Microbial diversity and cellulosic capacity in municipal waste sites

by

Rebecca Co

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

Waterloo, Ontario, Canada, 2019

Author's Declaration

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

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

Statement of Contributions

In Chapter 2, the sampling and DNA extraction and sequencing of samples (Section 2.2.1 - 2.2.2) were carried out by Dr. Aneisha Collins-Fairclough and Dr. Melessa Ellis. The work described in Section 2.2.3 Metagenomic pipeline and onwards was done by the thesis's author. Sections 2.2.1 Sample collection – 2.2.4 16S rRNA gene community profile were previously published in *Widespread antibiotic, biocide, and metal resistance in microbial communities inhabiting a municipal waste environment and anthropogenically impacted river* by Aneisha M. Collins-Fairclough, Rebecca Co, Melessa C. Ellis, and Laura A. Hug. 2018. mSphere: e00346-18. The writing and analyses incorporated into this chapter are by the thesis's author.

Abstract

Cellulose is the most abundant organic compound found on earth. Cellulose's recalcitrance to hydrolysis is a major limitation to improving the efficiency of industrial applications. The biofuel, pulp and paper, agriculture, and textile industries employ mechanical and chemical methods of breaking down cellulose. Enzymatic methods are attractive choices for industry due to their selectivity in their mode of action and high product yields. However, cellulases are not as economic as mechanical means of degrading cellulose, and few cellulases are optimized for large scales. Investigating the cellulolytic microbiome and functional potential of municipal waste sites, which house large amounts of paper waste, can identify novel cellulose degraders robust for industrial applications.

The microbial diversity and metabolic potential in landfills have not been well studied. In this thesis, the cellulose degradation capacity was investigated at two municipal waste sites (MWS). First, the microbial composition and the cellulose degradation capacity of a leachate pond from a dump in Jamaica and the river adjacent to the dump were assessed using metagenomics. The diversity of metagenome-assembled metagenomes (MAGs) was greater in the leachate compared to the river, with thirteen high-quality MAGs identified across seven phyla, including Bacteroidetes, Proteobacteria and Firmicutes. In contrast, two high-quality MAGs, both members of the Proteobacteria, were reconstructed from the river metagenome. A MAG assigned to the candidate phylum CPR2 is the first candidate phylum radiation MAG to be reported from a landfill. The metagenomes were screened for genes belonging to glycosyl hydrolase (GH) families containing cellulases as a measure of cellulolytic potential at the sites. Beta-glucosidases were detected at both sites. In the metagenomes, the taxonomic affiliation of most potential cellulases in the leachate metagenome were to the Bacteroidetes, Firmicutes, Actinobacteria, Spirochaetes,

and Tenericutes, whereas Bacteroidetes and Proteobacteria cellulases were most abundant in the river metagenome. The microbial composition of the leachate and river did not overlap based on read mapping, suggesting no contamination of the river by the leachate at the times and sites sampled.

Secondly, the cellulolytic microbial diversity was also analyzed in six metagenomes from a landfill in Southern Ontario. The samples included a composite leachate cistern (CLC), three leachate wells, and one groundwater well. Twelve GH families containing cellulases were detected across the six metagenomes, with genes from GH3 and GH5 being the most prevalent. Beta-glucosidases and endocellulases were detected across all sites, but exocellulases were only detected in some of the leachate sites and the groundwater well. A large number of hypothetical proteins and non-specifically annotated proteins were also detected across all sites, which likely represent novel carbohydrate-modifying enzymes. The majority of the potential cellulase genes across the six sites were affiliated with the Bacteroidetes and Firmicutes.

Thirdly, the potential cellulolytic capacity established from the metagenomes from the Ontario landfill was confirmed by enrichment cultivations of leachate biomass grown in synthetic leachate amended with cellulose. Several isolates from the enrichment cultures showed carboxymethylcellulose and cellobiose degradation capacities, signifying endocellulase and beta-glucosidase activities. Results from 16S rRNA gene amplicon sequencing of copy-paper, cardboard, newsprint, and filter paper-enriched cultures showed enrichment of exact sequence variants assigned to *Paenibacillus*, *Cytophaga*, and *Proteiniphilum* bacteria over time. The research in this thesis represents the first connections between the cellulolytic potential and relevant taxonomic groups in MWS to cellulose degradation by isolates enriched from landfill leachate.

Acknowledgements

I would like to thank Dr. Laura Hug for accepting me as her first graduate student, teaching me metagenomics in spurts, providing feedback in the darkest hours, enriching many aspects of my graduate experience, and trusting me. I couldn't have asked for more. Thank you to my committee members: Dr. Josh Neufeld for letting me use the equipment in his lab for library preparation and sequencing, providing robust feedback, and wonderful conversations. Thank you to Dr. Andrew Doxey for his valuable feedback in the bioinformatic side of my project. Thank you to Dr. Maria Strack for strengthening my project through new and refreshing viewpoints.

I would also like to thank Dr. Aneisha Collins-Fairclough for sharing the Riverton Dump data with us – our collaboration was a pleasure. Thank you to the Joint Genome Institute for sequencing our southern Ontario landfill samples. Thank you to the engineers and consulting company at the Ontario landfill for their help and supervision in sampling. Thank you to Dr. Kelly Wrighton and Richard Wolfe from Colorado State University for allowing us to use their in-house metagenomic pipeline in processing the Jamaican datasets. Thank you to Michael Mansfield for showing me how to navigate and bin in Anvi'o, sharing your wisdom, and your continual offers to help me. Thank you to Katja Engel who I've been incredibly grateful for her expertise and unyielding patience with her guidance in 16S rRNA gene amplicon sequencing and library preparation. Thank you to Ching-Lung Huang with her relentless cheerfulness for supporting me while we were preparing to sequence our cultures. Thank you to Dr. Trevor Charles for allowing me to use the spectrophotometer in his lab. Thank you to Karen Miinch for providing bacterial cultures from the American Type Culture Collection. Thank you to both Karen Miinch and Susanne Vesely for so ever kindly providing me with useful materials to support my lab work as our lab started up. Thank

you to Mishi Groh for allowing me to have some Congo red dye used for staining.

Thank you to previous and current Hug Lab members: Veronica, Angus, Alex, Nikhil, and Lisa, for sample collection, script sharing, support, and encouragement. You are a wonderful bunch. I would also like to thank Jackson Tsuji for his offer to help me in MetAnnotate (although I never ended up using it in my analyses) and sharing his valuable insights in bioinformatics and metagenomics.

Lastly, thank you to my family for all their love and support. I would also like to thank the following people for lending their ears and shoulders, always looking out for me, and sharing laughs: Barb B., Susanne, Karen, Joseph, Max, Donya, Emilie, Vivian, Barb K., and Bruce. Thank you to the amazing and generous people in the Biology Department. I am truly grateful from the bottom of my heart for all of you when the going got tough.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

ß	Beta	FP	filter paper
^{13}C	Carbon-13	g	gram
°C	degrees Celsius	Gb	gigabase pair
µl	microliter	GH	glycosyl hydrolase
µm	micrometre	H_2	hydrogen gas
%	percent	HCl	hydrochloric acid
$A_{260/280}$	Absorbance 260 nm:280 nm	HMM	hidden Markov model
ATCC	American Type Culture Collection	JUL	July
ATP	adenosine triphosphate	kb	kilobase pair
avg	average	L	litre
bp	base pair	L/s	litre per second
CB	cardboard	LB	Leachate Bin
CBM	carbohydrate binding module	LW	leachate well
CLC	composite leachate cistern	m	metre
cm	centimetre	M	molar
CMC	carboxymethylcellulose	MAG	metagenome-assembled genome
CO_2	carbon dioxide		
COG	Clusters of Orthologous Groups	mbps	million base pairs
CP	copy paper	MCC	microcrystalline cellulose
CPR	Candidate Phyla Radiation	min	minute
CTRL	control	mg	mg
dH_2O	deionized water	ml	millilitre
DNA	deoxyribonucleic acid	mm	millimetre
dNTP	deoxyribonucleotide triphosphate	N	normal
		NaCl	sodium chloride
dsDNA	double-stranded DNA	ng	nanogram
e	e-value	nM	nanomolar
EC	Enzyme Commission	No.	number
ESV	exact sequence variant	OP	Obsidian Pool

ORF	open reading frame	x	times (coverage)
OTU	operational taxonomic unit		
OW	Observation Well		
PCE	tetrachloroethene		
PCR	polymerase chain reaction		
pH	power of hydrogen		
PH1	Pumphouse 1		
PH3	Pumphouse 3		
PUL	polysaccharide utilizing loci		
RB	River Bin		
RpL	ribosomal protein large subunit		
rpm	revolutions per minute		
RpS	ribosomal protein small subunit		
rRNA	ribosomal ribonucleic acid		
S_x	subculture x		
sd	standard deviation		
SRB	sulfur-reducing bacteria		
str.	strain		
T	time		
Taq	*Thermus aquaticus*		
TAE	tris-acetate EDTA		
TB	terabyte		
TCE	tricholoroethene		
TIM	triose-phosphate isomerase		
UV	ultraviolet light		
v.	version		
V4	hypervariable region 4		
V5	hypervariable region 5		
w/v	weight/volume		

Chapter 1: Introduction

Cellulose is the most abundant organic compound found on earth. It has been widely exploited in industrial applications, namely for pulp and paper, biofuel, agriculture, textile, and pharmaceutical products (Delmer and Haigler, 2002; Kuhad *et al.*, 2011). Cellulose's abundance, availability, and renewability are invaluable for large-scale industrial and biotechnology applications. An important limitation of cellulose is its recalcitrance to decomposition by hydrolysis (Pérez *et al.*, 2002). Cellulose degradation has been a topic of study for more than 140 years, with a focus on improving the process's efficiency. Earliest research investigated the products resulting from cellulose hydrolysis by sulphuric acid (Scientific Intelligence, 1860). Physical and thermochemical pretreatments of cellulose are currently the preferred cellulose hydrolysis methods for their speed and ability to degrade a wide range of lignocellulosic feedstocks continuously (Kumar and Sharma, 2017). Enzymatic treatments, especially when combined with thermochemical pretreatments, are attractive alternatives because of increased specificity in the desired products, low energy usage, and lowered toxic waste production compared to chemical processes (Mathews *et al.*, 2015; Kumar and Sharma, 2017). Research targeting discovery of novel microbial cellulases that enhance efficiency in industrial processes is a much needed area of study to harness the abundance of cellulose available for industrial applications.

1.1 Cellulose and cellulases

Cellulases are a family of enzymes that hydrolyze the beta-1,4-glycosidic bonds linearly connecting glucose subunits in cellulose molecules. Cellulases may also hydrolyze the hydrogen bonds that laterally interlink multiple chains of cellulose forming microfibrils (Béguin and Aubert, 1994). Bundles of microfibrils form fibrils, which mostly exist in a crystalline form, interspersed

with amorphous regions where the structure of the cellulose is less organized and more weakly linked. In addition to cellulose, hemicellulose and lignin are the main molecules intricately complexed in the matrix that makes up the majority of plant biomass: lignocellulose. To degrade cellulose, lignin and hemicellulose must first be degraded by their respective enzymes (Pérez *et al.*, 2002). Cellulose contributes to the physical support and defensive role of the plant cell wall as an insoluble, ordered, and uniformly structured molecule (Béguin and Aubert, 1994). A cocktail of three types of cellulases are needed to completely hydrolyze this structured polymer: endocellulase [EC 3.2.1.4], which randomly cleaves the glycosidic bond in the middle of the glucose chain; exocellulase (cellobiohydrolase) [EC 3.2.1.91], which cleaves near the end of the chain, releasing cellobiose; and beta-glucosidase [EC 3.2.1.21], which cleaves cellobiose, releasing glucose (Mathews *et al.*, 2015). Cellulases can either occur as free exoenzymes that bind to cellulose or as cellulosomes, which are complexes of enzymes attached to the cell wall of certain anaerobic cellulose degraders (Cragg *et al.*, 2015).

1.2 Microbial diversity of cellulases

Biochemically characterized cellulose degraders have been identified across the three domains of life, with a historical focus on cellulases identified from organisms that consume plant materials such as ruminants, herbivores, termites, and fungi (Güllert *et al.*, 2016). From the CAZy database (Lombard *et al.*, 2014) cellulolytic enzymes have been characterized across 16 glycosyl hydrolase families, and are found in 118 genera of bacteria; 164 genera of eukaryotes, of which 82 are fungi; and 8 genera of archaea (Table 1.1). Fungi have historically been the dominant microorganisms

Table 1.1. Microorganisms that possess characterized cellulases across the tree of life. As of March 2018, a total of 18 phyla and 201 genera across the tree of life contain members that have characterized cellulases, from 16 glycosyl hydrolase (GH) families (GH1, -3, -5, -6, -7, -8, -9, -12, -30, -44, -45, -48, -51, -74, -116, and -124) according to the CAZy database (Lombard *et al.*, 2014). The number in brackets beside each domain or eukaryotic group indicates the number of phyla in that group that contains members possessing cellulases. The number in brackets beside each phylum indicates the number of genera in that phylum that contains members possessing cellulases.

	Phyla	**Genera**			
Bacteria (12)	Actinobacteria (23)	*Acidothermus* *Actinomyces* *Actinosynnema* *Aeromicrobium* *Bifidobacterium* *Cellulomonas*	*Cellulosimicrobium* *Clavibacter* *Gordonia* *Microbacterium* *Micrococcus* *Micromonospora*	*Mycobacterium* *Pseudarthrobacter* *Pseudonocardia* *Saccharopolyspora* *Sanguibacter* *Streptomyces*	*Terrabacter* *Thermobifida* *Thermobispora* *Thermomonospora* *Xylanimicrobium*
	Aquificae (1)	*Aquifex*			
	Bacteroidetes (8)	*Bacteroides* *Cellulophaga* *Cytophaga*	*Elizabethkingia* *Flavobacterium* *Mucilaginibacter*	*Prevotella* *Rhodothermus*	
	Chloroflexi (3)	*Roseiflexus*	*Thermobaculum*	*Thermomicrobium*	
	Deinococcus-Thermus (3)	*Deinococcus*	*Meiothermus*	*Thermus*	
	Dictyoglomi (1)	*Dictyoglomus*			
	Enterobacter (3)	*Pantoea*	*Pectobacterium*	*Salmonella*	
	Fibrobacteres (1)	*Fibrobacter*			
	Firmicutes (24)	*Acetivibrio* *Alicyclobacillus* *Anoxybacillus* *Bacillus* *Butyrivibrio* *Caldanaerobacter*	*Caldanaerobius* *Caldicellulosiruptor* *Cellulosilyticum* *Clostridium* *Eubacterium* *Exiguobacterium*	*Geobacillus* *Halothermothrix* *Klebsiella* *Lachnoclostridium* *Lactobacillus* *Oenococcus*	*Paenibacillus* *Ruminiclostridium* *Ruminococcus* *Salipaludibacillus* *Thermoanaerobacter* *Thermoanaerobacterium*
	Proteobacteria (42)	*Agrobacterium* *Azoarcus* *Azorhizobium* *Caulobacter* *Cellvibrio* *Desulfotalea* *Dickeya* *Enterobacter* *Erwinia* *Escherichia* *Hahella*	*Halomonas* *Jeongeupia* *Komagataeibacter* *Legionella* *Lysobacter* *Magnetospirillum* *Marinomonas* *Martelella* *Myxobacter* *Neisseria* *Niveispirillum*	*Novosphingobium* *Photobacterium* *Pseudoalteromonas* *Pseudomonas* *Ralstonia* *Rhizobium* *Rhodobacter* *Rhodopseudomonas* *Saccharophagus* *Salinivibrio* *Serratia*	*Sinorhizobium* *Sphingomonas* *Sphingopyxis* *Stigmatella* *Teredinibacter* *Vibrio* *Xanthomonas* *Xylella* *Zymomonas*
	Spirochaetes (1)	*Spirochaeta*			
	Thermogotae (5)	*Fervidobacterium* *Petrotoga*	*Pseudothermotoga* *Thermosipho*	*Thermotoga*	
Archaea (2)	Crenarchaeota (5)	*Acidilobus* *Caldivirga*	*Desulfurococcaceae* *Sulfolobus*	*Thermosphaera*	
	Euryarchaeota (3)	*Halorhabdus*	*Pyrococcus*	*Thermococcus*	

Table 1.1. Microorganisms that possess characterized cellulases across the tree of life (continued)

	Phyla	Genera			
Fungi (5)	Ascomycota (50)	*Acremonium*	*Fusarium*	*Paecilomyces*	*Scopulariopsis*
		Aspergillus	*Fusicoccum*	*Penicillium*	*Septoria*
		Aureobasidium	*Humicola*	*Pestalotiopsis*	*Stachybotrys*
		Bipolaris	*Hypocrea*	*Phaeosphaeria*	*Staphylotrichum*
		Bispora	*Kluyveromyces*	*Phialophora*	*Stilbella*
		Botrytis	*Komagataella*	*Pichia*	*Talaromyces*
		Candida	*Kuraishia*	*Podospora*	*Thermoascus*
		Chaetomium	*Lasiodiplodia*	*Rasamsonia*	*Thermothelomyces*
		Chrysosporium	*Macrophomina*	*Robillarda*	*Thielavia*
		Ciboria	*Ciboria*	*Saccharomyces*	*Trichoderma*
		Claviceps	*Claviceps*	*Saccharomycopsis*	*Wickerhamomyces*
		Clonostachys	*Clonostachys*	*Schizosaccharomyces*	
		Coccidioides	*Coccidioides*	*Sclerotinia*	
	Basidiomycota (22)	*Agaricus*	*Ganoderma*	*Ustilago*	*Trametes*
		Coprinopsis	*Gloeophyllum*	*Volvariella*	*Uromyces*
		Crinipellis	*Hamamotoa*	*Phanerochaete*	*Ustilago*
		Cryptococcus	*Heterobasidion*	*Polyporus*	*Volvariella*
		Dichomitus	*Irpex*	*Postia*	
		Flammulina	*Lentinula*	*Saitozyma*	
		Fomitopsis	*Uromyces*	*Schizophyllum*	
	Chytridiomycota (4)	*Anaeromyces*	*Neocallimastix*	*Orpinomyces*	*Piromyces*
	Mucoromycota (5)	*Mucor*	*Rhizomucor*	*Syncephalastrum*	
		Phycomyces	*Rhizopus*		

of study for cellulolysis, as they degrade the majority of cellulose-containing biomass on earth (Payne *et al.*, 2015). Filamentous, aerobic fungi such as *Aspergillus*, *Penicillium*, and *Trichoderma*, are classical workhorses for cellulase production on an industrial scale due to their ability to secrete high concentrations of cellulases (Sajith *et al.*, 2016).

Compared to fungi, bacteria generally have short generation times and thrive in a wide range of environments and conditions. Because of this, robust cellulolytic bacteria resistant to environmental stresses may be useful for industrial processes (Pourramezan *et al.*, 2012). Cellulases are secreted by free-living bacteria or bacteria in the rumen or gut microbiomes of eukaryotes to digest plant cell walls (Cragg *et al.*, 2015). Bacterial metabolism and physiology can partially delineate groups of cellulolytic bacteria. Fermentative anaerobes, including some representatives from *Clostridium*, *Ruminococcus*, *Butyrivibrio*, and *Fibrobacter*, occur at high

numbers in ruminants, and are involved in converting cellulose to organic acids, ethanol, carbon dioxide, and hydrogen (Güllert *et al.*, 2016; Dehority and Grubb, 1977; Hungate, 1950; Russell *et al.*, 2009). Other fermentative anaerobes include the thermophilic *Caldicellulosiruptor*, found in terrestrial hot springs (Scott *et al.*, 2015) and thermally heated mud flats (Huang *et al.*, 1998), and *Acetivibrio*, isolated from sewage sludge (Khan *et al.*, 1994). Other cellulolytic bacteria are aerobic Gram-positive bacteria from the *Cellulomonas* and *Thermobifida* genera (Lynd *et al.*, 2002). Aerobic bacteria differ from anaerobic cellulolytic bacteria in their cellulose-degrading mechanisms. Aerobic bacteria mainly excrete extracellular endocellulases, exocellulases, and beta-glucosidases that bind to the cellulosic substrate and work cooperatively - sometimes synergistically - to hydrolyze it without cell adherence to cellulose (Lynd *et al.*, 2008). Anaerobic bacteria, in contrast, possess cellulosomes, which are complexes of enzymes located on the cell wall that conduct cellulose degradation (Bayer *et al.*, 2004). However, some anaerobic bacteria can produce both free enzymes and cellulosomes (Berger *et al.*, 2007).

Compared to the diversity present in cellulolytic fungi and bacteria, relatively few archaea have been identified as cellulose degraders. Thermophilic archaea *Pyrococcus horikoshii* and *P. furiosus*, first isolated from a hydrothermal vent (González *et al.*, 1998), express active exogenous endocellulases, for which crystal structures have been resolved (Kim and Ishikawa, 2010b; Kim *et al.*, 2012; Kim and Ishikawa, 2011). Thermophiles *Sulfolobus solfataricus* MT4, *S. acidocaldarius*, and *S. shibatae* each produce high amounts of active beta-glucosidases (Grogan, 1991). Thermophilic archaeal cellulases are advantageous and valuable for industry use, as enzymatic reactions at higher temperatures promote greater cellulose solubility, faster reactions, and lower risk of unwanted bacterial contamination (Kim *et al.*, 2012; Grogan, 1991; Girfoglio *et al.*, 2012).

1.3 Cellulolytic microbial populations in municipal waste sites

Generation of municipal solid waste (MSW) has been steadily increasing in the United States since 1960. In 1960, 88.1 million tonnes of MSW were generated, and by 2015 that amount had almost tripled. In 2015, 262 million tonnes of MSW were generated, of which more than 137 million were landfilled (52.5%). Although recycling accounted for 25.8% of the total generated waste, paper waste still made up the largest fraction of organic material in landfills. Decomposition of organic materials in landfills generate gases such as methane and carbon dioxide, both of which are potent greenhouse gases that exacerbate global warming (Amritha and Anilkumar, 2016; Ontario). Of the waste landfilled, 13.3% or approximately 18 million tonnes was paper and cardboard (U.S. Environmental Protection Agency, 2018).

Although landfills are engineered to limit microbial degradation of stored waste, paper products such as office paper, cardboard, and newspapers in landfills are subjected to degradation by cellulolytic microbes. The past few years have shown an increase in research on microbial diversity in landfills, with a handful of studies also examining microbial metabolisms and functions (McDonald *et al.*, 2012; Staley *et al.*, 2012; Song *et al.*, 2015b, 2015a; Stamps *et al.*, 2016; Ransom-Jones *et al.*, 2017; Wang *et al.*, 2017b; Collins-Fairclough *et al.*, 2018). To date, a comprehensive understanding of the fate of cellulolytic materials in landfills is lacking.

Metagenomic sequencing to mine cellulolytic genes and microbes from landfills are starting steps to discovering novel, efficient cellulases for industrial applications such as biofuel production (Ransom-Jones *et al.*, 2017). The first large-scale sequence-based study on the microbial composition and species richness of municipal landfills used 16S rRNA gene amplicon sequencing on nineteen landfills from sixteen states across the United States of America (Stamps *et al.*, 2016). This study detected nearly 5,000 OTUs, with Proteobacteria and Firmicutes dominating the

microbial communities in nearly all the landfills. The most abundant family detected was Ruminococcaceae, whose members have demonstrated cellulolysis in ruminants (Julliand *et al.*, 1999). Archaea were present at relatively low abundances, and unclassified organisms represented up to 20% of the sampled communities. The microbiome of a leachate sample from one of the landfills also contained members of the candidate division OP9. Putative glycosyl hydrolases (enzymes that hydrolyze the glycosidic bond between a sugar group and a non-sugar group) and an endocellulase were identified in the core genomes of two OP9 members, suggesting members of this phylum may be involved in cellulose degradation (Dodsworth *et al.*, 2013). Stamps and colleagues (2016) provided insight into the microbial compositions of landfills, but a more focused study on cellulose degraders in landfills is needed to understand their diversity and activities in these sites.

In a recent study, a combination of metagenomics and 16S rRNA gene sequencing was applied to leachate microcosms supplemented with cotton cellulose (Ransom-Jones *et al.*, 2017). This work revealed abundant populations of Firmicutes, Bacteroidetes, Spirochaetes, and Fibrobacteres. This was the first study in landfills to report i) a *Fibrobacter* cellulase system, which involves secreting fibro-slime proteins, using pilli to attach to cellulose, and subsequently releasing hydrolytic cellulases, and ii) Bacteroidetes polysaccharide utilization loci (PULs): co-localized gene clusters that encode enzymes and proteins needed for hydrolysis of carbohydrates (Ransom-Jones *et al.*, 2017; Grondin *et al.*, 2017). Gene families containing cellulases and other carbohydrate-modifying enzymes were identified in metagenome-assembled genomes (MAGs) in the cellulose amended leachate microcosms. From identification of these features, a cellulolytic lifestyle was hypothesized for these Fibrobacteres and Bacteroidetes in the landfill (Ransom-Jones *et al.*, 2017).

1.4 Challenges in categorizing cellulases

With rapid sequencing technology improvements and lowered costs, metagenomics has become a standard method for examining microorganisms and their potential functions in an environment. From this, large amounts of data – annotated, but sometimes ambiguously so—are continuously being added to sequence databases. This is a problem for the growing number of genes encoding glycosyl hydrolases, including cellulases, because current methods of categorization of cellulases are not standardized.

Cellulases are non-homologous iso-functional enzymes (Sukharnikov *et al.*, 2012). It has been suggested that all known cellulases exhibit similar protein folds and amino acid sequences among homologs (Sukharnikov *et al.*, 2011). However, not all of these homologs demonstrate biochemical cellulose degradation, further complicating cellulase identification via sequence-based classification methods (Sukharnikov *et al.*, 2011). Categorizing cellulases is additionally challenging as it can be done in one or more of three main ways, based on sequence identity, function, and/or structure. Depending on the type of analysis and its end goal, one method may be more appropriate than others. Unfortunately, this means that there is not one universal convention for grouping cellulolytic enzymes. Due to cellulase structural, sequence, and functional diversities, it is additionally difficult to confidently predict activity from genes annotated as potential cellulases in newly available genomes and metagenomes.

One method for classification of enzymatic cellulases is the glycoside hydrolase (GHs) families, which is a grouping of enzymes that hydrolyze the glycosidic bonds in carbohydrates (Berlemont and Martiny, 2016). “Glycosyl Hydrolase” is one of several classes of enzyme in the Carbohydrate-Active enZyme (CAZy) database, which documents enzymes that anabolize, catabolize, or otherwise modify carbohydrates (Davies and Henrissat, 1995; Cragg *et al.*, 2015;

Lombard *et al.*, 2014). Since 1998, the online CAZy database has organized enzymes into classes, and each class further into families based on protein sequence similarity. Each family contains a minimum of one biochemically characterized enzyme. This system has been widely adopted for classifying carbohydrate-modifying enzymes (Lombard *et al.*, 2014). Each protein family is grouped based on significantly similar amino acid sequences in CAZyme active and catalytic sites, identified using gapped BLAST and HMMER using Hidden Markov Models (HMMs) with a threshold of >85% identity in ungapped alignments (Cantarel *et al.*, 2009). As GH families are based on sequence similarity rather than activity, enzymes within a protein family may act on different substrates, and enzymes that catalyze the same reaction may be found in different GH families. For example, of the 152 GH families in the CAZy database, 16 contain cellulases (Table 1.2) (Sharma and Yazdani, 2016). Beta-1,4-endoglucanase and β-glucosidase activities are in GH5 and GH9, but GH5 also contains mannosidases and chitosanases, and GH9 also contains lichenases and xyloglucanases (Lombard *et al.*, 2014). The variety of cellulose-acting and other carbohydrate substrate specificities present in a protein family with similar protein sequences suggests divergent evolution of the active sites to allow catalysis of different substrates. Conversely, cellulase activities found across a number of protein families suggests convergent evolution allowing the same substrate to be catalyzed by unrelated enzymes (Sharma and Yazdani, 2016). Biochemical characterization and protein modelling are needed to verify or reveal new functions (Cantarel *et al.*, 2009; Aspeborg *et al.*, 2012).

The disadvantage of the GH family classification system is that sequence identity does not necessarily indicate cellulolytic function. Protein structures within the same GH family suggests that structures are more conserved than their sequences. For example in GH7, phylogenetically different enzymes from fungi, protists, isopods, and water fleas share similar structures and

sequences, but enzyme surface properties such as electrostaticity vary, likely due to adaptations to different environments (Cragg *et al.*, 2015). Furthermore, protein structures can delineate cellulases from non-cellulases within a family with more than one activity. The GH48 family contains endocellulases, beta-1,4-glycosidases, and chitinases (Cantarel *et al.*, 2009). Within this family, cellulases can be differentiated from non-cellulases based on conserved amino acids and an omega-loop specific to the surface of GH48 cellulases, which can be used to mine for GH48 cellulases in genomic datasets (Sukharnikov *et al.*, 2012). For larger GH families, i.e., GH5, that contain many enzymes catalyzing different reactions, identifying substrate-specific structures may be complicated (Aspeborg *et al.*, 2012). Subfamilies within certain GH families (Table 1.2, e.g., GH5 and GH30) have been created to further narrow down the active site specificities via sequence identity and to attempt to group enzymes with shared functional properties. The GH5 family has been further classified into 51 subfamilies, restricting cellulases to certain subfamilies (Aspeborg *et al.*, 2012). However, to put subfamilies into practical use, current protein databases will need to annotate genes with subfamilies, which is not common practice.

A second option for categorizing cellulases is the Pfam (Protein family) database (Finn *et al.*, 2016), which assigns Pfam identification to cellulase families based on GH family classifications. Each Pfam is represented by a multiple sequence alignment and HMM covering the known diversity of the members of that Pfam (Finn *et al.*, 2011, 2016). All GH families containing cellulases have Pfam identifiers except three (GH 51, 74, 124). These Pfam identifiers can be used to screen for potential cellulases in genomic datasets (Table 1.2). The Pfam identifications have not kept pace with the exponentially growing genomic and metagenomic databases, which may

Table 1.2. Classification of cellulases by glycoside hydrolase (GH) families, GH subfamlies, Pfam, and Enzyme Commission (EC) numbers. Cellulases (endo-β-1,4-glucanase, exo-β-1,4-glucanase, β-glucosidase) are dominantly classified by the GH system into 16 families according to the CAZy database (Lombard *et al.*, 2014). The Pfam accession numbers classify cellulases based on the GH protein family and their domains (Finn *et al.*, 2016). The Enzyme commission (EC) numbers are assigned to enzymes that are characterized.

GH	Description	GH Subfamily	Pfam Accession	EC
1	β-glucosidase exo-β-1,4-glucanase	-	PF00232	3.2.1.21 3.2.1.74
3	β-glucosidase exo-β-1,4-glucanase	-	PF00933	3.2.1.21 3.2.1.74
5	endo-β-1,4-glucanase β-glucosidase exo-β-1,4-glucanase cellulose β-1,4-cellobiosidase	1, 2, 4, 5, 8, 25, 26, 37, 38 12 37, 52, 53 1,2	PF00150	3.2.1.4 3.2.1.21 3.2.1.74 3.2.1.91
6	endo-β-1,4-glucanase cellulose β-1,4-cellobiosidase	-	PF14871*	3.2.1.4 3.2.1.91
7	endo-β-1,4-glucanase reducing end-acting cellobiohydrolase	-	PF00840	3.2.1.4 3.2.1.176
8	endo-β-1,4-glucanase	-	PF01270	3.2.1.4
9	endo-β-1,4-glucanase β-glucosidase exo-β-1,4-glucanase cellulose β-1,4-cellobiosidase reducing end-acting cellobiohydrolase	-	PF00759	3.2.1.4 3.2.1.21 3.2.1.74 3.2.1.91 3.2.1.176^
12	endo-β-1,4-glucanase	-	PF01670	3.2.1.4
30	β-glucosidase	1	PF02055 (TIM-barrel domain) PF14587 (O-glycosyl) PF17189 (beta-sandwich domain)	3.2.1.21
44	endo-β-1,4-glucanase	-	PF12891	3.2.1.4
45	endo-β-1,4-glucanase	-	PF02015	3.2.1.4
48	endo-β-1,4-glucanase reducing end-acting cellobiohydrolase	-	PF14587	3.2.1.4 3.2.1.176
51	endo-β-1,4-glucanase	-	-	3.2.1.4
74	endo-β-1,4-glucanase	-	-	3.2.1.4
116	β-glucosidase	-	PF04685 (catalytic region) PF12215 (N-terminal)	3.2.1.21
124	endo-β-1,4-glucanase	-	-	3.2.1.4

– = non-existent.

* = family of hypothetical glycoside hydrolases.

^ = activity not listed as seen in the GH family by CAZy but is found under the characterized protein section of the GH family in the database

require re-curating HMM seeds to better depict the current sequence diversity for GH families.

A third method of classification for cellulases is the Enzyme Commission (EC) numbering system, which classifies biochemically characterized enzymes based on the type of chemical reaction they catalyze. Each four-number identifier groups enzymes into classes and subclasses. Enzymes are categorized based purely on function regardless of evolutionary relationship (McDonald *et al.*, 2009) which has both advantages and disadvantages. This classification system is useful to identify cellulases with higher specificity of function, distinguishing endo-β-1,4-glucanases (E.C. 3.2.1.4), β-glucosidases (E.C. 3.2.1.21), exo-β-1,4-glucanases (E.C. 3.2.1.74), cellulose (non-reducing end-acting) β-1,4-cellobiosidases (E.C. 3.2.1.91), and reducing end-acting cellobiohydrolases (E.C. 3.1.2.176). This classification system also prevents redundancy in categorizing an enzyme with the same catalytic function from different species (McDonald *et al.*, 2009). However, E.C. numbers are only assigned to biochemically characterized enzymes, which is a much smaller pool in databases compared to sequences with predicted functions.

The disparity between these three classification methods and existing vague annotations in databases make consistently classifying cellulases from genomic datasets difficult (Sukharnikov *et al.*, 2012).

1.5 Landfill design and operation

Landfills are designed to prevent or reduce contamination of the environment by municipal solid wastes and leachate (Environmental Protection Agency, 2000). There is engineered infrastructure for monitoring and controlling leachate generation and gas production, to protect the surrounding soil and water, and to minimize nuisances such as pests, odours, and fires. Landfill fires can occur due to ignition of combustible material or via decomposition of organic waste, such as paper products and food, which generates heat, carbon dioxide, and methane. If the methane and heat are

not dissipated. but instead are retained and increase in concentration, chemical oxidation of the waste will continue until combustion (National Environment and Planning Agency, 2010). In some modern landfills, gas capture systems are in place to trap gas produced at landfills for biogas conversion, while in less instrumented landfills, methane vents allow release of methane to prevent combustion. Leachate collection systems prevent leachate movement, which can mobilize unwanted and/or toxic compounds from refuse into the environment.

1.6 Scope of research and research objectives

The heterogeneity of substrates and complex environmental conditions in municipal waste sites make these environments of high interest for microbial research with a focus on industrial applications, bioremediation, and biotechnology. With metagenomic sequencing technology becoming more accessible, we can investigate at greater depth than previous 16S rRNA gene amplicon-based analyses (Song *et al.*, 2015a, 2015b; Stamps *et al.*, 2016; Ransom-Jones *et al.*, 2017). Through metagenomics, landfill microbial communities, key populations within these communities, and the functional potential that exists in these environments can be determined. The functional potential predicted from landfills can be corroborated through enrichment culturing and assaying enzymatic activities *in vitro*. As paper waste is the most abundant type of organic waste reported in landfills (U.S. Environmental Protection Agency, 2018), investigating paper degradation by cellulolytic microbial populations is valuable for improving downstream waste management decisions. Discovery of currently unknown microbial diversity may lead to improvements in the biofuel industry for the conversion of cellulose to cellulosic biofuel.

The microbial communities of landfills, and particularly their cellulolytic potential, have not been well characterized. For my research, microbial diversity and cellulolytic potential were analyzed in two municipal waste sites, one in Jamaica and one in Canada, with different levels of engineered

infrastructure. The first site is the Riverton City Dump in Kingston, Jamaica and the adjacent Duhaney River. The Riverton dump, and in fact all Jamaican municipal waste sites, are not sanitary landfills (National Environment and Planning Agency, 2010). Sanitary landfills have infrastructure that physically isolate waste from the environment to prevent contamination, along with other engineered mechanisms that may include daily covers, methane capture systems, compacting, and waste organization (World Health Organization, 1999). The Riverton dump has been reported as a human health concern, with additional concerns raised that its leachate may be contaminating the Duhaney River (Collins-Fairclough *et al.*, 2018). The Canadian municipal waste site is a landfill located in southern Ontario and its adjacent groundwater aquifer. This site is a sanitary landfill, with waste sorting, compacting, linings, daily covers, leachate capture systems, and a methane capture system.

