotide biosynthesis and ribonucleotide synthesis pathway completeness range from 40-60% in the majority of examined genomes. Uridine monophosphate biosynthesis is largely absent from the genomes, except for *Izimaplasma*, *Anaeroplasma*, *Dielma*, *Kandleria*, *Spiroplasma*, *Entomoplasma*, Bin_25_6, and Bin_35_5 having >80% completion. Thus, purine metabolism potential is possible in the majority of genomes, but pyrimidine metabolism appears limited to organisms with larger genome sizes.

3.3.5 Pangenome

The pangenome analysis included the twelve Tenericutes MAGs and five RFN20 MAGs with quality scores above 80% completion and below 5% redundancy (Figure 3.8). There is a core genome of 74 single copy genes that are conserved across all MAGs, which anchors a core pangenome of 171 gene clusters (black, Figure 3.8). There are also two sets of semi-conserved genes (Semi-Core Regions (SCR); grey, Figure 3.8). Given that these MAGs belong to five different taxonomic orders, it was expected that there would be low numbers of conserved gene clusters between all of the MAGs. Semi-Core Region 1 (SCR 1) contains 161 gene clusters, with more shared gene clusters between the

RFN20, Acholeplasmatales, Izimaplasmatales, and ML615J-28 MAGs with some shared gene clusters including the RF39 MAGs. Semi-Core Region 2 (SCR 2) contains 311 gene clusters largely found in the Izimaplasma and ML615J-28 MAGs, and showing lower representation of MAGs from the “Acholeplasma” group orders and the RFN20 and RF39 MAGs. The RFN20 and RF39 orders share a more recent common ancestor with each other than they do with the “Acholeplasma” group orders, which is reflected in the gene cluster differences between these clades in the SCR regions. There are several sets of gene clusters shared between pairs of MAGs, which reflect the phylogenetic relationship between the organisms (Figure 3.8). This includes the ML615J-28 (108 gene clusters), Acholeplasmatales (109 gene clusters), and RFN20 (14 gene clusters) cores. Analysis of the specific functions encoded within these core, semi-core, and order-conserved regions would be valuable future work to identify key similarities and differences between these organisms.

Order specific patterns for the MAGs are preliminary, given that there are only two MAGs for both the Izimaplasmatales and the RF39, and only four MAGs for each of the other orders. The number of gene clusters tracks with the total length of the MAGs. The number of singleton gene clusters is lower in the “Acholeplasma” group MAGs, as there are more gene clusters shared between MAGs within the included orders. An exception to this is Acholeplasmatales Bin_42_4 which has a gene cluster region not shared with the other Acholeplasmatales (purple, 7 o'clock on Figure 3.8). The RFN20 and RF39 MAGs each have singleton gene cluster regions. All of the MAGs have GC contents below 50%, shown in the GC content bar chart which ranges from 0 to 49% (green summary, Figure 3.8). All of the MAGs have relatively high numbers of genes per kb, in keeping with the reduced genome sizes of the landfill MAGs (Figure 3.6).

The geometric homogeneity index is higher than the functional homogeneity index for the gene clusters. The geometric homogeneity is a metric to measure gap/residue patterns such that high geometric homogeneity index values indicate mostly uniform gap/residue distribution within the gene cluster (Delmont and Eren, 2018). The high geometric homogeneity seen here indicates large insertions/deletions are relatively rare between these groups. In comparison, the functional homogeneity index considers aligned residues and their biochemical properties (Delmont and Eren, 2018). Given that our MAG gene clusters have lower functional homogeneity, we can infer that there are differences in amino acid composition between the MAGs with biochemical differences in polarity, charge, and whether they are acidic or basic (Delmont and Eren, 2018). These differences in

amino acid residues are likely driven by evolutionary changes over the time of divergence of the five orders.

