Microbial Composition and Functional Diversity in Waste Environments

by

Lisa Anne Johnson

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Richard Sparling Associate Professor, University of Manitoba
Supervisor(s)	Dr. Laura A. Hug Assistant Professor, University of Waterloo
Internal Members	Dr. Josh D. Neufeld Professor, University of Waterloo
	Dr. Andrew C. Doxey Associate Professor, University of Waterloo
Internal-external Member	Dr. Jenine McCutcheon Assistant Professor, University of Waterloo

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

Chapter 2

Biofilm samples were collected by employees and engineers at the regional waste management division. The bioinformatic pipeline used for the generation of metagenome-assembled genomes (MAGs) was built by Nikhil George.

Metaproteomics and metametabolomics data was generated at the Environmental Molecular Sciences Laboratory (EMSL; Department of Energy, Washington, USA) under an EMSL User Access Grant. Sample preparation for metaproteomics and metametabolomics was completed by Athena Schepmoes. Metaproteome data acquisition was performed by Isaac Kwame Attah, with protein identification and quantification performed by Nikola Tolic. Metametabolome data acquisition was performed by Vanessa Paurus, with metabolite identification and quantification by Chaevien Clendinen.

Chapter 3

A version of this chapter has been published as:

> Johnson, L.A. and Hug, L.A. (2022) Cloacimonadota metabolisms include adaptations in engineered environments that are reflected in the evolutionary history of the phylum. *Environmental Microbiology Reports*. 14: 520–529. https://doi.org/10.1111/1758-2229.13061

Chapter 4

A version of this chapter has been published as:

> Johnson, L.A. and Hug, L.A. (2019) Distribution of reactive oxygen species defense mechanisms across domain Bacteria. *Free Radical Biology & Medicine* 140: 93–102. https://doi.org/10.1016/j.freeradbiomed.2019.03.032

Abstract

Landfilling remains the most common endpoint of discarded waste. At landfill sites, waste is transformed and degraded by microbially mediated processes conducted by different trophic groups, comprised of members of a heterogeneous and diverse microbial community. Resulting metabolic products from these processes commonly include methane (CH_4) and carbon dioxide (CO_2), potent greenhouse gases (GHGs), leading to concerns about the long-term efficacy of landfilling as a waste management practice. These concerns arise over the increasing amount of waste being generated worldwide that will place a strain on landfills and, therefore, impact the level of contamination from landfill sites to the environment, particularly regarding emissions of GHGs. These climate change challenges at landfill sites are exacerbated by the limited understanding of the microbial community composition and functional diversity that are present in these heterogeneous ecosystems. Investigating the functional potential of microbial communities with omics technologies, such as metagenomics and metaproteomics, can clarify microbial lineages and processes impacting waste decomposition, stability, and greenhouse gas management.

This thesis examines microbial populations and their metabolic capacities using metagenomics, with a focus on waste management sites and anoxic environments. First, the microbial community within landfill-associated disruptive biofilms was investigated using a multi-omics approach (Chapter 2). Both 16S rRNA gene profiling and metagenomic datasets identified major biofilm phyla as the Proteobacteria, Firmicutes, Desulfobacterota, Campylobacterota, and Bacteroidota. However, the community composition across biofilm samples was very uneven, and varied at the genus-level across biofilm samples. Metaproteomics showed microbial lineages

highly abundant in genomic datasets were frequently not the most active. Each biofilm sample had unique keystone species that were detected at low abundances in the genomic DNA, but comprised relatively large proportions of the metaproteomic dataset. The most abundant MAG in Biofilm 1 (BF1) from the protein data classified to genus *Pseudodesulfuromonas* (phylum Desulfobacterota), and the most abundant MAG in Biofilm 2 (BF2) was classified to genus *Methanothrix* (phylum Halobacteriota). In Biofilm 4 (BF4) the genus *Thiothrix* represented 0.2% of 16S rRNA gene sequences and the MAG *Thiothrix* BF4 Bin 143 accounted for 0.2% of total coverage in MAGs, but made up 33.7% of the total normalized abundance of proteins identified from BF4, highlighting this bacterium as highly active. Evaluation of the proteins at the highest abundance in each sample revealed methyl-coenzyme M reductase and adenylylsulfate reductase as proteins common to each biofilm. Methane and sulfur cycling were important processes occuring in each biofilm and are potential targets for biofilm disruption.

As an enigmatic phylum with representatives found at landfill sites, the distribution and metabolic capacity of phylum Cloacimonadota was examined (Chapter 3) based on detection in the biofilm samples in Chapter 2. Cloacimonadota were abundant and diverse within metagenomic datasets from a municipal landfill, including biofilm metagenomes. A total of 24 new metagenome-assembled genomes (MAGs) were generated, expanding publicly available genomic datasets (30 genomes) by 80%. Cloacimonadota genomes were examined through phylogenetic relationships, metabolic capacity, and pangenome dynamics. Phylogenetic placement of genomes resulted in two distinct clades segregated by ecosystem. Clade I contained all genomes derived from engineered ecosystems, and Clade II contained the majority of genomes derived from environmental ecosystems. Metabolic reconstructions predict Cloacimonadota to be anaerobic and

acetogenic, with mixed fermentative and anaerobic respiration lifestyles for the majority of Cloacimonadota surveyed. Genomes from engineered ecosystems encode a suite of genes not typically found in genomes from natural environments, including acetate kinase, genes for cysteine degradation to pyruvate, an increased diversity of carbon utilization enzymes, and different mechanisms for generating membrane potential and ATP synthesis. This phylum-level examination also clarifies the distribution of functions previously observed for members of the phylum, demonstrating that propionate oxidation and reverse TCA cycles are not common components of Cloacimonadota metabolism.

Landfills contain heterogeneous and diverse habitats for microorganisms, and this includes both oxic and anoxic microenvironments. Therefore, microorganisms require mechanisms to manage damaging reactive oxygen species (ROS). To understand the distribution of enzymatic defenses against ROS, genomes spanning the domain Bacteria derived from diverse environments were surveyed for ROS defenses (Chapter 4). The diversity and distribution of enzymatic defenses were reviewed using peer-reviewed literature and publicly available genome databases. Distribution of bacterial genomic capacity for enzymatic ROS defenses did not identify gene complements that would allow for prediction of lifestyle (anaerobic or aerobic), or unique patterns of enzymes within and between bacterial phyla. Instead, the emerging trend was that many different strategies are employed. Further, specific strategies used by well-characterized organisms are described to highlight different oxidative stress responses by organisms using aerobic, facultatively anaerobic, and anaerobic lifestyles. The research in this thesis uses metagenomics as the methodological approach to describe microbial diversity and function. The use of metagenomics enabled the first description of landfill-associated biofilms, with a focus on

microbial diversity and key metabolic processes, and additional examination of the enigmatic phylum Cloacimonadota and the distribution of ROS defenses in domain Bacteria.

Acknowledgements

I would like to thank my advisor Dr. Laura Hug for her patience, guidance, and encouragement throughout my doctoral research. It has been an honor to work together. I am grateful for the opportunity to build my skillsets in microbial ecology and bioinformatics.

Thank you to both Dr. Josh Neufeld and Dr. Andrew Doxey, members of my advisory committee. I appreciate that you facilitated a space for me to learn, and for all your advice and dedication to advancing my research. And thank you to Dr. Richard Sparling for providing your expertise as my external examiner.

Thank you to Hug Lab members, both past and present, for your camaraderie and great scientific conversations.

I would like to thank employees and engineers at the waste management site for providing access to sample sites and assistance with sample collection. Thank you to our collaborators at the Environmental Molecular Sciences Laboratory (EMSL) User Facility for their assistance in generating metaproteomic and metametabolomic datasets, and all their invaluable discussions and assistance with analysis of these datasets.

Finally, thank you to my family and friends for all their encouragement, love, and support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA	Auxiliary activity
AGC	Automatic gain control
Ahp	Alkyl hydroperoxide reductase
ANI	Average nucleotide identity
Anvi'o	Analysis and visualization platform for 'omics data
ASV	Amplicon sequence variant
BF	Biofilm
BLAST	Basic local alignment search tool
BLASTp	BLAST-protein
Ca.	*Candidatus*
CAB	Chlorophyll *a/b*-binding
CAZy	Carbohydrate-active enzymes
CBM	Carbohydrate binding modules
CE	Carbohydrate esterases
CH_4	Methane
CLC	Composite leachate cistern
CO_2	Carbon dioxide
COGs	Clusters of orthologous groups
Cu	Copper
DNA	Deoxyribonucleic acid
DRAM	Distilled and Refined Annotation of Metabolism
DRep	Bioinformatic program for dereplicating genomes
EC	Enzyme commission number
EPS	Extracellular polymeric substance
E-value	Expect value
Fe	Iron
FT	Fourier transform
GC	Gas chromatography
GH	Glycoside hydrolases

GHGs	Greenhouse gases
GT	Glycosyltransferases
GTDB	Genome Taxonomy Database
GTDBtk	Genome Taxonomy Database toolkit
H_2	Hydrogen
H_2O_2	Hydrogen peroxide
HCD	Higher-energy collisional dissociation
HMM	Hidden Markov Model
JGI	Joint Genome Institute
Kat	Catalase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG orthology
LFG	Landfill gas
LW	Leachate well
LPS	Lipopolysaccharide
LUCA	Last universal common ancestor
MAG	Metagenome-assembled genome
Mbp	Megabase pair
MEROPS	Online database for peptidases and their inhibitors
Mn	Manganese
MS	Mass spectrometry
MSW	Municipal solid waste
mur	Gene encoding enzyme involved in peptidoglycan synthesis
NADH	Nicotinamide adenine dinucleotide
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
Ni	Nickel
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
NMOCs	Non-methane organic compounds
O_2	Superoxide anion

OH^-	Hydroxyl radical
ORF	Open reading frame
OCP	Orange carotenoid protein
OxyR	Hydrogen peroxide response regulon
PEP	Phosphoenolpyruvate
PFOR	Pyruvate ferredoxin oxidoreductase
PL	Polysaccharide lyases
PTOX	Plastoquinol terminal oxidase
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
ROS	Reactive oxygen species
SOD	Superoxide dismutase
SOR	Superoxide reductase
SoxRS	Superoxide response regulon
sp.	Species
SPE	Solid phase extraction
TCA	Tricarboxylic acid
TPM	Transcripts per million
Tpx	Thiol peroxidase
WWE1	Previous name for phylum Cloacimonadota (from Wastewater treatment plant in Evry, France)
Zn	Zinc

Chapter 1
Introduction to landfill microbiology

1.1 Landfills

1.1.1 Landfill impacts, and greenhouse gas emissions

Landfilling refers to the organized disposal of waste within an engineered environment. There are different ways a landfill can be designed to operate, including bioreactor landfills that encourage microbially mediated waste degradation, or landfills that limit decomposition by restricting exposure to moisture (Omar and Rohani, 2015). Irrespective of landfill design, the decomposition of discarded waste results in the generation of greenhouse gases (GHGs) (Broun and Sattler, 2016). The waste discarded into a landfill is referred to as municipal solid waste (MSW) and typically consists of residential, commercial, and institutional wastes (Hilger and Barlaz, 2006). The microbially mediated degradation of MSW results in the generation of methane (CH_4) and carbon dioxide (CO_2), which are potent GHGs whose release to the atmosphere exacerbates global warming (Hilger and Barlaz, 2006). The gases generated through degradation of MSW are already a concern, but the increasing volume of wastes being generated worldwide will intensify the impact landfills have on the environment.

Globally, growth in industrial production and a shift toward more urban lifestyles have resulted in increasing amounts of waste being generated, with continued growth projected for the foreseeable future (Hoornweg and Bhada-Tata, 2012). In Canada, the amount of solid waste generated increased by 16%, or 4.8 million tonnes, in the twelve years between 2002 and 2018 (Environment and Climate Change Canada, 2018). If there are limited interventions, and society continues to address waste with a "business as usual" perspective, we may not even hit the peak

amount of waste being generated this century (Hoornweg *et al.*, 2013). The increasing MSW will strain the capacity of current landfill sites, with concomitant intensification of GHG production. Even now in the United States, MSW is the third largest anthropogenic source of methane, accounting for 17% of total CH_4 emissions in 2019 (US Environmental Protection Agency, 2021). In the same year, 23% of Canada's CH_4 emissions were attributed to landfills (Environment and Climate Change Canada, 2021).

Management of MSW and subsequent management or diversion of GHGs is imperative for limiting environmental contamination from landfills. Gas collection systems are installed at some landfill sites to facilitate gas management, collecting emitted GHGs and converting CO_2, and CH_4 to biogas for energy generation. The disposal of wastes in Canada produced 37 Mt CO_2 equivalents of CH_4 in 2019, with 62% emitted and 31% captured (Environment and Climate Change Canada, 2021). The global warming potential of CH_4 is 25 times higher than that of CO_2, highlighting the significance of limiting the release of CH_4 to the atmosphere (US Environmental Protection Agency, 2021). Gas collection, as an intervention to mitigate environmental contamination from landfills, is important for reducing CH_4 emissions, and aligns with Canada's support for the Global Methane Pledge to reduce global methane emissions by 30% below 2020 levels by 2030 (Environment and Climate Change Canada, 2022). Ensuring gas collection systems are functioning at full capacity will increase the effectiveness of gas diversion and limit environmental contamination.

1.1.2 Landfills as an engineered, built environment

A typical municipal landfill site, or "sanitary landfill", is an engineered site designed to prevent or minimize contamination to the surrounding environment. Typically, these sites are

continually monitored even after waste deposition is complete to ensure the system is contained and not impacting the surrounding environment. The main components of landfill design are physical barriers to prevent movement of liquids and gases, removal and subsequent treatment of liquid leachate, and collection, and sometimes utilization, of biogas. Conventional landfills are designed with protective mechanisms, but bioreactor landfills are further designed to exploit anaerobic or aerobic processes to promote accelerated waste stabilization (Omar and Rohani, 2015). Bioreactor landfills have MSW deposited within cells for burial and containment. Over time, there will be cells containing waste at different ages, and there will be compacted waste layers within each cell (Meyer-Dombard *et al.*, 2020).

Protection of the surrounding soils, groundwater, and atmosphere from landfilled MSW includes physical barriers. To protect soils and groundwater neighboring the deposited waste, engineered landfill sites employ an impermeable liner to collect liquid that percolates downward (Figure 1.1). Liners typically consist of geomembrane and clay layers. To limit exposure of the top layers of MSW to the atmosphere, landfill operations include the use of daily, intermediate, and final covers. When a cell is filled with MSW to its capacity the final cover (or cap) is installed, composed of clay and geomembrane layers along with a vegetative layer on top. The final cap layers help to protect liners and the encompassed MSW, limit leachate generation by diverting surface waters, and maximize gas collection.

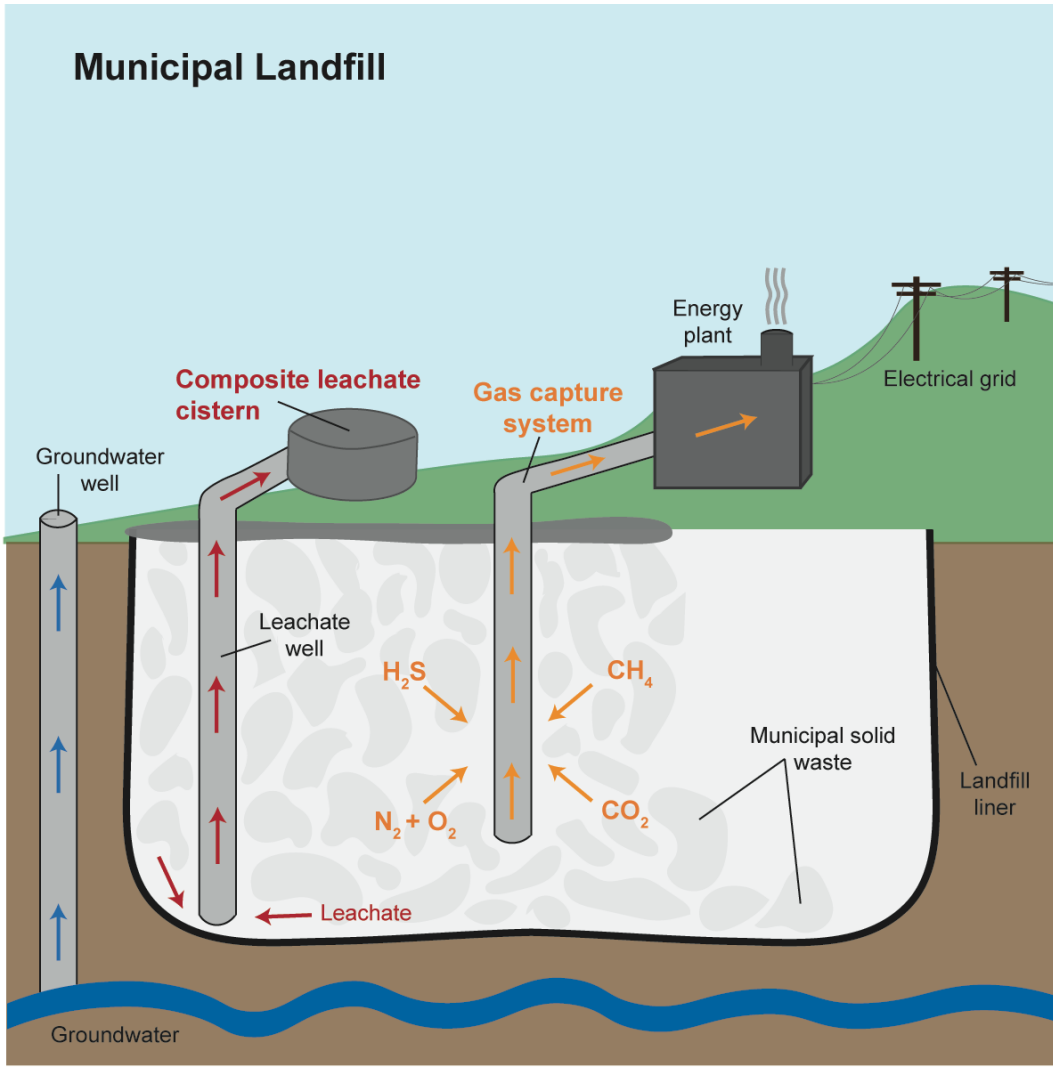


Figure 1.1 Schematic of a municipal landfill, or waste management site.

Leachate collection and diversion is an important environmental protection feature of landfill design. Leachate is the liquid that forms within landfill cells, originating from precipitation, surface run-off, or as a component of the deposited MSW. Leachate is collected within the impermeable liner in a leachate collection system, which is accessible via a series of leachate wells (LW) throughout the landfill site. At the landfill site sampled for research in Chapters 2 and 3 of this thesis, the leachate system diverts the liquid to a central location (*i.e.*,

composite leachate cistern; CLC) where leachate is collected and subsequently diverted from the landfill to a wastewater treatment facility where it is treated. Leachate is a complex liquid, whose composition is highly variable and influenced principally by climate (*i.e.*, precipitation, evaporation), age of the landfill, and composition of the originating MSW (Renou *et al.*, 2008). Leachate, as it percolates downward, is in contact with MSW, becoming contaminated with microorganisms, dissolved organic matter, inorganic components, heavy metals, and xenobiotic organic compounds (Christensen *et al.*, 2001; Slack *et al.*, 2005; Renou *et al.*, 2008). Analysis of leachate from 22 landfills in the United States found a median of 22 contaminants of emerging concern (CECs) per sample, with 101 CECs detected in total, including pharmaceuticals and industrial chemicals (Masoner *et al.*, 2016). Additionally, *in vitro* bioassays have demonstrated leachate as being a toxic, mutagenic, and estrogenic substance (Baderna *et al.*, 2019). Due to the heterogeneous and contaminated nature of leachate, it is an important feature of sanitary landfills to collect and divert leachate, preventing environmental contamination.

To prevent atmospheric and subsurface contamination due to gas migration, landfill sites are equipped with gas collection systems. Gas collection infrastructure typically consists of extraction wells placed throughout landfill cells, piping for gas transportation to a collection header, and a blower or vacuum induction system to power the movement of gas (US Environmental Protection Agency, 2017). Passive movement of gases will occur within waste due to diffusion, pressure, and permeability, though applying a vacuum is used to increase extraction efficiency and reduce gasses emitted to the atmosphere (Townsend *et al.*, 2015). The gas being produced by MSW degradation is typically warmer than ambient conditions, so the gas cools as it moves through the collection piping generating condensation. Condensate has the potential to gather at points of low elevation causing blockages so is removed with condensate traps

(Townsend *et al.*, 2015). Another important feature of the gas collection infrastructure is the flare, that is a device for igniting and burning landfill gas, which can be important during equipment downtime or when the amount of gas exceeds the facility's capacity. Flaring converts CH_4 to CO_2 via combustion with oxygen, reducing the potency of the GHG, but it can also generate other harmful gases (*i.e.*, SO_x, NO_x, and CO) (Omar and Rohani, 2015). Although flaring can be an important use or endpoint for landfill gas, the more desired outcome is for use in energy generation. Gas can be processed and cleaned to produce high value gas, and then turned into electricity using a combustion engine, fuel cell, or gas turbine, resulting in electricity that can be used on-site or sold onto the electrical grid.

1.1.3 Microbially mediated gas production

Methane production from MSW depends on a consortium of microorganisms, with MSW degradation occurring in four different stages, where the composition of the landfill gas (LFG) produced during each phase is distinct (Hilger and Barlaz, 2006; US Environmental Protection Agency, 2017). The two major components of landfill gas are CH_4 and CO_2, which typically comprise 45-60% and 40-60% of the total gas by volume, respectively (Agency for Toxic Substances and Disease Registry, 2001). Other components include sulfides and non-methane organic compounds (NMOCs).

The first phase of degradation is characterized by aerobic bacteria breaking down organic waste with CO_2 being the primary byproduct, and in an order of ease with which polymers are degraded. Here, cellulolytic and hydrolytic bacteria metabolize complex polymers into monomers and this phase continues until all the oxygen is depleted. Oxygen tolerance among microbial lineages is a significant driver of which metabolisms are active and which byproducts are accumulating or emitted from landfills at each decomposition stage. Microorganisms will use

defenses to protect themselves from oxidative stress and reactive oxygen species (ROS), and microbial lineages with varying capacities to tolerate oxygen can be indicative of which stage a landfill cell is in.

The second phase of degradation has anaerobic bacteria converting compounds created by aerobic bacteria into simple acids (acetic, lactic, formic) and alcohols (methanol, ethanol). During this phase the pH decreases because of an imbalance between fermentative activity with acetogenesis and methanogenesis (Hilger and Barlaz, 2006). The third phase begins when anaerobic bacteria consume the organic acids previously produced during acetate production which results in a more neutral pH. With the neutral pH, methanogenic archaea are able to grow, consuming CO_2 and acetate, and CH_4 production occurs at an accelerating rate. The fourth phase occurs when proportion of methane in the produced gas and the rate at which it is produced are decreasing (Barlaz *et al.*, 1989). Investigations with bioreactors mimicking landfill conditions have indicated that the early lifetime of a landfill generates the most NMOCs (Staley *et al.*, 2006), though the final phase can continue for 20 years (Agency for Toxic Substances and Disease Registry, 2001). Gasses are produced throughout the stages of MSW degradation, influenced by waste composition, age of refuse, presence of oxygen, moisture, and temperature (Omar and Rohani, 2015).

Throughout the stages of waste degradation and gas production there are different functional microbial guilds that dominate the active processes, yet the specific microorganisms that are vital at each stage is less well understood. For example, some guilds include starch degraders, pectin degraders, and protein degraders. But because cellulose and hemicellulose are the principal biodegradable component of MSW, cellulolytic microorganisms contribute significantly to degradation (Hilger and Barlaz, 2006). Cellulolytic microorganisms are a

functional guild of microorganisms hydrolyzing cellulose to soluble sugars that are dominant during the first aerobic stage of MSW degradation, but also under anoxic conditions. Several microorganisms have demonstrated cellulolytic degradation capacities under oxic conditions, such as microorganisms from the *Bacillaceae* family (phylum Firmicutes), the *Cytophagaceae* family (phylum Bacteroidota), and the phylum Actinobacteriota (Pourcher *et al*., 2006). For anaerobes, laboratory studies have indicated *Clostridium* in landfill leachate bioreactors to have cellulolytic degradation potential, and *Fibrobacter* species detected in MSW at landfill sites are predicted cellulose degraders (Burrell *et al.*, 2004; McDonald *et al.*, 2008). Recent studies utilizing high-throughput sequencing highlighted members of the Bacteroidota, Firmicutes, and Spirochaetota phyla in cellulose degradation in landfill leachates (Song, Wang, Zhao, *et al.*, 2015; Ransom-Jones *et al.*, 2017; Co and Hug, 2021). Next, using sugars generated by hydrolytic bacteria, fermentation by acidogenic and acetogenic microorganisms proceeds. Acidogens generate organic acids (*i.e.*, propionate, butyrate, acetate, and hydrogen) through sugar fermentation, while acetogens generate acetate, often through CO_2 reduction. The microbial members who provide these functional capacities, or how these microbes interact with each other, at landfill sites have not been clearly delineated (Semrau, 2011). However, studies using bioreactors to mimic temporal degradation processes of MSW have frequently highlighted Bacteroidota, Firmicutes, and Proteobacteria as prevalent community members (Cardinali-Rezende *et al.*, 2012; Bareither *et al.*, 2013; Staley *et al.*, 2018). Bacteroidota were identified as putative acetogenic bacteria that were likely to support the growth of acetoclastic methanogens such as *Methanosarcinaceae* (Bareither *et al.*, 2013). *Ruminococcus* was the dominant genus (phylum Firmicutes) during the anaerobic acid stage (Staley *et al.*, 2018). Similarly, with 16S rRNA gene sequencing, *Ruminococcus* along with

Pseudomonas were found to be abundant in earlier stages of MSW degradation at a landfill site in China (Song *et al.*, 2015).

The final functional guild of microorganisms in the degradation of MSW is the methanogens, responsible for the terminal step in the anaerobic decomposition process, and dependent on previous functional guilds' activities. Methanogens are anaerobic Archaea that produce methane, either through the acetoclastic, hydrogenotrophic, or methylotrophic pathways. In laboratory bioreactors, *Methanomicrobiales*, *Methanosarcinales*, and *Methanobacteriales* were often identified in later methanogenic stages (Cardinali-Rezende *et al.*, 2012; Bareither *et al.*, 2013; Staley *et al.*, 2018), and from 16S rRNA sequencing of landfill leachates, *Methanomicrobiales* and *Methanobacteriales* were frequently the dominant archaea at landfill sites (Song *et al.*, 2015a; Stamps *et al.*, 2016).

Microbial ecological succession is substantial during decomposition of MSW, and high-throughput sequencing technologies have provided more resolution for identifying microbial community members at landfill sites (Cardinali-Rezende *et al.*, 2012; Bareither *et al.*, 2013; Song *et al.*, 2015; Stamps *et al.*, 2016). Functional guilds of microorganisms shift in abundance and activity in response to the heterogeneous nature of MSW, leachate, and environmental conditions they are exposed to in the landfill ecosystem. High-throughput sequencing has identified under-described lineages, such as Cloacimonadota (previously WWE1; Cloacimonetes), in landfill ecosystems (Cardinali-Rezende *et al.*, 2012; Sauk and Hug, 2022). Further investigation through metagenomic sequencing offers the opportunity to probe microbial diversity and function at landfill sites. Several microorganisms at landfill sites also make the switch from a free-living, planktonic lifestyle to a sessile, biofilm lifestyle in gas collection systems and leachate infrastructure. Biofilms in gas collection systems are likely seeded from microorganisms occurring

in MSW; however biofilm-specific microbial diversity and function have not yet been described in the literature and remains a novel landfill ecosystem to study. Clarifying the role of microorganisms that reside within landfill sites will aid the implementation of sustainable landfilling practices.

1.2 Biofilms

1.2.1 Stages of biofilm development, and emergent properties

Biofilms are a widely distributed and successful form of life, existing in many different ecosystems, because biofilm attachment provides microorganisms with an environment conducive to their success and survival. Life for microbial aggregates in a biofilm is distinct from life as free-floating, planktonic cells, largely due to the protection that is provided by the self-produced matrix in which microorganisms are embedded. The matrix has properties that support biodiversity, protect from the surrounding environment, and facilitate synergistic relationships between microorganisms (Burmølle *et al.*, 2014). Biofilms are found in a wide variety of environments, and play important roles in biogeochemical cycling (Battin *et al.*, 2003; Flemming and Wuertz, 2019). They are part of natural ecosystems, such as sediments (Foulquier *et al.*, 2011; Jägevall *et al.*, 2011), hydrothermal vents (Stokke *et al.*, 2015; Sciutteri *et al.*, 2022), streams (Besemer *et al.*, 2009; Battin *et al.*, 2016), and soils (Bystrianský *et al.*, 2019; Wu *et al.*, 2019). Biofilms can provide beneficial outcomes. For example, biofilms provide an environment conducive for initiating larval development in sponges (Webster *et al.*, 2004; Whalan and Webster, 2014), for filtering drinking water (Kasuga *et al.*, 2007), and for facilitating wastewater treatment (Galván *et al.*, 2000; Yuan *et al.*, 2020). Biofilms can also be detrimental with severe economic consequences, such as biofilms that form on medical devices (Lleo *et al.*, 2007; Francolini and Donelli, 2010), biofouling of manufactured surfaces in the ocean (Dang *et al.*, 2011; Papadatou *et al.*, 2021), and

biofouling of oil and gas pipelines (Lenhart *et al.*, 2014; Liduino *et al.*, 2019). With biofilms as a dominant form of life in many different ecosystem types, cells in biofilms may account for up to 80% of the total cells on Earth (Flemming and Wuertz, 2019).

Biofilms differ from planktonic communities because they are heterogeneous assemblages of microbial cells encased within a self-produced matrix that adhere to a surface, and to themselves. The generated matrix is made of extracellular polymeric substances (EPS), and consists largely of polysaccharides along with proteins, lipids, and DNA (Flemming *et al.*, 2016). The matrix in biofilms is multilayered, resulting in spatial organization with microcolonies and gradients (*i.e.*, electron donors and acceptors, pH, redox conditions), and zones within a biofilm that are morphologically and physiologically distinct, even at the micrometer scale, that are called microenvironments (Serra *et al.*, 2013). For attached biofilms, the inner layer maintains an irreversible attachment to the substratum. Microorganisms residing in microcolonies experience enhanced physical and social interactions, including some cell-to-cell contact. Overall, biofilms are heterogeneous, often containing many microbial species with high cell densities, with the generation and remodeling of the self-produced matrix being a continuous and dynamic process.

The biofilm lifestyle or formation can be summarized in steps that include attachment, early development, maturation, and dispersal (Kirchman, 2018). The initial step of biofilm formation requires a planktonic cell to adhere to a surface, first reversibly and then irreversibly. At the liquid-surface interface, movement creates a hydrodynamic boundary layer near the surface that is a zone of negligible flow (Donlan, 2002). The size of the hydrodynamic boundary layer will depend on the movement of liquid, where higher velocities result in a thinner layer, more turbulence, and require a cell to adhere rapidly. Properties of the substratum and cell both influence the rate and efficiency at which cells can attach to a surface. Cellular appendages and proteins

influence attachment, such as flagella, fimbriae, pili, and curli proteins (Friedlander *et al.*, 2013; Berne *et al.*, 2015). To strengthen attachment cells may increase production of adhesions, alter cell surface structures, or make surface-contact dependent changes in gene regulation (Berne *et al.*, 2018). Once cells have become irreversibly attached to the substratum they can replicate and initiate matrix production, cementing cells to the surface and other cells, and allowing the biofilm to enter the maturation stage.

During biofilm maturation the density and complexity of the biofilm increases. Cells are replicating and generating the EPS matrix, forming the biofilm architecture in a continuous and dynamic process. The matrix provides structural and functional benefits – such as hydration, nutrient capture, digestive capacity, and protection, while facilitating intercellular interactions (Kirchman, 2018). Biofilm-associated cells benefit over free-living bacterial cells because they are in contact with either the substratum or flowing water that can be continuous sources of nutrients. Biofilms can use the passive sorption properties of the EPS to capture nutrients, gases, and other molecules, which can be especially beneficial in oligotrophic environments. However, passive sorption is not specific, and a variety of molecules can be sequestered, some contributing to nutrient cycling and other recalcitrant molecules will accumulate, and so biofilms can act as both a sink and a source of contaminants (Flemming *et al.*, 2016). Cellular debris from decaying cells can be recycled and used by surviving cells as a nutrient source. Microorganisms in biofilms can make more efficient use of their extracellular degradative enzymes compared to their planktonic counterparts, because the enzymes are retained in the matrix in close proximity to cells, acting as an external digestive system (Flemming *et al.*, 2016). The structure of the matrix provides protection from exposure to the environment and other molecules. For example, the EPS matrix is composed of hydrated polymers that acts as a hydrogel, protecting embedded cells from

desiccation. And further, the matrix can quench the activity of antimicrobials that diffuse into the biofilm, creating sublethal concentrations, and promoting selection for antimicrobial resistance (Flemming *et al.*, 2016). Another mechanism that enhances resistance to antimicrobials in biofilms is the uptake of resistance genes by horizontal gene transfer. The high cell densities with cells in close proximity inherent to biofilm structure allow for efficient gene transfer, which has been demonstrated to occur at higher rates in biofilms than in free-living cells (Flemming *et al.*, 2016).

A biofilm structure facilitates close interactions between cells in embedded microcolonies. This can include different types of communication, metabolic cooperation, and competition. Cells that are proximal to each other can communicate using chemical signaling, such as with quorum sensing, or with electrical signals, like nanowires. With quorum sensing, chemical signal molecules called autoinducers are released, increasing in concentration with increasing cell density, to communicate with other cells. Metabolic interactions will occur, allowing metabolic cross-feeding where metabolic by-products of one species are metabolized further by other species (Joshi *et al.*, 2021). There will also be competition between cells residing in the biofilm, such as with killing mechanisms, like antibiotics or bacteriocins, competition for space and resources, or negative interactions like predator-prey or parasitic dynamics. Viral infections can spread rapidly within a biofilm and immobilized cells are easy targets for grazing protists (Nadell *et al.*, 2016).

A mature biofilm is dynamic. The biofilm is constantly reconfiguring, with cells replicating, building the EPS matrix, and opening new channels for nutrient flow. The final stage of biofilm development is called dispersion, where cells are shed from the biofilm either as clumps of biofilm or in a return to their planktonic life cycle, sometimes to colonize new sites. This can occur passively, where cells are released due to mechanical or shear stress, as seen in erosion or sloughing (Rumbaugh and Sauer, 2020). Abrasion can lead to detachment due to the collision of

particles from bulk fluid with the biofilm. Dispersion typically refers to an active event where cells are leaving the biofilm matrix to become free-living cells. Dispersion occurs as cells respond to cues, signals, and environmental factors, including self-produced signaling molecules, steep gradients, or as a response to sensing factors in the surrounding environment (Rumbaugh and Sauer, 2020).

1.3 Multi-omic methods to investigate microbial diversity and function

1.3.1 Leveraging DNA sequencing for evaluation of microbial diversity and function

The diversity of bacteria and archaea in microbial ecosystems have primarily been explored using 16S rRNA gene sequencing. The 16S rRNA gene is an excellent marker gene, as it is present ubiquitously among bacteria and archaea and has conserved and hypervariable regions that can be used to differentiate microorganisms taxonomically (Janda and Abbott, 2007; Mizrahi-Man *et al.*, 2013). This method of sequencing provides insight into microbial diversity; however, it does not provide information regarding the metabolic capacity harbored by the microbial community. An alternative type of DNA sequencing is untargeted, or shotgun, sequencing, which, when applied to microbial community DNA, is called metagenomics.

Metagenomics utilizes shotgun sequencing to evaluate the genetic content of the entire community of microorganisms in an environment, and provides a broader description of the microbial community compared to 16S rRNA gene analyses (Thomas *et al.*, 2012). Metagenomic surveys allow uncultured organisms to be identified and their metabolic capacities predicted, while also allowing organisms to be investigated in the context of their environment (Sharon and Banfield, 2013). Metagenomic sequence data can be used to reconstruct genomes for populations within the entire microbial community without any preliminary cultivation or enrichment steps,

generating Metagenome-Assembled Genomes (MAGs). Briefly, reads from a sequencing run using community DNA are assembled into larger contiguous sequences (contigs) or scaffolds based on sequence overlaps. These scaffolds can further be binned, which is the process of sorting scaffold sequences based on a combination of sequence characteristics and abundance data to delineate MAGs from environmental populations (Thomas *et al.*, 2012). Genes can then be annotated from the obtained draft genomes, where annotation pipelines first identify where genes are located (gene calling) and then assign putative functions based on similarity searches to global databases of microbial genes (Huntemann *et al.*, 2016). The obtained gene content information can then be assessed as to presence and completion of metabolic pathways, determining the metabolic potential of these organisms.

Sequencing environmental DNA bypasses the requirement for cultivation or enrichment, enabling the exploration of microbial dark matter. As such, in recent years exploring the genetic content within MAGs has expanded our understanding of the tree of life (Hug *et al.*, 2016; Parks *et al.*, 2017; Castelle and Banfield, 2018). And metagenomics has provided opportunity for exploration of diverse, and sometimes extreme ecosystems, from hot springs within an active volcanic ecosystem in Russia (Wilkins *et al.*, 2019) to the primate gut (Manara *et al.*, 2019). Beyond simply expanding the tree of life, the gene content of novel lineages can be surveyed to predict metabolic potential. For example, these methods have been used to describe a novel lineage *Candidatus* Kryptonia, derived from a geothermal spring. Metabolic reconstruction predicted nutritional deficiencies, and therefore it was posited these organisms may rely on microbial partners (Eloe-Fadrosh *et al.*, 2016). Another phylum, Aquificae, and specifically the genus *Sulfurihydrogenibium* was the dominant lineage within a biofilm near a lake hydrothermal vent. With metabolic predictions obtained from MAGs, *Sulfurihydrogenibium* metabolism is predicted

to use oxidation of sulfur compounds to fuel carbon fixation (McKay *et al.*, 2022). Metabolic reconstructions obtained through the annotation of MAGs aids our understanding of metabolic potential in diverse, extreme environments, and modern sequencing technologies have been recognized as a promising avenue for elucidation of important microbial processes at landfill sites (Meyer-Dombard *et al.*, 2020).

1.3.2 Integrated multi-omics approaches

Metatranscriptomics (total RNA sequencing) and metaproteomics (total protein sequencing) can be paired with metagenomes to portray the metabolism and other functions that are active within a specific environment at a specific time. Metatranscriptomics and metaproteomics extend metagenomic analyses beyond gene content to genes that are expressed and translated, respectively (Warnecke and Hugenholtz, 2007), and can be used to validate observations from metagenomic studies or to clarify the roles of organisms whose activity differs from their overall abundance (*e.g.*, highly active keystone species whose functional impact is outsized compared to their population abundance) (Roume *et al.*, 2015).

Metatranscriptomics includes the study of expressed RNA, which provides insight into genes that are being expressed, but still does not always mean the expressed genes are translated into proteins. Metatranscriptomic studies are further complicated by the fact that RNA molecules are highly labile, and RNases are ubiquitously present – therefore, sample processing and preservation are critical steps to ensure RNA profiles accurately depict the expression of the cells at the time of sampling, as well as to ensure RNA quantity and quality.

Proteins are more stable than RNA, but metaproteomics is lower throughput and a more expensive type of analysis. A typical metaproteomics workflow consists of protein extraction, tryptic digestion into peptides, separation via liquid chromatography, and then tandem mass

spectrometry analyses, with protein identification from comparison of peptide experimental mass spectra to theoretical mass spectra (Heyer *et al.*, 2017). As the half-life of proteins is longer than that of RNA and proteins are less fragile during extraction processes, metaproteomics can give distinctly different metabolic profiles to paired metatranscriptomic datasets. The proteomics-based profiles are likely a more accurate depiction of the activity of the abundant organisms in the community under the sampled environmental conditions.

Metametabolomics analyzes metabolites, or small molecules, in the environment. The process can use a combination of chromatography separation, such as liquid chromatography (LC) or gas chromatography (GC) coupled with a detection method like mass spectrometry (MS), or nuclear magnetic resonance (NMR) that creates spectra that can be compared to spectral databases for identification. LC-MS and GC-MS have more sensitivity and lower limits of detection than NMR, though there are benefits to NMR in resolving power, so a dual approach can be beneficial when exploring metabolites (Markley *et al.*, 2017; Emwas *et al.*, 2019; Bauermeister *et al.*, 2022).

A multi-omics approach can be applied to understand the metabolism of specific molecules. For example, polyphenol metabolism in soils was evaluated with anoxic soil bioreactors to mimic wetland soils, where polyphenols are predicted accumulate due to low levels of degradation. Within bioreactors metametabolomic data revealed evidence of polyphenol depolymerization, and genome-resolved metaproteomic data indicated a functional soil microbiome, challenging previous ideas that polyphenols inhibit microbial activity (McGivern *et al.*, 2021). Another use of multi-omics can be to elucidate important organisms or processes occurring in a system. For example, an integrated omics approach was used to evaluate seasonal fluctuations in a wastewater treatment plant that identified differences between seasons in glycerolipid metabolism encoded by *Candidatus* Microthrix parvicella, and “keystone genes”

which were more highly expressed than expected based on gene content (Roume *et al.*, 2015). Integration of multiple omics datasets can provide insight into important microbial lineages and processes in a system.

1.4 Research objectives and thesis outline

Due to the immense impact of landfills to current and future climate change initiatives, paired with the lack of a robust understanding of microbial processes at these sites, the research in this thesis aims to broaden our understanding of landfill microbial diversity and ecology. Specifically, problematic biofilms that form within gas capture and leachate infrastructure are occurring at an increasing number of landfill sites (Mission Black Goo: Figuring out what it is and how to tackle it, 2022; Lozecnik *et al*., 2012); however, the microbial composition of these biofilms have largely not yet been explored. Though our understanding of landfill microbiology is imperfect, landfills are known to harbor many microbial lineages (Song *et al.*, 2015; Stamps *et al.*, 2016; Sauk and Hug, 2022). Several lineages, such as Cloacimonadota, remain enigmatic despite detection in other environments. And due to the wide range of processes that occur throughout the lifecycle of a landfill, ranging from oxic to anoxic, understanding the resiliency of microbes to ROS is important.

There are three main objectives for my research which all use metagenomics as an approach to describe microbial diversity and function. The first objective was to characterize the microbial diversity and function in landfill-associated biofilms using a multi-omics approach (Chapter 2). The purpose of this research was to identify important microorganisms and microbial metabolic processes within the biofilms that could be used to inform on management practices to limit biofilm persistence at landfill sites. A total of five biofilm samples, from separate biofouling events at a Southern Ontario landfill, were evaluated using 16S rRNA profiling, metagenomics,

metaproteomics, and metametabolomics. The second objective was to describe the metabolic potential of phylum Cloacimonadota, an understudied phylum detected in metagenomic datasets from the study site (Chapter 3). MAGs were assembled and complemented by genomes from public databases to explore phylum-level diversity and metabolic reconstructions. The third objective was to delineate the distribution of ROS defenses across bacteria (Chapter 4). Defenses against oxygen are important for microbial survival in oxygenated environments and given the varied oxic and anoxic phases throughout the lifecycle of a landfill, are important for organisms in this environment. Distribution of genes was evaluated through annotated genomes that are publicly available and peer-reviewed literature.

Chapter 2
Life in black goo: multi-omic description of microbial diversity and function in biofouling at a landfill site

2.1 Introduction

In this chapter, microbial diversity and function within landfill-associated biofilms is described using a multi-omics approach that was centered around metagenomics. Biofilms found at landfill sites can be detrimental by causing blockages in the gas and leachate collection systems from which they are found. Gas collection systems are designed within landfill sites to redirect landfill gases (biogas), including methane (CH_4), that are by-products of waste degradation and transformation. The methane within biogas can subsequently be converted into electricity and used as an energy source (Omar and Rohani, 2015). Gas collection systems are highly variable, often containing high concentrations of methane (CH_4, up to 70% of the atmospheric gases) and carbon dioxide (CO_2, up to 15% of the atmospheric gases) (landfill engineers, personal communication). Biofouling causes blockages in the system from overgrowth of a thick viscous biofilm, colloquially referred to as "black goo", that reduces gas collection efficiency and limits biogas usage (Mission Black Goo: Figuring out what it is and how to tackle it, 2022). Biofouling within landfill gas and leachate collection systems is detrimental to efficient landfill operation and gas capture. Despite this black goo being a recognized problem at many landfills along the North American east coast and central regions since 2009 (Dr. Craig Benson, University of Wisconsin-Madison, personal communication), the physical composition and microbial community of the black goo has not been thoroughly investigated until now, in a collaboration between Dr. Craig Benson, the Region of Waterloo, SiREM laboratories, and the Hug Research Group.

Biofilms are a common form of microbial life. Biofilms are microbial cells and assemblages that are encased within a self-produced matrix of extracellular polymeric substances (EPS), and often consist of multiple microbial species (Madigan *et al.*, 2012). Biofilms can comprise single or multiple species microbial assemblages, though interactions in multi-species biofilms provide protective effects and the opportunity for synergistic interactions (Burmølle *et al.*, 2014). Biofilms are a common occurrence in aquatic environments, including oceans and streams. There, biofilms drive important ecological processes and contribute to biogeochemical fluxes (Battin *et al.*, 2016). Metagenomic sequencing has been beneficial in expanding our understanding of microbial diversity in ocean biofilms, highlighting genes that comprise a functional core across biofilms (Zhang *et al.*, 2019).

An example of a multi-species biofilms occurring in an industrial setting with similarly detrimental effects is growth that occurs in oil and gas and pipelines. These biofilms are problematic due to microbially mediated corrosion that is often caused by sulfate-reducing bacteria (Li *et al.*, 2013). A common solution to reduce corrosion is "pigging", a mechanical cleaning of the interior of the pipe, followed by a biocide treatment (Vigneron *et al.*, 2018). Knowledge of microbial metabolisms has also been leveraged as a solution for biofouling in these systems. Because sulfate reducing bacteria and the production of hydrogen sulfide (H_2S) are linked with biocorrosion, limiting those microbial guilds may limit biofouling and biocorrosion. Therefore, nitrate injections have been used to enable nitrate reducers to outcompete sulfate reducers, as there is an increased energy yield associated with nitrate reduction (Schwermer *et al.*, 2008). However, in practice nitrate injections have varying rates of success due to reemergence of sulfate reducers once nitrate has been depleted, or through continued metabolism of other sulfur intermediates

(Vigneron *et al.*, 2018; Nicoletti *et al.*, 2022). Understanding the critical metabolic processes occurring in a biofilm can strengthen remediation or prevention strategies.

Microbial communities have been examined at other areas within landfill sites, such as leachate wells and composite leachate cisterns (Sauk and Hug, 2022). Biofouling within the landfill gas and leachate collection systems is a unique phenomenon, where descriptions of the microbial diversity present and metabolisms supporting the community is limited (Lozecnik *et al*., 2012). A multi-omics approach was used to explore the microbial diversity and function within biofilms from the gas capture system and leachate pumps, and to assess the dominant metabolic strategies underlying these industrially disruptive biofilms.