There are three main objectives to my research. The first objective of my research was to investigate the microbial diversity and cellulolytic populations and their potential from the Riverton City dump and the Duhaney River (Chapter 2). The second objective was to examine the diversity of cellulolytic enzymes from both the Riverton City dump (Chapter 2) and the southern Ontario landfill (Chapter 3). My third objective was to confirm the presence of the cellulolytic activity in the leachate at the southern Ontario landfill through culture-based approaches (Chapter 4).

Chapter 2: Microbial Diversity and Cellulolytic Diversity at the Riverton Dump

2.1 Introduction

Municipal waste sites house highly heterogeneous waste and thus are complex environments. Paper is the most discarded organic waste in landfills and is likely subjected to degradation by resident microorganisms (U.S. Environmental Protection Agency, 2018). Cellulose, the base constituent of paper, is the most abundant organic compound on earth and has been leveraged as a form of sustainable energy via cellulosic biofuels (Balan, 2014). Investigating degradation of paper waste in landfills can shed light on potentially novel cellulolytic microorganisms, whose genes may be used to improve the robustness and efficiency of biotechnology and industrial processes.

Studies investigating the microbial compositions in landfills, predominantly using 16S rRNA gene amplicon sequencing, have become more frequent recently (Song *et al.*, 2015b, 2015a; Stamps *et al.*, 2016; Remmas *et al.*, 2017b; Ransom-Jones *et al.*, 2017). However, there has only been one study to date investigating the microbial composition and potential function in microcosms from cellulose-amended landfill leachate (Ransom-Jones *et al.*, 2017). Through 16S rRNA gene and metagenomic analyses, Ransom-Jones et al. detected an increase in enrichment of microorganisms belonging to Firmicutes, Bacteroidetes, Fibrobacteres, and Spirochaetes in cellulose-enriched leachate microcosms when compared to raw leachate. Four metagenome-assembled genomes (MAGs) associated with each of these phyla and a proteobacterial MAG possessed carbohydrate active enzymes (CAZymes) in their genomes.

The Riverton City dump in Jamaica and the Duhaney River were sites of interest (Figure 2.1). Most of the waste generated by weight within the Riverton wasteshed (the area including the four parishes that the Riverton City dump services) is compostable. In 2013, ~228 kg of paper was generated over a 3.5-day period, which accounted for ~9% of the waste by weight

Figure 2.1. Sampling sites in Jamaica. (A) Map of Jamaica. Grey-shaded area indicates the parishes the Riverton City dump (red star) services. (B) Physical map showing the leachate pond sampling site (red star), the Duhaney River sampling site (blue star), and its surrounding vegetation. (C) Satellite-viewed map (courtesy of Google Maps) showing the sampling sites and its distance. Red star = location of leachate pond sampling in the Riverton City dump, blue star = location of river sampling, white arrow indicates direction of river flow. Figure courtesy of Aneisha Collins-Fairclough (Collins-Fairclough *et al.*, 2018).

(National Solid Waste Management Authority, 2013). There is little to no sorting of waste in Jamaica; thus, there is often uncontrolled combustion at the dumpsites and the dumps pose human health risks (Planning Institute of Jamaica, 2007). There is also concern of contamination of the Duhaney river from leachate run-off from the dump (Collins-Fairclough *et al.*, 2018). Investigating microbial communities and their cellulose potential in the Riverton waste site can potentially reveal novel cellulolytic microbes that are able to withstand the heterogeneous and changing environmental conditions of landfills and therefore may be robust for use in industrial processes.

The objective of this chapter was to investigate the microbial composition of the leachate from the Riverton City dump and in the Duhaney River, and determine whether the detected microorganisms and the genes in the community have cellulolytic potential.

2.2 Materials and Methods

2.2.1 Sample collection

Collaborators from the University of Technology Jamaica sampled two locations in Jamaica—a leachate pond that passively develops within the Riverton City dump and a site in the Duhaney River adjacent to the dump, approximately 600 m from the leachate pond (Figure 2.1). There is no obvious fluid flow between the two sampling sites. The two locations were chosen as representative sites for the two environments. Leachate samples were collected from the periphery of a perennial leachate pond at the Riverton City dump (18.01052 N, 76.85667 W) in Jamaica. An autoclaved disposable jug was first rinsed with a surface sample from the leachate pond. The jug was used to scoop leachate from the top 1 m layer of one edge of the pond into two autoclaved 2.5-L conical flasks. Water samples were collected from the Duhaney River in Jamaica, which passes through the Riverton City dump (18.012292 N, 76.850922 W). Autoclaved conical flasks were rinsed with surface water from the river. The flasks were then dipped in the river and filled to two-thirds capacity with surface water from the periphery of the river. Flasks were sealed with Parafilm M and capped with aluminum foil prior to transportation to the laboratory. All samples were transported directly to the laboratory and stored at 4°C, and DNA was extracted within four days of sample collection.

2.2.2 DNA extraction and sequencing

Our collaborators prepared the leachate and river samples for sequencing. Leachate samples (approximately 100 ml) were filtered using six 0.22-µm polyethersulfone membranes (diameter, 47 mm; Sterilitech), switching filters upon clogging. DNA was extracted from the membranes using the MoBio PowerWater DNA isolation kit according to the manufacturer's instructions. In parallel, the MoBio PowerSoil DNA isolation kit was used to extract DNA from 18 ml of leachate

as follows. First, 1.0-ml aliquots of leachate samples were centrifuged at 14,000 rpm for 30 min. Supernatants from the tubes were transferred to new tubes. The supernatant and pellet from each tube were processed separately for DNA extraction using the MoBio PowerSoil DNA isolation kit, according to the manufacturer's instructions, with the supernatant and pellet as input in place of soil. The resulting eluates from all leachate DNA extractions (supernatants and pellets) were pooled and purified using a phenol-chloroform extraction.

The Duhaney River water samples were filtered through a total of nine 0.22-µm polyethersulfone membranes using a vacuum pump attached to a Buchner funnel. The filtrate was further filtered through seven 0.03-µm polyethersulfone filters to trap small cell sizes. Approximately 10 L of river water was filtered in total. Filters were used for DNA extraction using the MoBio PowerWater DNA extraction kit. DNA eluates from all river extractions were pooled for sequencing.

Shotgun metagenomic sequencing was conducted by the McMaster University Farncombe Metagenomics Facility in Canada. Prior to sequencing, the pooled leachate DNA was further purified using AMPure beads. The NEBNext Ultra DNA kit was used for library preparation from 100 ng from each DNA sample. The leachate library and the river library were pooled, and a sequencing library was prepared using the Illumina MiSeq 250-bp paired-end read v2 kit.

2.2.3 Metagenomic pipeline

Metagenomes were assembled and annotated as described previously (Daly *et al.*, 2016). Briefly, reads from both metagenomes were quality trimmed with Sickle (Joshi and Fass, 2011). Paired-end reads were assembled using IDBA-UD (Peng *et al.*, 2012) under default parameters, with each metagenome assembled separately. Open reading frames (ORFs) were predicted using

MetaProdigal (Hyatt *et al.*, 2012), and annotated via USEARCH (Edgar, 2010) against KEGG (Kanehisa *et al.*, 2012), UniRef90 (Suzek *et al.*, 2007), and InterproScan (Jones *et al.*, 2014) databases. Annotations were ranked A to E and reported as follows: (A) reciprocal best hits with bit score of >350, then (B) reciprocal best hit to UniRef with a bit score of >350, (C) best hit to KEGG with a bit score of >60 or best hit to UniRef90 with a bit score of >60. Proteins with InterproScan matches but no other hits were ranked as (D), and hypothetical proteins that were predicted open reading frames but no further annotations were ranked (E). Reads from both leachate and river were then mapped to both leachate and river assemblies using Bowtie 2 v. 2.2.6 (Langmead and Salzberg, 2012) to determine contig coverage statistics and enable abundance-based binning metrics.

Anvi'o v. 2.0.2 (Eren *et al.*, 2015) was used to bin the scaffolds, manually refine the bins, and visualize the data. First, a contig database was created from the respective metagenome's contig file, and open reading frames were identified through Prodigal v. 2.6.2. These ORFs were used solely for assessing bin completion, while MetaProdigal annotations described above were used for metabolic reconstructions and phylogenetic inferences. Genes corresponding to single-copy core gene bacterial (Alneberg *et al.*, 2014; Campbell *et al.*, 2013; Creevey *et al.*, 2011; Dupont *et al.*, 2012) and archaeal (Rinke *et al.*, 2013) gene collections were identified using HMMER v. 3.1b2 (Finn *et al.*, 2011). Coverage information for each contig was determined via samtools (Li *et al.*, 2009a). Contigs were binned using CONCOCT (Alneberg *et al.*, 2014), leveraging nucleotide frequency information as well as differential coverage. All programs were used under default parameters as implemented by Anvi'o. The bins were manually refined based on completion and redundancy statistics in the Anvi'o interactive interface. A bin was considered a high-quality MAG if it was greater than 70% complete and had less than 10% contamination.

Organism taxonomic placement was inferred from a phylogenetic tree built from concatenated protein alignments of 15 conserved, single-copy ribosomal proteins (RpL2, -3, -4, -5, -6, -14, -16, -18, -22, and -24, and RpS3, -8, -10, -17, and -19). The MAGs containing more than 50% of these marker genes were included in the phylogenetic inference alongside a reference set comprising one member of each genus for which sequenced genomes are available (from Hug *et al.*, 2016). Each protein data set was aligned individually using MUSCLE v 3.8.31 (Edgar, 2004), and then the 15 alignments were concatenated. Alignments were edited using Geneious v. 10.0.5 (Kearse *et al.*, 2012). Alignment positions with greater than 95% gaps were removed, and C- and N-termini with non-conserved regions were trimmed. Taxa with information for less than 50% of the trimmed concatenated alignment were removed. The final concatenated alignment contained 2,786 sequences and 2,470 amino acid positions. A maximum likelihood tree was constructed using RAxML-HPC v. 8.2.10 (Stamatakis, 2014) on the public web server CIPRES Science Gateway v. 3.3 (Miller *et al.*, 2011) using the LG+gamma protein substitution matrix and with automatic bootstopping to determine the optimal number of bootstrap replicates. The phylogenetic tree was visualized in FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). All software programs used were operated under default parameters unless otherwise stated.

2.2.4 16S rRNA gene community profiles

We used the SILVA database core alignment to search for 16S rRNA genes within our data sets. The SILVA alignment contained 592,605 bacterial and 25,026 archaeal 16S rRNA genes. Reads with best hits to eukaryotes were removed from analyses from this point onwards. A hidden Markov model (HMM) was built using HMMER 3.1b2 for the nonredundant small-subunit reference data set (nonredundant at 99% identity) from the SILVA 132 release (Quast *et al.*, 2012).

The 16S rRNA gene HMM was searched against the trimmed leachate and river reads using the per-target output, with an E-value of $1e^{-5}$. Reads identified from this search were BLASTn searched against the RefSeq RNA database (release 87) (Altschul *et al.*, 1990) and the NCBI taxonomy database (November 2017). Top hits with a minimum E-value of $1e^{-40}$ were used to identify taxonomy at the phylum level for the 16S rRNA gene-containing reads. For the assembled data, the same 16S rRNA gene HMM was searched against the leachate and river assembled scaffolds using the same pipeline and parameters as for the read search. All plots were created with ggplot2 v. 3.0.0 (Wickham, 2016) in RStudio v. 1.0.136 (RStudio Team, 2015).

2.2.5 Data availability

The Riverton City dump leachate data are available under BioProject PRJNA475763 and biosample SAMN09401598. The Duhaney River data are available under BioProject PRJNA475764 and biosample SAMN09401599. The Leachate and the River reads are available in the Sequence Read Archive under accession no. SRR7299214 and no. SRR7346984, respectively.

2.2.6 Abundance of potential cellulase genes in the metagenomes

The leachate and river metagenomes (contigs >5,000 bases) were annotated using the dbCAN CAZyme database (released on July 31, 2018) on a local linux computer. The following steps were followed according to the readme.txt available on the dbCAN website. In summary, the dbCAN HMM database release 7.0 was downloaded and formatted using hmmpress. The function hmmscan from HMMER v. 3.1b2 was used to search the formatted HMMs against the leachate and river metagenomes. The resulting data were parsed using hmmscan-parser.sh available on the

dbCAN website, with the following parameters as suggested by dbCAN: if alignment >80 aa, use E-value threshold $1e^{-5}$ and coverage threshold 0.3. The resulting data was further screened for hits that had E-value >$1e^{-15}$ and coverage >0.35. These parameters would give the same results as the dbCAN web interface, which is more stringent than those of the hmmscan-parser.sh.
All hits to GH families containing cellulases were then annotated as described below.

2.2.7 Potential cellulase community profile

The dbCAN hits from the leachate and river contigs >5,000 bases were annotated taxonomically and functionally using blastp against the RefSeq protein database. Genes were considered potential cellulases if they were i) annotated as one of the three types of cellulases, i.e., endocellulase, exocellulase, or beta-glucosidase, including names that are synonymous with these, ii) annotated as "cellulase", and iii) ambiguously annotated, i.e., "hypothetical protein" or "glycosyl hydrolase family n" where n is the GH family number. Hits whose protein annotations do not fall into one of the listed categories were not considered as potential cellulases. All plots were created with ggplot2 v. 3.0.0 (Wickham, 2016) in RStudio v. 1.0.136 (RStudio Team, 2015).

2.3 Results and Discussion

2.3.1 Metagenome and metagenome-assembled genome statistics

Total community shotgun sequencing from both the leachate DNA and the river DNA yielded metagenomes of 4.2 and 3.7 Gb, respectively. Metagenomes were assembled, and scaffolds longer than 2,500 bases were used for binning to reconstruct metagenome-assembled genomes (MAGs) for the high-abundance populations. The two metagenomes had similar total read numbers and

assembly sizes, indicating that the communities were sampled to approximately equivalent depths (Table 2.1). Using Anvi'o, assembled scaffolds were assigned to MAGs, leveraging differential abundance information combined with tetranucleotide frequencies by CONCOCT. Differential abundance-based binning was of limited use, as fewer than 0.5% of reads mapped to the non-source assembly in each case. MAGs were manually refined prioritizing completion and redundancy statistics. After refinement, 13 of 55 leachate MAGs and 3 of 33 river MAGs were high-quality (>70% complete and <10% redundant) (Table 2.2).

Table 2.1. Statistics for Riverton City dump leachate and Duhaney river metagenomes. Scaffolds >2,500 bp were used for binning.

Sample	# Reads	Read length	# Scaffolds total; >2500 bp	Scaffold N50 total; >2500 bp	Max scaffold length (bp)	% Reads assembled total; >2500 bp
Leachate	16,673,648	250	555,592; 5,391	12,753; 1,011	532,373	42.7; 16.4
River	14,615,770	250	455,023; 3,348	16,750; 882	511,705	65.2; 24.0

2.3.2 Microbial community composition of the Riverton City Dump leachate and the Duhaney River

Microbial populations in the leachate and river metagenomes were identified via 16S rRNA genes and/or a set of 15 conserved, single-copy, co-located ribosomal proteins (Table 2.2; Figure. 2.2). The community composition of the reads, assembled scaffolds, and MAGs were compared using predicted 16S rRNA genes to assess whether the binned populations were representative of the total microbial diversity sampled from the sites (Figure 2.3).

Reads containing 16S rRNA gene fragments were identified using a hidden Markov model (HMM) search using an HMM built from the SILVA database 16S rRNA genes. A total of 8,762 and 7,731

reads from the leachate and river data sets contained predicted 16S or 18S rRNA gene sequences, of which 7,423 and 6,978 could be taxonomically placed into bacterial or archaeal phyla via BLASTn against the NCBI RefSeq RNA database (Figure 2.3).

From the reads, the leachate pond contained three most abundant phyla: Proteobacteria (34.6%, with 16.7% Gammaproteobacteria and 12.2% Deltaproteobacteria, with other classes of Proteobacteria occurring at less than 2%), Firmicutes (22.9%), and Bacteroidetes (20.3%) (Figure 2.3). Low-abundance phyla, with relative abundances between 1 and 10%, include the Tenericutes, Spirochaetes, Actinobacteria, and Chloroflexi phyla. There were 30 rare phyla occurring at less than 1% abundance in the leachate community. The river sample was dominated by Proteobacteria, with 69.7% of 16S rRNA gene-containing reads classified to that phylum (25.4% Alphaproteobacteria, 20.6% Betaproteobacteria, 20.6% Gammaproteobacteria, 2.6% Epsilonproteobacteria, and less than 1% Deltaproteobacteria) (Figure 2.3). The second most abundant river phylum was Bacteroidetes at 28.0%, with the remaining 22 detected phyla occurring at less than 1% of the total community.

From assembled scaffolds, a total of 412 scaffolds from the leachate metagenome and 372 scaffolds from the river metagenome contained predicted 16S rRNA gene sequences (Figure 2.3). The microbial composition in the assembled data sets showed that the leachate assembly was dominated by the same major groups as in the reads: 32.3% Firmicutes, 15.1% Bacteroidetes, and Proteobacteria, with 9.7% Gammaproteobacteria and 9.4% Deltaproteobacteria. Tenericutes were also present at 9.7%. In the river assembly, Proteobacteria again dominated (82.2%), with 30.7% Gammaproteobacteria, 29.1% Alphaproteobacteria, and 21.4% Betaproteobacteria, with Epsilonproteobacteria and Deltaproteobacteria at less than 1%. Phyla above 1% abundance in the river also included the Bacteroidetes (14.6%) and the Firmicutes (1.3%). Many of the identified

16S rRNA genes were present on short scaffolds (~1 kb), which means that they were not included in the binning process. Of the leachate and river MAGs identified via the concatenated ribosomal protein tree, all 13 high-quality leachate MAGs and 2 of 3 high-quality river MAGs were included on the tree (Table 2.2, Figure 2.2, and Figure 2.4).

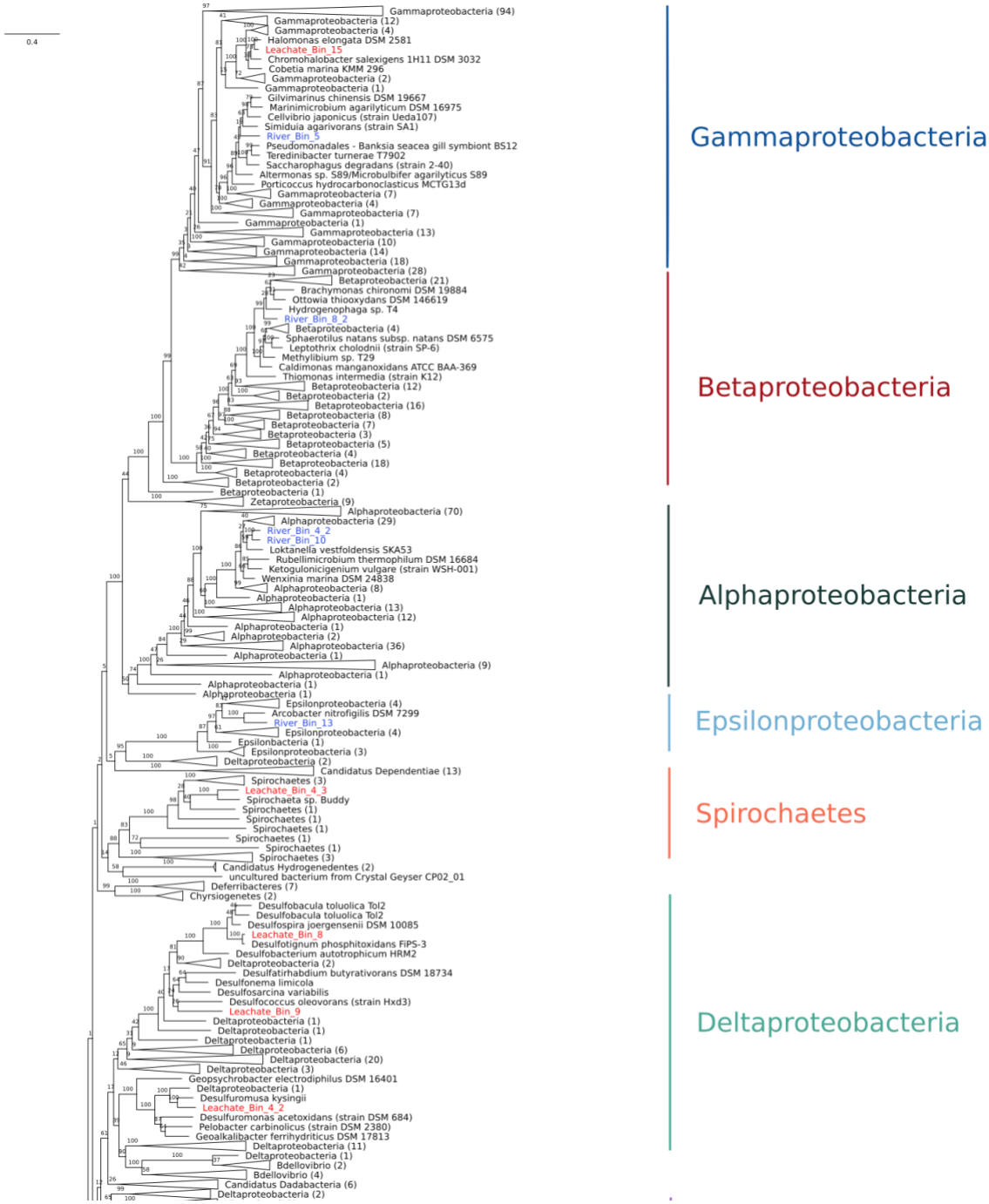
From the taxonomically assigned MAGs, the dominant phylum in the leachate community was Bacteroidetes, with 35.4% relative abundance across five MAGs. The next most abundant phyla were Proteobacteria (4 MAGs, 25.7%, where Deltaproteobacteria contributed 23.4%) and Firmicutes (4 MAGs, 16.4%). Other phyla within the leachate community were represented by one MAG each, including Tenericutes, Spirochaetes, Chloroflexi, and a member of the candidate phylum CPR2. In contrast, the five MAGs from the river metagenome were all affiliated with the Proteobacteria, with Alphaproteobacteria dominating (68.7%) over Betaproteobacteria (8.5%), Gammaproteobacteria (9.6%), and Epsilonproteobacteria (13.3%). Across the reads, assemblies, and MAGs, archaea were a minor proportion of the communities, with their highest abundance at 1.4% in the leachate assembly. Members of the Crenarchaeota, Euryarchaeota, and Thaumarchaeota were present at low abundance in both metagenomes.

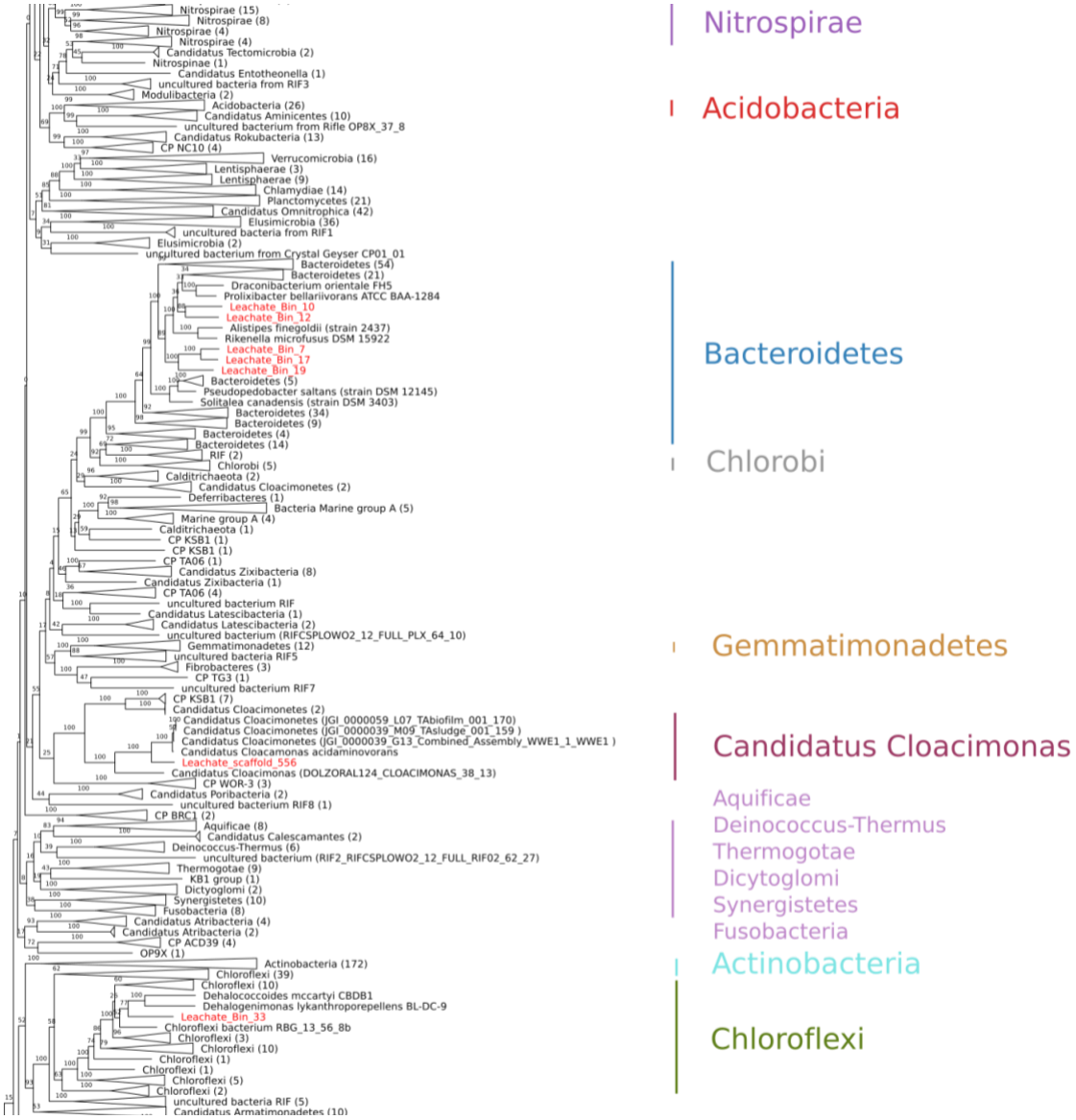
2.3.3 Key microbial populations

The MAGs revealed microorganisms from seven phyla in the leachate and one from the river. Most of these phyla have been reported in previous landfill studies. Based on ribosomal protein marker gene abundances and the 16S rRNA gene analysis, Bacteroidetes and Firmicutes were the two most abundant phyla in the leachate, with 35.4% and 16.4% relative abundance in the reads, respectively, and included the most abundant MAGs (coverage of ~140x and 65x, respectively). Bacteroidetes and Firmicutes have frequently been detected in landfills irrespective of landfill age

Table 2.2. Statistics on Riverton metagenome-assembled genomes that contained the 15 ribosomal proteins (RpL2, -3, -4, -5, -6, -14, -16, -18, -22, -24, and RpS3, -8, -10, -17, -19) in their scaffolds. Scaffolds >2,500 bp were used for binning. Good quality MAGs had completion >70% and redundancy <10%. LB=Leachate Bin RB = River Bin

No.	Bin	Phylum	Closest Relative/Clade	Total length (mbps)	Number of Contigs	GC Content	Abundance/ Coverage	Completion (%)	Redundancy (%)
Leachate									
1	LB_9	Proteobacteria	*Desulfococcus oleovorans* Hxd3	3.46	198	51.68	7.41	96.44	6.73
2	LB_19	Bacteroidetes	LB_7, LB_17	3.26	50	52.94	10.33	95.63	3.93
3	LB_32	Tenericutes	LB_18_1	1.18	52	28.68	16.46	94.99	3.55
4	LB_22	Firmicutes	*Tepidimicrobium xylanilyticum* DSM 23310	1.58	137	43.32	8.12	94.23	6.17
5	LB_18_1	Tenericutes	LB_32	1.10	75	33.44	26.17	92.06	3.65
6	LB_10	Bacteroidetes	LB_12	2.31	204	32.84	7.99	91.32	3.02
7	LB_7	Bacteroidetes	LB_19, LB_17	2.48	223	42.19	99.51	88.51	5.01
8	LB_27_3	candidate division CPR2		0.74	8	37.52	22.23	88.24	2.22
9	LB_8	Proteobacteria	*Desulfotignum phosphitoxidans* FiPS 3	2.56	207	53.33	7.89	85.26	4.16
10	LB_12	Bacteroidetes	LB_10	2.23	182	43.54	12.44	79.84	2.95
11	LB_26	Firmicutes	Firmicutes, LB_26	0.98	102	40.31	8.23	79.11	7.77
12	LB_16	Firmicutes	Clostridiaceae	1.85	152	46.97	8.93	71.89	1.88
13	LB_4_2	Proteobacteria	*Desulfuromusa kysingii* DSM 7343	1.13	115	57.58	77.38	70.69	0.80
14	LB_17	Bacteroidetes	LB_7, LB_19	1.25	157	41.44	9.50	66.53	2.72
15	LB_33	Chloroflexi	Dehalococcoidetes	0.42	59	43.99	6.01	50.24	0.63
16	LB_18_3	Firmicutes	*Erysipelothrix rhsiopathiae* str. Fujisawa	0.57	55	33.92	39.46	49.70	0.92
17	LB_15	Proteobacteria	*Halomonas elongata* DSM 2581	1.23	158	57.53	8.87	41.23	2.90
18	LB_4_3	Spirochaetes	*Spirochaeta* sp. Buddy	1.58	154	51.95	18.01	26.45	0.18
River									
20	RB_10	Proteobacteria	RB_4_2	3.53	14	60.94	22.36	86.40	5.41
21	RB_5	Proteobacteria	Gammaproteobacteria	4.27	177	42.11	9.34	85.41	3.64
22	RB_8_2	Proteobacteria	Comamonadaceae	0.82	99	53.31	8.39	68.56	3.24
23	RB_13	Proteobacteria	*Acrobacter nitrofigilis* DSM 7299	1.37	130	37.39	13.12	45.29	8.79
24	RB_4_2	Proteobacteria	RB_10	1.36	159	59.63	45.46	38.98	1.99


0.4
Gammaproteobacteria (94)
Gammaproteobacteria (12)
Gammaproteobacteria (4)
Halomonas elongata DSM 2581
Leachate_Bin_15
Chromohalobacter salexigens 1H11 DSM 3032
Cobetia marina KMM 296
Gammaproteobacteria (2)
Gammaproteobacteria (1)
Gilvimarinus chinensis DSM 19667
Marinimicrobium agarilyticum DSM 16975
Cellvibrio japonicus (strain Ueda107)
Simiduia agarivorans (strain SA1)
River_Bin_5
Pseudomonadales - Banksia seacea gill symbiont BS12
Teredinibacter turnerae T7902
Saccharophagus degradans (strain 2-40)
Altermonas sp. S89/Microbulbifer agarilyticus S89
Porticoccus hydrocarbonoclasticus MCTG13d
Gammaproteobacteria (7)
Gammaproteobacteria (4)
Gammaproteobacteria (7)
Gammaproteobacteria (1)
Gammaproteobacteria (13)
Gammaproteobacteria (10)
Gammaproteobacteria (14)
Gammaproteobacteria (18)
Gammaproteobacteria (28)
Betaproteobacteria (21)
Brachymonas chironomi DSM 19884
Ottowia thiooxydans DSM 146619
Hydrogenophaga sp. T4
River_Bin_8_2
Betaproteobacteria (4)
Sphaerotilus natans subsp. natans DSM 6575
Leptothrix cholodnii (strain SP-6)
Methylibium sp. T29
Caldimonas manganoxidans ATCC BAA-369
Thiomonas intermedia (strain K12)
Betaproteobacteria (12)
Betaproteobacteria (2)
Betaproteobacteria (16)
Betaproteobacteria (8)
Betaproteobacteria (7)
Betaproteobacteria (3)
Betaproteobacteria (5)
Betaproteobacteria (4)
Betaproteobacteria (18)
Betaproteobacteria (4)
Betaproteobacteria (2)
Betaproteobacteria (1)
Zetaproteobacteria (9)
Alphaproteobacteria (70)
Alphaproteobacteria (29)
River_Bin_4_2
River_Bin_10
Loktanella vestfoldensis SKA53
Rubellimicrobium thermophilum DSM 16684
Ketogulonicigenium vulgare (strain WSH-001)
Wenxinia marina DSM 24838
Alphaproteobacteria (8)
Alphaproteobacteria (1)
Alphaproteobacteria (13)
Alphaproteobacteria (12)
Alphaproteobacteria (1)
Alphaproteobacteria (2)
Alphaproteobacteria (36)
Alphaproteobacteria (1)
Alphaproteobacteria (9)
Alphaproteobacteria (1)
Alphaproteobacteria (1)
Alphaproteobacteria (1)
Epsilonproteobacteria (4)
Arcobacter nitrofigilis DSM 7299
River_Bin_13
Epsilonproteobacteria (4)
Epsilonbacteria (1)
Epsilonproteobacteria (3)
Deltaproteobacteria (2)
Candidatus Dependentiae (13)
Spirochaetes (3)
Leachate_Bin_4_3
Spirochaeta sp. Buddy
Spirochaetes (1)
Spirochaetes (1)
Spirochaetes (1)
Spirochaetes (1)
Spirochaetes (1)
Spirochaetes (3)
Spirochaetes (3)
Candidatus Hydrogenedentes (2)
uncultured bacterium from Crystal Geyser CP02_01
Deferribacteres (7)
Chyrsiogenetes (2)
Desulfobacula toluolica Tol2
Desulfobacula toluolica Tol2
Desulfospira joergensenii DSM 10085
Leachate_Bin_8
Desulfotignum phosphitoxidans FiPS-3
Desulfobacterium autotrophicum HRM2
Deltaproteobacteria (2)
Desulfatirhabdium butyrativorans DSM 18734
Desulfonema limicola
Desulfosarcina variabilis
Desulfococcus oleovorans (strain Hxd3)
Leachate_Bin_9
Deltaproteobacteria (1)
Deltaproteobacteria (1)
Deltaproteobacteria (1)
Deltaproteobacteria (6)
Deltaproteobacteria (20)
Deltaproteobacteria (3)
Geopsychrobacter electrodiphilus DSM 16401
Deltaproteobacteria (1)
Desulfuromusa kysingii
Leachate_Bin_4_2
Desulfuromonas acetoxidans (strain DSM 684)
Pelobacter carbinolicus (strain DSM 2380)
Geoalkalibacter ferrihydriticus DSM 17813
Deltaproteobacteria (11)
Deltaproteobacteria (1)
Bdellovibrio (2)
Bdellovibrio (4)
Candidatus Dadabacteria (6)
Deltaproteobacteria (2)
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Epsilonproteobacteria
Spirochaetes
Deltaproteobacteria


Nitrospirae (15)
Nitrospirae (8)
Nitrospirae (4)
Nitrospirae (4)
Candidatus Tectomicrobia (2)
Nitrospinae (1)
Candidatus Entotheonella (1)
uncultured bacteria from RIF3
Modulibacteria (2)
Acidobacteria (26)
Candidatus Aminicentes (10)
uncultured bacterium from Rifle OP8X_37_8
Candidatus Rokubacteria (13)
CP NC10 (4)
Verrucomicrobia (16)
Lentisphaerae (3)
Lentisphaerae (9)
Chlamydiae (14)
Planctomycetes (21)
Candidatus Omnitrophica (42)
Elusimicrobia (36)
uncultured bacteria from RIF1
Elusimicrobia (2)
uncultured bacterium from Crystal Geyser CP01_01
Bacteroidetes (54)
Bacteroidetes (21)
Draconibacterium orientale FH5
Prolixibacter bellariivorans ATCC BAA-1284
Leachate_Bin_10
Leachate_Bin_12
Alistipes finegoldii (strain 2437)
Rikenella microfusus DSM 15922
Leachate_Bin_7
Leachate_Bin_17
Leachate_Bin_19
Bacteroidetes (5)
Pseudopedobacter saltans (strain DSM 12145)
Solitalea canadensis (strain DSM 3403)
Bacteroidetes (34)
Bacteroidetes (9)
Bacteroidetes (4)
Bacteroidetes (14)
RIF (2)
Chlorobi (5)
Calditrichaeota (2)
Candidatus Cloacimonetes (2)
Deferribacteres (1)
Bacteria Marine group A (5)
Marine group A (4)
Calditrichaeota (1)
CP KSB1 (1)
CP KSB1 (1)
CP TA06 (1)
Candidatus Zixibacteria (8)
Candidatus Zixibacteria (1)
CP TA06 (4)
uncultured bacterium RIF
Candidatus Latescibacteria (1)
Candidatus Latescibacteria (2)
uncultured bacterium (RIFCSPLOWO2_12_FULL_PLX_64_10)
Gemmatimonadetes (12)
uncultured bacteria RIF5
Fibrobacteres (3)
CP TG3 (1)
uncultured bacterium RIF7
CP KSB1 (7)
Candidatus Cloacimonetes (2)
Candidatus Cloacimonetes (JGI_0000059_L07_TAbiofilm_001_170)
Candidatus Cloacimonetes (JGI_0000039_M09_TAsludge_001_159)
Candidatus Cloacimonetes (JGI_0000039_G13_Combined_Assembly_WWE1_1_WWE1)
Candidatus Cloacamonas acidaminovorans
Leachate_scaffold_556
Candidatus Cloacimonas (DOLZORAL124_CLOACIMONAS_38_13)
CP WOR-3 (3)
Candidatus Poribacteria (2)
uncultured bacterium RIF8 (1)
CP BRC1 (2)
Aquificae (8)
Candidatus Calescamantes (2)
Deinococcus-Thermus (6)
uncultured bacterium (RIF2_RIFCSPLOWO2_12_FULL_RIF02_62_27)
Thermogotae (9)
KB1 group (1)
Dictyoglomi (2)
Synergistetes (10)
Fusobacteria (8)
Candidatus Atribacteria (4)
Candidatus Atribacteria (2)
CP ACD39 (4)
OP9X (1)
Actinobacteria (172)
Chloroflexi (39)
Chloroflexi (10)
Dehalococcoides mccartyi CBDB1
Dehalogenimonas lykanthroporepellens BL-DC-9
Leachate_Bin_33
Chloroflexi bacterium RBG_13_56_8b
Chloroflexi (3)
Chloroflexi (10)
Chloroflexi (1)
Chloroflexi (1)
Chloroflexi (5)
Chloroflexi (2)
uncultured bacteria RIF (5)
Candidatus Armatimonadetes (10)
Nitrospirae
Acidobacteria
Bacteroidetes
Chlorobi
Gemmatimonadetes
Candidatus Cloacimonas
Aquificae
Deinococcus-Thermus
Thermogotae
Dicytoglomi
Synergistetes
Fusobacteria
Actinobacteria
Chloroflexi