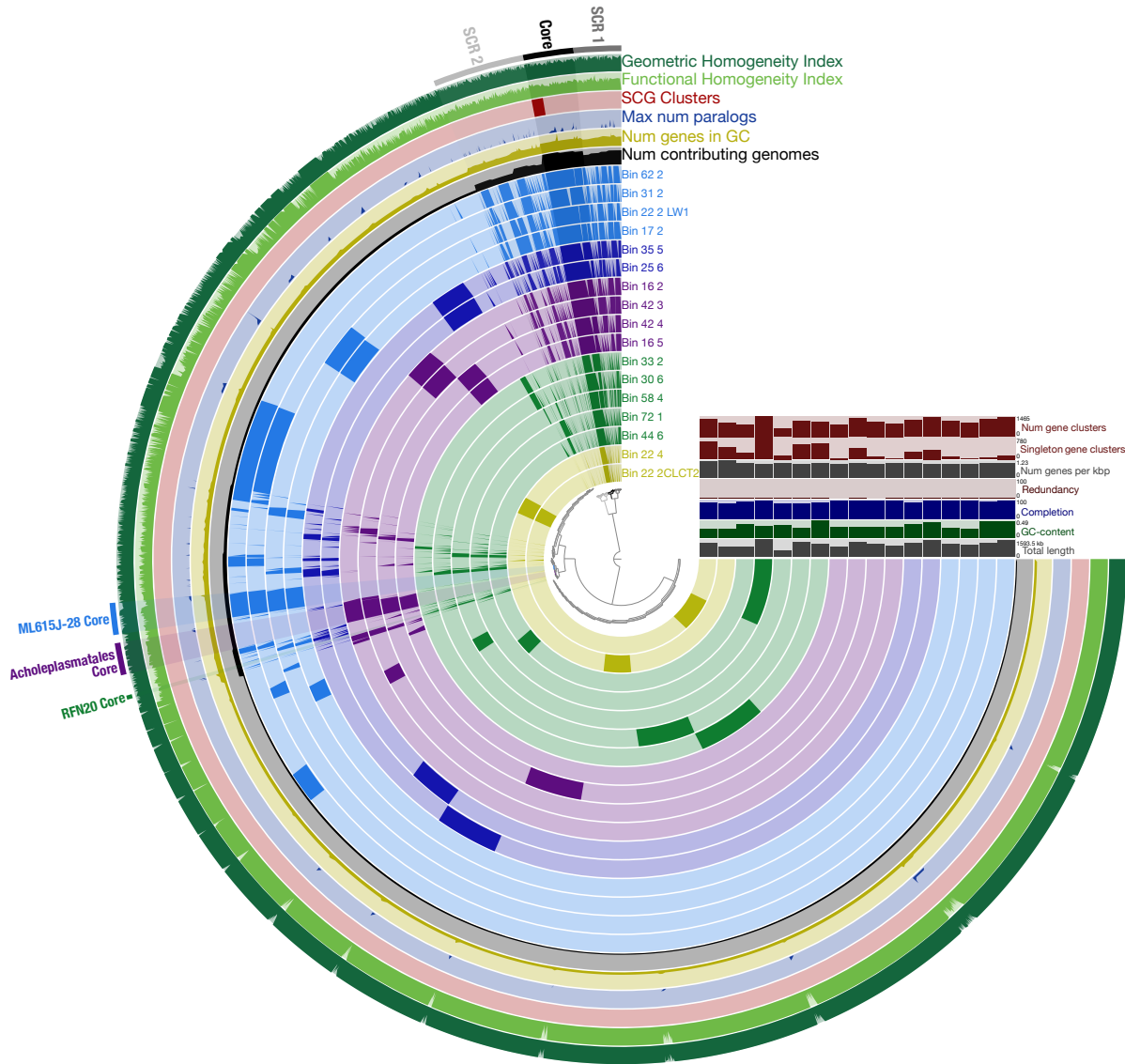


Figure 3.8: Pangenome analysis of the 17 landfill metagenome assembled genomes showing gene cluster presence/absence and shared gene cluster profiles. The 17 inner rings represent the 17 MAGs, from inside to outside: two RF39 MAGs in yellow, five RFN20 MAGs in green, four Achleplasmatales MAGs in purple, two Izimaplasmatales MAGs in dark blue, and four ML615J-28 MAGs in light blue. The outer six rings are bar charts for the following metrics for the gene clusters,

outside to inside: geometric homogeneity index (dark green), functional homogeneity index (light green), single copy gene (SCG) clusters (red), maximum number of paralogs (blue), number of genes in gene cluster (yellow), and number of contributing genomes (black). The bar charts at 3 o'clock on the circle have metrics for the MAGs, top to bottom: number of gene clusters (red), singleton gene clusters (red), number of genes per kpb (grey), redundancy (red), completion (blue), GC content (green), and total MAG length (grey). The core genome (Core), including the SCG cluster, and the two semi-conserved regions (SCR 1 and SCR2) are noted in black, dark grey, and light grey, respectively. Gene clusters shared within ML615J-28, Acholeplasmatales, and RFN20 are noted in blue, purple, and green, respectively. Pangenome analysis was completed and visualized using Anvi'o 5 Margaret (Delmont and Eren, 2018; Eren et al., 2015). MAG amino acid sequences were aligned using MUSCLE (Edgar, 2004).

3.3.6 Cultivation trials

Bacteria were cultivated from the landfill CLC_T2 sample (see Chapter 2 for sample collection and information); however, none of our characterized isolates were Tenericutes. Five colony morphologies were observed in the cultivation trials: 1) fried egg-like with a denser center and white; 2) small, dense, and white; 3) globular and colourless; 4) small and yellow; and 5) white and flat (Table 3.4). The five different isolate colony morphologies were each identified to the genus level based on 16S rRNA gene amplification, cloning, sequencing, and classification using Silva (Table 3.4). Isolates from the genera *Acinetobacter*, *Ochrobactrum*, and *Paracoccus* all belong to the phylum Proteobacteria, which is consistent with Proteobacteria being one of the most abundant phyla in the landfill leachate (see Chapter 2). I was curious whether any of the bacteria isolated were at high enough abundance in the landfill to have been detected in the metagenomes or 16S rRNA amplicon data. Three of the isolates' 16S rRNA genes (M2_C1_P2_C3, M3_C1_P3_M5, and M3_C1_P4_F1), matched to 16S rRNA gene sequences in the assembled metagenomes for samples CLC_T1 and CLC_T2 (Table 3.5) as well as to an ESV from the 16S rRNA gene amplicon sequencing (Tables 3.5 and 3.6). None of the other isolates have 99% matches for either metagenome sequences or ESVs. The inclusion of penicillin in the media was effective in preventing the growth of Gram-positive bacteria as all of the cultivated bacteria are Gram-negative except for *Corynebacterium*, a genus that contains multiple strains exhibiting antibiotic resistance to penicillin (Neemuchwala et al., 2018; Weiss et al., 1996).

Table 3.4: Cultivation trial isolate taxonomic affiliations based on 16S rRNA gene comparisons to the SILVA database (release 132) and their morphologies. Isolate codes refer to growth media number (M), colony morphology (C), streak plate number (P), and cloning plate isolate codes.

<i>Isolate</i>	<i>Silva identity (%)</i>	<i>Silva identification</i>	<i>Colony morphology</i>
<i>M2_C1_P2_C3</i>	97.81	<i>Acinetobacter</i>	White, denser centre, fried-egg like
<i>M3_C1_P3_M5</i>	97.81		
<i>M3_C1_P4_F1</i>	97.81		
<i>M3_C2_P1_A2</i>	99.93	<i>Paracoccus</i>	Small, white
<i>M3_C2_P2_E3</i>	98.60		
<i>M2_C3_P3_D2</i>	100	<i>Ochrobactrum</i>	Colourless, mucoid
<i>M4_C3_P1_H5</i>	99.58		
<i>M3_C4_P2_J1</i>	99.16	<i>Corynebacterium</i>	Small, yellow
<i>M3_C4_P4_G4</i>	99.43		
<i>M4_CU_P1_B6</i>	99.67	<i>Alcaligenes</i>	White, flat

Table 3.5: Isolates with representation in the metagenome sequences, based on BLAST of isolate 16S rRNA genes to assembled metagenomes. A threshold of 1×10^{-20} was used to filter BLAST hits. Abundance indicates the average fold-coverage of the scaffold in that metagenome assembly.