2.2 Materials and Methods

2.2.1 Sample collection

Biofilm samples were collected at a Southern Ontario active sanitary landfill site. The samples were collected from biofouling events within the gas collection system. The first sample, Biofilm 1 (BF1), was extracted from a valve in a condensate trap on October 25, 2017. All other biofilms were obtained from separate biofouling events at leachate or gas pump stations, with Biofilm 2 (BF2) collected on October 4, 2018, Biofilm 3 (BF3) collected on August 23, 2019, Biofilm 4 (BF4) collected on January 14, 2020, and Biofilm 5 (BF5) collected on July 9, 2019. Biofilm numbers are ordered temporally except for BF5. The BF5 sample was numbered last due to different handling during collection (*i.e*., placed on ground, hosed off with water), and, consequently, the BF5 sample was not included in multi-omic analyses. Biofilms were stored in a -80°C freezer until processed further. Genomic DNA from each biofilm was extracted using the PowerSoil DNA isolation kit (MoBio) following the manufacturer's instructions with the modification of adding 10 g of biofilm in place of soil. All five biofilms were subject to 16S rRNA

gene sequencing. Multi-omic analyses were performed on four biofilms: BF1, BF2, BF3, and BF4. Insufficient DNA was extracted from BF3 for metagenomic sequencing, as quantified using Qubit dsDNA High Sensitivity assay kit (Thermo Fisher Scientific Inc.), despite multiple attempts and after exhausting the available sample. Protein and metabolites were successfully extracted and processed for metaproteomics and metametabolomic analyses.

2.2.2 16S rRNA gene sequencing and analyses

Genomic DNA was input for 16S rRNA gene sequencing on a MiSeq (Illumina) for all five biofilm samples (BF1 – BF5) at MetagenomBio (https://metagenom.co/). The 16S rRNA gene was amplified through the V4 region using the forward primer 515FB (5'-GTGYCAGCMGCCGCGGTAA-3') and the reverse primer 806RB (5'-GGACTACNVGGGTWTCTAAT-3') (Walters *et al.*, 2016). For each sample, PCR was performed in triplicate to a volume of 25 µL, where each reaction contained: 2.5 µL of 10X standard *Taq* buffer, 0.5 µL of 10mM dNTP, 0.25 µL of BSA (20 mg/mL), 5 µL of 1 µM forward primer, 5 µL of 1 µM reverse primer, 5 µL DNA, and 0.2 µL of *Taq* DNA polymerase (5U/µL), and 6.55 µL of PCR water. In the thermocycler, the reaction was performed with denaturation at 95°C for 5 minutes, followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 30°C, and 30 seconds at 72°C, and a final extension at 72°C for 10 minutes. Triplicate PCR products were pooled in equal amounts and resolved with a 2% TAE agarose gel, followed by gel purification. Amplicons were quantified using Qubit dsDNA High Sensitivity assay kit (Thermo Fisher Scientific Inc.). Amplified DNA was sequenced on the MiSeq platform with preparation using the MiSeq Reagent Kit v2 (2 x 250 cycles), and FASTQ files were generated for taxonomic analyses.

Sequence reads were analyzed using Quantitative Insights Into Microbial Ecology 2 (QIIME2) version 2020.2.0 (Bolyen *et al.*, 2019). Paired end sequences were imported and

demultiplexed, and adapters were removed from the sequences with *qiime cutadapt trim-paired*. Next, DADA2 was used to control for sequence quality and to generate amplicon sequence variants (ASVs) (Callahan *et al.*, 2016). This step used *qiime dada2 denoise-paired* to remove low quality sequences and truncate the forward and reverse reads to 180 base pairs, where the kept length was chosen based on visualization of quality scores produced within QIIME2. Sequence variants were summarized using *qiime feature-table summarize.* Taxonomic classification of ASVs was determined using a naïve Bayes classifier trained against SILVA release 138 (Pruesse *et al.*, 2007).

Beta- and alpha-diversity metrics were calculated with *qiime diversity core-metrics-phylogenetic* with a rarefied sampling depth of 5000 sequences per sample. The phylogenetic tree used for diversity analyses was generated within the QIIME2 workflow, where first an alignment was generated using *qiime alignment mafft*, sequences were filtered to remove highly variable positions with *qiime alignment mask*, and *qiime phylogeny fasttree* was used to generate the tree. Diversity values were exported and visualized in R.

2.2.3 Metagenomic sequencing and analyses

Three biofilm samples had sufficient DNA extracted to enable metagenomic sequencing: BF1, BF2, and BF4. BF1 was sequenced at the US Department of Energy's (DOE) Joint Genome Institute (JGI) on a HiSeq platform (Illumina). BF2 and BF4 were sequenced at The Center for Applied Genomics (TCAG) in Toronto, Ontario, where BF2 was sequenced on a HiSeq (Illumina), and BF4 on a NovaSeq (Illumina) platform. All metagenomic sequencing generated 2 x 150 bp paired end reads. All metagenomes were subsequently processed with the same pipeline. Sequence reads were trimmed with the adaptive read trimmer, Sickle (Joshi and Fass, 2011). Sequences were assembled *de novo* with MetaSPAdes (Nurk *et al.*, 2017) and output scaffolds were size selected for further analyses to include scaffolds over a size of 2.5 Kb. All-by-all read mapping was

completed with Bowtie2 version 2.3.5 (Langmead and Salzberg, 2012) and read mapping information was used to calculate scaffold coverage within a sample.

Scaffolds were binned into metagenome-assembled genomes (MAGs). Three binning algorithms, CONCOCT version 1.1.0 (Alneberg *et al.*, 2014), MaxBin2 version 2.2.6 (Wu *et al.*, 2016), and MetaBAT2 version 2.12.1 (Kang *et al.*, 2019), were run individually. Output from these algorithms were subsequently integrated with DASTool version 1.1.1 (Sieber *et al.*, 2018) to select the highest proportion of high quality, near-complete MAGs. The quality of the generated MAGs was assessed with CheckM (version 1.1.3) by using the *lineage_wf* workflow to generate estimates of genome completeness and contamination (Parks *et al.*, 2015). Any MAGs with >70% completeness and <10% contamination were retained for further analyses. The coverage for each MAG was calculated from scaffold coverages previously obtained through read mapping. MAG coverage was calculated as the average from all corresponding scaffold coverages, while accounting for the length of each scaffold. The relative abundance of a MAG, or proportion, was calculated as percent of a MAG's coverage from the total coverage for all MAGs within a biofilm sample. Taxonomy of MAGs was assigned with the Genome Taxonomy Database (GTDB) version 1.0.2, release 89 (Chaumeil *et al.*, 2020). MAGs were dereplicated with DRep version 2.6.2 with clustering at 95% ANI to roughly represent species level differences. A phylogenetic tree was generated using the dereplicated MAGs. The tree was generated using the GToTree program with the "Universal_Hug_et_al" HMM set and default parameters (Lee, 2019), and was visualized using the interactive tree of life (iTOL) (Letunic and Bork, 2016).

To evaluate the metabolic potential encoded within the metagenomes, open reading frames were predicted from the assembled scaffolds using the gene prediction tool Prodigal (Hyatt *et al.*, 2010). The genes identified were then annotated using the JGI-IMG Annotation Pipeline (IMGAP;

version 5.0.23), and metabolic function was predicted based on the presence of key enzymes within metabolic pathways. To supplement annotations for genes involved in defense against reactive oxygen species (ROS), hmmsearch with hidden Markov models (HMMs) was used to search for the rubrerythrin family (PF02915) and superoxide reductase (desulfoferrodox) family (PF01880), which were downloaded from the Pfam website. Hits were considered when corresponding e-values were below $1e^{-20}$.

2.2.4 Metaproteomics and metametabolomics

Metaproteomic and metametabolomic analyses were completed at the Environmental Molecular Sciences Laboratory User Facility (EMSL; Department of Energy, Washington, USA) for four biofilm samples: BF1, BF2, BF3, and BF4. From each biofilm, a 1 g aliquot was used for extraction. The MPLEx extraction protocol was used because it is designed for multi-omics analyses and allows single-sample extraction of both proteins and metabolites (Nakayasu *et al.*, 2016).

For metaproteomics, proteins from the extraction were then resolubilized in an 8M urea solution and digested with Trypsin. Peptides were purified through a C18 solid phase extraction (SPE) cartridge and diluted to 0.1 μg per μL for mass spectrometry (MS). The MS analysis was executed with a Q-Exactive HF-X mass spectrometer (Thermo Scientific) that was outfitted with a homemade nano-electrospray ionization interface, where the electrospray emitters were created using 150 μm o.d. x 20 μm i.d. chemically etched fused silica (Kelly *et al.*, 2006). The ion transfer tube temperature was 300°C and the spray voltage was 2.2 kV. Data were collected for 120 minutes, following a 10-minute delay after completion of sample trapping and the start of gradient. Fourier Transform Mass Spectrometry (FT-MS) spectra were acquired from 300 to 1800 m/z at a resolution of 60 k (AGC target 3e6; Automatic Gain Control). The top 12 FT-HCD-MS/MS (HCD;

higher-energy collisional dissociation) spectra were acquired in data-dependent mode with an isolation window of 0.7 m/z at a resolution of 45 k (AGC target 1e5) using a normalized collision energy of 30, dynamic exclusion time of 45 seconds, and detected charge state of an ion 2 – 6.

Database searching of generated MS/MS spectra was completed with MS-GF+ (Kim and Pevzner, 2014), where best matches were filtered at 1% false detection rate (FDR). The database used for searching was an aggregated, non-redundant set of predicted and annotated genes from the metagenomes of BF1, BF2, and BF4. Proteins were quantified using MASIC (MS/MS Automated Selected Ion Chromatogram Generator) (Monroe *et al.*, 2008). The abundance values generated were used to calculate a normalized protein abundance (NPA) for each protein within a biofilm sample. The NPA was calculated as:

$$NPA = 10^6 * \frac{\text{(protein abundance/protein length)}}{\text{sum(protein abundance/protein length)}}$$

This calculation is similar to the transcripts per million (TPM) calculation that is often used in metatranscriptomic studies (Zhao *et al.*, 2020). This NPA normalization accounts for protein length and total abundance values within a sample, allowing for proportional comparison of protein abundance across biofilm samples per million spectral counts. The datasets contained matching metaproteomes and metagenomes for BF1, BF2, and BF4. A metagenome was not obtained for BF3, and therefore, there was no database for matching peptides spectra. Consequently, BF3 was not included in further metaproteomic analyses.

Nitrogen, sulfur, and methane cycling, along with enzymatic defenses against reactive oxygen species (ROS), were evaluated to provide a summary of main metabolic strategies detected in the metagenome and metaproteome. For each pathway, specific marker genes were used to represent the presence or absence of nitrogen, sulfur, and methane cycling functions (Table 2.1).

Table 2.1 Genes used as markers for functions of interest.

Metabolic pathway	Gene name	Protein name	KO number
Denitrification	nitrate reductase alpha subunit	NarG	K00370
Denitrification	nitrate reductase (cytochrome)	NapA	K02567
Denitrification	nitrite reductase (NO-forming)	NirS	K15864
Denitrification	nitric oxide reductase subunit B	NorB	K04561
Denitrification	nitrous oxide reductase	NosZ	K00376
Nitrification	ammonia monooxygenase subunit A	AmoA	K10944
Nitrification	nitrate reductase beta subunit	NxrB	K00371
Nitrogen fixation	nitrogenase iron protein	NifH	K02588
Dissimilatory nitrate reduction to ammonia (DNRA)	nitrite reductase (cytochrome c-552)	NrfA	K03385
Anammox	hydrazine synthase alpha subunit	Hzs	K20932
Anammox	hydrazine dehydrogenase	Hdh	K20935
Dissimilatory sulfate reduction	sulfate adenylyltransferase	Sat	K00958
Dissimilatory sulfate reduction	adenylylsulfate reductase subunit A	AprA	K00394
Dissimilatory sulfate reduction	dissimilatory sulfite reductase alpha subunit	DsrA	K11180
Sulfide oxidation	sulfide:quinone oxidoreductase	Sqr	K17218
Sulfide oxidation	sulfide dehydrogenase [flavocytochrome c]	FccB	K17229
Sulfur oxidation	sulfur oxygenase reductase	Sor	K16952
Sulfite oxidation	sulfite dehydrogenase (cytochrome) subunit A	SorA	K05301

Table 2.1, cont.

Metabolic pathway	Gene name	Protein name	KO number
Thiosulfate oxidation	sulfur-oxidizing protein	SoxB	K17224
Thiosulfate disproportionation	thiosulfate reductase/polysulfide reductase chain A	PhsA	K08352
Methanogenesis	methyl-coenzyme M reductase alpha subunit	McrA	K00399
Methanogenesis	methyl-coenzyme M reductase beta subunit	McrB	K00401
Methanogenesis	methyl-coenzyme M reductase gamma subunit	McrG	K00402
Methanotrophy	methane monooxygenase regulatory protein B	MmoB	K16160
Methanotrophy	methane monooxygenase component C	MmoC	K16161
Methanotrophy	methane monooxygenase component A alpha chain	MmoX	K16157
Methanotrophy	methane monooxygenase component A beta chain	MmoY	K16158
Methanotrophy	methane monooxygenase component A gamma chain	MmoZ	K16159
Methanotrophy	methane/ammonia monooxygenase subunit A	PmoA	K10944
Methanotrophy	methane/ammonia monooxygenase subunit B	PmoB	K10945
Methanotrophy	methane/ammonia monooxygenase subunit C	PmoC	K10946

Genes used to evaluate the capacity for defenses against ROS are described in Chapter 4 (Johnson and Hug, 2019). Each metabolic cycle was visualized as a pathway. If there was more than one marker gene to represent a function, the NPA for all relevant proteins was summed to give an overview of that function's abundance in each sample. Heatmaps were created with ggplot2 (Ginestet, 2011) to further discern the taxonomic affiliations for the functions of interest in the MAGs.

Metabolite extractions were completed as part of the MPLEx extraction, along with the proteins as mentioned above, and input to GC-MS for metametabolomics. Metabolites were derivatized using a previously described protocol (Fiehn, 2016), followed by GC/MS analysis. An Agilent GC 7890A was coupled with a single quadrupole MSD 5975C (Agilent Technologies) to collect GC/MS data over a mass range of 50 – 550 m/z. Raw GC/MS data files were processed using Metabolite Detector software version 2.5 beta (Hiller *et al.*, 2009). Metabolites were first identified through matching experimental spectra to the FiehnLib library (Kind *et al.*, 2009), and then manual curation was used to validate identifications with the NIST 14 GC-MS library (NIST; National Institute of Standards and Technology). Heatmaps were created by log10 transforming metabolite raw abundance values and then performing hierarchical clustering and visualization with the "pheatmap" package in R (https://github.com/raivokolde/pheatmap).

2.3 Results and Discussion

2.3.1 Microbial heterogeneity within biofilms

Assessing microbial community composition within biofouling of landfill gas capture systems is critical for understanding the key players involved in biofilm dynamics and may offer insight into how to prevent or control biofouling events. In this study, five biofilms were sourced

Figure 2.1 Examples of biofouling occurring at the municipal waste management site.

Table 2.2 Daily averages of leachate cumulative flow, landfill gas (LFG) concentrations, and LFG flowrate for days which biofilm samples were collected. Daily leachate cumulative flow averages were collected from two pump stations, labelled PS(A) and PS(B). LFG concentrations were measured as percentages for CH_4 and CO_2. Averages were collected and provided by engineers and employees at the landfill site.

Date	Biofilm	Leachate cumulative flow (m^3)		LFG concentration		Gas plant LFG flowrate (cfm)
		PS(A)	PS(B)	CH_4 (%)	CO_2 (%)	
2017-10-25	BF1	134	58	50.8	0.4	1,140
2018-10-04	BF2	82	33	50.7	0.7	1,047
2019-08-23	BF3	54	18	48.1	0.7	1,082
2020-01-14	BF4	624	105	52.3	0.6	1,092
2019-07-09	BF5	63	18	48.8	1.4	1,109

from separate biofouling events within a gas collection system and leachate pumps at a municipal landfill site (Figure 2.1; Table 2.2). The microbial communities forming or existing within these biofilms were analyzed using a multi-omics approach to assess microbial diversity and function.

The first biofilm sample, BF1, was obtained from clogging of a condensate trap in the gas collection system. The samples BF2, BF4, and BF5 were collected from pump stations, which are locations near where leachate from multiple wells is collected, with BF2 and BF5 being collected from the same station, although at separate times and biofouling events. The BF3 sample was obtained from within the leachate collection system near a leachate well. The exact conditions surrounding each biofouling event (*i.e.*, atmospheric conditions at each site, or leachate composition) were not available. However, engineers at the landfill site were able to provide daily averages across the entire landfill site for leachate cumulative flow, landfill gas (LFG) concentrations, and LFG flowrate at the gas plant (Table 2.2). The LFG concentrations for CH_4 and CO_2 were consistent, with averages of 50.1% and 0.8%, respectively. The cumulative flow of leachate on the days of biofilm collection were measured from two different pump stations (Table 2.1; PS(A) and PS(B)). The cumulative flow rate from PS(A) ranged from 54 to 624 m^3, and PS(B) ranged from 13 to 105 m^3. But each pump station had the highest cumulative flow rate on the day BF4 was collected. This information provides insight to the average landfill conditions on the dates of biofilm collection. However, the current data does not provide information regarding conditions that are present in early biofilm development or throughout biofilm developmental stages. Nor does it provide information specific to the location of biofilm collection, but rather is a daily average across the landfill site. Future work would be enhanced with metadata more specific to

the region of the landfill where a biofilm sample is collected, and temporally throughout biofilm growth.

The structure of the microbial community was assessed using both 16S rRNA gene amplicon sequencing and metagenomics, with results compared between methods. Biofilms were physically similar and originated from the same sanitary landfill leachate and gas collections systems, and the corresponding microbial communities consisted of similar taxonomic lineages with variation in community composition and diversity (Figures 2.2 – 2.7). Evaluating the 16S rRNA gene profiles, there were fourteen phyla that were detected in all five biofilm samples (Figure 2.8). Proteobacteria, Firmicutes, Desulfobacterota, Campylobacterota, and Bacteroidota were the most prominent phyla detected in all five biofilms, comprising different proportions of the total profile. Proteobacteria dominated BF3, accounting for 49.6% of the total community. Firmicutes ranged from 5.1% of the total community in BF3 to 23.7% of BF1. Similarly, Desulfobacterota ranged from 3.4% of BF2 and BF3, to 23.9% of BF1. Campylobacterota accounted for 18.4% to 28.8% of the community in four biofilms, but only a small proportion of BF4, at 1.9% of the 16S rRNA gene profile. Bacteroidota was the most dominant phylum in BF4 (45.7%), but represented smaller proportions in the remaining biofilms, at 24.4% of the profile in BF5, and less than 5.5% of BF1, BF2, and BF3. BF4 was the only biofilm to contain ASVs from phyla Hydrogenedentes, LCP-89, Margulisbacteria, Myxococcota, Nitrospirota, Sumerlaeota, and Zixibacteria, though each accounted for less than 0.2% of the total community (Figure 2.8).

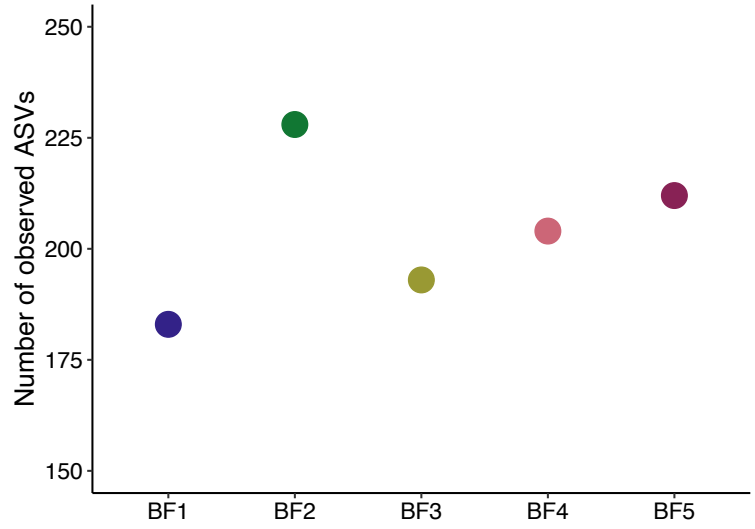


Figure 2.2 Observed number of ASVs for five biofilm samples from 16S rRNA gene sequencing.

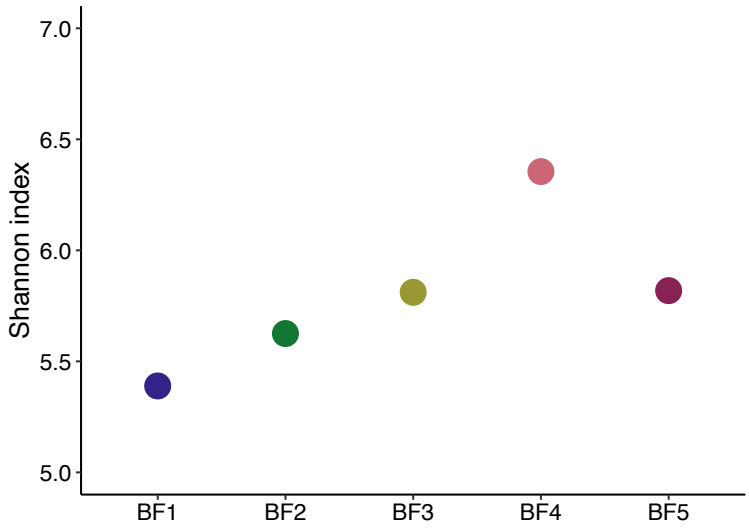


Figure 2.3 Shannon index as a measure of alpha-diversity from 16S rRNA gene sequences from five biofilm samples.

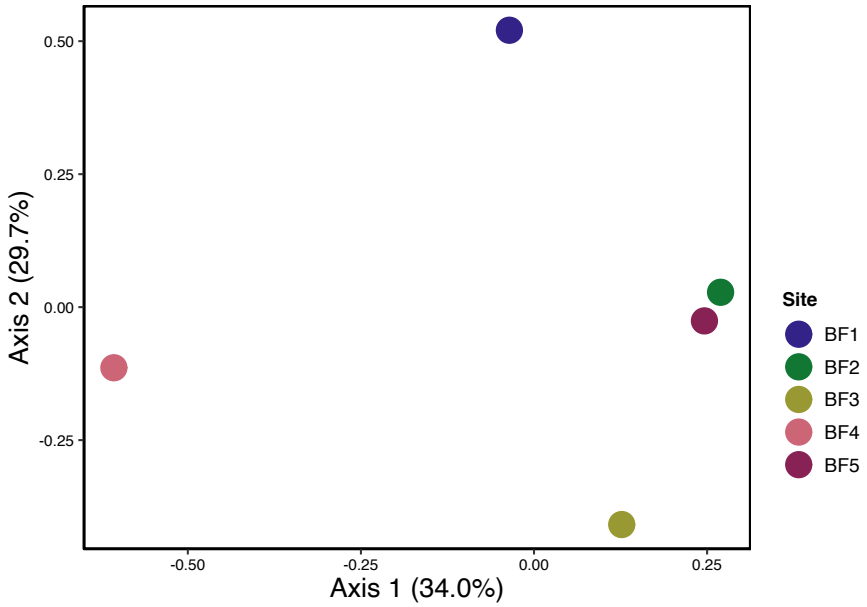


Figure 2.4 Principle-coordinate analysis (PCoA) ordination based on Bray-Curtis metric for beta-diversity of 16S rRNA gene sequences.

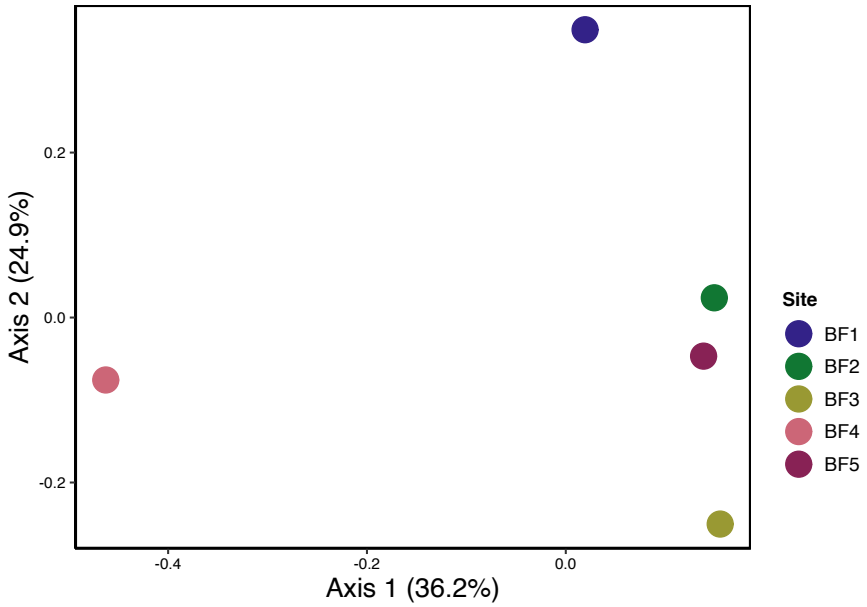


Figure 2.5 Principle-coordinate analysis (PCoA) ordination based on unweighted UniFrac metric for beta-diversity of 16S rRNA gene sequences.

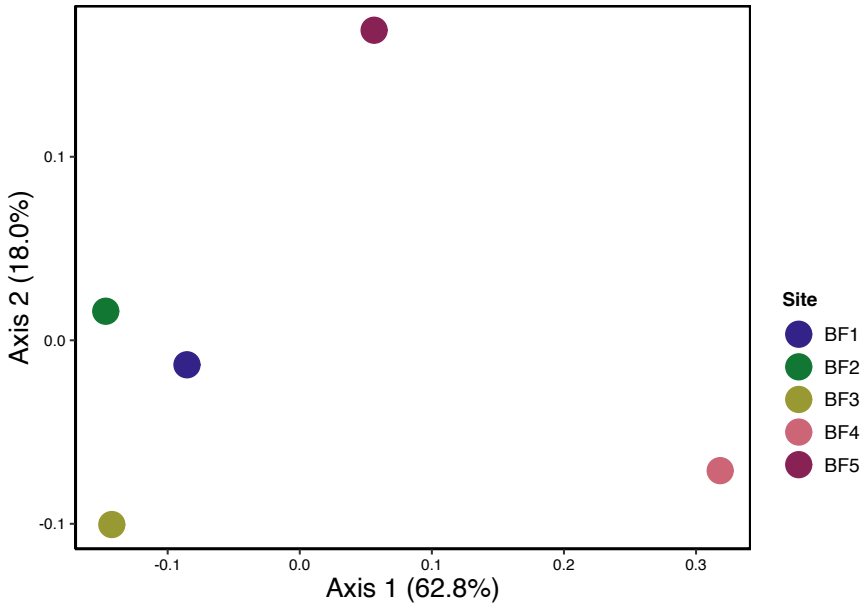


Figure 2.6 Principle-coordinate analysis (PCoA) ordination based on weighted UniFrac metric for beta-diversity of 16S rRNA gene sequences.

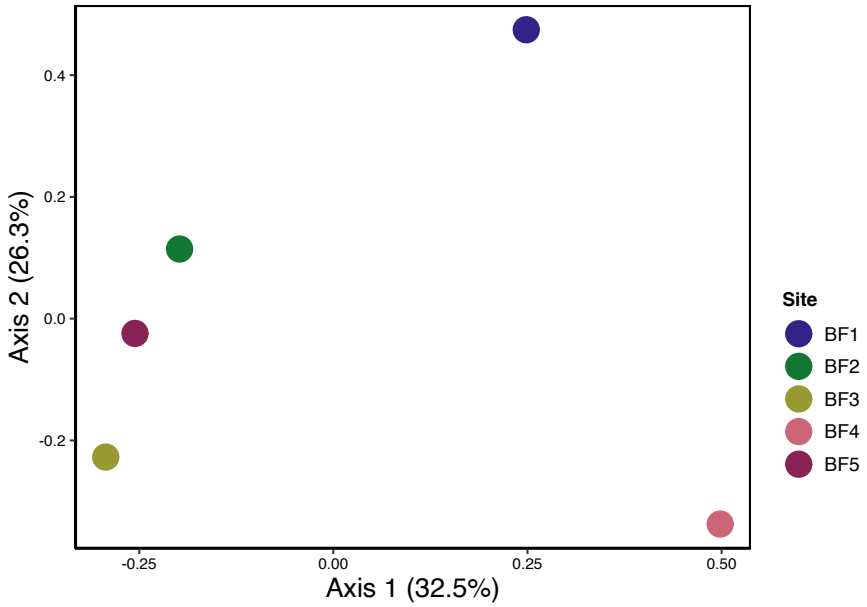


Figure 2.7 Principle-coordinate analysis (PCoA) ordination based on Jaccard metric for beta-diversity of 16S rRNA gene sequences.

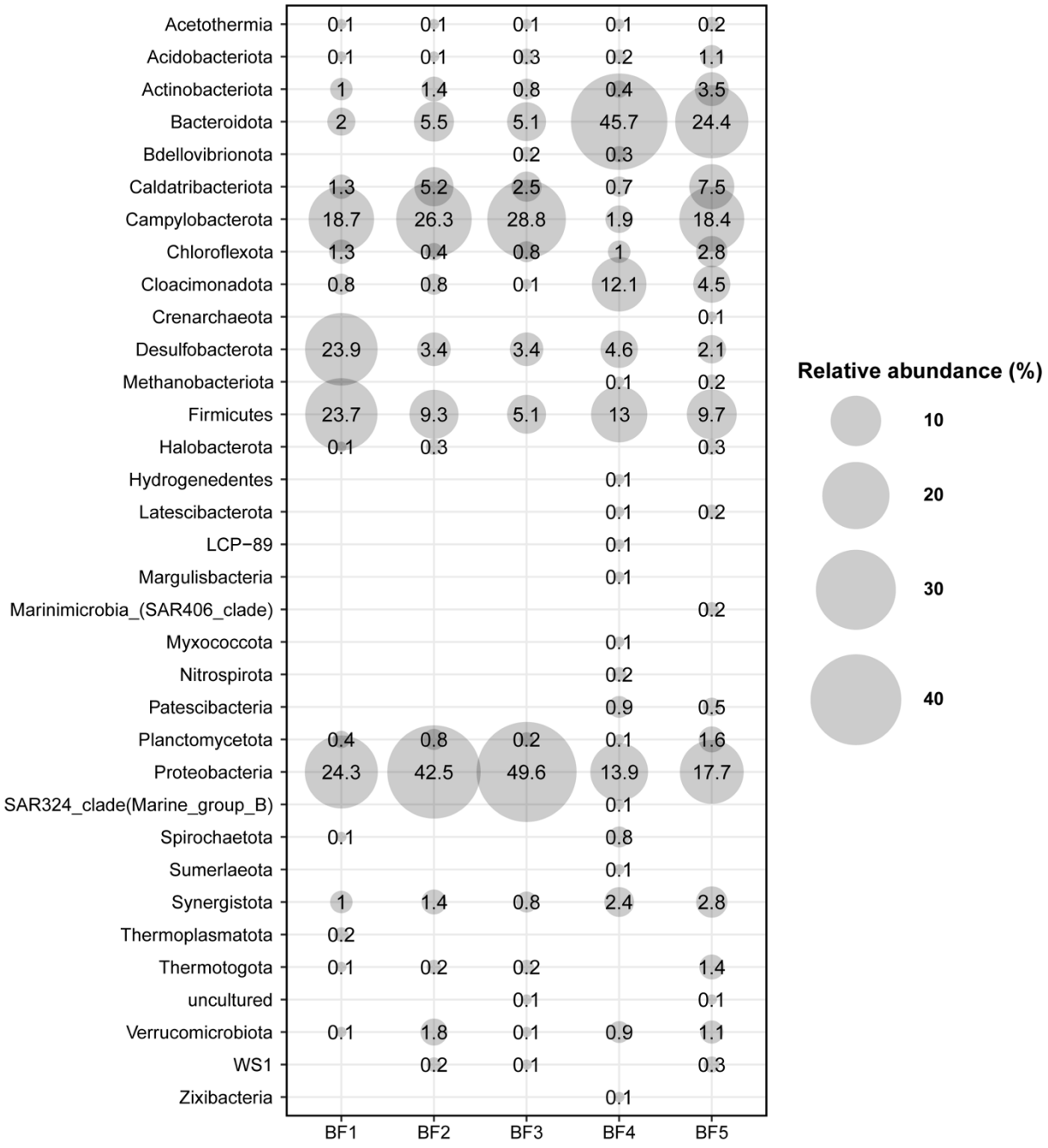


Figure 2.8 Phylum-level proportional abundance of amplicon sequence variants (ASVs). The number in each bubble represents the relative abundance of ASVs taxonomically classified to different phyla in a given sample, for the five biofilm samples.

At the genus level, community profiles were different between biofilms. Biofilms were not dominated by specific genera, with the highest genus-level proportion from *Acinetobacter* at 27.9% of BF2 (Figure 2.9). ASVs did not have species-level classification, and *Acinetobacter* is a metabolically diverse genus (Touchon *et al.*, 2014). *Acinetobacter baumannii* has received research attention because of its role in nosocomial infections and biofilm formation (Yang *et al.*, 2019), though as aerobic microorganisms are not expected to thrive in an anoxic environment. The most abundant genus in BF1 was *Pseudomonas* (21.3%), an organism known for its biofilm forming capabilities on a wide variety of surfaces and environments (Masák *et al.*, 2014). In BF3, the most abundant genus was *Sulfuricurvum,* containing the representative *Sulfuricurvum kujiense* known for oxidizing reduced sulfur compounds (Han *et al.*, 2012). *Sulfurimonas* was the genus with the highest abundance in BF5, a genus known for metabolic versatility in use of electron donors for energy generation (Han and Perner, 2015). BF4 did not contain any individual genera at high abundance that were common to the other biofilms. The highest abundance family was Rikenellaceae (13.3%; phylum Bacteroidota), which lacked taxonomic classification at the genus-level. BF5 had a slightly more even community than BF1, BF2, and BF3, with no genera above 20% abundance. *Sulfurimonas* was the most abundant genus in BF5 at 16.8%, similar to the 14.1% in BF2. On average, the biofilms contained 204 unique ASVs but BF4 had the highest Shannon diversity with a value of 6.4 and 204 unique ASVs (Table 2.3; Figures 2.2 and 2.3). With highest alpha diversity and containing distinct genera, BF4 was most dissimilar to other biofilms, which was reflected in visualizations of beta-diversity values (Figures 2.4 – 2.7). Taxonomic classification of metagenome-assembled genomes (MAGs) provided an alternate measure of community composition. As has been seen previously, amplicon sequencing of the 16S rRNA gene

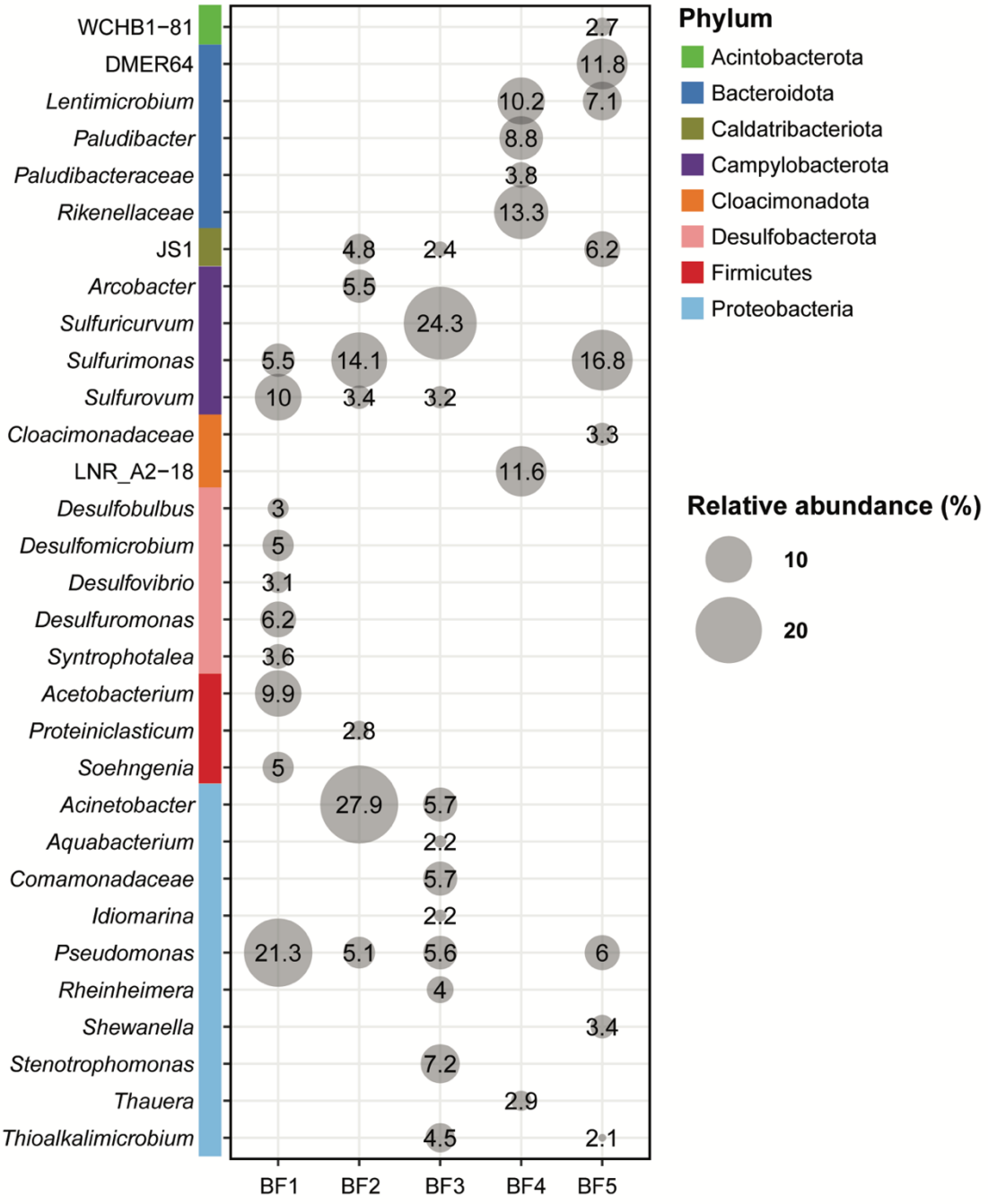


Figure 2.9 Genus-level proportional abundance of amplicon sequence variants (ASVs) present at over 1% abundance. The number in each bubble represents the relative abundance of ASVs taxonomically classified to different genera in the five biofilm samples. Only genera that were detected at over 1% abundance are displayed. The colored bar indicates the phylum-level classification for each genus.

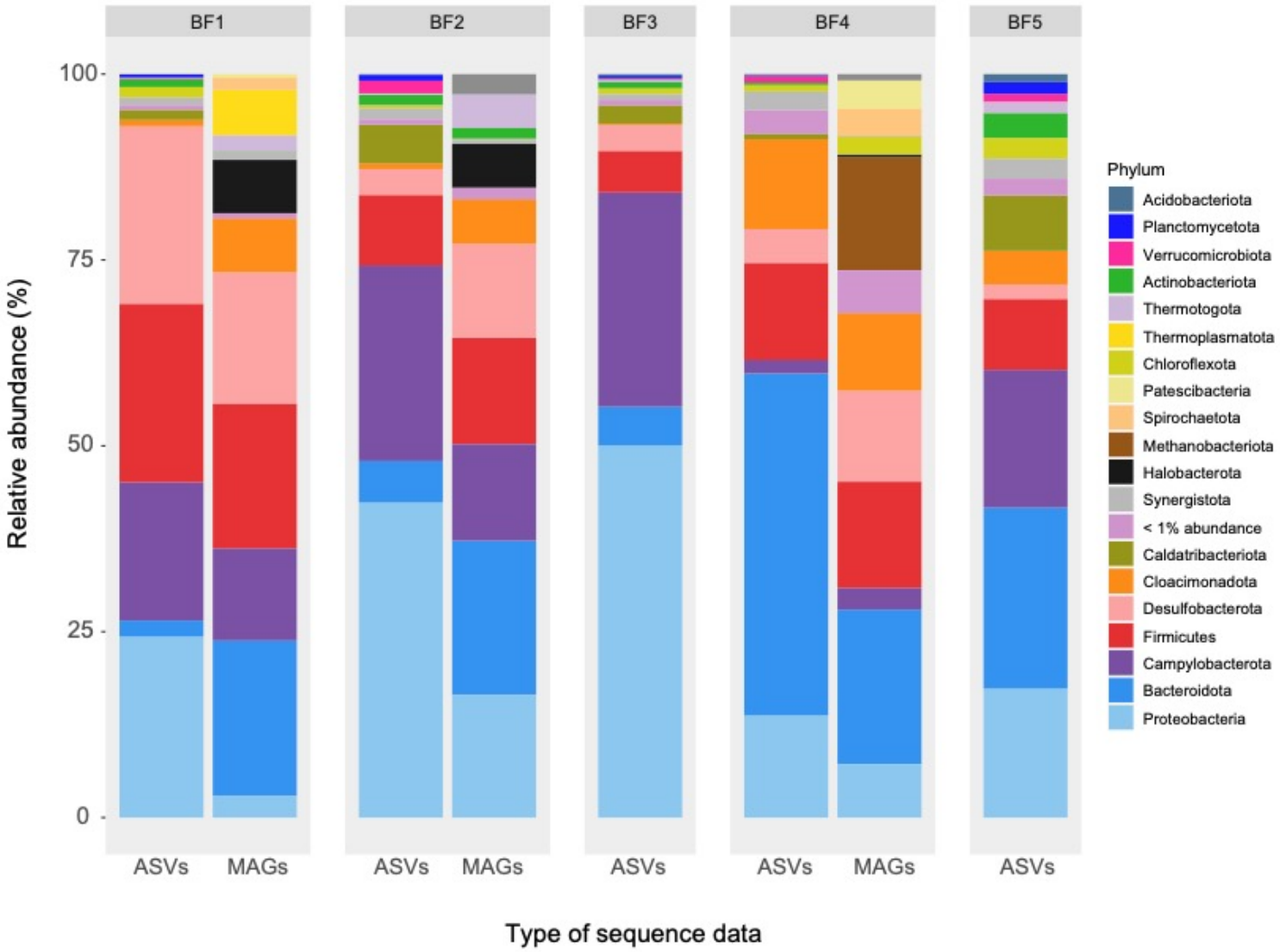
Table 2.3 Total number of unique ASVs and assembled MAGs in each biofilm sample. Metagenomic sequencing was not available for BF3 and BF5.

	ASVs	MAGs
BF1	183	106
BF2	228	51
BF3	193	-
BF4	204	187
BF5	212	-

and metagenome-based analyses yielded differences in community composition, despite both approaches using the same genomic DNA (Hassa *et al.*, 2018).

A total of 344 metagenome-assembled genomes (MAGs) were generated from metagenomic sequence data, with 106 generated from BF1, 51 from BF2, and 187 from BF4 (Table 2.3; Appendix A, Supplementary Table A1). To assess the overlap of similar organisms across biofilm samples, MAGs were dereplicated at 95% ANI with DRep, which approximates the expected DNA divergence at the species level. The original 344 MAGs were dereplicated to a total of 296 unique species-level MAGs. For each MAG, genome coverage was calculated based on the length of each scaffold and the scaffold's depth of read coverage in a metagenome. The abundance of each MAG was estimated as the proportion of coverage of each MAG from the total coverage of MAGs in the metagenome. Proteobacteria, Bacteroidota, Campylobacterota, Firmicutes, and Desulfobacterota were, as seen in the 16S rRNA gene profiles, the five most prominent phyla detected in the MAGs (Figure 2.10).

Figure 2.10 Relative abundance of phyla detected at greater than 1% abundance in at least one sample, comparing 16S rRNA gene amplicons and MAGs. Within the five biofilm samples, proportion was determined for amplicon sequence variants (ASVs) and metagenome-assembled genomes (MAGs). ASVs are based on ASV proportion within each biofilm sample. MAG abundance was calculated as proportion of an individual MAG's coverage from the total coverage of all MAGs within a biofilm sample. MAGs were only available for three biofilm samples, BF1, BF2, and BF4. Phyla detected below 1% abundance in all samples were combined into the category '<1% abundance'.


BF1
BF2
BF3
BF4
BF5
100
75
50
25
0
Relative abundance (%)
ASVs
MAGs
Type of sequence data
Phylum
Acidobacteriota
Planctomycetota
Verrucomicrobiota
Actinobacteriota
Thermotogota
Thermoplasmatota
Chloroflexota
Patescibacteria
Spirochaetota
Methanobacteriota
Halobacterota
Synergistota
< 1% abundance
Caldatribacteriota
Cloacimonadota
Desulfobacterota
Firmicutes
Campylobacterota
Bacteroidota
Proteobacteria

Unlike in the 16S rRNA amplicon data, in each metagenome Bacteroidota MAGs comprise the largest proportion of each community (Figure 2.11). In an additional divergence from the 16S rRNA amplicon community profiles, the second-most abundant phylum in BF4 was Methanobacteriota (classified as Euryarchaeota by NCBI). An individual MAG, BF4 Bin 74, belonging to the phylum Euryarchaeota accounted for 15% of the total community (Figure 2.11C). BF4 Bin 74 was classified to the genus *Methanothermobacter*, where *Methanothermobacter* are typically thermophilic, hydrogenotrophic methanogens (Wasserfallen *et al.*, 2000; Kaster *et al.*, 2011). Methanobacteriota accounted for 0.1% of the 16S rRNA gene profile in BF4 (Figure 2.8). There were no Methanobacteriota MAGs assembled or ASVs detected from BF1 or BF2. Instead, in BF1 and BF2, archaeal MAGs were present but were classified as Halobacteriota. Genera for these MAGs included *Methanosarcina, Methanothrix, Methanocorpusculum, Methanolobus,* and *Methanimicrococcus*, all exclusively methanogenic lineages. These assembled MAGs comprised 7% and 6% of the MAG abundances in BF1 and BF2 respectively, despite a low abundance of Halobacteriota in 16S rRNA gene profiles (BF1 = 0.1%; BF2 = 0.3%).

Biofilms BF1 and BF2 each contained two MAGs, BF dRep Bin 5 and BF dRep Bin 4, at high abundance. Classified to phylum Desulfobacterota and genus *Pseudodesulfuromonas*, BF dRep Bin 5 accounted for a relative abundance of 8% within the BF1 MAGs and 11% for BF2. The other MAG, BF dRep Bin 4, classified to phylum Campylobacterota and genus *Sulfurovum,* comprised 2% and 7% of the BF1 and BF2 metagenomes, respectively. Members of these genera are known to play a role in sulfur-cycling, and specifically, oxidation of sulfur compounds in *Sulfurovum* and reduction of elemental sulfur in *Pseudodesulfuromonas* (Kuever *et al.*, 2015; Stokke *et al.*, 2015).