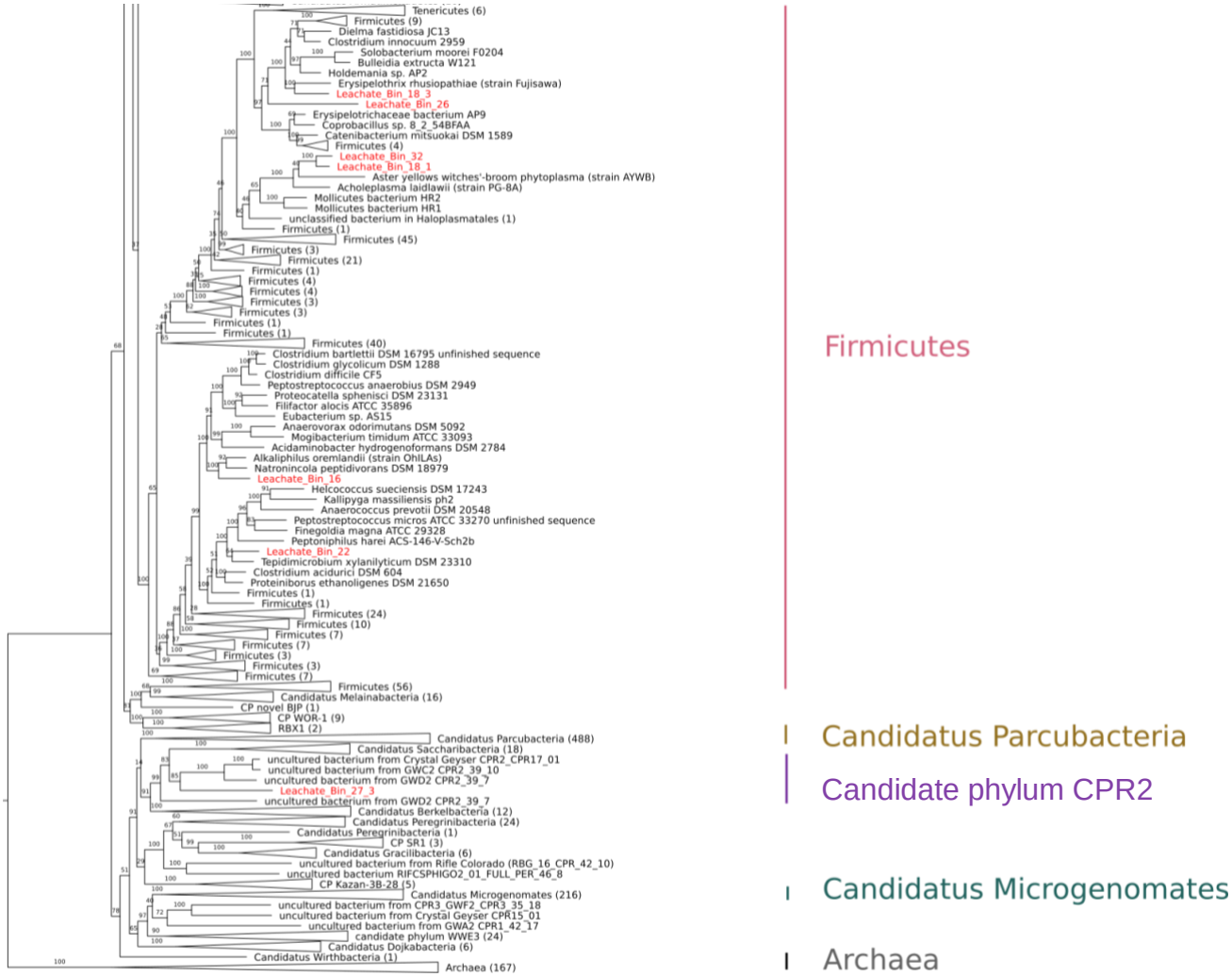


Figure 2.2. Concatenated ribosomal protein tree of the tree of life and metagenome-assembled genomes from the Riverton City dump leachate and Duhaney River. The maximum likelihood tree was constructed from a concatenated alignment of 15 ribosomal proteins (RpL2, -3, -4, -5, -6, -14, -16, -18, -22, -24, and RpS3, -8, -10, -17, -19), including 2,762 bacterial and archaeal reference organisms, 18 MAGs from the leachate, 5 MAGs from the river, and one ribosomal-protein-containing scaffold from the leachate. Organisms in red are from the Riverton City dump leachate and those in blue are from the Duhaney River. The closest relative to each Jamaican organism is named on the tree and taxonomic groups are collapsed where appropriate.

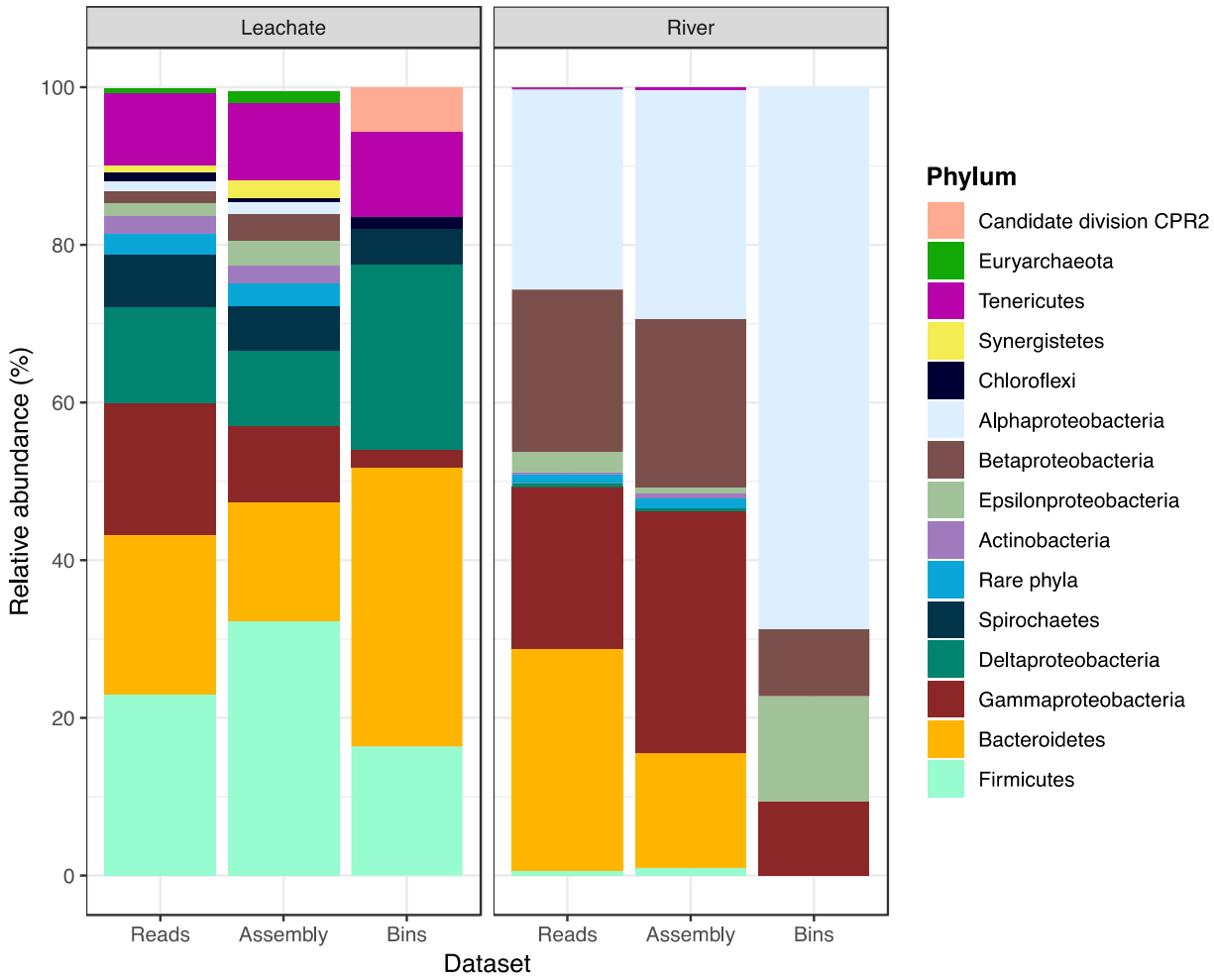


Figure 2.3. Stacked bar comparison of relative abundance of microorganisms from the leachate and river datasets through the sequencing, assembly, and binning pipeline based on marker genes. Phylum-level assignments for unassembled reads, assembled scaffolds, and reconstructed MAGs are displayed for the leachate and river metagenomes. The abundance affiliated with a particular phylum was calculated as follows: i) for reads: the percentage of reads affiliated with the phylum out of all identified 16S rRNA gene-containing reads; ii) for assembled scaffolds: the percentage of scaffolds affiliated with the phylum out of all 16S rRNA-gene encoding scaffolds, iii) the average fold coverage for scaffolds within the MAG, which was taxonomically classified based on the concatenated 15 ribosomal protein tree. Microorganisms that occurred at less than 1% were summed together and labeled as rare phyla for clarity of community proportional abundance visualization.

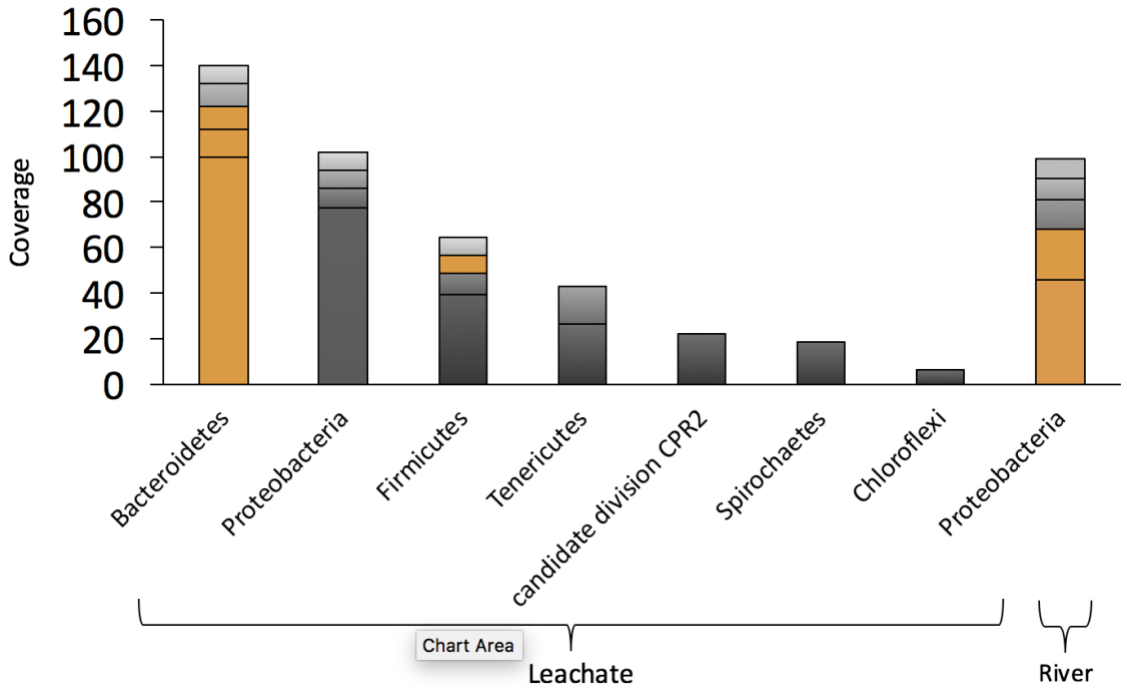


Figure 2.4. MAG-associated diversity of the Riverton City Dump leachate and Duhaney River. The height of each bar represents the summed average coverage of a phylum and each box represents the average coverage of each MAG as a proxy for abundance. Taxonomic assignments were based on a phylogeny inferred from a concatenated alignment of 15 ribosomal proteins (RpL2, -3, -4, -5, -6, -14, -16, -18, -22, -24, RpS3, -8, -10, -17, -19) (Figure 2.2). Genome bins encoding cellulase-containing GH families are highlighted in gold.

and geographic location (Song *et al.*, 2015a; Remmas *et al.*, 2017b). They have been detected in landfills from different continents, including in the United States (Stamps *et al.*, 2016), China (Wang *et al.*, 2017b), and the United Kingdom (Ransom-Jones *et al.*, 2017). Temporal analyses of landfill leachate have revealed that Bacteroidetes and Firmicutes are present in the methanogenic phases of waste decomposition, which suggests that they are involved in anaerobic decomposition of waste (Song *et al.*, 2015b). In a study of cellulose-enriched microcosms in landfill leachate, Bacteroidetes and Firmicutes were enhanced by 5-10% compared to unenriched leachate (Ransom-Jones *et al.*, 2017). The Bacteroidetes and Firmicutes MAGs recovered by Ransom-Jones *et al.* (2017) revealed a variety of CAZyme families involved in polysaccharide degradation, including

those related to cellulose-degradation. In our study, the Bacteroidetes leachate MAGs (Leachate Bin (LB)_10, 8x coverage and LB_12, 12x coverage; as well as LB_7, 100x coverage; LB_17, 10x coverage; and LB_19; 10x coverage) are most closely related to each other to the exclusion of other major Bacteroides groups. In contrast, the closest relatives of the two Firmicute MAGs were *Tepidimicrobium xylanilyticum* (LB_22, 8x coverage) and *Erysipelothrix rhusiopathiae* (LB_18_3; 39x coverage) (Figure 2.1). Members of *Tepidimicrobium* (class Tissierellia) have been isolated from a freshwater hot spring (Slobodkin *et al.*, 2006) and a thermophilic anaerobic digester treating municipal solid waste (MSW) (Niu *et al.*, 2009). *Erysipelothrix rhusiopathiae* is chiefly considered to be an animal pathogen causing erysipelas in pigs and erysipeloid in humans. Animals have been seen foraging in the Riverton Dump from nearby pig and cattle farms (Hamilton, 2012), which may contribute to *Erysipelothrix* presence. No virulence factors were identified from the *Erysipelothrix* bin LB_18_3 (Collins-Fairclough *et al.*, 2018), making its potential for pathogenicity unclear.

Two Desulfobacterales MAGs (LB_8, 8x coverage; LB_9, 7x coverage) and one Desulfuromonadales MAG (LB_4_2, 77x coverage) (order level) were identified, whose closest relatives were *Desulfotignum phosphitoxidans* FiPS 3, *Desulfococcus oleovorans* Hxd3, and *Desulfuromusa kysingii* DSM 7343, respectively (Figure 2.4, Figure 2.2). Deltaproteobacteria have been found at high abundance in landfills (Song *et al.*, 2015a; Stamps *et al.*, 2016). From a survey of 19 landfills across the United States, Desulfuromonadales were identified as one of the dominant classes of Deltaproteobacteria and Desulfobacterales were identified in three landfills that contained low concentrations of sulfate and high concentrations of barium (Stamps *et al.*, 2016). Members of Desulfobacterales and Desulfuromonadales are capable of sulfate reduction (Ontiveros-Valencia *et al.*, 2013) and are likely contributors to this process in landfills.

The moderately high coverage of the two Tenericute MAGs (LB_18_1, coverage 33x; LB_32, coverage 32x) suggest that these populations are abundant in the Riverton City dump leachate (Figure 2.4). The closest relatives of these MAGs are each other, and they were assigned to the family Acholeplasmataceae (Figure 2.2). Tenericutes were reported at high abundance in landfills sampled by Song and colleagues, with *Acholeplasma* as the dominant genus (2015a). *Acholeplasma laidlawii*, the type strain for this group, was isolated from wastewater and has been detected in pig manure slurry (Laidlaw and Elford, 1936; Hanajima *et al.*, 2015). *Acholeplasma* tenericutes are free-living, unlike their parasitic cousins, the Mollicutes (Atobe *et al.*, 1983; Kisary *et al.*, 1976; Maejima *et al.*, 2014).

One MAG (LB_27_3) belonged to the candidate phylum CPR2 and had a coverage of 22x. Populations of candidate divisions are infrequently seen in landfill microbiomes, but candidate divisions OP3, OP9, and OP11 have been detected (Stamps *et al.*, 2016; Huang *et al.*, 2004). Candidate division CPR2 have never been reported in any landfill microbiome studies. At the time of writing, there were only eight public CPR2 draft genomes in the Joint Genome Institute's Integrated Microbial Genomes (IMG) database, all from a groundwater aquifer microbiome study from Rifle, Colorado, USA (Brown *et al.*, 2015). The metabolic functions of candidate phylum CPR2 are not well understood, as they encode small, highly reduced genomes. This new MAG from a landfill environment is a valuable addition to genomic databases.

One MAG (LB_4_3) belonged to the Spirochaetes and had a coverage of 18x. Its closest relative was *Spirochaeta* sp. Buddy (Figure 2.2). Spirochaetes are known to cause syphilis and Lyme disease but some species are also free-living in waters and sediment (Gerbase *et al.*, 1998; Dworkin *et al.*, 2008; Canale-Parola *et al.*, 1968; Canale-Parola, 1984). They have also been detected in landfills (Song *et al.*, 2015b; Stamps *et al.*, 2016; Remmas *et al.*, 2017b). Ransom-Jones and

colleagues identified the bacterial composition of raw landfill leachate and cellulose-enriched microcosms, and found that Spirochaetes were present in both. Read counts of Spirochaetes specifically from the family Spirochaetaceae were greater in the cellulose enrichment by an average of ~5% (Ransom-Jones *et al.*, 2017). The study posits Spirochaetes are one of the major phyla in degrading cellulose in the dump microbial community, possibly as a symbiotic partner to cellulose-degrading lineages. Some members of the Spirochaetes, such as those in the rumen, enhance cellulose degradation despite not being cellulose degraders (Stanton and Canale-Parola, 1980; Kudo A N *et al.*, 1987). The Spirochaete in the Riverton dump may be playing a similar role in supporting cellulose degradation.

The gammaproteobacterial MAG (LB_15) had a coverage of 9x. Its closest relative was *Halomonas elongata* DSM 2581 (Figure 2.2). Gammaproteobacteria have been detected in relatively high abundance in landfill leachate (Xie *et al.*, 2012; Song *et al.*, 2015a; Stamps *et al.*, 2016; Remmas *et al.*, 2017b). Members of *Halomonas* have been isolated from a variety of environments including marine, municipal sewage, hypersaline soils and waters, and alkaline soda lakes (Rafael R. de la Haba *et al.*, 2006; Remmas *et al.*, 2017a). *Halomonadaceae* are generally halophilic, and thus perhaps well suited to the high salt concentrations of mature landfill leachate (Remmas *et al.*, 2017a).

We reconstructed one Chloroflexi MAG (LB_33) with a coverage of 6x, which was placed in the class Dehalococcoidetes (Figure 2.2). Chloroflexi have been previously detected in landfill leachate at low abundance (Song *et al.*, 2015a, 2015b; Remmas *et al.*, 2017b; Bareither *et al.*, 2013; Stamps *et al.*, 2016). Chloroflexi are associated with both aerobic and anaerobic metabolisms and have been found in freshwater sediments and anaerobic sludge (Inagaki *et al.*, 2003; Hug *et al.*, 2013). *Dehalococcoides* are the only known organisms able to anaerobically

respire the groundwater contaminants tetrachloroethenes (PCE) and trichloroethenes (TCE) to the non-toxic product ethene (Maymó-Gatell *et al.*, 1997). These chloroethenes are in dry-cleaning solvents, adhesives, paints, and caulking. PCE contaminants have been detected in groundwater aquifers near a landfill and a special waste compound in a landfill in Ontario (Canadian Council of Ministers of the Environment, 1999; Jackson, R.E. *et al.*, 1991). The metabolic niche of an aerobic Dehalococcoidetes in the leachate is unclear; our leachate sample was taken from the surface and is presumed to be oxic.

In the Duhaney River metagenome, all reconstructed genomes were from the Proteobacteria. Two alphaproteobacterial MAGs (RB_10 and RB_4_2) were each other's closest relative, both affiliated with the family Rhodobacteraceae. These were the most abundant river MAGs, with 22x and 45x coverage, respectively. Marine Rhodobacteraceae encode various proteins that defend against toxic compounds and heavy metals (Simon *et al.*, 2017). These proteins include (S)-2-haloacid dehalogenase which detoxifies organohalogens largely produced by macroalgae, and mercury(II) reductase which reduces toxic mercury compounds to volatile mercury(0). Work led by Dr. Collins-Fairclough showed that RB_10's genome encoded genes against some predicted metal resistance and two classes of antibiotics; no similar traits were found for RB_4_2 (Collins-Fairclough *et al.*, 2018).

One river MAG (RB_5, 9x coverage) was only able to be identified at the class level, suggesting it is a novel lineage of Gammaproteobacteria. Gammaproteobacteria have been detected at relatively high abundance in tropical rivers and a polluted urban river through 16S rRNA gene sequencing (Thoetkiattikul *et al.*, 2017; García-Armisen *et al.*, 2014). Some members of the class are opportunistic heterotrophs that exploit nutrient-rich microniches and irregular availability of organic matter (Pernthaler and Amann, 2005; Lau *et al.*, 2013). Gammaproteobacterial abundance

has also been found in tandem with phytoplankton blooms, where they are likely involved in the cycling of dissolved organic matter (Buchan *et al.*, 2014). Dr. Collins-Fairclough's virulence analysis predicted that RB_5 is broadly resistant to metals and resistant to five antibiotic classes (Collins-Fairclough *et al.*, 2018).

The Epsilonproteobacteria MAG (RB_13) was present at 13x coverage, and was most closely related to *Arcobacter nitrofigilis* DSM 7299. *Arcobacter nitrofigilis* was originally called *Campylobacter nitrofigilis* but was later reclassified (Vandamme *et al.*, 1991). The strain of free-living *A. nitrofigilis* was isolated from the roots of cordgrass in coastal salt marshes. *A. nitrofigilis* has also been found in meat and shellfish, but none of these strains were pathogenic (Collado *et al.*, 2009), unlike *Campylobacter* spp. which are usually pathogens of warm-blooded animals (McClung and Patriquin, 1980).

One MAG (RB_8_2, 8x coverage) belonged to the betaproteobacterial family Comamonadaceae. Members of the Comamonadaceae have been found in soil and water (Willems *et al.*, 1991). They are aerobic chemoorganotrophs or chemolithotrophs, and most are motile via flagella (Willems *et al.*, 1991). *Comamonas* sp. have been shown to mineralize nitrobenzene and release nitrite (Nishino and Spain, 1995). Nitrobenzenes are largely produced industrially for aniline and motor oil production, and can also occur in small quantities in the manufacture of pesticides, synthetic rubber, and dyes (Agency for Toxic Substances and Disease Registry, 1990).

It is interesting to note that there is no explicit evidence of contamination of the Duhaney River from the Riverton dump from our samples: there is very little overlap in the populations detected (<0.5% of the reads from the leachate mapped to the river assembly and vice-versa) (Collins-Fairclough *et al.*, 2018). Due to a lack of spatial and temporal samples of the leachate and river, findings here cannot speculate on the presence or absence of leachate contamination in the river.

More spatial and temporal samples downstream of the dump would help to cover a greater portion of the microbial populations in the river, identify the core microbial composition, and investigate whether the microbial populations overlap with that of the leachate.

2.3.4 Disparities between the community composition as defined by reads, assembled scaffolds, and MAGs

In general, the taxonomic composition and relative abundances of the leachate microbial community in the assembled metagenomes reflected that of the reads, varying by less than 3% (+/- 2.8%) for a given taxonomic group (Figure 2.3). Exceptions were the Bacteroidetes, Gammaproteobacteria, and Firmicutes which differed by - 5.2%, -7%, and +9.3%, respectively in the assembly compared to the reads. In the river reads and assembly, the taxonomic group abundances were within +/- 3.8%, except for Gammaproteobacteria and Bacteroidetes reads, which exhibited +10.2% and -13.5%, respectively in the assembly compared to the reads. The MAGs in both leachate and river captured most of the prominent phyla in the scaffold, except that no Bacteroidetes river MAGs were assembled despite 14.6% of scaffolds in the assembly affiliating with the Bacteroidetes. Higher coverage for more abundant organisms facilitates generation of relatively good quality draft genomes (>70% complete and <10% redundant). This is seen in the recovered MAGs, which are mostly affiliated with the dominant phyla from the reads and assemblies.

2.3.5 Cellulase diversity and abundance

The same number of putative cellulases were identified in the leachate and river metagenomes. There were 46 genes identified from the leachate across six GH families (GH1, -3, -5, -8, -30, and -51) and 46 from the river spanning seven GH families (GH1, -3, -5, -8, -9, -30, and -51)

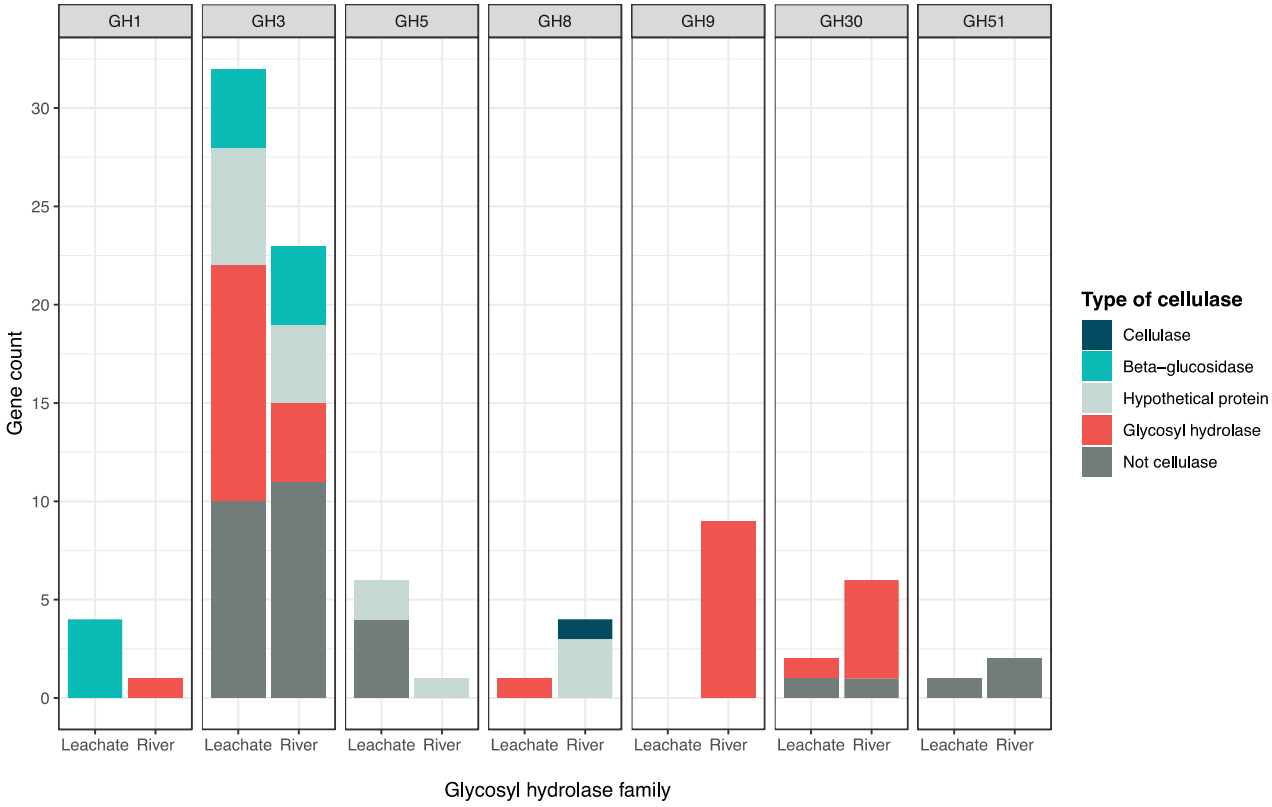


Figure 2.5. Abundance of cellulase types by glycosyl hydrolase families in the leachate and river metagenomes after manual curation. Genes whose protein sequences were annotated as glycoside hydrolase family n (where n is the family number) are grouped under "Glycosyl hydrolase." Protein sequences that were not annotated as of the three types of cellulases, annotated as "cellulase," hypothetical proteins, or "glycosyl hydrolase n", were grouped under "Not cellulase."

(Figure 2.5). Genes belonging to these seven GH families have previously each been identified in at least one of five MAGs constructed from a cellulose-amended microcosm in landfill leachate (Ransom-Jones *et al.*, 2017). Putative beta-glucosidases were identified in both the leachate and river metagenomes. A putative cellulase was only identified in the river. No genes explicitly annotated as "endocellulase" nor "exocellulases" were identified from either metagenome. A total of eight genes in the leachate and four in the river were annotated as beta-glucosidases from the GH1, -3, and -5 families. The beta-glucosidase protein sequences had identities ranging from 47-57% and E-values from 0 - $3.52e^{-76}$ to their respective best hits. One gene from the river was

annotated as "cellulase." Cellulase is used synonymously with endoglucanase in the CAZy database and in some literature (Alvarez *et al.*, 2013; Hug *et al.*, 2013) so this gene is potentially an endocellulase (identity: 84%, E-value: 0). A small number of genes (22 from leachate, 27 from river) were annotated as glycosyl hydrolase family n (where n is the family number) or as hypothetical proteins. From leachate, the "Glycosyl hydrolase" genes ranged from 43-73% identity to their respective closest-related proteins with E-values of 0 - $3.85e^{-87}$, whereas the hypothetical proteins ranged from 33-57% identity and had E-values of 0 - $6.37e^{-34}$ (and one gene whose match to its closest-related protein had an E-value of 6.1). The low identities (and a relatively large E-value of 6.1) suggest that these proteins have either not been recorded in public databases, are novel cellulases, and/or are novel GHs, possibly from undescribed microorganisms. Investigation through protein modelling could provide more insight as to how similar these proteins' structures are to existing cellulases. Genes that were best hits to non-cellulases (16 from leachate, 14 from river) were recovered from the HMMs because the GH families contain multiple enzymatic activities.

There is no published metagenomic study of landfills examining the distribution and diversity of the three types of cellulases. Cellulolytic function and cellulase genes have instead been identified through enzymatic assays and one genomic analysis of isolates from landfill leachate (Pourcher *et al.*, 2001; Masngut and Manap, 2017; Chua *et al.*, 2014). The leachate community contained predicted beta-glucosidases and the river community contained potential beta-glucosidases and cellulases/endocellulases. This suggests that the cellulolytic capacity in leachate is low and that cellulose degradation may be possible in the river. Usually, endocellulases, exocellulases, and beta-glucosidases are needed for the complete degradation of cellulose. However, studies have shown that combinations of microbial endocellulases and beta-glucosidases can generate glucose

from cellulose (Jeon *et al.*, 2009; Kim and Ishikawa, 2010a). Our results suggest that cellulose may be degraded in the landfill leachate and in the river, but this degradation would be more efficient if exocellulases were present as well. The wealth of functionally unknown proteins in the "Other" category provides the possibility that uncharacterized cellulases or exocellulases are present at the sampled sites.

2.3.6 Taxonomic diversity of cellulases

Bacteroidetes and Firmicutes were affiliated with putative cellulases on scaffolds from both metagenomes (Figure 2.6). Scaffolds affiliated with Actinobacteria, Spirochaetes, Synergistetes, and Tenericutes were only recovered in the leachate metagenome, whereas those affiliated with Proteobacteria (Alpha-, Beta-, Gamma-, and Epsilonproteobacteria) were recovered in the river. Of the 62 total potential cellulases (does not including proteins annotated as non-cellulases) detected across both environments, 17 genes (29%) belonged to Bacteroidetes; 8/30 (27%) of those were from leachate and 9/32 (33%) were from the river. Bacteroidetes-assigned cellulases were mainly from GH3 in both metagenomes, and additionally from GH30 in river. A total of 9/62 (15%) potential cellulases were associated with Firmicutes, with 8/30 (27%) in the leachate and only 1/32 (3%) in the river. Firmicute-associated potential cellulases from leachate were from GH3 and GH1, whereas the river one was from GH3. Potential cellulase genes affiliated with the Spirochaetes, Tenericutes, Actinobacteria, Synergistetes, and Balneolaeota were only identified in leachate and totalled to 14/30 (47%) genes, with less than four potential cellulase genes were affiliated with each of these phyla. Proteobacterial cellulases were identified only in the river. Alphaproteobacteria- and Gammaproteobacteria-associated potential cellulases were at a

relatively high abundance in the river, both were at 10/32 (32%). Beta- and Epsilonproteobacteria-associated potential cellulases were present in the river at 1/32 genes each (3%).

Taxonomic identification of potential cellulases was largely to lineages implicated in cellulase degradation. It was not surprising to identify potential cellulases associated with Bacteroidetes and Firmicutes, as several classes within these phyla are known to degrade cellulose, CMC, and/or cellobiose (e.g., members of the *Anaerophaga, Marinilabilia, , Ohtaekwangia,* and *Clostridium*) (Denger *et al.*, 2002; Suzuki *et al.*, 1999; Yoon *et al.*, 2011; Koeck *et al.*, 2014). The Bacteroidetes-affiliated cellulases from the leachate and river shared highest similarity with *Blautia, Hydrotalea, and Lentimicrobium* genes. None of these Bacteroidetes genera have been shown to degrade cellulose or any of its smaller constituents. However, some members of *Ruminococcus* were renamed to *Blautia,* and members of *Ruminococcus* are known cellulose degraders (Liu *et al.*, 2008; Koeck *et al.*, 2014). Members of the Bacteroidetes and Firmicutes possessing cellulose-active enzymes are quite common and are usually found at high abundance in anaerobic environments, such as the rumen of cows, mangrove ecosystems, and soils (Güllert *et al.*, 2016; Thompson *et al.*, 2013; López-Mondéjar *et al.*, 2016).

The potential cellulases annotated as beta-glucosidases affiliated with the Spirochaetes were most closely related to proteins from *Sphaerochaeta globusa* strain Buddy, *Sphaerochaeta pleomorpha,* and *S. caldaria* (previously known as *Treponema caldaria*) (Abt *et al.*, 2013). *Sphaerochaeta caldaria* has been shown to enhance cellulose degradation in co-cultures with cellulolytic bacteria (Caro-Quintero *et al.*, 2012; Pohlsehroeder *et al.*, 1994); however, the glycosyl hydrolases these organisms possess have yet to be investigated for their substrates and catalytic functions.

The potential cellulases that were most closely related to Tenericutes were related to genes from members of the *Acholeplasma*. Although occurring here at low abundance, Tenericutes have been identified in 16S rRNA gene surveys from the gut microbiomes of termites, where cellulolytic protists ferment cellulose to generate H_2 and CO_2. Endosymbiotic bacteria convert these products to acetate, and provide amino acids via nitrogen fixation to the protists (Brune, 2014; Hongoh *et al.*, 2008; Sabree and Moran, 2014; Ohkuma *et al.*, 2015).