<i>Isolate</i>	<i>Scaffold Id</i>	<i>Sample</i>	<i>Identity (%)</i>	<i>Average scaffold coverage</i>
<i>M2_C1_P2_C3</i>	Ga0172378_10776477	CLC_T1	100	130.4
	Ga0172378_10929939	CLC_T1	99	105.6
	Ga0172377_10823095	CLC_T2	99	46.2
<i>M3_C1_P3_M5</i>	Ga0172378_10929939	CLC_T1	100	130.4
	Ga0172378_10776477	CLC_T1	100	105.6
	Ga0172377_10823095	CLC_T2	100	46.2
<i>M3_C1_P4_F1</i>	Ga0172378_10776477	CLC_T1	100	130.4
	Ga0172378_10929939	CLC_T1	99	105.6
	Ga0172377_10823095	CLC_T2	99	46.2

Table 3.6: Isolates with representation in the 16S rRNA gene amplicon exact sequence variants. All three isolates with a match were to the same ESV.

<i>Isolate</i>	<i>Pairwise identity (%)</i>	<i>Length</i>	<i>ESV_code</i>	<i>Sample</i>	<i>Relative Abundance</i>	<i>Silva Taxonomy</i>
<i>M2_C1_P2_C3</i>	99.3	271/273	ESV_761	CLC_T2	0.03% (in CLC_T2)	<i>Acinetobacter</i>
<i>M3_C1_P3_M5</i>	99.1	271/273				
<i>M3_C1_P4_F1</i>	99.1	271/273				

3.4 Conclusions

There were twelve *Tenericutes* MAGs and five *RFN20* MAGs that met our conservative quality criteria. The MAGs placed within four orders of the *Tenericutes* and within the order *RFN20*, sister to the *Erysipelotrichales*. The reconstructed genomes for these landfill populations demonstrated a range of genome sizes and gene numbers, with a commensurate range of associated metabolic

capacities. Based on the combination of genome characteristics, predicted metabolic capacities, and phylogenetic placements of these MAGs, we have generated hypotheses about their potential lifestyles. The RFN20 MAGs share some metabolic similarities to the *Erysipelotrichia* and “*Acholeplasma*” group in terms of sugar metabolisms, but have more limited amino acid synthesis capabilities. Many of the *Erysipelotrichia*, sister clade to the RFN20, have host-associated lifestyles, so it is possible that the RFN20 organisms are also host-associated and reliant on those hosts for metabolites. If the RFN20 populations in the landfill do not have hosts, it would be interesting to see how they are obtaining required metabolites given their more reduced genomes and metabolisms. Two of the Tenericutes MAGs, Bin_35_5 and Bin_25_6, placed with the *Izimaplasmatales* and have similar metabolic profiles to the available *Izimaplasma* genomes. As such, the two populations represented by these genomes are predicted to be free-living organisms. Four of the Tenericutes placed within the *Acholeplasmatales* order, and, although they do not share fatty acid synthesis capacities with the *Acholeplasmatales* reference genomes, they are otherwise predicted to be metabolically similar. Given that many *Acholeplasma* species are commensal, it is possible that these Tenericute MAGs represent commensal microorganisms, with partnerships with other microorganisms in the leachate or possibly with eukaryotic hosts within or prior to deposition in the landfill. The two Tenericutes assigned to the RF39 order are the closest phylogenetically to the *Mycoplasmatales* and have similar predicted metabolic capacities to the *Phytoplasma*, suggesting that these organisms may have parasitic lifestyles. There are no cultivated organisms from this order, and thus this group is still poorly characterized. Similarly, the ML615J-28 is another under-characterized group, into which four of our Tenericute MAGs were placed. Given the ML615J-28 phylogenetic placement close to the *Izimaplasmatales*, it is possible that these populations may also be free-living, as they possess partial pathways for amino acid synthesis and purine and pyrimidine synthesis. As well, the ML615J-28 MAGs Bin_17_2 and Bin_22_2_LW possess a complete non-mevalonate pathway for C5 isoprenoid biosynthesis, indicating the potential to make sterols. More research, including successful isolation of members of these under-studied lineages, is required to better characterize these new groups and understand their life histories.

Chapter 4

Conclusions and Future Directions

Landfills are unique built environments that harbour diverse communities of microorganisms. Previous studies of landfill microbial communities have begun to examine this diversity using 16S rRNA gene sequencing (Cardinali-Rezende *et al.*, 2016; Stamps *et al.*, 2016; Xu *et al.*, 2017) but have been unable to link the taxonomy of the bacteria and archaea present in landfills to potential functions in waste degradation. My study using metagenomics allowed for a more in-depth analysis of the microbial community to determine their potential roles and lifestyles. Landfill leachate contains a variety of metals, volatiles, and other compounds which can be hazardous to human and environmental health (Stamps *et al.*, 2016). The microbial populations within the leachate are often capable of degrading these compounds into less complex substances; however, this does not always result in lowered toxicity (Stamps *et al.*, 2016). Understanding the metabolic capabilities of the microbial community will better inform municipal waste management on what is being degraded in the landfill, what is being produced, and how these processes can be encouraged or controlled.

I conducted a survey of microbial community diversity across a Southern Ontario municipal landfill (Chapter 2). The microbial communities at the different leachate sampling sites appear very similar at the phylum level, with high abundances of Patescibacteria, Bacteroidota, Firmicutes, and Proteobacteria. Of the two groundwater sites, GW1, the impacted groundwater well, appears more like the leachate wells at the phylum level whereas GW2, the unimpacted groundwater well, exhibits a different phylum-level profile, with a higher abundance of Nanoarchaeota and Omnitrochota. Although the sites share similar phylum level profiles, the individual populations represented by the ESVs vary. Only one of the ESVs is present at greater than 1% abundance across the landfill, a species of *Shewanella*, and 99.8% of the ESVs are present at less than 0.05% abundance across the landfill. The number of ESVs above 1% relative abundance per sampling site is very low, ranging from zero for CLC_T2 to twelve for GW1. The vast majority of microbial populations in the landfill are present exclusively at one sample site, with only 37 of the 2,989 ESVs present at two or more sites. This suggests that each well catchment area within the landfill may act as its own ecosystem with its own unique microbial community, comprised of many low abundance organisms with a select few that are more abundant.