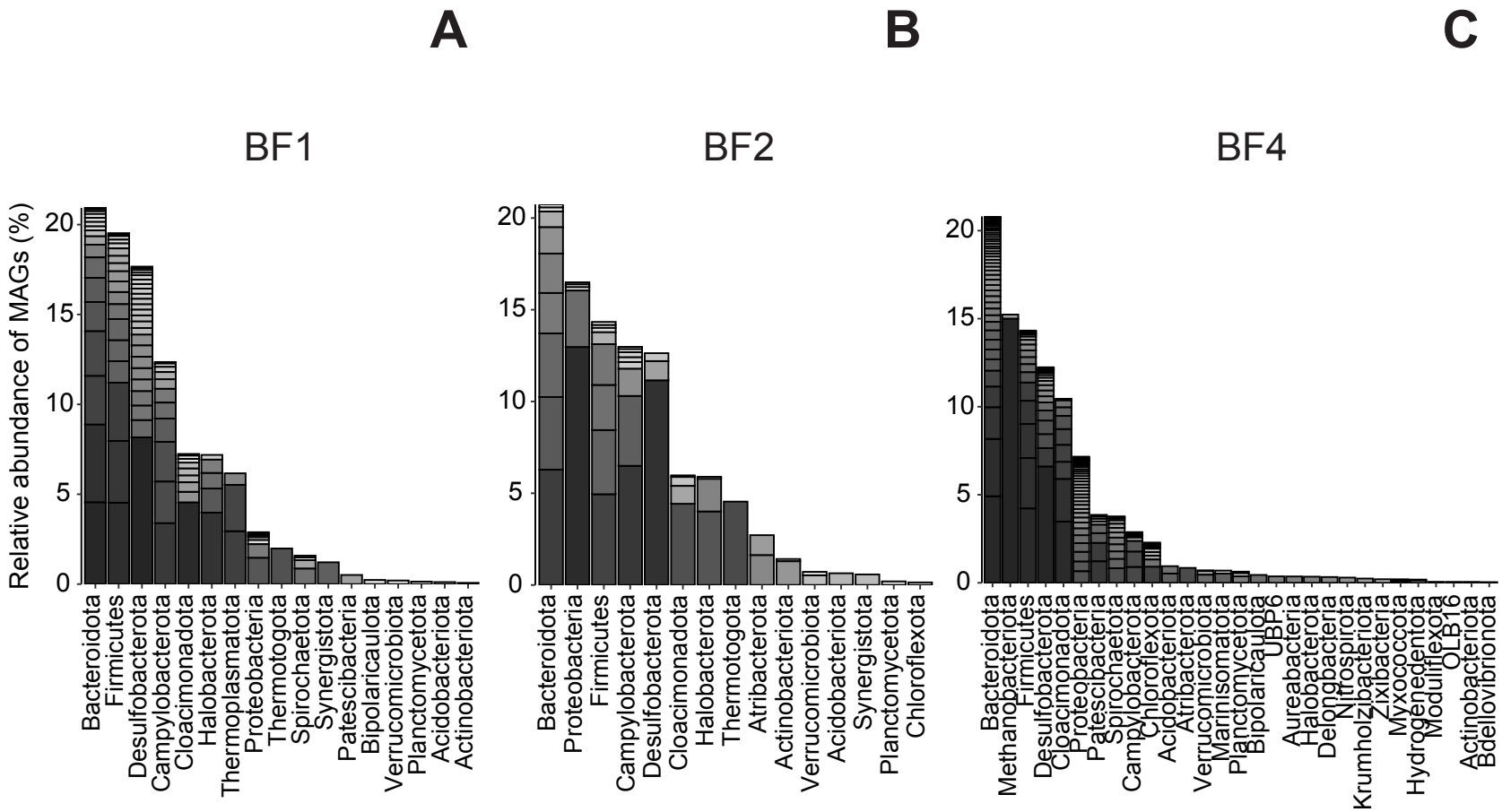


Figure 2.11 Rank abundance curve of the metagenome-assembled genomes (MAGs) from (A) BF1, (B) BF2, and (C) BF4. Bars correspond to a summed abundance for each phylum, with stacked boxes within a bar denoting individual MAG abundances. Relative abundances were calculated as a MAG's coverage from the total coverage of all MAGs within a biofilm sample. Stacked boxes are colored on a gradient with most abundant MAGs within a dataset shaded darkest, shading to lighter grey as abundance decreases.

2.3.2 Metabolic capacity of the microbial communities within biofilms

Metaproteomics and metametabolomics were used to complement the metagenome in assessing metabolic activity. Protein abundances were calculated accounting for sample depth and protein length and referred to as the normalized protein abundance (NPA), where each sample had a total NPA of 1,000,000. Metaproteomic determination identified a total of 5,466 proteins, or 0.12% of annotated genes. The detected proteins were associated to specific MAGs, as the dereplicated gene content of the metagenomes was used as the database to identify peptides. An average NPA of 422,052 was associated with unbinned proteins in the database for the three biofilm samples, with BF2 having the highest unbinned value with a NPA of 596,270 (Table 2.4). Alternatively, associated with MAGs, BF1 had an NPA of 634,686, BF2 with 403,730, and BF4 with 695,428 (63%, 40%, and 70% of the total NPA).

For proteins associated with MAGs, the taxonomic classification of MAGs then allowed for extrapolation of the taxonomic affiliation of actively expressed proteins. The distribution of phyla in the metagenomes and metaproteomes show differing proportions within and across biofilm samples (BF1 in Figure 2.12; BF2 in Figure 2.13; and BF4 in Figure 2.14). Desulfobacterota comprise the largest proportion of the proteins in BF1 with a NPA of 458,222, or 45% of the total NPA, compared to 18% of the relative abundance of Desulfobacterota MAGs

Table 2.4 Total NPA value that belongs to unbinned or MAGs within in proteins.

	Unbinned	Associated with MAGs
BF1	365,314	634,686
BF2	596,270	403,730
BF4	304,572	695,428

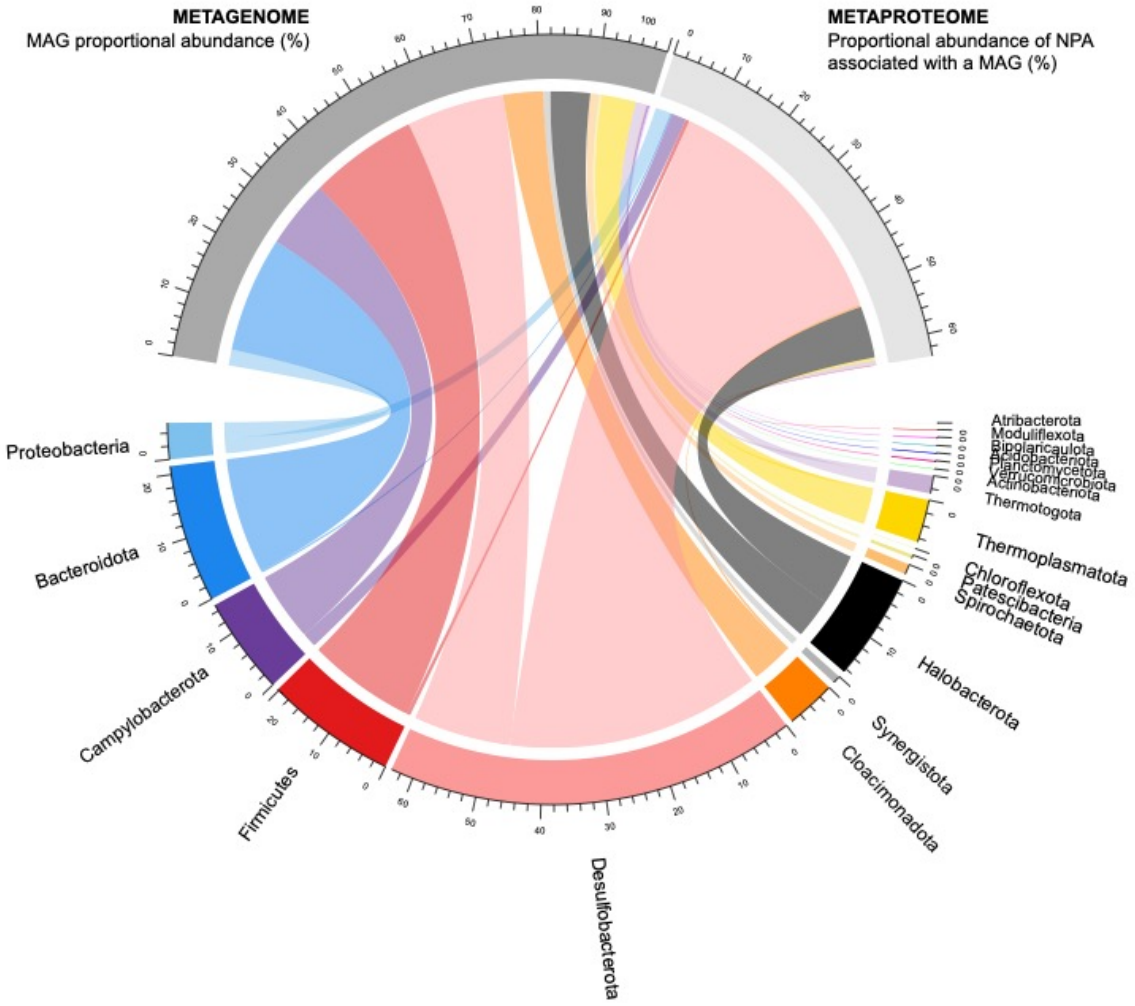


Figure 2.12 Distribution of abundance differences between the metagenome-assembled genomes (MAGs) and associated metaproteome in BF1. Dark grey scale on upper left quadrant corresponds to the percent abundance of MAGs in the metagenome grouped by phylum, and light grey scale on the upper right quadrant displays proportion of normalized protein abundances (NPA) in the metaproteome that are associated with MAGs, also grouped by phylum. The cross-circle ribbons are colored by phylum-level classification, and the width of the ribbon indicates the proportion of the metagenome and/or metaproteome that phylum comprises.

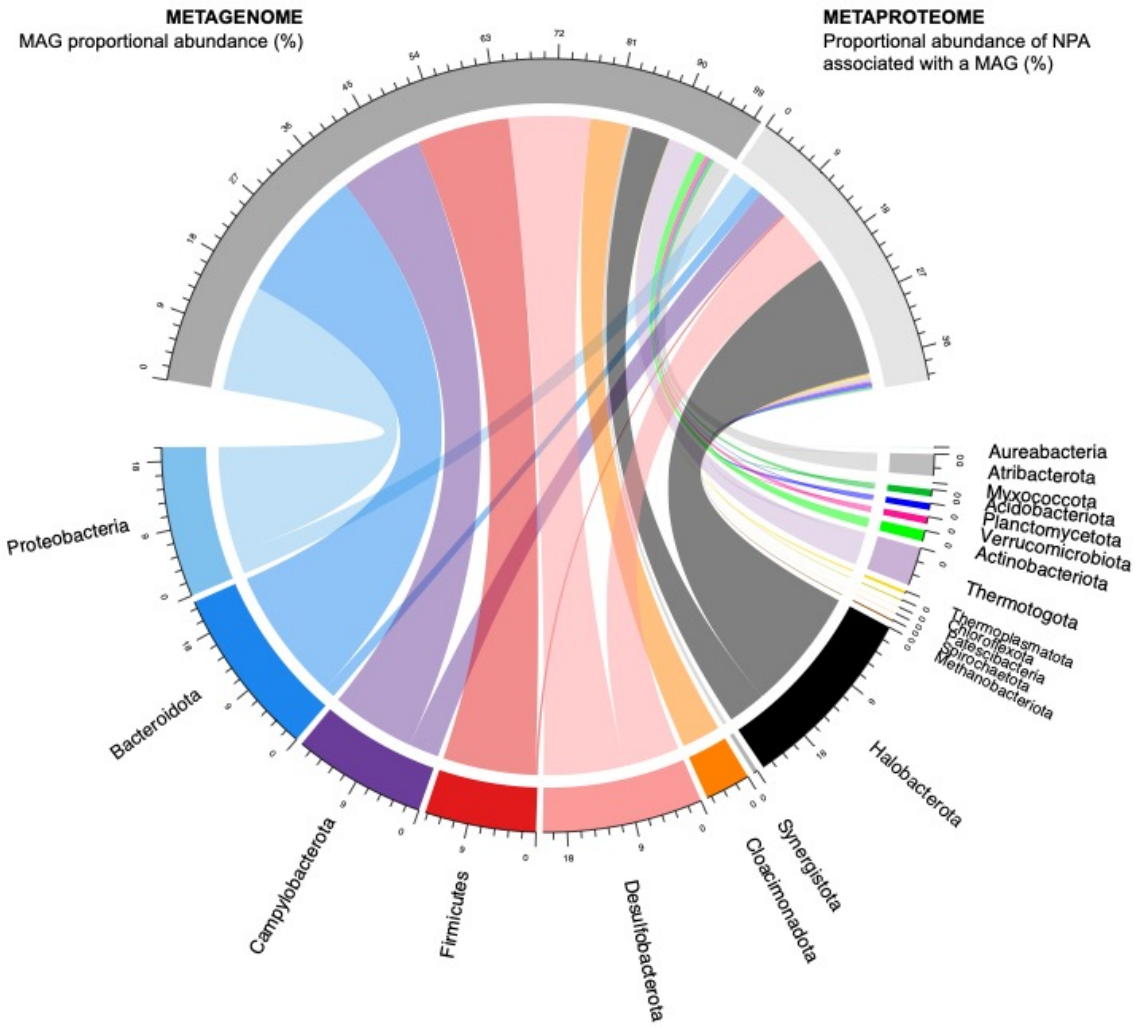


Figure 2.13 Distribution of abundance differences between the metagenome-assembled genomes (MAGs) and associated metaproteome in BF2. Dark grey scale on upper left quadrant corresponds to the percent abundance of MAGs in the metagenome grouped by phylum, and light grey scale on the upper right quadrant displays proportion of normalized protein abundances (NPA) in the metaproteome that are associated with MAGs, also grouped by phylum. The cross-circle ribbons are colored by phylum-level classification, and the width of the ribbon indicates the proportion of the metagenome and/or metaproteome that phylum comprises.

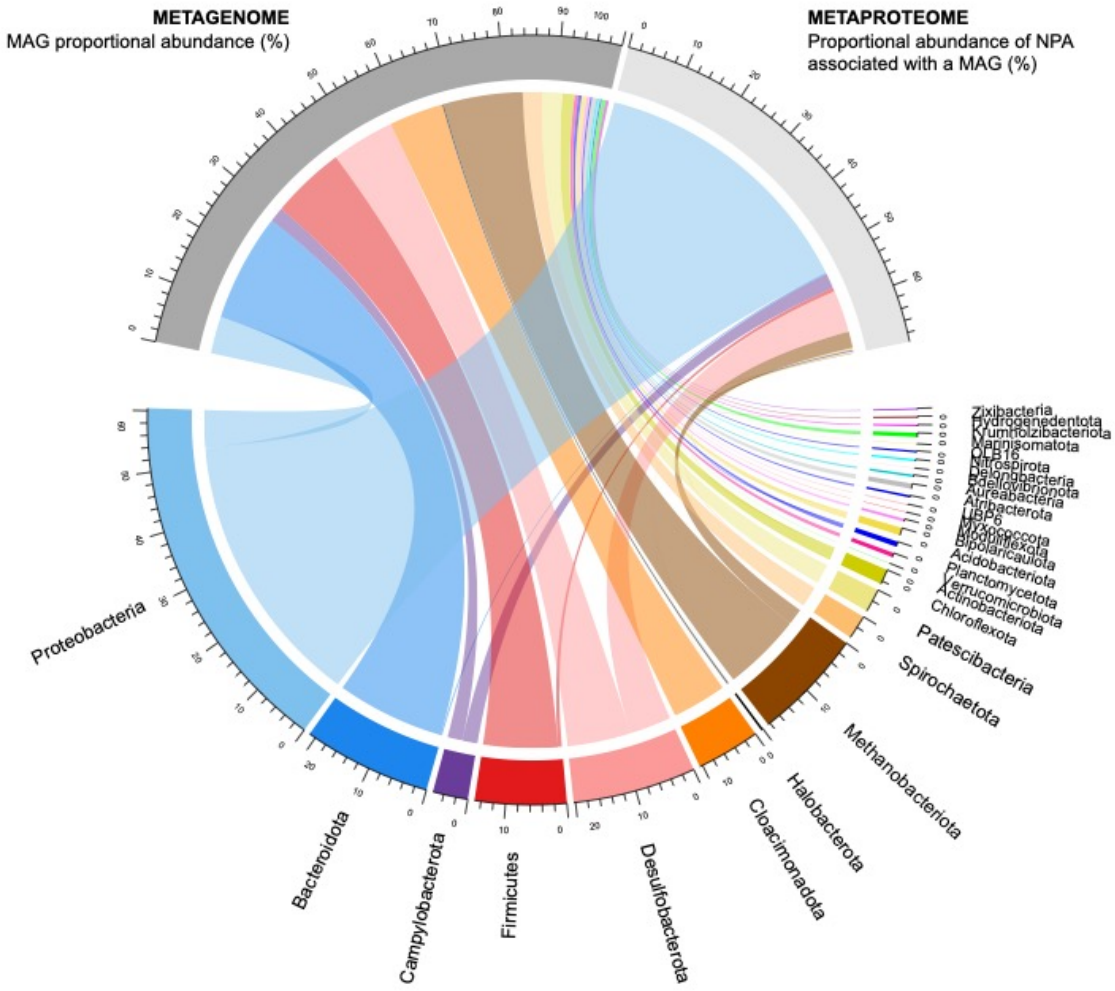


Figure 2.14 Distribution of abundance differences between the metagenome-assembled genomes (MAGs) and associated metaproteome in BF4. Dark grey scale on upper left quadrant corresponds to the percent abundance of MAGs in the metagenome grouped by phylum, and light grey scale on the upper right quadrant displays proportion of normalized protein abundances (NPA) in the metaproteome that are associated with MAGs, also grouped by phylum. The cross-circle ribbons are colored by phylum-level classification, and the width of the ribbon indicates the proportion of the metagenome and/or metaproteome that phylum comprises.

from the coverage of all MAGs in BF1. Desulfobacterota were also detected in BF2 proteins with a NPA of 85,751 (8.5% of total NPA compared to 13% of coverage in MAGs), and in BF4 with a NPA of 80,634 (8.0% of total NPA compared to 12% of coverage in MAGs). Halobacteriota was the second most abundant phylum in BF1 proteins (NPA of 93,484 (9.3%); 7.2% of coverage in MAGs) and was the most abundant by protein detection in BF2 with a NPA of 188,927 (19%), though only accounted for 5.9% of MAG relative abundance from total MAG coverage in BF2. Halobacteriota were sparsely detected in BF4 (NPA of 173 (0.02%); 0.3% of MAGs). In BF4 the most prominent phylum by protein detection was Proteobacteria with a NPA of 532,430 (53% total NPA), while only accounting for 7.1% of MAG relative abundance. Proteobacteria were present in BF1 with a NPA of 26,965 (2.3%; 2.9% of coverage in MAGs) and in BF2 with a NPA of 33,989 (3.4% NPA compared to 17% of coverage in MAGs). Bacteroidota, Firmicutes, and Cloacimonadota were phyla detected in all three biofilms with moderate to high abundance with the metagenome but comprised a small proportion of the identified proteins. Phylum Bacteroidota consisted of the highest relative abundance of MAGs in each biofilm's metagenomic data, accounting for an average of 21% of total coverage in MAG-associated DNA, but in the proteins had an average NPA of 8,030 (0.8%). Similarly, Firmicutes accounted for an average abundance of 16% from total coverage in MAGs across the three biofilms, but only an average NPA of 6,092 (0.6%). Cloacimonadota, while not dominant within the proteomics (average NPA of 1,604; 0.2%), were consistently in each biofilm sample, with an average abundance of 8% of total MAG coverage.

Beyond phylum-level affiliation of proteins, proteins were associated with specific MAGs. The distribution of MAGs was visualized on a phylogenetic tree, with outer rings displaying information on the relative abundance of MAGs in the metagenome and NPA in the metaproteome

(Figure 2.15). The phylogenetic tree was created with the 296 dereplicated MAGs using the GToTree workflow, where 32 genomes were removed due to having too few hits to the targeted single copy genes, resulting in 264 genomes on the tree (Lee, 2019). Two genomes that did not pass thresholds to make it onto the phylogenetic tree were BF1 Bin 64 and BF dRep Bin 10, both taxonomically classified as Halobacteriota. Both genomes were notable because of their association with a large number of proteins (Table 2.5). BF1 Bin 64 was the second most abundant MAG in BF1 based on protein abundance with a NPA of 69,261 (6.9% total NPA), and BF dRep Bin 10 was the most abundant bin in BF2 based on proteomics with a NPA of 146,409 (14.6% total NPA). BF2 is thus underrepresented in Figure 2.15's proteomic bar plot because the dominant MAG, BF dRep Bin 10, is not displayed on the tree.

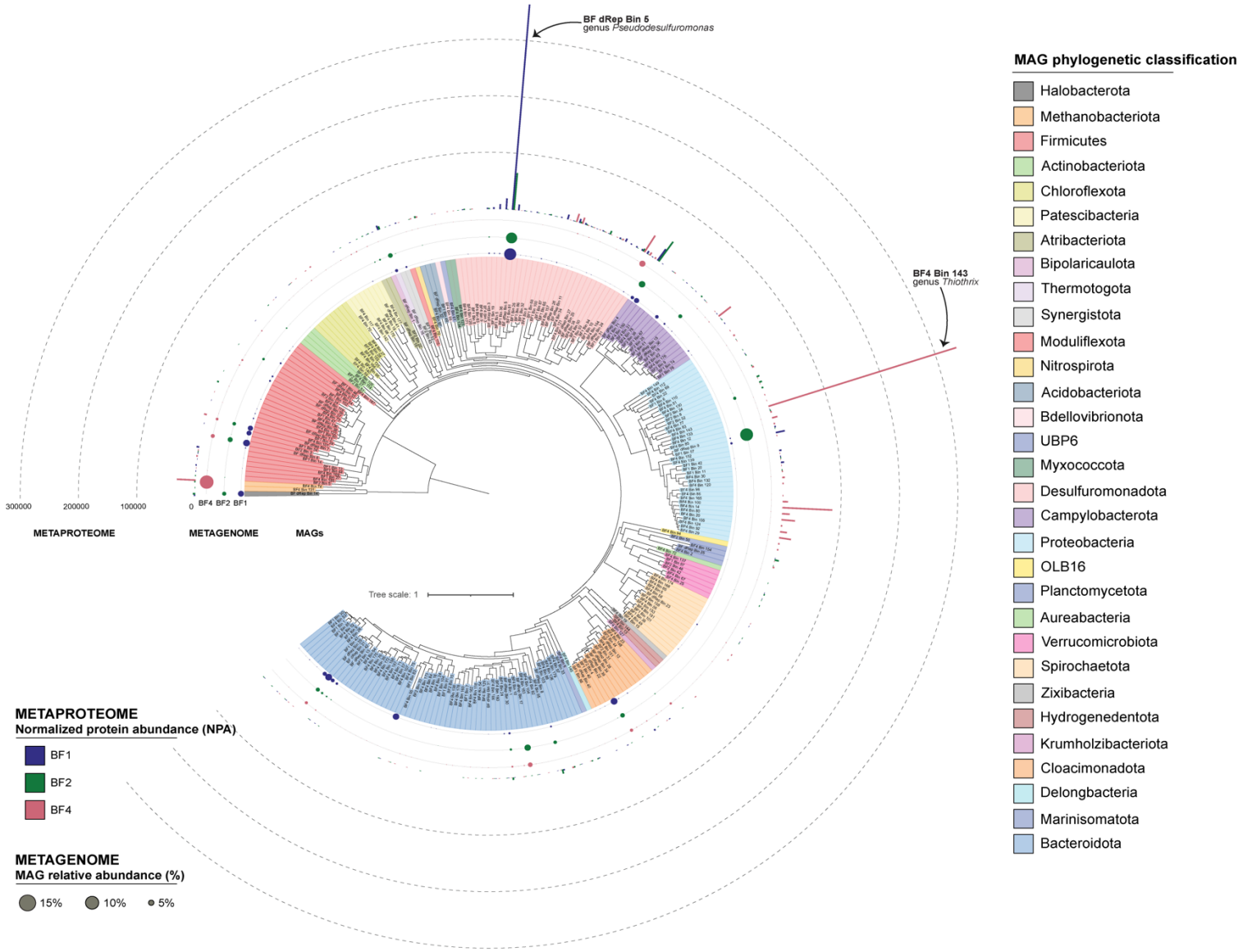
Table 2.5 Total NPA for the top 5 MAGs by protein abundance.

	MAG	Phylum	BF1	BF2	BF4
Top 5 MAGs in BF1	BF dRep Bin 5	Desulfobacterota	363,855		
	BF1 Bin 64	Halobacteriota	69,261		
	BF dRep Bin 4	Campylobacterota	24,035		
	BF1 Bin 36	Desulfobacterota	19,080		
	BF1 Bin 57	Proteobacteria	15,136		
Top 5 MAGs in BF2	BF dRep Bin 10	Halobacteriota		146,409	
	BF dRep Bin 5	Desulfobacterota		64,991	
	BF dRep Bin 4	Campylobacterota		41,760	
	BF2 Bin 39	Halobacteriota		34,762	
	BF2 Bin 14	Proteobacteria		17,686	
Top 5 MAGs in BF4	BF4 Bin 143	Proteobacteria			337,711
	BF4 Bin 100	Proteobacteria			85,078
	BF4 Bin 164	Desulfobacterota			33,900
	BF4 Bin 74	Methanobacteriota			31,601
	BF4 Bin 176	Campylobacterota			26,051

Each biofilm sample contains a dominant microorganism in terms of protein abundance (Table 2.5). The abundant organisms within the metaproteome are not always abundant in the metagenome (Appendix A, Supplementary Figure A1). In BF1, the most abundant MAG within the metaproteome was BF dRep Bin 5 (phylum Desulfobacterota, genus *Pseudodesulfuromonas*) with a NPA of 363,855 (36.3% of total NPA). In BF2 the most abundant MAG in the proteins was BF dRep Bin 10 (phylum Halobacteriota, genus *Methanothrix*) with an NPA of 146,409 (14.6% of total NPA), and in BF4 the most abundant MAG in the proteins was BF4 Bin 143 (phylum Proteobacteria, genus *Thiothrix*) with a NPA of 337,711 (33.7% of total NPA). These same MAGs that were dominant in the proteins of a biofilm sample corresponded to low relative abundances in the matching metagenome. In the metagenomic data from BF1, the coverage of BF dRep Bin 5 accounted for 8.2% of total MAG coverage (36.3% of total NPA). In BF2, BF dRep Bin 10 was associated with of 14.6% of total NPA with a corresponding relative abundance of MAG coverage in BF2 of 1.8%. Within BF4, Bin 143 comprised a total of 33.8% of NPA and only 0.2% of the total coverage of MAGs.

Within the proteins from BF dRep Bin 5 (phylum Desulfobacterota, genus *Pseudodesulfuromonas*) in the BF1 sample, the most abundant protein detected was citrate synthase (NPA of 160,712), followed by hypothetical proteins (NPA of 114,713). A total of 162 proteins were detected, and within the top 15 proteins, five were ribosomal proteins. Along with citrate synthase, malate dehydrogenase was the fourth most abundant protein and highlights functionality of the citric acid cycle. Acetyl-CoA hydrolase, which catalyzes the conversion of acetyl-CoA to acetate, is the sixth most abundant protein. Despite the representatives from the genus *Pseudodesulfuromonas* typically being associated with energy generation through the

Figure 2.15 Phylogenetic placement of dereplicated MAGs generated from biofilm metagenomes. MAGs are colored by phylum-level classification. Metagenome rings display dots whose circumferences are scaled to the relative abundance of the MAG within a biofilm sample and are colored by biofilm sample (BF1 = blue; BF2 = green; BF4 = pink). The outer metaproteome ring displays bars for each MAG, where bars are also colored by sample. Bars display the sum for normalized protein abundances (NPA) per MAG.


BF dRep Bin 5
genus Pseudodesulfuromonas
BF4 Bin 143
genus Thiothrix
300000
200000
100000
0
BF4 BF2 BF1
METAPROTEOME
METAGENOME
MAGs
Tree scale: 1
METAPROTEOME
Normalized protein abundance (NPA)
BF1
BF2
BF4
METAGENOME
MAG relative abundance (%)
15%
10%
5%
MAG phylogenetic classification
Halobacterota
Methanobacteriota
Firmicutes
Actinobacteriota
Chloroflexota
Patescibacteria
Atribacteriota
Bipolaricaulota
Thermotogota
Synergistota
Moduliflexota
Nitrospirota
Acidobacteriota
Bdellovibrionota
UBP6
Myxococcota
Desulfuromonadota
Campylobacterota
Proteobacteria
OLB16
Planctomycetota
Aureabacteria
Verrucomicrobiota
Spirochaetota
Zixibacteria
Hydrogenedentota
Krumholzibacteriota
Cloacimonadota
Delongbacteria
Marinisomatota
Bacteroidota

reduction of sulfur to hydrogen sulfide, the detected proteins involved in this pathway were not associated with BF dRep Bin 5 in BF1 (Kuever *et al.* 2015, Waite *et al.* 2020).

For BF2, the abundant MAG based on protein detected was BF dRep Bin 10 (phylum Halobacteriota, genus *Methanothrix*). The most abundant protein was acetyl-CoA decarbonylase/synthase complex beta (NPA of 21,331), followed by methyl-coenzyme M reductase beta subunit (NPA of 14,936), and acetyl-CoA synthetase (NPA of 11,846). In the top 15 proteins, were four subunits of a V/A-type H+-transporting ATPase, and three components of methyl-coenzyme M reductase (alpha, beta, and gamma). The acetyl-CoA decarbonylase/synthase complex is indicative of acetoclastic methanogenesis, and is likely playing a similar role here (Gencic and Grahame 2003).

In BF4, the abundant MAG was BF4 Bin 143 (phylum Proteobacteria, genus *Thiothrix*), where the protein with the highest NPA was a hypothetical protein (NPA of 52,637), OOP family OmpA-OmpF porin (NPA of 28,758), and phasin family protein (NPA of 14,663). The OOP family are outer membrane proteins, and phasin family proteins are associated with polyhydroxyalkanoate (PHA) storage (Confer and Ayalew 2013, Mezzina and Pettinari 2016). Within the top 15 proteins, four were ribosomal proteins. There was also a F-type H+-transporting ATPase subunit b with a NPA of 6,068. Adenylylsulfate reductase subunit A and B had a NPA value of 7,425 and 8,823, respectively, consistent with other *Thiothrix* species (Mardanov *et al.* 2020). Though often associated with dissimilatory sulfate reduction, adenylylsulfate reductase (*aprAB*) in sulfur-oxidizing bacteria, such as *Thiothrix*, catalyzes the indirect oxidation of sulfite along with sulfate adenylyltransferase (*sat*) (Nguyen *et al*., 2022). However, to ensure this gene was not inaccurately binned within the MAG BF4 Bin 143, the quality of the scaffold containing

the gene was examined. The scaffold was 803,526 base pairs long, with 829 predicted genes, and in the JGI-IMG database the scaffold was also taxonomically identified as belonging to the *Thiothrix* genus. The scaffold contained both *aprAB* subunits, followed by the *sat* gene. Together, the quality of the scaffold does not suggest misplacement of the scaffold within BF4 Bin 143, or misclassification of the MAG as *Thiothrix*, but rather that this sulfur-oxidizing microorganisms uses *aprAB* for sulfite oxidation.

Though the taxonomic distribution in the metagenome and metaproteome are variable across biofilms, some metabolic processes are consistent. Here, the presence of metabolic pathways was based on the presence or detection of marker genes for key enzymes of each metabolic pathway of interest (Table 2.1). Within the top proteins by function in BF1 was methyl-coenzyme M reductase alpha, beta, and gamma subunits, with the gamma subunit being the fourth most abundant protein. Together the complex catalyzes the last step of the methanogenesis pathway and comprised a total NPA of 69,232 in BF1. Similarly, the complex was prominent in the top proteins of BF2 with a total NPA of 203,859, where the gamma subunit was the second most abundant protein. Proportionally less significant in BF4, methyl-coenzyme M reductase's alpha subunit was still the 16th most abundant protein, and the complex comprised a total NPA of 16,893 (1.7%).

Another set of proteins abundant in BF1 was adenylylsulfate reductase subunit A and B, an enzyme in the dissimilatory sulfate reduction pathway. In BF1, subunit A was the sixth most abundant protein, both subunits comprised a total NPA of 33,285. In BF2, adenylylsulfate reductase was not the most abundant protein, but subunit B was the 23rd most abundant, but both subunits together had a total NPA of 11,007. In BF4 this enzyme was again very highly expressed,

being the third and fifth most abundant proteins and a total NPA of 76,367. When evaluating the top proteins in each biofilm sample, methyl-coenzyme M reductase and adenylylsulfate reductase were common between all three biofilms. However, to evaluate other enzymes within nitrogen, sulfur, methane, and reactive oxygen species (ROS) defense pathways, gene content in the metagenome and proteins in the metaproteome were evaluated.

2.3.2.1 Nitrogen cycle

The capacity for nitrogen metabolism in the biofilms was limited. In the absence of oxygen, nitrate is a common terminal electron acceptor via denitrification (Lam and Kuypers 2011). Denitrification was the main pathway in the nitrogen cycle that was observed, with a stronger proteomic signal in BF4 compared to BF1 or BF2 (Figure 2.16; Figure 2.17). In BF4, the mean NPA across the pathway was 1,508 with an average of 35 MAGs encoding genes across the pathway. In each biofilm sample, the entire denitrification pathway was encoded and, therefore, the potential for this function exists. In BF1 nitrogen cycling was limited, where nitrous oxide (NosZ) was the only protein that was expressed and associated with a MAG, with an NPA of 12. Catalyzing the terminal step of denitrification converting nitrous oxide (N_2O) to nitrogen gas (N_2), NosZ expression was associated with BF1 Bin 20, classified to the genus *Castellaniella* in the Burkholderiaceae family. Members of the *Castellaniella* have been described as facultatively anaerobic, denitrifying microorganisms (Kämpfer *et al.* 2006). In BF4, expressed proteins for each step of denitrification were detected except for nitric oxide reductase (NorB). Nitrate reductase (NarG and NapA), nitrite reductase (NirS), and NosZ proteins were associated with MAGs. BF4 Bin 176, classified to genus *Aliarcobacter* (phylum Campylobacterota), expressed the most complete denitrification pathway with NapA, NirS, and NosZ. Other MAGs with expressed

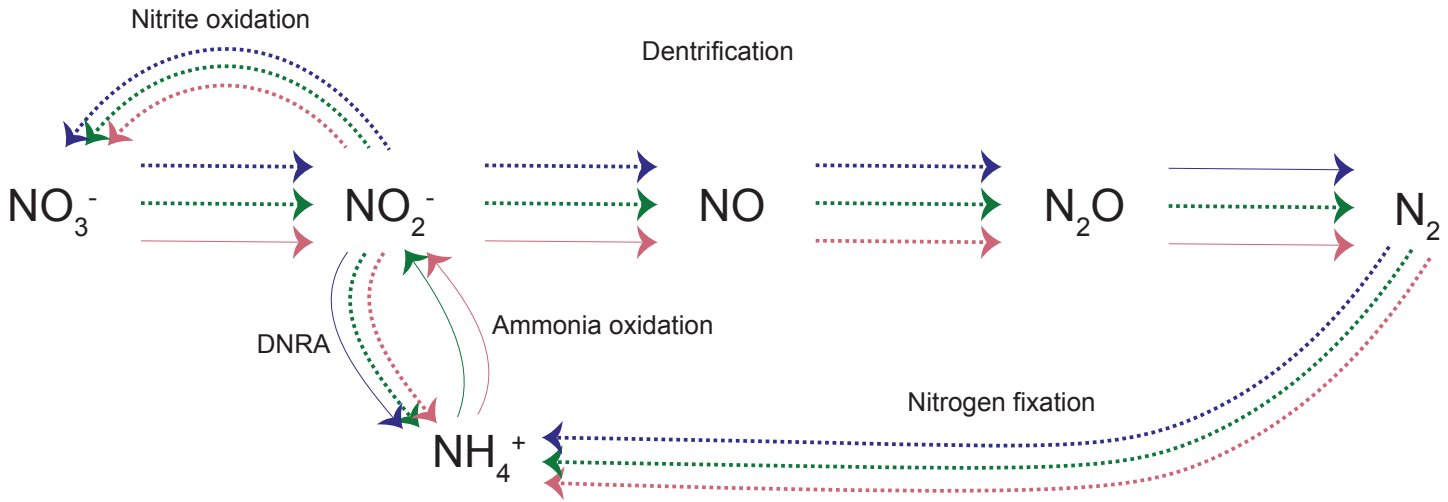
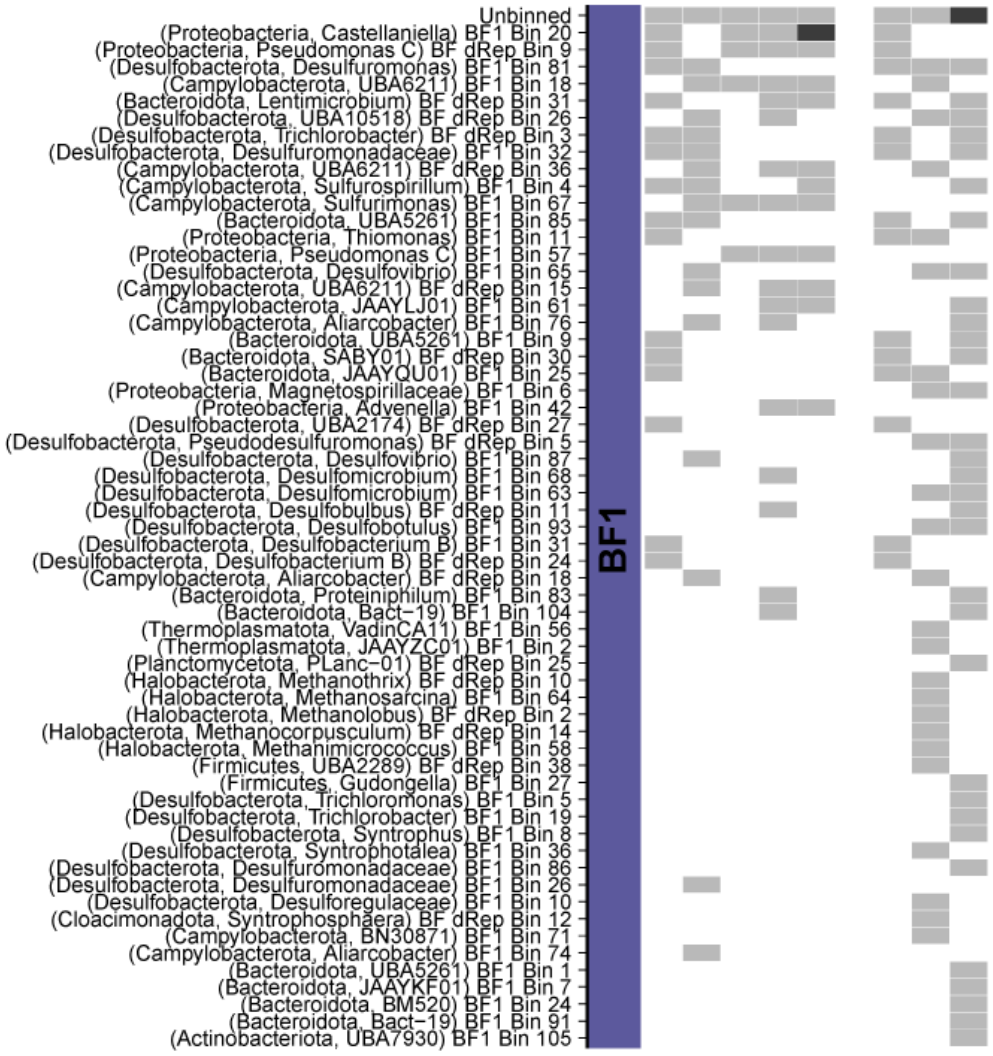
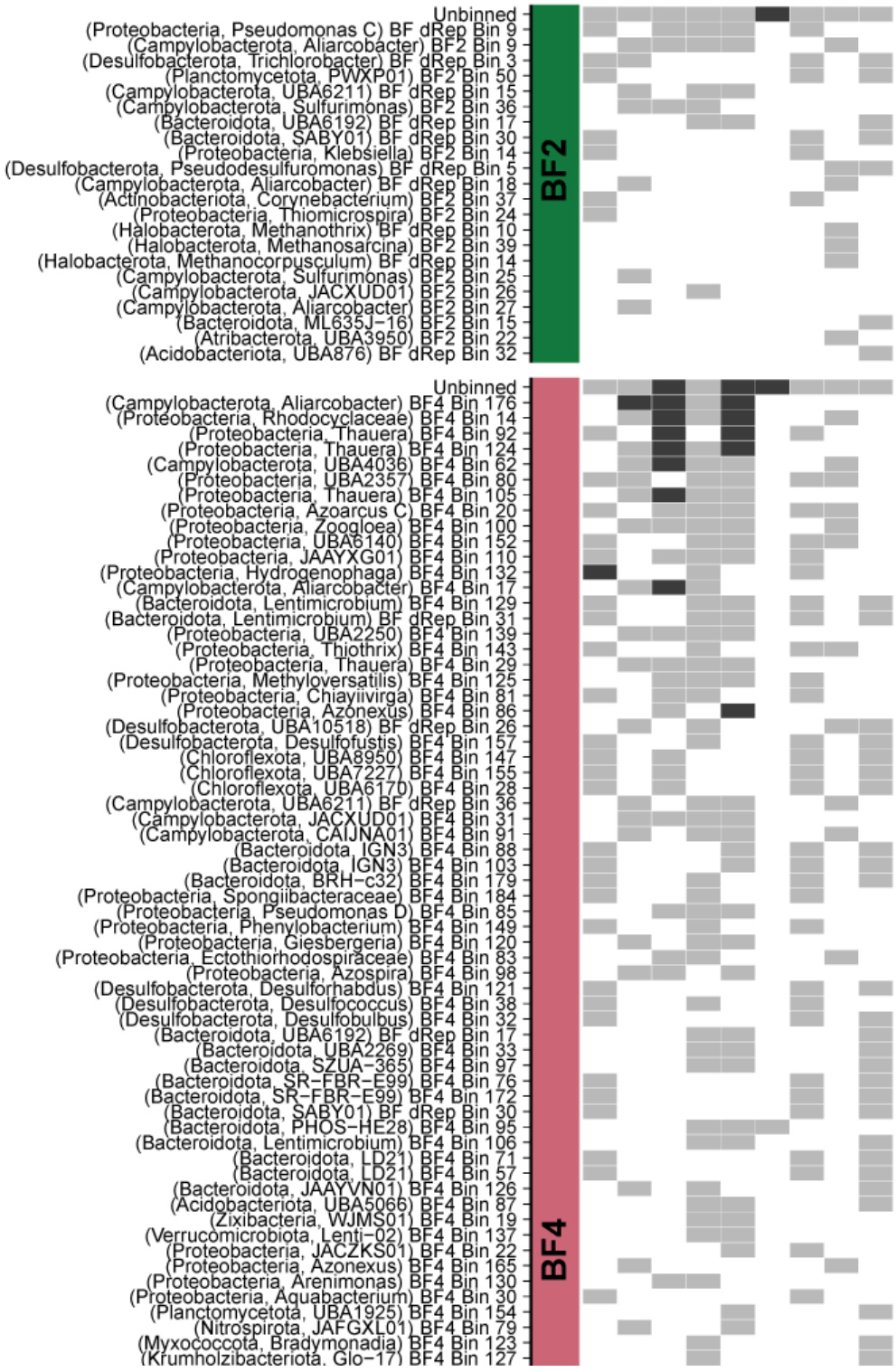
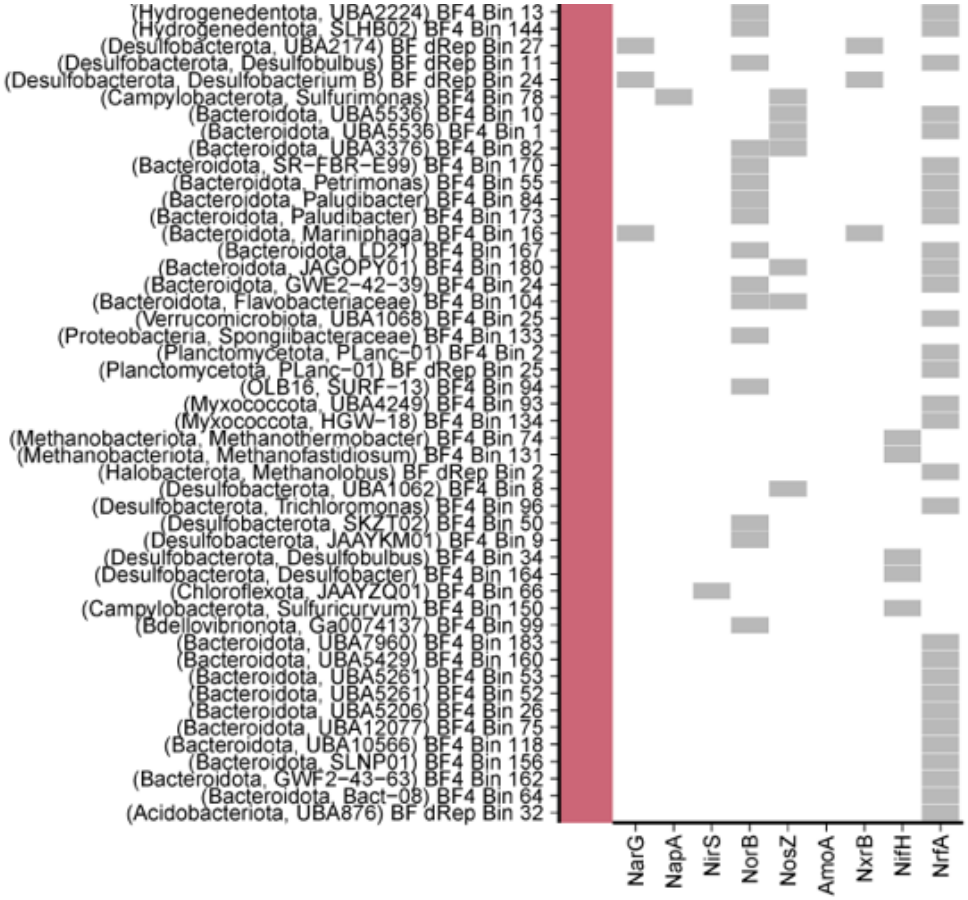


Figure 2.16 Predicted activity for nitrogen cycling in biofilms. Arrows were included if the important enzymes were detected in the metagenome or metaproteome, where matches could be associated with specific metagenome-assembled genomes (MAGs) or unbinned. Arrows are colored by biofilm sample (BF1 = blue; BF2 = green; BF4 = pink). Solid arrows indicate presence in the metaproteome and are weighted to the sum of normalized protein abundance (NPA) within a biofilm sample. Dotted arrows are not weighted and indicate detection in the metagenome only. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1. DNRA = Dissimilatory nitrate reduction to ammonia.