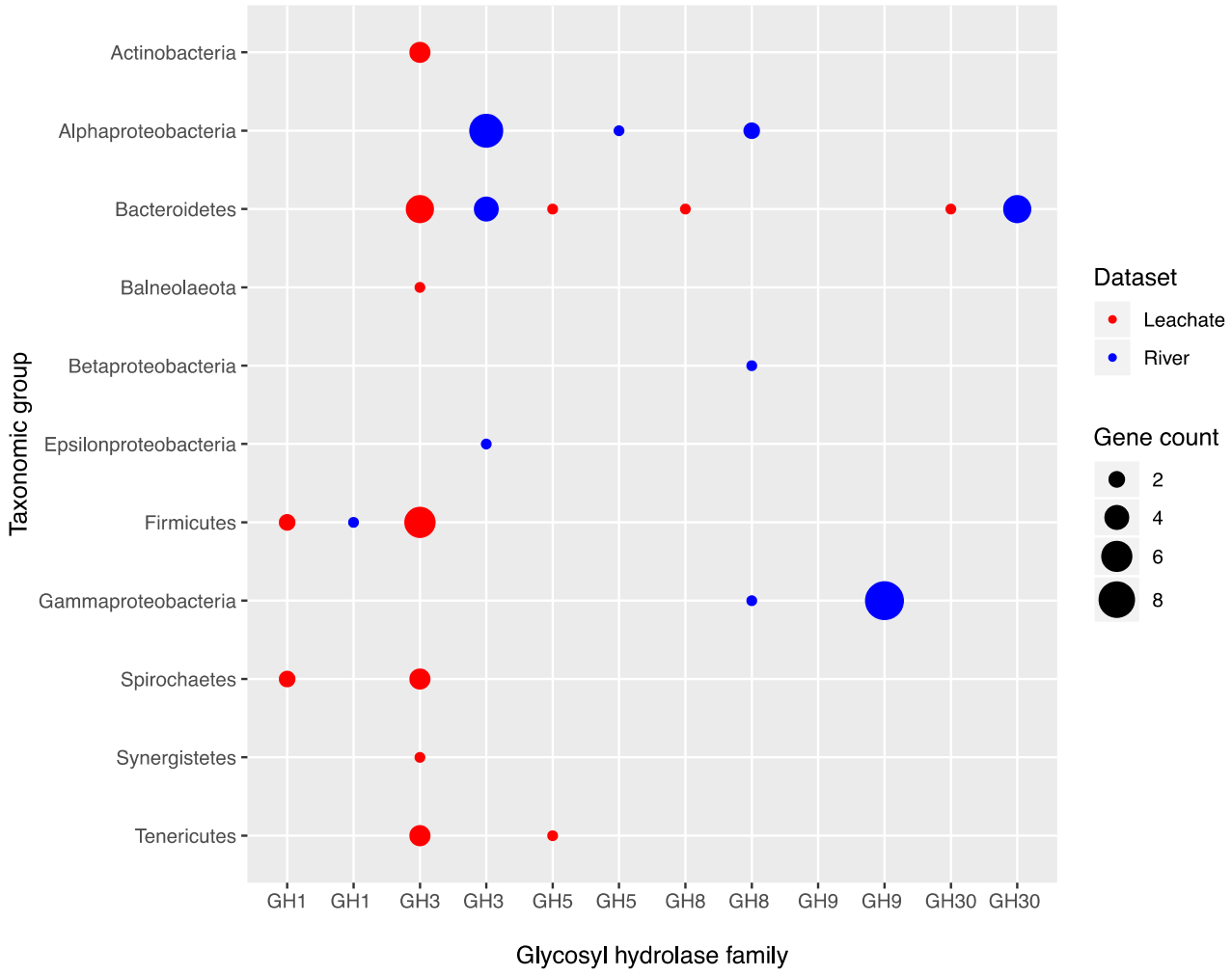


Figure 2.6. Taxonomic composition of GH families containing cellulases in the leachate and river metagenomes. Glycosyl hydrolases along the x-axis indicate those for leachate (first) and river (second).

The potential cellulases affiliated with the Actinobacteria were most closely related to genes from *Lysinimicrobium* sp., *Nitriliruptor alkaliphilus*, and *Demequina aurantiaca*. *Lysinimicrobium*

species isolated from mangrove forest soil were able to use cellobiose (Hamada *et al.*, 2015). A strain of *N. alphaliphilus* has been isolated from an isobutyronitrile enrichment culture inoculated with soda lake sediments and can utilize cellobiose but not cellulose (Sorokin *et al.*, 2009). *D. aurantiaca* was isolated from seaweed from a lake in Japan and was shown to use cellobiose (Ue *et al.*, 2011). The genes from our survey were related to ambiguously annotated glycosyl hydrolases, with identities from 56-60% to their best hits, thus they may be new genes either belonging to these species or their relatives.

A potential cellulase affiliated with the Synergistetes was most closely related to a gene from *Aminiphlilus circumscriptus* which a strain has been isolated from an anaerobic sludge reactor treating wastewater (Díaz *et al.*, 2007). There is no evidence that *A. circumscriptus* can utilize cellobiose, and cellulose has not been tested as a substrate.

The predicted cellulases in the river annotated as "hypothetical proteins" most closely related to genes from the Alphaproteobacteria affiliated with *Martelella mediterranea*, *Martelella* sp., *Marivita cryptomonadis*, *Sphingobium lactosutens*, *Tropicimonas isoalkanivorans*, *Roseomonas stagni*, and *Wenxinia marina*. *M. mediterranea* conducts cellobiose and carboxymethylcellulose (CMC) degradation (Dong *et al.*, 2010). A strain of S. *lactosutens isolated from a hexachlorocyclohexane dump site* has been shown to utilize cellobiose (Kumari *et al.*, 2009). T. *isoalkanivorans* isolated from seawater from a port in Indonesia was also positive for cellobiose utilization (Harwati *et al.*, 2009). *M. cryptomonadis* was isolated from a marine phytoplankton culture of *Cryptomonas* sp. and is able to utilize cellobiose (Hwang *et al.*, 2009). *R. stagni* has not been tested for cellulose or cellobiose utilization but other species of *Roseomonas* are not able to hydrolyze CMC (Kim *et al.*, 2003; Sik Baik *et al.*, 2012). A strain of *W. marina* was isolated from oilfield sediments and also demonstrated CMC and cellobiose hydrolysis (Riedel *et al.*, 2014;

Tokuda *et al.*, 2005). Two genes showed relatively high identity (89-96%) to glycosyl hydrolases associated with Rhodobacteraceae, suggesting that these genes are promising potential cellulases from this family or close relatives of it. Further examinations of hypothetical proteins (i.e. gene cloning and activity assays) are needed to verify whether the predicted cellulases from our datasets are enzymes involved in cellulose degradation.

In the river dataset, potential cellulases were closely related to glycosyl hydrolases from the Gammaproteobacteria *Agaribacterium haliotis* and *Simiduia agorivorans*, as well as a hypothetical protein from *Methylomarinum vadi*. Gammaproteobacteria have also been found in marine biofilms on cellulose baits (Edwards *et al.*, 2010). *A. haliotis* has been isolated from abalone feces and is not able to degrade cellulose (Huang *et al.*, 2017). *S. agorivans* is a cellulose degrader that has been isolated from coastal waters (Shieh *et al.*, 2008). *M. vadi* is a methanotroph that has been isolated from marine environments, but its capacity for cellulose and cellobiose degradation has not been tested (Hirayama *et al.*, 2013).

A potential cellulase was closely related to a glycosyl hydrolase from Epsilonproteobacteria belonging to *Campylobacter showae*. Cellulose degradation by *C. showae* has not been tested.

It is important to note that the literature presented here regarding the ability of these genera to degrade cellulose substrates are depended on the experimental conditions under which the microorganisms were tested. Thus, those that were not able to degrade cellulose at certain conditions may be active under other conditions, including different temperatures and the presence or absence of oxygen.

2.3.7 Cellulase genes identified in MAGs

Potential cellulase genes were identified in six of the twenty-three taxonomically classified

MAGs from the leachate and river metagenomes (Figure 2.4, Figure 2.7). The six MAGs included three Bacteroidetes (LB_7, LB_12, and LB_19), one Firmicute (LB_26), and two

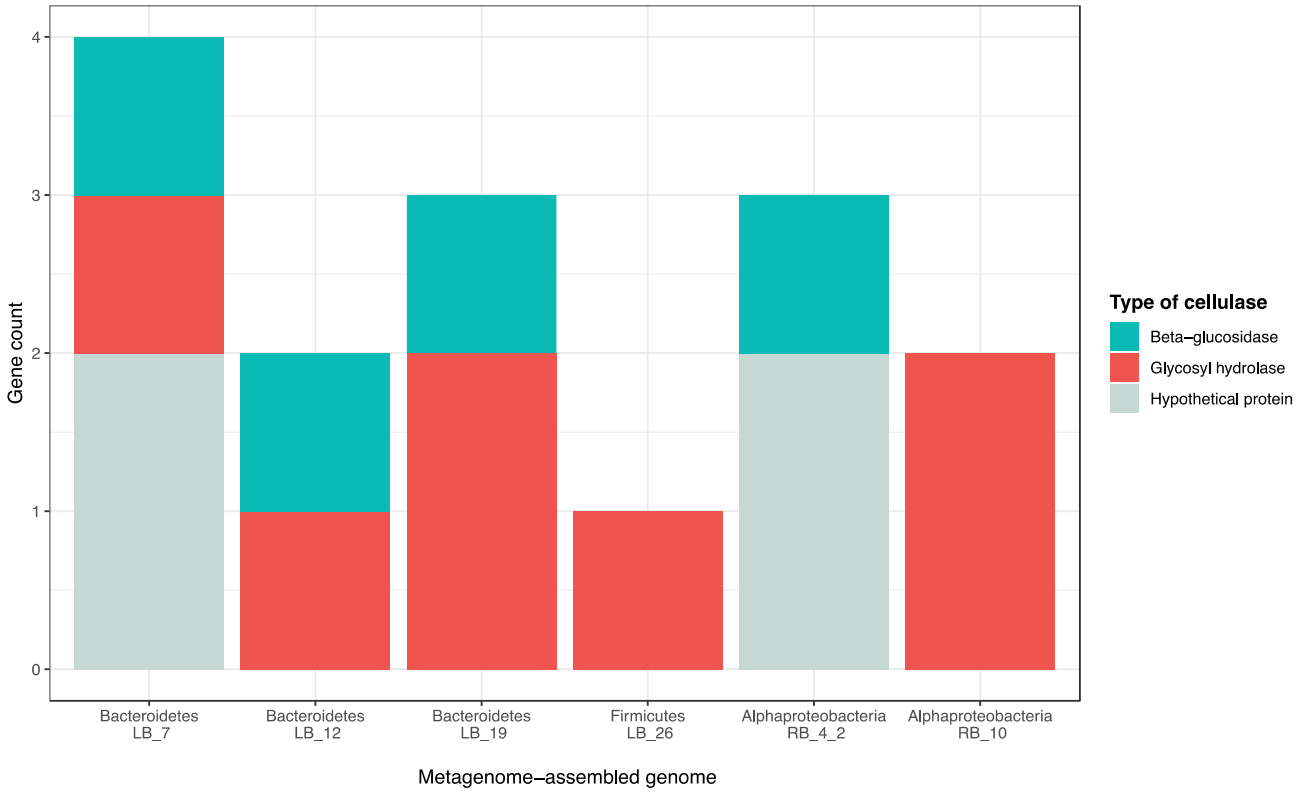


Figure 2.7. Types of potential cellulases detected in high-quality MAGs. The MAGs are listed in order of highest to lowest mean genome coverage for each dataset. The "Glycosyl hydrolase" category includes proteins that were annotated with or without a specific GH family. LB = Leachate Bin, RB = River Bin.

Alphaproteobacteria (RB_4_2 and RB_10). Genes whose closest hits were to beta-glucosidases were in 4/6 MAGs and glycosyl hydrolases were identified in five of the putatively cellulolytic bins, with the exception of Alphaproteobacteria RB_4_2. The potential functions encoded by genes with closest hits to GHs and hypothetical proteins may encompass other carbohydrate enzymatic activities found within the respective GH families. The populations for which we have high-quality MAGs are not predicted to hydrolyze cellulose effectively based on the presence of only beta-glucosidase, as endo- and exocellulases are typically necessary for this process.

2.4 Conclusions

Leachate MAGs possessed a greater diversity at the phylum level than river MAGs. In leachate MAGs, Bacteroidetes, Proteobacteria, Firmicutes, and Tenericutes were the abundant phyla, whereas only Proteobacteria were recovered from the river metagenome. Although there was little overlap between the microbial profiles based on the classes of the MAGs recovered between leachate and river, neither presence nor absence of contamination of the river from the leachate can be concluded.

From the pool of potential cellulases identified in the metagenomes, the majority of beta-glucosidases identified were classified as GH1 and GH3, and were mostly affiliated with members of the Bacteroidetes and Firmicutes. The absence of endocellulases and exocellulases suggests that cellulose degradation is unlikely by the sampled communities. The identified hypothetical proteins and the ambiguously annotated glycosyl hydrolases may be a source of novel cellulases in these environments. Cellulose decomposition or potential cellulases have been detected in mangrove ecosystems, river sediments near a sewage treatment plant, and in lotic ecosystems fueled by foliage (Thompson *et al.*, 2013; Friedmann *et al.*, 1979; Yue *et al.*, 2016). These environments have either natural or engineered sources of organic material input. The presence of potential enzymes involved in cellulose degradation in the river metagenome suggests the presence of microorganisms active on cellulose and/or cellobiose, or that can enhance cellulose degradation. The sources of the organic matter in the river may be from leachate contamination, foliage falling into the water, or organic material run-off from farms upstream.

A greater sample size including temporal and spatial samples of the leachate and river would be needed to elucidate the prevalent microbial populations over time and the possibility of contamination of the Duhaney river by the Riverton City Dump leachate. Spatial samples of the

river downstream of the Riverton City Dump and analyses of the chemical composition of the leachate and the river water could be compared to further ascertain the presence and scope of contamination of the river by the leachate.

Chapter 3: Microbial Diversity of Predicted Cellulases in the Southern Ontario Landfill

3.1 Introduction

Paper waste, which includes copy paper, newspaper, and cardboard, makes up the greatest fraction of waste in landfills. The rate of waste decomposition depends on environmental conditions, where the process is much slower in dry climates, and on leachate circulation through the refuse, which accelerates decomposition (Kjeldsen *et al.*, 2002). Waste is degraded by microorganisms. The microbial diversity of landfills has been examined via 16S rRNA gene sequencing but microbial functional potential still remains understudied. Specifically, studying cellulose decomposition in landfills would be a promising approach to discovering novel cellulases.

Recent studies have predicted cellulose degradative capacity in landfills based on cellulase-containing GH families and/or based on the putative cellulolytic microbes identified by 16S rRNA gene amplicon sequencing and metagenomic analyses (Song *et al.*, 2015a; Ransom-Jones *et al.*, 2017; Stamps *et al.*, 2016). There are no published metagenomic studies that define the functional composition of genes belonging to cellulase-containing glycosyl hydrolase families in landfill refuse or landfill leachate. This distinction is important, as it separates non-cellulases from cellulases, allowing us to more precisely estimate the abundance and the taxonomy of the cellulases at a site. The common practice of reporting GH families in place of cellulases likely overestimates the abundance of cellulases in an environment. Instead, surveying true cellulolytic potential has been done by biochemical characterizations or genome analyses from selected isolates cultured from landfill refuse (Pourcher *et al.*, 2001; Li *et al.*, 2009b). From a phenotypic and biochemical survey of cellulolytic isolates cultured from refuse, 355 cellulolytic bacteria were isolated (Pourcher *et al.*, 2001). In another study, a type of endocellulase that degrades carboxymethylcellulose, a more soluble derivative of cellulose, was detected in *Bacillus* species

isolated from a landfill in India (Korpole *et al.*, 2011). Draft genomes of two *Paenibacillus* strains from landfill leachate revealed thirteen genes encoding endocellulases, exocellulases, and beta-glucosidases (Chua *et al.*, 2014). Although these studies demonstrated cellulolytic activity from landfills, they worked only with culturable isolates, most of which belong to phyla and genera whose members were already known to degrade cellulose. These genera are therefore not representative of the novel cellulolytic potential in landfills. More metagenomic studies targeting cellulolytic function in landfills are needed to expand our current knowledge of cellulolytic capacity in novel taxa. The discovery of new cellulolytic microbial diversity and their potential metabolic functions are then available to be optimized and applied to industrial and biotechnology processes.

The cellulolytic microbial potential at an inactive area of the southern Ontario landfill was examined through analysis of metagenomes derived from leachate and groundwater wells. The sampling area actively received waste from 1972-2001, with leachate and groundwater monitoring as well as gas collection continuing to date. Sampling sites included one composite leachate cistern, three leachate wells transecting the landfill area, and groundwater from a perched water table. The objective of this chapter is to examine the diversity of cellulolytic enzymes from the southern Ontario landfill.

3.2 Materials and Methods

3.2.1 Sampling sites and sample collection

Five sites of interest from the landfill in southern Ontario, Canada were sampled (Table 3.1; Figure 3.1, Appendix A for original names of sampling sites at the landfill). The pumphouse is a composite leachate cistern (CLC), a reservoir where all the leachate in the landfill is pooled before

it is transported to a wastewater treatment plant. Pumphouse 3 was sampled twice, a week apart, where July 14th, 2016 (CLC1_T1) was a preliminary sampling trip and July 20th, 2016 (CLC1_T2) was the official expedition when the remaining sites were also sampled. The leachate wells were chosen to form a transect across the central, inactive landfilling area. A groundwater well on the periphery of the inactive landfill area was additionally sampled.

Table 3.1. Sampling sites and dates at the inactive landfilling area of the Ontario landfill.

Site_time point	Type	Date sampled
CLC1_T1	Composite leachate cistern	14JUL2016
CLC1_T2		20JUL2016
LW1	Leachate well	20JUL2016
LW2	Leachate well	20JUL2016
LW3	Leachate well	20JUL2016
GW1	Perched water table/groundwater well	20JUL2016

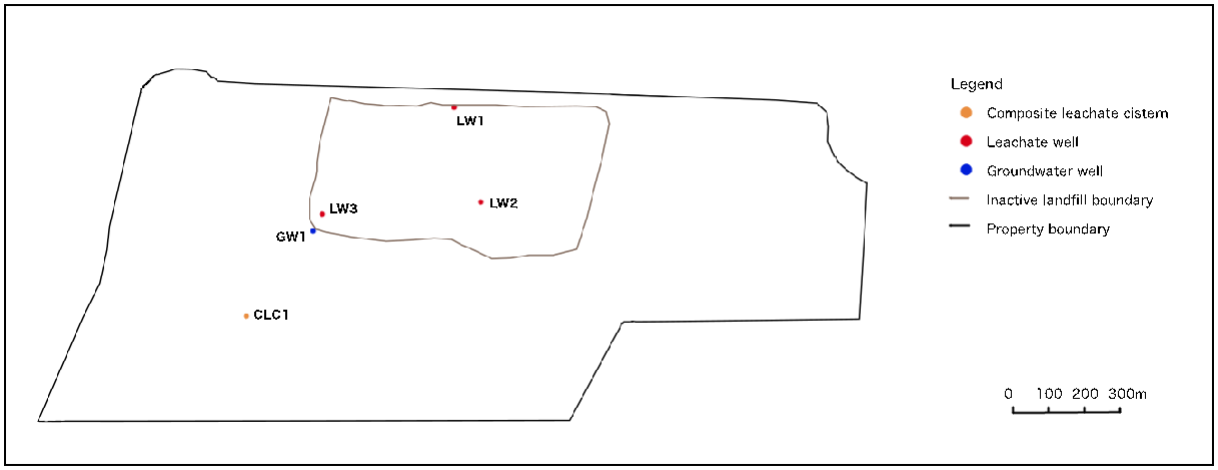


Figure 3.1. Schematic of the sampling sites at the Ontario landfill. The map was modified from the 2014 landfill management report.

Leachate and groundwater were collected by Hug Lab members using carboys that were first rinsed with the sample to be collected and then were used to collect 4-5 gallons of samples. The protocol to filter samples differed slightly between the two sampling days. On July 14th, the CLC1_T1 sample was filtered through 144 mm Supor polyethersulfone membranes of 0.2 µm and 0.1 µm

pore sizes in series (Pall Corporation, New York, New York) using an electrical pump. Approximately 150 mL of leachate was filtered through each of three pairs of filters prior to clogging. On July 20th, samples were filtered through a 3.1 µm glass fibre pre-filter (Pall Corporation, New York, New York) and a 144 mm Supor® polyethersulfone 0.1 µm membrane in series. Sample volumes ranged from 150 mL (CLC) to ~10 L (GW1), and were determined based on time to filter clogging or filtering of all collected liquid.

3.2.2 Genomic DNA extraction

Total community genomic DNA was extracted from one half of a 0.1 µm filter from each July 20th location and one half of a 0.2 µm filter and a 0.1 µm filter from the CLC1_T1 using the MO BIO PowerSoil DNA Kit according to manufacturer's instructions, with the filter cut up into small pieces and used as direct input in place of soil. The eluted DNA was further concentrated using an overnight ethanol precipitation in 0.3 M sodium acetate and 2.5 volumes of ice cold ethanol. Following pelleting and resuspension of the DNA in molecular grade water, the concentrations and $A_{260/280}$ ratio of the DNA was measured using the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The CLC1_T1 0.1 µm filter extraction resulted in insufficient DNA for metagenomic sequencing.

Six samples were sent for sequencing: leachate: CLC1_T1, CLC1_T2, LW1, LW2, LW3, and GW1. Library preparation and shotgun metagenome sequencing was done on a HiSeq 2500 1 TB platform (Illumina, Hayward, CA) with subsequent assembly and annotation also conducted by the U.S. Department of Energy Joint Genome Institute (JGI), (Walnut Creek, CA).

3.2.3 Data availability

The Southern Ontario metagenomes are deposited on IMG with the following IMG Genome IDs (Taxon Object IDs): 3300014203 (CLC1_T1), 3300014206 (CLC1_T2), 3300014204 (LW1), 3300015214 (LW2), 3300014205 (LW3), and 3300014208 (GW1).

3.2.4 Identifying potential cellulase genes in the Ontario landfill metagenomes

Through the IMG-ER online interface, a functional query for Pfam IDs (Finn *et al.*, 2016) of glycosyl hydrolase (GH) families containing cellulases was used to search the six metagenomes for potential cellulases. The GH families searched were GH1, -3, -5, -6, -7, -8, -9, -12, -30, -44, -45, -48, and -116. Search hits for each GH family were then identified taxonomically using blastp with an e-value threshold of $<1e^{-40}$. All hits with e-values above $1e^{-40}$, regardless of whether they were annotated as cellulases or not, were excluded in the analysis of potential cellulases. The hits were then filtered for putative cellulolytic function, and categorized. For simplicity, the extensive synonymous names for different types of cellulases were simplified to three general categories: endocellulases, exocellulases, and beta-glucosidases. For example, genes that were annotated as 1,4-beta-D-glucan glucohydrolases or cellobiohydrolases were categorized as exocellulases, and endoglucanases were categorized as endocellulases. Other gene annotations that did not belong to those three categories included cellulases, hypothetical proteins, and glycosyl hydrolase family n proteins (where n is the GH family number). The uncategorized annotation hits were kept in the subset used for further analyses. Genes that were affiliated with the respective Pfam IDs but were not kept as part of the cellulolytic subset included glycosyl transferases, xylanases, mannanases, and membrane proteins. Genes affiliated with Eukaryotes which were not of interest were removed from analyses as well.

3.3 Results and Discussion

3.3.1 Statistics of southern Ontario landfill metagenomes

Genes annotated with Pfam IDs of cellulase-containing glycosyl hydrolase families were extracted from the six Ontario landfill metagenomes. The metagenomes ranged from 1.8-2.5 billion bases and 3.2-5.0 million genes (Table 3.2). The groundwater well metagenome produced the smallest metagenome and the second CLC time point the largest. Genes associated with Pfam GH families 1, -3, -5, -6, -7, -8, -9, -12, -30, -44, -45, -48, and -116 were identified from the CLC,

Table 3.2: Statistics for the southern Ontario landfill metagenomes.

	CLC1_T1	**CLC1_T2**	**LW1**	**LW2**	**LW3**	**GW1**
Size of metagenome (millions of bases)	2,244	2,497	2,297	2,151	2,274	1,835
No. of genes in assembly	4,628,938	5,019,461	4,947,135	3,941,155	4,713,313	3,307,093
Subset of genes associated with GH families containing cellulases (A); % of assembly	6,153; 0.13%	4,729; 0.09%	1,689; 0.03%	2,447 0.06%	2,127 0.05%	2,524 0.08%
Subset of genes (A) that are potential cellulases*; % of (A)	5281; 86%	4006; 85%	1482; 88%	1874; 77%	1728; 81%	1842; 79%
Subset of genes (A) annotated as specific types of cellulases; % of A	1,326; 25%	1,018; 25%	340; 23%	499; 27%	491; 28%	307; 17%

*Genes annotated as endocellulase, exocellulase, beta-glucosidase, and cellulase were included, as well as genes that could potentially be cellulases but not specifically annotated as such which included hypothetical proteins and GH family n protein (where n is a particular GH family).

LW, and GW metagenomes to analyze the proportion of glycosyl hydrolases with potential cellulolytic function. These families also include enzymes with activities other than cellulolysis. Less than 0.2% of genes in all metagenomes were annotated as belonging to the families surveyed. Of the total genes in the metagenomes, 77-88% of proteins were considered to be potential cellulases at each site. The criteria for inclusion as a potential cellulase were proteins annotated as

endocellulases, exocellulases, beta-glucosidases, and cellulases. The selected genes also included those annotated as predicted proteins, hypothetical proteins, and glycosyl hydrolase family n (where n is the particular GH family), as their annotations do not disqualify the genes as potential cellulases. Of the potential cellulases, 23-28% from CLC and LW samples and 17% from GW were explicitly annotated as one of the three types of cellulases or as "cellulase." The remaining ~72-77% of putative cellulases lacked clear annotations, suggesting that there is substantial carbohydrate active enzyme (CAZyme) novelty in the landfill system, including potentially unannotated cellulases in the twelve GH families examined.

3.3.2 Presence and abundance of cellulase-containing glycosyl hydrolase families

Of the twelve cellulase-containing GH families, eight (GH1, -3, -5, -8, -9, -30, -44, and -116) were present at all sites (Figure 3.2A; Appendix B, Table B1 for full gene count and relative abundance for all GH families across all sites). Genes belonging to GH7, which is composed of exclusively eukaryotic cellulases (Cragg *et al.*, 2015), were not detected at any of the sites prior to filtering out eukaryotic genes. Glycosyl hydrolase families 5, 3, 9, and 1 were the most abundant, representing ~87-92% of hits across all sites. Approximately 74% of all of the predicted GH genes belonged to GH3 and GH5 (Figure 3.2B). No GH family was unique to leachate nor groundwater.

The two CLC metagenomes had the largest number of genes (CLC1_T1 = 6,153 and CLC1_T2 = 4,729) belonging to the glycosyl hydrolase families of interest, where CLC1_T1 had ~2.5 times the number of genes as the highest represented LW (LW2) and the GW metagenome (Figure 3.2A). In the CLC samples, although the first time point (CLC1_T1) had 23% more genes than the later time point (CLC1_T2), the relative proportions of GH families within the two metagenomes showed consistency across the week of sampling, with a maximum of 3.2% change in relative

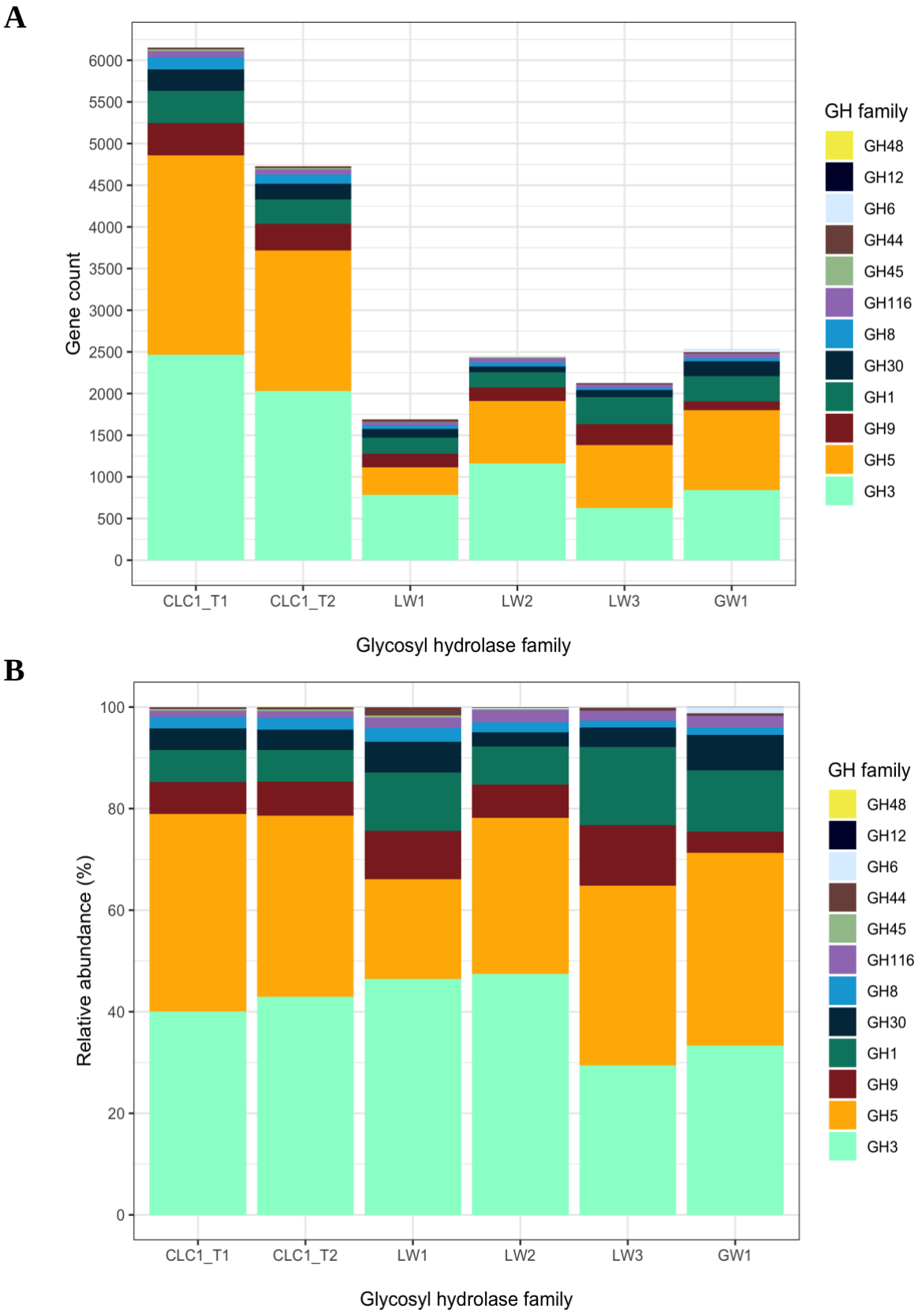


Figure 3.2. Distribution of the total cellulolytic potential by GH families containing cellulases families across the southern Ontario landfill sites (before filtering) by **A)** gene counts and **B)** relative abundance of genes belonging to relevant GH families. Gene counts are total hits, prior to filtering to remove non-cellulase annotations.

abundance across GH families (Figure 3.2B). The consistent GH family proportions in the CLC was unexpected, as the CLC is a reservoir of pooled leachate where leachate from across the landfill is constantly mixed and then discharged at 15.2 L/s to the local wastewater treatment centre (Stantec Consulting Ltd., 2015). Even though the cistern experiences constant changes through inflow and discharge of leachate, very minor change was seen in the overall relative abundance of the GH families. It could also be that the constant mixing of leachate in the CLC could homogenize the pooled leachate, resulting in similar compositions over time. Either way, the data suggests that CLC1, and, by extension the landfill, harbours a stable community of potential cellulolytic microorganisms over the timescale of at least a week.

In contrast to the CLC samples, the relative abundance of GH families amongst the three leachate wells were not as similar. The leachate wells varied both in terms of the total number of relevant GH genes and their GH family composition. The three LWs had 1,600-2,500 GH genes, which represent 0.03-0.06% of the total genes in the metagenomes (Table 3.2). The differences in the proportions of GH families (Figure 3.2) suggest that the range of carbohydrate-hydrolyzing enzymes are quite different between leachate wells in the same landfill at a given time. Sample series from each site would clarify if the observed differences in GH compositions are stable over time.

Most of the cellulase-containing GH families detected in the leachate samples (CLC and LW) have also been found in cellulolytic environments. A number of CAZymes were identified in metagenome-assembled genomes (MAGs) generated from a cellulose-amended landfill leachate microcosm, including GH families associated with lignocellulose degradation (i.e., GH1, -3, -5, -8, -9, -30, -48, -51, -74, and -94, and -116) (Ransom-Jones *et al.*, 2017). Similarly, GH1, -3, -5, -6, -8, -9, -12, -30, -44, -45, -48, -51, and -94 have been identified in cellulolytic environments such

as biogas fermenters, forest soil, and/or cow rumens (Güllert *et al.*, 2016; López-Mondéjar *et al.*, 2016; Wang *et al.*, 2013). To our current knowledge, the genes from GH6 identified in the Ontario landfill are the first from this GH family reported from landfills. The GH6 family contains proteins with characterized endocellulase and exocellulase activities from bacteria and eukaryotes (Lombard *et al.*, 2014).

Lastly, the metagenome from the groundwater well had approximately 2,500 genes from the GH families of interest, which accounted for ~0.08% of the GW annotated genes (Table 3.2). The two most abundant GH families, GH3 and GH5, made up 71%, a distribution between that of the CLC and LWs (Figure 3.2B). One notable difference was the majority of genes identified as GH6 were from the groundwater metagenome (31/42, 74%). The GW metagenome also contained one of two genes from GH12 detected across all sites. There is little investigation on the presence of glycosyl hydrolase families in groundwater, but one study profiled more than 2,000 Candidate Phyla Radiation (CPR) genomes in groundwater wells and detected genes from GH1, -3, -5, -6, -8, -9, -12, -30, -44, and -116 (Danczak *et al.*, 2017). Although the profiling of the CPR genomes revealed a wide range of GH families encompassing potential activities that can process carbon compounds, it is unclear what the proportion of these genes were cellulases.

The most prevalent families across all samples were GH3 and GH5, even though the absolute abundance of glycosyl hydrolases were much greater in the CLCs than in the LWs and GW. Glycosyl hydrolase families -3 and -5 are amongst the largest and most functionally diverse CAZy families, even more so when considered within those associated with cellulose degradation. Beta-glucosidase and exocellulase activities are present in GH3 whereas GH5, formerly "cellulase family A", includes all three types of enzymes needed for hydrolytic breakdown of cellulose (Dodd *et al.*, 2010; Aspeborg *et al.*, 2012; Lombard *et al.*, 2014). Glycosyl hydrolase families -3 and -5

also contain hydrolytic cellulases and other related enzymes such as xylanases, xylosidases, alpha-mannanases, and alpha-L-arabinofuranosidases, which promote the degradation of lignocellulose (Lombard *et al.*, 2014; Pérez *et al.*, 2002; Béguin and Aubert, 1994). At the time of writing (September 2018), the CAZy database had 19,211 bacterial and 104 archaeal proteins in GH3, and 9,990 bacterial and 93 archaeal proteins in GH5 (Lombard *et al.*, 2014). One reason for the abundance of genes belonging to GH3 and GH5 across the different metagenomes may be the sizes of these families and the microbial diversity that they encompass.

Smaller GH families such as GH1, -8, -9, -30, -44, and -116 were present in all samples. The relative abundances of these families also remained quite stable in the CLC samples (maximum of 0.3% difference in GH30) but fluctuated amongst the LW samples (minimum difference was 0.4% in GH116 and the greatest difference was 8% for GH1) (Figure 3.2B). Glycosyl hydrolase family -6 was only detected in LW2, LW3, and GW1; GH12 was detected only in LW2 and GW1, making these two of the rarer GH families in the metagenomes. GH6 contains endocellulases and cellobiohydrolases, whereas GH12 contains only endocellulases (Lombard *et al.*, 2014). The reference sets available for these two families are smaller than those for the other GH families, thus the detection of potential cellulases belonging to these GH families may be lower due to database bias.