Of the studied sites, five had volatile compound concentration data for the sampling year and four had non-volatile compound data. From this data, we were able to determine that the geochemical

composition of these sites varied significantly; however, we were not able to correlate the geochemical data with the taxonomic data from ESVs due to the limited number of sites – there was insufficient sampling to allow robust statistical testing. As well, we used an average of April and October volatile and non-volatile compound concentrations to correlate with July biomass sampling, as the biomass and geochemistry sampling events were not conducted at the same time, which introduces potential inaccuracies to observed correlations. New biomass samples were collected from an expanded range of sites at the same landfill in October of 2017, coinciding exactly with geochemistry sampling. This expanded dataset and synced biomass and geochemistry data will give a more accurate assessment of the composition of the leachate and groundwater that the microbial communities are living in, and may allow for more accurate analyses to determine correlations between microbial community members and site geochemistry.

There is evidence of disturbance in the impacted groundwater well, both in terms of its microbial community and its geochemistry. The leachate leak from the area of LW3 to the groundwater at the GW1 site brings with it microbes and contaminants from the landfill. This influx resulted in lower evenness and richness for the GW1 community. Some of these microbes are able to survive in the more dilute groundwater, such that the phylum level profile for GW1 is similar to that of the leachate wells rather than that of the unimpacted groundwater well, GW2. However, there is little overlap between GW1 and either LW3 or GW2 at the ESV level, suggesting that the mixing of groundwater and leachate is a disturbance for both the native and introduced microbes. It is of future interest to monitor GW1 to see how this mixed community changes over time, as community succession and potential rebound will inform on the consequences of any other leaks, should they occur. Whether the GW1 community remains disturbed or switches to an alternate stable state will depend on the influx of microbes and contaminants into the system, and which, if any, of the native and introduced microbes are able to adapt to the mixed ecosystem. If the ecosystem is able to reach an alternate stable state, the microbial community may stabilize over time to an increased richness and evenness of species than is currently seen at this site. If the inputs from the leachate catchment around LW3 change in composition, the community at GW1 may experience additional disturbances that further alter the community or slow development of a stable state. Thus, the comparison between the leachate wells and the groundwater wells gives useful insight into the effects of leachate on groundwater ecosystems, and future monitoring of the GW1 microbial community will be valuable for understanding how these communities respond to disturbance.

The analyses demonstrate that the microbial community in the composite leachate cistern appears to turnover in a week. While the CLC_T1 and CLC_T2 samples had similar phylum level profiles, they did not share any ESVs. In fact, CLC_T1, collected on July 14, did not share any ESVs with any of the samples collected on July 20. This rate of microbial population turnover in the landfill was unexpected, as previous studies have not resampled communities on as short a time frame before (Cardinali-Rezende *et al.*, 2016). This quick turnover will need to be considered for future studies if attempting to compare community structure across samples from different years. It would be beneficial for future studies to determine if this rate of turnover is dependent on the mixing and aerating of the leachate in the cistern or whether it is a consistent pattern across the landfill. As well, it may be of interest to determine if this turnover rate varies seasonally. The average temperatures for the leachate wells and GW2 measured by the monitoring company were 11.9°C and 17.6°C in April and October 2016, respectively. This seasonal change in temperature may affect the growth rates of the microbes. Further, any seasonal changes in waste inputs to the landfill may introduce new microbes to the landfill that would further affect the community structure. Thus, the discovery of a rapid turnover of leachate microbial communities, if consistent across the landfill, will have implications for comparing microbial communities sampled at different times.

In a survey of Tenericute diversity and metabolic potential (Chapter 3), we found that our Tenericute MAGs were taxonomically placed across several different orders and that the MAGs initially classified as Erysipelotrichia belong to the uncharacterized order RFN20. The MAGs identified to the Izimaplasmatales and ML615J-28 are predicted to be free-living based on their phylogenetic placement and the greater biosynthetic potential encoded on their genomes. The MAGs identified to the Acholeplasmatales are predicted to be commensal microorganisms, either with partnerships with other microorganisms or with eukaryotic hosts. It is not clear if these commensal relationships occur within or prior to deposition in the landfill. The two MAGs assigned to the RF39 order are potentially parasitic, as they are closest phylogenetically to the Mycoplasmatales and are metabolically similar to the *Phytoplasma*, a group of plant parasites. The RFN20 MAGs are not as easily classified for predicted lifestyles, but, given their phylogenetic placement and metabolic potential, they may be host-associated with a greater reliance on their host(s) than members of the Erysipelotrichales or Acholeplasmatales. Thus, I predict the landfill Tenericutes and associated organisms possess a range of lifestyles from free-living to parasitic.

Given the phylogenetic and metabolic ranges of the Tenericutes, it is not surprising that we were unsuccessful in cultivating any representatives of this group. With the genomic expansion of this

radiation with the addition of our landfill MAGs, and our detailed information on their phylogenetic relationships, future work can investigate predicted metabolisms and the metabolisms of related organisms more closely, to better understand growth requirements for the different orders. Additionally, a liquid culture may be needed to cultivate Tenericutes, as they were harvested from a liquid environment. Notably, Skennerton *et al.* (2016) were able to cultivate *Izimaplasma* representatives in a liquid medium. For the non-Tenericute organisms that were cultivated, future directions include metabolic assays (*e.g.*, Biolog plates or similar broad assay) to determine what these organisms may be metabolizing in the landfill. For the isolates that matched to metagenome scaffolds, a genome-based metabolic analysis could also be conducted if a high-quality MAG exists for this organism. In future, improvements would need to be made to the cultivation trial procedures in order to attempt to cultivate landfill Tenericutes. However, metabolic studies can be conducted on the organisms that were cultivated to better understand their roles in the landfill.

Overall, the Southern Ontario landfill hosts a large diversity of bacteria and archaea that vary between sites and likely temporally in the landfill. The differences in geochemistry at the different sites are likely influencing the microbial heterogeneity, but this needs to be further studied. The metagenomic approach allowed for a more in-depth phylogenetic and metabolic analysis of the Tenericutes group. While there is still much to uncover about the diversity of microbes in landfills, including the lifestyles and roles of the Tenericutes in this unique environment, the research presented in this thesis provides an important piece to understanding landfill microbial diversity.