Figure 2.17 Presence or absence of important nitrogen metabolism enzymes corresponding to MAGs or unbinned in biofilm metagenomes and metaproteomes. Detection within MAGs and unbinned data are grouped and colored by biofilm samples (BF1 = blue; BF2 = green; BF4 = pink). MAG labels include taxonomy in brackets (Phylum, Genus) and the MAG ID. Light grey boxes indicate detection in the metagenome only, medium grey boxes indicate detection in metaproteome only, and dark grey indicates detection in both metagenome and metaproteome. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1. Protein abbreviations are: NarG = nitrate reductase alpha subunit; NapA = nitrate reductase (cytochrome); NirS = nitrite reductase (NO-forming); NorB = nitric oxide reductase subunit B; NosZ = nitrous oxide reductase; AmoA = ammonia monooxygenase subunit A; NxrB = nitrate reductase beta subunit; NifH = nitrogenase iron protein; NrfA = nitrite reductase (cytochrome c-552). Figure 2.11 is presented across three pages.




Unbinned
(Proteobacteria, Pseudomonas C) BF dRep Bin 9
(Campylobacterota, Aliarcobacter) BF2 Bin 9
(Desulfobacterota, Trichlorobacter) BF dRep Bin 3
(Planctomycetota, PWXP01) BF2 Bin 50
(Campylobacterota, UBA6211) BF dRep Bin 15
(Campylobacterota, Sulfurimonas) BF2 Bin 36
(Bacteroidota, UBA6192) BF dRep Bin 17
(Bacteroidota, SABY01) BF dRep Bin 30
(Proteobacteria, Klebsiella) BF2 Bin 14
(Desulfobacterota, Pseudodesulfuromonas) BF dRep Bin 5
(Campylobacterota, Aliarcobacter) BF dRep Bin 18
(Actinobacteriota, Corynebacterium) BF2 Bin 37
(Proteobacteria, Thiomicrospira) BF2 Bin 24
(Halobacterota, Methanothrix) BF dRep Bin 10
(Halobacterota, Methanosarcina) BF2 Bin 39
(Halobacterota, Methanocorpusculum) BF dRep Bin 14
(Campylobacterota, Sulfurimonas) BF2 Bin 25
(Campylobacterota, JACXUD01) BF2 Bin 26
(Campylobacterota, Aliarcobacter) BF2 Bin 27
(Bacteroidota, ML635J-16) BF2 Bin 15
(Atribacterota, UBA3950) BF2 Bin 22
(Acidobacteriota, UBA876) BF dRep Bin 32
BF2
Unbinned
(Campylobacterota, Aliarcobacter) BF4 Bin 176
(Proteobacteria, Rhodocyclaceae) BF4 Bin 14
(Proteobacteria, Thauera) BF4 Bin 92
(Proteobacteria, Thauera) BF4 Bin 124
(Campylobacterota, UBA4036) BF4 Bin 62
(Proteobacteria, UBA2357) BF4 Bin 80
(Proteobacteria, Thauera) BF4 Bin 105
(Proteobacteria, Azoarcus C) BF4 Bin 20
(Proteobacteria, Zoogloea) BF4 Bin 100
(Proteobacteria, UBA6140) BF4 Bin 152
(Proteobacteria, JAAYXG01) BF4 Bin 110
(Proteobacteria, Hydrogenophaga) BF4 Bin 132
(Campylobacterota, Aliarcobacter) BF4 Bin 17
(Bacteroidota, Lentimicrobium) BF4 Bin 129
(Bacteroidota, Lentimicrobium) BF dRep Bin 31
(Proteobacteria, UBA2250) BF4 Bin 139
(Proteobacteria, Thiothrix) BF4 Bin 143
(Proteobacteria, Thauera) BF4 Bin 29
(Proteobacteria, Methyloversatilis) BF4 Bin 125
(Proteobacteria, Chiayiivirga) BF4 Bin 81
(Proteobacteria, Azonexus) BF4 Bin 86
(Desulfobacterota, UBA10518) BF dRep Bin 26
(Desulfobacterota, Desulfofustis) BF4 Bin 157
(Chloroflexota, UBA8950) BF4 Bin 147
(Chloroflexota, UBA7227) BF4 Bin 155
(Chloroflexota, UBA6170) BF4 Bin 28
(Campylobacterota, UBA6211) BF dRep Bin 36
(Campylobacterota, JACXUD01) BF4 Bin 31
(Campylobacterota, CAIJNA01) BF4 Bin 91
(Bacteroidota, IGN3) BF4 Bin 88
(Bacteroidota, IGN3) BF4 Bin 103
(Bacteroidota, BRH-c32) BF4 Bin 179
(Proteobacteria, Spongiibacteraceae) BF4 Bin 184
(Proteobacteria, Pseudomonas D) BF4 Bin 85
(Proteobacteria, Phenylobacterium) BF4 Bin 149
(Proteobacteria, Giesbergeria) BF4 Bin 120
(Proteobacteria, Ectothiorhodospiraceae) BF4 Bin 83
(Proteobacteria, Azospira) BF4 Bin 98
(Desulfobacterota, Desulforhabdus) BF4 Bin 121
(Desulfobacterota, Desulfococcus) BF4 Bin 38
(Desulfobacterota, Desulfobulbus) BF4 Bin 32
(Bacteroidota, UBA6192) BF dRep Bin 17
(Bacteroidota, UBA2269) BF4 Bin 33
(Bacteroidota, SZUA-365) BF4 Bin 97
(Bacteroidota, SR-FBR-E99) BF4 Bin 76
(Bacteroidota, SR-FBR-E99) BF4 Bin 172
(Bacteroidota, SABY01) BF dRep Bin 30
(Bacteroidota, PHOS-HE28) BF4 Bin 95
(Bacteroidota, Lentimicrobium) BF4 Bin 106
(Bacteroidota, LD21) BF4 Bin 71
(Bacteroidota, LD21) BF4 Bin 57
(Bacteroidota, JAAYVN01) BF4 Bin 126
(Acidobacteriota, UBA5066) BF4 Bin 87
(Zixibacteria, WJMS01) BF4 Bin 19
(Verrucomicrobiota, Lenti-02) BF4 Bin 137
(Proteobacteria, JACZKS01) BF4 Bin 22
(Proteobacteria, Azonexus) BF4 Bin 165
(Proteobacteria, Arenimonas) BF4 Bin 130
(Proteobacteria, Aquabacterium) BF4 Bin 30
(Planctomycetota, UBA1925) BF4 Bin 154
(Nitrospirota, JAFGXL01) BF4 Bin 79
(Myxococcota, Bradymonadia) BF4 Bin 123
(Krumholzibacteriota, Glo-17) BF4 Bin 127
BF4



denitrification proteins in BF4 included Proteobacteria from the genera *Thauera*, which includes known denitrifiers (Pal *et al.* 2018), and *Hydrogenophaga,* a genus with members capable of hydrogenotrophic denitrification (Liu *et al.* 2020). In BF2 there were no proteins associated with the denitrification pathway, and an average of 5 MAGs encoded each step of the pathway.

The dissimilatory nitrate reduction to ammonia (DNRA) activity in the biofilms was limited. Nitrite reductase (NrfA) was only expressed in BF1 (NPA of 99), where it was not associated with a particular MAG, and was only detected in the metagenomes for BF2 and BF4. *nrfA* was encoded in 32 MAGs from BF1, 7 MAGs in BF2, and 56 MAGs in BF4. Additionally, the nitrogen fixation activity was very limited because the nitrogenase iron proteins (NifH) were

not found in any biofilm sample in the proteins, though *nifH* genes were encoded in the metagenomes, from 24, 7, and 17 genomes in samples BF1, BF2, and BF4, respectively. Genes involved in anaerobic ammonia oxidation (anammox), hydrazine synthase (*hzs*) and hydrazine dehydrogenase (*hdh*) genes, were not detected in the biofilm metagenomes or metaproteomes.

There was one MAG from the phylum Nitrospirota, BF4 Bin 79, that was assembled from the BF4 metagenome. Several members from this phylum, specifically members of the genus *Nitrospira,* are known to implement complete ammonia oxidation (comammox) (Daims *et al.* 2015). However, BF4 Bin 79 was classified to the Dissulfurispiraceae family and did not encode the enzymes needed for comammox. Some members from Dissulfurispiraceae utilize thiosulfate disproportionation, and the thiosulfate reductase (*pshA*) gene was encoded in BF4 Bin 79, though not detected in the expressed proteins (Umezawa *et al.* 2021) (Figure 2.18; Figure 2.19).

Overall, genes involved in nitrogen metabolism were most often detected in BF4 and from the denitrification pathway. However, nitrogen cycle-related proteins comprised a small NPA in each sample (99 in BF1, 5,435 in BF2, and 6,573 in BF4). Compared to other metabolisms, nitrogen cycling does not appear to be an important function for the microbial communities within these biofilms.

2.3.2.2 Sulfur cycle

In contrast to nitrogen cycling, sulfur cycling was a large component of the metabolic activity demonstrated through expressed proteins in the biofilms (Figure 2.18; Figure 2.19). Specifically, dissimilatory sulfate reduction was prominent. Dissimilatory sulfate reduction is an anaerobic process which reduces sulfate (SO_4^{2-}) to sulfide (H_2S) and is inhibited by oxygen. In this pathway, ATP sulfurylase (Sat) first activates sulfate to adenosine-5'-phosphosulfate (APS,

also called adenylylsulfate). The second enzyme, adenylylsulfate reductase (Apr) converts APS to sulfite, and dissimilatory sulfite reductase (Dsr) executes the final step converting sulfite (SO_3^{2-}) to sulfide (H_2S). Apr was one of the highest expressed function in the biofilms. In BF1, BF2, and BF4, AprA accounted for NPA values in BF1 of 21,669, BF2 of 3,377, and BF4 of 48,695. In the metagenome, *aprA* was encoded in 13 MAGs in BF1 and 16 MAGs in BF4, but not encoded in any MAGs associated with BF2. Instead, both *aprA* and *dsrA* in BF2 were associated with unbinned sequences. Thought typically associated with dissimilatory sulfate reduction, Apr and Sat in sulfur-oxidizing bacteria is associated with sulfite oxidation (Nguyen *et al*., 2022). Proteins involved in the other steps of the pathways were at lower abundance in the metaproteome. Sat was present in biofilms with an average NPA of 2,532 and DsrA with an average NPA of 2,034, with detection in an average of 22 and 9 MAGs.

Sulfur cycling processes beyond dissimilatory sulfate reduction were detected in the metagenomes and metaproteomes, but at lower overall abundances. Conversion of sulfide to elemental sulfur (S_0), or sulfide oxidation, is catalyzed by either sulfide:quinone oxidoreductase (Sqr) or sulfide dehydrogenase flavoprotein chain (Fcc). Detection or prediction of this function is based on presence of either Sqr or FccB, or both. In the metagenome, *fccB* was encoded in an average of four MAGs, and *sqr* encoded in an average of 29 MAGs. The average NPA of these proteins in the biofilm samples was 2,565, with the highest expression occurring in BF4 (6,643 NPA). Sulfite dehydrogenases (Sor and Soe) catalyze sulfite oxidation, converting sulfite to sulfate. Soe was not detected in any metagenomes or metaproteomes. However, *sorA* was encoded in all three biofilm metagenomes, associated with 10 MAGs in BF4 and the unbinned sequences of BF1 and BF2. In the proteins, SorA was detected only in BF4 with a NPA of 1,265.

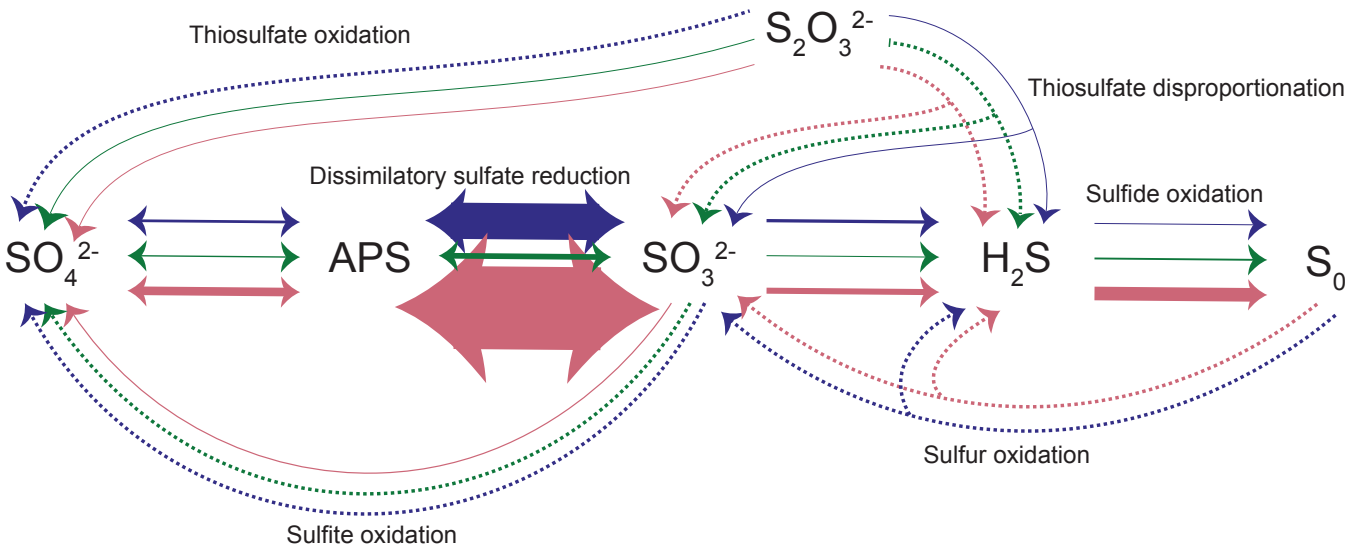
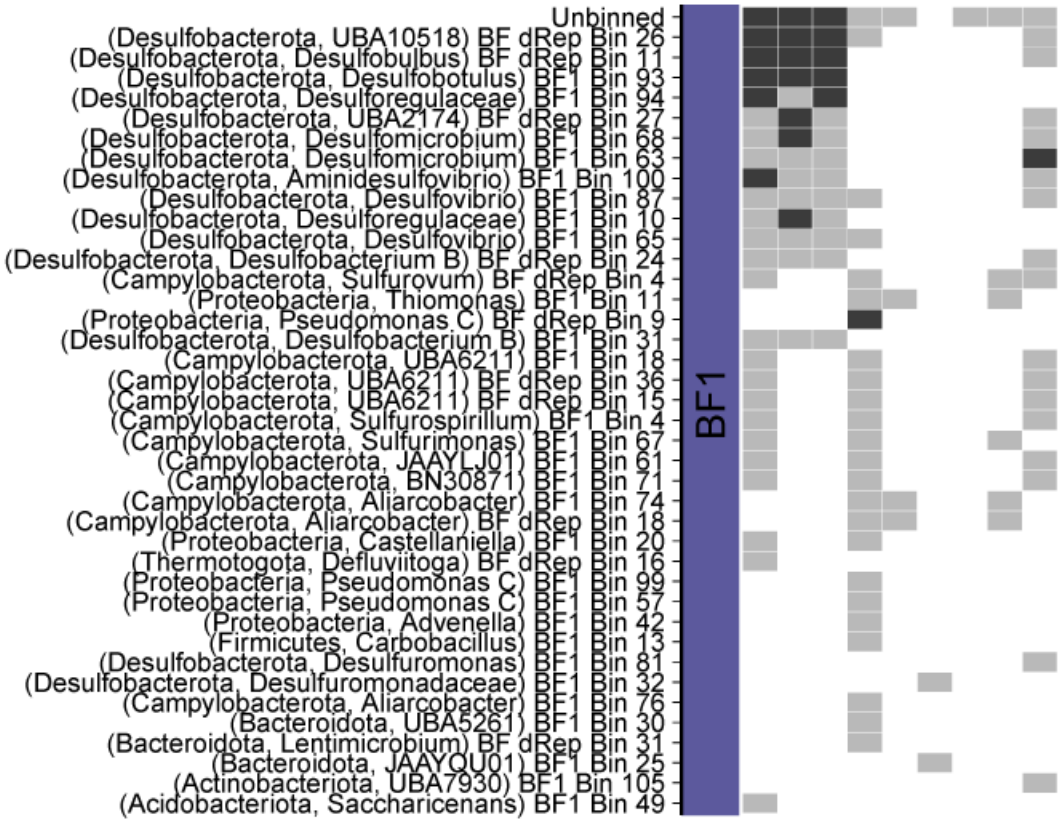
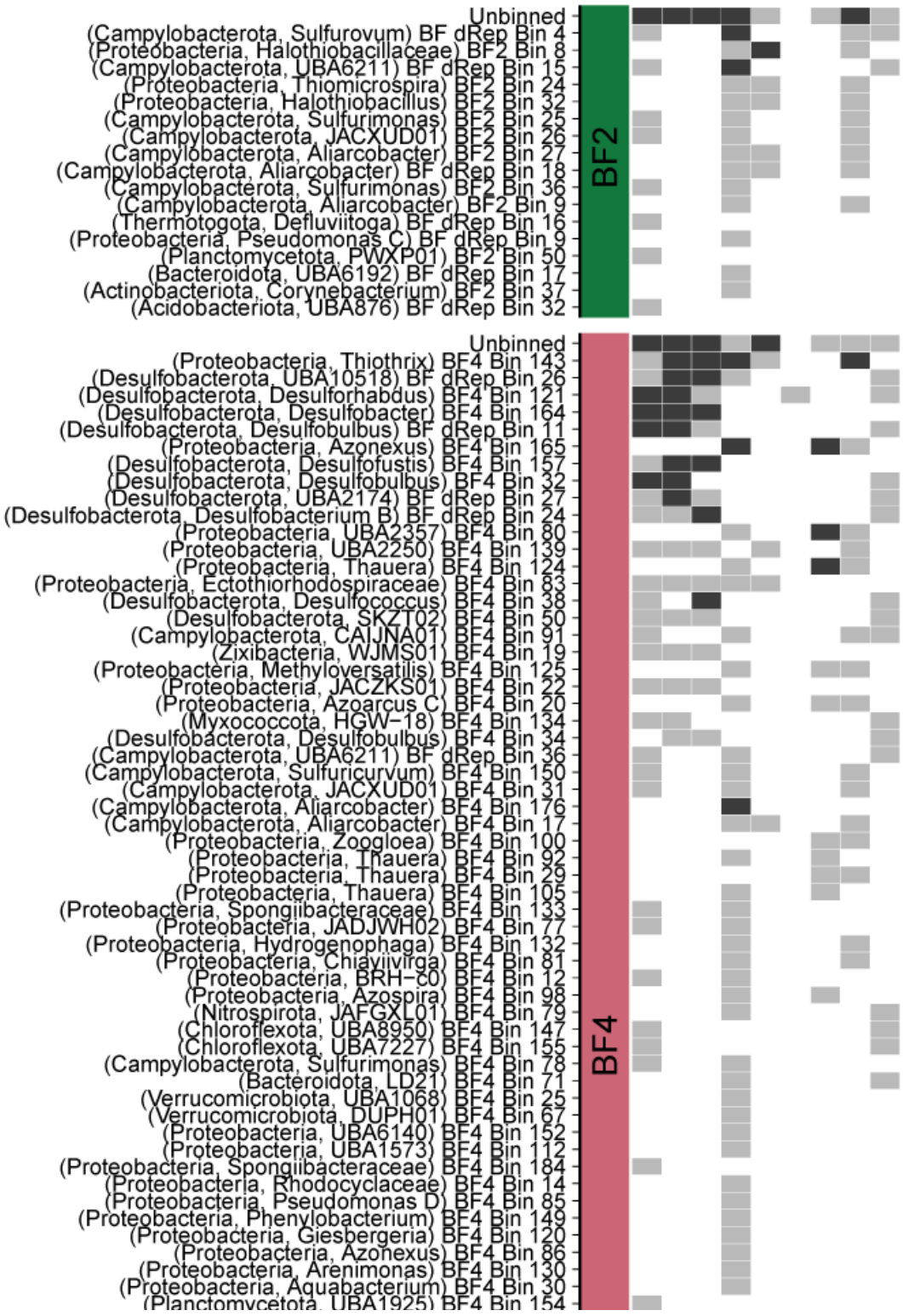
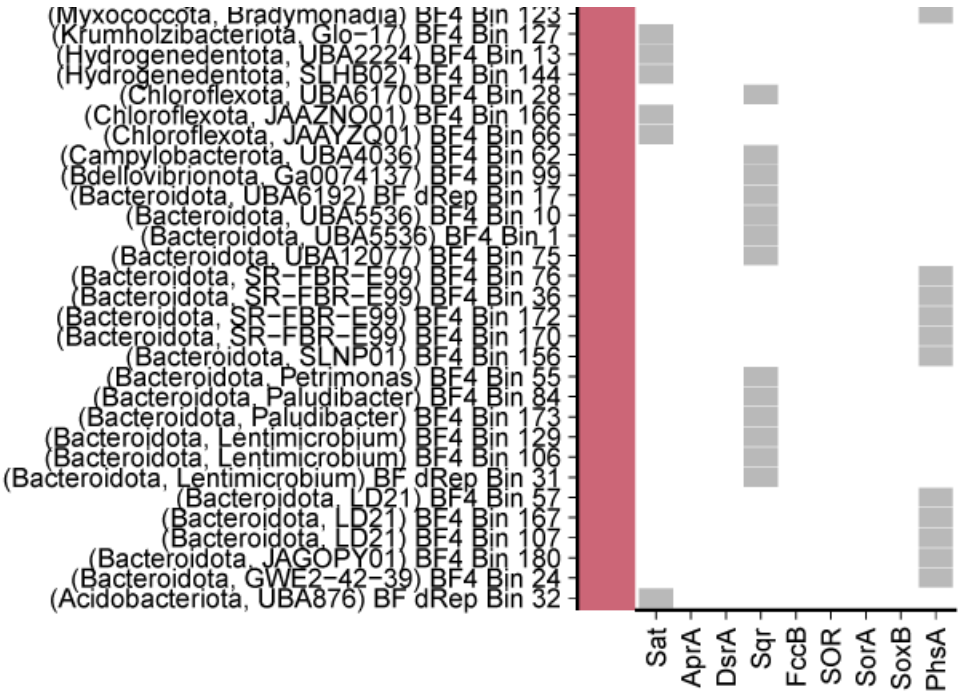


Figure 2.18 Predicted activity for sulfur cycling in biofilms. Arrows were included if the important enzymes were detected in the metagenome or metaproteome, where matches could be associated with specific metagenome-assembled genomes (MAGs) or unbinned. Arrows are colored by biofilm sample (BF1 = blue; BF2 = green; BF4 = pink). Solid arrows indicate presence in the metaproteome and are weighted to the sum of normalized protein abundance (NPA) within a biofilm sample. Dotted arrows are not weighted and indicate detection in the metagenome only. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1.

Figure 2.19 Presence or absence of important sulfur metabolism enzymes corresponding to MAGs or unbinned in biofilm metagenomes and metaproteomes. Detection within MAGs and unbinned data are grouped and colored by biofilm samples (BF1 = blue; BF2 = green; BF4 = pink). MAG labels include taxonomy in brackets (Phylum, Genus) and the MAG ID. Light grey boxes indicate detection in the metagenome only, medium grey boxes indicate detection in metaproteome only, and dark grey indicates detection in both metagenome and metaproteome. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1. Protein abbreviations are: Sat = sulfate adehylyltransferase; AprA = adenylylsulfate reductase subunit A; DsrA = dissimilatory sulfite reductase alpha subunit; Sqr = sulfide:quinone oxidoreductase; FccB = sulfide dehydrogenase [flavocytochrome c]; SOR = sulfur oxygenase reductase; SorA = sulfite dehydrogenase (cytochrome) subunit A; SoxB = sulfur oxidizing protein; PhsA = thiosulfate reductase/polysulfide reductase chain A. Figure 2.13 is presented across three pages.




Unbinned
(Campylobacterota, Sulfurovum) BF dRep Bin 4
(Proteobacteria, Halothiobacillaceae) BF2 Bin 8
(Campylobacterota, UBA6211) BF dRep Bin 15
(Proteobacteria, Thiomicrospira) BF2 Bin 24
(Proteobacteria, Halothiobacillus) BF2 Bin 32
(Campylobacterota, Sulfurimonas) BF2 Bin 25
(Campylobacterota, JACXUD01) BF2 Bin 26
(Campylobacterota, Aliarcobacter) BF2 Bin 27
(Campylobacterota, Aliarcobacter) BF dRep Bin 18
(Campylobacterota, Sulfurimonas) BF2 Bin 36
(Campylobacterota, Aliarcobacter) BF2 Bin 9
(Thermotogota, Defluviitoga) BF dRep Bin 16
(Proteobacteria, Pseudomonas C) BF dRep Bin 9
(Planctomycetota, PWXP01) BF2 Bin 50
(Bacteroidota, UBA6192) BF dRep Bin 17
(Actinobacteriota, Corynebacterium) BF2 Bin 37
(Acidobacteriota, UBA876) BF dRep Bin 32
BF2
Unbinned
(Proteobacteria, Thiothrix) BF4 Bin 143
(Desulfobacterota, UBA10518) BF dRep Bin 26
(Desulfobacterota, Desulforhabdus) BF4 Bin 121
(Desulfobacterota, Desulfobacter) BF4 Bin 164
(Desulfobacterota, Desulfobulbus) BF dRep Bin 11
(Proteobacteria, Azonexus) BF4 Bin 165
(Desulfobacterota, Desulfofustis) BF4 Bin 157
(Desulfobacterota, Desulfobulbus) BF4 Bin 32
(Desulfobacterota, UBA2174) BF dRep Bin 27
(Desulfobacterota, Desulfobacterium B) BF dRep Bin 24
(Proteobacteria, UBA2357) BF4 Bin 80
(Proteobacteria, UBA2250) BF4 Bin 139
(Proteobacteria, Thauera) BF4 Bin 124
(Proteobacteria, Ectothiorhodospiraceae) BF4 Bin 83
(Desulfobacterota, Desulfococcus) BF4 Bin 38
(Desulfobacterota, SKZT02) BF4 Bin 50
(Campylobacterota, CAIJNA01) BF4 Bin 91
(Zixibacteria, WJMS01) BF4 Bin 19
(Proteobacteria, Methyloversatilis) BF4 Bin 125
(Proteobacteria, JACZKS01) BF4 Bin 22
(Proteobacteria, Azoarcus C) BF4 Bin 20
(Myxococcota, HGW−18) BF4 Bin 134
(Desulfobacterota, Desulfobulbus) BF4 Bin 34
(Campylobacterota, UBA6211) BF dRep Bin 36
(Campylobacterota, Sulfuricurvum) BF4 Bin 150
(Campylobacterota, JACXUD01) BF4 Bin 31
(Campylobacterota, Aliarcobacter) BF4 Bin 176
(Campylobacterota, Aliarcobacter) BF4 Bin 17
(Proteobacteria, Zoogloea) BF4 Bin 100
(Proteobacteria, Thauera) BF4 Bin 92
(Proteobacteria, Thauera) BF4 Bin 29
(Proteobacteria, Thauera) BF4 Bin 105
(Proteobacteria, Spongiibacteraceae) BF4 Bin 133
(Proteobacteria, JADJWH02) BF4 Bin 77
(Proteobacteria, Hydrogenophaga) BF4 Bin 132
(Proteobacteria, Chiayiivirga) BF4 Bin 81
(Proteobacteria, BRH−c0) BF4 Bin 12
(Proteobacteria, Azospira) BF4 Bin 98
(Nitrospirota, JAFGXL01) BF4 Bin 79
(Chloroflexota, UBA8950) BF4 Bin 147
(Chloroflexota, UBA7227) BF4 Bin 155
(Campylobacterota, Sulfurimonas) BF4 Bin 78
(Bacteroidota, LD21) BF4 Bin 71
(Verrucomicrobiota, UBA1068) BF4 Bin 25
(Verrucomicrobiota, DUPH01) BF4 Bin 67
(Proteobacteria, UBA6140) BF4 Bin 152
(Proteobacteria, UBA1573) BF4 Bin 112
(Proteobacteria, Spongiibacteraceae) BF4 Bin 184
(Proteobacteria, Rhodocyclaceae) BF4 Bin 14
(Proteobacteria, Pseudomonas D) BF4 Bin 85
(Proteobacteria, Phenylobacterium) BF4 Bin 149
(Proteobacteria, Giesbergeria) BF4 Bin 120
(Proteobacteria, Azonexus) BF4 Bin 86
(Proteobacteria, Arenimonas) BF4 Bin 130
(Proteobacteria, Aquabacterium) BF4 Bin 30
(Planctomycetota, UBA1925) BF4 Bin 154
BF4



Thiosulfate oxidation presence was based on SoxB which was detected in the proteomes of BF2 and BF4, but only found in the metagenome for BF1. In the metagenome, *soxB* was encoded in an average of 10 MAGs. While identified in the proteins of BF2 and BF4, SoxB was a very small proportion of the proteins with a NPA of 4 in BF2 and a NPA of 72 in BF4. Thiosulfate disproportionation to sulfite and hydrogen sulfide by thiosulfate reductase (PhsA) was only detected in BF1's metaproteome, with a very low NPA of 3, but was detected in all three metagenomes, in an average of 15 MAGs. Sulfur dioxygenase (Sdo) and sulfur reduction via sulfur reductase (Sre) were not detected in any metagenomes or metaproteomes.

Sulfate reduction was a key process occurring in the biofilms, detected in both the unbinned proteins, and in abundant MAGs. The complete sulfate reduction pathway was expressed and associated with MAGs in BF1 and BF4. In BF1, taxonomic classification of the MAGs included

phylum Desulfobacterota, and *Desulfobulbus* and *Desulfobotulus* genera. In BF4, the only MAG expressing proteins for the entire sulfate reduction pathway was BF4 Bin 164, classified to genus *Desulfobacter*. Sulfate reducing bacteria may reside in this biofilm due sulfur-containing waste being deposited within the landfill site, feeding into the biofilm ecosystem (Sun *et al.* 2016).

2.3.2.3 Methane cycling

Methanogenesis was another process that was highly active based on detection within the metaproteome (Figure 2.20; Figure 2.21). Methanogenesis is the biological production of methane, that occurs in a group of strictly anaerobic archaea called methanogens where the reaction occurs via methyl-coenzyme M reductase (Mcr). The alpha subunit of Mcr (McrA) is often used as a marker for the function of methanogenesis, and McrA accounted for 11,771, 25,795, and 8,180 of the NPA in BF1, BF2, and BF4. Within the metagenome, *mcrA* was encoded in 9, 2, and 2 MAGs, but 57 additional genes were associated with the unbinned sequences in BF1, 13 unbinned genes in BF2, and 3 unbinned genes in BF4. The results show that in BF2, the gamma subunit of Mcr (McrG) accounted for a NPA of 151,005, compared to 45,775 and 6,419 in BF1 and BF4, respectively. The beta subunit accounted for a NPA of 11,688 in BF1, 27,060 in BF2, and 2,295 in BF4. In each biofilm, proteins were found for the entire Mcr complex that associated with MAGs, with BF1 containing four MAGs with the full complex expressed. BF dRep Bin 10 was the only MAG with proteins expressed for Mcr subunits in all biofilms, though only McrB was detected in BF4. BF dRep Bin 10 belongs to phylum Halobacteriota and genus *Methanothrix*. In BF4, phylum Euryarchaeota genus *Methanothermobacter* from BF4 Bin 74 expressed the Mcr complex and was the fourth most abundant MAG in the BF4 metaproteome. Halobacteriota (BF1 Bin 64) from the genus *Methanosarcina* expressed McrA and McrC and was the second most abundant MAG in the BF1 metagenome. *Methanosarcina* was also one of the main methanogens

detected at a landfill site in China (Tang *et al.* 2016), and *Methanothermobacter* belongs to the order Methanobacteriales, common at landfill sites (Song *et al.* 2015, Sauk and Hug 2022). Despite the high methane concentrations and anoxic conditions in the gas collection system, there was no evidence for anaerobic methanotrophy (ANME), including no identified MAGs associated with ANME lineages (Guerrero-Cruz *et al.* 2021).

For aerobic methanotrophy, two enzymes can facilitate this activity: particulate methane monooxygenase (Pmo) and soluble methane monooxygenase (Mmo). Given the largely anoxic nature of the gas collection system and leachate reservoirs within the landfill, aerobic methanotrophy would not be expected to be a large factor in the biofilm community metabolism.

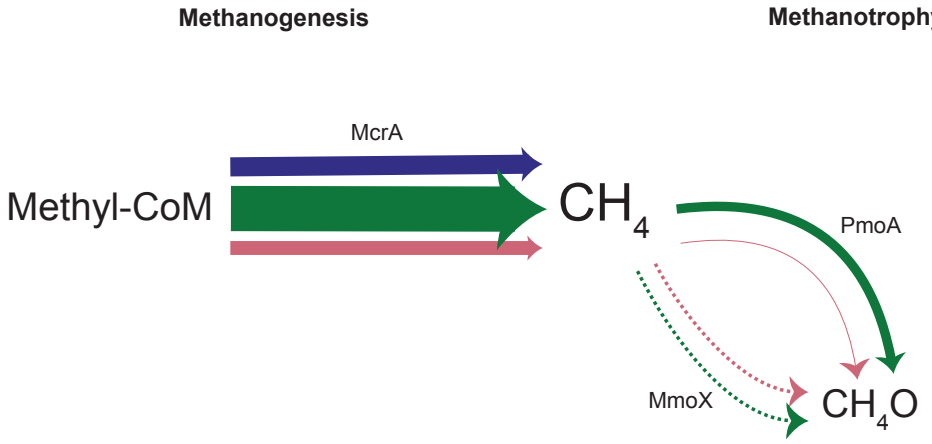


Figure 2.20 Predicted activity for methane metabolism in biofilms. Arrows were included if the important enzymes were detected in the metagenome or metaproteome, where matches could be associated with specific metagenome-assembled genomes (MAGs) or unbinned. Arrows are colored by biofilm sample (BF1 = blue; BF2 = green; BF4 = pink). Solid arrows indicate presence in the metaproteome and are weighted to the sum of normalized protein abundance (NPA) within a biofilm sample. Dotted arrows are not weighted and indicate detection in the metagenome only. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1. Protein abbreviations are: PmoA = methane/ammonia monooxygenase subunit A; MmoX = methane monooxygenase component A alpha chain.

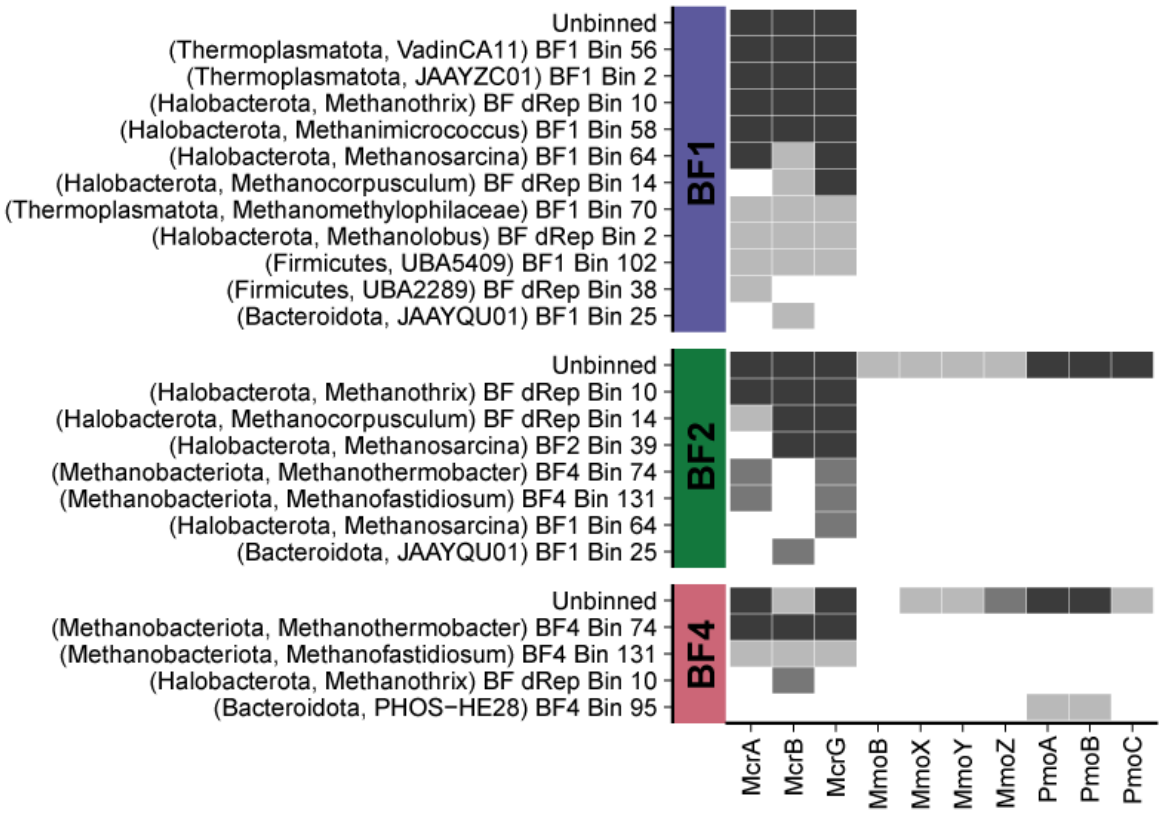


Figure 2.21 Presence or absence of enzymes involved in methane metabolism corresponding to MAGs or unbinned in biofilm metagenomes and metaproteomes. Detection within MAGs and unbinned data are grouped and colored by biofilm samples (BF1 = blue; BF2 = green; BF4 = pink). MAG labels include taxonomy in brackets (Phylum, Genus) and the MAG ID. Light grey boxes indicate detection in the metagenome only, medium grey boxes indicate detection in metaproteome only, and dark grey indicates detection in both metagenome and metaproteome. Activity predictions were determined through detection of key marker genes for metabolic processes and are summarized in Table 2.1. Protein abbreviations are: McrA = methyl-coenzyme M reductase alpha subunit; McrB = methyl-coenzyme M reductase beta subunit; McrG = methyl-coenzyme M reductase gamma subunit; MmoB = methane monooxygenase regulatory protein B; MmoX = methane monooxygenase component A alpha chain; MmoY = methane monooxygenase component A beta chain; MmoZ = methane monooxygenase component A gamma chain; PmoA = methane/ammonia monooxygenase subunit A; PmoB = methane/ammonia monooxygenase subunit B; PmoC = methane/ammonia monooxygenase subunit C.

While individual MAGs did not encode *pmo* and *mmo* genes, there were expressed Pmo and Mmo proteins in the metaproteome that corresponded to *pmo* and *mmo* genes from the unbinned fraction of the metagenomes in BF2 and BF4. MmoX was expressed in BF2 with an NPA of 5,435 and in BF4 with a NPA of 543. PmoA was only expressed in BF4 with a NPA of 74.

Overall, methanogenesis and specifically the methyl-coenzyme M reductase protein were abundantly detected in each biofilm. Despite high concentrations of methane within the landfill environment, methanotrophy was a function that was limited in the biofilms. This is consistent with methanogens also being detected in methane seeps, environments with high methane concentrations (Tarnovetskii *et al.* 2019), where methanotrophy and methanogenesis processes have the potential to be co-occurring in close proximity (Inagaki *et al.* 2004).

2.3.2.4 Other metabolisms and metabolites

To evaluate the distribution of enzymatic defenses against reactive oxygen species (ROS), a set of previously identified proteins (see Chapter 4) was evaluated in the biofilm metagenomic and metaproteomic datasets. ROS defenses were encoded in MAGs with sparse detection in the metaproteomic dataset (Figure 2.22; Figure 2.23). All major ROS functions were encoded in some capacity in each biofilm sample (Figure 2.22). The defenses most often encoded with MAGs include thioredoxin reductase, associated with an average of 111 MAGs, thioredoxin 1 with 175 MAGs, and alkyl hydroperoxide reductase with 95 MAGs. Thioredoxin reductase was not detected in the proteins, and thioredoxin was detected in biofilms with an average NPA of 1,191, and alkyl hydroperoxide reductase proteins were found with an average NPA of 3,434. Proteins were detected, but at low abundance with no ROS function occurring at more than 1% of total NPA in any biofilm sample, suggesting a largely anoxic environment within the biofilms.