3.3.3 Composition of types of cellulases across the sites

The glycosyl hydrolase genes described in the previous section included all genes assigned to the GH families surveyed, regardless if they were annotated as cellulases or not. This encapsulated the total potential cellulases belonging to these families, as is commonly presented in the literature

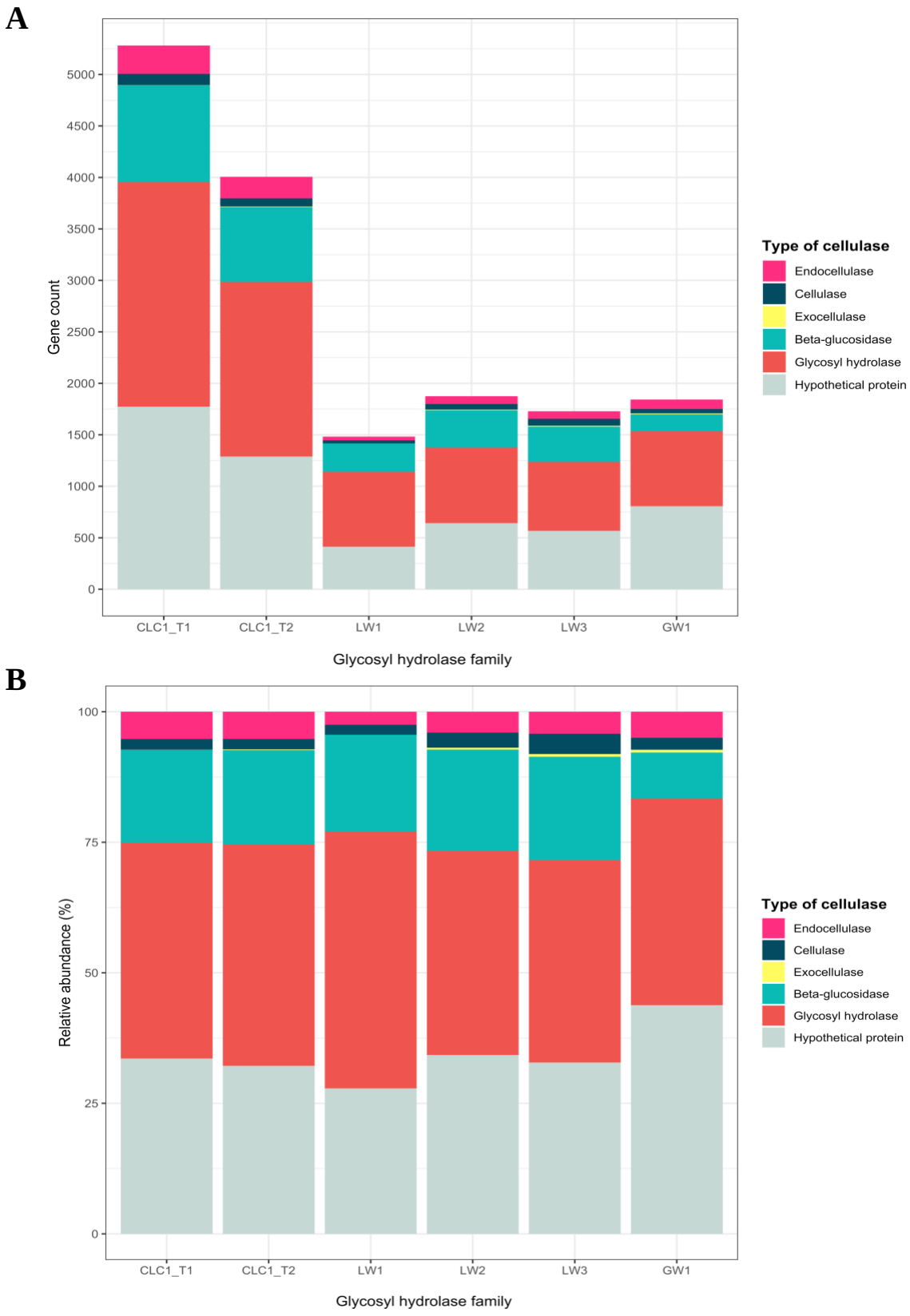


Figure 3.3. Composition of the types of cellulases across the composite leachate cistern, leachate wells, and groundwater well A) by gene count and **B**) by relative abundance. Genes annotated as "cellulase", "glycosyl hydrolase", and "hypothetical protein" are ambiguous in function thus have been included as separate categories.

when potential function is inferred from metagenomes (Danczak *et al.*, 2017; Ransom-Jones *et al.*, 2017). In this section, a more thorough estimate of the total cellulolytic potential of the landfill system is provided. The total hits presented in the previous section were filtered to only include those whose protein annotation was $<1e^{-40}$ by blastp against the RefSeq protein database. The threshold was imposed to ensure we were working with high-quality protein annotations.

The curated potential cellulase genes were then further screened by annotation for one of the three cellulase categories, "cellulases", hypothetical proteins (to account for potential novel cellulases), and "glycosyl hydrolase family n" in order to specifically focus on genes with potential cellulolytic activities. Genes confidently annotated as cellulases in the screened dataset constituted 22%, 20-22%, and 12%, in the CLCs, the LWs, and the GW samples, respectively (Figure 3.3B; see Appendix B, Table B2 for full gene count and relative abundance for all GH families across all sites). The majority of curated genes comprised the hypothetical and vaguely annotated genes, which are potential cellulases (Figure 3.3B). Beta-glucosidases were the most abundant type of unambiguously annotated cellulase, making up 18%, 8-20%, and 9% in the screened genes of CLCs, LWs, and GW, respectively (Figure 3.3B). As beta-glucosidase activity was present in half of the GH families analyzed (GH1, -3, -5, -8, -30, and -116), the predominance of this activity in the cellulase subset may be a result of database bias. Endocellulases made up approximately 5% of both CLC samples, 2-5% in the LWs, and ~5% in the GW sample. Exocellulases occurred at <1% across all sites except in LW1, in which none were detected. Genes annotated as "cellulases" made up 1-4% of each site. It is unclear which specific type of enzyme the annotation "cellulase" refers to, as cellulase is a collective term referring to all three categories of cellulolytic enzymes. To avoid incorrect assumptions, analyses of genes annotated as "cellulase" were not combined

with any specific cellulolytic enzyme but rather kept as a category on their own, contributing to the collective cellulolytic potential of the site.

The composition of cellulase types generally shows remarkable consistency of relative abundances between the CLC time points, amongst the spatially distributed LWs, and between the CLC and LWs (Figure 3.3B). Comparing the CLC metagenomes, the relative abundances of all cellulase types were <1.1% different. The cellulase composition of LW2 and LW3 was the most similar among all LWs with a difference of <1.5%. The consistency in relative abundance of the cellulases seen across the CLC and LW sites suggests that there exists a stable suite of cellulase functionality that may degrade the cellulosic material in the landfill.

In contrast, there were a couple of differences in the proportional abundance of the cellulases between the leachate samples and the groundwater sample. The lower proportion of confidently annotated cellulases in the GW sample (12%, ~10% lower than other samples) was balanced by an increase in hypothetical proteins, which made up 44% of the cellulase subset in GW, 9-16% more than those in the CLC and LW. This signifies that there are more unknown carbohydrate-modifying activities and genes in the groundwater community, which may or may not be cellulolytic. It is uncertain whether the CAZymes detected in GW are abundant due to contamination from the landfill or from an alternative carbohydrate or cellulose source outside of the landfill. The GW well surveyed is impacted by landfill leachate, but the cellulose concentration moving with the leachate is not documented.

3.3.4 Microbial composition of potential cellulases

The microbial compositions of the cellulase-specific datasets were analyzed to identify the taxonomy and trends in the predicted cellulolytic bacteria and archaea in the leachate and

groundwater. Across all sites, Bacteroidetes was the most abundant phylum, accounting for, on average, ~47% in the CLC samples, ~36% in the LWs, and 44% in the GW (Figure 3.4; see Appendix B, Table B3 for full gene count and relative abundance for all taxonomic groups across all sites). Bacteroidetes have been found at high abundance in many cellulolytic environments, such as cellulose microcosms from landfill leachate, cow rumens, elephant feces, and biogas fermenters (Ransom-Jones *et al.*, 2017; Song *et al.*, 2015a; Güllert *et al.*, 2016; Lo *et al.*, 2010). It was also not surprising that Firmicutes was an abundant group, representing ~23% in CLC samples, 10-21% in the LWs, and ~7% in GW of the curated cellulase genes. Firmicutes have been detected at high abundance in landfills and other cellulolytic environments through both 16S rRNA gene sequencing and metagenomics (Song *et al.*, 2015a, 2015b; Stamps *et al.*, 2016; Güllert *et al.*, 2016; Ransom-Jones *et al.*, 2017). Culture-based studies also identified *Bacillus* and *Paenibacillus*, both Firmicutes, as major cellulolytic groups in landfills (Pourcher *et al.*, 2001). In the curated sets of cellulases, the other phyla and proteobacterial classes that were present at >1% at all sites included Verrucomicrobia, Spirochaetes, Actinobacteria, and Gammaproteobacteria. These phyla have also been detected at relatively low abundance in landfills, cellulolytic microcosms from landfill leachate, and cellulolytic microcosms from landfill leachate biogas fermenters, (Song *et al.*, 2015b; Stamps *et al.*, 2016; Huang *et al.*, 2004; Ransom-Jones *et al.*, 2017; Güllert *et al.*, 2016). Phyla that occurred at <1% at all sites were categorized as rare phyla. The pooled abundance of rare phyla was 5-7% across all sites, consisting of 22 bacteria phyla, 2 archaeal phyla, and several unclassified bacteria. Archaeal cellulases affiliated with the Euryarchaeota were detected from all sites, with crenarchaeotal genes only detected at CLC1_T2, LW2, and LW3. Archaea were recovered at low abundance (0.8-4%) in a previous study, which examined 19 landfills using 16S rRNA gene sequencing (Stamps *et al.*, 2016).

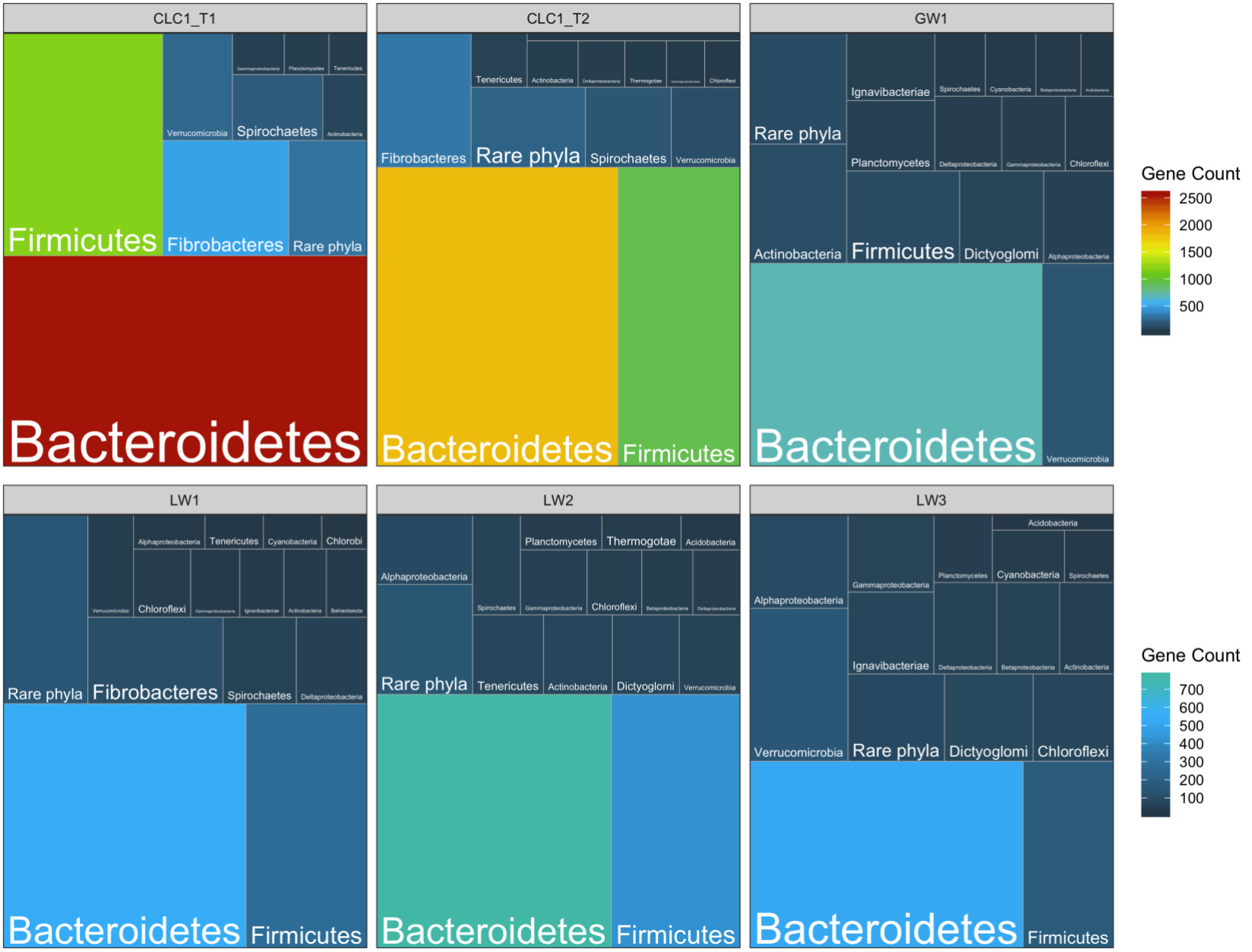


Figure 3.4. Taxonomic composition of predicted cellulases across the composite leachate cistern, leachate wells, and groundwater well. The size of the box with a taxon name indicates the taxon's relative abundance and its colour represents its absolute abundance of the genes at a particular site. The unlabelled box in CLC1_T2 is Epsilonproteobacteria. The Rare phyla category consists of the highest-level of identifiable taxonomic group and proteobacterial classes that occurred at <1% at each site. A phylum that has a designated box (present at >1%) in one site and not in others means it occurred at <1% in the other sites (not necessarily absent nor undetectable).

From a taxonomic perspective, the CLC samples were largely consistent, with a maximum phylum abundance difference of 2.8% seen in the Bacteroidetes. Beyond the Bacteroidetes and Firmicutes, the LW curated cellulases were assigned to 13 phyla at less than 10% abundance, whereas the GW curated cellulases were assigned to 14 phyla with abundances between 1-10%. Fibrobacteres had a varied distribution across the landfill and was present as one of the top three most abundant phyla in the CLC samples (8-10%) and LW1 (8%) but was at less than 1% in all other samples. All potential cellulases detected belonging to GH45 were affiliated with Fibrobacteres, in contrast to the GH45 family in the CAZy database, which, though dominated by Eukaryotic sequences, contains an even mix of Fibrobacteres and Proteobacteria sequences. Fibrobacteres have been implicated as contributors to bovine rumen cellulolytic communities, as *Fibrobacter succinogenes* is a known cellulose degrader and is the most commonly isolated Fibrobacteres species from bovine rumen (Henderson *et al.*, 2015; Suen *et al.*, 2011; Russell *et al.*, 2009). Fibrobacteres have been observed to decrease in abundance in cattle rumen when protozoa were removed. Fibrobacteres have also been detected in landfills (Song *et al.*, 2015b; Ransom-Jones *et al.*, 2017). A Fibrobacteres MAG from a landfill leachate microcosm contained potential cellulases belonging to GH5, -9, -12, and, exclusive to this MAG, GH45 (Ransom-Jones *et al.*, 2017). GH45 proteins have been found in all representatives of the Fibrobacteres phylum except *Chitinivibro alkaliphilus* (Abdul Rahman *et al.*, 2016). Fibrobacteres may be degrading cellulose in the CLC and the LW1 sites but their functions require confirmation.

Several phyla identified here were not previously associated with cellulose degradation at landfills. Dictyglomi cellulase genes were detected at ~3-5% in LW2, LW3, and GW1, and at <1% at all other sites (Figure 3.4). Members of the Dictyoglomi are thermophilic, anaerobic bacteria (Saiki *et al.*, 1985). Both *Dictyloglomi turgidum* and *Dictyloglomi thermophilum* encode endoglucanases

used to break down carboxymethylcellulose (CMC) (Brumm *et al.*, 2011; Shi *et al.*, 2013) and *D. thermophilum* has been shown to use cellobiose as well (Saiki *et al.*, 1985). Additionally unexpected were the Ignavibacteria, whose cellulase genes occurred at ~2-4% in LW1, LW3, and GW1, and at <1% at all other sites. Iganvibacteria have not previously been described at municipal landfills. Ignavibacteria are generally facultatively anaerobic thermophilic bacteria related to Bacteroidetes and Chlorobi. Within the Ignavibacteria, a strain of *Melioribacter roseus* is able to use microcrystalline cellulose and CMC (Podosokorskaya *et al.*, 2013b).

3.4 Conclusions

Glycosyl hydrolase families 5, 3, 9, and 1 were the most prevalent, making up 87-93% of detected GHs. Within this, GH3 and GH5 accounted for 64-79% of all glycosyl hydrolases detected across all sites. Beta-glucosidases were the most abundant type of unambiguously annotated cellulases. The relative abundances of the different types of potential cellulases were generally consistent between the two CLC samples, suggesting a stable presence of cellulases in the CLC. In contrast, relative abundances of the larger GH families, such as GH3 and GH5 varied across the three leachate wells in the landfill at a given time, suggesting differences in the range of carbon-processing potential in leachate wells across the sampled landfill area. This speaks to the heterogeneity of the waste composition in landfills that may enrich for activities belonging to different CAZyme protein families. A large proportion of the putative cellulases were annotated as hypothetical proteins or ambiguously annotated genes. These represent a considerable amount of currently unknown carbohydrate-modifying function and diversity across the five sites. Potential cellulases affiliated with bacteria belonging to Bacteroidetes, Firmicutes, Verrucomicrobia, and Spirochaetes, were most abundant within the CLC and LW metagenomes'

annotated genes. Dictyoglomi and Ignavibacteria-associated putative cellulases were observed at greater than 1% abundance in two LW sites and the GW site in the Ontario landfill. Dictyoglomi and Ignavibacteria have not been previously reported in landfill leachate studies; our work suggests they may be contributors to cellulose degradation or carbohydrate-active function in landfills.

A generally low proportion of cellulase genes was detected in the LW metagenomes, well below 1% of all annotated genes for any of the metagenomes. The low numbers may be due to two factors. First, the LW sites sampled are in an inactive part of the landfill that stopped receiving waste in 2001. In a study comparing older (5 years) and newer (1 year) landfill sections, a lower absolute number of cellulolytic bacteria was reported in the older parts of the landfill, which could be attributed to decreased levels of cellulose waste (Pourcher *et al.*, 2001). Cellulolytic bacteria are predicted to be most active at the beginning of anaerobic digestion of landfill waste, converting waste biomass to sugars and acids (Lynd *et al.*, 2002; Li *et al.*, 2009b). Second, only cellulases belonging to glycosyl hydrolase families were surveyed in the Ontario landfill, but profiling other protein families associated with cellulose degradation could increase the gene counts and therefore increase the predicted cellulose degradation potential.

There are several potential additional future directions for this research. Although HMM profiling of GH families based on conserved domains is a way to identify potential GHs, the sequences used to construct the HMM profiles are composed of both cellulases and non-cellulases for many of the relevant GH families (Sukharnikov *et al.*, 2011). This is because the GH families are categorized based on sequence similarity, not substrate specificity nor enzyme activity. As suggested by Sukharnikov *et al.* (2011), generation of new HMMs consisting of cellulase sequences, making use of conserved domains and protein folds for each GH family, would be an improved approach for cellulase identification. Profiling cellulases in the leachate and

groundwater can be expanded to include other types of cellulases and protein structures involved with cellulose degradation. Cellulose phosphorylases, which use phosphate instead of water to break down cellulose, can be surveyed. Other informative targets include carbohydrate-binding modules (CBMs), scaffoldin, and dockerin structures associated with cellulosomes (Ohmiya *et al.*, 1997). Carbohydrate-binding modules in cellulases are believed to increase hydrolytic activity by increasing the surface area exposed to cellulose (Karita *et al.*, 1996). Scaffoldin proteins anchoring CBMs and dockerins can form cellulosomes, complexes that exists on the cell surface of anaerobic, cellulolytic bacteria and degrade cellulose in close proximity to the cell (Várnai *et al.*, 2013). Also, polysaccharide-utilization loci (PUL) could be mined; PULs have been specifically studied in Bacteroidetes, are co-localized and co-regulated genes involved in detection, take up, catabolism, and import of complex carbohydrates (Cragg *et al.*, 2015; Grondin *et al.*, 2017). Lastly, because of the temporal consistency seen in the CLC, it may be better to take more spatial samples in different individual leachate wells when investigating for more novel cellulases.

Chapter 4: Cellulose degradation capacity and microbial diversity of cellulose- and paper- enriched cultures

4.1 Introduction

It is estimated that 1% or less of bacteria and archaea in the environment can be cultured and isolated in the lab using current techniques (Amann *et al.*, 1990). As sequencing technologies have been steadily improving and becoming more economically feasible, many researchers have instead turned to 16S rRNA gene amplicon sequencing and metagenomics to gain insight into the microbial functional potential in the environment. However, prior to the advent of high-throughput sequencing, there were several studies that investigated the cellulolytic capabilities of isolates cultured from landfill waste or leachate (Bagnara *et al.*, 1985; Westlake *et al.*, 1995; Pourcher *et al.*, 2001; Shiratori *et al.*, 2006). Isolates from municipal waste sites have been characterized to confirm the functions and microbial populations predicted from metagenomes and 16S rRNA gene analyses (Korpole *et al.*, 2011; Ransom-Jones *et al.*, 2017). There has not yet been a study connecting the potential cellulolytic activities detected in landfill metagenomes through isolation and characterization of isolates from the same landfill. My research links the cellulolytic potential predicted by metagenome analyses of a landfill with cellulolytic activity detected through enrichment of the microbial populations in the landfill leachate via enrichment culture and isolations.

The objective of this chapter was to enrich for cellulolytic microorganisms from the leachate in CLC1, a composite leachate cistern at the southern Ontario landfill, identify the microbial communities in the enrichments, and test for their ability to degrade cellulose.

4.2 Materials and Methods

4.2.1 Sampling sites and sample collection

Biomass collected from CLC1 at the Ontario landfill on July 14th, 2016 (CLC1_T1) was used as inoculum for the cellulose-degrading enrichment cultures. Please refer to section 2.2.1 for site and sample collection information.

4.2.2 Developing synthetic leachate recipe

The composition of the synthetic leachate (Table 4.1) was based on the protocol developed by Rowe and colleagues (2002), and adjusted to reflect the chemical composition of the landfill leachate as reported by the Southern Ontario landfill's site engineers in the 2014 annual report (see Table C1 for a comparison between the synthetic leachate composition developed for this research and that of Rowe *et al.*, 2002). Each solution was made and autoclaved separately. Solution 1 was autoclaved at 95°C for 15 min and all other solutions were autoclaved at 121°C for 20 min. When making synthetic leachate medium amended with microcrystalline cellulose (average particle size = 90 µm), Solutions 1-5 were combined. When making synthetic leachate medium without microcrystalline cellulose, Solutions 2-5 were combined. When making synthetic leachate agar medium with microcrystalline cellulose or without cellulose, Solutions 1-6 or Solutions 2-6 were combined, respectively. All synthetic leachate media without agar were combined at room temperature and the pH adjusted to 6.6-6.8 using 0.22 µm filter-sterilized 2.5 M hydrochloric acid while stirring. Media components for media with agar were added while the agar was still molten, with slow stirring.

Table 4.1. Compositions of synthetic leachate and metal stock solution

	Amount	Unit (/L)	dH_2O
Solution 1*			*
Microcrystalline cellulose	10	g	
Solution 2			200 ml
$CaCl_2$ x $2H_2O$	870	mg	
Solution 3			200 ml
$MgSO_4$	54	mg	
$MgCl_2$ x $6H_2O$	1,083	mg	
Solution 4			100 ml
K_2HPO_4	30	mg	
Solution 5			500 ml
$KHCO_3$	312	mg	
K_2CO_3	324	mg	
NaCl	745	mg	
$NaHCO_3$	1,558	mg	
$NaNO_3$	26	mg	
NH_4HCO_3	1,430	mg	
$CO(NH_2)_2$ (urea)	407	mg	
Solution 7 - Metal Stock Solution	1	ml	
Solution 6**			* *
Agar	15	g	
Solution 7 - Metal Stock Solution			top up to 1 L
$Al_2(SO_4)_3$ x 16 H_2O	30	mg	
$CoSO_4$ x 5 H_2O	150	mg	
$CuSO_4$ x 5 H_2O	40	mg	
$FeSO_4$	4,000	mg	
H_3BO_3	19,446	mg	
$MnSO_4$ x H_2O	2,453	mg	
$(NH_4)_6Mo_7O_{24}$ x 4 H_2O	50	mg	
$NiSO_4$ x 6 H_2O	500	mg	
$ZnSO_4$ x 7 H_2O	50	mg	
96% concentrated H_2SO4	2.3	ml	

* Only added when making synthetic leachate amended with microcrystalline cellulose. Autoclave Solution 1 separately and combine with Solutions 2-6 at room temperature. Adjust total volumes of Solutions 1-6 to total 1 L.

** Only added when making synthetic leachate agar media or synthetic leachate with microcrystalline cellulose agar media. Adjust total volume of synthetic leachate to 1 L with the amendment.

Adjusted pH to ~6.6-6.8 with HCl

4.2.3 Enriching for cellulose degraders on microcrystalline cellulose

Approximately half of one 0.2 µm filter was diced into little pieces, and the biomass was then re-suspended in 10 ml of synthetic leachate, and vortexed thoroughly at high speed in order to maximize biomass resuspension off of the filter. The S_0 microcrystalline cultures (MCC S_0) contained 5 ml of the resuspended biomass and 95 ml of synthetic leachate amended with microcrystalline cellulose (MCC) in two clear 120 ml glass serum bottles capped with butyl stoppers. One bottle of MCC S_0 culture was incubated in the dark at 25°C with shaking at 170 RPM. The other bottle was maintained in the dark at room temperature without shaking. A negative control containing only synthetic leachate with MCC and no biomass was incubated under both conditions. After two weeks, the MCC S_0 cultures were transferred at 1:10 dilution into synthetic leachate amended with MCC for a total of 100 ml, in duplicate (MCC S_1 cultures), which were incubated at 25°C and at room temperature. MCC S_1 bottles were subcultured into MCC S_2 in the same way after two weeks of incubation. Subculturing of MCC in the same way was continued up until S_5. All cultures were batch-refreshed every two weeks by removing 15 ml of spent culture and adding 15 ml of fresh synthetic leachate amended with MCC. A negative control for each sub-culture was incubated in parallel, containing the same synthetic leachate used as diluent but without any MCC.

4.2.4 Enriching for cellulose degraders on paper substrates

Copy paper (CP), cardboard (CB, Thermo Fisher box), inkless newsprint (NP, sourced from the Imprint campus newspaper), and filter paper (FP, Whatman filter paper No. 1, Whatman, Dassel, Germany) were cut into 1 x 6 cm strips, individually wrapped in aluminum foil, and autoclaved. A paper degradation trial was designed as follows: i) paper controls contained one strip of paper

and 10 ml of synthetic leachate in a test tube (in duplicate); ii) paper-free controls (CTRL) contained 9 ml of synthetic leachate in a test tube and 1 ml of a biomass suspension from MCC S_2 (in duplicate); and iii) test cultures contained a strip of one paper (CP, CB, FP, or NP) combined with 9 ml of synthetic leachate and 1 ml of MCC S_2 culture in a test tube (in triplicate). Test tubes were capped with plastic test tube caps.

Table 4.2. Composition and source of inoculum in samples. Incubation times indicate length of enrichment prior to DNA extraction for a given culture. S = Subculture, MCC = microcrystalline cellulose, CTRL = paper-free control, CP = copy paper, CB = cardboard, FP = filter paper, NP = newsprint.

Sample	Biomass inoculum	Incubation time of sample before subculture(weeks)	Time at inoculation of sample (total weeks)
Microcrystalline cellulose enrichment cultures			
S0 MCC	biomass from leachate	2	0
S1 MCC	S0 MCC	2	2
S2 MCC	S1 MCC	2	4
S3 MCC*	S2 MCC	2	6
S4 MCC*	S3 MCC	2	8
S5 MCC*	S4 MCC	2	10
Paper enrichment cultures			
S0 CTRL	S2 MCC	5	0
S0 CP	S2 MCC	5	0
S0 CB	S2 MCC	5	0
S0 FP	S2 MCC	5	0
S0 NP	S2 MCC	5	0
S1 CTRL	S0 CTRL	13	11
S1 CP	S0 CP	13	11
S1 CB	S0 CB	13	11
S1 FP	S0 FP	13	11
S1 NP	S0 NP	13	11
S2 CTRL	S1 CTRL	8	24
S2 CP	S1 CP	8	24
S2 CB	S1 CB	8	24
S2 FP	S1 FP	8	24
S2 NP	S1 NP	8	24

* this sample was not sequenced

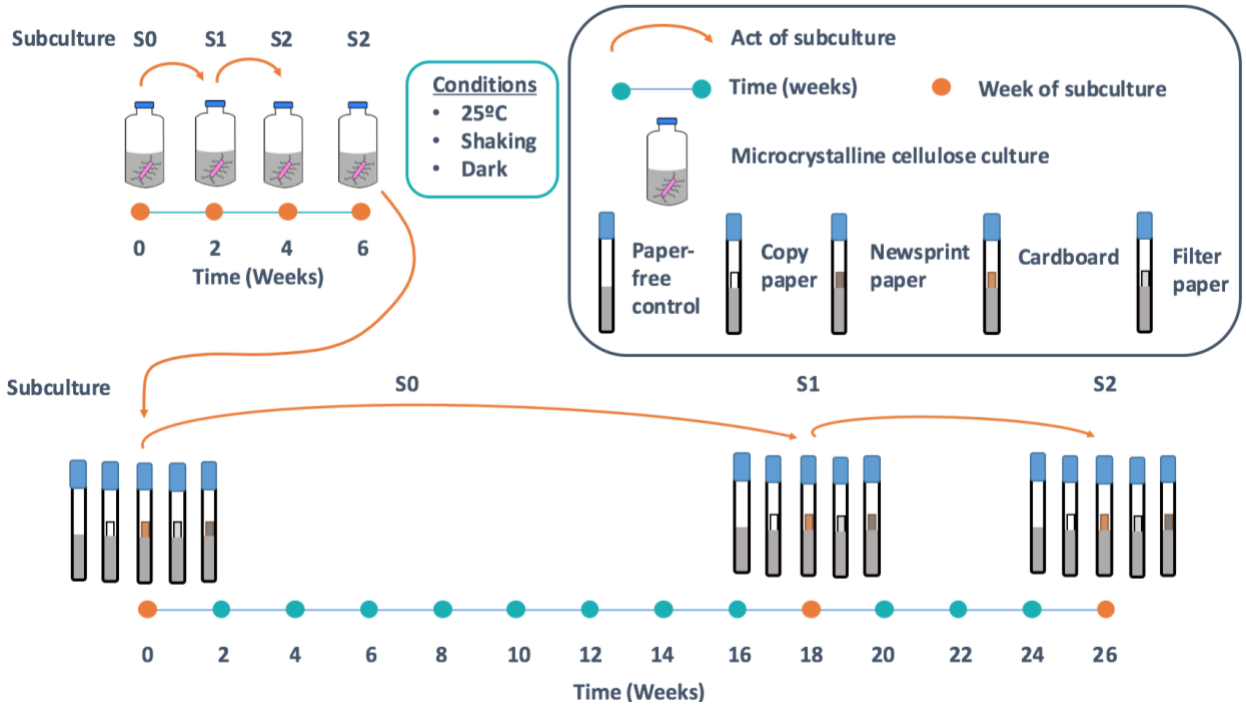


Figure 4.1. Timeline of MCC and paper subcultures. Microcrystalline cultures were done in duplicate except S_0. A media control without biomass was prepared fresh with each MCC subculture. Paper culture series were done in triplicate for each subculture and its paper-free control was done in duplicate. DNA was extracted at each subculture time point. The paper control and the media control are not shown.

Media controls in test tubes contained 10 ml of synthetic leachate and no paper nor biomass (in duplicate). All cultures were incubated in the dark at 37°C with shaking at 170 rpm. Cultures were refreshed every 5-7 days by removing 1 ml of spent culture and adding 1 ml of fresh synthetic leachate medium. All paper cultures were sub-cultured to S_1 and then to S_2 subcultures when the copy paper and filter paper trials showed physical degradation (Figure 4.1, Table 4.2).

4.2.5 Detecting cellulose degrading isolates from the MCC enrichment culture

4.2.5.1 Testing for carboxymethylcellulose degradation

Presence of endocellulase activity in the MCC enrichment culture was tested by using MCC S_5 bottles #1 and #2, subcultures that had been incubated for 59 weeks at 25°C with shaking with

periodic media refreshment. A modified version of carboxymethyl cellulose (CMC) (w/v) agar medium (American Type Culture Collection (ATCC) medium 2720 protocol (American Type Culture Collection)) (Table 4.3) was used to test isolates for endocellulase activity, where $FeCl_3$ was omitted and casitone was replaced by Trypticase Peptone (Catalog No. 211921, BD Biosciences, Mississauga, ON). The medium was autoclaved at 115°C for 15 mins and K_2HPO_4 was added right before plates were poured.

A loopful of MCC S_5 liquid culture from bottle #1 and #2 were each used to make one straight streak on CMC agar plate, which was incubated at 30°C for one week to encourage faster colony growth. The plate was stained with 0.1% Congo red solution for 15 min, then the dye was poured off before adding 1 M NaCl for 15 min. Carboxymethyl cellulose hydrolysis was visualized by chromogenic clearing on the plate. Clearing zones, where there was little to no Congo red stain, indicate zones of cellulose hydrolysis (Teather and Wood, 1982).

To isolate potential cellulose-degrading colonies, two classical streak plates were made using MCC S_5 bottle #1 only (as the previous Congo red staining identified more cellulose degradation activity in bottle #1 than #2). These streak plates were incubated at 30°C and 45°C until colonies formed. Eight colonies of different morphologies were chosen from the streak plates on CMC and further purified by restreaking on fresh CMC agar. Once isolated colonies were obtained, a colony was picked and streaked back and forth five times on fresh CMC agar. Plates were incubated at 30°C and 45°C for seven days. *Bacillus subtilis* (ATCC 6633), a known endocellulose degrader, was used as a positive control. After seven days, the CMC agar was stained with Congo red as before. Endocellulase activity of isolates was assessed through CMC hydrolysis by computing the ratios of the widths of the hydrolysis zones to the widths of the growth streaks. Isolates were purified on nutrient agar to phenotypically characterize cell morphologies and Gram-reaction.

Table 4.3 Carboxymethyl cellulose medium. The medium composition was modified from ATCC medium 2720.

Component	Mass (g) / Volume (ml)
$(NH_4)_2SO_4$	1.0 g
$MgSO_4$ x $7H_2O$	1.0 g
$CaCl_2$ X $2H_2O$	1.0 g
K_2HPO_4 (filtered)	1.0 g in 50 ml of distilled water
Peptone	2.0 g
Carboxymethyl cellulose	15.0 g
Agar	15.0 g
Distilled water	950 ml

4.2.5.2 Testing for microcrystalline cellulose degradation

Exocellulase activity was tested by the degradation of microcrystalline cellulose. The isolates grown on CMC were streaked on the ATCC 2720 medium with microcrystalline cellulose substituted as the cellulose source and supplemented with 0.2% Congo-Red in the medium. Plates were incubated at 30℃ and 45℃ for eleven days. Degradation was visualized by chromogenic clearing of Congo red.

4.2.5.3 Testing for cellobiose degradation

Beta-glucosidase activity of isolates was tested on minimal medium amended with cellobiose. The isolates were streaked on modified ATCC Medium: 2511 M9 Minimal agar (American Type Culture Collection) and on M9 where the glucose was replaced with 1% D(+)- cellobiose (w/v) as the sole carbon source. The ATCC Medium 2511 M9 minimal agar without glucose was used as a control medium. Both cultures were incubated at 30℃ and 45℃ for seven days, after which growth on the two media was compared. Cells able to grow on cellobiose as the sole carbon source would have enhanced growth compared to those that cannot. To quantitatively assess beta-glucosidase activity, the ratio of the widths of the growth streak on M9 amended with cellobiose

to that of the M9 control medium was calculated. Enhanced growth on cellobiose signifies beta-glucosidase activity. *Bacillus subtilis* (ATCC 6633) was used as a positive control.

4.2.6 16S rRNA gene amplicon sequencing on time series cultures

The MCC S_0-S_2 cultures and the paper cultures S_0-S_2 (Table 4.2) maintained over ~7.5 months were selected for 16S rRNA gene amplicon sequencing to analyze microbial population enrichment dynamics over time on various cellulose substrates.