Bibliography

- Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* *11*, 1144.
- Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* *7*, 13219.
- Anwar, Y., El-Hanafy, A.A., Sabir, J.S.M., Al-Garni, S.M.S., Al-Ghamdi, K., Almehdar, H., and Waqas, M. (2017). Characterization of Mesophilic Bacteria Degrading Crude Oil from Different Sites of Aramco, Saudi Arabia. *Polycycl. Aromat. Compd.* *82*, 1–9.
- Atobe, H., Watabe, J., and Ogata, M. (1983). *Acholeplasma parvum*, a new species from horses. *Int. J. Syst. Bacteriol.* *33*, 344–349.
- Baker, B.J., and Banfield, J.F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* *44*, 139–152.
- Bareither, C.A., Wolfe, G.L., McMahon, K.D., and Benson, C.H. (2013). Microbial diversity and dynamics during methane production from municipal solid waste. *Waste Manag.* *33*, 1982–1992.
- Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* *37*, 852–857.
- Broun, R., and Sattler, M. (2016). A comparison of greenhouse gas emissions and potential electricity recovery from conventional and bioreactor landfills. *J. Clean. Prod.* *112*, 2664–2673.
- Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* *523*, 208.
- Burrell, P.C., Song, H., Clarke, W.P., and Blackall, L.L. (2004). Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. *Appl. Environ. Microbiol.* *70*, 2414–2419.

- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* *13*, 581.
- Cardinali-Rezende, J., Rojas-Ojeda, P., Nascimento, A.M.A., and Sanz, J.L. (2016). Proteolytic bacterial dominance in a full-scale municipal solid waste anaerobic reactor assessed by 454 pyrosequencing technology. *Chemosphere* *146*, 519–525.
- Chen, S., and Dong, X. (2005). *Proteiniphilum acetatigenes* gen. nov., sp. nov., from a UASB reactor treating brewery wastewater. *Int. J. Syst. Evol. Microbiol.* *55*, 2257–2261.
- Co, R. (2019). Microbial diversity and cellulosic capacity in municipal waste sites. University of Waterloo.
- Crusoe, M.R., Alameldin, H.F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., Charbonneau, A., Constantinides, B., Edvenson, G., Fay, S., et al. (2015). The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* *4*, 900.
- Davis, J.J., Xia, F., Overbeek, R.A., and Olsen, G.J. (2013). Genomes of the class Erysipelotrichia clarify the firmicute origin of the class Mollicutes. *Int. J. Syst. Evol. Microbiol.* *63*, 2727–2741.
- Delmont, T.O., and Eren, A.M. (2018). Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* *6*, e4320.
- Dutta, C., and Paul, S. (2012). Microbial Lifestyle and Genome Signatures. *Curr. Genomics* *13*, 153–162.
- Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* *32*, 1792–1797.
- Edward, D.G. (1971). Determination of sterol requirement for Mycoplasmatales. *J. Gen. Microbiol.* *69*, 205–210.
- Edward, D.G., and Freundt, E.A. (1956). The classification and nomenclature of organisms of the pleuropneumonia group. *J. Gen. Microbiol.* *14*, 197–207.
- Elmnasri, K., Hamdi, C., Ettoumi, B., Crotti, E., Guesmi, A., Najjari, A., Doudoumis, V., Boudabous, A., Daffonchio, D., Tsiamis, G., et al. (2018). Highly divergent Mollicutes symbionts coexist in the scorpion *Androctonus australis*. *J. Basic Microbiol.* *58*, 827–835.

- Epelde, L., Lanzén, A., Blanco, F., Urich, T., and Garbisu, C. (2015). Adaptation of soil microbial community structure and function to chronic metal contamination at an abandoned Pb-Zn mine. *FEMS Microbiol. Ecol.* *91*, 1–11.
- Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* *3*, e1319–e1319.
- Faith, D.P., and Baker, A.M. (2007). Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol. Bioinform. Online* *2*, 121–128.
- Ferris, F.G., Tazaki, K., and Fyfe, W.S. (1989). Iron oxides in acid mine drainage environments and their association with bacteria. *Chem. Geol.* *74*, 321–330.
- Freundt, E.A. (1975). Present Status of the Classification of the Order Mycoplasmatales, Class Mollicutes. *Proc. Indian Natl. Sci. Acad.* *41*, 285–297.
- Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* *13*, 133–146.
- Gibbons, S.M., and Gilbert, J.A. (2015). Microbial diversity--exploration of natural ecosystems and microbiomes. *Curr. Opin. Genet. Dev.* *35*, 66–72.
- Gilbert, J.A., and Stephens, B. (2018). Microbiology of the built environment. *Nat. Rev. Microbiol.* *16*, 661–670.
- Gupta, R.S., Sawnani, S., Adeolu, M., Alnajjar, S., and Oren, A. (2018). Correction to: Phylogenetic framework for the phylum Tenericutes based on genome sequence data: proposal for the creation of a new order Mycoplasmoidales ord. nov., containing two new families Mycoplasmoidaceae fam. nov. and Metamycoplasmataceae fam. nov. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* *111*, 2485–2486.
- Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L., et al. (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* *330*, 204–208.
- Hedlund, B.P., Dodsworth, J.A., Murugapiran, S.K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile 'microbial dark matter'. *Extremophiles* *18*, 865–875.

- Henson, M.W., Hanssen, J., Spooner, G., Fleming, P., Pukonen, M., Stahr, F., and Thrash, J.C. (2018). Nutrient dynamics and stream order influence microbial community patterns along a 2914 kilometer transect of the Mississippi River. *Limnol. Oceanogr.* *63*, 1837–1855.
- Hilger, H.H., and Barlaz, M.A. (2006). Anaerobic decomposition of refuse in landfills and methane oxidation in landfill covers. In *Manual of Environmental Microbiology*, (Washington, DC: American Society of Microbiology), pp. 818–842.
- Hoornweg, D., and Bhada-Tata, P. (2012). What a Waste: A Global Review of Solid Waste Management. *Urban Dev. Ser. – Knowl. Pap.* *15*, 1–116.
- Hoornweg, D., Bhada-Tata, P., and Kennedy, C. (2013). Waste production must peak this century. *Nature* *502*, 615–617.
- Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams, K.H., Tringe, S.G., and Banfield, J.F. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* *1*, 22.
- Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hermsdorf, A.W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* *1*, 1–6.
- Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Tennessen, K., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.M.A., et al. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand. Genomic Sci.* *11*, 1–5.
- ICSB Subcommittee on the Taxonomy of Mollicutes. (1995). Revised minimum standards for description of new species of the Class Mollicutes (division Tenericutes). *Int. J. Syst. Bacteriol.* *45*, 605–612.
- Idris, A., Inanc, B., and Hassan, M.N. (2004). Overview of waste disposal and landfills/dumps in Asian countries. *J. Mater. Cycles Waste Manag.* *6*, 104–110.
- Iehata, S., Valenzuela, F., and Riquelme, C. (2015). Analysis of bacterial community and bacterial nutritional enzyme activity associated with the digestive tract of wild Chilean octopus (*Octopus mimus* Gould, 1852). *Aquac. Res.* *46*, 861–873.