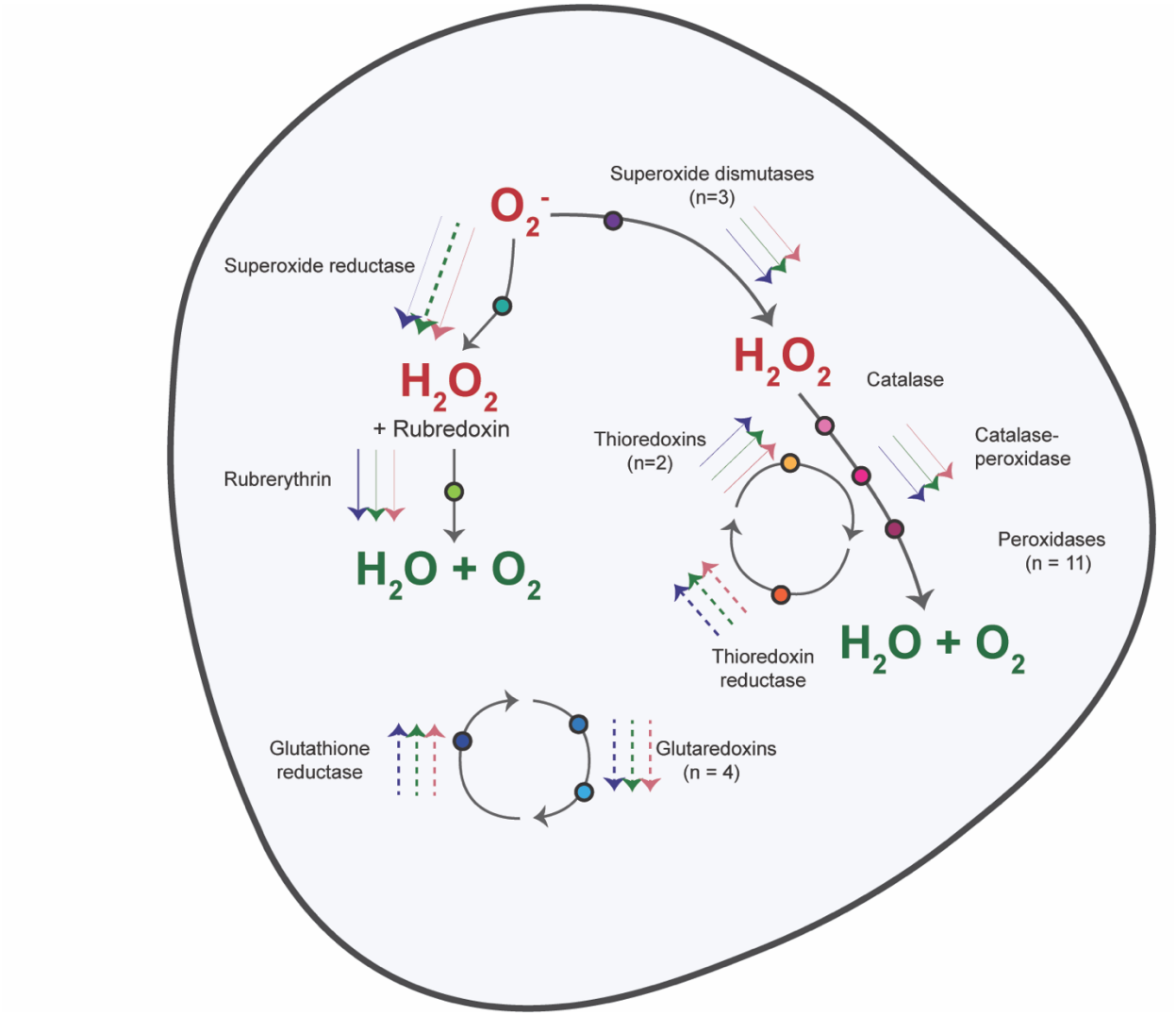
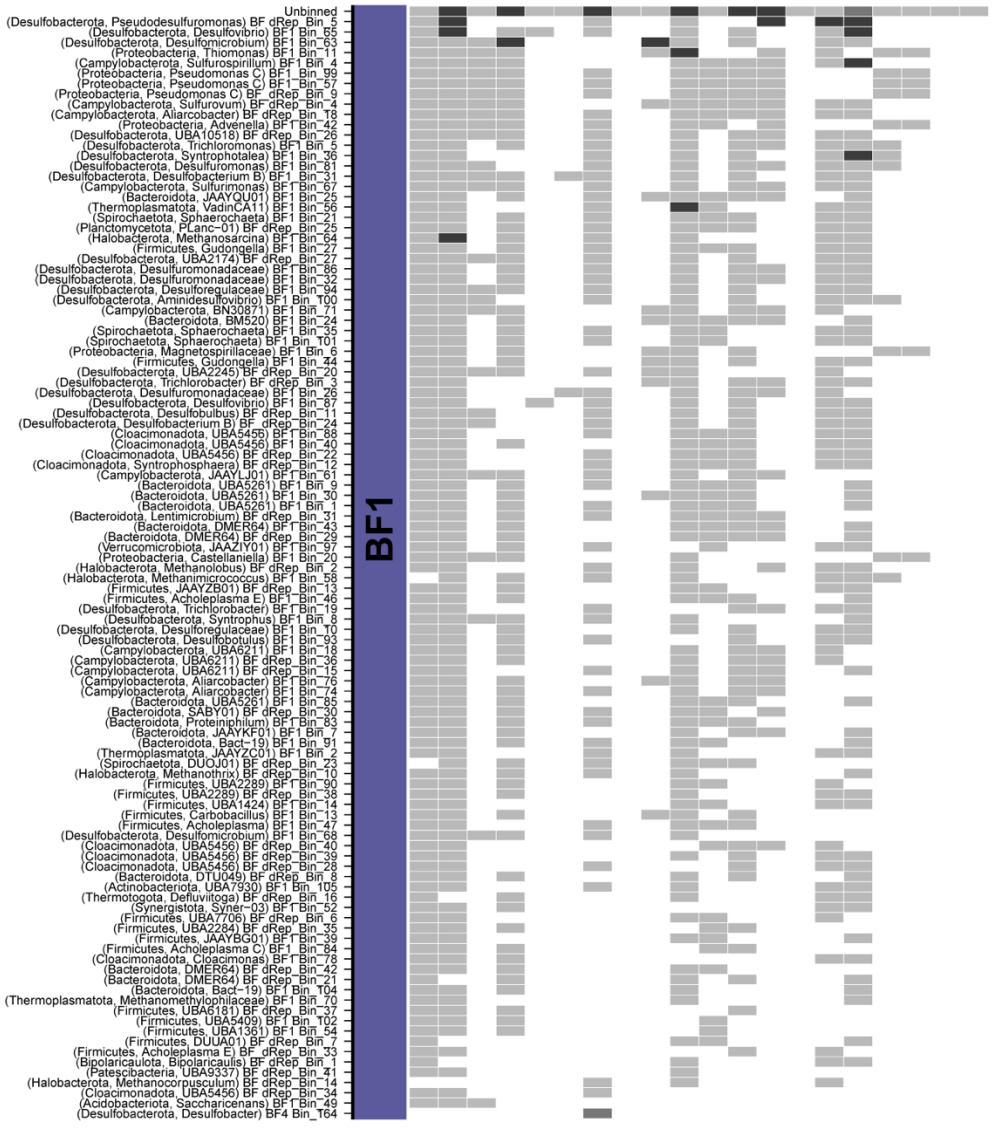
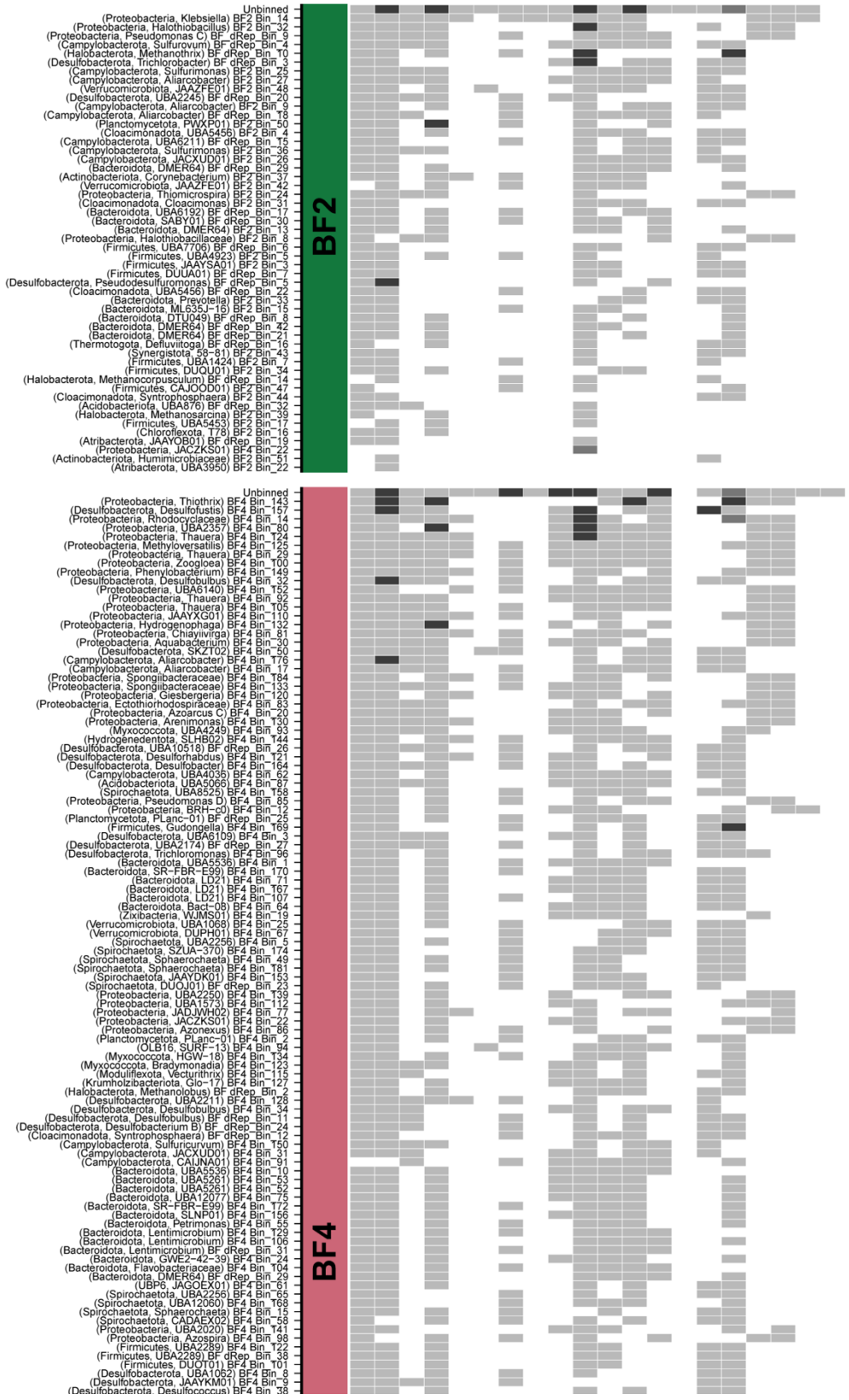
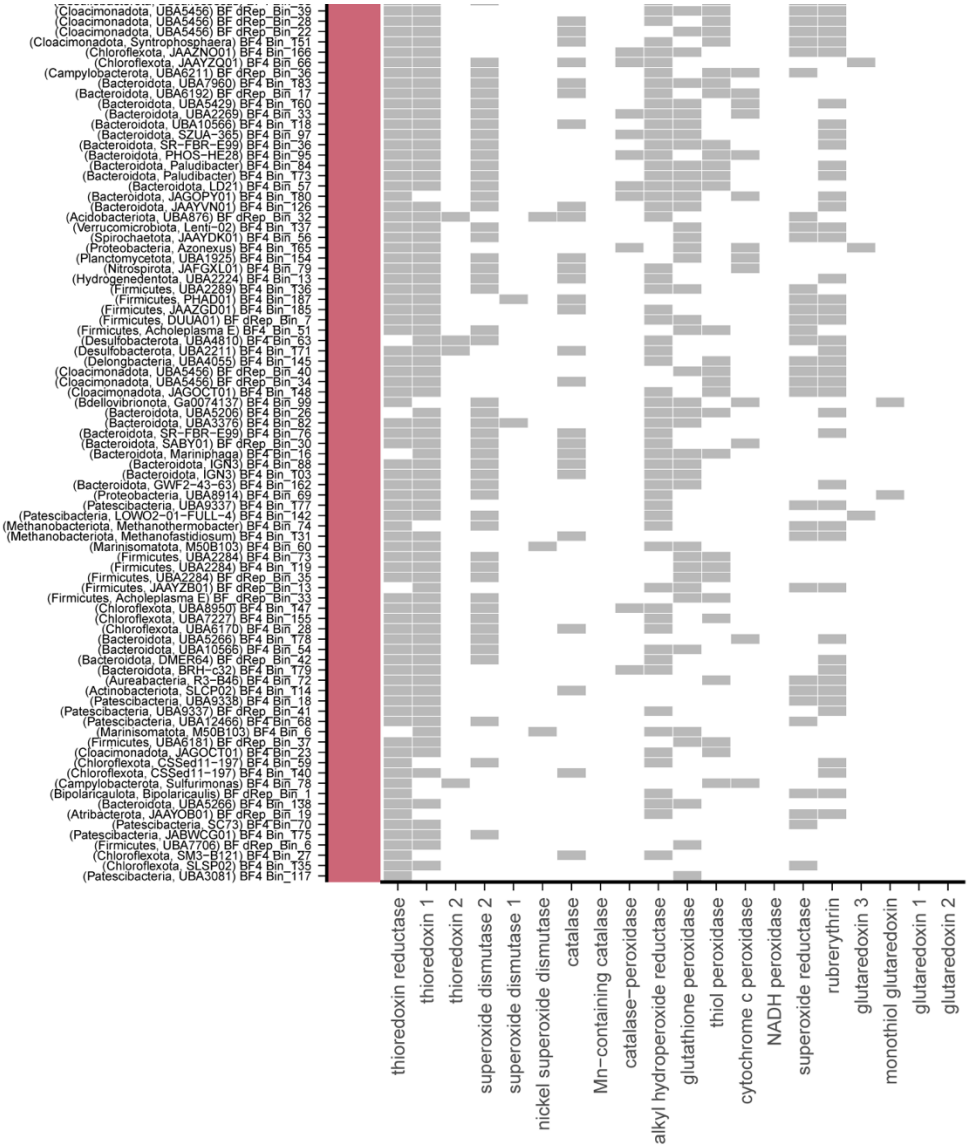


Figure 2.22 Predicted activity for enzymatic ROS defense mechanisms. Arrows were included if the important enzymes were detected in the metagenome or metaproteome, where matches could be associated with specific metagenome-assembled genomes (MAGs) or unbinned. Arrows are colored by biofilm sample (BF1 = blue; BF2 = green; BF4 = pink). Solid arrows indicate presence in the metaproteome and are weighted to the sum of normalized protein abundance (NPA) within a biofilm sample. Dotted arrows are not weighted and indicate detection in the metagenome only. Activity was determined through detection of key marker genes in enzymatic ROS defense, as previously identified (Chapter 4).

Figure 2.23 Presence or absence of important enzymes involved in defense against reactive oxygen species (ROS) corresponding to MAGs or unbinned in biofilm metagenomes and metaproteomes. Detection within MAGs and unbinned data are grouped and colored by biofilm samples (BF1 = blue; BF2 = green; BF4 = pink). MAG labels include taxonomy in brackets (Phylum, Genus) and the MAG ID. Light grey boxes indicate detection in the metagenome only, medium grey boxes indicate detection in metaproteome only, and dark grey indicates detection in both metagenome and metaproteome. Activity was determined through detection of key marker genes in enzymatic ROS defense, as previously identified (Chapter 4). Figure 2.17 is presented across three pages.




BF2
BF4



Metametabolomic datasets were also evaluated for four biofilm samples, and identifiable metabolites were visualized on a heatmap (Figure 2.24). The three most abundant metabolites across all four biofilms were bisphenol A, carbonate, and palmitic acid. Bisphenol A is a hazardous component of plastic, that has been demonstrated to alter soil microbiomes and disrupt soil

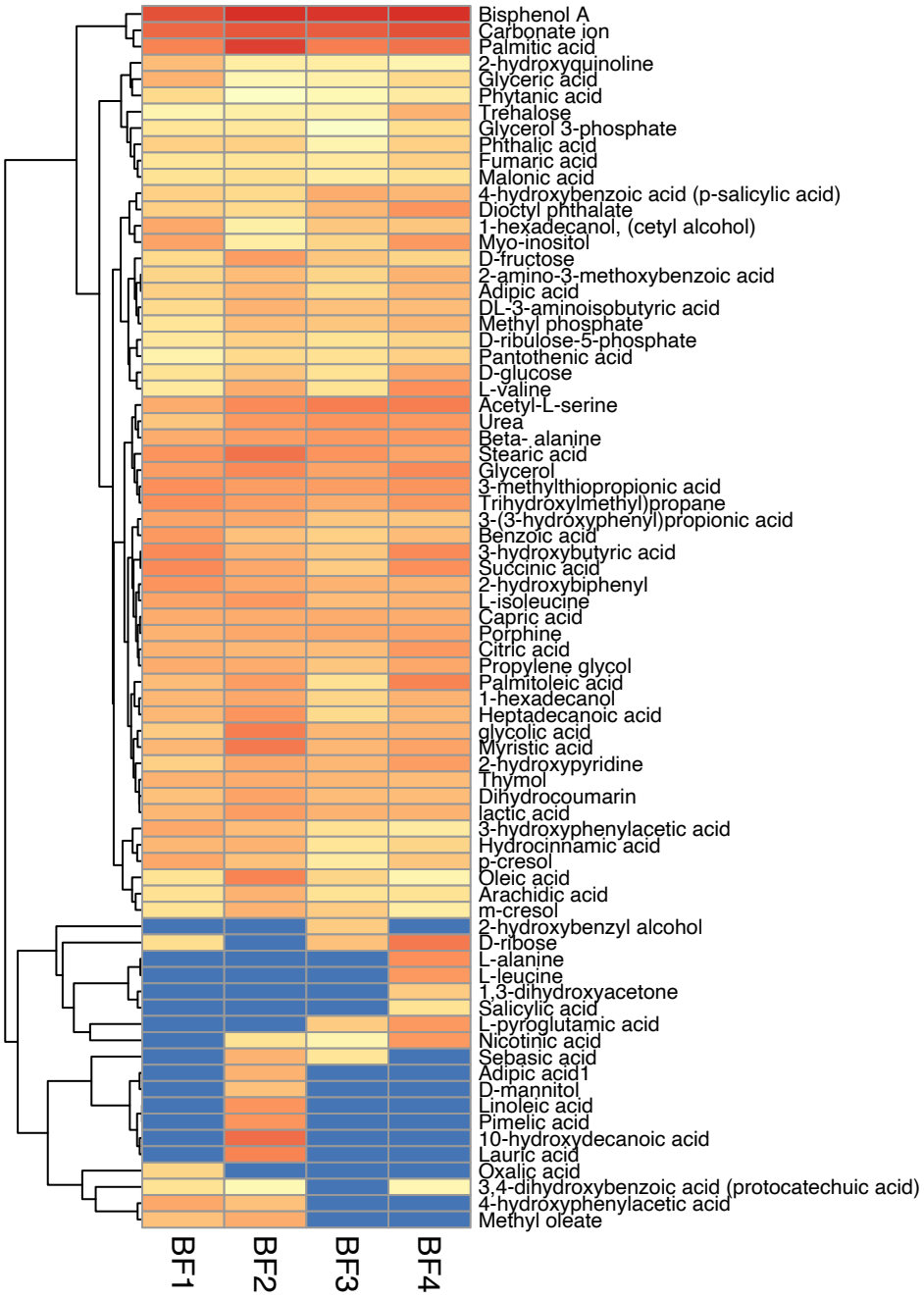


Figure 2.24 Heatmap depicting detection of identified metabolites. Abundance values were log10 transformed, and hierarchical clustering and visualization were completed with the "pheatmap" package in R.

homeostasis (Zaborowska *et al.* 2020). Palmitic acid is a saturated fatty acid, and it may play a role in cell survival within a biofilm, as it has been demonstrated that some microorganisms in biofilms have higher levels of saturated fatty acids in their membranes compared to their planktonic counterparts (Dubois-Brissonnet *et al.* 2016). Overall, a major limitation of untargeted metametabolomics is driven by lack of metabolite identification (Chaleckis *et al.* 2019). Metabolite identification is a bottleneck due to the complexity of LC-MS data, along with the often-needed manual refinement of annotations. In the metametabolomic dataset, 79 known metabolites were identified, along with 309 unknown metabolites – highlighting the lack of metabolite identification in this system (Figure 2.24; Appendix A, Supplemental Figure A2).

2.4 Conclusions

The microbial communities within landfill gas collection system biofilms are complex, heterogeneous, and uneven. Communities were dominated by several lineages within amplicon and metagenomic datasets, with other lineages dominating the metaproteomic data. Despite taxonomic differences between biofilm samples, key enzymes involved in dissimilatory sulfate reduction and methanogenesis are major components of the expressed proteins in the system. These processes are metabolic features that occur in anoxic conditions by organisms that are anaerobic. Due to the limited capacity for defense against reactive oxygen species in these biofilm environments, along with primarily anaerobic metabolic processes being dominant, oxygen may be detrimental or disruptive to biofilm formation, and may be a direction for landfills to reduce impacts of biofouling.

Chapter 3

Cloacimonadota metabolisms include adaptations for engineered environments that are reflected in the evolutionary history of the phylum

A version of this chapter has been published as:

> Johnson, L.A. and Hug, L.A. (2022) Cloacimonadota metabolisms include adaptations in engineered environments that are reflected in the evolutionary history of the phylum. *Environmental Microbiology Reports*. 14: 520–529. https://doi.org/10.1111/1758-2229.13061

The published version is a short report with an extensive supplemental materials component. The below version integrates much of the supplemental materials into the main thesis chapter, matching the original manuscript format prior to short-form condensing.

3.1 Introduction

Cloacimonadota is an enigmatic bacterial phylum, and therefore, in the literature this lineage has a limited description of broad metabolic capacities. The lack of metabolic information, along with detection at landfill sites, led to the selection of phylum Cloacimonadota for further investigation in this thesis chapter with metagenomics being used to examine metabolic diversity and adaptation to engineered environments. Frequently observed in engineered systems like wastewater and sludge bioreactors, their metabolic diversity and adaptations to these engineered environments have not been examined. The first record of phylum Cloacimonadota was as unidentified 16S rRNA gene sequences from clone libraries derived from two unrelated samples: an anoxic flooded rice microcosm (Derakshani *et al.*, 2001), and a reductive dechlorinating enrichment culture inoculated with material from a methanogenic reactor (Gu *et al.*, 2004). Originally associated with the phylum Spirochaetota due to limited 16S rRNA gene reference

sequences, Chouari *et al.*, (2005) expanded the clade with sequences recovered from an anaerobic sludge digester to describe a well-supported monophyletic lineage distinct from Spirochaetota and other bacterial phyla (Chouari *et al.*, 2005). The group was initially named candidate division WWE1, for the wastewater treatment plant in Evry, France from which the original 16S rRNA gene sequences were recovered, and as the first of two novel groups described in their study. The first, and only, complete genome for the phylum was sequenced from an anaerobic wastewater digestor (Pelletier *et al.*, 2008). Provisionally classified as deriving from "*Candidatus* Cloacimonas acidaminovorans", the genome was reconstructed using an iterative assembly process of genome extension from a metagenomic dataset (Pelletier *et al.*, 2008). The genus name *Ca*. Cloacimonas derives from Latin terms translating to "a unit from the sewer", with the species acidaminovorans translating to "amino acid eating". In 2013, the phylum was reclassified from WWE1 to Cloacimonetes following expansion of the group with single-cell genomes from a brackish lake, deep sediment in a lagoon, and terephthalate degrading bioreactors (Rinke *et al.*, 2013). Cloacimonetes were a focus due to the phylum's relative underrepresentation on the tree of life and the absence of a cultivated representative (Rinke *et al.*, 2013), characteristics which hold true for the phylum today. Cloacimonetes were recently renamed to Cloacimonadota in the Genome Taxonomy Database (GTDB; http://gtdb.ecogenomic.org/) (Parks *et al.*, 2018, 2020), a genome-based system attempting to standardized microbial taxonomy. The GTDB database currently contains 58 representative genomes for phylum Cloacimonadota, representing 23 species-level groups (Mendler *et al.*, 2019). This level of available genomic diversity puts Cloacimonadota in the middle (48th percentile) of the named bacterial phyla, further highlighting their comparative absence from the literature to date.

Cloacimonadota have frequently been identified from wastewater treatment plants (Westerholm *et al.*, 2016; Calusinska *et al.*, 2018; Jankowska *et al.*, 2018; Theuerl *et al.*, 2018; Shakeri Yekta *et al.*, 2019) and biogas plants (Solli *et al.*, 2014; Ahlert *et al.*, 2016), often via 16S rRNA gene amplicon sequencing. Anaerobic digesters hosting Cloacimonadota span diverse wastewater substrates, including sugar beet molasses (Chojnacka *et al.*, 2015), coffee processing (Botello Suárez *et al.*, 2018), starch processing (Antwi *et al.*, 2017; Qin *et al.*, 2018), lipidic waste (Saha *et al.*, 2019), cow manure/straw (Sun *et al.*, 2015; Dong *et al.*, 2019), municipal solid waste (Cardinali-Rezende *et al.*, 2012; Akyol *et al.*, 2019), sewage (Braz *et al.*, 2019), and food waste (Moestedt *et al.*, 2020). Although Cloacimonadota are predominantly found in wastewater environments, they have also been found in the coal-bearing strata of the Cherokee basin (Kirk *et al.*, 2015), an oilfield enrichment consortium (Toth and Gieg, 2018), and bodies of water including the meromictic Sakinaw Lake (Gies *et al.*, 2014), a cold seep brine pool of the Red Sea (Zhang *et al.*, 2016), the sulfidic zone in the Black Sea (Suominen *et al.*, 2021), and the euxinic waters of Ursu Lake (Baricz *et al.*, 2021).

The current understanding of Cloacimonadota metabolisms, lifestyles, and ecosystem roles largely derives from the "*Candidatus* Cloacimonas acidaminovorans" genome, where metabolism predictions indicate *Ca.* C. acidaminovorans is likely a syntrophic bacterium using fermentation of amino acids for carbon and energy (Pelletier *et al.*, 2008). Another important study highlighting Cloacimonadota metabolism used enrichment cultures that included a member of Cloacimonadota, *Ca.* Syntrophosphaera thermopropionivorans (Dyksma and Gallert, 2019). Enrichment cultures were able to produce methane from propionate, and quantitative polymerase chain reaction (qPCR) and 16S rRNA gene sequencing analyses validated the presence of *Ca.* S. thermopropionivorans

from enrichments. Further, a draft genome of *Ca.* S. thermopropionivorans encoded several genes involved in propionate oxidation via methylmalonyl-CoA (Dyksma and Gallert, 2019). Other Cloacimonadota metagenome-assembled genomes (MAGs) derived from metagenomes of thermophilic biogas plants encoded the capacity to ferment amino acids with the production of acetate, H_2, and carbon dioxide (CO_2) (Stolze *et al.*, 2016), activities that were supported by metatranscriptomics for one of the Cloacimonadota MAGs (Stolze *et al.*, 2018). Three Cloacimonadota MAGs from deep terrestrial subsurface sediments contain [FeFe] hydrogenases, suggesting members of this group have potential roles in H_2 cycling (Hernsdorf *et al.*, 2017). A recent examination of seven novel MAGs from Cloacimonadota identified the capacity for carbon fixation using the reverse tricarboxylic acid cycle (Williams *et al.*, 2021). Other studies indicated that members of the Cloacimonadota have cellulolytic potential (Limam *et al.*, 2014; Xia *et al.*, 2018), and quantitative DNA stable isotope probe incubations along with a MAG recovered from the sulfidic zone of the Black sea highlighted that one Cloacimonadota member is a generalist capable of using different organic matter substrates (Suominen *et al.*, 2021; Williams *et al.*, 2021).

Consistent with strong representation in engineered waste environments, Cloacimonadota have also been identified at municipal landfills. Cloacimonadota have been observed at mature landfill sites where microbial communities are involved in humification (Liu *et al.*, 2011; Remmas *et al.*, 2017), and as part of landfill leachate microcosms enriched for lignocellulose degradation (Ransom-Jones *et al.*, 2017). However, several investigations of bacterial diversity at municipal landfills did not detect Cloacimonadota (Song 2015b; Stamps *et al.*, 2016), which could be due to the heterogeneous nature of municipal solid waste (MSW) at landfills. The drivers that influence the presence or distributions of Cloacimonadota within municipal landfills are not known.

The study site is an active municipal landfill in Southern Ontario. The landfill is instrumented with several hundred wells for monitoring leachate chemical composition. Leachate is the liquid component of landfill waste, derived from microbial degradation of organic wastes, rainwater inflow, and groundwater infiltration, where leachate typically contains many organic and inorganic compounds, with the composition varying with the make-up of deposited waste (Renou *et al.*, 2008). These compounds can have adverse impacts on surrounding environments if leachate is not properly contained or treated. Alongside the formation of leachate, microbially mediated decomposition of MSW results in gas formation, including the greenhouse gases CO_2 and methane (CH_4) (Hilger and Barlaz, 2006). To mediate the negative impact of gas emissions, many modern landfills, including the study site, are instrumented with gas capture systems that redirect emitted methane to biogas generation plants, subsequently converting methane to electricity (Spokas *et al.*, 2006). At the Southern Ontario landfill, the gas capture system experiences periodic fouling by overgrowth of a thick biofilm (personal communication, landfill engineers). Biofouling blocks the methane-capture infrastructure, consequently reducing methane gas capture and sustainable energy production. The infrastructure at the landfill site uses leachate wells to monitor leachate, composite leachate cisterns to collect leachate formed across the site prior to transport to a wastewater treatment facility, and a gas capture system to collect methane and other gases for bioenergy conversion. Diverse and abundant Cloacimonadota have been detected across the site in leachate wells (LW), a composite leachate cistern (CLC), and a biofilm (BF) within the gas capture system (Sauk and Hug, 2022).

Here the genomic and metabolic diversity of the understudied phylum Cloacimonadota is expanded using MAGs derived from landfill metagenomes. I examined the metabolic capacities

and roles for this lineage within engineered environments. In a pangenome comparison of all currently available Cloacimonadota genomes, to assess traits specific to microorganisms found in anthropogenically impacted or pristine environments. The results delineate the ecosystem services the members of this phylum contribute to an environment and clarify why these microorganisms are so frequently prevalent in bioreactors and related anoxic engineered environments.

3.2 Materials and methods

3.2.1 Sampling, DNA sequencing, assembly, and binning

Samples were collected from an active municipal landfill site in Southern Ontario. Samples included landfill leachate wells (n = 5), groundwater wells (n = 1), and a biofilm (n = 1) collected over three separate sampling events. In the first sampling in July 2016, leachate biomass was collected from a composite leachate cistern (CLC_T1) by filtration of the leachate through a 0.2 µm and a 0.1 µm polyethersulfone filter in series. One week later, leachate biomass was collected from the same composite leachate cistern (CLC_T2) along with three leachate wells (LW1, LW2, and LW3), and one groundwater well (GW1) by filtration through a 3 µm glass fiber filter followed by a 0.1 µm polyethersulfone filter in series. This change in filter pore sizes was made to reduce filter clogging. Glass fiber pre-filters were discarded, and 0.1 µm filters were frozen on dry ice and transported to the lab where they were stored at -80ºC until processed. A biofilm sample (BF1) was collected from overgrowth in the landfill's gas capture system in October 2017. The sample was extricated from a clogged valve using sterile scalpels, as it resisted any other form of manipulation.

DNA was extracted from the landfill biomass on filters and the gas intake biofilm sample using a DNeasy PowerSoil Pro DNA isolation kit (QIAGEN) following the manufacturer's

instructions, with the modification of slicing the filters into pieces and adding the filter directly to the extraction in place of soil, or with the addition of 10 g of biofilm in place of soil. DNA was shotgun sequenced on an HiSeq platform (Illumina) with 150 bp paired-end reads (U.S. Department of Energy Joint Genome Institute).

For all samples collected in 2016, the metagenomic sequences were assembled with IDBA_UD (Peng *et al.*, 2012), and binned with CONCOCT (Alneberg *et al.*, 2014) using tetranucleotide frequencies and cross-sample abundance patterns based on read mapping. Bins were manually refined on the anvi'o analysis platform (Eren *et al.*, 2015). The biofilm metagenome from 2017 was assembled with metaSPAdes (Nurk *et al.*, 2017), and binned using DasTool (Sieber *et al.*, 2018) taking input from the binning algorithms CONCOCT, MaxBin2 (Wu *et al.*, 2016), and MetaBAT2 (Kang *et al.*, 2019). All MAGs were required to have thresholds of completion above 70% and contamination below 10% to be considered for metabolic analyses. MAGs were identified as belonging to the members of the phylum Cloacimonadota based on placement within a concatenated ribosomal protein tree that contained representatives across the tree of life (Sauk and Hug, 2022). The average coverage was calculated for each MAG based on read mapping of the input reads to the relevant metagenome using Bowtie2 version 2.3.5.1 (Langmead and Salzberg, 2012).

3.2.2 Mining public databases for Cloacimonadota genomes

All publicly available genomes for members of the phylum Cloacimonadota were downloaded from the Joint Genome Institute's Integrated Microbial Genomes and Microbiomes (http://img.jgi.doe.gov) or the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) (July 2019). Three genomes for members of the

Gemmatimonadota, five genomes for members of the Fibrobacterota and six genomes for members of the Zixibacteria were also downloaded as phylogenetically related yet distinct outgroups for phylogenetic trees. Genome quality was evaluated with CheckM (Parks *et al.*, 2015) for all collected genomes, and genomes that did not meet the greater than 70% completion and less than 10% contamination criteria were not included in metabolic analyses. A total of 30 reference Cloacimonadota of sufficient quality were included in metabolic analyses.

3.2.3 Phylogenetic tree

Landfill Cloacimonadota MAGs and all reference genomes were placed on a phylogenetic tree. GToTree was used with the bacterial HMM single copy gene set of 74 genes (Lee, 2019). Briefly, this workflow identifies target genes in each genome and uses genome completeness and contamination values to filter gene hits and genomes. Sequences are aligned and concatenated into a full alignment, and the phylogeny was inferred with FastTree with default settings. The phylogenetic tree was visualized using iTOL (http://itol.embl.de) (Letunic and Bork, 2016).

3.2.4 Metabolic analyses: KEGG, DRAM, CAZy, and MEROPS

Cloacimonadota genomes, both reference genomes from public databases and the landfill MAGs, meeting the thresholds of completion above 70% and contamination below 10% were included in a survey of phylum Cloacimonadota's metabolic potential (n = 46). Genomes were evaluated using the KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/kegg/kaas/) (Moriya *et al.*, 2007). KAAS was used with bi-directional best hit information based on sequence similarities to provide KEGG orthology identifiers for genes within each genome. Enzyme Commission (EC) numbers or KO numbers were used to cross reference against metabolic functions of interest to ascertain presence or absence.

Distilled and Refined Annotation of Metabolism (DRAM) was used under default parameters and without access to the KEGG database, for annotation of the 46 Cloacimonadota genomes in order to assess specific metabolic predictions in greater detail (e.g., amino acid degradation pathways) (Shaffer *et al.*, 2020).

To search for carbohydrate active enzymes, Cloacimonadota genomes were compared to the CAZy database (Lombard *et al.*, 2014) using dbCAN2 (version July 31, 2019) (Zhang *et al.*, 2018). Annotations were considered if the coverage was greater than 0.5 and e-value was below $1e^{-20}$ for a hmmer hit. The MEROPS database (https://www.ebi.ac.uk/merops/; merops_scan.lib) was used to search for peptidases and peptidase inhibitors (Rawlings *et al.*, 2018). The database was searched with BLASTp version 2.8.1+ and hits were considered if the e-value was below $1e^{-10}$. Beyond these two databases, manual BLASTp searches were conducted to search for superoxide reductase (WP_133200558.1), rubrerythrin (WP_133200557.1) and a pilin protein (WP_133201237.1) from *Candidatus* Syntrophosphaera thermopropionivorans to supplement annotations (Dyksma and Gallert, 2019).

3.2.5 Pangenome analysis

The pangenome analysis tool from the anvi'o package was used to examine core genes across the phylum, as well as across Cloacimonadota groups associated with engineered and natural environments (Eren *et al.*, 2015; Delmont and Eren, 2018). In anvi'o version 6.2, the 46 Cloacimonadota genomes were used to generate anvi'o genomes storage databases, where open reading frames were identified with Prodigal (Hyatt *et al.*, 2010), and genes were annotated using the Clusters of Orthologous Groups (COGs) database (Tatusov, 2000). Next, the "anvi-pangenome" program was used to create the pangenome with NCBI BLASTP for searching

(Altschul *et al.*, 1990), a minbit parameter of 0.5, a MCL inflation parameter of 2 (for family-level or higher divergence), and a gene cluster min occurrence parameter of 3 (excluding singleton and doubleton clusters). This step identifies "gene clusters" from the similarity searches of each amino acid sequences in each genome to every other amino acid sequence, and further the occurrence of these gene clusters across genomes. Hierarchical clustering analyses for both gene clusters and genomes used Euclidean distance and Ward clustering. In the pangenome visualization, genomes were organized based on the frequency of the gene clusters that they share. Heatmaps were created from the assigned COG categories for gene clusters for each gene cluster in the pangenome, and were visualized using ggplot2 (Ginestet, 2011).

3.3 Results and discussion

3.3.1 Characteristics of Cloacimonadota genomes

The genome quality of all landfill-derived and publicly available Cloacimonadota genomes was evaluated using CheckM (Appendix B, Supplementary Table B1) (Parks *et al.*, 2015). Of the 29 Cloacimonadota MAGS from the landfill datasets, 24 were of sufficient quality (>70% completion, <10% contamination). From public databases, 30 Cloacimonadota genomes and MAGS were obtained, with 22 genomes being used in metabolic analyses. The highest quality genome is from the sole sequenced isolated, "*Candidatus* Cloacimonas acidaminovorans" strain Evry, with a completion of 100% and predicted contamination of 1.1% (Pelletier *et al.*, 2008). Within the set of Cloacimonadota genomes kept for further analyses, the number of scaffolds ranges from 1 to 443, with a median of 135 (Appendix B, Supplementary Table B1). The average genome size was 2.4 Mbp, with an average of 2,028 genes per genome. Generally, the number of predicted genes increases linearly with genome size, irrespective of the originally sampled

environment (Figure 3.1). The largest genome belongs to *Candidatus* Cloacimonetes bacterium NORP72 from marine subsurface, with a size of 4.6 Mbp (Tully *et al.*, 2018), and the smallest genome belongs to *Candidatus* Cloacimonetes bin NIOZ-UU4 from an epipelagic marine environment, with a size of 1.4 Mbp (Villanueva *et al.*, 2021).

3.3.2 Environmental distribution of Cloacimonadota

Cloacimonadota were detected at different locations with the waste management system in Southern Ontario (Appendix B, Supplementary Table B1). At the landfill site, Cloacimonadota MAGs were represented in metagenomes from three leachate wells (LW1, n = 2; LW2, n = 8;

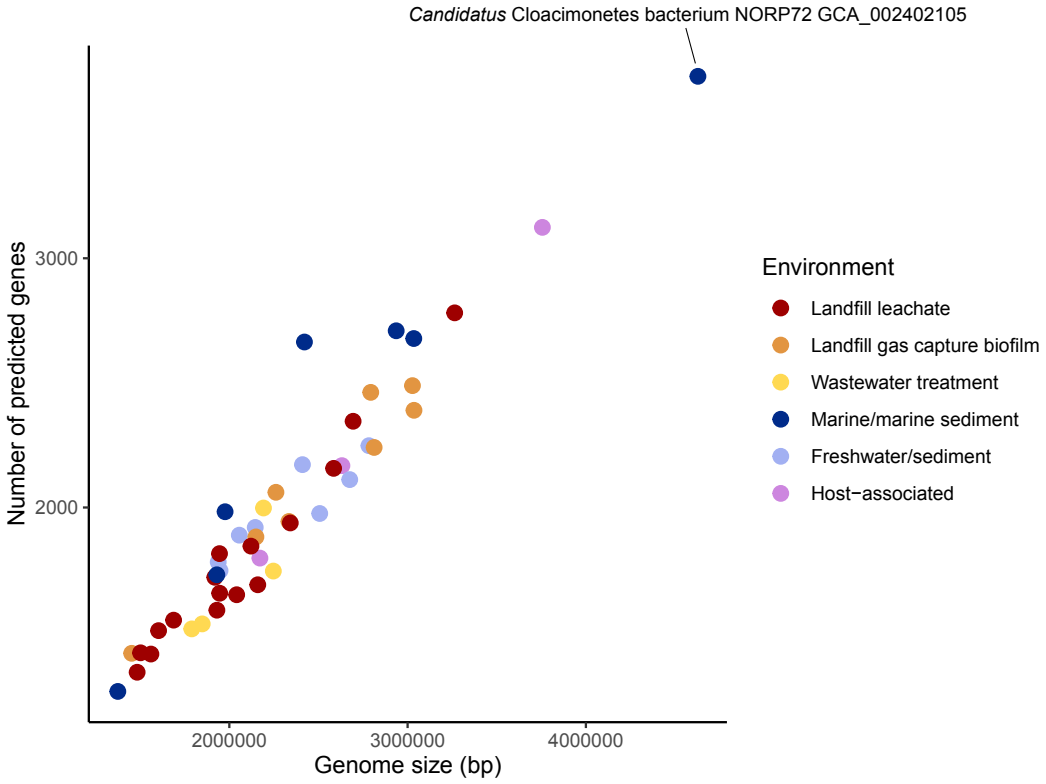


Figure 3.1 Number of predicted genes compared to genome size for Cloacimonadota genomes of sufficient quality. Each Cloacimonadota genome (n = 46) is represented by a dot, colored by the environment of origin.

LW3, n = 1), the composite leachate cistern at both time points (CLC_T1, n = 3; CLC_T2, n = 7), and the biofilm sample (BF1, n = 8) (Appendix B, Supplementary Table B1). Cloacimonadota were not represented in the MAGs from the groundwater well metagenome at the landfill site. The 24 MAGs passing the CheckM quality metrics were analyzed further. The proportion of reads from the total assembled dataset attributed to each MAG provides an estimate of relative abundance across landfill sites. Across the landfill sites, Cloacimonadota MAGs accounted for 0.06% to 4.72% of the assembled metagenome, with an average of 0.4%, making them numerous but not relatively abundant members of the landfill microbial communities. The biofilm MAG, BF1 Bin 72, had the highest proportional abundance within a sample, with 4.72% of the reads included in the assembled metagenome mapping to this genome. The next highest proportionally abundant MAG, BF1 Bin 89, contained 0.71% of the assembled metagenome. A total of 22 moderate quality Cloacimonadota genomes were publicly available, generated from diverse locations and ecosystems (Figure 3.2; Appendix B, Supplementary Table B1). The addition of 24 MAGs from the Southern Ontario landfill expands the available genomic representation of Cloacimonadota by 80%, a substantial increase for this radiation.

Phylogenetic analyses based on 74 bacterial single-copy genes, identified in all genomes (n = 54) as belonging to the monophyletic phylum Cloacimonadota. The phylogenetic tree displays Cloacimonadota genomes falling into two clades differentiated by ecosystem type, where one clade contains genomes derived predominantly from engineered ecosystems (Clade I), and the other clade contains genomes derived from natural, though not necessarily unimpacted, environmental ecosystems (Clade II) (Figure 3.2). Clade I is a monophyletic lineage (bootstrap = 100, total n = 41), composed primarily of genomes isolated from engineered ecosystems (29/41)

and includes all landfill-derived MAGs. Other genomes within this engineered clade were reconstructed from wastewater treatment plants (n = 5) of varying types: municipal

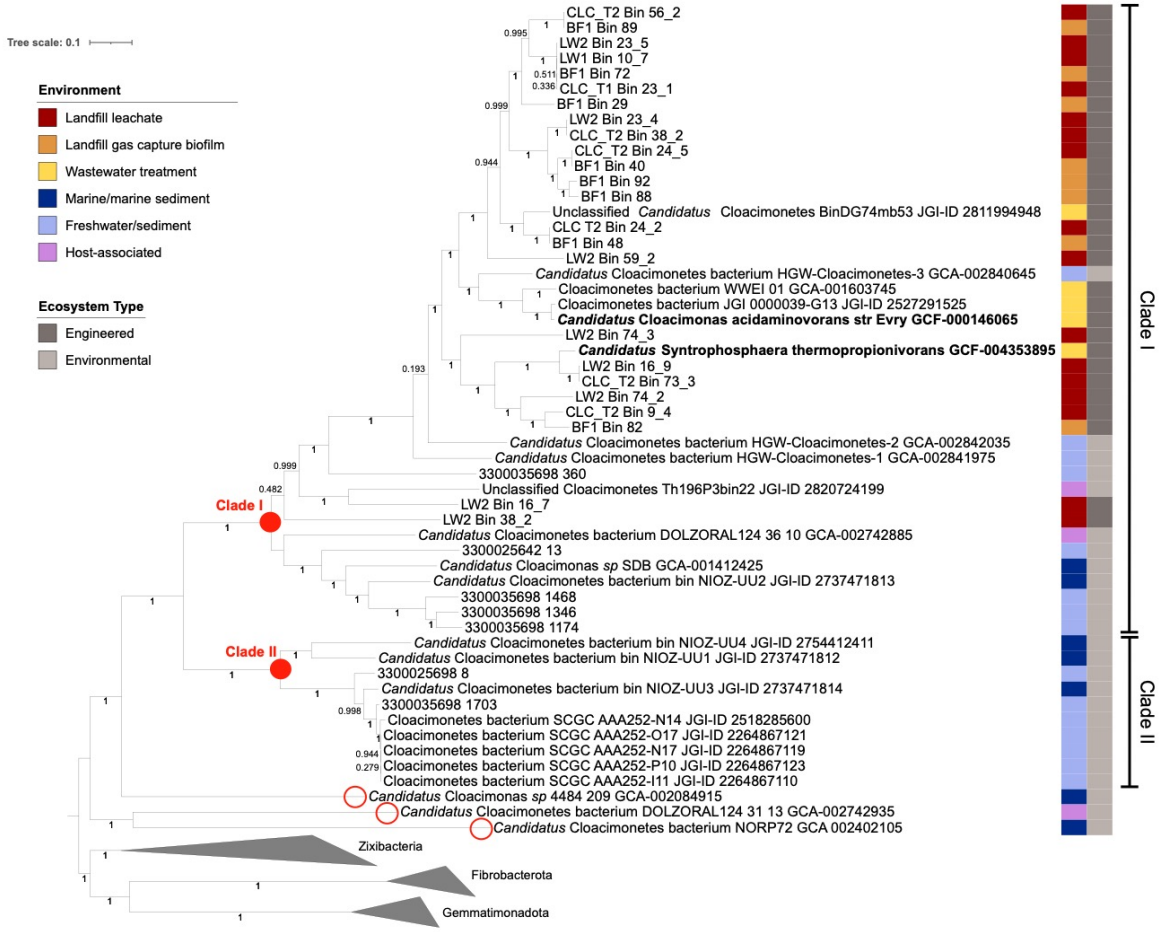


Figure 3.2 Phylogenomic placement of Cloacimonadota genomes. Gemmatimonadota, Fibrobacterota, and Zixibacteria genomes were included as outgroups. The outside greyscale bar indicates the type of ecosystem from which the genome was obtained, and the inner colored bars further discern the specific environment of origin. The two bolded genome names denote two genomes which have been previously investigated regarding metabolic function (Pelletier *et al.*, 2008; Dyksma and Gallert, 2019). The nodes that anchor Clade I and Clade II are highlighted with red circles. Three genomes did not fall into either clade, which are indicated with red circles on the tips beside the genome name.

(Pelletier *et al.*, 2008), terephthalate-degrading (Rinke *et al.*, 2013), cellulose-degrading (unpublished), and solid organic waste (unpublished). Two genomes were from a biogas plant reactor metagenome (unpublished). Clade I also contained nine genomes from natural ecosystems, and three that are host-associated (Figure 3.2). Three groundwater genomes were derived from a sediment ecosystem dedicated to radioactive waste disposal research (Hernsdorf *et al.*, 2017). Although this groundwater system is considered a natural environment, the sampled location is anthropogenically affected. Another two genomes in Clade I were isolated from animal hosts: one from a termite gut sample (unpublished) and one from a dolphin's oral cavity (Dudek *et al.*, 2017). Five genomes were isolated from an Antarctic lake, encompassing three *Candidatus* genera: Zophobacter, Stygibacter, and Tenebribacter (Williams *et al.*, 2021). The remaining two genomes were *Candidatus* Cloacimonas sp. SDB, and *Candidatus* Cloacimonetes bin NIOZ-UU2, which originated from a methanogenic culture enriched from contaminated sediments (Callaghan *et al.*, 2010), and the Black sea at a depth of 70 m (Villanueva *et al.*, 2021), respectively. Placed basal within Clade I, the environmental genomes were of specific interest when examining the evolution or acquisition of traits affiliated with adaptation to engineered environments.

Clade II is a monophyletic group of 10 genomes from environmental ecosystems (Figure 3.2; bootstrap = 100). Of those genomes, five are single-cell genomes from Sakinaw Lake, collected from a depth of 120 m, a location that is brackish, devoid of oxygen, and has high levels of hydrogen sulfide (1,000 μM) (Rinke *et al.*, 2013). Two genomes belong to the *Candidatus* genus Celanobacter, and the genomes originated from an Antarctic lake (Williams *et al.*, 2021). The

remaining three genomes from this clade originate from the water column from deep sea dives in the Black Sea (Villanueva *et al.*, 2021).

Beyond these two clades, there are three genomes that were more distantly related to other Cloacimonadota genomes than to each other, at the base of the Cloacimonadota radiation though still clearly associated with the phylum (Figure 3.2). *Candidatus* Cloacimonetes bacterium DOLZORAL124_31_13 originated from a dolphin's oral cavity (Dudek *et al.*, 2017), *Candidatus* Cloacimonetes bacterium NORP72 from an oxic subseafloor aquifer (Tully *et al.*, 2018), and *Candidatus* Cloacimonas sp. 4484_209 from marine deep-sea hydrothermal vent sediments (Dombrowski *et al.*, 2017). For metabolic comparisons, these three genomes were considered separate from both Clade I and Clade II.

3.3.3 Metabolic potential of Cloacimonadota

The Cloacimonadota landfill MAGs and reference genomes passing quality standards were analyzed with the KEGG Automatic Annotation Server (KAAS) (Appendix B, Supplementary Table B2) (Moriya *et al.*, 2007), and the program called Distilled and Refined Annotation of Metabolism (DRAM) (Shaffer *et al.*, 2020), each of which provides an overview of the metabolic functions encoded by each genome (Appendix B, Supplementary Table B3). More detailed examinations for specific functions of interest were carried out through manual curation. Newly described genomes from Williams *et al.* (2021) and *Candidatus* Syntrophosphaera thermopropionivorans (Dyksma and Gallert, 2019) were not subject to this pipeline, but their metabolic predictions were compared when assessing global trends within the Cloacimonadota.