4.2.6.1 DNA extraction

A 1 ml aliquot of each culture was centrifuged at 13.5 rpm for 2 minutes. Cell pellets were stored at -20°C until extraction. An extraction kit control with no added biomass was carried out alongside the samples. Genomic DNA was extracted using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) as per manufacturer's instructions except the pellets were re-suspended in the bead-beating fluid pipetted out of the bead-beating tubes and then the suspensions were transferred back to their respective bead-beating tubes, in place of the suggested soil samples. Bead beating was done using a MO BIO Vortex Adapter (MO BIO). The elution volume for DNA was 50 µl in kit solution S6 (10 mM Tris). Extracted DNA was stored at -20°C.

4.2.6.2 DNA amplification, purification, and quantification

The variable region V4-V5 of the 16S rRNA gene was amplified using the primers 515F (5'GTGYCAGCMGCCGCGGTAA-3') and 926R (3'-CCGYCAATTYMTTTRAGTTT-5') (Parada et al. 2016) using *Taq* DNA polymerase (New England Biolabs Ltd., Whitby, ON). The PCR amplification was done on the Bio-Rad T100™ Thermal Cycler (Bio-Rad Laboratories,

Hercules, CA) using the following protocol modified from Bartram *et al.'s* (2011) protocol: initial denaturation at 95˚C for 3 min, 35 cycles of 95˚C for 30 s, 50˚C for 30 s, 68˚C for 1 min, and the last extension at 68˚C for 7 min. The primers included 5' Illumina sequencing adapters composed of a unique barcode corresponding to each sample for multiplexing and the Illumina adapter sequences that bind to the sequencing primers. For this analysis, we used two forward primers with different barcodes, 515fi14 and 515fi6, and a suite of 32 uniquely barcoded reverse primers. The composition of the PCR master mix per reaction consisted of: 2.5 µl of 10X ThermoPol Buffer buffer, 1.5 µl of 10 mg/ml Bovine Serum Albumin, 0.05 µl of 100 µM V4 forward primer, 1 µl of 5 µM V5 reverse primer, 0.05 µl of 100 mM dNTPs, 0.125 µl of *Taq* DNA polymerase (0.6 units/µl), 1 µl of 1-10 ng of template DNA (1-10 ng), and 18.775 µl of nuclease-free molecular grade water (UV-treated). The water was UV-treated for 20 min. The final reaction volume was 25 µl. The PCR amplification for each sample was done in triplicate to lower PCR bias and reduce contamination effects. The amplification included four non-template controls (containing the master mix and enzyme but without template DNA). Triplicate amplicon reactions from each sample were pooled and the concentrations were quantified by 1% gel electrophoresis (with 1X TAE buffer). The 1 kb Sigma DirectLoad DNA ladder (10 µl loaded) (Sigma Aldrich Oakville, ON, Canada) was used as reference for DNA amplicon size. The DNA was visualized using GelRed (Biotium, Hayward, CA) on the AlphaImager HP (Bio-Techne, Minneapolis, MN) and the accompanying AlphaView software was used to quantify the DNA concentration of the bands. A stock PCR master mix was made without the uniquely barcoded reverse primers and then combined with the individually pipetted reverse primers and DNA templates in a 96-well plate. Amplicons from each sample were pooled at equal concentrations. The pooled library was spiked with a positive control of a 1ng/µl mixture of *Allivibrio fischeri*'s and *Thermus thermophiles*'s

341F – 926R fragment of the 16S rRNA gene cloned into vectors at a 1:1 ratio. Amplicons were run on a 1% agarose gel alongside 50 ng of a 1 Kb Plus DNA ladder (Invitrogen, Carlsbad, CA) and visualized using ethidium bromide. The pooled sequencing library was excised from the gel using a razor blade and purified using the New England Biolabs Monarch DNA Gel Extraction Kit (New England Biolabs Ltd., Whitby, ON) with the following modifications to the protocol to accommodate for the larger size of the gel slice (~2.67 g): i) 10.7 ml of gel dissolving buffer was added to the gel in a 15 ml centrifuge tube and the gel dissolved as indicated in the manufacturer's protocol; ii) two spin columns were used, with each processing half the volume of the dissolved DNA gel solution; and iii) the final product was eluted in a total of 20 µl of elution buffer. The purified amplicon library was quantified using a Qubit dsDNA HS Assay kit on the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). A portion of the library was diluted to 6 nM in 10 nM Tris and 0.05% Tween 20, pH 8. The concentration of the diluted amplicon library was verified using Qubit and 1% agarose gel quantification.

4.2.6.3 Sequencing

The MiSeq (Illumina, Hayward, CA) platform was used to sequence the pooled library of amplicons from all samples and controls. The MiSeq v2 Reagent Kit for 500 cycles (2 x 250 bp) (Illumina, Hayward, CA) was used per manufacturer's instructions. The library had an average final template size of 411 base pairs. The library was also spiked with 10% PhiX for run quality control and to increase library diversity in the case of low-diversity libraries. The library and PhiX were denatured using 0.2 N sodium hydroxide, diluted to 5 pM, and combined in a 10:1 ratio for a total volume of 600 µl. The library pool was loaded in one lane of a MiSeq flow cell. After sequencing, indexed reads for each sample were demultiplexed by the MiSeq Reporter Generate

FASTQ software (Illumina). Read metrics and data quality were analyzed using the online interface BaseSpace Sequence Hub (https://basespace.illumina.com).

After demultiplexing, ~65% of the clusters were not assigned to a sample. Unassigned clusters are due to barcode mismatches between the sequenced clusters and the assigned barcodes (known primer sequences). In an attempt to increase the number of assigned clusters, the samples were demultiplexed again with modification of the 6-base barcode sequence where the first nucleotide was replaced with an "N" for all of the forward and reverse primers.

4.2.7 Bioinformatic pipeline and sequence analyses

Quality control, denoising, merging of paired end reads, and training of a classifier to assign taxonomy to sequences were done by using plugins in QIIME2 v. 2018.2 (Caporaso *et al.*, 2010). Bases were trimmed at base position 25 of the forward read and at position 23 of the reverse reads. Reads were then denoise and merged using the DADA2 (Callahan *et al.*, 2016) plugin. Taxonomic assignments for the individual features in Qiime2 were generated by first training a Naive Bayes classifier on the SILVA (release 132) 99% OTUs reference dataset (Quast *et al.*, 2012) using the q2-feature-classifier plugin, after which the classifier was applied to the sample dataset. Individual features were collapsed at the species level (level 7) at the taxonomic analysis step of Qiime2 and species-level taxonomic abundances were summed. From this point forward, the term "ESV" (exact sequence variant) is used to represent a taxonomic level at which individual features sharing the same taxa at the species level have been collapsed. The ESVs that were present only in the positive controls (*A. fischeri* and *T. thermophiles*), the non-template controls, and extraction controls were removed from further analysis. Samples that had a lower number of reads than any non-template control or extraction control were also removed from analysis, which included one

replicate of copy paper S_2 and one replicate of the paper-free control S_1. Taxonomy and abundance patterns of the most abundant ESVs in each paper trial were analyzed and visualized by bar plots and line plots using ggplot2 (Wickham, 2016), and heat maps using the Phyloseq v. 1.24.2 (McMurdie and Holmes, 2013) in RStudio (RStudio Team, 2015). The core microbiome of each culture was determined as follows: i) the relative abundance of an ESV across all replicates in a paper trial were summed and the ESVs were ranked from greatest to least abundance; ii) a cut-off threshold was applied at which the last ESV ranked that was included in the core microbiome was >1% relative abundance in any S2 replicate. Any ESVs whose relative abundance was <1% in all replicates and all subcultures but were above this cut-off threshold were removed from the core microbiome. To investigate enrichment of the core microbiota from the MCC S_2 (T=0 inoculum) over the 26-week enrichment trial on different papers, ESVs whose relative abundance were >5% in any replicate from any culture were screened for an upward trend in relative abundance by 18 weeks (S_1) in any one of the paper enrichments.

4.3 Results and Discussion

4.3.1 Cellulolytic isolates from the microcrystalline cellulose culture

Biomass filtered from the CLC1_T1 leachate was cultured in synthetic leachate enriched with microcrystalline cellulose (MCC) (S_0) at 25-27°C. Subcultures were set up using a 1:10 dilution in fresh MCC enrichment medium every two weeks up to S_5. The subcultures were additionally refreshed every two weeks by removing 15% of spent culture volume and replacing with fresh media. The MCC S_5 enrichment at 59 weeks was streaked onto carboxymethyl cellulose (CMC) medium to isolate cellulose-degrading colonies from the mixed culture and incubated at 30°C and 45°C for one week to test the growth capacity and cellulase production at different temperatures.

Eight phenotypically different isolates (Table 4.4) were randomly chosen from the MCC S_5 mixed culture grown on CMC medium and were purified on the same medium.

Irrespective of incubation temperature, six colonies (I1, I2, I3, I4, I6, I7) demonstrated CMC hydrolysis (ratio >1) (Table 4.5; Figure 4.2A). At 30°C and 45°C, four isolates (I1, I3, I4, and I6)

Table 4.4. Phenotypic characteristics of isolates.

Isolate	Gram-reaction	Cellular morphology
I1	Gram-negative	rod
I2	Gram-negative	coccobacillus
I3	Gram-negative	rod
I4	Gram-negative	rod
I5	Gram-negative	coccobacillus
I6	Gram-negative	rod
I7	Gram-negative	rod
I8	Gram-negative	rod

Table 4.5. Assessment of CMC and cellobiose hydrolysis by isolates. For carboxymethyl cellulose, the ratios of the widths (mm) of hydrolysis zones to the widths (mm) of the growth streaks are shown. For cellobiose, the ratio of the widths of the growth streak on M9 amended with cellobiose to that of M9 (control) was computed (Cellobiose:Control ratio).

Medium	Temp-erature (°C)	Ratios							
		I1	I2	I3	I4	I5	I6	I7	I8
CMC	30	3.7	1.1	4.3	5.7	0.0	6.0	1.2	1.0
CMC	45	3.8	0.0	9.7	14.5	1.0	6.0	0.0	1.0
Cellobiose: Control Ratio	30	0.0	0.8	0.2	0.0	1.7	0.8	1.7	1.7
Cellobiose: Control Ratio	45	0.0	0.6	7.0	1.0	1.3	1.3	1.4	2.0

demonstrated strong CMC hydrolysis (ratio: >1.5). The endocellulases secreted by these organisms are active at mesophilic and at the low end of thermophilic microorganisms' growth temperatures. Two isolates (I2, I7) showed weaker hydrolysis of CMC (ratio = 1-1.2) at 30°C and no hydrolysis at 45°C (Figure 4.2A). Two isolates (I5, I8) showed no hydrolysis (ratio ≤1) at

either temperature. Isolates that can hydrolyze CMC have endocellulases that randomly cleave the cellulose chain at internal bonds of amorphous regions (Sajith *et al.*, 2016). These results indicate that I1, I3, I4, and I6 are mesophilic or potentially thermophilic cellulose degraders that possess endocellulases. This evidence demonstrates that microorganisms exhibiting endocellulase activity are present in the MCC S_5 enrichment cultures and in the Southern Ontario landfill as well. None of the isolates definitively showed exocellulase activity on microcrystalline cellulose (MCC) at 30ºC nor 45ºC. Although growth of all isolates was observed for at least one temperature, chromogenic clearing of Congo red by any of these isolates was not seen (Figure 4.2B). The difference in the tones of the plates were due to the difference in the thickness of the agar where two plates, each incubated at 30ºC and 45ºC were less thick (top two) and thus allowed more light to pass through than the other (bottom two). Growth was possible without hydrolysis zones because the microorganisms could have been growing on the peptone in the medium instead of the MCC.

Several reasons may explain the absence of hydrolysis zones on MCC: first, these isolates may not be able to hydrolyze microcrystalline cellulose in the allotted incubation time (11 days). No positive control pure culture was available to confirm the validity of this test. Secondly, the MCC may have settled to the bottom of the plate as the agar was solidifying, preventing the microorganisms from accessing most of the MCC. To remedy this problem for future tests, an underlay of the MCC agar containing no MCC can be poured before adding a thin layer of agar containing MCC. Alternatively, MCC degradation can be qualitatively detected by culturing in liquid MCC medium and the surface morphology of MCC after incubation with the isolates can observed by scanning electron microscopy for degradation (Wang *et al.*, 2017a). Thus,

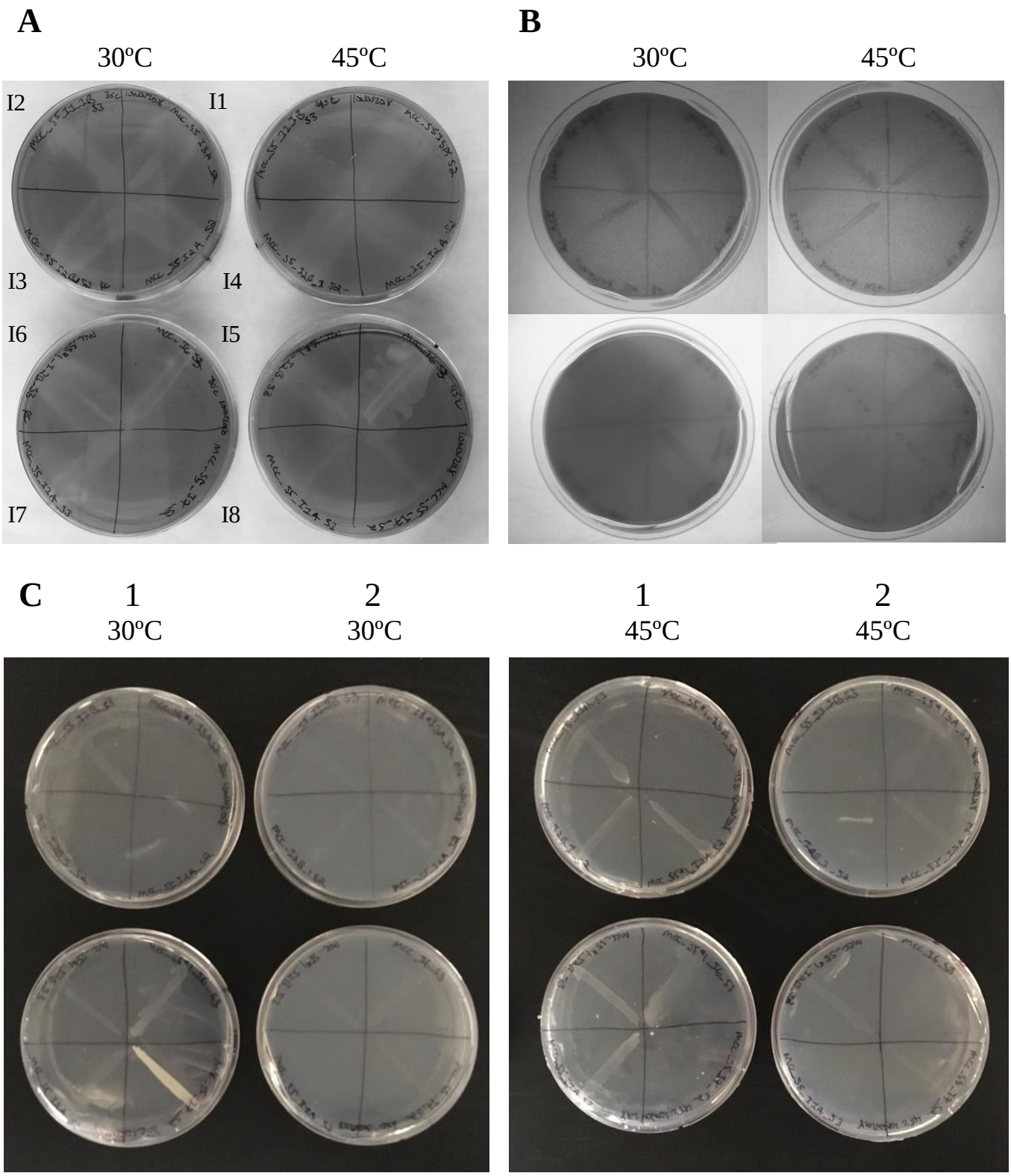


Figure 4.2 Colonies grown on cellulose-containing media at different temperatures. A) Clearing zones formed on carboxymethyl cellulose stained by Congo red (CR) **B)** Growth on microcrystalline cellulose containing CR. **C)** M9 medium amended with cellobiose (1) and unamended M9 medium (2). Orientation of isolates as specified in **A)** are the same on all plates.

exocellulase activity was not demonstrated in the MCC S_5 enrichment culture, but cannot be ruled out at this time.

Beta-glucosidase activity was demonstrated by five isolates (I3, I5, I6, I7, I8) through enhanced growth on cellobiose-amended medium. At 30ºC, three isolates (I5, I7, and I8) grew better (ratio >1) on cellobiose-amended medium compared to growth on the control (M9 minimal medium) (Table 4.5, Figure 4.2C). At 45ºC, four isolates (I3, I5, I6, I7) were able to grow better on cellobiose-containing M9 minimal medium. These observations suggest that the beta-glucosidases of two isolates (I3 and I6) operate only at a warmer, thermophilic temperature. The optimal temperature for beta-glucosidases has been reported to be 45ºC for i) a beta-glucosidase isolated from soil that was closely related to beta-glucosidase from *Methanocella paludicola* (Euryarchaeota), ii) a putative beta-glucosidase from *Caldiliniea aerophila* (Chloroflexi), and iii) a beta-glucosidase from *Anaerolinea thermophila* (Chloroflexi), as well as at 50ºC for *Bacillus licheniformis* from the rumen of a goat (Bergmann *et al.*, 2014; Seo *et al.*, 2013).

My results indicate that I1, I2, and I4 contain endocellulase(s), while I5 and I8 contain betaglucosidase(s). Isolates I3, I6, and I7 likely encode both of these cellulolytic enzyme types.

4.3.2 Frequency of ESVs in enrichment cultures and the paper-free control

A cumulative total of 319 ESVs were detected across the microcrystalline cellulose enrichments, paper enrichments, and the paper-free control cultures. The greatest number of ESVs were detected in MCC S_0 (199 ESVs), which was expected because this enrichment represents the time zero for the paper enrichment trial, and would theoretically represent the maximum number of populations present under the given conditions (Table 4.6). The MCC S_0 enrichment was sequenced two weeks after incubation meaning microorganisms recovered from the leachate filter were already enriched

for cellulolytic activity through exposure to MCC as a substrate. The number of ESVs detected in the microcrystalline cellulose enrichment remain stable between S_1 and S_2. It is uncertain whether the microbial populations have actually stabilized: sequencing of subsequent subcultures would be needed to verify this. The microcrystalline S_2 enrichment culture was used as the inoculum (in a 1:10 dilution) to create the paper enrichments with either copy paper, cardboard, newsprint, or filter paper as the amended substrate. The average number of ESVs detected and the average number of ESVs occurring at >1% relative abundance in the paper enrichment cultures declined from S_0 to S_2, indicating a general reduction in microbial population richness over the enrichment trial (Table 4.6). Two exceptions were the Newsprint culture, where the average number of ESVs increased from S_1 to S_2 and the Filter Paper culture where the number of ESVs remained unchanged from S_1 to S_2. In the paper-free control (CTRL), of the 75 ESVs detected in S_0 and 77 ESVs in S_2, 49 ESVs were shared between the two subcultures. The relative abundance of the 20 and 28 other ESVs in S_0 and S_2, respectively, were at <1% relative abundance. It is possible that these organisms fluctuate at abundances near our limit of detection. Six ESVs in the CTRL S_0 sample present at 1-5.6% were undetectable in S_2, representing organisms that were lost or at lower abundance via the enrichment process. The relatively stable composition of ESVs from S_0 to S_2 in the paper-free control samples suggests that the synthetic leachate on its own is not strongly affecting the microbial community when considering the reductions in richness seen in the paper enrichments. A note that the S_1 time point for the paper-free cultures was discarded as one replicate was removed after filtering by read number, and the other was considered unreliable without a replicate.

Table 4.6. Number of ESVs detected and the number of ESVs with >1% relative abundance in the MCC and paper enrichment trials. All numbers come from averages over three replicates per culture except where noted.

Culture	Average frequency of ESVs detected per subculture (±sd of replicates)			Average frequency of ESVs >1%		
	Subculture			Subculture		
	0	1	2	0	1	2
Microcrystalline cellulose	196[a]	89[a]	90[a]	17[a]	3[a]	7[a]
Copy paper	122±22	94[b]±9	51±7	14	15[b]	15
Cardboard	114±9	68±14	50±7	21	14	7
Filter paper	119±11	94±7	94±6	17	7	10
Newsprint	91±8	67±2	83±9	20	15	11
Paper-free control	75±9	28[c]	77±22	27	16[c]	13

sd = standard deviation

[a] No replicates

[b] Duplicates

[c] Only one replicate available

4.3.3 Microbial composition of enrichment cultures

4.3.3.1 Core microbiota

Of the 319 ESVs cumulatively detected across all cultures, analysis was focused on ESVs that form the core microbiota in each enrichment as well as the ESVs that enriched within each culture's core microbiome over time and transfers. In this context, the core microbiome in a particular enrichment represents the consistently present, relatively abundant organisms across an enrichment trial. Core microbiome members were ranked based on summed relative abundance across the trial. A cut-off threshold at which the last ranked ESV of the core microbiome was detected above 1% in any replicate of the S_2 cultures (final time point) was applied to separate the core microbiota potentially involved in cellulose degradation from other low abundance organisms. The core microbiota for the different trials ranged from 20-37 ESVs (MCC = 21 ESVs; CP = 26; CB = 20; FP = 30; NP = 37, and paper-free control = 35) (Figure 4.3), and included 96

unique ESVs in total. Across all cultures, members of the Proteobacteria, with representatives from Gammaproteobacteria and Alphaproteobacteria, as well as Bacteroidetes, Firmicutes, and Chloroflexi were predominant (Figure 4.3). Actinobacteria were present in the core microbiome of the paper-free control but not in the cellulose enrichments. As mentioned in previous chapters, these phyla and classes have been detected in landfills (Song *et al.*, 2015a, 2015b; Collins-Fairclough *et al.*, 2018). A comparison between raw landfill leachate and cellulose-enriched landfill leachate microcosms showed members of Bacteroidetes and Firmicutes increased 5 and 8% respectively, and Proteobacteria decreased 15% in the enriched leachate (Ransom-Jones *et al.*, 2017). Metagenome-assembled genomes of Proteobacteria, Bacteroidetes, and Firmicutes from the cellulose-enriched microcosms in landfill leachate were generated and they were shown to encode glycosyl hydrolase families containing cellulases (Ransom-Jones *et al.*, 2017). It was surprising that representatives of Chloroflexi were present in the core microbiota of the paper enrichments. Although not usually implicated in cellulose degradation, Chloroflexi possess thermophilic representatives from the classes Anaerolineae and Ktedonobacteria that exhibit cellulose degradation in culture (Podosokorskaya *et al.*, 2013a; King and King, 2014). It is interesting to note that the strains cultured in these studies were grown at 47°C and 40-65°C respectively, but our paper enrichments were cultured at 37°C. It is possible that the incubation temperature restricted the Chloroflexi population from growing.

At the genus or lowest classifiable taxonomic level, the 96 core microbiota ESVs comprised 14 genera in MCC, 19 in CP, 13 in CB, 18 in FP, 21 NP enrichments, and 20 in the paper-free control culture (CTRL), with some ESVs present in multiple culture trials (Figure 4.4, Figure 4.5). The majority of ESVs are between 0-10% relative abundance in each enrichment (Figure 4.4). Two

ESVs at >1% relative abundance were common in all the paper enrichments and the paper-free control: ESV171-Caulobacteraceae and ESV28-*Proteiniphilum*, (Figure 4.4).

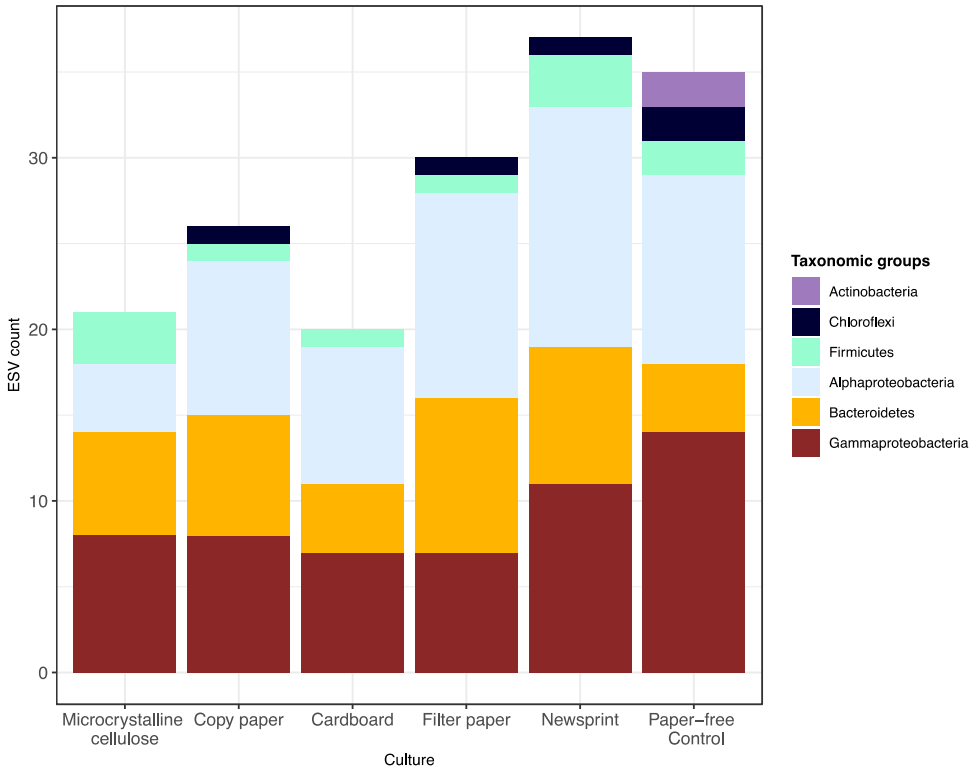
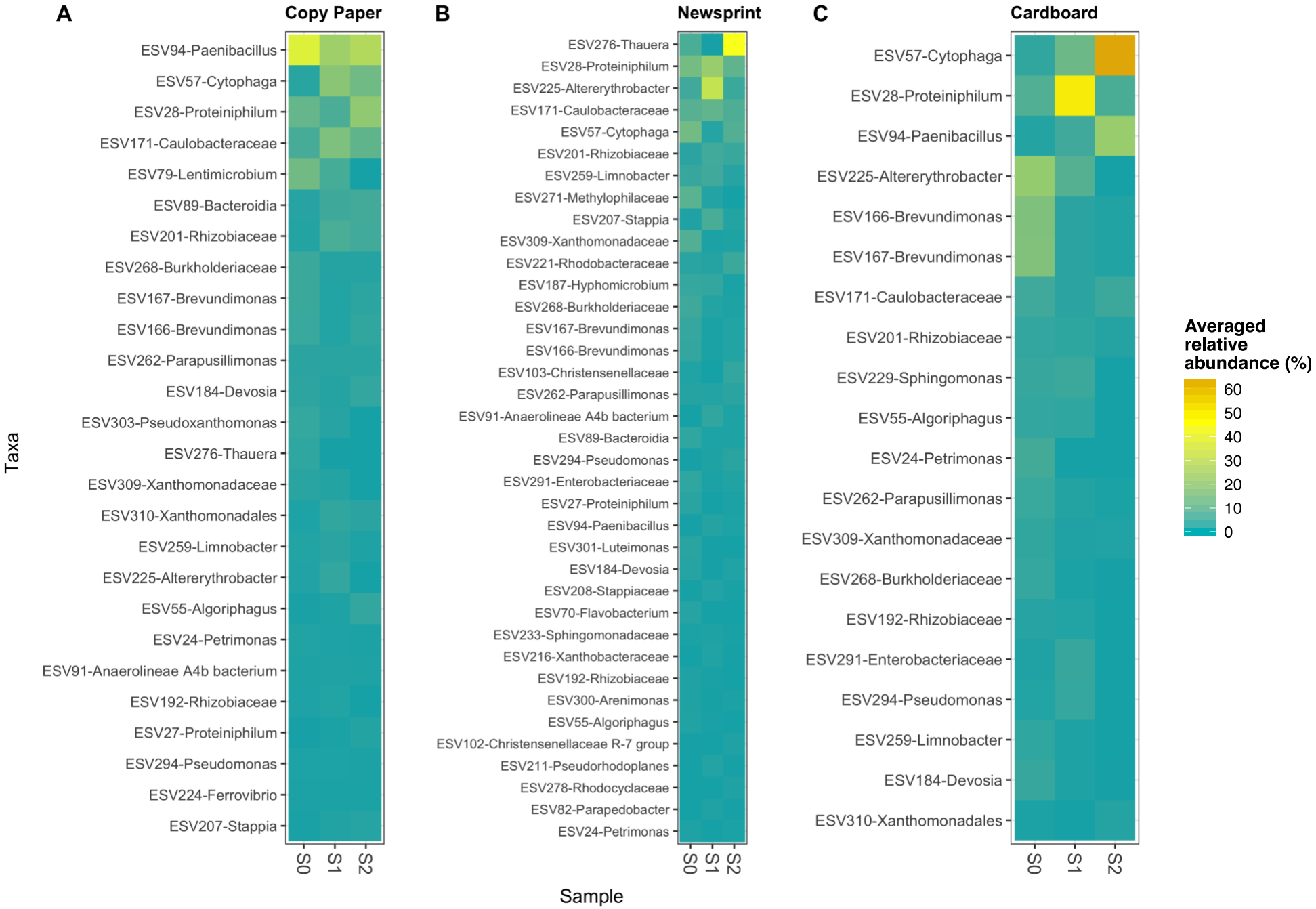


Figure 4.3. Core microbiome ESVs coloured by their taxonomic group (phylum and proteobacterial-class) affiliations for the microcrystalline cellulose enrichment, the paper enrichment cultures, and the paper-free control.


A
Copy Paper
ESV94-Paenibacillus
ESV57-Cytophaga
ESV28-Proteiniphilum
ESV171-Caulobacteraceae
ESV79-Lentimicrobium
ESV89-Bacteroidia
ESV201-Rhizobiaceae
ESV268-Burkholderiaceae
ESV167-Brevundimonas
ESV166-Brevundimonas
ESV262-Parapusillimonas
ESV184-Devosia
ESV303-Pseudoxanthomonas
ESV276-Thauera
ESV309-Xanthomonadaceae
ESV310-Xanthomonadales
ESV259-Limnobacter
ESV225-Altererythrobacter
ESV55-Algoriphagus
ESV24-Petrimonas
ESV91-Anaerolineae A4b bacterium
ESV192-Rhizobiaceae
ESV27-Proteiniphilum
ESV294-Pseudomonas
ESV224-Ferrovibrio
ESV207-Stappia
S0
S1
S2
B
Newsprint
ESV276-Thauera
ESV28-Proteiniphilum
ESV225-Altererythrobacter
ESV171-Caulobacteraceae
ESV57-Cytophaga
ESV201-Rhizobiaceae
ESV259-Limnobacter
ESV271-Methylophilaceae
ESV207-Stappia
ESV309-Xanthomonadaceae
ESV221-Rhodobacteraceae
ESV187-Hyphomicrobium
ESV268-Burkholderiaceae
ESV167-Brevundimonas
ESV166-Brevundimonas
ESV103-Christensenellaceae
ESV262-Parapusillimonas
ESV91-Anaerolineae A4b bacterium
ESV89-Bacteroidia
ESV294-Pseudomonas
ESV291-Enterobacteriaceae
ESV27-Proteiniphilum
ESV94-Paenibacillus
ESV301-Luteimonas
ESV184-Devosia
ESV208-Stappiaceae
ESV70-Flavobacterium
ESV233-Sphingomonadaceae
ESV216-Xanthobacteraceae
ESV192-Rhizobiaceae
ESV300-Arenimonas
ESV55-Algoriphagus
ESV102-Christensenellaceae R-7 group
ESV211-Pseudorhodoplanes
ESV278-Rhodocyclaceae
ESV82-Parapedobacter
ESV24-Petrimonas
S0
S1
S2
C
Cardboard
ESV57-Cytophaga
ESV28-Proteiniphilum
ESV94-Paenibacillus
ESV225-Altererythrobacter
ESV166-Brevundimonas
ESV167-Brevundimonas
ESV171-Caulobacteraceae
ESV201-Rhizobiaceae
ESV229-Sphingomonas
ESV55-Algoriphagus
ESV24-Petrimonas
ESV262-Parapusillimonas
ESV309-Xanthomonadaceae
ESV268-Burkholderiaceae
ESV192-Rhizobiaceae
ESV291-Enterobacteriaceae
ESV294-Pseudomonas
ESV259-Limnobacter
ESV184-Devosia
ESV310-Xanthomonadales
S0
S1
S2
Averaged relative abundance (%)
60
50
40
30
20
10
0
Taxa
Sample

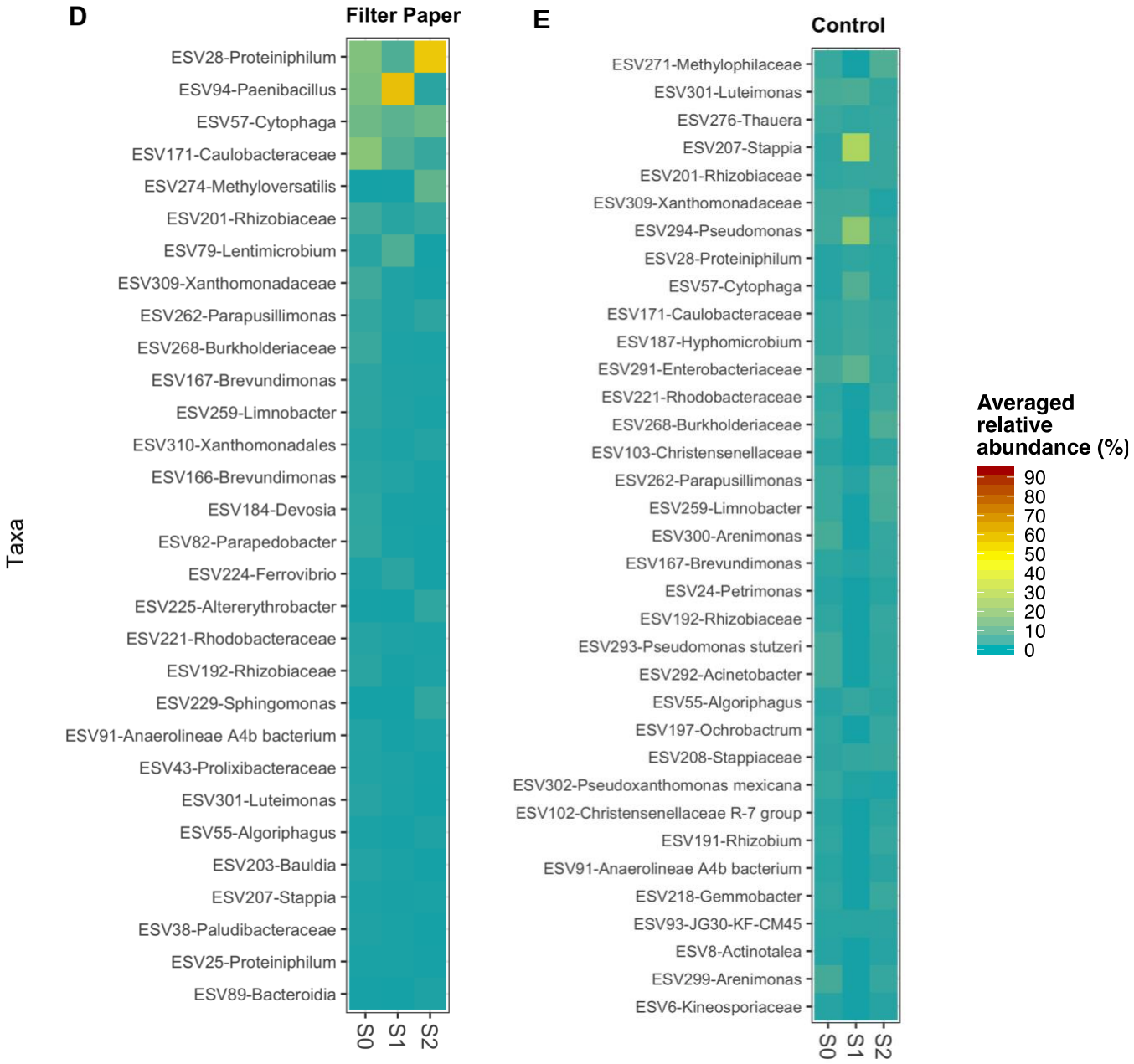


Figure 4.4. Heat maps showing the average relative abundance of ESVs at >1% in the A) copy paper, B) newsprint, C) cardboard, D) filter paper, and E) paper-free control cultures.