- Janda, J.M., and Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* *45*, 2761–2764.
- Joblin, K.N., and Naylor, G.E. (2002). The ruminal mycoplasmas : A review. *J. Appl. Anim. Res.* *21*, 161–179.
- Kanno, M., Katayama, T., Morita, N., Tamaki, H., Hanada, S., and Kamagata, Y. (2015). *Catenisphaera adipataaccumulans* gen. nov., sp. nov., a member of the family Erysipelotrichaceae isolated from an anaerobic digester. *Int. J. Syst. Evol. Microbiol.* *65*, 805–810.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* *30*, 772–780.
- Keller, M., and Zengler, K. (2004). Tapping into microbial diversity. *Nat. Rev. Microbiol.* *2*, 141–150.
- King, G.M. (2014). Urban microbiomes and urban ecology: How do microbes in the built environment affect human sustainability in cities? *J. Microbiol.* *52*, 721–728.
- Köchling, T., Sanz, J.L., Gavazza, S., and Florencio, L. (2015). Analysis of microbial community structure and composition in leachates from a young landfill by 454 pyrosequencing. *Appl. Microbiol. Biotechnol.* *99*, 5657–5668.
- Konopka, A., Zakharova, T., Bischoff, M., Oliver, L., Nakatsu, C., and Turco, R.F. (1999). Microbial biomass and activity in lead-contaminated soil. *Appl. Environ. Microbiol.* *65*, 2256–2259.
- Koskella, B., and Vos, M. (2015). Adaptation in Natural Microbial Populations. *Annu. Rev. Ecol. Evol. Syst.* *46*, 503–522.
- Kowarsky, M., Camunas-Soler, J., Kertesz, M., De Vlaminc, I., Koh, W., Pan, W., Martin, L., Neff, N.F., Okamoto, J., Wong, R.J., et al. (2017). Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. *Proc. Natl. Acad. Sci. U. S. A.* *114*, 9623–9628.
- Kuzuyama, T., and Seto, H. (2012). Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* *88*, 41–52.
- Lazarev, V.N., Levitskii, S.A., Basovskii, Y.I., Chukin, M.M., Akopian, T.A., Vereshchagin, V. V., Kostjukova, E.S., Kovaleva, G.Y., Kazanov, M.D., Malko, D.B., et al. (2011). Complete genome and proteome of *Acholeplasma laidlawii*. *J. Bacteriol.* *193*, 4943–4953.

- Leclercq, S., Dittmer, J., Bouchon, D., and Cordaux, R. (2014). Phylogenomics of “candidatus Hepatoplasma crinochetorum,” a lineage of mollicutes associated with noninsect arthropods. *Genome Biol. Evol.* *6*, 407–415.
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* *47*, W256–W259.
- Leung, M.H.Y., and Lee, P.K.H. (2016). The roles of the outdoors and occupants in contributing to a potential pan-microbiome of the built environment: A review. *Microbiome* *4*, 1–15.
- Lin, B., Monreal, C.M., Tambong, J.T., and Miguez, C.B. (2009). Phylogenetic analysis of methanotrophic communities in cover soils of a landfill in Ontario. *1112*, 1103–1112.
- Lo, W., Chen, L., Chung, W., Gasparich, G.E., and Kuo, C. (2013). Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* *14*, 1.
- Locey, K.J., and Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 5970–5975.
- Lu, Z., He, Z., Parisi, V.A., Kang, S., Deng, Y., Van Nostrand, J.D., Masoner, J.R., Cozzarelli, I.M., Suflita, J.M., and Zhou, J. (2012). GeoChip-based analysis of microbial functional gene diversity in a landfill leachate-contaminated aquifer. *Environ. Sci. Technol.* *46*, 5824–5833.
- Ludwig, W., and Schleifer, K. (2005). Molecular phylogeny of bacteria based on comparative sequence analysis of conserved genes. *Microb. Phylogeny Evol. Concepts Controv.* 70–98.
- Ludwig, W., Euzeby, J., and Whitman, W.B. (2010). Road map of the phyla Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes. In *BERGEY’S MANUAL® OF Systematic Bacteriology Second Edition*, (New York, NY: Springer), pp. 1–19.
- Lutz, S., Anesio, A.M., Edwards, A., and Benning, L.G. (2017). Linking microbial diversity and functionality of arctic glacial surface habitats. *Environ. Microbiol.* *19*, 551–565.
- Lynch, M.D.J., and Neufeld, J.D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* *13*, 217.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* *17*, 10–12.

- Martini, M., Marcone, C., Lee, I.-M., and Firrao, G. (2014). The Family Achleplasmataceae (Including Phytoplasmas). In *The Prokaryotes: Firmicutes and Tenericutes*, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 469–504.
- McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* *10*, 13–26.
- McDonald, J.E., Allison, H.E., and McCarthy, A.J. (2010). Composition of the landfill microbial community as determined by application of domain- And group-specific 16S and 18S rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* *76*, 1301–1306.
- McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* *8*, 1–11.
- Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A., and Doxey, A.C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* *47*, 4442–4448.
- Merhej, V., Royer-Carenzi, M., Pontarotti, P., and Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* *4*, 13.
- von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J., Ward, N., and Bork, P. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* *315*, 1126–1130.
- Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pp. 1–8.
- Mollicutes, I.S. on the T. of (1995). Revised minimal standards for description of new species of the class Mollicutes (division Tenericutes). *Int. J. Syst. Bacteriol.* *45*, 605–612.
- Müller, A.K., Westergaard, K., Christensen, S., and Sørensen, S.J. (2002). The diversity and function of soil microbial communities exposed to different disturbances. *Microb. Ecol.* *44*, 49–58.
- Murry, R.G.E. (1984). No Title. In *BERGEY'S MANUAL® OF Systematic Bacteriology Volume 1*, N.R. Krieg, and J.G. Holt, eds. (Baltimore: Williams & Wilkins), pp. 31–36.