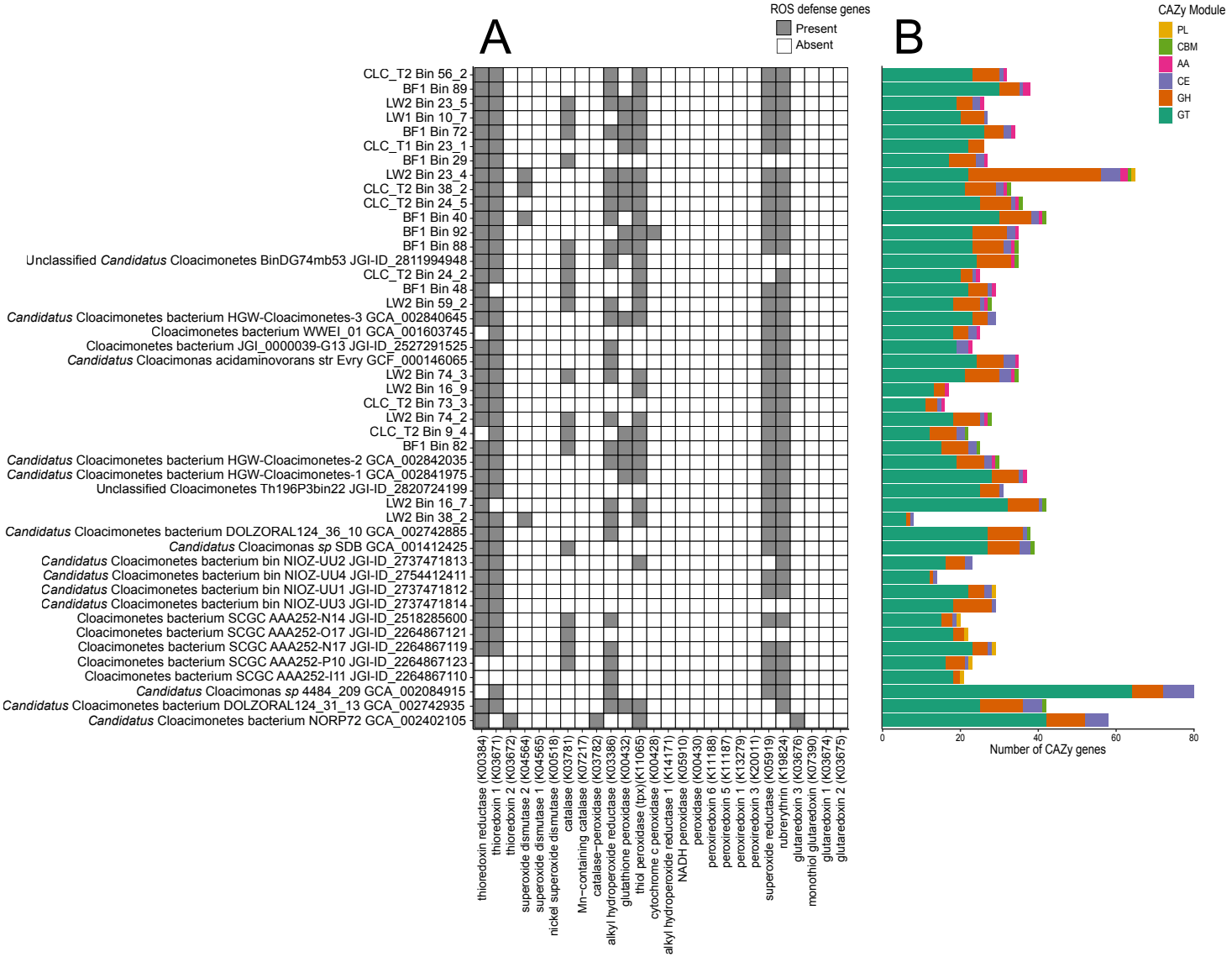
3.3.3.1 General lifestyle

Cloacimonadota are predicted to be anaerobic, Gram-negative bacteria (Pelletier *et al.*, 2008). Gram-negative bacteria have two membranes: an inner membrane surrounding the cytoplasm and an outer membrane protecting the cell from the environment that is made from lipopolysaccharide (LPS) molecules (Bertani and Ruiz, 2018). Based on KEGG annotations, all 46 genomes encoded partial pathways involved in lipopolysaccharide biosynthesis, such as the Kdo_2-lipid A biosynthesis pathway (Raetz pathway), indicating these organisms are likely Gram-negative. Cloacimonadota have been hypothesized to synthesize archaeal-like membrane lipids based on the presence of phosphoglycerol geranylgeranyltransferase (K17104) and geranylgeranylglycerol-phosphate geranylgeranyltransferase (K17105) (Villanueva *et al.*, 2021). These two enzymes were encoded in only 24% (11/46) and 30% (14/46) of Cloacimonadota genomes, respectively, with 24% of genomes encoding both genes and the potential ability to generate mixed archaeal/bacterial membranes. The genomes encoding both enzymes in Clade I included LW2 Bin 16_7, Unclassified Cloacimonetes Th196P3bin22, *Candidatus* Cloacimonetes bacterium DOLZORAL124_36_10, LW2 Bin 38_2, *Candidatus* Cloacimonas sp. SDB, and *Candidatus* Cloacimonetes bacterium bin NIOZ-UU2, all of which are deep-branching within Clade I, suggesting this trait was lost in later-diverging lineages. Another important cellular component is peptidoglycan, which provides structural support for maintaining cell shape, and the *mur* enzymes involved in peptidoglycan biosynthesis are partially represented in all Cloacimonadota genomes. The complete pathway was only detected in 6% of genomes from the engineered Clade I (2/35 genomes), 63% of the environmental Clade II genomes (5/8), and 100% of the remaining, basal Cloacimonadota genomes (3/3).

Cloacimonadota genomes encoded genes for type IV pili, including genes involved in the components for the major pilin, assembly ATPase, retraction ATPase, platform protein, and alignment protein (Craig *et al.*, 2019). Additionally, many genes involved in flagellar systems were encoded by a select group of five Cloacimonadota genomes, including LW2 Bin 16_7, Unclassified Cloacimonetes Th196P3bin22, *Candidatus* Cloacimonetes bacterium DOLZORAL124_36_10, *Candidatus* Cloacimonetes bacterium DOLZORAL124_31_13, and *Candidatus* Cloacimonetes bacterium NORP72. The flagellar genes encoded in these five genomes are consistent with genes found to be "core" in diverse and flagellated bacteria from 11 different phyla (Liu and Ochman, 2007).

The most common defenses against reactive oxygen species (ROS) encoded in Cloacimonadota genomes are thioredoxin reductase (K00384), and thioredoxin 1 (K03671), which were each present in 89% (41/46) of genomes. Alkyl hydroperoxide reductase (K03386) was encoded in 59% (27/46) of Cloacimonadota genomes. Thiol peroxidase (K11065) was detected in 74% (26/35) of Cloacimonadota genomes from Clade I, and none of the genomes from Clade II. Similarly, glutathione peroxidase (K00432) was encoded in 40% (14/35) of genomes from Clade I, and none from Clade II. Other ROS defenses were detected sporadically across Cloacimonadota genomes or were not detected at all (Figure 3.3). Based on the KEGG annotations, superoxide reductase (K05919) and rubrerythrin (K19824) were not detected in Cloacimonadota genomes though these genes were encoded in *Ca.* C. acidaminovorans and *Ca.* S. thermopropionivorans (Dyksma and Gallert, 2019). Therefore, a manual BLAST search was conducted using superoxide reductase (WP_133200558.1) and rubrerythrin (WP_133200557.1) genes annotated in the *Ca.* S. thermopropionivorans genome, where the genes were detected in 83% (38/46) and 89% (41/46) of

Figure 3.3 Reactive oxygen species (ROS) defense gene presence or absence (A), and count of carbohydrate-active genes (B), identified in Cloacimonadota genomes. Genomes are ordered as in the phylogeny in Figure 3.2. (A) Dark grey boxes denote ROS defense genes that were encoded, and white indicates ROS genes that were not detected. (B) Carbohydrate-active enzyme (CAZy) modules found in each genome colored by function. PL (yellow) = polysaccharide lyases; CBM (green) = carbohydrate binding modules; AA (pink) = auxiliary activity; CE (purple) = carbohydrate esterases; GH (orange) = glycoside hydrolases; GT (teal) = glycosyltransferases


A
ROS defense genes
Present
Absent
CLC_T2 Bin 56_2
BF1 Bin 89
LW2 Bin 23_5
LW1 Bin 10_7
BF1 Bin 72
CLC_T1 Bin 23_1
BF1 Bin 29
LW2 Bin 23_4
CLC_T2 Bin 38_2
CLC_T2 Bin 24_5
BF1 Bin 40
BF1 Bin 92
BF1 Bin 88
Unclassified Candidatus Cloacimonetes BinDG74mb53 JGI-ID_2811994948
CLC_T2 Bin 24_2
BF1 Bin 48
LW2 Bin 59_2
Candidatus Cloacimonetes bacterium HGW-Cloacimonetes-3 GCA_002840645
Cloacimonetes bacterium WWEI_01 GCA_001603745
Cloacimonetes bacterium JGI_0000039-G13 JGI-ID_2527291525
Candidatus Cloacimonas acidaminovorans str Evry GCF_000146065
LW2 Bin 74_3
LW2 Bin 16_9
CLC_T2 Bin 73_3
LW2 Bin 74_2
CLC_T2 Bin 9_4
BF1 Bin 82
Candidatus Cloacimonetes bacterium HGW-Cloacimonetes-2 GCA_002842035
Candidatus Cloacimonetes bacterium HGW-Cloacimonetes-1 GCA_002841975
Unclassified Cloacimonetes Th196P3bin22 JGI-ID_2820724199
LW2 Bin 16_7
LW2 Bin 38_2
Candidatus Cloacimonetes bacterium DOLZORAL124_36_10 GCA_002742885
Candidatus Cloacimonas sp SDB GCA_001412425
Candidatus Cloacimonetes bacterium bin NIOZ-UU2 JGI-ID_2737471813
Candidatus Cloacimonetes bacterium bin NIOZ-UU4 JGI-ID_2754412411
Candidatus Cloacimonetes bacterium bin NIOZ-UU1 JGI-ID_2737471812
Candidatus Cloacimonetes bacterium bin NIOZ-UU3 JGI-ID_2737471814
Cloacimonetes bacterium SCGC AAA252-N14 JGI-ID_2518285600
Cloacimonetes bacterium SCGC AAA252-O17 JGI-ID_2264867121
Cloacimonetes bacterium SCGC AAA252-N17 JGI-ID_2264867119
Cloacimonetes bacterium SCGC AAA252-P10 JGI-ID_2264867123
Cloacimonetes bacterium SCGC AAA252-I11 JGI-ID_2264867110
Candidatus Cloacimonas sp 4484_209 GCA_002084915
Candidatus Cloacimonetes bacterium DOLZORAL124_31_13 GCA_002742935
Candidatus Cloacimonetes bacterium NORP72 GCA_002402105
thioredoxin reductase (K00384)
thioredoxin 1 (K03671)
thioredoxin 2 (K03672)
superoxide dismutase 2 (K04564)
superoxide dismutase 1 (K04565)
nickel superoxide dismutase (K00518)
catalase (K03781)
Mn−containing catalase (K07217)
catalase−peroxidase (K03782)
alkyl hydroperoxide reductase (K03386)
glutathione peroxidase (K00432)
thiol peroxidase (tpx)(K11065)
cytochrome c peroxidase (K00428)
alkyl hydroperoxide reductase 1 (K14171)
NADH peroxidase (K05910)
peroxidase (K00430)
peroxiredoxin 6 (K11188)
peroxiredoxin 5 (K11187)
peroxiredoxin 1 (K13279)
peroxiredoxin 3 (K20011)
superoxide reductase (K05919)
rubrerythrin (K19824)
glutaredoxin 3 (K03676)
monothiol glutaredoxin (K07390)
glutaredoxin 1 (K03674)
glutaredoxin 2 (K03675)
B
CAZy Module
PL
CBM
AA
CE
GH
GT
0
20
40
60
80
Number of CAZy genes

the genomes, respectively. The distribution of genes in Cloacimonadota genomes is generally consistent with a previous broad search of ROS defenses across bacterial phyla (Johnson and Hug, 2019), with the exception of rubrerythrin, which was not detected in the broader search, but was identified here under a targeted search. It is possible there was enough sequence dissimilarity for superoxide reductase and rubrerythrin to be missed with the automated KEGG annotation, though both were detected with manual BLAST searches. Collectively, for both Cloacimonadota clades, management of damaging free radicals appears limited, likely constraining these microorganisms to anoxic environments.

3.3.3.2 Carbon metabolism

The Cloacimonadota genomes generally contain a full glycolysis pathway, with the exception of pyruvate kinase which catalyzes the formation of pyruvate. The other nine enzymes of the glycolysis pathway were found in 72% to 98% of genomes (33 to 45/46), whereas pyruvate kinase (EC: 2.7.1.40) was only identified in 26% of the genomes (12/46), and not associated with any specific phylogenetic radiation. To ensure pyruvate kinase was not missed during automated annotation or due to partial hits, additional BLAST searches of Cloacimonadota genomes for pyruvate kinase were conducted, which did not locate any additional genes. While 74% of Cloacimonadota genomes were missing pyruvate kinase, pyruvate orthophosphate dikinase (EC: 2.7.9.1), which catalyzes a reversible reaction between pyruvate and phosphoenolpyruvate (PEP), was present in 91% of the genomes (42/46). Pyruvate orthophosphate dikinase has been shown to mediate pyruvate flux in *Clostridium thermocellum* (Olson *et al.*, 2017), and therefore it is predicted to play a similar role in the Cloacimonadota. The presence of either pyruvate kinase or pyruvate orthophosphate dikinase was not detected in three genomes (Cloacimonetes bacterium

SCGC AAA252-P10, *Candidatus* Cloacimonas sp. 4484-209, and CLC_T1 Bin 23_1). Pyruvate is predicted to be converted to acetyl-CoA by pyruvate ferredoxin oxidoreductase (PFOR) (EC: 1.2.7.1), which was encoded in 98% of the genomes (45/46). Additionally, 48% (22/46) of the Cloacimonadota genomes encode a predicted pyruvate dehydrogenase (EC: 1.2.4.1), underscoring the likely importance of acetyl-CoA generation within Cloacimonadota metabolism. None of the genomes contained a gene for pyruvate formate lyase, indicating similar anaerobic processes are not active in these organisms.

The Cloacimonadota genomes encode highly incomplete tricarboxylic acid (TCA) cycles, with only two of the genes in the TCA cycle found in more than 22% of the genomes. Fumarate hydratase (EC: 4.2.1.2), which catalyzes the conversion of fumarate to malate, was found in 89% of the genomes (41/46), and oxoglutarate ferredoxin oxidoreductase (EC: 1.2.7.3 or 1.2.7.11), catalyzing the formation of succinyl-CoA from 2-oxoglutarate, was identified in 96% of the genomes (44/45). The Cloacimonadota genomes typically possess part of the pentose phosphate pathway, particularly the genes transketolase (EC: 2.2.1.1), ribulose-phosphate 3-epimerase (EC: 5.1.3.1), ribose 5-phosphate isomerase A (EC: 5.3.1.6), ribose-phosphate pyrophosphokinase (EC: 2.7.6.1), and phosphoglucomutase (EC: 5.4.2.2), with each being present in 80 to 96% of genomes (37 to 44/46) highlighting a capacity for sugar transformation.

Sugars may derive from more complex carbohydrates - Cloacimonadota genomes encode a suite of carbohydrate-active enzymes (CAZy), the enzymes involved in cleavage of specific sugar linkages (Lombard *et al.*, 2014). Prominent CAZy enzymes encoded in Cloacimonadota genomes include starch and chitin cleavage proteins from the DRAM annotation (Appendix B, Supplementary Figure B1). The prevalence of these CAZy enzymes was higher in the genomes

from Clade I, suggesting Cloacimonadota in engineered environments can each individually catabolize a wider variety of substrates than their Clade II counterparts, perhaps reflective of the higher heterogeneity of the available substrates in these engineered systems. To further investigate CAZy enzymes the genomes were annotated with dbCAN2 (Zhang *et al.*, 2018). The CAZy modules with the highest number of genes in Cloacimonadota genomes were glycosyltransferases and glycoside hydrolases (Figure 3.3; Figure 3.4). Unique to Clade I are auxiliary activities (Figure 3.3, pink). Specifically, CAZy family AA6 was encoded in 69% (24/35) of Clade I genomes and is classified as 1,4-benzoquinone reductase (EC: 1.6.5.6) in the CAZy database, where in *Burkholderia* sp. the enzyme plays a role in 4-aminophenol metabolism (Takenaka *et al.*, 2011).

Another important aspect of Cloacimonadota carbon metabolism involves acetate. The acetogenic pathway requires phosphate acetyltransferase and acetate kinase, with concomitant substrate-level phosphorylation of ADP to ATP. Phosphate acetyltransferase (EC: 2.3.1.8) was present in 91% of genomes (42/46). Acetate kinase (EC: 2.7.2.1) was present in 72% of all Cloacimonadota genomes (33/46). However, the presence of acetate kinase is not evenly distributed across the phylogenetic tree, where 91% of the Clade I genomes (32/35) encode acetate kinase while this gene was not detected in the Clade II genomes. Acetate kinase was detected in one of the three more basal genomes outside Clade I and Clade II (Candidatus Cloacimonetes bacterium DOLZORAL_124_31_13). Some Cloacimonadota genomes encode the potential to regenerate acetyl-CoA through acetyl-CoA synthetase (EC: 6.2.1.1), which was present in 22% of genomes (10/46). The analyses suggest that Cloacimonadota are not capable of generating acetyl-CoA via the Wood Ljundahl pathway, as the CO-methylating acetyl-CoA synthetase (EC: 2.3.1.169) was not detected in any genome. Intermediate steps of the Wood-Ljungdahl pathway,

methenyltetrahydrofolate cyclohydrolase (EC: 3.5.4.9) and methylenetetrahydrofolate dehydrogenase (EC: 1.5.1.5), were detected in all Cloacimonadota genomes, and formate-tetrahydrofolate ligase (EC: 6.3.4.3) was detected in 61% of genomes (28/46). Though these genes can be associated with carbon fixation in conjunction with CO-methylating acetyl-CoA synthetase, I expect these genes are indicative of folate uptake for DNA synthesis in these organisms given dihydrofolate reductase (EC: 1.5.1.3) was detected in 65% of genomes (30/46).

Ca. S. thermopropionivorans was identified within a propionate-degrading, methanogenic enrichment culture, where it may play an important role in anaerobic mineralization of organic matter (Dyksma and Gallert, 2019). The genome of *Ca.* Syntrophosphaera thermopropionivorans encodes genes for propionate oxidation via methylmalonyl-CoA (Dyksma and Gallert, 2019), though the capacity for propionate oxidation in this organism has not been directly determined. Several genes from this pathway were not detected in *Ca.* S. thermopropionivorans' genome, including the critical gene propionate-CoA transferase (EC: 2.8.3.1), which converts propionate to propionyl-CoA, and succinyl-CoA synthase (EC: 6.2.1.5). Similarly, these genes were detected in 0 and only 17% of Cloacimonadota genomes (8/46), respectively. Other genes in this pathway were prevalent in Cloacimonadota genomes, such as propionyl-CoA carboxylase (EC: 6.4.1.3; 89% of genomes), methyl malonyl-CoA/ethyl malonyl-CoA epimerase (EC: 5.1.99.1; 78%), and methylmalonyl-CoA mutase (EC: 5.4.99.2; 93%). Several enzymes in this pathway were present more inconsistently, such as succinate dehydrogenase (EC: 1.3.5.1; 22%) and malate

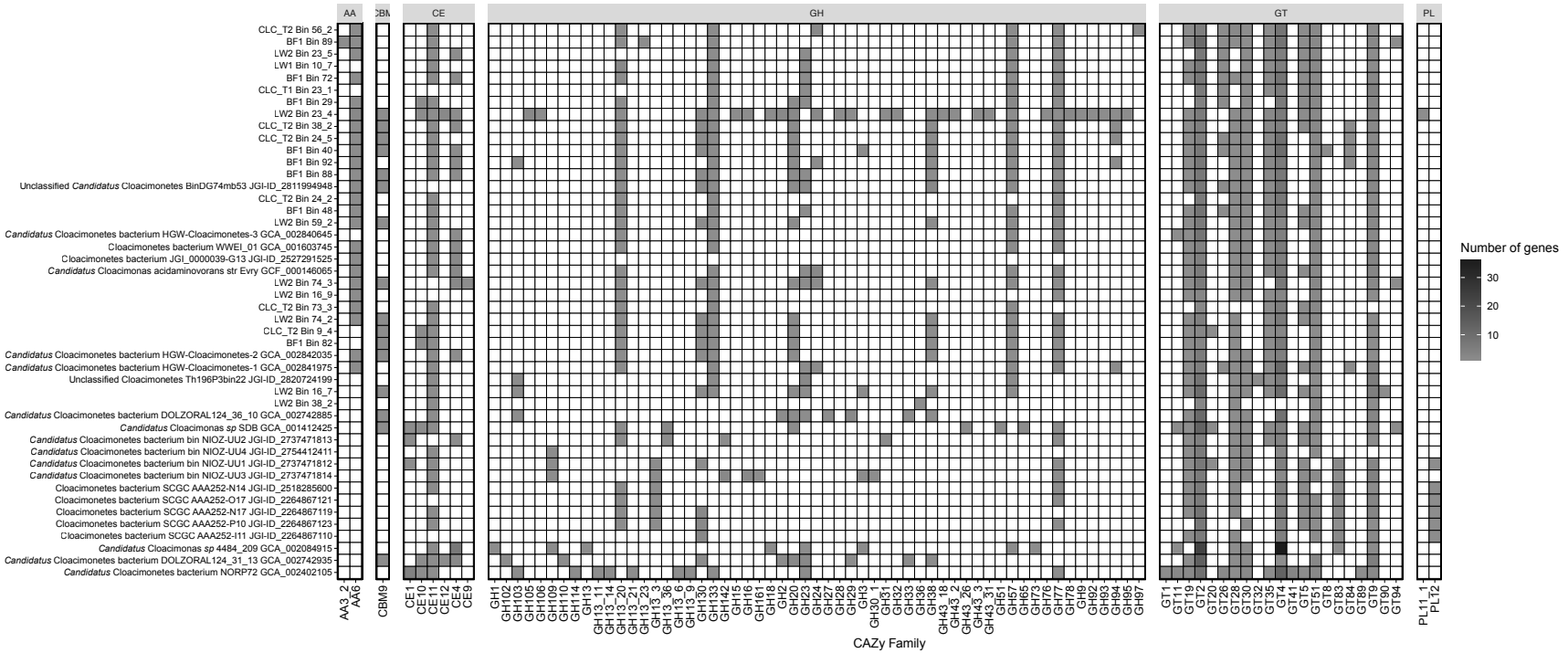


Figure 3.4 Distribution of genes in each CAZy family per Cloacimonadota genome. The CAZy families are: AA = auxiliary activity; CBM = carbohydrate binding modules; CE = carbohydrate esterases; GH = glycoside hydrolases; GT = glycosyltransferases, and PL = polysaccharide lyases.

dehydrogenase (EC: 1.1.1.37; 9%). Pyruvate carboxylase (EC: 6.4.1.1) was encoded in 96% of genomes (44/46), and pyruvate ferredoxin oxidoreductase (PFOR; EC: 1.2.7.1), as mentioned above, was encoded in 98% of genomes (45/46). Collectively, Cloacimonadota genomes encoded a fragmented pathway for propionate oxidation, calling into question their suggested role in propionate fermentation.

Overall, most Cloacimonadota genomes encode some capacity for sugar cleavage from more complex carbohydrates, sugar fermentation to acetate, and encode a fragmented set of genes involved in the methylmalonyl-CoA pathway for propionate oxidation.

3.3.3.3 Energy metabolism

Cloacimonadota are likely fermenters, though for five genomes (11%), fermentation is not the sole predicted mechanism for energy generation. All of the genomes surveyed lack complexes from the oxidative electron transport chain, including absence of a canonical Complex I (NADH dehydrogenase, EC: 7.1.1.2). None of the genomes encoded a full 14-subunit NADH-quinone oxidoreductase (NuoA, B, C, D, E, F, G, H, I, J, K, L M, and N). One genome, CLC_T2 Bin 9_4, did not encode any subunits, while all other genomes encoded at least one subunit. The most complete NADH dehydrogenases were in the genomes *Ca.* Cloacimonetes bacterium HGW Cloacimonetes-3 (encoding nuoBCDEFGHL), and *Ca.* Cloacimonetes bacterium NORP72 (encoding nuoBCDHILMN), each with eight subunits (57%). Given the high completion threshold (70% complete) required for the genomes under consideration, the NADH dehydrogenase is thus interpreted as non-functional in most Cloacimonadota genomes, with one exception. *Ca.* Cloacimonetes bacterium HGW Cloacimonetes-3 contains the NuoEFG module and therefore has

the potential for NADH oxidation independent of the rest of the complex (Deusch *et al.*, 2019). Complex II, succinate dehydrogenase (EC: 1.3.5.1), was detected in 88% of genomes in the environmental Clade II (7/8) but was found in only one of the genomes in the engineered Clade I (*Ca.* Cloacimonetes NIOZ-UU2). Complex III, the cytochrome bc_1 complex (EC: 7.1.1.8), was absent from all genomes. Complex IV, cytochrome *c* oxidase (EC: 7.1.1.9), was absent in all genomes except for *Ca*. Cloacimonetes bacterium NORP72. Complex V, ATP synthase, occurs in two forms. The F-type ATPase (alpha, beta, gamma, delta, epsilon, a, b, c – 8 subunits) was partially present in 23% (8/35) of the Clade I genomes. In contrast, 88% (7/8) of the environmental Clade II genomes contained at least partial complexes, with five of eight genomes encoding complete F-type ATPases (Appendix B, Supplementary Figure B1). The V/A-type ATPase (A, B, C, D, E, F, G/H, I, K – 9 subunits), was partially detected in 86% (30/35) of the engineered Clade I genomes and was completely absent in the Clade II genomes, though, no complete V/A-type ATPases were detected in any Cloacimonadota genome (Appendix B, Supplementary Figure B1). Clade I genomes encoding a V/A-type ATPase is consistent with earlier metabolic predictions from Cloacimonadota genomes from engineered ecosystems (Dyksma and Gallert, 2019).

ATP synthases rely on generation of a membrane potential, and the absence of a canonical electron transport chain indicates an alternate mechanism is used by Cloacimonadota. The Rnf electron transport complex is a ferredoxin:NAD oxidoreductase that oxidizes reduced ferredoxin and reduces NAD, generating a transmembrane ion gradient, in a reversible reaction (Hess *et al.*, 2016). Each of the genes from the Rnf complex (*rnfABCDEG*) were encoded in 85 to 93% (39 to 43/46) of Cloacimonadota genomes, with 70% (32/46) of genomes encoding the complete complex. The Rnf complex can be important for bacteria in terms of energy conservation, and

along with an ATP synthase, creates a simple two-step respiratory chain in the fermentative *Thermotoga maritima* (Kuhns *et al.*, 2020). Two genomes, LW2 Bin 23_5 and *Ca.* Cloacimonetes bacterium NORP72, did not encode any components of the complex. The NORP72 genome contained a more-complete canonical electron transport chain than any other Cloacimonadota, suggesting this basal lineage may generate a membrane potential through NADH dehydrogenase and cytochrome *c* oxidase. For LW2 Bin 23_5, this may indicate an absence due to incomplete assembly or binning, because it is predicted to be 84% complete.

Hydrogenases are an alternative mechanism for generating a membrane potential. Ferredoxin hydrogenase (EC: 1.12.7.2) was present in 89% of the Clade I Cloacimonadota genomes (31/35), and none of the Clade II genomes. Of the three genomes that did not fall into either clade, a ferredoxin hydrogenase was detected only in *Ca*. Cloacimonetes bacterium DOLZORAL_124_31_13. There was no overlap between the presence of ferredoxin hydrogenases and complete ATP synthases. Additionally, formate dehydrogenase (EC: 1.17.1.9) was encoded in 74% of Clade I Cloacimonadota genomes (26/35), one Clade II genome (Cloacimonetes bacterium SCGC AAA252-N17), and one genome not in either clade (*Ca*. Cloacimonetes sp. 4484_209). Group A3 [FeFe] hydrogenases were only present in Clade I organisms above the basal clade of environmentally-sourced genomes (32/33). Basal environmentally-sourced Clade I genomes and Clade II genomes instead encode Group 3c [NiFe] hydrogenases (15/16 genomes). Group 3c [NiFe] hydrogenases are implicated in electron transfer from H_2 to ferredoxin and heterodisulfide, and so may play the same role in regeneration of electron donors to the respiratory complexes. The Group A3 [FeFe] and Group 3c [NiFe] hydrogenases can be part of flavin-based electron bifurcation, where the reduction of two electron acceptors is completed using a single

electron donor, improving energy capture for anaerobes where energy yields are low (Poudel *et al*., 2018). The functionality of complex V in the presence of membrane-potential-generating mechanisms is possible due to the presence of the Rnf complex, but the significance of hydrogenases in Cloacimonadota energy generation is unclear due to their inconsistent distribution across genomes and clades.

Cloacimonadota genomes largely lack any genes involved in nitrogen, sulfur and methane transformations for energy generation. For nitrogen cycling, 80% of the engineered Clade I genomes (28/35) have predicted hydroxylamine reductases (EC: 1.7.99.1), or ammonia:acceptor oxidoreductases, which were absent in all Clade II genomes. The majority of genomes lack genes involved in denitrification or dissimilative reduction of nitrate. The only exception was the genome *Ca.* Cloacimonetes bacterium NORP72, which encodes both a predicted nitrate reductase and a predicted nitric oxide reductase – these may connect to the encoded partial electron transport chain, allowing this organism, uniquely among the Cloacimonadota, to respire using nitrate as a terminal electron acceptor. Cloacimonadota genomes generally lacked genes involved in sulfur metabolism and sulfur cycling, but the environmental Clade II genomes did show the presence of several relevant genes. Specifically, 88% (7/8) of genomes had both an anaerobic sulfite reductase (*asr*) and a sulfur reductase (EC: 1.12.98.4). The Clade II genome that lacked both of these genes was Cloacimonetes bacterium SCGC AAA252-P10. These predicted functions point to a role for Cloacimonadota from marine and anoxic lake environments in hydrogen sulfide production. All Cloacimonadota genomes lacked haloacid dehalogenases (E.C: 3.8.1.2), reductive dehalogenases (EC: 1.21.99.5), and arsenic reductases (EC: 1.20.4.1), and only 17% of genomes (8/47) contained predicted selenate reductases (EC:1.97.1.9).

Previous analyses have suggested Cloacimonadota ferment amino acids to generate energy, pointing to five ferredoxin oxidoreductases as key markers for this activity (Pelletier *et al.*, 2008; Williams *et al.*, 2021). However, neither study discussed specific amino acid degradation pathways, which generate the necessary 2-oxoacid substrates. I identified limited, but likely energetically significant, capacity to degrade asparagine, aspartate, and glutamate to oxaloacetate or 2-oxoglutarate. Specifically, Cloacimonadota nearly universally contain the capacity to convert asparagine to aspartate via asparagine synthase (EC: 3.5.1.1), and to ferment aspartate to oxaloacetate and/or fumarate via aspartate aminotransferase (EC: 2.6.1.1). Similarly, the capacity to convert glutamate to either glutamine (and subsequently to sugar/purine metabolic pathways) or to 2-oxoglutarate is encoded in most genomes, encoded by glutamine synthetase (EC: 6.3.1.2) and glutamate dehydrogenase (EC: 1.4.1.4), respectively. Aspartate and glutamate degradation are the main pathways observed for anaerobic amino acid degradation tied to ferredoxin oxidoreductases. Cysteine showed the only phylogenetically associated pattern of degradation, with Clade I genomes encoding the potential for degradation to pyruvate via cysteine-S-conjugate beta-lyase (EC: 4.4.1.13), which was absent in Clade II and basal genomes. For other amino acids, degradation potential was limited. Serine and threonine are predicted to be converted to glycine with glycine hydroxymethyl transferase (EC: 2.1.2.1) and threonine aldolase (EC: 4.1.2.48). Only three genomes were predicted to convert serine to pyruvate via L-serine ammonia lyase (EC: 4.3.1.17). Glycine is predicted to be metabolized to pyruvate in three genomes, and otherwise feeds into 1-carbon metabolism through formation of 5,10-methylenetetrahydrofolate. A complete absence of metabolic pathways for alanine, valine, leucine, isoleucine, lysine, arginine, proline, tyrosine, phenylalanine (to phenylpyruvate only), and tryptophan (some to indole) were observed.

Methionine is predicted to be degraded to S-methyl-5'-thioinosine, which is not associated with energy or electron transfer.

Downstream oxidation of 2-oxoacids relies on five enzymes, including pyruvate ferredoxin oxidoreductase (K03737), 2-ketoglutarate ferredoxin oxidoreductase (K00174, K00175, K00176, and K00177), aldehyde ferredoxin oxidoreductase (K03738), branched-chain ketoacid ferredoxin oxidoreductase (K00186, K00187, K00188, and K00189), and indolepyruvate ferredoxin oxidoreductase (K00179, and K00180). All five ferredoxin oxidoreductases were encoded in at least some Cloacimonadota genomes, though with varying distributions across clades. Abundant ferredoxin oxidoreductases in all genomes included pyruvate ferredoxin oxidoreductase (K03737, 98% of genomes (45/46)) and indolepyruvate ferredoxin oxidoreductase (K00179, 91% of genomes (42/46)). Three subunits of 2-ketoglutarate ferredoxin oxidoreductase (*korA*, *korB* and *korD*) were each encoded in over 93% of genomes, but one subunit (*korC*) was only encoded in 24% of genomes (11/46). Aldehyde ferredoxin oxidoreductase was encoded in 66% of Clade I genomes (23/35) but none of the Clade II genomes. In contrast, branched chain ketoacid ferredoxin oxidoreductase (*vor*) was not detected in Clade I genomes but was patchily present in Clade II genomes. Two of the four subunits, *vorA* and *vorC*, were both encoded in 88% of Clade II environmental genomes (7/8), while the other two subunits, *vorB* and *vorG*, were not encoded in any genome. Overall, the presence of ferredoxin oxidoreductases suggests the capability of amino acid fermentation encoded within Cloacimonadota genomes, with some variation between clades. To evaluate the distribution of peptidases and peptidase inhibitors, BLASTp was used to search each of the genomes against the MEROPS database (Rawlings *et al.*, 2018). The peptidase families with the highest number of genes in Cloacimonadota genomes were the metallo, serine, and

cysteine peptidases (Figure 3.5). Cloacimonadota genomes encoded a variety of peptidases with no strong delineation between clades. The observed Cloacimonadota auxotrophy for most amino acids may explain the importance of peptidases despite only a few amino acids being directly degraded for energy.

From the genomic assessments of metabolic capacity, it is predicted that these Cloacimonadota are obligate anaerobic acetogenic organisms, without major roles in the sulfur and nitrogen cycles (Figure 3.6). The examined Cloacimonadota genomes contain the genes PFOR and acetate kinase, involved in conversion of pyruvate to acetate. However, acetate kinase was only prevalent in Clade I, suggesting differences in energy generation between the two clades defined by the phylogenetic tree. For energy production via ATPases, the environmental Clade II organisms encode the F-type ATPase, whereas the engineered Clade I organisms lack a complete ATPase but have partial V/A type ATPases. Succinate dehydrogenase was frequently identified in Clade II genomes, while formate dehydrogenase was consistently detected in Clade I genomes, indicating different mechanisms of generating a membrane potential. Most Cloacimonadota encoded the RNF complex, which also may help create a transmembrane proton gradient, influencing energy generation. The presence of ferredoxin oxidoreductases and a variety of peptidases suggest Cloacimonadota may play a role in the degradation of proteins and/or amino acids.

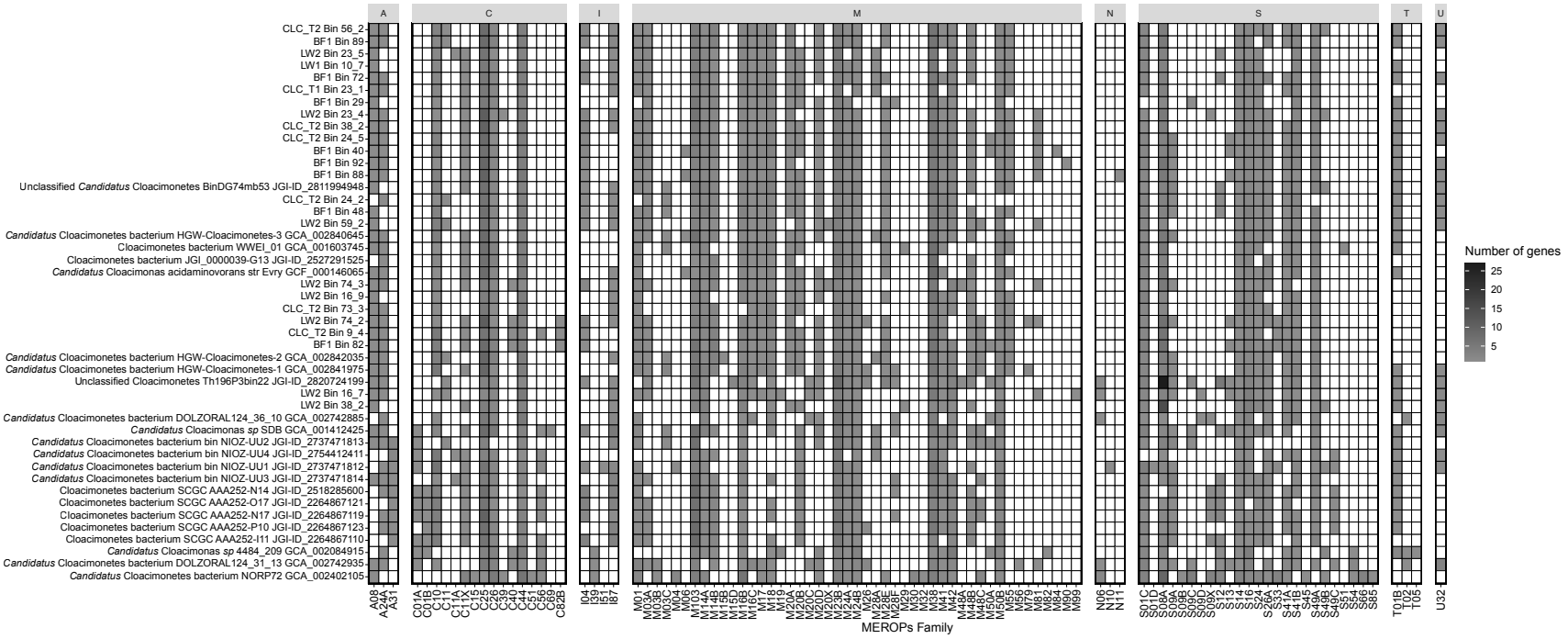


Figure 3.5 Distribution of peptidases per Cloacimonadota genome based on annotation with the MEROPs database. The MEROPS families are: A = aspartic; C = cysteine; I = inhibitors; M = metallo; N = asparagine; S = serine; T = threonine; and U = unknown.

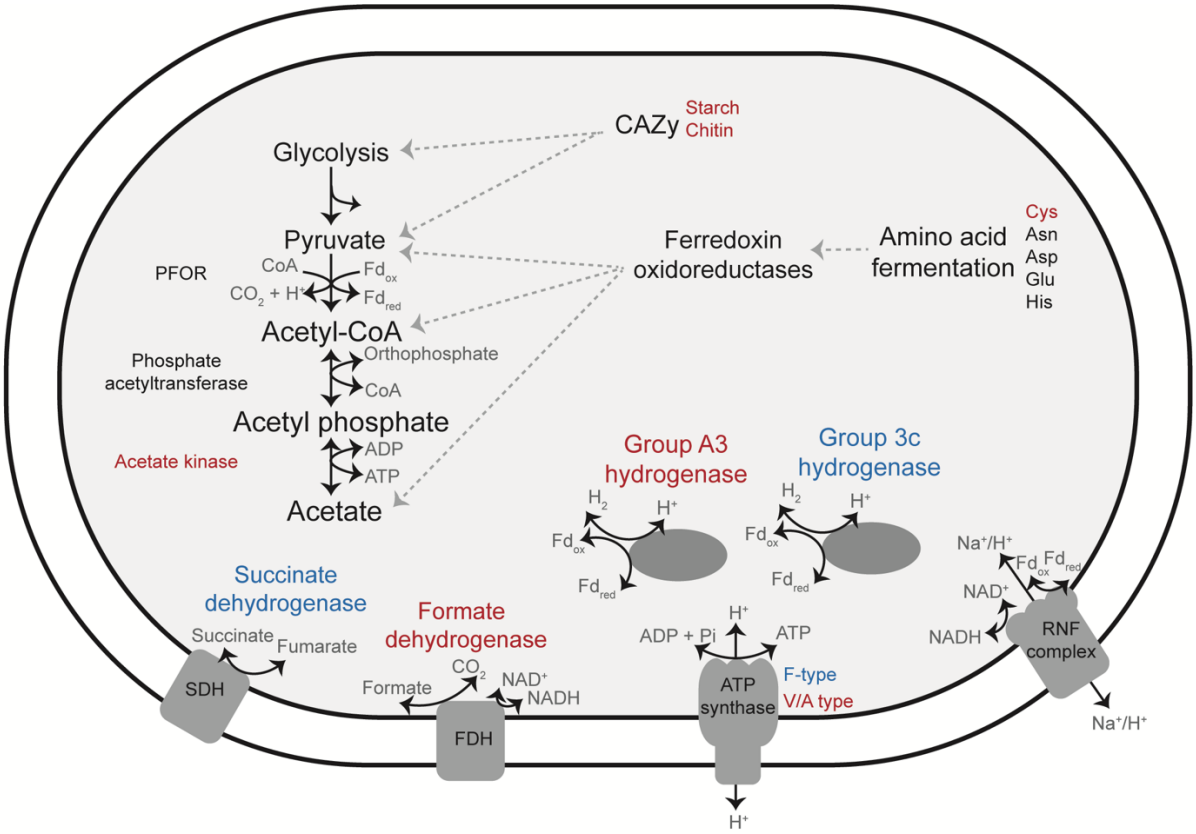


Figure 3.6 Simplified cell diagram of carbon and energy metabolism in Cloacimonadota. Reconstructed from annotations of assembled MAGs and reference genomes. Red labels indicated genes encoded predominantly in Clade I genomes, largely from engineered environments, and blue labels indicated genes encoded predominantly in Clade II genomes. Black labels specify proteins encoded in all Cloacimonadota genomes. Dotted lines denote possible paths for substrate degradation. Not all genes were detected in the glycolysis or propionate oxidation pathways (see text for details). Amino acid degradation is indicated by the presence of ferredoxin oxidoreductases. CAZy = carbohydrate-active enzymes; PFOR = pyruvate ferredoxin oxidoreductase; SDH = succinate dehydrogenase; FDH = formate dehydrogenase.

3.3.4 Cloacimonadota pangenome

The phylogenetic separation of engineered and environmental ecosystem-derived organisms was unexpected, so the pangenome of the phylum was examined to assess traits that separated these groups and might explain adaptations specific to each habitat type.

The anvi'o analysis platform was used to create a pangenome for the 46 high quality genomes included in the metabolic analyses (Figure 3.7). The pangenome analysis used pairwise amino acid similarity searches to identify gene clusters (Delmont and Eren, 2018). The pangenome analysis resulted in 5,149 gene clusters from a total of 64,300 genes across the 46 genomes, capturing 70% of genes in a gene cluster. Each of the gene clusters contain genes from a minimum of three genomes, as singletons and doubletons were excluded from this pangenome analysis (parameter --min_occurrence = 3). To highlight patterns of shared genomic profiles, genomes were ordered based on the frequency of the gene clusters that they share (Figure 3.7). The genomes clustered by the ecosystem type from which the genome originated, similar to the phylogenetic tree (Figure 3.2). Groupings of gene clusters were manually selected based on highly conserved gene cluster patterns across the Cloacimonadota genomes (Groups 1-4, Figure 3.8). These groups contained a total of 2,719 gene clusters of the original 5,149, where the median proportion of an individual genome's genes captured by these four groups was 70%.

Each gene cluster group was examined based on the COG categories for the represented gene clusters (Figure 3.8). Group 1, labelled "core", contains gene clusters that are near-universal and highly conserved across all Cloacimonadota, and contains 347 gene clusters.

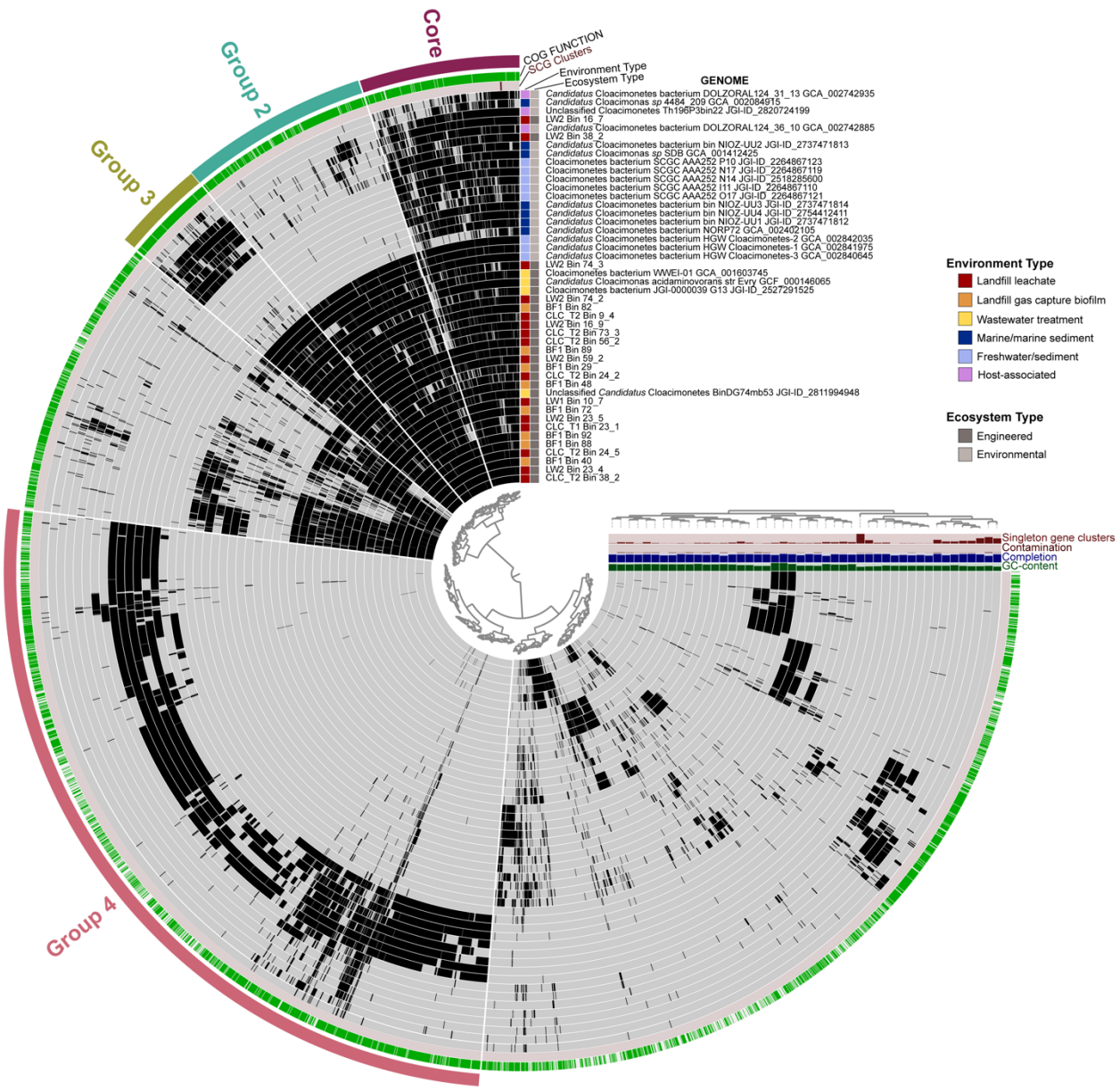


Figure 3.7 Cloacimonadota pangenome with 5,149 gene clusters from 64,300 genes identified from 46 genomes. Genomes are organized based on the frequency of the gene clusters that they share, with each genome represented as a ring on the plot and genes marked as present (black) or absent (grey) within gene clusters (rays around the circle). Bar graphs on the right indicate parameters for each genome: GC content (%), completion (%), contamination (%), and number of singleton gene clusters. Colored boxes to the left of the genome names indicate the environment type and ecosystem type from which the genome originated.