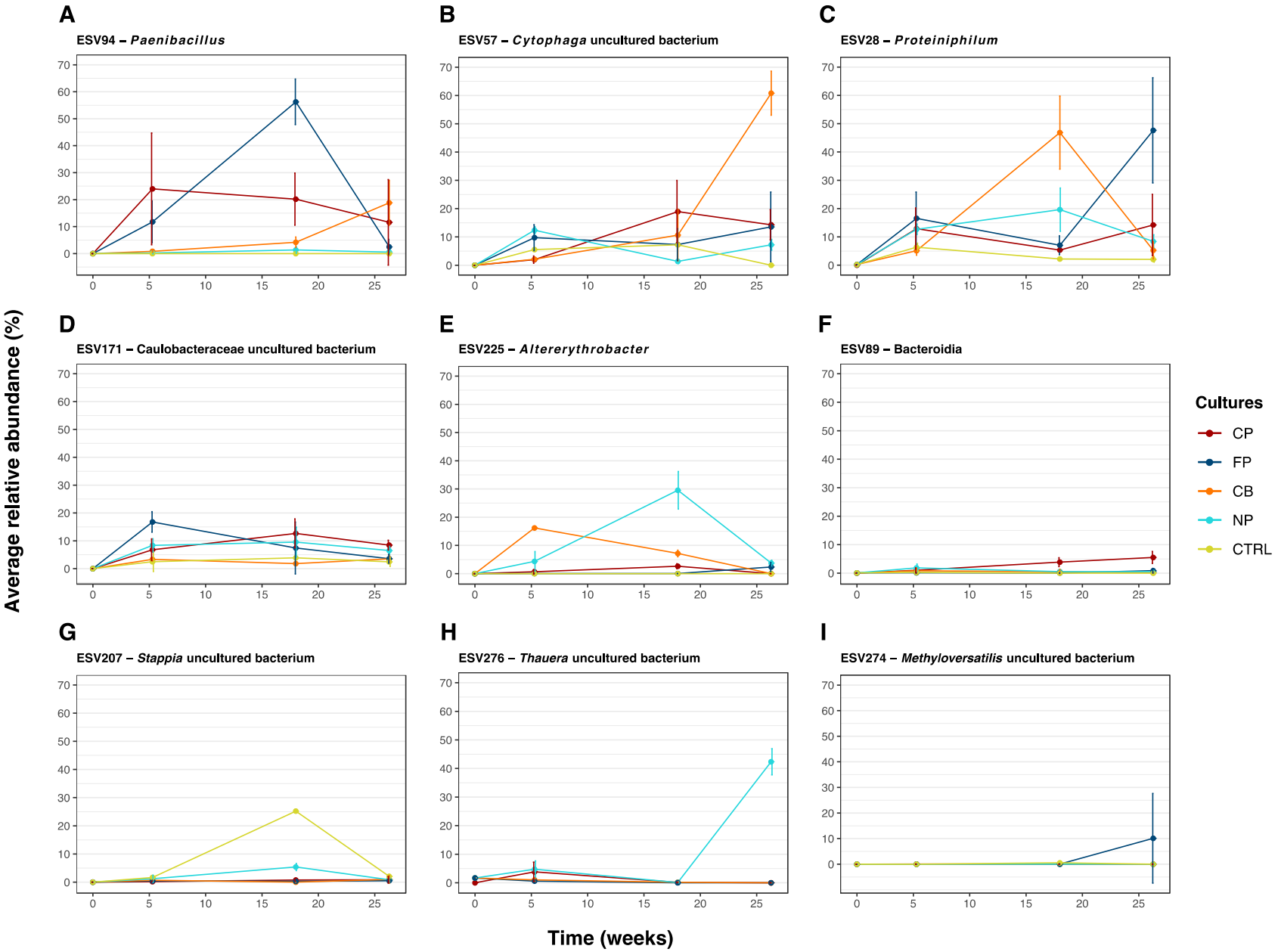


Figure 4.5. Line plots showing the change in the average relative abundance across all cultures for ESVs (A-I) with >5% relative abundance in any time point, in any culture, that showed an increase in relative abundance by week 18 (S_1). CP = copy paper, FP = filter paper, CB = cardboard, NP = newsprint, NP = newsprint, CTRL = paper-free control.

4.3.3.2 Population dynamics in the microbial communities of paper enrichment cultures

A subset of the most abundant ESVs (>5% relative abundance in any replicate, 21 ESVs) was selected to examine enrichment patterns over the experiment. This group was screened for an increase in relative abundance from T=0 (relative abundance taken from MCC S_2) over at least 18 weeks (S_1). The ESVs that showed increases in relative abundance only from T=0 to 5 weeks (S_0) were not considered potential cellulose degraders and were not considered further. Four ESVs were present at >5% in the MCC and paper enrichments that showed an increase in their average relative abundance by at least 18 weeks: ESV94-*Paenibacillus*, ESV57-*Cytophaga* uncultured bacterium, ESV-*Proteiniphilum*, and ESV161-Caulobacteraceae uncultured bacterium (Figure 4.5A-D).

ESV94-*Paenibacillus* (phylum Firmicutes) was not detectable in the MCC S_2 enrichment, increased to 56% in FP by 18 weeks, but decreased by 26 weeks to 2.5% (Figure 4.5A). In the CP culture, *Paenibacillus* peaked at 24% but gradually decreased to 12% by 26 weeks. In CB, *Paenibacillus* slowly increased to peak at 18% at 26 weeks. Its abundance remained relatively stable, and low, in NP and the CTRL. The filter paper we used has a minimum 98% cellulose content, with other trace constituents such as ash and minerals (Sigma-Aldrich). The cellulose content of waste office paper was reported as 87%, with newsprint at 48%, and cardboard at 57% (Palmisano and Barlaz, 1996). The relative increase in ESVs affiliated with *Paenibacillus* in the paper cultures, especially those with higher cellulose content such as filter paper and copy paper, strongly suggests that is involved in cellulose degradation. *Paenibacillus* have been found in landfill leachate (Chua *et al.*, 2014; Remmas *et al.*, 2017b) and isolated strains of the genus from forest soil have demonstrated degradation of cellulose (Wang *et al.*, 2008; López-Mondéjar *et al.*, 2016). ESV94 – *Paenibacillus* is the most likely cellulose degrader enriched in the paper cultures.

ESV57-*Cytophaga* uncultured bacterium (phylum Bacteroidetes) was also not detectable in the MCC enrichment S_2 culture. It increased in abundance to 61% by 26 weeks in the CB culture (Figure 4.5B). In CP, *Cytophaga* also increased to 19% by 18 weeks but declined by 26 weeks. In FP, *Cytophaga* increased to 10% by 5 weeks and then hovered around 10% for the remainder of the trial. *Cytophaga* in NP increased to 12% by 5 weeks then decreased to 1% by 18 weeks, and increased again to 7% by 26 weeks suggesting it was not stably enriched and potentially not a strong competitor for cellulose. ESV57 in the paper-free control increased to 7% by 18 weeks but was not detectable by 26 weeks. This suggests it was able to survive on the synthetic leachate, which has urea as the only organic compound, as well as potentially low amounts of microcrystalline cellulose subcultured from the original MCC bottle. By S_2, the cellulose was likely diluted out or consumed and ESV57 could no longer sustain itself without a carbon source. A member of the *Cytophaga*, *Cytophaga hutchinsonii*, is a well-known aerobic cellulolytic soil bacterium. *C. hutchinsonii*'s genome encodes CAZymes with predicted involvement in hemicellulose hydrolysis (Xie *et al.*, 2007). Waste corrugated cardboard is composed of 57% cellulose, 9.9% hemi-cellulose, and 20.8% lignin, whereas copy paper is 87.4% cellulose, 8.4% hemi-cellulose, and 2.3% lignin (Staley *et al.*, 2012). The high relative abundance of *Cytophaga* in CB by week 26 suggests it can hydrolyze lignin and/or hemicellulose in the complex lignocellulose matrix, allowing cellulose to be physically accessed for hydrolysis. A recent study reported enrichment of *Cytophaga* from forest soil amended with ^{13}C-lignin and ^{13}C-cellulose, suggesting an ability to break down these substrates (Wilhelm *et al.*, 2019). No *Cytophaga* populations have been reported in landfill leachate, but *C. hutchinsonii* exhibits contact-dependent degradation of crystalline cellulose (as opposed to releasing extracellular cellulases) and quick gliding motility (Walker and Warren, 1938; Xie *et al.*, 2007). These features can potentially aid in

rapid colonization of cellulolytic substrates and immediate uptake of glucose released from cellulose. In the paper cultures, it seems possible *Cytophaga* is degrading lignin and hemicellulose, making cellulose available. Populations that are capable of cellulose degradation and are at relatively high abundance could then hydrolyze the available cellulose, and these populations would therefore further increase. This is likely the case for ESV94-*Paenibacillus* in CB, as an increase is seen from weeks 18-26 concurrent with an increase in *Cytophaga* abundance.

ESV28-*Proteiniphilum* (phylum Bacteroidetes) was detected at 0.2% in the MCC enrichment. This ESV increased in to 48% in FP by week 26, but non-linearly: it had increased to 17% by week 5 and then dropped to 7% in week 18 (Figure 4.5C). In CB, *Proteiniphilum* steadily increased to 47% then dropped to 5% by week 26. In CP and NP, *Proteiniphilum* increased to similar abundances by 5 weeks, to 13%. The two trials then diverged by 18 weeks, with its abundance in CP decreasing to 6% whereas it increased to 20% in NP. Then by 26 weeks, *Proteiniphilum* increased to 14% in CP and decreased to 8% in NP. The relative abundance of ESV28 in the CTRL remained relatively even, nearing 2% from week 18 and onwards, suggesting no enrichment of *Proteiniphilum* from the synthetic leachate alone. These observations generally suggest that *Proteiniphilum* grows well on cardboard, which is high in cellulose and lignin content and on filter paper, which is mostly cellulose. One study found that *Proteiniphilum sacchorofermentans* str. M3/6^{T} demonstrated weak enzymatic activity against CMC and phosphoric acid-swollen cellulose (Hahnke *et al.*, 2016). However, another research group sequenced the M3/6^{T} genome and found that it possessed genes encoding all three types of cellulases as well as enzymes predicted to degrade sugar compounds in hemicellulose (Tomazetto *et al.*, 2018). However, Tomazetto *et al.* (2018) suggested that the environmental conditions from which M3/6^{T} was isolated were different from isolation conditions and thus the metabolisms observed differed from what was predicted

from the genome. *Proteiniphilum* have not been reported from landfill microbiota but *P. acetatigenes* has been found in an anaerobic reactor (Chen and Dong, 2005). It is also interesting to note that *Proteiniphilum* is reported to be an obligate anaerobe and ferments glucose and cellobiose producing a variety of acids (Chen and Dong, 2005; Tomazetto *et al.*, 2018). However, our cultures were incubated with shaking, thus promoting oxygen diffusion. However, it is possible that anoxic micro-niches in the crevices of the cardboard and the environment near the bottom of the test tube persisted. The proliferation of *Proteiniphilum* could mean a high amount of acid was produced, which may have self-impeded the growth of that population as well as others until eventually a negative feedback loop in acid production was created. There is an interesting dynamic between ESV-94-*Paenibacillus* and ESV28-*Proteiniphilum*. It is unclear as to why the abundance of *Paenibacillus* in FP decreased between 18-26 weeks. However, in this interval of time, ESV28-*Proteiniphilum* increased 5-18 weeks. The growth pattern of these two organisms switched after being at similar abundances after five weeks of enrichment. This dynamic suggests resource competition between these two groups.

ESV171-Caulobacteraceae uncultured bacterium (Alphaproteobacteria) was not detected in the MCC S_2 enrichment, but was weakly enriched in CP and NP (Figure 4.5D). Recent study shows members of Caulobacteraceae were enriched from forest soil amended with lignin, hemicellulose, and cellulose, and have been shown to assimilate these ^{13}C-labeled substrates (Wilhelm *et al.*, 2019). In the same study, members of *Caulobacter* were able to assimilate the three substrates. ESV171-Caulobacteraceae may break down these components of lignocellulose in the paper enrichments. It is a bit surprising that this population is not more abundant compared to previously discussed ESVs. One reason may be the production of three types of cellulases by the Caulobacteraceae bacterium may require a higher energy expenditure than the other cellulolytic

organisms which only have one or two types of cellulases, thus hampering the proliferation of the Caulobacteraceae population.

The other ESVs that showed an increase in relative abundance over one or more subcultures included ESV225-*Altererythrobacter*, ESV89 - an organism from the class Bacteroidia, ESV207-*Stappia*, ESV276-*Thauera*, and ESV224-*Methyloversatillis* (Figure 4.5E-I). Some members of the *Altererythrobacter* have demonstrated cellulose degradation in culture (Xue *et al.*, 2012; Yuan *et al.*, 2017). *Thauera* isolated from coal have been shown to degrade lignin (Wang *et al.*, 2016). To our present knowledge, no study has looked at cellulose degradation by *Stappia* and *Methyloversatillis*. Our data shows *Stappia* slightly increased in abundance at 18 weeks in NP but decreased again by 26 weeks, suggesting a weak enrichment (Figure 4.5G). These microorganisms seem to be slightly to moderately enriched in the presence of a paper-based cellulose source compared to their abundance in MCC S_2, whereas their stable low abundance in the CTRL cultures do not suggest enrichment in the synthetic leachate alone. One exception to this was the ESV207-*Stappia* CTRL at 18 weeks, which is quite high. However, this data point was based on only one replicate (the other replicate was removed after sequencing because its read counts were lower than the non-template PCR controls), thus it should be interpreted with caution. This data point does suggest caution in inferring a cellulose degrading role for ESV207-*Stappia*.

4.4 Conclusions

Members of the microbial community in the leachate from CLC1_T1 at the southern Ontario landfill showed cellulolytic activity in culture and, based on 16S rRNA gene amplicon sequencing, enrichment of specific populations when amended with microcrystalline cellulose, copy paper, cardboard, filter paper, and newsprint. Select isolates were able to degrade carboxymethyl

cellulose (CMC) and cellobiose at 30°C and 45°C, thus demonstrating endocellulase and exocellulase activities. Microcrystalline cellulose degradation was not observed by any of the isolates either because this activity was absent from these organisms, insufficient incubation time was allowed, or a methodological improvement in plating MCC is needed. Hence, exocellulase potential in the MCC enrichment cultures was not definitively determined. However, the detected CMC and cellobiose degradation in the MCC enrichment culture may be sufficient to partially hydrolyze MCC in the culture. Microcrystalline cellulose has both crystalline and amorphous regions on which endocellulases can act and release cellodextrins (polymers of glucose with greater than two subunits that have been released from cellulose hydrolysis). The cellodextrins can then be hydrolyzed by both endo- and exocellulases, and subsequently beta-glucosidases (Wood and Bhat, 1988; Annamalai *et al.*, 2016).

The ESVs identified from the microcrystalline cellulose and paper enrichment cultures were from members of the Proteobacteria including the Alphaproteobacteria and Gammaproteobacteria classes, Bacteroidetes, Firmicutes, and Chloroflexi. Of the abundant (>1%) ESVs, only two ESVs of nearly 100 found collectively across the MCC and paper enrichments were shared in the core microbiota of the cultures. One member each from the *Paenibacillus*, *Cytophaga*, *Proteiniphilum*, and Caulobacteraceae were the most abundant ESVs showing enrichment in at least one paper culture, to a minimum of 5% abundance. *Paenibacillus*, *Cytophaga*, and *Caulobacter* from Caulobacteraceae are known cellulose degraders, however only *Paenibacillus* and *Caulobacter* have been detected in landfill leachate (Wang *et al.*, 2008; Remmas *et al.*, 2017b; Xie *et al.*, 2007; Wilhelm *et al.*, 2019; Abraham *et al.*, 1999). *Proteiniphilum* has neither been reported to degrade cellulose nor been found in landfills. The shifts in population dynamics seen in the paper cultures may be a result of some microorganisms' abilities to hydrolyze lignin and hemicellulose thus

freeing cellulose in the paper, giving rise to other microbial populations in the enrichment cultures. Taken together, these data demonstrate that microbial cellulose degradation is present in the composite leachate cistern at the southern Ontario landfill.

Chapter 5: Conclusions and Future Directions

Municipal waste sites are complex and heterogeneous engineered sites that house substantial microbial diversity. Past landfill research has used culture-based methods to identify and characterize isolates, as well as 16S rRNA gene sequencing to identify the microbial communities (Pourcher *et al.*, 2001; Song *et al.*, 2015a, 2015b; Remmas *et al.*, 2017a; Stamps *et al.*, 2016). Our research advances on these techniques with the application of metagenomics to a landfill environment in order to connect the functional potential with its community composition. Although landfills are engineered to minimize and delay the degradation of waste, the results of this study show that there are potential cellulolytic microbes harbouring different types of cellulases in the leachate of two landfills, and that this cellulose degradation is confirmed in landfill leachate biomass grown in enrichment cultures at the bench scale. These are encouraging preliminary results that emphasize the value of investigating municipal solid waste for novel microorganisms that degrade cellulose under challenging conditions, and which in future could be leveraged, optimized, and scaled up for industrial processes.

Investigation of the Riverton City dump in Jamaica and the adjacent Duhaney River showed that the microbial community represented by MAGs in leachate is more diverse than that of the river. Bacteroidetes, Proteobacteria, Firmicutes, and Tenericutes were the most abundant phyla recovered in the eighteen metagenome-assembled genomes (MAGs) identifiable by ribosome proteins. Only five MAGs were recovered from the Duhaney river sample, all associated with the Proteobacteria. Six of the high-quality leachate MAGs and one alphaproteobacterial river MAG possessed potential cellulase genes. The total number of genes belonging to glycosyl hydrolase (GH) families containing cellulases was similar in the leachate and the river, whereas beta-glucosidases were detected in both systems, endocellulases were only detected in the river.

Cellulose degradation may occur in the river environment; however, it would perhaps be more effective if exocellulases were also present. The potential cellulases detected in the leachate belonged mostly to Bacteroidetes, Firmicutes, and Spirochaetes, whereas Bacteroidetes and Proteobacteria (Alphaproteobacteria and Gammaproteobacteria) encoded the prevalent potential cellulases in the river. A large majority of potential GH genes were either vaguely annotated or labeled as hypothetical proteins, suggesting novel carbohydrate-modifying enzymes not exclusive to cellulases are present in these systems. These two sites do not seem to have overlapping community members, as <1% of the reads of one metagenome mapped to the other metagenome and vice-versa. However, this does not dismiss the possibility of contamination of the river by the leachate, a growing concern in Jamaica.

The data presented in Chapter 2 was interpreted with caution as one metagenome each from leachate and the river was used to define the community composition and cellulolytic capacity at these sites. Future directions would include re-sampling of the landfill leachate and the river to gather more temporal and spatial samples. Multiple samples would help support the interpretations here, and possibly increase the detection of taxonomic groups that appear seasonally and at different sampling locations, especially important given the heterogeneity of landfills. Furthermore, enrichment cultures can be set up to capture the cellulolytic populations in landfill in order to confirm the low cellulolytic potential that was seen in the original metagenomes.

The results from the southern Ontario landfill also suggested cellulolytic capacity by members of Bacteroidetes, Firmicutes, Verrucomicrobia, and Spirochaetes being prevalent across leachate-containing wells. This study is the first report of potential cellulases affiliated with Dictyoglomi and Ignavibacteria in landfill leachate, suggesting that they may have cellulose degradation capacities or carbohydrate-modifying activities. Putative endocellulase genes and beta-glucosidase

activities were detected at all sites, but putative exocellulases genes made up a small fraction of genes identified from only three of six sites (LW2, LW3, and GW1). As was seen in the Jamaican sites, the relative abundance of genes annotated as glycosyl hydrolases and hypothetical proteins in the Ontario landfill metagenomes make up the majority of the dataset upon screening for cellulase genes. This speaks to the growing abundance of novel genes in databases, where the genes identified in the landfill here add to those that need to be better annotated. The data also showed remarkable consistency in the relative abundance of GH families containing cellulases in the composite leachate cistern (CLC) in the difference of a week, suggesting a stable presence of cellulases in an open system where the landfill leachate is constantly circulated and pumped to a wastewater treatment plant. In contrast, the GH compositions of the three leachate wells sampled varied more strongly amongst each other, indicating the carbohydrate-active enzymes in the spatially separated leachate wells are different at a given time. Genes annotated as beta-glucosidases, exocellulases, endocellulases, and cellulases were identified in the groundwater well. The groundwater sample also had the highest proportion of glycosyl hydrolase and hypothetical proteins, indicating a large amount of novel cellulose-degradation capacity and carbohydrate-modifying activity.

The research of the GH composition and its taxonomic affiliations presented in Chapter 3 can be strengthened by repeated sampling of these sites. It would be interesting to see if other CLCs at the Ontario landfill are consistent with CLC1 and whether the relative abundance of the GH families at each CLC are consistent over time in order to confirm the trend that was seen here. Repeated sampling of the LWs may reveal whether the LWs maintain a heterogeneity in the relative abundance of GH families. Moreover, sampling of a pristine aquifer ("pristine" based on chemical composition, total organic carbon measurements, or total dissolved solids) at the landfill

site could help to establish baseline microbial communities and functions. The microbial communities of this pristine aquifer could then be compared to GW1 in order to gauge whether the GH composition seen in this metagenome is driven by leachate contamination. Lastly, in addition to glycosyl hydrolase families, other protein families associated with cellulose degradation could be screened to increase the cellulolytic potential we identified from our metagenomes. These include protein families such as carbohydrate-binding modules; cellulases that use another medium besides water, such as cellulose phosphorylases; and other non-enzymatic proteins that are associated with anaerobic cellulose degradation (i.e., scaffoldin, dockerin, and cohesin in cellulosomes, and/or polysaccharide-utilization loci) (Cragg *et al.*, 2015; Grondin *et al.*, 2017).

Cellulose degradation, with carboxymethyl cellulose and cellobiose as specific substrates, was demonstrated by isolates in enrichments cultured from biomass from landfill leachate (CLC1_T1). Enrichments were maintained in synthetic leachate amended with microcrystalline cellulose over 59 weeks. The 16S rRNA gene amplicon sequencing of enrichments on different paper sources (copy paper, cardboard, filter paper, and newsprint) revealed 96 unique exact variance sequences (ESVs) in the core microbiota across all conditions. Four ESVs: *Cytophaga*, *Proteiniphilum*, and Caulobacteraceae showed general increased relative abundance over 26 weeks and were the most abundant in the paper-enriched cultures. *Proteiniphilum* has neither been reported to degrade cellulose nor previously been found in landfills.

Future directions of the culture-based research in Chapter 4 could include developing a method to detect or measure microcrystalline cellulose degradation definitively, such using scanning electron microscopy to observe the change in crystalline structure of microcrystalline cellulose after incubation with enrichment or isolate cultures. Furthermore, sequencing the metagenomes of the

microcrystalline cellulose cultures could confirm whether potential cellulase genes are enriched, what additional potential functions are present in the cellulolytic communities, and which microorganism(s) may be performing these potential functions. 16S rRNA gene sequencing of the isolates derived from MCC agar plates would identify these culturable microorganisms. Quantitative assessment of their cellulase activity, such as measuring the specific activity of their enzymes, could indicate their cellulase efficiency *in vitro*. Lastly, it would be interesting to test the synergistic effects of a mixed culture of the cellulolytic isolates grown on carboxymethylcellulose, microcrystalline cellulose, and cellobiose, as well as on the different types of paper.

Bibliography

Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. (2016). A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Front Microbiol* **6**: 1469.

Abraham W-R, Strompl C, Meyer H, Lindholst S, Moore ERB, Christ R, *et al.* (1999). Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundirnonas* and *Caulobacter*. *Int J Syst Bacteriol* **49**: 1053–1054.

Abt B, Göker M, Scheuner C, Han C, Lu M, Misra M, *et al.* (2013). Genome sequence of the thermophilic fresh-water bacterium *Spirochaeta caldaria* type strain (H1T), reclassification of *Spirochaeta caldaria*, *Spirochaeta stenostrepta*, and *Spirochaeta zuelzerae* in the genus *Treponema* as *Treponema caldaria* comb. nov., *Treponema stenostrepta* comb. nov., and *Treponema zuelzerae* comb. nov., and emendation of the genus *Treponema*. *Stand Genomic Sci* **8**: 88–105.

Agency for Toxic Substances and Disease Registry. (2017). Toxicological profile for toluene. Atlanta, GA. https://www.atsdr.cdc.gov/toxprofiles/tp56.pdf.

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, *et al.* (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Alvarez TM, Paiva JH, Ruiz DM, Cairo JPLF, Pereira IO, Paixão DAA, *et al.* (2013). Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PLoS One* **8**: e83635.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919–25.

American Type Culture Collection. ATCC Medium: 2511 M9 minimal agar/broth. https://www.atcc.org/~/media/B3590BBD1DF74203B0E8F0890ECF974F.ashx (Accessed November 22, 2018a).

American Type Culture Collection. ATCC medium: 2720 Carboxymethyl cellulose medium. https://www.atcc.org/~/media/E7E374F27EEB41B284EA577D950CBD25.ashx.

Amritha PK, Anilkumar PP. (2016). Development of landscaped landfills using organic waste for sustainable urban waste management. *Procedia Environ Sci* **35**: 368–376.

Annamalai N, Rajeswari MV, Nallusamy S. (2016). Chapter 2 Cellobiohydrolases: Role, mechanism,and recent developments. In: Gupta VK (ed). *Microbial Enzymes in Bioconversions of Biomass*. Springer International Publishing, pp 29–35.

Aspeborg H, Coutinho PM, Wang Y, Brumer III H, Henrissat B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **12**: 186.

Atobe H, Watabe J, Ogata M. (1983). *Acholeplasma parvum*, a new species from horses. *Int J Syst Bacteriol* **33**: 344–349.

Bagnara C, Toci R, Gaudin C, Belaich JP. (1985). Isolation and characterization of a cellulolytic microorganism, *Cellulomonas fermentans* sp. nov. *Int J Syst Bacteriol* **35**: 502–507.

Balan V. (2014). Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN Biotechnol* **2014**: 1–31.

Bareither CA, Wolfe GL, McMahon KD, Benson CH. (2013). Microbial diversity and dynamics during methane production from municipal solid waste. *Waste Manag* **33**: 1982–1992.

Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol* **77**: 3846–52.

Bayer EA, Belaich J-P, Shoham Y, Lamed R. (2004). The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharide. *Annu Rev Microbiol* **58**: 521–554.

Béguin P, Aubert J-P. (1994). The biological degradation of cellulose. *FEMS Microbiol Rev* **13**: 25–58.

Berger E, Zhang D, Zverlov V V., Schwarz WH. (2007). Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiol Lett* **268**: 194–201.

Bergmann JC, Costa OYA, Gladden JM, Singer S, Heins R, D'haeseleer P, *et al.* (2014). Discovery of two novel β-glucosidases from an Amazon soil metagenomic library. *FEMS Microbiol Lett* **351**: 147–155.

Berlemont R, Martiny AC. (2016). Glycoside hydrolases across environmental microbial communities. *PLOS Comput Biol* **12**: e1005300.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, *et al.* (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208–211.

Brumm P, Hermanson S, Luedtke J, Mead D. (2011). Identification, cloning and characterization of *Dictyoglomus turgidum* CelA, an endoglucanase with cellulose and mannanase activity. *J Life Sci* **5**: 488–496.

Brune A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol* **12**: 168–180.

Buchan A, LeCleir GR, Gulvik CA, González JM. (2014). Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: 686–698.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, *et al.* (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci U S A* **110**: 5540–5545.

Canadian Council of Ministers of the Environment. (1999). Canadian soil quality guidelines for the protection of environmental and human health: Tetrachloroethylene (1997). In: *Canadian Environmental Quality Guidelines*. Canadian Council of Ministers of the Environment: Winnipeg. http://ceqg-rcqe.ccme.ca/download/en/281 (Accessed July 16, 2018).

Canale-Parola E. (1984). Genus I: Spirochaeta. In: Krieg NR (ed). *Bergey's Manual of Systematic Bacteriology*. Williams & Wilkins: Baltimore, MD, p 1:39-46.

Canale-Parola E, Udris Z, Mandel M. (1968). The classification of free-living Spirochetes. *Arch Mikrobiol* **63**: 385–397.

Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233–D238.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Caro-Quintero A, Ritalahti KM, Cusick KD, Löffler FE, Konstantinidis KT. (2012). The chimeric genome of *Sphaerochaeta*: nonspiral spirochetes that break with the prevalent dogma in spirochete biology. *MBio* **3**: e00025-12.

Chen S, Dong X. (2005). *Proteiniphilim* from a UASB reactor treating brewery wastewater. *Int J Syst Evol Microbiol* **55**: 2257–2261.

Chua P, Yoo H-S, Gan HM, Lee S-M. (2014). Draft genome sequences of two cellulolytic *Paenibacillus* sp. strains, MAEPY1 and MAEPY2, from Malaysian landfill leachates. *Genome Announc* **2**: e00065-14.

Collado L, Guarro J, Figueras MJ. (2009). Prevalence of *Arcobacter* in meat and shellfish. *J Food Prot* **72**: 1102–1106.

Collins-Fairclough AM, Co R, Ellis MC, Hug LA. (2018). Widespread antibiotic, biocide, and metal resistance in microbial communities inhabiting a municipal waste environment and anthropogenically impacted river. *mSphere* **3**: e00346-18.

Cragg SM, Beckham GT, Bruce NC, Distel DL, Dupree P, Etxabe AG, *et al.* (2015). Lignocellulose degradation mechanisms across the Tree of Life. *Curr Opin Chem Biol* **29**: 108–119.

Creevey CJ, Doerks T, Fitzpatrick DA, Raes J, Bork P. (2011). Universally distributed single-copy genes indicate a constant rate of horizontal transfer Liberles D (ed). *PLoS One* **6**: e22099.

Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, *et al.* (2016). Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nat Microbiol* **1**: 16146.

Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ. (2017). Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome* **5**: 112.

Davies G, Henrissat B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* **3**: 853–9.

Dehority BA, Grubb JA. (1977). Characterization of the predominant bacteria occurring in the rumen of goats (*Capra hircus*). *Appl Env Microbiol* **33**: 1030–1036.

Delmer DP, Haigler CH. (2002). The regulation of metabolic flux to cellulose, a major sink for carbon in plants. *Metab Eng* **4**: 22–28.

Denger K, Warthmann R, Ludwig W, Schink B. (2002). *Anaerophaga thermohalophila* gen. nov., sp. nov., a moderately thermohalophilic, strictly anaerobic fermentative bacterium. *Int J Syst Evol Microbiol* **52**: 173–178.

Díaz C, Baena S, Fardeau M-L, Patel BKC, Baena CS. (2007). *Aminiphilus circumscriptus* gen. nov., sp. nov., an anaerobic amino-acid-degrading bacterium from an upflow anaerobic sludge reactor. *Int J Syst Evol Microbiol* **57**: 1914–1918.

Dodd D, Kiyonari S, Mackie RI, Cann IKO. (2010). Functional diversity of four glycoside hydrolase family 3 enzymes from the rumen bacterium Prevotella bryantii B14. *J Bacteriol* **192**: 2335–2345.

Dodsworth JA, Blainey PC, Murugapiran SK, Swingley WD, Ross CA, Tringe SG, *et al.* (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat Commun* **4**: 1854.

Dong J, Hong Y, Shao Z, Liu Z. (2010). Molecular cloning, purification, and characterization of a novel, acidic, pH-stable endoglucanase from *Martelella mediterranea*. *J Microbiol* **48**: 393–398.

Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Alexander Richter R, Valas R, *et al.* (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**: 1186–1199.

Dworkin MS, Schwan TG, Anderson DE, Borchardt SM. (2008). Tick-borne relapsing fever. *Infect Dis Clin North Am* **22**: 449–468.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edwards JL, Smith DL, Connolly J, McDonald JE, Cox MJ, Joint I, *et al.* (2010). Identification of carbohydrate ,etabolism genes in the etagenome of a marine biofilm community shown to be dominated by Gammaproteobacteria and Bacteroidetes. *Genes (Basel)* **1**: 371–384.

Environmental Protection Agency I. (2000). Environmental Protection Agency landfill manuals: Landfill site design. Wexford. www.epa.ie (Accessed January 21, 2019).

Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, *et al.* (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: e1319.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29-37.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, *et al.* (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Friedmann M, Williams SC, Simpson HJ. (1979). A new enzymatic method for analysis of cellulose in sediments.

García-Armisen T, İnceoğlu Ö, Ouattara NK, Anzil A, Verbanck MA, Brion N, *et al.* (2014). Seasonal variations and resilience of bacterial communities in a sewage polluted urban river Woo PCY (ed). *PLoS One* **9**: e92579.

Gerbase AC, Rowley JT, Heymann DH, Berkley SF, Piot P. (1998). Global prevalence and incidence estimates of selected curable STDs. *Sex Transm Infect* **74 Suppl 1**: S12-6.

Girfoglio M, Rossi M, Cannio R. (2012). Cellulose degradation by *Sulfolobus solfataricus* requires a cell-anchored endo-β-1-4-glucanase. *J Bacteriol* **194**: 5091–100.

González JM, Masuchi Y, Robb FT, Ammerman JW, Maeder DL, Yanagibayashi M, *et al.* (1998). *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa trough. *Extremophiles* **2**: 123–130.

Grogan DW. (1991). Evidence that beta-galactosidase of *Sulfolobus solfataricus* is only one of several activities of a thermostable beta-d-glycosidase. *Appl Environ Microbiol* **57**: 1644–1649.

Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. (2017). Polysaccharide utilization loci: Fueling microbial communities. *J Bacteriol* **199**: e00860-16.

Güllert S, Fischer MA, Turaev D, Noebauer B, Ilmberger N, Wemheuer B, *et al.* (2016). Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol Biofuels* **9**: 121.

Hahnke S, Langer T, Koeck DE, Klocke M. (2016). Description of *Proteiniphilum saccharofermentans* sp. nov., *Petrimonas mucosa* sp. nov. and *Fermentimonas caenicola* gen. nov., sp. nov., isolated from mesophilic laboratory-scale biogas reactors, and emended description of the genus *Proteiniphilum*. *Int J Syst Evol Microbiol* **66**: 1466–1475.

Hamada M, Shibata C, Saitou S, Tamura T, Komaki H, Ichikawa N, *et al.* (2015). Proposal of nine novel species of the genus *Lysinimicrobium* and emended description of the genus *Lysinimicrobium*. *Int J Syst Evol Microbiol* **65**: 4394–4402.

Hamilton C. (2012). From the dump to the dinner table. *Jamaica Obs*. http://www.jamaicaobserver.com/news/From-the-dump-to-your-dinner-table_10863928.

Hanajima D, Aoyagi T, Hori T. (2015). Survival of free-living Acholeplasma in aerated pig manure slurry revealed by ^{13}C-labeled bacterial biomass probing. *Front Microbiol* **6**: 1206.

Harwati TU, Kasai Y, Kodama Y, Susilaningsih D, Watanabe K. (2009). *Tropicimonas isoalkanivorans* gen. nov., sp. nov., a branched-alkane-degrading bacterium isolated from Semarang Port in Indonesia. *Int J Syst Evol Microbiol* **59**: 388–391.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH, *et al.* (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* **5**: 14567.

Hirayama H, Fuse H, Abe M, Miyazaki M, Nakamura T, Nunoura T, *et al.* (2013). *Methylomarinum vadi* gen. nov., sp. nov., a methanotroph isolated from two distinct marine environments. *Int J Syst Evol Microbiol* **63**: 1073–1082.

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, *et al.* (2008). Genome of an endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. *Science (80-)* **322**: 1108–1109.

Huang C-Y, Patel BK, Mah RA, Baresi L. (1998). *Caldicellulosiruptor owensensis* sp. nov., an anaerobic, extremely thermophilic, xylanolytic bacterium. *Int J Syst Bacteriol* **48**: 91–97.

Huang L-N, Zhou H, Zhu S, Qu L-H. (2004). Phylogenetic diversity of bacteria in the leachate of a full-scale recirculating landfill. *FEMS Microbiol Ecol* **50**: 175–183.

Huang Z, Lai Q, Zhang D, Shao Z. (2017). *Agaribacterium haliotis* gen. nov., sp. nov., isolated from abalone faeces. *Int J Syst Evol Microbiol* **67**: 3819–3823.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, *et al.* (2016). A new view of the tree of life. *Nat Microbiol* **1**: 16048.

Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, *et al.* (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.

Hungate RE. (1950). The anaerobic mesophilic cellulolytic bacteria. *Bacteriol Rev* **14**: 1–49.

Hwang CY, Bae GD, Yih W, Cho BC, Byung Cho CC. (2009). *Marivita cryptomonadi*s gen. nov., sp. nov. and *Marivita litorea* sp. nov., of the family Rhodobacteraceae, isolated from marine habitats. *Int J Syst Evol Microbiol* **59**: 1568–1575.

Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223–2230.

Inagaki F, Suzuki M, Takai K, Oida H, Sakamoto T, Aoki K, *et al.* (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. *Appl Environ Microbiol* **69**: 7224–35.

Intelligence S. (1860). American Journal of Science and Arts. *Am J Sci Arts* **29**: 272.