Neemuchwala, A., Soares, D., Ravirajan, V., Marchand-Austin, A., Kus, J. V., and Patel, S.N. (2018). In vitro antibiotic susceptibility pattern of non-diphtheriae *Corynebacterium* isolates in Ontario, Canada, from 2011 to 2016. *Antimicrob. Agents Chemother.* *62*, e01776-17.

Ogawa, Y., Ooka, T., Shi, F., Ogura, Y., Nakayama, K., Hayashi, T., and Shimoji, Y. (2011). The genome of *Erysipelothrix rhusiopathiae*, the causative agent of swine erysipelas, reveals new insights into the evolution of Firmicutes and the organism's intracellular adaptations. *J. Bacteriol.* *193*, 2959–2971.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2018). *vegan: Community Ecology Package*. R package version 2.5-1.

Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep : a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* *11*, 2864–2868.

Oremland, R.S., and Stolz, J.F. (2005). Arsenic, microbes and contaminated aquifers. *Trends Microbiol.* *13*, 45–49.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003). Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* *5*, 650–659.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* *36*, 996.

Pearce, D.A., Bridge, P.D., Hughes, K.A., Sattler, B., Psenner, R., and Russell, N.J. (2009). Microorganisms in the atmosphere over Antarctica. *FEMS Microbiol. Ecol.* *69*, 143–157.

Pielou, E.C. (1966). The measurement of diversity in different types of biological collections. *J. Theor. Biol.* *13*, 131–144.

Polak-Vogelzang, A.A., Brugman, J., and Reijgers, R. (1987). Comparison of two methods for detection of Mollicutes (*Mycoplasmatales* and *Acholeplasmatales*) in cell cultures in The Netherlands. *Antonie Van Leeuwenhoek* *53*, 107–118.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* *28*, 1823–1829.

- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* *41*, D590–D596.
- Rakoczy, J., Feisthauer, S., Wasmund, K., Bombach, P., Neu, T.R., Vogt, C., and Richnow, H.H. (2013). Benzene and sulfide removal from groundwater treated in a microbial fuel cell. *Biotechnol. Bioeng.* *110*, 3104–3113.
- Ramasamy, D., Lagier, J.C., Nguyen, T.T., Raoult, D., and Fournier, P.E. (2013). Non contiguous-finished genome sequence and description of *Dielma fastidiosa* gen. nov., sp. nov., a new member of the Family Erysipelotrichaceae. *Stand. Genomic Sci.* *8*, 336–351.
- Rappé, M.S., and Giovannoni, S.J. (2003). The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* *57*, 369–394.
- Razin, S. (2006). The Genus *Mycoplasma* and Related Genera (Class Mollicutes). In *Prokaryotes*, (Springer US), pp. 836–904.
- Reese, A.T., Pereira, F.C., Schintlmeister, A., Berry, D., Wagner, M., Hale, L.P., Wu, A., Jiang, S., Durand, H.K., Zhou, X., et al. (2018). Microbial nitrogen limitation in the mammalian large intestine. *Nat. Microbiol.* *3*, 1441–1450.
- Rose, D.L., Tully, J.G., and Del Giudice, R.A. (1980). *Acholeplasma morum*, a new non-sterol-requiring species. *Int. J. Syst. Bacteriol.* *30*, 647–654.
- Ross, A.A., Doxey, A.C., and Neufeld, J.D. (2017). The Skin Microbiome of Cohabiting Couples. *MSystems* *2*, 1–15.
- Savage, A.M., Hills, J., Driscoll, K., Fergus, D.J., Grunden, A.M., and Dunn, R.R. (2016). Microbial diversity of extreme habitats in human homes. *PeerJ* *4*, e2376.
- Sawamura, H., Yamada, M., Endo, K., Soda, S., Ishigaki, T., and Ike, M. (2010). Characterization of microorganisms at different landfill depths using carbon-utilization patterns and 16S rRNA gene based T-RFLP. *J. Biosci. Bioeng.* *109*, 130–137.
- Scholz, M.B., Lo, C.C., and Chain, P.S.G. (2012). Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* *23*, 9–15.

- Schuler, C.G., Havig, J.R., and Hamilton, T.L. (2017). Hot Spring Microbial Community Composition, Morphology, and Carbon Fixation: Implications for Interpreting the Ancient Rock Record. *Front. Earth Sci.* *5*, 1–17.
- Schulz, F., Eloë-Fadrosch, E.A., Bowers, R.M., Jarett, J., Nielsen, T., Ivanova, N.N., Kyrpides, N.C., and Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. *Microbiome* *5*, 140.
- Seitz, K.W., Dombrowski, N., Eme, L., Spang, A., Lombard, J., Sieber, J.R., Teske, A.P., Ettema, T.J.G., and Baker, B.J. (2019). Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* *10*, 1822.
- Siewert, C., Hess, W.R., Duduk, B., Huettel, B., Reinhardt, R., Büttner, C., and Kube, M. (2014). Complete genome determination and analysis of *Acholeplasma oculi* strain 19L, highlighting the loss of basic genetic features in the *Acholeplasmataceae*. *BMC Genomics* *15*, 1–16.
- Skennerton, C.T., Haroon, M.F., Briegel, A., Shi, J., Jensen, G.J., Tyson, G.W., and Orphan, V.J. (2016). Phylogenomic analysis of Candidatus “*Izimaplasma*” species: Free-living representatives from a *Tenericutes* clade found in methane seeps. *ISME J.* *10*, 2679–2692.
- Smith, D.G., Russell, W.C., Ingledew, W.J., and Thirkell, D. (1993). Hydrolysis of urea by *Ureaplasma urealyticum* generates a transmembrane potential with resultant ATP synthesis. *J. Bacteriol.* *175*, 3253–3258.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored rare biosphere. *Proc. Natl. Acad. Sci.* *103*, 12115–12120.
- Song, L., Wang, Y., Tang, W., and Lei, Y. (2015). Bacterial community diversity in municipal waste landfill sites. *Appl. Microbiol. Biotechnol.* *99*, 7745–7756.
- Srinivasan, R., Karaoz, U., Volegova, M., MacKichan, J., Kato-Maeda, M., Miller, S., Nadarajan, R., Brodie, E.L., and Lynch, S. V (2015). Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS One* *10*, e0117617.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* *30*, 1312–1313.