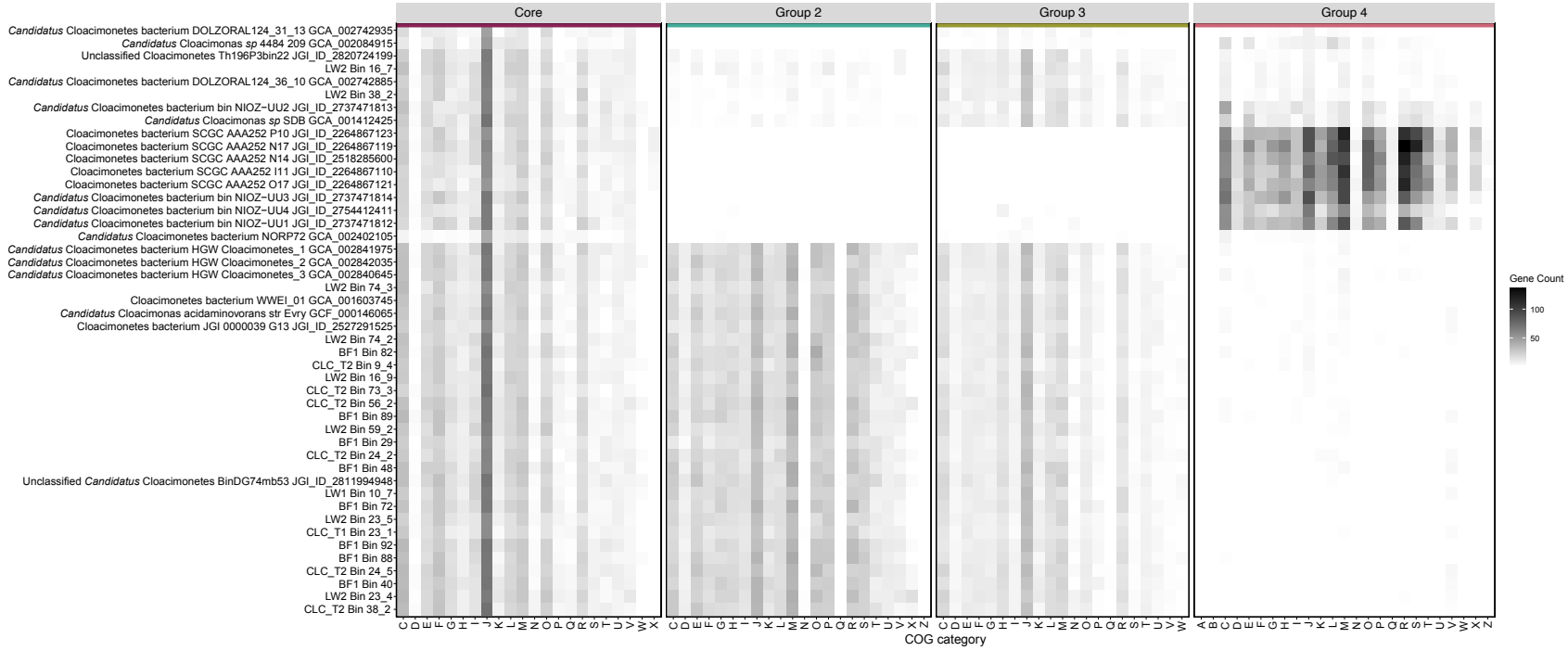


Figure 3.8 COG category distribution per genome for groups identified in the pangenome. Groups correspond to those visually selected from the pangenome and are colored to match Figure 4. Clusters of Orthologous Groups (COGs) were annotated in the anvi'o pipeline. Heatmaps were created from the assigned COG categories for gene clusters for each gene cluster in the pangenome, and were visualized using ggplot2 (Ginestet, 2011) with the genomes ordered to match the clustering from the pangenome analysis. COG categories by letter are: (A) RNA processing and modification; (B) Chromatin structure and dynamics; (C) Energy production and conversion; (D) Cell cycle control and mitosis; (E) Amino acid metabolism and transport; (F) Nucleotide metabolism and transport; (G) Carbohydrate metabolism and transport; (H) Coenzyme metabolism; (I) Lipid metabolism; (J) Translation; (K) Transcription; (L) Replication and repair; (M) Cell wall/membrane/envelope biogenesis; (N) Cell motility; (O) Post-translational modification; (P) Inorganic ion transport and metabolism; (Q) Secondary structure; (R) General function prediction only; (S) Function unknown; (T) Signal transduction; (U) Intracellular trafficking; (V) Defense mechanisms; (W) Extracellular structures; (X) Mobliome: prophages, transposons; (Z) Cytoskeleton.

The “core” section contained the highest number of COG category “J – translation” genes. Other categories of essential functions were also well represented, such as “C - energy production and conversion”, “F – nucleotide metabolism and transport”, “L – replication and repair”, “M – cell wall/membrane/envelope biogenesis”, and “O – post-translational modification”. Interestingly, “R – general function prediction only” encompasses 6% of the “core” wedge, identifying highly conserved genes with no known function that are likely important for Cloacimonadota growth or survival.

Group 2 contained 405 gene clusters that were nearly universal in genomes originating from engineered ecosystems but largely absent in genomes from other environments. This section had a comparatively high proportion of genes in “M – cell wall/membrane/envelope biogenesis” and “P – inorganic ion transport” (Figure 3.8). The Group 2 gene clusters included a copper chaperone (*copZ*) and a copper oxidase (laccase) domain. Copper chaperones are responsible for transportation and transfer of copper (Rosenzweig and O’Halloran, 2000), and laccase is a copper oxidase that catalyzes oxidation of a broad range of compounds (Reiss *et al.*, 2013). Engineered environments typically have higher metal concentrations, which may select for metal tolerance mechanisms. Another interesting gene cluster was chromate transport protein (*chrA*). Chromate transport is an important microbial heavy metal resistance mechanism for managing chromium (Díaz-Pérez *et al.*, 2007), and may be of importance to Cloacimonadota in engineered environments. Other gene clusters, such as ABC-type Fe^{3+} transporters and divalent metal cation (Fe/Co/Zn/Cd) transporter are present in both the “Engineered” and “Environmental” sections, indicating these functions are important across the Cloacimonadota, but had diverged sufficiently to form separate gene clusters within the pangenome analysis.

Group 3 contained 202 gene clusters that were nearly universal in genomes that correspond to Clade I on the phylogenetic tree. The pangenome Group 3 gene clusters have representation from genomes from engineered ecosystems (interior rings) as well as an additional six genomes (exterior rings) (Figure 3.7), which all correspond to Clade I (Figure 3.2). The COG category that is comparatively abundant in this group was "J – translation", similar to the "core" Group 1. It is possible for functional overlap across the groups or with gene clusters outside of the defined groups, due to sequence dissimilarity that drives the generation of separate gene clusters. In the case of Group 3, no key functions were identified that were specifically associated with the genomes represented in the group – this group appears to indicate higher sequence conservation for these functions among the Clade I organisms rather than specific gains or losses of function.

The last identified grouping, Group 4, contains 1,765 gene clusters from genomes originating from environmental ecosystems, and is thus the largest identified Group. A notable category in Group 4 is "V - defense mechanisms". The "core", Group 2, and Group 3 sections each contained a small number of genes, including ABC-type antimicrobial transport systems and a multidrug efflux pump. In contrast, Group 4 contained gene clusters for components of the CRISPR/Cas system. However, CRISPR/Cas system genes were also detected in genomes originating from engineered environments and the landfill MAGs. The presence of a CRISPR/Cas system in Cloacimonadota genomes does not appear to be clearly linked to the functional or phylogenetic distribution of Cloacimonadota, but the CRISPR/Cas system does appear more tightly conserved in the environmentally derived genomes, allowing for identification of gene clusters from Clade II genomes.

The presence of the same genes across multiple gene clusters is due to sequence dissimilarity breaking gene clusters into two or more clusters with higher sequence identities. This research evaluated a phylum-level lineage, where pangenome comparisons are better targeted to species- or strain-level comparisons (Delmont and Eren, 2018). However, the high number of identified gene clusters, and the connections between ecosystem and gene complements do provide interesting insights into Cloacimonadota evolution and specialization. There were eight genomes (Figure 3.7, exterior circles) clustered differently based on gene clusters compared to their placement on the phylogenetic tree (Figure 3.2). First, the six genomes *Ca.* Cloacimonas sp. SDB, *Ca.* Cloacimonetes bacterium NIOZ-UU2, *Ca.* Cloacimonetes bacterium DOLZORAL_124_36_10, Unclassified Cloacimonetes Th196P3bin22, along with landfill MAGs LW2 Bin 38_2 and LW2 Bin 16_7, clustered together due to a unique pattern of shared gene clusters, with strong representation in Group 3 (gene clusters shared by engineered ecosystem-derived genomes), a partial presence in Group 4 (gene clusters from environmentally-derived genomes), and a third, discrete cluster of genes that is more patchily represented across the six genomes (Figure 3.7, 4 o'clock). This set of genomes contains two of the three host-associated Cloacimonadota, as well as two landfill-derived MAGs and two marine-associated MAGs, suggesting there may be a shared host-associated lifestyle for these organisms. All of these genomes were affiliated with Clade I on the phylogenetic tree. Further, the three host-associated genomes (*Ca.* Cloacimonetes bacterium DOLZORAL_124_31_13, *Ca.* Cloacimonetes bacterium DOLZORAL_124_36_10, Unclassified Cloacimonetes Th196P3bin22) along with landfill MAG LW2 Bin 16_7 and *Ca.* Cloacimonetes bacterium NORP7, were five genomes that encoded many genes involved in a flagellar system as determined through the KEGG annotation (Appendix B,

Supplementary Table B2). Flagella may be a trait of host-associated Cloacimonadota, whose gene clusters are contributing to the organization of genomes in the pangenome.

Three genomes have the lowest number of overall shared gene clusters and are not affiliated with any group. These three genomes are also the three genomes that fall outside Clade I and Clade II, placing as basal and most divergent on the phylogenetic tree (Figure 3.2). Of these, *Ca.* Cloacimonas sp. 4484_209 and *Ca.* Cloacimonetes bacterium DOLZORAL_124_31_13 (furthest exterior rings, Figure 3.7), seem to be outliers because they share "core" gene clusters but lack gene clusters from other major groupings. In contrast, *Ca.* Cloacimonetes bacterium NORP72 (middle of genome rings, Figure 3.7), does not share many gene clusters with the other genomes and contains the largest number of singleton genes. *Ca.* Cloacimonetes bacterium NORP72 also contains fewer hits across all the COG categories for Groups 1-4 compared to the other genomes (Figure 3.8). *Ca.* Cloacimonetes bacterium NORP72 was the only genome to encode certain genes, such as nitrate reductase. Furthermore, compared to other Cloacimonadota, *Ca.* Cloacimonetes bacterium NORP72 had the largest genome, the largest number of genes, and the longest branch lengths (Figure 3.2). Collectively these differences suggest *Ca.* Cloacimonetes bacterium NORP72 is either a representative of a distantly related Cloacimonadota clade (Class level difference), or possibly belongs within a different phylum.

3.4 Conclusions

Here 24 Cloacimonadota MAGs were reconstructed from a landfill site, expanding the number of Cloacimonadota genomes by 80% and nearly doubling the number of good quality genomes available. Cloacimonadota MAGs were assembled from landfill leachate wells, a

composite leachate cistern, and a biofilm sample from the landfill site, often at low proportional abundances. The MAG with the highest proportional abundance was obtained from the biofilm sample, where it accounted for 4.72% of the assembled metagenome (BF1 Bin 72). In a phylum-level comparison of phylogenetic diversity, Cloacimonadota genomes formed two distinct clades segregated by ecosystem, with one clade containing all engineered-environment derived genomes and one clade containing the majority of environmentally derived genomes. From metabolic predictions, Cloacimonadota are predicted to be anaerobic, acetogenic, and use a mixed fermentative and anaerobic respiratory lifestyle. Substrate utilization may differ across lineages; for example, carbohydrate cleavage enzymes were more diverse in genomes from engineered environments, suggesting different substrate capacities. Propionate oxidation via methylmalonyl-CoA was hypothesized to be an important function based on culture trials (Dyksma and Gallert, 2019), but the pathway was only partially encoded in most genomes suggesting this function may have been conducted by other consortia members. Amino acid fermentation of asparagine, aspartate, glutamate, and, in Clade I Cloacimonadota, cysteine to 2-oxoacids or pyruvate with subsequent oxidation by ferredoxin oxidoreductases, though not by pyruvate-formate lyase, consistent with an anaerobic lifestyle. Most Cloacimonadota genomes encode a RNF complex, which may be important in energy generation. ATPase type was phylogenetically linked, where Clade I genomes lacked complete ATPase but had a partial V/A-type ATPase, and some genomes from Clade II encoded a F-type ATPase. This indicates there may be clade specific mechanisms for energy generation. Cloacimonadota genomes from engineered environments were enriched in metal transport and resistance genes and share substantially higher sequence identities compared to their environmentally derived counterparts. Pangenome and phylogenetic analyses indicate a

shared evolutionary history for Cloacimonadota from engineered environments, where this lineage is frequently identified.

Chapter 4
Distribution of reactive oxygen species defense mechanisms across domain Bacteria

A version of this chapter has been published as:

Johnson, L.A. and Hug, L.A. (2019) Distribution of reactive oxygen species defense mechanisms across domain Bacteria. *Free Radical Biology & Medicine* **140**: 93–102. https://doi.org/10.1016/j.freeradbiomed.2019.03.032

4.1 Introduction

Microorganisms have enzymatic defenses against reactive oxygen species (ROS) which are important for microbial survival, especially in environments, such as landfills, where anoxic and oxic microenvironments exist. In this chapter, metagenomics was used as a tool to survey diverse bacterial lineages for the presence of genes involved in ROS defense. ROS are formed as intermediates of O_2 reduction, and include the superoxide anion (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radical (OH^-). These ROS can cause damage to cellular macromolecules (DNA, RNA, proteins, and lipids), leading to membrane damage, loss of protein function, blocking of DNA replication, and mutations (Cabiscol *et al.*, 2000). ROS represent a significant pervasive threat to microbial life, and bacteria have evolved several mechanisms to mitigate this threat, such as scavenging enzymes (Inupakutika *et al.*, 2016), transcription factors (Imlay, 2015), and repair systems (Arts *et al.*, 2015). This research focuses on the diversity of ROS scavenging and defense enzymes and their prevalence across the bacterial domain.

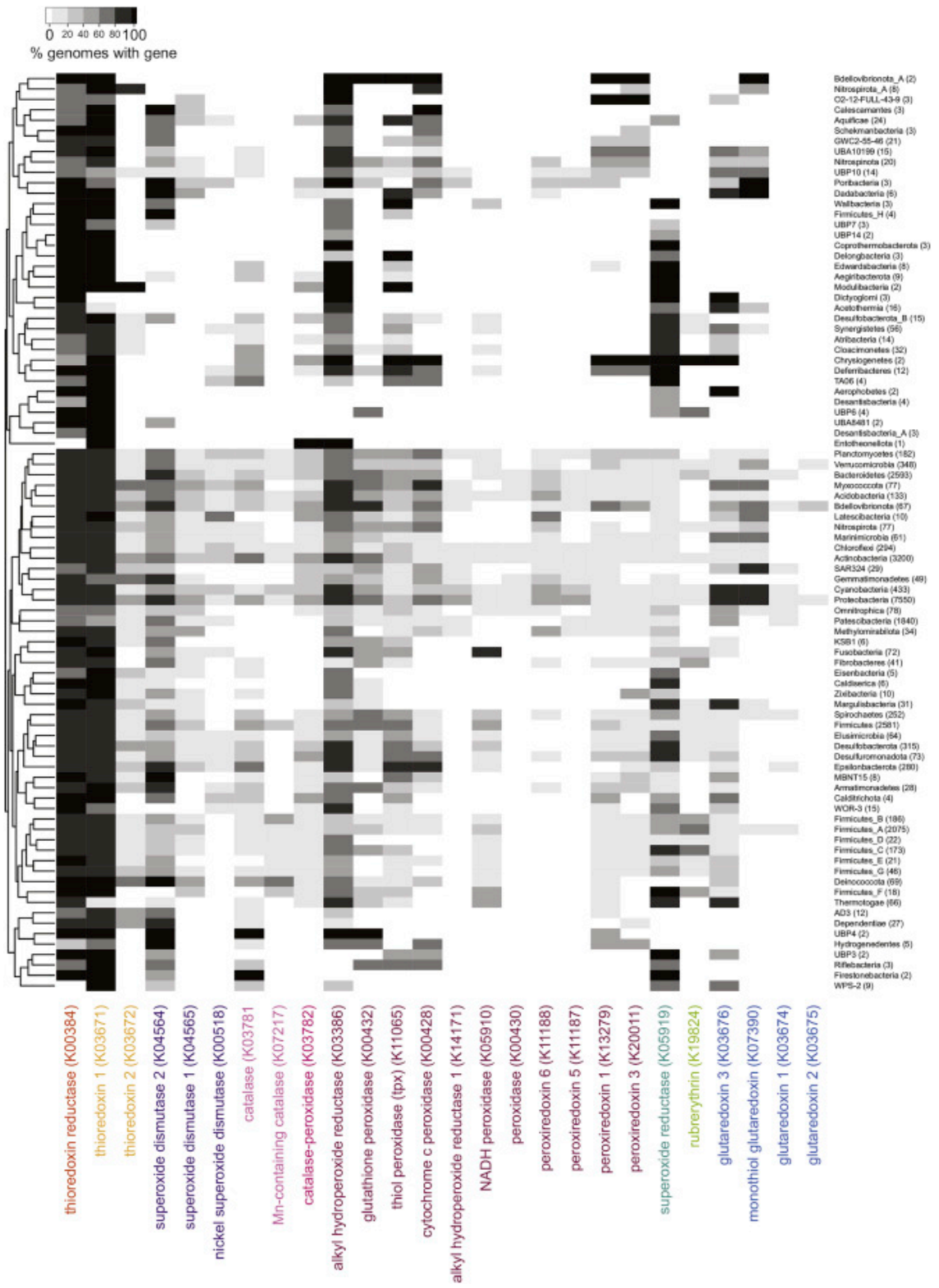
Life on Earth began under anoxic conditions. With the advent of oxygenic photosynthesis in the Cyanobacteria, Earth's atmosphere underwent a geologically rapid shift to oxidizing

conditions across all surface environments (Brocks *et al.*, 1999). As oxygen accumulated, some bacteria evolved to use oxygen as their obligate electron acceptor for high energy growth (aerobes), others evolved flexible metabolisms allowing growth in the presence or absence of oxygen (facultative anaerobes), and a third class retreated to subsurface environments where oxygen is rarely, if ever, present (obligate anaerobes). In this research, I explore whether these metabolic divisions within the Bacteria are mirrored in their ROS defense mechanisms, and whether hallmark defenses for each bacterial lifestyle can be identified.

4.2 Distribution of ROS-active enzymes across the Bacteria

The recent expansion of available genomic data presented an opportunity to assess the universality of ROS defenses across the bacterial Domain. This work summarized the distribution of 26 known ROS defense enzymes, assessing their presence and prevalence across bacterial phyla (Figure 4.1). A total of 23,913 bacterial genomes representing 88 different phyla were included. Identification of ROS-active enzymes was conducted using the AnnoTree tool, which maps genome annotation data on the bacterial tree of life (Mendler *et al.*, 2019). The reported ROS defenses are based largely on genome annotations using the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification system. Because these are not *in situ* or *in vitro* observations of

Figure 4.1 ROS scavenging enzymes' distribution across the domain Bacteria. Heatmap depicts prevalence of a given ROS enzyme (columns) across bacterial phyla (rows), where black indicates 100% of sequenced organisms in that lineage contain the gene of interest, white indicates zero genomes per clade, and the greys are 20% increments in a gradient from light (1–20%) to dark (80–99%). Enzyme names are colored by function as in Figure 4.2. Columns are ordered by type of ROS defense (*e.g.*, superoxide dismutases), with the members ranked by prevalence. Phyla were hierarchically clustered using the Canberra distance method and the "complete" clustering approach as implemented in the R package "heatmap2". Gene distributions were determined from 23,913 bacterial genomes in the Genome Taxonomy Database (Parks *et al.*, 2018) using the AnnoTree software tool (Mendler *et al.*, 2019).


0 20 40 60 80 100
% genomes with gene
Bdellovibrionota_A (2)
Nitrospirota_A (8)
C0-12-FULL-43-9 (3)
Calescamantes (3)
Aquificae (24)
Schekmanbacteria (3)
GWC2-55-46 (21)
UBA10199 (15)
Nitrospinota (20)
UBP10 (14)
Poribacteria (3)
Dadabacteria (6)
Wallbacteria (3)
Firmicutes_H (4)
UBP7 (3)
UBP14 (2)
Coprothermobacterota (3)
Delongbacteria (3)
Edwardsbacteria (8)
Aegiribacterota (9)
Modulibacteria (2)
Dictyoglomi (3)
Acetothermia (16)
Desulfobacterota_B (15)
Synergistetes (56)
Atribacteria (14)
Cloacimonadota (32)
Chrysiogenetes (2)
Deferribacteres (12)
TA06 (4)
Aerophobetes (2)
Desantisbacteria (4)
UBP6 (4)
UBA5461 (2)
Desantisbacteria_A (3)
Entotheonellota (1)
Planctomycetota (182)
Verrucomicrobiota (348)
Bacteroidetes (2593)
Myxococcota (77)
Acidobacteriota (133)
Bdellovibrionota (67)
Latescibacteria (10)
Nitrospirota (77)
Marinimicrobia (61)
Chloroflexi (294)
Actinobacteriota (3200)
SAR324 (29)
Gemmatimonadetes (49)
Cyanobacteria (433)
Proteobacteria (7550)
Omnitrophica (78)
Patescibacteria (1840)
Methylomirabilota (34)
KSB1 (6)
Fusobacteriota (72)
Fibrobacteres (41)
Eisenbacteria (5)
Caldisericota (6)
Zixibacteria (10)
Margulisbacteria (31)
Spirochaetes (252)
Firmicutes (2581)
Elusimicrobia (64)
Desulfobacterota (315)
Desulfuromonadota (73)
Epsilonbacterota (280)
MBNT15 (8)
Armatimonadetes (28)
Caldithrichota (4)
WOR-3 (15)
Firmicutes_B (186)
Firmicutes_A (2075)
Firmicutes_D (22)
Firmicutes_C (173)
Firmicutes_E (21)
Firmicutes_G (46)
Deinococcota (69)
Firmicutes_F (18)
Thermotogae (66)
AD3 (12)
Dependentiae (27)
UBP4 (2)
Hydrogenedentes (5)
UBP3 (2)
Riflebacteria (3)
Firestonebacteria (2)
WPS-2 (9)
thioredoxin reductase (K00384)
thioredoxin 1 (K03671)
thioredoxin 2 (K03672)
superoxide dismutase 2 (K04564)
superoxide dismutase 1 (K04565)
nickel superoxide dismutase (K00518)
catalase (K03781
Mn-containing catalase (K07217)
catalase-peroxidase (K03782)
alkyl hydroperoxide reductase (K03386)
glutathione peroxidase (K00432)
thiol peroxidase (tpx) (K11065)
cytochrome c peroxidase (K00428)
alkyl hydroperoxide reductase 1 (K14171)
NADH peroxidase (K05910)
peroxidase (K00430)
peroxiredoxin 6 (K11188)
peroxiredoxin 5 (K11187)
peroxiredoxin 1 (K13279)
peroxiredoxin 3 (K20011)
superoxide reductase (K05919)
rubrerythrin (K19824)
glutaredoxin 3 (K03676)
monothiol glutaredoxin (K07390)
glutaredoxin 1 (K03674)
glutaredoxin 2 (K03675)

activity, this summary should therefore be taken as a prediction of functional potential rather than confirmed activities. This caution is especially important for novel lineages, including candidate phyla, which may have more divergent proteins, resulting in less confidence in automatic annotations. For visualization of the distribution of ROS defenses, the taxonomy from the Genome Taxonomy Database was adopted, as the tree from the GTDB study underlies the AnnoTree interface (Parks *et al.*, 2018). AnnoTree allows the user to select Pfam or KEGG-based annotations; I chose to use the KEGG database annotations as they allowed greater resolution of proteins compared to Pfam (Ogata *et al.*, 1999; Finn *et al.*, 2016). KEGG annotations in AnnoTree were derived from a similarity search against all UniRef100 clusters with a KO identifier, with hits assigned to a KO based on an e-value cutoff of $\leq 1e^{-5}$, percent identity ≥30%, and subject-query percent alignment ≥70%.

The distribution of ROS defense mechanisms across the bacterial domain varies considerably (Figure 4.1; Appendix C, Supplementary Figure C1). The most prevalent enzymes are thioredoxin 1 and thioredoxin reductase, which play roles in intracellular thiol/disulphide balance and in the repair of oxidized proteins through catalytic reduction of disulfide bridges or sulfenic acids (Vido *et al.*, 2005). Thioredoxins can directly reduce hydrogen peroxide (Spector *et al.*, 1988; Kang *et al.*, 1998), donate hydrogen to peroxidases (Chae *et al.*, 1994), and scavenge hydroxyl radicals (Das and Das, 2000). Both thioredoxin 1 (K03671) and thioredoxin reductase (K00384) were predicted from 93% of the ~24K genomes. Thioredoxin 1 is present in all but one bacterial phylum (Dictyoglomi), while thioredoxin reductase is present in all but six phyla (Entotheonellota and five unnamed candidate phyla), though it is important to note that absences from incomplete draft genomes may not be true absences of these genes in these organisms.

Thioredoxin and thioredoxin reductase have been implicated as important within the oxidative stress response (Zeller and Klug, 2006) – their near-universality across the bacterial domain strengthens this importance. However, thioredoxins and thioredoxin reductases have additional functions beyond scavenging ROS, such as, balancing cellular redox potentials, that may contribute to their near universality in bacterial genomes (Valette *et al*., 2017). At the other extreme, glutaredoxin 2 (K03675) is found in only three phyla – the Proteobacteria, Bdellovibrionata, and the Verrucomicrobia, and was predicted from only 3.7% of bacterial genomes overall. Glutaredoxins are electron carriers that reduce glutathione as part of the glutathione system, and act as antioxidants by reducing metals and peroxides (Deponte, 2013). There are four families of glutaredoxins in the KEGG database: glutaredoxin 1 (K03674), glutaredoxin 2 (K03675), glutaredoxin 3 (K03676), and monothiol glutaredoxin (K07390). Glutaredoxin 1, 3, and monothiol glutaredoxin are more prevalent than glutaredoxin 2, predicted in 6.0, 36.5 and 33.7% of bacterial genomes respectively. The other 19 ROS-active enzyme families generate a distribution from highly prevalent to largely absent (Figure 4.1).

Superoxide dismutase, catalase, and peroxidase are three of the best characterized ROS defense mechanisms, so it was surprising to observe their distributions across the bacterial domain are inconsistent and less abundant than several other ROS-active enzymes. Superoxide dismutase (SOD) catalyzes the dismutation of superoxides into hydrogen peroxide and oxygen ($2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$) (McCord and Fridovich, 1969). SOD enzymes are classified by their metal cofactors, where the SOD 1 family (K04565) contains Cu or Zn, SOD 2 (K04564) contain Fe or Mn, and the Ni-dependent SODs form a third family (*sodN*, K00518) (Fridovich, 1975). Of the three SOD families, SOD 2 is most abundant within the bacterial domain, with 72.9% of bacterial

genomes predicted to encode this gene. SOD 1 is a distant second (29.2%) and nickel SOD the least common (7.9%). In total, 17 of the 88 bacterial phyla do not have any members with a SOD, of which eight are candidate phyla (Figure 4.1). The SOD activity produces hydrogen peroxide, a source of highly toxic hydroxyl radicals (Fridovich, 1975). Hydrogen peroxide is typically removed by catalase or peroxidase (Chapman *et al.*, 1983). Catalase converts hydrogen peroxide to oxygen and water ($2H_2O_2 \rightarrow 2H_2O + O_2$). Catalases are present in two forms in the KEGG database: catalase (K03781) and Mn-containing catalase (K07217). Neither form of catalase is particularly abundant within bacterial genomes, with predicted prevalence's of 41.4% and 13.1%, respectively, and both are predominantly found in the well-sampled Proteobacteria and Firmicutes. The results who that 35 of 88 bacterial phyla had no catalase genes identified, and only two poorly sampled phyla, the candidate phyla UBP4 and Firestonebacteria, have universally present catalase genes. Catalase-peroxidases (K03782) contain heme cofactors, and are present in a variety of bacterial, archaeal, and fungal genomes (Nicholls *et al.*, 2000). They have substantial peroxidatic activities, and function as H_2O_2 scavengers (Nicholls *et al.*, 2000). Catalase-peroxidase (K03782) was predicted in 33.4% of genomes, and 26 of the 35 phyla lacking catalase genes also lack any members with a catalase-peroxidase.

Beyond catalase-peroxidase, I summarized predicted gene presence/absence for seven additional peroxidase types: peroxidase (K00430), thiol peroxidase (K11065), cytochrome *c* peroxidase (K00428), glutathione peroxidase (K00432), NADH peroxidase (K05910), alkyl hydroperoxide reductases (ahp1: K14171, ahpC: K03386), and peroxiredoxins 1 (K13279), 3 (K20011), 5 (K11187) and 6 (K11188). Thiol peroxidase (K11065, Tpx, also scavengase p20) is a thioredoxin-linked peroxidase found in all three domains of life (Van Vliet *et al.*, 2002). Tpx

protects enzymes like glutamine synthase against inactivation by oxidative stress (Comtois *et al.*, 2003). Catalase is solely cytoplasmic, but H_2O_2 can also be produced in the periplasm, where periplasmic cytochromes require protection against peroxidation (Zapata *et al.*, 2017). Periplasmic cytochrome *c* peroxidases mediate removal of this periplasmic H_2O_2. Glutathione peroxidases are selenoprotein oxoreductases found in plants, animals, protists, and diverse bacterial species (Arthur, 2000; Brenot *et al.*, 2004; Arenas *et al.*, 2011; Mishra and Imlay, 2012). They mediate the glutathione-dependent detoxification of H_2O_2, play a role in stress resistance, and, in bacteria, contribute to virulence (Arthur, 2000; Brenot *et al.*, 2004; Arenas *et al.*, 2011; Mishra and Imlay, 2012). NADH peroxidases inactivate endogenous H_2O_2, but may also protect against exogenous H_2O_2 (Gordon *et al.*, 1953). Alkyl hydroperoxide reductases convert reactive hydroperoxides to their corresponding alcohols ($ROOH + NADH + H^+ \rightarrow ROH + NAD^+ + H_2O$) (Van Vliet *et al.*, 2002). Alkyl hydroperoxide reductase C was the most abundant peroxidase, predicted from 72.6% of all genomes, followed by glutathione peroxidase (54.5%), thiol peroxidase (34.9%), cytochrome *c* peroxidase (33.8%), alkyl hydroperoxide reductase 1 (18.6%), NADH peroxidases (11.1%), and finally peroxidase (3.3%). A 2007 survey of the TIGR database determined that NADH peroxidases were restricted to a few bacterial genera, including *Streptococcus*, *Lactobacillus*, and *Listeria* (La Carbona *et al.*, 2007). This updated survey identified putative NADH peroxidases in 30 bacterial phyla, including the Chloroflexi and Actinobacteria.

Peroxiredoxins are cysteine-dependent peroxidases active in regulating peroxide levels within the cell. The thioredoxin system provides reducing equivalents to peroxiredoxins to enable reduction of H_2O_2 (Kang *et al.*, 1998; Zeller and Klug, 2006). Considered critical defense and regulatory enzymes, Perkins and colleagues describe peroxiredoxins as a ubiquitous family of

enzymes with complex patterns of regulation, including post-translational modifications and oligomeric states (Perkins *et al.*, 2015). Far from ubiquitous, peroxiredoxins were never predicted at more than 20% abundance across the Bacteria (peroxiredoxin 6: 18.7%; peroxiredoxin 5: 18.5%, peroxiredoxin 1: 6.7%; peroxiredoxin 3: 4.8%). Putative peroxiredoxins were identified in 56 of 88 phyla. Four phyla have no representatives encoding a single peroxidase, including catalase-peroxidase (Acetothermia, Dictyoglomi, and two candidate phyla). Of these, the two named phyla completely lack all of the typical textbook ROS defense enzymes (SOD, catalase, peroxidase). The recognition of entire phyla lacking these three enzyme types, including phyla with closed genome representatives, suggests there may be a significant number of uncultured microbes that do not use SOD, catalase, or peroxidase.

Superoxide reductase (SOR, K05919) is a redox-active metalloenzyme similar in nature to the superoxide dismutases, from an independent evolutionary origin (Sheng *et al.*, 2014). SORs catalyze a reduction reaction in contrast to the disproportionation reaction catalyzed by the three SODs, where both reactions ultimately convert superoxide to H_2O_2. Characterized SORs universally contain an iron co-factor, compared to the diversity of metal cofactors found in the SOD enzymes (Sheng *et al.*, 2014). Occasionally described as an enzyme found only in anaerobic lineages (*e.g.*, (Kitagawa *et al.*, 2006)), SORs have been identified from aerobic and anaerobic lineages. In this research, predicted superoxide reductases were widely distributed but not particularly abundant in the bacterial domain, present in 76% of phyla but only 13.6% of genomes. This discrepancy was partly driven by a low prevalence in the Proteobacteria, which are the best sampled phylum by a wide margin. The distribution of superoxide reductase was patchy in most phyla, with a median prevalence of 50% for the phyla in which it was found.

The final ROS enzyme that was profiled is rubrerythrin (K19824), a non-heme di-iron protein involved in the oxidative stress response as a peroxide scavenger (Andrews, 2010). The majority of rubrerythrins contain a rubredoxin domain that facilitates H_2O_2 reduction in anoxic conditions, but there is a second, distantly related rubrerythrin type that functions in aerobes (Cardenas *et al.*, 2016). The aerobic rubrerythrin type is hypothesized to have originated in thermophilic archaea and later transferred to aerobic bacteria (Cardenas *et al.*, 2016). Rubrerythrin was predicted in 25 out of 88 bacterial phyla, and in only 10.1% of genomes.

From a phylum perspective, the Proteobacteria has the highest proportion of genomes encoding the largest number of ROS defense enzymes. There were only three phyla with consistent distributions of enzymes within their members: the Dictyoglomi, Coprothermobacterota, Bdellovibrionota_A, and Entotheonellota (previously the Tectomicrobia). These phyla are represented by only three, three, two, and a single genome, respectively, and so further sampling of these lineages is needed to assess whether the current pattern of universal presence and absence for ROS active enzymes will prove characteristic for these lineages. All other bacterial phyla have variation in the distribution of ROS defenses within their members. All of the better sampled phyla show patchy distributions of enzymes (Figure 4.1, bottom clade). As an example, the Chloroflexi phylum's distribution of ROS enzymes ranges from 0 to 89% of genomes encoding various ROS-active enzymes, with no enzyme present in 100% of all Chloroflexi genomes surveyed. For phyla with more than 10 representative genome sequences (those less likely to be strongly affected by missing data from partial genomes), only superoxide dismutase 2, thioredoxin 1, and superoxide reductase are present at 100% in any lineage, and only in one, ten, and two phyla, respectively (Figure 4.1).

Hierarchical clustering grouped the bacterial phyla into two main clusters (Figure 4.1). While there are no ROS enzymes that neatly delineate the groups, there is a substantial difference in the prevalence of ROS enzymes. One cluster has a punctuated pattern of presence/absence, where ROS enzymes are frequently present at 100% or entirely absent for the phyla. In contrast, the other cluster is distinguished by less absolute presence/absence patterns, where most ROS enzymes were detected in a portion of the genomes from those phyla. This difference is, unfortunately, likely driven by sampling bias rather than an underlying biological principle: the first clade (Figure 4.1, upper clade) is largely composed of undersampled phyla, with 32 of 36 phyla in the clade represented by 20 or fewer genomes. The Synergistetes is the best represented phylum within the first cluster, with 56 genomes. In contrast, only 16 of 52 phyla in the second cluster had fewer than 20 genome representatives, and all highly sampled lineages were included in this cluster (*e.g*, Proteobacteria (7,550 genomes), Actinobacteria (3,200 genomes), and Bacteroidetes (2,593 genomes)). It is anticipated that as additional genomes for the undersampled phyla are sequenced, the punctuated presence/absence pattern of ROS defenses will likely adjust to a gradient similar to the more deeply sampled phyla.

This compilation was generated in the hopes of identifying gene complements that would allow prediction of anaerobic and aerobic lifestyles from genomic information, and to identify shared or unique patterns of ROS-active enzymes within and between bacterial phyla. There was also interest in whether any of the known ROS defenses were necessary or sufficient to determine a microorganism's lifestyle – aerobic, facultative anaerobic, or obligately anaerobic. No strong trends were apparent from this compilation, including for candidate phyla and other radiations that have been exclusively identified from anoxic environments, even when the phyla were clustered

based on phylogenetic relatedness (Appendix C, Supplementary Figure C1). This may also be impacted by the fact that phyla contain a wide range of classes, families and genera that contain microorganisms with varying lifestyles. Instead, what emerges from this view of bacterial genomic capacity for ROS defense is an appreciation of the many different strategies employed, including a realization that the "canonical" defenses of superoxide dismutase, catalase, and peroxidase are not widespread outside of commonly cultivated phyla.

4.3 Bacterial lifestyles and ROS defenses

In the absence of distinct gene complements to guide a discussion of bacterial lifestyles and the ROS defenses employed under oxic and anoxic conditions, I instead present a series of vignettes from well-characterized bacterial groups to highlight the diversity of microbial responses to oxygen stress, anchored by the environmental conditions each microorganism can tolerate (Table 4.1).

4.3.1 Aerobic with defenses

4.3.1.1 Cyanobacteria: oxygen progenitors

The Cyanobacteria are a phylum of widely distributed, photosynthetic, aerobic or facultatively anaerobic bacteria. In terms of oxidative stress, Cyanobacteria are unique from other lineages of bacteria as they are exposed to increased ROS generated from the photosynthetic electron transport chain (Latifi *et al.*, 2009). Cyanobacteria must mitigate these ROS generated during photosynthesis in addition to the background levels of ROS that heterotrophs generate via respiration. As a result of this increased stress, Cyanobacteria have multiple defense mechanisms

Table 4.1 Example bacterial lifestyles and the ROS defenses maintained by selected well-characterized organisms. Organisms in bold are depicted in Figure 4.2. There are no known organisms, aerobic or anaerobic, that completely lack ROS defense mechanisms.

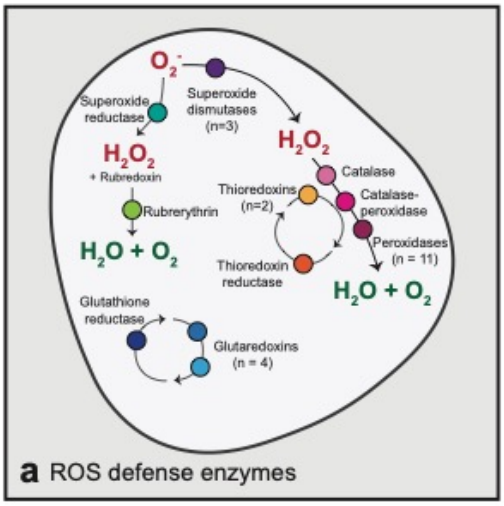
Lifestyle	Lineage	ROS defenses	Notes
Aerobe	**Cyanobacteria**	SOD Thioredoxin Glutaredoxin Monothiol glutaredoxin Peroxiredoxin	Exposed to additional oxidative stress from photosynthesis, mitigated by energy dissipation and alternative electron transport pathways
	Mycobacterium	SOD Catalase Thioredoxin Alkyl-hydroperoxidase reductase Peroxiredoxin	ROS defense mechanisms are drug targets *Mycobacterium*-specific cell wall constituents provide additional protection
Facultative anaerobe	*Campylobacter*	SOD Catalase Cytochrome *c* peroxidase Alkyl hydroperoxide reductase	Must survive transitional periods in oxic conditions between hosts
	Lactobacillus	SOD Catalase Thioredoxin NADH-peroxidase	Can be microaerophilic or facultatively anaerobic, some are able to switch to an aerobic lifestyle

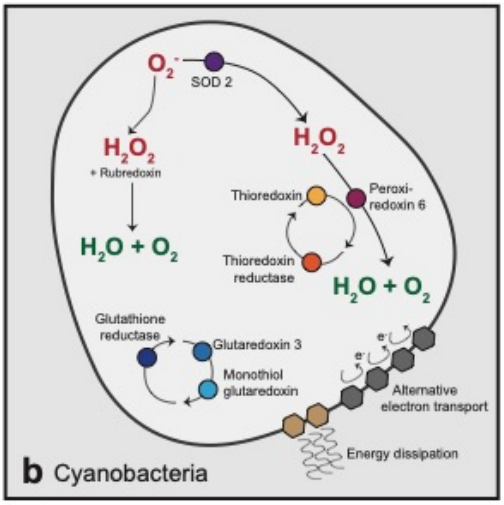
Table 4.1, cont.

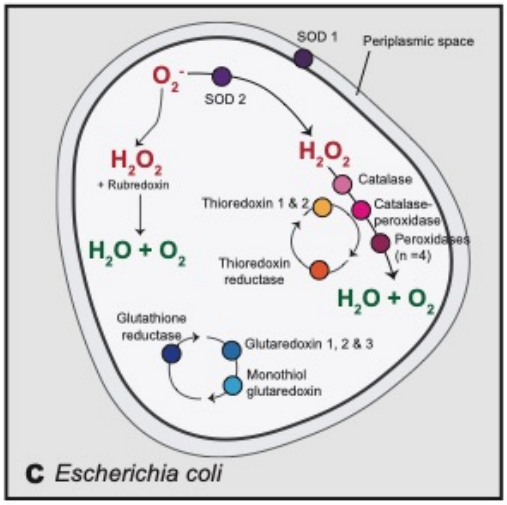
Lifestyle	Lineage	ROS defenses	Notes
	Desulfovibrio	SOD Catalase Superoxide reductase Thioredoxin (Trx) Thioredoxin reductase NAD(P)H-dependent peroxidases	Sulfate reducing bacteria often found in sediments, subject to shifts in oxygen concentrations
Obligate anaerobe	***Dehalococcoides***	SOD Superoxide reductase Thioredoxin Thioredoxin reductase Alkylhydroperoxide reductase	Oxygen exposure disrupts the activity of essential metabolic proteins
	Minimum genome (*Mycoplasma*)	Thioredoxin-disulfide reductase	A theorized minimal genome did not include any ROS defenses Thioredoxin-disulfide reductase was required for robust growth *in vivo*

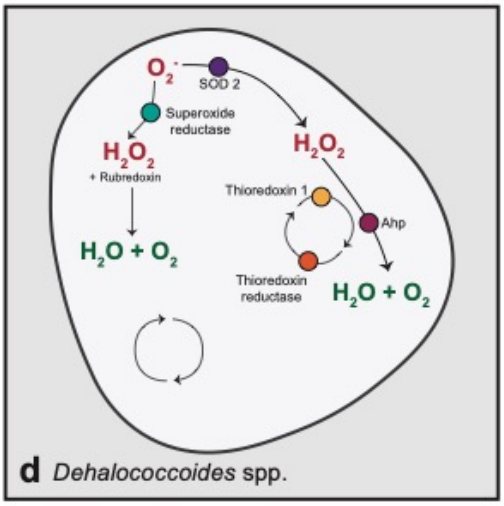
to manage oxidative stress, including detoxifying enzymes, alternative electron transport pathways, energy dissipation, and non-enzymatic antioxidants (Figure 4.2b) (Bukhov and Carpentier, 2004; Wilson *et al.*, 2007; Kelman *et al.*, 2009). Superoxide anions and H_2O_2 produced by the photosynthetic electron transport chain are mainly scavenged by cellular enzymes, including superoxide dismutases and catalases (such as glutathione peroxidase, peroxiredoxins, and rubrerythrin). Additional photosynthesis-related defense mechanisms include energy dissipation, which reduces the efficiency of photosynthesis and thus prevents ROS production (Latifi *et al.*, 2009). Energy dissipation is mediated by Cyanobacteria-specific proteins, including orange carotenoid protein (OCP) (Kirilovsky, 2007), iron stress induced protein (Havaux *et al.*, 2005), and high light-inducible proteins (or small CAB-like proteins) (He *et al.*, 2001; Yao *et al.*, 2007). Along with enzymatic scavenging and energy dissipation, Cyanobacteria are protected by alternative electron transport pathways in the photosynthetic electron transport chain. These pathways help protect against oxidative stress through removal of excess electrons by cytochrome oxidases and plastoquinol terminal oxidase (PTOX) (Schubert *et al.*, 1995; Bailey *et al.*, 2008). A metagenomic study revealed PTOX gene distributions in marine prokaryotes. There

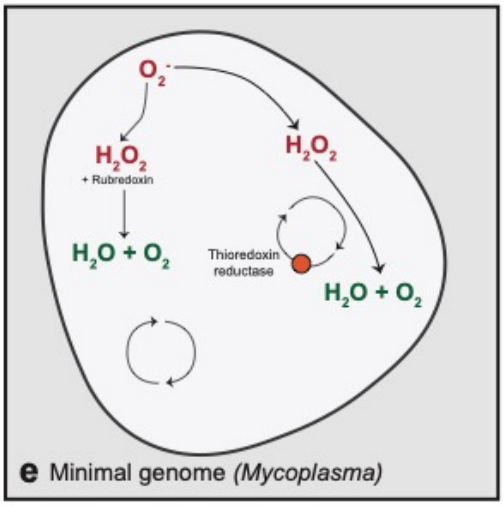
Figure 4.2 Diagrammatic comparison of ROS defense complements in the highlighted bacterial lineages and constructs. (a) Representative cell depicting all ROS defense enzymes described in this review, (b) Cyanobacteria, (c) *Escherichia coli*, (d) *Dehalococcoides* spp., and (e) a minimal synthetic genome derived from *Mycoplasma*. See Fig. 1 for list of each enzyme examined (*e.g.*, SOD1, SOD2, and Ni-SOD for the three superoxide dismutases). The four peroxidases in *E. coli* are thiol peroxidase, cytochrome *c* peroxidase, glutathione peroxidase, and alkylhydroperoxide reductase.