Jackson, R.E., Lesage S, Priddle MW, Crowe AS, S. S. (1991). Contaminant hydrogeology of toxic organic chemicals at a disposal site, Gloucester, Ontario 2: Remedial investigation. Sci. Ser. Environment Canada, Inland Waters Directorate: Ottawa.

Jeon E, Eun Hyeon J, Eun LS, Park B-S, Kim SW, Lee J, *et al.* (2009). Cellulosic alcoholic fermentation using recombinant *Saccharomyces cerevisiae* engineered for the production of *Clostridium cellulovorans* endoglucanase and *Saccharomycopsis fibuligera* β-glucosidase. *FEMS Microbiol Lett* **301**: 130–136.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, *et al.* (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). https://github.com/najoshi/sickle.

Julliand V, de DVaux A, Milet L, Fonty G. (1999). Identification of *Ruminococcus flavefaciens* as the predominant cellulolytic bacterial species of the equine cecum. *Appl Env Microbiol* **65**: 3738–3741.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109-14.

Karita S, Sakka K, Ohmiya K. (1996). Cellulose-binding domains confer an enhanced activity against insoluble cellulose to *Ruminococcus albus* endoglucanase IV. *J Ferment Bioeng* **81**: 553–556.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Drummond A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Khan AW, Meek E, Sowden LC, Colvin JR. (1994). Emendation of genus *Acetivibrio* and description of *Acetivibrio cellulosolvens*, new species, of nonmotile cellulolytic mesophile. *Int J Syst Bacteriol* **34**: 410–422.

Kim D-U, Correspondence J-OK, Ka J-O. (2003). *Roseomonas soli* sp. nov., isolated from an agricultural soil cultivated with Chinese cabbage (*Brassica campestris*). *Int J Syst Evol Microbiol* **64**: 1024–1029.

Kim H-W, Ishikawa K. (2010a). Complete saccharification of cellulose at high temperature using endocellulase and β-glucosidase from *Pyrococcus* sp. *J Microbiol Biotechnol* **20**: 889–892.

Kim H-W, Ishikawa K. (2011). Functional analysis of hyperthermophilic endocellulase from *Pyrococcus horikoshii* by crystallographic snapshots. *Biochem J* **437**: 223–230.

Kim H-W, Ishikawa K. (2010b). Structure of hyperthermophilic endocellulase from *Pyrococcus horikoshii*. *Proteins Struct Funct Bioinforma* **78**: 496–500.

Kim H-W, Kataoka M, Ishikawa K. (2012). Atomic resolution of the crystal structure of the hyperthermophilic family 12 endocellulase and stabilizing role of the DxDxDG calcium-binding motif in *Pyrococcus furiosus*. *FEBS Lett* **586**: 1009–1013.

King CE, King GM. (2014). Description of *Thermogemmatispora carboxidivorans* sp. nov., a carbon-monoxideoxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int J Syst Evol Microbiol* **64**: 1244–1251.

Kisary J, El-Ebeedy AA, Stipkovits L. (1976). *Mycoplasma* infection of geese II. Studies on pathogenicity of mycoplasmas in goslings and goose and chicken embryos. *Avian Pathol* **5**: 15–20.

Kjeldsen P, Barlaz MA, Rooker AP, Baun A, Ledin A, Christensen TH. (2002). Present and long-term composition of MSW landfill leachate: A review. *Crit Rev Environ Sci Technol* **32**: 297–336.

Koeck DE, Pechtl A, Zverlov V V, Schwarz WH. (2014). Genomics of cellulolytic bacteria. *Curr Opin Biotechnol* **29**: 171–183.

Korpole S, Sharma R, Verma D. (2011). Characterization and phylogenetic diversity of carboxymethyl cellulase producing *Bacillus* species from a landfill ecosystem. *Indian J Microbiol* **51**: 531–5.

Kudo A N H, Cheng DK-J, Costerton JW. (1987). Interactions between *Treponema bryantii* and cellulolytic bacteria in the in vitro degradation of straw cellulose. *Can J Microbiol* **33**: 244–248.

Kuhad RC, Gupta R, Singh A. (2011). Microbial cellulases and their industrial applications. *Enzyme Res* **2011**: 280696.

Kumar AK, Sharma S. (2017). Recent updates on different methods of pretreatment of lignocellulosic feedstocks: a review. *Bioresour Bioprocess* **4**: 7.

Kumari H, Kumar Gupta S, Jindal S, Katoch P, Correspondence RL, Lal R. (2009). *Sphingobium lactosutens* sp. nov., isolated from a hexachlorocyclohexane dump site and *Sphingobium abikonense* sp. nov., isolated from oil-contaminated soil. *Int J Syst Evol Microbiol* **59**: 2291–2296.

Laidlaw PP, Elford WJ. (1936). A new group of filterable organisms. *Proc R Soc B Biol Sci* **120**: 292–303.

Langmead B, Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lau SCK, Zhang R, Brodie EL, Piceno YM, Andersen G, Liu W-T. (2013). Biogeography of bacterioplankton in the tropical seawaters of Singapore. *FEMS Microbiol Ecol* **84**: 259–269.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li T, Mazéas L, Sghir A, Leblon G, Bouchez T. (2009b). Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions. *Environ Microbiol* **11**: 889–904.

Liu C, Finegold SM, Song Y, Lawson PA. (2008). Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*, *Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinki*i as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **58**: 1896–1902.

Lo N, Tokuda G, Watanabe H. (2010). Evolution and function of endogenous termite cellulases. In: *Biology of Termites: a Modern Synthesis*. Springer Netherlands: Dordrecht, pp 51–67.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495.

López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P, Gessner MO, *et al.* (2016). Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci Rep* **6**: 25279.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, *et al.* (2008). How biotech can transform biofuels. *Nat Biotechnol* **26**: 169–172.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**: 506–577.

Maejima K, Oshima K, Namba S. (2014). Exploring the phytoplasmas, plant pathogenic bacteria. *J Gen Plant Pathol* **80**: 210–221.

Masngut N, Manap S. (2017). Bacteria isolation from landfill for production of industrial enzymes for waste degradation. *Indian J Sci Technol* **10**: 1–5.

Mathews SL, Pawlak J, Grunden AM. (2015). Bacterial biodegradation and bioconversion of industrial lignocellulosic streams. *Appl Microbiol Biotechnol* **99**: 2939–2954.

Maymó-Gatell X, Chien Y, Gossett JM, Zinder SH. (1997). Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**: 1568–71.

McClung CR, Patriquin DG. (1980). Isolation of a nitrogen-fixing *Campylobacter* species from the roots of *Spartina alterniflora* Loisel. *Can J Microbiol* **26**: 881–886.

McDonald AG, Boyce S, Tipton KF. (2009). ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res* **37**: D593–D597.

McDonald JE, Houghton JNI, Rooks DJ, Allison HE, McCarthy AJ. (2012). The microbial ecology of anaerobic cellulose degradation in municipal waste landfill sites: evidence of a role for fibrobacters. *Environ Microbiol* **14**: 1077–1087.

McMurdie PJ, Holmes S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

Miller MA, Pfeiffer W, Schwartz T. (2011). The CIPRES science gateway. In: *Proceedings of the 2011 TeraGrid Conference on Extreme Digital Discovery - TG '11*. ACM Press: New York, New York, USA, p 1.

National Environment and Planning Agency. (2010). State of the environment report. http://nepa.gov.jm/new/media_centre/publications/docs/State-of-The-Environment-Report-2010-Jamaica.pdf.

National Solid Waste Management Authority. (2013). Waste characterization and per capita generation rate report 2013. http://www.nswma.gov.jm/wastegenerationandcompositionsurveyreport%0A2013.pdf (Accessed November 21, 2018).

Nishino SF, Spain JC. (1995). Oxidative pathway for the biodegradation of nitrobenzene by *Comamonas* sp. strain JS765. *Appl Environ Microbiol* **61**: 2308–2313.

Niu L, Song L, Liu X, Dong X. (2009). *Tepidimicrobium xylanilyticum* sp. nov., an anaerobic xylanolytic bacterium, and emended description of the genus *Tepidimicrobium*. *Int J Syst Evol Microbiol* **59**: 2698–2701.

Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, *et al.* (2015). Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proc Natl Acad Sci U S A* **112**: 10224–30.

Ohmiya K, Sakka K, Karita S, Kimura T. (1997). Structure of cellulases and their applications. *Biotechnol Genet Eng Rev* **14**: 365–414.

Ontario. Landfill gas capture : a guideline on the regulatory and approval requirements for landfill gas | Ontario.ca. https://www.ontario.ca/page/landfill-gas-capture-guideline-regulatory-and-approval-requirements-landfill-gas (Accessed January 17, 2019).

Ontiveros-Valencia A, Ilhan ZE, Kang D-W, Rittmann B, Krajmalnik-Brown R. (2013). Phylogenetic analysis of nitrate- and sulfate-reducing bacteria in a hydrogen-fed biofilm. *FEMS Microbiol Ecol* **85**: 158–167.

Palmisano A, Barlaz M. (1996). Microbiology of Solid Waste. 1st ed. CRC Press: Boca Raton.

Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, *et al.* (2015). Fungal cellulases. *Chem Rev* **115**: 1308–1448.

Peng Y, Leung HCM, Yiu SM, Chin FYL. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Pérez J, Muñoz-Dorado J, De La Rubia T, Martínez J. (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *Int Microbiol* 53–63.

Pernthaler J, Amann R. (2005). Fate of heterotrophic microbes in pelagic habitats: Focus on populations. *Microbiol Mol Biol Rev* **69**: 440–61.

Planning Institute of Jamaica SD and RPD. (2007). Management of hazardous and solid wastes in Jamaica.
https://www.pioj.gov.jm/Portals/0/Sustainable_Development/Management_of_Wastes.pdf (Accessed November 21, 2018).

Podosokorskaya OA, Bonch-Osmolovskaya EA, Novikov AA, Kolganova TV, Kublanov I. (2013a). *Ornatilinea apprima* gen. nov., sp. nov., a cellulolytic representative of the class Anaerolineae. *Int J Syst Evol Microbiol* **63**: 86–92.

Podosokorskaya OA, Kadnikov V V., Gavrilov SN, Mardanov A V., Merkel AY, Karnachuk O V., *et al.* (2013b). Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria , and a proposal of a novel bacterial phylum Ignavibacteriae. *Environ Microbiol* **15**: 1759–1771.

Pohlsehroeder M, Leschine SB, Canale-Parola E. (1994). *Spirochaeta caldaria* sp. nov., a thermophilic bacterium that enhances cellulose degradation by *Clostridium thermocellum*. *Arch Microbiol* **161**: 17–24.

Pourcher A-M, Sutra L, Hébé I, Moguedet G, Bollet C, Simoneau P, *et al.* (2001). Enumeration and characterization of cellulolytic bacteria from refuse of a landfill. *FEMS Microbiol Ecol* **34**: 229–241.

Pourramezan Z, Ghezelbash GR, Romani B, Ziaei S, Hedayatkhah A. (2012). Screening and identification of newly isolated cellulose-degrading bacteria from the gut of xylophagous termite Microcerotermes diversus (Silvestri). *Microbiology* **81**: 736–742.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Rafael R. de la Haba, Arahal DR, Sánchez-Porro C, Ventosa A. (2006). The family Halomonadaceae. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. Springer, Berlin, Heidelberg, pp 811–835.

Ransom-Jones E, McCarthy AJ, Haldenby S, Doonan J, McDonald JE. (2017). Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. *mSphere* **2**: e00300-17.

Remmas N, Melidis P, Voltsi C, Athanasiou D, Ntougias S. (2017a). Novel hydrolytic extremely halotolerant alkaliphiles from mature landfill leachate with key involvement in maturation process. *J Environ Sci Heal Part A* **52**: 64–73.

Remmas N, Roukouni C, Ntougias S. (2017b). Bacterial community structure and prevalence of *Pusillimonas*-like bacteria in aged landfill leachate. *Environ Sci Pollut Res* **24**: 6757–6769.

Riedel T, Fiebig A, Han J, Huntemann M, Spring S, Petersen J, *et al.* (2014). Genome sequence of the *Wenxinia marina* type strain (DSM 24838T), a representative of the Roseobacter group isolated from oilfield sediments. *Stand Genomic Sci* **9**: 855–865.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Rowe K, VanGulck JF, Millward SC. (2002). Biologically induced clogging of a granular medium permeated with synthetic leachate. *J Environ Eng Sci* **1**: 135–156.

RStudio Team. (2015). RStudio: Integrated development for R. http://www.rstudio.com/.

Russell JB, Muck RE, Weimer PJ. (2009). Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiol Ecol* **67**: 183–197.

Sabree ZL, Moran NA. (2014). Host-specific assemblages typify gut microbial communities of related insect species. *Springerplus* **3**: 138.

Saiki T, Kobayashi Y, Kawagoe K, Beppu T. (1985). *Dictyoglomus thermophilum* gen. nov., sp. nov., a chemoorganotrophic, anaerobic, thermophilic bacterium. *Int J Syst Evol Microbiol* **27**: 253–259.

Sajith S, Priji P, Sreedevi S, Benjamin S. (2016). An overview on fungal cellulases with an industrial perspective. *J Nutr Food Sci* **6**: 1–13.

Scott IM, Rubinstein GM, Lipscomb GL, Basen M, Schut GJ, Rhaesa AM, *et al.* (2015). A new class of tungsten-containing oxidoreductase in Caldicellulosiruptor, a genus of plant biomass-degrading thermophilic bacteria. *Appl Environ Microbiol* **81**: 7339–7347.

Seo JK, Park TS, Kwon IH, Piao MY, Lee CH, Ha JK. (2013). Characterization of cellulolytic and xylanolytic enzymes of *Bacillus licheniformis* JK7 isolated from the rumen of a native korean goat. *Asian-Australasian J Anim Sci* **26**: 50–58.

Sharma S, Yazdani SS. (2016). Chapter 6 – Diversity of microbial cellulase system. In: Gupta VK (ed). *New and Future Developments in Microbial Biotechnology and Bioengineering: Microbial Cellulase System Properties and Applications*. Elsevier: Amsterdam, pp 49–64.

Shi R, Li Z, Ye Q, Xu J, Liu Y. (2013). Heterologous expression and characterization of a novel thermo-halotolerant endoglucanase Cel5H from *Dictyoglomus thermophilum*. *Bioresour Technol* **142**: 338–344.

Shieh WY, Liu TY, Lin SY, Jean WD, Chen J-S. (2008). *Simiduia agarivorans* gen. nov., sp. nov., a marine, agarolytic bacterium isolated from shallow coastal water from Keelung, Taiwan. *Int J Syst Evol Microbiol* **58**: 895–900.

Shiratori H, Ikeno H, Ayame S, Kataoka N, Miya A, Hosono K, *et al.* (2006). Isolation and characterization of a new *Clostridium* sp. that performs effective cellulosic waste digestion in a thermophilic methanogenic bioreactor. *Appl Environ Microbiol* **72**: 3702–9.

Sigma-Aldrich. Whatman® Cellulose Filter Paper - Whatman® Filter Paper | Sigma-Aldrich. https://www.sigmaaldrich.com/labware/labware-products.html?TablePage=17207008 (Accessed November 7, 2018).

Sik Baik K, Park SC, Choe HN, Kim SN, Moon J-H, Seong CN. (2012). *Roseomonas riguiloci* sp. nov., isolated from wetland freshwater. *Int J Syst Evol Microbiol* **62**: 3024–3029.

Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, *et al.* (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* **11**: 1483–1499.

Slobodkin AI, Tourova TP, Kostrikina NA, Lysenko AM, German KE, Bonch-Osmolovskaya EA, *et al.* (2006). *Tepidimicrobium ferriphilum* gen. nov., sp. nov., a novel moderately thermophilic, Fe(III)-reducing bacterium of the order Clostridiales. *Int J Syst Evol Microbiol* **56**: 369–372.

Song L, Wang Y, Tang W, Lei Y. (2015a). Bacterial community diversity in municipal waste landfill sites. *Appl Microbiol Biotechnol* **99**: 7745–7756.

Song L, Wang Y, Zhao H, Long DT. (2015b). Composition of bacterial and archaeal communities during landfill refuse decomposition processes. *Microbiol Res* **181**: 105–111.

Sorokin DY, Van Pelt S, Tourova TP, Evtushenko LI, Dimitry C, Sorokin Y. (2009). *Nitriliruptor alkaliphilus* gen. nov., sp. nov., a deeplineage haloalkaliphilic actinobacterium from

soda lakes capable of growth on aliphatic nitriles, and proposal of Nitriliruptoraceae fam. nov. and Nitriliruptorales ord. nov. *Int J Syst Evol Microbiol* **59**: 248–253.

Staley BF, de los Reyes FL, Barlaz MA. (2012). Comparison of bacteria and archaea communities in municipal solid waste, individual refuse components, and leachate. *FEMS Microbiol Ecol* **79**: 465–473.

Stamatakis A. (2014). RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stamps BW, Lyles CN, Suflita JM, Masoner JR, Cozzarelli IM, Kolpin DW, *et al.* (2016). Municipal solid waste landfills harbor distinct microbiomes. *Front Microbiol* **7**: 534.

Stantec Consulting Ltd. (2015). Waterloo Sanitary Master Plan: Volume 1. Waterloo. https://www.waterloo.ca/en/contentresources/resources/government/rpt_MP_final_20150820.pdf (Accessed September 5, 2018).

Stanton TB, Canale-Parola E. (1980). *Treponema bryantii* sp. nov., a rumen Spirochete that interacts with cellulolytic bacteria. *Arch Microbiol* **127**: 145–156.

Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, *et al.* (2011). The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist Li W (ed). *PLoS One* **6**: e18814.

Sukharnikov LO, Alahuhta M, Brunecky R, Upadhyay A, Himmel ME, Lunin V V, *et al.* (2012). Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. *J Biol Chem* **287**: 41068–77.

Sukharnikov LO, Cantwell BJ, Podar M, Zhulin IB. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol* **29**: 473–479.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**: 1282–1288.

Suzuki M, Nakagawa Y, Harayama' S, Yamarnotol S. (1999). Phylogenetic analysis of genus *Marinilabilia* and related bacteria based on the amino acid sequences of GyrB and emended description of *Marinilabilia salmonicolor* with *Marinilabilia agarovorans* as its subjective synonym. *Int J Syst Bacteriol* **49**: 1551–1557.

Teather RM, Wood PJ. (1982). Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* **43**: 777–780.

Thoetkiattikul H, Mhuantong W, Pinyakong O, Wisawapipat W, Yamazoe A, Fujita N, *et al.* (2017). Culture-independent study of bacterial communities in tropical river sediment. *Biosci Biotechnol Biochem* **81**: 200–209.

Thompson CE, Beys-da-Silva WO, Santi L, Berger M, Vainstein MH, Guima Rães JA, *et al.* (2013). A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. *AMB Express* **3**: 65.

Tokuda G, Lo N, Watanabe H. (2005). Marked variations in patterns of cellulase activity against crystalline- vs. carboxymethyl-cellulose in the digestive systems of diverse, wood-feeding termites. *Physiol Entomol* **30**: 372–380.

Tomazetto G, Hahnke S, Wibberg D, Pühler A, Klocke M, Schlüter A. (2018). *Proteiniphilum saccharofermentans* str. M3/6T isolated from a laboratory biogas reactor is versatile in polysaccharide and oligopeptide utilization as deduced from genome-based metabolic reconstructions. *Biotechnol Reports* **18**: e00254.

U.S. Environmental Protection Agency. (2018). Advancing sustainable materials management: 2015 fact sheet. Assessing trends in material generation, recycling, composting, combustion with energy recovery and landfilling in the United States. https://www.epa.gov/sites/production/files/2018-07/documents/2015_smm_msw_factsheet_07242018_fnl_508_002.pdf (Accessed November 18, 2018).

Ue H, Matsuo Y, Kasai H, Yokota A. (2011). *Demequina globuliformis* sp. nov., *Demequina oxidasica* sp. nov. and *Demequina aurantiaca* sp. nov., actinobacteria isolated from marine environments, and proposal of Demequinaceae fam. nov. *Int J Syst Evol Microbiol* **61**: 1322–1329.

Vandamme P, Falsen E, Rossau R, Hoste B, Segers P, Tytgat R, *et al.* (1991). Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: Emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *Int J Syst Bacteriol* **41**: 88–103.

Várnai A, Siika-aho M, Viikari L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnol Biofuels* **6**: 30.

Walker E, Warren FL. (1938). Decomposition of cellulose by *Cytophaga*. *Biochem J* **32**: 31–43.

Wang C-M, Shyu C-L, Ho S-P, Chiou S-H. (2008). Characterization of a novel thermophilic, cellulose-degrading bacterium *Paenibacillus* sp. strain B39. *Lett Appl Microbiol* **47**: 46–53.

Wang L, Hatem A, Catalyurek U V, Morrison M, Yu Z. (2013). Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One* **8**: e78507.

Wang L, Nie Y, Tang Y-Q, Song X-M, Cao K, Sun L-Z, *et al.* (2016). Diverse bacteria with lignin degrading potentials isolated from two ranks of coal. *Front Microbiol* **7**: 1428.

Wang S, Zhao D, Bai X, Zhang W, Lu X. (2017a). Identification and characterization of a large protein essential for degradation of the crystalline region of cellulose by *Cytophaga hutchinsonii*. *Appl Environ Microbiol* **83**: e02270-16.

Wang X, Cao A, Zhao G, Zhou C, Xu R. (2017b). Microbial community structure and diversity in a municipal solid waste landfill. *Waste Manag* **66**: 79–87.

Westlake K, Archer DB, Boone DR. (1995). Diversity of cellulolytic bacteria in landfill. *J Appl Bacteriol* **79**: 73–78.

Wickham H. (2016). ggplot2: Elegant graphics for data analysis.

Wilhelm RC, Singh R, Eltis LD, Mohn WW. (2019). Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J* **13**: 413–429.

Willems A, De Ley J, Gillis M, Kersters DK. (1991). Comamonadaceae, a new family encompassing the acidovorans rRNA complex, including *Variovorax paradoxus* gen. nov. , comb. nov. for *Alcaligenes paradoxus* (Davis 1969). *Int J Syst Evol Microbiol* **41**: 445–450.

Wood TM, Bhat KM. (1988). Methods for measuring cellulase activities. *Methods Enzym* **160**: 87–117.

World Health Organization. (1999). Safe management of wastes from health-care activities. World Health Organization Geneva.

Xie B, Xiong S, Liang S, Hu C, Zhang X, Lu J. (2012). Performance and bacterial compositions of aged refuse reactors treating mature landfill leachate. *Bioresour Technol* **103**: 71–77.

Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P, *et al.* (2007). Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol* **73**: 3536–46.

Xue X, Zhang K, Cai F, Dai J, Wang Y, Rahman E, *et al.* (2012). *Altererythrobacter xinjiangensis* sp. nov., isolated from desert sand, and emended description of the genus *Altererythrobacter*. *Int J Syst Evol Microbiol* **62**: 28–32.

Yoon J-H, Kang S-J, Lee S-Y, Lee J-S, Correspondence SP. (2011). *Ohtaekwangia koreensis* gen. nov., sp. nov. and *Ohtaekwangia kribbensis* sp. nov., isolated from marine sand, deep-branching members of the phylum Bacteroidetes. *Int J Syst Evol Microbiol* **61**: 1006–1072.

Yuan C-G, Chen X, Jiang Z, Chen W, Liu L, Xian W-D, *et al.* (2017). *Altererythrobacter lauratis* sp. nov. and *Altererythrobacter palmitatis* sp. nov., isolated from a Tibetan hot spring. *Antonie Van Leeuwenhoek* **110**: 1077–1086.

Yue K, Peng C, Yang W, Peng Y, Zhang C, Huang C, *et al.* (2016). Degradation of lignin and cellulose during foliar litter decomposition in an alpine forest river. *Ecosphere* **7**: e01523.

Appendices

Appendix A

Given sample names used in thesis corresponding to wells in the southern Ontario landfill with their sampling dates

CLC1_T1: Pumphouse 3 (PH3) sampled on July 14, 2016
CLC1_T2: Pumphouse 3 (PH3) sampled on July 20, 2016
LW1: LW64_88 sampled on July 20, 2016
LW2: LW138R sampled on July 20, 2016
LW3: LW168 sampled on July 20, 2016
GW1: OW334 sampled on July 20, 2016

Appendix B

Table B1. Gene counts and relative abundance of genes from glycosyl hydrolase families containing cellulases (pre-filtered dataset) representing the total cellulolytic potential in the sampled sites at the Ontario landfill. RA = relative abundance within all GHs identified.

GH family	CLC1_T1	CLC1_T2	LW1	LW2	LW3	GW1
GH1	385	295	194	182	327	305
RA (per site)	0.06	0.06	0.11	0.07	0.15	0.12
GH3	2463	2030	784	1161	626	842
RA (per site)	0.40	0.43	0.46	0.47	0.29	0.33
GH5	2394	1688	332	752	753	957
RA (per site)	0.39	0.36	0.20	0.31	0.35	0.38
GH6	0	0	0	8	3	31
RA (per site)	0.00	0.00	0.00	0.00	0.00	0.01
GH8	146	110	45	49	30	39
RA (per site)	0.02	0.02	0.03	0.02	0.01	0.02
GH9	389	316	161	161	254	105
RA (per site)	0.06	0.07	0.10	0.07	0.12	0.04
GH12	0	0	0	1	0	1
RA (per site)	0.00	0.00	0.00	0.00	0.00	0.00
GH30	261	188	103	69	81	176
RA (per site)	0.04	0.04	0.06	0.03	0.04	0.07
GH44	21	20	28	2	12	12
RA (per site)	0.00	0.00	0.02	0.00	0.01	0.00
GH45	24	16	7	1	0	0
RA (per site)	0.00	0.00	0.00	0.00	0.00	0.00
GH48	0	3	0	3	0	0
RA (per site)	0.00	0.00	0.00	0.00	0.00	0.00
GH116	70	63	35	58	41	56
RA (per site)	0.01	0.01	0.02	0.02	0.02	0.02

Table B2. Gene counts and relative abundance of genes annotated as types of cellulases (screened dataset) representing the specific cellulolytic potential in the sampled sites at the Ontario landfill. RA = relative abundance.

Type	CLC1_T1	RA	CLC1_T2	RA	LW1	RA	LW2	RA	LW3	RA	GW1	RA
Endocellulase	276	0.0523	209	0.0522	37	0.0250	75	0.0400	73	0.0422	92	0.0499
Exocellulase	1	0.0002	6	0.0015	0	0.0000	7	0.0037	9	0.0052	10	0.0054
Beta-glucosidase	941	0.1782	723	0.1805	275	0.1856	363	0.1937	342	0.1979	163	0.0885
Cellulase	108	0.0205	80	0.0200	28	0.0189	54	0.0288	67	0.0388	42	0.0228
Glycosyl hydrolase	2182	0.4132	1699	0.4241	729	0.4919	733	0.3911	670	0.3877	728	0.3952
Hypothetical protein	1773	0.3357	1289	0.3218	413	0.2787	642	0.3426	567	0.3281	807	0.4381
Total	5281	-	4006	-	1482	-	1874	-	1728	-	1842	-

Table B3. Gene count and relative abundance of putative cellulase genes (filtered) across the sites sampled at the southern Ontario landfill. Highest-level of identifiable taxonomic group and proteobacterial classes at <1% relative abundance are categorized as Rare Phyla, with all rare phylum abundances summed. CLC1 = Composite leachate cistern, LW = Leachate well, GW = Groundwater well, T# = Time point, RA = relative abundance.

Taxonomic group	CLC1_T1	RA	CLC1_T2	RA	LW1	RA	LW2	RA	LW3	RA	GW1	RA
Bacteroidetes	2563	0.4862	1829	0.4579	533	0.3913	772	0.3779	516	0.3239	688	0.3762
Firmicutes	1187	0.2252	927	0.2321	266	0.1953	424	0.2075	171	0.1073	121	0.0662
Fibrobacteres	485	0.0920	322	0.0806	105	0.0771	18	0.0088	1	0.0006	2	0.0011
Verrucomicrobia	251	0.0476	140	0.0351	42	0.0308	63	0.0308	152	0.0954	169	0.0924
Spirochaetes	199	0.0377	174	0.0436	57	0.0419	62	0.0303	26	0.0163	37	0.0202
Actinobacteria	98	0.0186	60	0.0150	26	0.0191	71	0.0348	50	0.0314	134	0.0733
Alphaproteobacteria	32	0.0061	17	0.0043	22	0.0162	87	0.0426	93	0.0584	75	0.0410
Gammaproteobacteria	72	0.0137	45	0.0113	30	0.0220	56	0.0274	67	0.0421	55	0.0301
Deltaproteobacteria	45	0.0085	55	0.0138	55	0.0404	40	0.0196	58	0.0364	58	0.0317
Dictyoglomi	22	0.0042	27	0.0068	1	0.0007	69	0.0338	78	0.0490	90	0.0492
Chloroflexi	30	0.0057	42	0.0105	35	0.0257	46	0.0225	71	0.0446	42	0.0230
Planctomycetes	62	0.0118	36	0.0090	8	0.0059	37	0.0181	40	0.0251	72	0.0394
Tenericutes	53	0.0101	77	0.0193	18	0.0132	73	0.0357	3	0.0019	1	0.0005
Ignavibacteriae	13	0.0025	17	0.0043	27	0.0198	14	0.0069	71	0.0446	69	0.0377
Betaproteobacteria	15	0.0028	10	0.0025	10	0.0073	43	0.0210	58	0.0364	33	0.0180
Cyanobacteria	14	0.0027	11	0.0028	18	0.0132	15	0.0073	38	0.0239	37	0.0202
Acidobacteria	14	0.0027	18	0.0045	10	0.0073	27	0.0132	19	0.0119	24	0.0131
Rare taxonomic group	302	0.0573	232	0.0581	89	0.0653	137	0.0671	85	0.0534	125	0.0683
Thermogotae	23	0.0044	49	0.0123	11	0.0081	36	0.0176	11	0.0069	3	0.0016
Epsilonproteobacteria	17	0.0032	42	0.0105	0	0.0000	18	0.0088	7	0.0044	2	0.0011
Armatimonadetes	16	0.0030	14	0.0035	11	0.0081	7	0.0034	4	0.0025	17	0.0093
Balneolaeota	3	0.0006	9	0.0023	25	0.0184	6	0.0029	6	0.0038	5	0.0027
Deinococcus-Thermus	5	0.0009	9	0.0023	12	0.0088	8	0.0039	10	0.0063	5	0.0027
unclassified Haloplasmatales	15	0.0028	20	0.0050	5	0.0037	7	0.0034	1	0.0006	1	0.0005
Euryarchaeota	8	0.0015	10	0.0025	5	0.0037	8	0.0039	6	0.0038	6	0.0033
Chlorobi	2	0.0004	4	0.0010	14	0.0103	12	0.0059	2	0.0013	9	0.0049
unclassified bacteria	17	0.0032	13	0.0033	3	0.0022	2	0.0010	2	0.0013	3	0.0016
Elusimicrobia	3	0.0006	0	0.0000	8	0.0059	0	0.0000	1	0.0006	18	0.0098
Candidatus Kryptonia	0	0.0000	0	0.0000	2	0.0015	0	0.0000	5	0.0031	17	0.0093
Gemmatimonadetes	1	0.0002	0	0.0000	0	0.0000	2	0.0010	1	0.0006	13	0.0071
Calditrichaeota	1	0.0002	3	0.0008	1	0.0007	2	0.0010	4	0.0025	6	0.0033
Kiritimatiellaeota	0	0.0000	1	0.0003	0	0.0000	4	0.0020	2	0.0013	10	0.0055
Crenarchaeota	0	0.0000	3	0.0008	0	0.0000	1	0.0005	6	0.0038	0	0.0000
Chrysiogenetes	2	0.0004	4	0.0010	1	0.0007	1	0.0005	1	0.0006	0	0.0000
Rhodothermaeota	0	0.0000	1	0.0003	0	0.0000	1	0.0005	4	0.0025	2	0.0011
Enterobacteria	2	0.0004	1	0.0003	0	0.0000	2	0.0010	1	0.0006	1	0.0005
Chlamydiae	0	0.0000	0	0.0000	0	0.0000	0	0.0000	3	0.0019	3	0.0016

Table B1. continued

Rare taxonomic group	CLC1_T1	RA	CLC1_T2	RA	LW1	RA	LW2	RA	LW3	RA	GW1	RA
Synergistetes	0	0.0000	1	0.0003	0	0.0000	4	0.0020	0	0.0000	0	0.0000
Candidatus Saccharibacteria	0	0.0000	0	0.0000	0	0.0000	2	0.0010	2	0.0013	0	0.0000
Lentisphaerae	1	0.0002	1	0.0003	0	0.0000	1	0.0005	0	0.0000	0	0.0000
Fusobacteria	1	0.0002	1	0.0003	1	0.0007	0	0.0000	0	0.0000	0	0.0000
Deferribacteres	0	0.0000	1	0.0003	0	0.0000	0	0.0000	0	0.0000	1	0.0005
Nitrospirae	0	0.0000	0	0.0000	0	0.0000	1	0.0005	0	0.0000	0	0.0000

Table C1. Comparison of compositions of synthetic leachate recipes**.** Synthetic leachate composition was modified from (Rowe *et al.*, 2002) based on leachate compositions from the surveyed leachate wells from the Ontario landfill.

Synthetic leachate				Rowe et al., 2002	
	Amount	Unit (/L)	Deionized H_2O	Component	Unit(/L)
Solution 1*			*		
Microcrystalline cellulose	10	g		N/A	N/A
Solution 2			200 ml		
$CaCl_2$ x $2H_2O$	870	mg		$CaCl_2$	2,882 mg
Solution 3			200 ml		
$MgSO_4$	54	mg		-	156 mg
$MgCl_2$ x $6H_2O$	1,083	mg		-	3,114 mg
Solution 4			100 ml		
K_2HPO_4	30	mg		-	30
Solution 5			500 ml		
$KHCO_3$	312	mg		-	NC
K_2CO_3	324	mg		-	NC
NaCl	745	mg		-	1,440 mg
$NaHCO_3$	1,558	mg		-	3,012 mg
$NaNO_3$	26	mg		-	50 mg
NH_4HCO_3	1,430	mg		-	2,439 mg
$CO(NH_2)_2$ (urea)	407	mg		-	695 mg
Solution 7 - Metal Stock Solution	1	ml		-	1 ml
Solution 6**			* *		
Agar	15	g		N/A	N/A
Solution 7 - Metal Stock Solution			top up to 1 L	Distilled H_2O	top up to 1L
$Al_2(SO_4)_3$ x 16 H_2O	30	mg		-	NC
$CoSO_4$ x 5 H_2O	150	mg		$CoSO_4$ x 7 H_2O	-
$CuSO_4$ x 5 H_2O	40	mg		-	-
$FeSO_4$	4,000	mg		-	2,000 mg
H_3BO_3	19,446	mg		-	50 mg
$MnSO_4$ x H_2O	2,453	mg		-	500 mg
$(NH_4)_6Mo_7O_{24}$ x 4 H_2O	50	mg		-	NC
$NiSO_4$ x 6 H_2O	500	mg		-	NC
$ZnSO_4$ x 7 H_2O	50	mg		-	NC
96% concentrated H_2SO4	2.3	ml		-	1 ml

- Same component used in the Ontario leachate (column 1)

N/A Not applicable

NC Component amount is the same in both leachate protocol

* Only added when making synthetic leachate amended with microcrystalline cellulose. Autoclave Solution 1 separately and combine with Solutions 2-6 at room temperature. Adjust total volumes of Solutions 1-6 to total 1 L.

** Only added when making synthetic leachate agar media or synthetic leachate with microcrystalline cellulose agar media. Adjust total volume of synthetic leachate to 1 L with the amendment.

The synthetic leachate was adjusted pH to ~6.6-6.8 with HCl.

The synthetic leachate (Rowe et al., 2002) was adjusted to pH 5.8-6.0 with NaOH.

Other modifications to Rowe *et al.*'s protocol:

i) Acetic acid, proprionic acid, and butyric acid were omitted because they would have been additional carbon sources beyond the the microcrystalline cellulose.

ii) Na_2S x $9H_2O$ was omitted because of its toxicity.

iii) All of the components that are not in the Metal Stock Solution (Rowe et al., 2002) are prepared in multiple solutions then combined after autoclaving, instead of mixed together in one solution. This modification was to avoid precipitation.

iv) NaOH was omitted because the microcrystalline cellulose generated a basic initial pH, so HCl was used to adjust the pH instead.

The leachate well chemical composition data was derived from four sampling dates at the Southern Ontario landfill, reported by the site engineers in the 2014 annual report. This information was pooled and used to modify concentrations such that the synthetic leachate more closely matched conditions at the site. The following sampling dates were used:

LW1: April 2013, October 2013, April 2014, October 2014

LW2: October 2012, April 2013, October 2013, April 2014

LW3: April 2013, October 13, April 2014, October 2014