- Stamps, B.W., Lyles, C.N., Suflita, J.M., Masoner, J.R., Cozzarelli, I.M., Kolpin, D.W., and Stevenson, B.S. (2016). Municipal solid waste landfills harbor distinct microbiomes. *Front. Microbiol.* 7, 534.
- Terrat, S., Horrigue, W., Dequietd, S., Saby, N.P.A., Lelièvre, M., Nowak, V., Tripied, J., Régnier, T., Jolivet, C., Arrouays, D., et al. (2017). Mapping and predictive variations of soil bacterial richness across France. *PLoS One* 12, 5–8.
- Tremblay, J., Singh, K., Fern, A., Kirton, E., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J., and Tringe, S. (2015). Primer and platform effects on 16S rRNA tag sequencing . *Front. Microbiol.* 6, 771.
- Trosvik, P., and de Muinck, E.J. (2015). Ecology of bacteria in the human gastrointestinal tract--- identification of keystone and foundation taxa. *Microbiome* 3, 44.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V. V, Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43.
- Vavourakis, C.D., Ghai, R., Rodriguez-Valera, F., Sorokin, D.Y., Tringe, S.G., Hugenholtz, P., and Muyzer, G. (2016). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front. Microbiol.* 7, 1–18.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* (80-.). 304, 66–74.
- Ventorino, V., Pascale, A., Adamo, P., Rocco, C., Fiorentino, N., Mori, M., Faraco, V., Pepe, O., and Fagnano, M. (2018). Comparative assessment of autochthonous bacterial and fungal communities and microbial biomarkers of polluted agricultural soils of the Terra dei Fuochi. *Sci. Rep.* 8, 1–13.
- Verbarg, S., Rheims, H., Emus, S., Fruhling, A., Kroppenstedt, R.M., Stackebrandt, E., and Schumann, P. (2004). *Erysipelothrix inopinata* sp. nov., isolated in the course of sterile filtration of vegetable peptone broth, and description of *Erysipelotrichaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 54, 221–225.
- Větrovský, T., and Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* 8, e57923.

- Vitorino, L., and Bessa, L. (2018). Microbial Diversity: The Gap between the Estimated and the Known. *Diversity* 10, 46.
- Vogt, C., Kleinsteuber, S., and Richnow, H.H. (2011). Anaerobic benzene degradation by bacteria. *Microb. Biotechnol.* 4, 710–724.
- De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.-H., and Whitman, W.B. (2009). Volume Three The Firmicutes.
- Weiss, J. V., Rentz, J.A., Plaia, T., Neubauer, S.C., Merrill-Floyd, M., Lilburn, T., Bradburne, C., Megonigal, J.P., and Emerson, D. (2007). Characterization of neutrophilic Fe(II)-oxidizing bacteria isolated from the rhizosphere of wetland plants and description of *Ferritrophicum radicolica* gen. nov. sp. nov., and *Sideroxydans paludicola* sp. nov. *Geomicrobiol. J.* 24, 559–570.
- Weiss, K., Laverdière, M., and Rivest, R. (1996). Comparison of antimicrobial susceptibilities of *Corynebacterium* species by broth microdilution and disk diffusion methods. *Antimicrob. Agents Chemother.* 40, 930–933.
- Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185.
- Will, C., Thürmer, A., Wollherr, A., Nacke, H., Herold, N., Schrumpf, M., Gutknecht, J., Wubet, T., Buscot, F., and Daniel, R. (2010). Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Appl. Environ. Microbiol.* 76, 6751 LP – 6759.
- Xu, S., Lu, W., Liu, Y., Ming, Z., Liu, Y., Meng, R., and Wang, H. (2017). Structure and diversity of bacterial communities in two large sanitary landfills in China as revealed by high-throughput sequencing (MiSeq). *Waste Manag.* 63, 41–48.
- Zhao, S., Guo, Y., Sheng, Q., and Shyr, Y. (2014). Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinformatics* 15, P16.

Appendix A

Synthetic Leachate

Table A1: Synthetic Leachate recipe used for Tenericute growth media during cultivation trials. Recipe was developed by Co (2019).

	Amount	Units	Volume of dH ₂ O
Solution 1		ml	200
CaCl ₂ x 2H ₂ O	870	mg	-
Solution 2		ml	200
MgSO ₄	54	mg	-
MgCl ₂ x 6H ₂ O	1083	mg	-
Solution 3		ml	100
K ₂ HPO ₄	30	mg	-
Solution 4		ml	500
KHCO ₃	312	mg	-
K ₂ CO ₃	324	mg	-
NaCl	745	mg	-
NaHCO ₃	1558	mg	-
NaNO ₃	26	mg	-
NH ₄ HCO ₃	1430	mg	-
CO(NH ₂) ₂ (urea)	407	mg	-
Metal Stock Solution	1	ml	-
Solution 5 - Metal Stock Solution		-	-
Al ₂ (SO ₄) ₃ x 16 H ₂ O	30	mg	-
CoSO ₄ x 5 H ₂ O	150	mg	-
CuSO ₄ x 5 H ₂ O	40	mg	-
FeSO ₄	4000	mg	-
H ₃ BO ₃	19446	mg	-
MnSO ₄ x H ₂ O	2453	mg	-
(NH ₄) ₆ Mo ₇ O ₂₄ x 4 H ₂ O	50	mg	-
NiSO ₄ x 6 H ₂ O	500	mg	-
ZnSO ₄ x 7 H ₂ O	50	mg	-
96% concentrated H ₂ SO ₄	2	ml	-
dH ₂ O	top up to 1 L	-	-