O_2^-
Superoxide reductase
Superoxide dismutases (n=3)
H_2O_2
+ Rubredoxin
Rubrerythrin
$H_2O + O_2$
H_2O_2
Catalase
Thioredoxins (n=2)
Catalase-peroxidase
Peroxidases (n = 11)
Thioredoxin reductase
$H_2O + O_2$
Glutathione reductase
Glutaredoxins (n = 4)
a ROS defense enzymes


O_2^-
SOD 2
H_2O_2
+ Rubredoxin
$H_2O + O_2$
H_2O_2
Thioredoxin
Peroxi-redoxin 6
Thioredoxin reductase
$H_2O + O_2$
Glutathione reductase
Glutaredoxin 3
Monothiol glutaredoxin
e^-
e^-
Alternative electron transport
Energy dissipation
b Cyanobacteria


SOD 1
Periplasmic space
O_2^-
SOD 2
H_2O_2
+ Rubredoxin
$H_2O + O_2$
H_2O_2
Catalase
Thioredoxin 1 & 2
Catalase-peroxidase
Peroxidases (n =4)
Thioredoxin reductase
$H_2O + O_2$
Glutathione reductase
Glutaredoxin 1, 2 & 3
Monothiol glutaredoxin
c Escherichia coli


O_2^-
SOD 2
Superoxide reductase
H_2O_2
+ Rubredoxin
$H_2O + O_2$
H_2O_2
Thioredoxin 1
Ahp
Thioredoxin reductase
$H_2O + O_2$
d Dehalococcoides spp.


O_2^-
H_2O_2
+ Rubredoxin
$H_2O + O_2$
H_2O_2
Thioredoxin reductase
$H_2O + O_2$
e Minimal genome (Mycoplasma)

was no evidence of PTOX genes in any prokaryotes other than cyanobacteria, leaving the currently understood distribution of PTOX to be exclusively in photosynthetic organisms (McDonald and Vanlerberghe, 2005).

4.3.1.2 *Mycobacteria*: ROS defenses impact drug susceptibility

Mycobacterium is an aerobic genus from the phylum Actinobacteria that contains the human pathogen *Mycobacterium tuberculosis*. *M. tuberculosis* infects the lungs and represents a significant and persistent infectious disease in humans (Russell *et al.*, 2010), while other *Mycobacteria* live as saprophytes in the environment (Schulze-Röbbecke, 1993). Pathogenic *Mycobacteria* can also be found in environmental reservoirs (King *et al.*, 2017). Pathogenic *Mycobacteria* generate ROS via aerobic metabolism and are also exposed to oxygen and ROS that are generated by their hosts, so must deal with oxygen stress from both sources. *M. tuberculosis* uses antioxidant enzymes SOD (Piddington *et al.*, 2001) and catalase (KatG) (Ng *et al.*, 2004), thioredoxin, alkyl-hydroperoxide reductase, and peroxiredoxin. A conventional redox couple using glutathione is absent in *Mycobacteria*. Intracellular redox homeostasis is maintained with alternate redox couples, including thioredoxin, mycothiol, and ergothionine. Mycothiol and ergothionine are mycobacterial redox buffers and provide a barrier of protection against oxidative stress (Patel and Blanchard, 1999; Paul and Snyder, 2010). Mycothiol is a thiol produced by many Actinomycetes, including *Mycobacteria,* that functions as a redox couple similar to glutathione (Kumar *et al.*, 2011).

Mycobacterial ROS defenses are targets for infection control. Mycobacterial catalase (KatG) is required for activation of the drug isoniazid, an antibiotic used to treat tuberculosis. In the on-going arms race between pathogen and medicine, mutations in the *katG* gene result in

resistance to the drug (Zhang *et al.*, 1992; Barry *et al.*, 1998; Ramaswamy *et al.*, 2003; Cade *et al.*, 2010). Exposure to superoxide also increases susceptibility of *M. tuberculosis* to isoniazid (Bulatovic *et al.*, 2002). Beyond drug activation, mutations in the *katG* gene and reduced levels of alkyl-hydroperoxide reductase subunit C affect virulence and fitness of the *M. tuberculosis* (Chen *et al.*, 1998; Nieto *et al.*, 2016). Research has demonstrated that cytoplasmic redox cycling in *M. tuberculosis* is sensitive to endogenous ROS levels and may be a future target for controlling tuberculosis infection (Tyagi *et al.*, 2015).

4.3.2 Aerobic without defenses

There are no described aerobic organisms completely lacking ROS defense mechanisms. While this is not surprising, the obligate requirement for multiple ROS defenses underlines the cost of inhabiting oxic environments. Maintenance of an effective ROS defense system, including damage repair, must be balanced by the energetic potential of oxygen as a terminal electron acceptor. ROS-scavenging enzymes are thought to have evolved 4.1-3.5 billion years ago, well before the Great Oxidation Event (Inupakutika *et al.*, 2016). These defenses likely evolved in response to ROS formed by anaerobic respiratory processes (Inupakutika *et al.*, 2016).

4.3.3 Anaerobic lifestyles with defenses

4.3.3.1 Facultatively anaerobic lifestyles

The degree of aerotolerance in anaerobes varies greatly, from strict anaerobes to those which retain viability after exposure to oxic conditions. Anaerobes do not generate ROS through respiration but do require protection from molecules formed in other metabolic processes and, depending on the environment, protection from exposure to different levels of oxygen. As discussed above, the distribution of ROS defense systems varies across and within bacterial phyla,

and there were no specific defenses identified that are predictive of taxonomy, aerotolerance, or lifestyle.

4.3.3.1.1 Proteobacteria

The Proteobacteria are the best sampled phylum, and also the phylum with the strongest ROS defense signature from the above compilation. The Proteobacteria are a group of Gram-negative bacteria with diverse metabolisms and lifestyles. Many Proteobacteria are facultative anaerobes – I have profiled three important genera as examples of this lifestyle below.

4.3.3.1.1.1 *Campylobacter* – survival through infection stages

Campylobacter are Epsilonproteobacterial members of the intestinal flora of domestic and wild animals. Contamination of meat during the slaughter process can lead to human infection and food-borne illness (Van Vliet and Ketley, 2001). *Campylobacter* are pathogenic in the anoxic environment of the digestive tract, and require methods to neutralize ROS encountered during the initial, aerobic, stages of infection. Survival during exposure to oxygen is vital to the persistence of these anaerobic organisms in the long term. *Campylobacter* encodes Fe-dependent SOD (SodB, SOD1) which is located in the cytoplasm. *C. jejuni* and *C. coli* express SodB, which dismutates superoxide (Pesci *et al.*, 1994), as well as a catalase, KatA. Another important enzyme is cytochrome *c* peroxidase, which provides protection from H_2O_2 in the periplasm (Van Vliet *et al.*, 2002). Oxidative stress was determined to affect survival and morphological changes of *C. jejuni* under aerobic conditions through examination of *katA,* alkyl hydroperoxide reductase (*ahpC*) and *sodB* mutants (Oh *et al.*, 2015).

4.3.3.1.1.2 *Desulfovibrio* – free-living, sediment associated

Desulfovibrio are sulfate-reducing bacteria that are often found in marine and freshwater sediments and can be abundant in both high-organic material environments as well as in extremely oligotrophic habitats. Originally thought to be obligately anaerobic, *Desulfovibrio* have been shown to survive and grow under oxic conditions (Sigalevich *et al.*, 2000). *Desulfovibrio* genomes encode SOD enzymes and superoxide reductase (SOR)-like activity, allowing for efficient superoxide scavenging. Iron-dependent SOD have been characterized in *D. gigas* (Dos Santos *et al.*, 2000). SOR activity is encoded by desulfoferrodoxin and neelaredoxin genes (Dolla *et al.*, 2006). Catalase and NAD(P)H-dependent peroxidases (such as rubrerythrin) are used for hydrogen peroxide elimination (Dolla *et al.*, 2006). In *D. vulgaris*, thioredoxin reductase reduces thioredoxin, which is important in regulation of the redox environment and protecting cells against superoxide (Valette *et al.*, 2017).

4.3.3.1.1.3 *Escherichia coli* – a model system's ROS defenses

Escherichia coli is a member of the Gammaproteobacteria and is a model organism of study in the field of microbiology. The first purification of bacterial SOD was from *E. coli* (Keele *et al.*, 1970), and an *E. coli* genome was one of the first bacterial genomes sequenced (Blattner *et al.*, 1997). *E. coli* is a facultative anaerobe often found in the vertebrate gut. Many strains are pathogenic, and *E. coli* is routinely used as a biomarker for fecal contamination in environmental samples, where it is found in the interim between host infections.

E. coli contains three isoforms of SOD: manganese SOD and iron SOD (SOD2 forms) and copper/zinc SOD (SOD1 form). The manganese SOD and iron SOD reside in the cytoplasm, whereas the copper/zinc SOD is in the periplasm of the cell (Niederhoffer *et al.*, 1990; Compan

and Touati, 1993; Benov and Fridovich, 1994; Battistoni and Rotilio, 1995; Korshunov and Imlay, 2006). *E. coli* is also equipped with two catalases, KatE and KatG (Loewen and Switala, 1986; Farr *et al.*, 1988; Triggs-Raine *et al.*, 1988; Von Ossowski *et al.*, 1991), and an alkyl hydroperoxide reductase (Ahp) (Seaver and Imlay, 2001) (Figure 2c). These enzymes are vital for scavenging endogenous ROS, but any change in environmental conditions requires additional oxygen stress responses. This includes a transcription factor, OxyR, and the SoxRS signaling system, which have been heavily studied in *E. coli* but are also found in other bacterial lineages (Cabiscol *et al.*, 2000; Imlay, 2013). OxyR provides defense for the cell against peroxides while the SoxRS system defends against superoxide. Increases in the cellular concentration of hydrogen peroxide trigger OxyR, which promotes transcription of a set of operons that influence defense against ROS, such as increased levels of catalase (KatG) (Chiang and Schellhorn, 2012; Imlay, 2013). Hydrogen peroxide is a small, uncharged molecule that crosses cellular membranes easily, while superoxide is a charged molecule which will not easily permeate across the cell membrane, though is often produced in the cytoplasm of bacteria. In *E. coli*, the SoxRS system is a two-step regulatory system. First, SoxR is oxidized and binds upstream of and incites transcription of the *soxS* gene. Second, this SoxS protein acts as a secondary transcription factor and enhances the expression of numerous genes involved in ROS stress response, including MnSOD (the *sodA* gene) (Chiang and Schellhorn, 2012; Imlay, 2013). Transcription factors add another layer of complexity to understanding ROS defense in bacteria and have only been studied at sufficient depth in model systems like *E. coli*.

4.3.3.1.2 *Lactobacillus* – flexible metabolism in a food fermentation microbe

Lactobacillus is a genus belonging to the Firmicutes that is found in a range of environments, including the human reproductive tract and oral cavity, and which is used extensively in food production. Accumulation of ROS in fermented foods can be detrimental to *Lactobacillus* activity (Choe and Min, 2006). *Lactobacillus* are generally considered to be facultatively anaerobic or microaerophilic, with a fermentative lifestyle. There are some lactobacilli strains, along with other lactic acid-producing bacteria, that can switch between a fermentative and an aerobic respiration metabolism (Murphy and Condon, 1984; Zotta *et al.*, 2014, 2017). The lactic acid producing bacteria with both fermentative and aerobic lifestyles use NADH as the electron donor in their aerobic respiratory chain via NADH dehydrogenase, a quinone to deliver electrons, and a heme-dependent cytochrome oxidase as the terminal electron transport protein delivering electrons to nitrate or fumarate, typically (Zotta *et al.*, 2014). Exogenous sources of heme and menaquinone stimulate *Lactobacillus* aerobic respiration (Brooijmans *et al.*, 2009). *Lactobacillus* are equipped with a number of ROS-removing enzymes to facilitate survival under fluctuating ROS and oxygen conditions: two catalase families (haem-dependent and Mn-dependent), Mn-dependent SOD, NADH-peroxidase, and thioredoxin (Zotta *et al.*, 2014).

4.3.3.2 Obligate anaerobic lifestyles

This work analyzed anaerobic-specific ROS defense mechanisms within the bacterial domain, and summarized the oxygen defense mechanisms in six well-studied genera characterized as obligate anaerobes: *Dehalococcoides*, *Veillonella*, *Prevotella*, *Peptostreptococcus*, *Fusobacterium*, and *Bacteroides*. The genera considered are members of four different phyla (Chloroflexi, Firmicutes (2), Fusobacteria, and Bacteroidetes (2)). This work did not reveal any

specific trends in ROS defenses compared to the general ROS defense distribution across bacterial genomes (Figure 4.1) and did not identify any anaerobe-specific mechanisms of defense from the literature.

Dehalococcoides was chosen as a case study for the strict anaerobes. This genus comprises a group of small celled and small genome-containing obligate organohalide respirers. *Dehalococcoides* spp. have the ability to dechlorinate recalcitrant contaminants, including chlorinated benzenes (Hölscher *et al.*, 2003), ethenes (Duhamel and Edwards, 2006; Cheng and He, 2009), and ethanes (Grostern and Edwards, 2006). As a result, the *Dehalococcoides* have been used for bioremediation at contaminated sites worldwide (*e.g.*, (Pérez-De-Mora *et al.*, 2014)). A limitation to *Dehalococcoides*' use in bioremediation efforts is their obligately anaerobic nature. *Dehalococcoides* spp. can withstand oxygen exposure, however, exposure routinely results in extended lag periods in halide reduction (89-200 days) once anoxic conditions are reestablished (Amos *et al.*, 2008; Yoshikawa *et al.*, 2017). More severely, oxygen exposure can permanently disrupt *Dehalococcoides*-mediated dehalogenation reactions for some halogenated substrates (Amos *et al.*, 2008; Yoshikawa *et al.*, 2017).

Dehalococcoides are obligate anaerobes despite encoding a number of ROS defense mechanisms, a pattern that was observed in the strict anaerobic lineages examined here (Figure 4.2). All available *Dehalococcoides* genomes in the GTDB database (6 total) encode identical ROS defense mechanisms: thioredoxin reductase, thioredoxin 1, superoxide dismutase 2, alkyl hydroperoxide reductase 1, and superoxide reductase (Figure 4.2d). These five ROS enzymes are all found at comparatively high prevalence across the bacterial domain (Table 4.2), suggesting the *Dehalococcoides* spp. contain a core set of defenses rather than an anaerobe-specific suite of

enzymes. The ROS defenses encoded on *Dehalococcoides* genomes are sufficient to confer oxygen stress tolerance, lending weight to the current model that *Dehalococcoides* spp.'s obligate anaerobic lifestyle is due to reductive dehalogenase enzyme oxygen intolerance. In the environment, or in the mixed microbial cultures commonly used to grow *Dehalococcoides*, other organisms are also involved in oxygen scavenging, maintaining the strict anoxic growth conditions required to facilitate dechlorination (Hug *et al.*, 2012).

4.3.3.3 Anaerobic lifestyle with no defenses

It is conceivably possible for an organism to survive without any ROS defenses. While the ROS burden is substantially reduced in the absence of oxygen and aerobic metabolic processes, peroxides, superoxides, and other ROS can still inflict cellular damage in anoxic environments. In literature and genome-based searches, there were no identified cultured organisms lacking all known ROS defense mechanisms. Theoretical research has posited that the last universal common ancestor (LUCA) was able to tolerate oxygen exposure, and could detoxify ROS in the primordial environment (Slesak *et al.*, 2012). This is consistent with other research that predicted ROS defense systems have been present in bacteria for 4.1 to 3.5 billion years (Inupakutika *et al.*, 2016).

The discussion does not need to be constrained to the hypothetical, however, as researchers engineering the smallest possible microbial genome have attempted to exhaustively identify all redundant and non-essential genes. The resulting genomes can be examined for ROS defense enzymes to ascertain their importance under anaerobic growth conditions. The original minimal gene set was defined through comparison of the *Mycoplasma genitalium* and *Haemophilus*

Table 4.2 Distribution of ROS defense enzymes in obligately anaerobic bacterial genera compared to the median distribution across all bacterial phyla. Prevalence above 80% is highlighted in grey. Numbers in brackets correspond to the number of genomes available for that genus. Veillonella combines the genera Veillonella and Veillonella_A in the GTDB. ROS defenses not detected in the listed anaerobic genera were omitted from this table. ROS defenses omitted that were present at a median abundance greater than zero in bacterial phyla were: peroxiredoxin 1 (K13279; 2.3%) and glutaredoxin 3 (K03676; 8.3%).

	Median of all phyla	*Dehalococcoides* (6)	*Peptostreptococcus* (6)	*Veillonella* (16)	*Fusobacterium* (24)	*Bacteroides* (70)	*Prevotella* (257)
thioredoxin 1 (K03671)	98.1	100	100	18.8	100	100	91.4
thioredoxin reductase (K00384)	93.9	100	100	100	100	100	93.4
alkyl hydroperoxide reductase (K03386)	67.2	0	0	0.0	0	4.3	3.1
superoxide dismutase 2 (K04564)	49.6	100	16.7	68.8	0	97.1	31.1
superoxide reductase (K05919)	25.7	0	0	43.8	0	45.7	0.0
thiol peroxidase (tpx) (K11065)	19.0	0	0	0	0	18.6	1.9
catalase (K03781)	6.9	0	0	6.3	0	0	0
thioredoxin 2 (K03672)	3.3	0	83.3	0	100	87.1	35.8
glutathione peroxidase (K00432)	2.0	0	16.7	43.8	95.8	55.7	95.3
cytochrome c peroxidase (K00428)	1.4	0	100	100	0	98.6	3.1
catalase-peroxidase (K03782)	0.6	0	0	0	0	24.3	0.0
Mn-containing catalase (K07217)	0	100	0	100	0	0	0.0
alkyl hydroperoxide reductase 1 (K14171)	0	0	100	0	100	0	0.0
NADH peroxidase (K05910)	0	100	100	100	16.7	0	41.6
rubrerythrin (K19824)	0	0	0	0	0	92.9	43.2

influenzae genomes, which revealed a conserved core of 256 genes (Mushegian and Koonin, 1996). This conserved core did not contain any oxygen radical defense mechanisms (Mushegian and Koonin, 1996), but this core set has not been proven sufficient for growth. Later work with transposon mutagenesis reduced the 525 genes in the *Mycoplasma genitalium* genome to 375 essential genes in a living organism (Glass *et al.*, 2006), a gene complement that included thioredoxin and thioredoxin-disulfide reductase. Next, using the *Mycoplasma mycoides* genome as a blueprint, a fully synthetic minimal genome comprising 473 genes was generated – the shortest genome of an autonomously replicating organism to date (Hutchison *et al.*, 2016). In this synthetic minimal genome, quasi-essential genes that affected robust growth were maintained. Only one ROS defense mechanism was essential in all iterations of genome reduction: thioredoxin-disulfide reductase, a protein involved in defense against oxidative damage as well as in redox signaling (Hutchison *et al.*, 2016) (Figure 4.2e).

4.4 Summary and conclusions

Bacteria have successfully colonized nearly every environment on Earth, and can be found in anoxic, microoxic, and oxic conditions. Work examining the evolutionary origins of ROS and ROS defenses postulate the origin for ROS detoxification systems (Slesak *et al.*, 2012; Inupakutika *et al.*, 2016; Napolitano *et al*., 2019), which is impacted by the observation that all known bacteria contain ROS defenses, regardless of the oxygen concentrations within their habitat. This research described publicly available genomes identified a diverse suite of ROS scavenging enzymes that are patchily distributed both within individual phyla and across the bacterial domain. The distribution of ROS defenses suggests the prior focus on superoxide dismutase, catalase, and peroxidase as the primary enzymatic mechanisms for ROS detoxification is incomplete for many

lineages, including newly identified candidate phyla. In highlighting the ROS defenses of well-characterized aerobes, facultative anaerobes, and obligate anaerobes, the commonalities, as well as the diversity of approaches bacteria have developed for mitigating the damage caused by reactive oxygen species are revealed.

Chapter 5
Conclusions

5.1 Summary

Microbial ecology focuses on the study of microorganisms and their interactions with their environments and the interactions between microbes within a given environment. Investigations to describe and characterize microbial ecology are both limited and driven by the methods that are available. For example, studies in microbial ecology used to be limited to only culture-based methods. However, this resulted in all investigations being biased towards cultivable organisms and entirely overlooking non-culturable microorganisms within an environment (Boughner and Singh, 2016). Innovative molecular methods, such as new DNA sequencing technologies, revolutionized the gold standard for analyzing microbial communities, and allowed exploration of uncultured microorganisms (Boughner and Singh, 2016). One molecular method that has recently changed the scale and the depth with which microbial communities can be explored is metagenomics. Metagenomics, or shotgun sequencing, is the sequencing of all the environmental DNA within a sample, which allows for investigation of the entire microbial community beyond what is only culturable. Moreover, metagenomics has advantages over sequencing specific taxonomic marker genes (e.g., 16S rRNA gene amplicon sequencing) as the entire genetic content of the community is sequenced. Taxonomic markers are captured, and the remaining community DNA that is sequenced provides insight into the predicted metabolic capacity of microorganisms in that environment.

Metagenomic datasets can be evaluated in multiple ways to explore the breadth of microbial diversity and function in an environment – this richness of information and the

opportunity to address different microbial ecology questions is a recurring theme in my thesis work. For example, evaluating the genetic content can provide context for microbial lineages that are not well described or uncultivated, and phylogenetic relationships can be inferred to expand our understanding of microbial life and diversity (Rinke *et al.*, 2013; Hug *et al.*, 2016; Parks *et al.*, 2017), and to describe novel lineages (Méheust *et al.*, 2020). I exploit this capacity in Chapter 3, in expanding the genomic diversity of the phylum Cloacimonadota, and in describing the lifestyles and geochemical roles members of this phylum exhibit. The in-depth genetic information can also be used to explore the functional capacity of a microbial community within a specific environment, such as the exploration of the genetic content from hot spring biofilms to elucidate community members involved in sulfur cycling (McKay *et al.*, 2022). I leveraged deep metagenomic sequencing to explore the microbial community diversity and metabolic potential in a landfill-associated novel biofilm in Chapter 2. I coupled this metagenomic exploration with emerging tools, metaproteomics and metametabolomics, in order to strengthen hypotheses around community function. Finally, large metagenomic databases can be used to explore the distribution of a gene of interest, such as plastic degrading enzymes (Viljakainen and Hug, 2021). I took advantage of global metagenomic databases in an exploration of the distribution of enzymes involved in bacterial defenses to oxygen species in Chapter 4.

5.1.1 Microbial diversity and metabolic processes in landfill biofilms

In Chapter 2, I provide the first examination of landfill biofilm microbial communities by characterizing the microbial composition associated with the “black goo” and identifying key metabolic processes conserved across different biofilm samples. By surveying five biofilms obtained from a landfill site in Southern Ontario, my research provided the first information on the

composition of microorganisms in landfill biofilms. Rather than a consistent set of microorganisms occurring in each biofilm, biofilms isolated from separate biofouling events but at the same landfill site showed distinct microbial composition. At the study site, biofilm formation occurs periodically at different locations within the landfill leachate and gas collection systems. It was not known whether biofilm formation results from a central source (*e.g.,* a location with a mature biofilm that sloughs and seeds pieces of biofilm to other locations), or whether biofilm formation develops independently in separate colonizing events. The dissimilar distribution of microorganisms between the biofilm samples, for example at the genus level (Figure 2.9), suggests the latter model may be correct, where each detection of black goo stems from an independent formation event. The structure of the microbial community as detected through 16S rRNA gene sequencing, metagenomic, and metaproteomic datasets varied across biofilm samples, but also within individual biofilms depending on the dataset examined (Figure 2.10; Figure 2.12; Figure 2.13; Figure 2.14). Metaproteomics revealed "keystone" species, or organisms that are not abundant numerically but metabolically very active. One example is *Thiothrix* in BF4, where the MAG named BF4 Bin 143 had a relative abundance of 0.23% from the total MAG coverage but accounted for 34% of the total protein abundance in the BF4 sample (Appendix A, Supplementary Figure A1). Further, each of the three biofilm samples that underwent metaproteomic analyses had an individual MAG that dominated in the community based on expressed proteins (Table 2.5; Figure 2.15).

Despite taxonomic differences, evaluating the mostly highly expressed proteins in each sample identified commonalties at the functional level. Specifically, enzymes involved in sulfur metabolism and methanogenesis were consistently detected at high abundance (Figure 2.18; Figure

2.20); both processes are executed by anaerobic microorganisms. The leachate and gas collection systems have highly variable gas concentrations, with a likelihood for fluctuations in methane and oxygen concentrations. However, the prevalence of enzymes involved in obligately anaerobic processes within each biofilm suggest an anoxic environment exists within the biofilms. Therefore, the protective properties of the biofilm matrix may be providing a barrier from the highly variable gas capture and leachate system atmospheres, allowing anaerobic microorganisms to thrive in the goo. The prominence of sulfur cycling and methanogenesis as metabolic processes in the biofilms provides insight into what the biofilm community needs to survive and thrive. It is known that municipal solid waste contains sulfur compounds, and that under anoxic conditions microorganisms mediate conversion of sulfate to hydrogen sulfide (Sun *et al.*, 2016). As a significant component of the biofilm communities, reducing the success of anaerobic microorganisms that participate in sulfur metabolism could reduce biofilm establishment and growth.

5.1.2 Metabolic potential of phylum Cloacimonadota genomes

My research expanded on sparse knowledge regarding the metabolic potential of enigmatic microbial taxa, specifically the phylum Cloacimonadota (Chapter 3). Examination of genomes belonging to phylum Cloacimonadota clarified the distribution and metabolic capacities of this under-described lineage that is often found in engineered environments. Using metagenomic shotgun sequencing and generation of metagenome-assembled genomes, my research resulted in 24 Cloacimonadota MAGs being generated, supplementing Cloacimonadota genomes available in public databases, and expanding the number of available high-quality genomes by 80%. Phylogenetic analysis of all available Cloacimonadota genomes revealed two distinct clades

segregated by ecosystem of origin, where the first clade contained genomes derived from engineered environments, and the second clade contained the majority of genomes derived from environmental ecosystems (Figure 3.2). Overall, the genetic content of genomes predicts Cloacimonadota to be anaerobic and acetogenic, with a mixed fermentative and anaerobic respiratory lifestyle, with some metabolic functions differing across lineages (Figure 3.6). Although Cloacimonadota had been predicted to use fermentation of amino acids to generate energy (Pelletier *et al.*, 2008; Williams *et al.*, 2021), my research provided the first full description of amino acid degradation pathways encoded in Cloacimonadota genomes. For energy generation, Cloacimonadota genomes encoded two different kinds of ATPases, where the type was phylogenetically linked. The engineered clade encoded components of a V/A-type ATPase, and the environmental clade encoded a F-type ATPase. The differences in metabolic capacity between clades suggest clade specific mechanisms for energy generation and metabolism.

5.1.3 Distribution of enzymatic defenses against reactive oxygen species in bacteria

This thesis explored the distribution of enzymatic defenses against reactive oxygen species (ROS) in bacteria (Chapter 4). Understanding how bacteria defend against oxygen exposure clarifies bacterial survival in different environments and is specifically interesting for landfill sites where oxygen concentrations can be highly variable. By examining genomes from public databases with respect to enzymatic defenses to reactive oxygen species (ROS), my research provided a framework to evaluate the content of ROS defense in the bacterial domain. Defenses against ROS have been studied before (Imlay, 2013; Inupakutika *et al.*, 2016), especially those used by well-studied bacteria lineages (*e.g.*, superoxide dismutase in *Escherichia coli*).

Nonetheless, the broad distribution of enzymatic defenses, with regard to taxonomic groups or lifestyle, had not been previously assessed. My research provided a suite of enzymatic defenses against ROS compiled from the literature that can be leveraged to assess the potential for ROS defenses. Understanding the mechanisms bacteria use to tolerate oxygen is important to understand the mechanisms bacteria use to colonize diverse environments ranging from completely anoxic to oxic. The evaluation of ROS defenses across the bacterial domain did not identify any gene complements that could predict either an aerobic or anaerobic lifestyle, and the complement of ROS defenses across bacterial lineages are varied. And furthermore, in recently described bacterial lineages the "canonical" defenses of superoxide dismutase, catalase, and peroxidase are limited.

5.2 Future directions

The research presented in this thesis generates avenues for additional research related to landfill microbiology and our understanding of landfill microbial metabolisms. My research on microbial communities within problematic landfill biofilms provides a fundamental understanding of key microbial lineages and processes. This sets the stage to further investigate these biofilms, with the goal of biofilm prevention, reduction, or removal. Currently our team is working with a polymer scientist (Dr. Craig Benson, University of Wisconsin-Madison) to determine the composition of the biofilm matrix. This partnership has sourced a suite of biofilm samples from 13 landfills in the United States, which will be undergoing microbiome profiling. The larger set of biofilm samples will provide a broader understanding of microbial communities within biofilms, and comparison to work presented in Chapter 2 provides a framework to evaluate commonalties or differences between biofilms. Further, this collaboration provided the opportunity to share biofilm samples analyzed in Chapter 2 with Dr. Benson's laboratory where they will be evaluated

with Fourier-transform infrared (FTIR) spectroscopy to evaluate biofilm composition. These analyses may provide a better understanding into the major drivers of biofilm formation and clarify whether the polymer composing the biofilm matrix is microbially generated or a waste product that microorganisms have colonized.

Related to my research, a collaboration with industry partners at SiREM Laboratories involves establishment of a large-scale flow reactor system on the landfill site to mimic landfill infrastructure where biofilms form, in hopes of growing biofilm *in vitro*. The system contains cartridges where biofilm formation can occur that are easily sampled at different points of biofilm formation. The system is also built with six flow reactors to allow for replication and to test different treatment options in tandem. To date, leachate has been pumped through the system for eight months and has resulted in formation of solid materials. Leachate flow is on-going, aiming to achieve a mature, substantial biofilm. Once the system is fully functional it will enable studies of biofilm formation and/or growth, and samples will be evaluated through microbiome analysis. My research has provided information into typical biofilm microbial community members and major processes, which can be applied in this study system. For example, due to major active proteins in the biofilms being part of anaerobic metabolic processes, it suggests a flushing of the system with oxygen could be detrimental to the microorganisms thriving in the biofilm. Alternatively, limiting the substrate availability used by sulfate reducing bacteria may negatively impact biofilm growth. Hypotheses generated through my research can be experimentally validated in this system.

Another route to evaluate biofilm growth is through smaller-scale cultivation in the laboratory. Specifically, the BioFlux system (Fluxion Biosciences) uses plates with micro-fluidic

channels that provide the fluid shear forces for biofilm development. The benefits of this system include ease of replication for testing different treatments or conditions, including exposure to oxygen and a range of biocides. This would be helpful in testing a variety of media for biofilm growth, such as a synthetic medium designed to mimic landfill leachate or a minimal media. These experiments are comparatively quicker and easier, and therefore could effectively be used for preliminary testing of conditions, types of media, or biocides before these measures are applied to the larger, more complex flow system described above.

Beyond cultivation of the biofilm, which may also capture members of the phylum Cloacimonadota, cultivation efforts for Cloacimonadota would greatly benefit our understanding of this enigmatic phylum. My work presented in Chapter 3 provides an understanding of the metabolism used by Cloacimonadota lineages and can guide cultivation parameters. Previous cultivation attempts in other laboratories were unsuccessful at obtaining a pure culture of Cloacimonadota (Pelletier *et al.*, 2008), though they have been found in consortia cultures and this may be due to their requirement for a syntrophic partner (Pelletier *et al.*, 2008; Dyksma and Gallert, 2019). The consortia approach may be more effective at obtaining Cloacimonadota growth in the laboratory. Confirming metabolic predications provided in Chapter 3 would be fundamental to understanding why Cloacimonadota lineages thrive in engineered environments, and potentially expand or clarify the understanding of their roles in landfill microbial communities.

Due to the importance of controlling oxygen toxicity, especially in environments with varying concentrations of oxygen, understanding how anaerobic and aerobic microorganisms defend against oxygen is important. Since the publication of a version of Chapter 4 three years ago, databases for genomic content have been continually expanding. The increased genomic

content, especially with regard to novel lineages, provides more information to evaluate the distribution of enzymatic ROS defenses and reassess whether there is shared prevalence of specific enzymatic strategies between anaerobic or aerobic lifestyles. Next steps could include building tools to easily summarize the distribution of ROS defenses in genomes so other researchers can evaluate this information with their own datasets. This could be done by building a software tool similar to FeGenie, a tool designed for detection of genes specific to iron metabolism (Garber *et al.*, 2020). Or this could be achieved through implementation of a module for ROS defenses in an established annotation program, such as DRAM (Distilled and Refined Annotation of Metabolism) or METABOLIC (Shaffer *et al.*, 2020; Zhou *et al.*, 2022). Further research avenues include evaluating minimal genomes for ROS defenses to determine what ROS defenses are required for life, or whether certain defenses may be required for anaerobic or aerobic lifestyles.

References

Agency for Toxic Substances and Disease Registry (2001) Landfill gas basics. *Landfill Gas Prim* 3–14.

Ahlert, S., Zimmermann, R., Ebling, J., and König, H. (2016) Analysis of propionate-degrading consortia from agricultural biogas plants. *Microbiologyopen* **5**: 1027–1037.

Akyol, Ç., Ozbayram, E.G., Demirel, B., Onay, T.T., Ince, O., and Ince, B. (2019) Linking nano-ZnO contamination to microbial community profiling in sanitary landfill simulations. *Environ Sci Pollut Res* **26**: 13580–13591.

Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., *et al.* (2014) Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–10.

Amos, B.K., Ritalahti, K.M., Cruz-Garcia, C., Padilla-Crespo, E., and Löffler, F.E. (2008) Oxygen effect on *Dehalococcoides* viability and biomarker quantification. *Environ Sci Technol* **42**: 5718–5726.

Andrews, S.C. (2010) The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochim Biophys Acta - Gen Subj* **1800**: 691–705.

Antwi, P., Li, J., Opoku Boadi, P., Meng, J., Shi, E., Xue, C., *et al.* (2017) Functional bacterial and archaeal diversity revealed by 16S rRNA gene pyrosequencing during potato starch processing wastewater treatment in an UASB. *Bioresour Technol* **235**: 348–357.

Arenas, F.A., Covarrubias, P.C., Sandoval, J.M., Pérez-Donoso, J.M., Imlay, J.A., and Vásquez, C.C. (2011) The *Escherichia coli* BtuE protein functions as a resistance determinant against reactive oxygen species. *PLoS One* **6**: e15979.

Arthur, J.R. (2000) The glutathione peroxidases. *Cell Mol Life Sci* **57**: 1825–1835.

Arts, I.S., Gennaris, A., and Collet, J.F. (2015) Reducing systems protecting the bacterial cell envelope from oxidative damage. *FEBS Lett* **589**: 1559–1568.

Baderna, D., Caloni, F., and Benfenati, E. (2019) Investigating landfill leachate toxicity *in vitro*: A review of cell models and endpoints. *Environ Int* **122**: 21–30.

Bailey, S., Melis, A., Mackey, K.R.M., Cardol, P., Finazzi, G., van Dijken, G., *et al.* (2008) Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim Biophys Acta - Bioenerg* **1777**: 269–276.

Bareither, C.A., Wolfe, G.L., McMahon, K.D., and Benson, C.H. (2013) Microbial diversity and dynamics during methane production from municipal solid waste. *Waste Manag* **33**: 1982–1992.

Baricz, A., Chiriac, C.M., Andrei, A. Ștefan, Bulzu, P.A., Levei, E.A., Cadar, O., *et al.* (2021) Spatio-temporal insights into microbiology of the freshwater-to-hypersaline, oxic-hypoxic-euxinic waters of Ursu Lake. *Environ Microbiol* **23**: 3523–3540.

Barlaz, M.A., Schaefer, D.M., and Ham, R.K. (1989) Bacterial population development and chemical characteristics of refuse decomposition in a simulated sanitary landfill. *Appl Environ Microbiol* **55**: 55–65.

Barry, C.E., Slayden, R.A., and Mdluli, K. (1998) Mechanisms of isoniazid resistance in *Mycobacterium tuberculosis*. *Drug Resist Updat* **1**: 128–134.

Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., and Packmann, A.I. (2016) The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* **14**: 251–263.

Battin, T.J., Kaplan, L.A., Newbold, J.D., and Hansen, C.M.E. (2003) Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* **426**: 439–442.

Battistoni, A. and Rotilio, G. (1995) Isolation of an active and heat-stable monomeric form of Cu,Zn superoxide dismutase from the periplasmic space of *Escherichia coli*. *FEBS Lett* **374**: 199–202.

Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A.K., and Dorrestein, P.C. (2022) Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol* **20**: 143–160.

Benov, L.T. and Fridovich, I. (1994) *Escherichia coli* expresses a copper- and zinc-containing superoxide dismutase. *J Biol Chem* **269**: 25310–25314.

Berne, C., Ducret, A., Hardy, G.G., and Brun, Y. V. (2015) Adhesins involved in attachment to abiotic surfaces by gram-negative Bacteria. *Microbiol Spectr* **3**: MB-0018-2015.

Berne, C., Ellison, C.K., Ducret, A., and Brun, Y. V. (2018) Bacterial adhesion at the single-cell level. *Nat Rev Microbiol* **16**: 616–627.

Bertani, B. and Ruiz, N. (2018) Function and biogenesis of Lipopolysaccharides. *EcoSal Plus* **8**: 1.

Besemer, K., Singer, G., Hödl, I., and Battin, T.J. (2009) Bacterial community composition of stream biofilms in spatially variable-flow environments. *Appl Environ Microbiol* **75**: 7189–7195.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science (80-)* **277**: 1453–1462.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., *et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**: 852–857.

Botello Suárez, W.A., da Silva Vantini, J., Duda, R.M., Giachetto, P.F., Cintra, L.C., Tiraboschi Ferro, M.I., and de Oliveira, R.A. (2018) Predominance of syntrophic bacteria, *Methanosaeta* and *Methanoculleus* in a two-stage up-flow anaerobic sludge blanket reactor treating coffee processing wastewater at high organic loading rate. *Bioresour Technol* **268**: 158–168.

Boughner, L.A. and Singh, P. (2016) Microbial Ecology: Where are we now? *Postdoc J* **4**: 3–17.

Braz, G.H.R., Fernandez-Gonzalez, N., Lema, J.M., and Carballa, M. (2019) Organic overloading affects the microbial interactions during anaerobic digestion in sewage sludge reactors. *Chemosphere* **222**: 323–332.

Brenot, A., King, K.Y., Janowiak, B., Griffith, O., and Caparon, M.G. (2004) Contribution of glutathione peroxidase to the virulence of *Streptococcus pyogenes*. *Infect Immun* **72**: 408–413.

Brocks, J.J., Logan, G.A., Buick, R., and Summons, R.E. (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* **285**: 1033–1036.

Brooijmans, R., Smit, B., Santos, F., van Riel, J., de Vos, W.M., and Hugenholtz, J. (2009) Heme and menaquinone induced electron transport in lactic acid bacteria. *Microb Cell Fact* **8**: 28.

Broun, R. and Sattler, M. (2016) A comparison of greenhouse gas emissions and potential electricity recovery from conventional and bioreactor landfills. *J Clean Prod* **112**: 2664–2673.

Bukhov, N. and Carpentier, R. (2004) Alternative photosystem I-driven electron transport routes: Mechanisms and functions. *Photosynth Res* **82**: 17–33.

Bulatovic, V.M., Wengenack, N.L., Uhl, J.R., Hall, L., Roberts, G.D., Cockerill, F.R., and Rusnak, F. (2002) Oxidative stress increases susceptibility of *Mycobacterium tuberculosis* to isoniazid. *Antimicrob Agents Chemother* **46**: 2765–2771.

Burmølle, M., Ren, D., Bjarnsholt, T., and Sørensen, S.J. (2014) Interactions in multispecies biofilms: Do they actually matter? *Trends Microbiol* **22**: 84–91.

Burrell, P.C., O'Sullivan, C., Song, H., Clarke, W.P., and Blackall, L.L. (2004) Identification, detection, and spatial resolution of *Clostridium* populations responsible for cellulose degradation in a methanogenic landfill leachate bioreactor. *Appl Environ Microbiol* **70**: 2414–2419.

Bystrianský, L., Hujslová, M., Hršelová, H., Řezáčová, V., Němcová, L., Šimsová, J., *et al.* (2019) Observations on two microbial life strategies in soil: Planktonic and biofilm-forming microorganisms are separable. *Soil Biol Biochem* **136**: 107535.

Cabiscol, E., Tamarit, J., and Ros, J. (2000) Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int Microbiol* **3**: 3–8.

Cade, C.E., Dlouhy, A.C., Medzihradszky, K.F., Salas-Castillo, S.P., and Ghiladi, R.A. (2010) Isoniazid-resistance conferring mutations in *Mycobacterium tuberculosis* KatG: Catalase, peroxidase, and INH-NADH adduct formation activities. *Protein Sci* **19**: 458–474.

Callaghan, A. V., Davidova, I.A., Savage-Ashlock, K., Parisi, V.A., Gieg, L.M., Suflita, J.M., *et al.* (2010) Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287–7294.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

Calusinska, M., Goux, X., Fossépré, M., Muller, E.E.L., Wilmes, P., and Delfosse, P. (2018) A year of monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but different core microbiomes in bio-waste and wastewater anaerobic digestion systems. *Biotechnol Biofuels* **11**: 196.

Cardenas, J.P., Quatrini, R., and Holmes, D.S. (2016) Aerobic lineage of the oxidative stress response protein rubrerythrin emerged in an ancient microaerobic, (hyper)thermophilic environment. *Front Microbiol* **7**: 1822.

Cardinali-Rezende, J., Colturato, L.F.D.B., Colturato, T.D.B., Chartone-Souza, E., Nascimento, A.M.A., and Sanz, J.L. (2012) Prokaryotic diversity and dynamics in a full-scale municipal solid waste anaerobic reactor from start-up to steady-state conditions. *Bioresour Technol* **119**: 373–383.

Castelle, C.J. and Banfield, J.F. (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**: 1181–1197.

Chae, H.Z., Chung, S.J., and Rhee, S.G. (1994) Thioredoxin-dependent peroxide reductase from yeast. *J Biol Chem* **269**: 27670–27678.

Chaleckis, R., Meister, I., Zhang, P., and Wheelock, C.E. (2019) Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Curr Opin Biotechnol* **55**: 44–50.

Chapman, S.K., Davies, D.M., Watson, A.D., and Sykes, A.G. (1983) Metalloproteins and electron transfer. In *Inorganic Chemistry: Toward the 21st Century*. pp. 177–197.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020) GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**: 1925–1927.

Chen, L., Xie, Q.W., and Nathan, C. (1998) Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol Cell* **1**: 795–805.

Cheng, D. and He, J. (2009) Isolation and characterization of "*Dehalococcoides*" sp. strain MB, which dechlorinates tetrachloroethene to *trans*-1,2-dichloroethene. *Appl Environ Microbiol* **75**: 5910–5918.

Chiang, S.M. and Schellhorn, H.E. (2012) Regulators of oxidative stress response genes in *Escherichia coli* and their functional conservation in bacteria. *Arch Biochem Biophys* **525**: 161–169.

Choe, E. and Min, D.B. (2006) Chemistry and reactions of reactive oxygen species in foods. *Crit Rev Food Sci Nutr* **46**: 1–22.

Chojnacka, A., Szczęsny, P., Błaszczyk, M.K., Zielenkiewicz, U., Detman, A., Salamon, A., and Sikora, A. (2015) Noteworthy facts about a methane-producing microbial community processing acidic effluent from sugar beet molasses fermentation. *PLoS One* **10**: e0128008.

Chouari, R., Le Paslier, D., Dauga, C., Daegelen, P., Weissenbach, J., and Sghir, A. (2005) Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl Environ Microbiol* **71**: 2145–2153.

Christensen, T.H., Kjeldsen, P., Bjerg, P.L., Jensen, D.L., Christensen, J.B., Baun, A., *et al.* (2001) Biogeochemistry of landfill leachate plumes. *Appl Geochemistry* **16**: 659–718.

Co, R. and Hug, L.A. (2021) Prediction, enrichment and isolation identify a responsive, competitive community of cellulolytic microorganisms from a municipal landfill. *FEMS Microbiol Ecol* **97**: fiab065.

Compan, I. and Touati, D. (1993) Interaction of six global transcription regulators in expression of manganese superoxide dismutase in *Escherichia coli* K-12. *J Bacteriol* **175**: 1687–1696.

Comtois, S.L., Gidley, M.D., and Kelly, D.J. (2003) Role of the thioredoxin system and the thiol-peroxidases Tpx and Bcp in mediating resistance to oxidative and nitrosative stress in *Helicobacter pylori*. *Microbiology* **149**: 121–129.

Confer, A.W. and Ayalew, S. (2013) The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. *Vet Microbiol* **163**: 207–222.

Craig, L., Forest, K.T., and Maier, B. (2019) Type IV pili: dynamics, biophysics and functional consequences. *Nat Rev Microbiol* **17**: 429–440.

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., *et al.* (2015) Complete nitrification by *Nitrospira* bacteria. *Nature* **528**: 504–509.

Dang, H., Chen, R., Wang, L., Shao, S., Dai, L., Ye, Y., *et al.* (2011) Molecular characterization of putative biocorroding microbiota with a novel niche detection of Epsilon- and Zetaproteobacteria in Pacific Ocean coastal seawaters. *Environ Microbiol* **13**: 3059–3074.

Das, K.C. and Das, C.K. (2000) Thioredoxin, a singlet oxygen quencher and hydroxyl radical scavenger: Redox independent functions. *Biochem Biophys Res Commun* **277**: 443–447.

Delmont, T.O. and Eren, E.M. (2018) Linking pangenomes and metagenomes: The *Prochlorococcus* metapangenome. *PeerJ* **2018**: e4320.

Deponte, M. (2013) Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. *Biochim Biophys Acta - Gen Subj* **1830**: 3217–3266.

Derakshani, M., Lukow, T., and Liesack, W. (2001) Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl Environ Microbiol* **67**: 623–631.

Deusch, S., Bok, E., Schleicher, L., Seifert, J., and Steuber, J. (2019) Occurrence and function of the NA^+-translocating NADH:quinone oxidoreductase in *Prevotella* spp. *Microorganisms* **7**: 117.

Díaz-Pérez, C., Cervantes, C., Campos-García, J., Julián-Sánchez, A., and Riveros-Rosas, H. (2007) Phylogenetic analysis of the chromate ion transporter (CHR) superfamily. *FEBS J* **274**: 6215–6227.

Dolla, A., Fournier, M., and Dermoun, Z. (2006) Oxygen defense in sulfate-reducing bacteria. *J Biotechnol* **126**: 87–100.

Dombrowski, N., Seitz, K.W., Teske, A.P., and Baker, B.J. (2017) Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* **5**: 106.

Dong, L., Cao, G., Guo, X., Liu, T., Wu, J., and Ren, N. (2019) Efficient biogas production from cattle manure in a plug flow reactor: A large scale long term study. *Bioresour Technol* **278**: 450–455.

Donlan, R.M. (2002) Biofilms: Microbial life on surfaces. *Emerg Infect Dis* **8**: 881–890.

Dos Santos, W.G., Pacheco, I., Liu, M.Y., Teixeira, M., Xavier, A., and LeGall, J. (2000) Purification and characterization of an iron superoxide dismutase and a catalase from the sulfate-reducing bacterium *Desulfovibrio gigas*. *J Bacteriol* **182**: 796–804.

Dubois-Brissonnet, F., Trotier, E., and Briandet, R. (2016) The biofilm lifestyle involves an increase in bacterial membrane saturated fatty acids. *Front Microbiol* **7**: 1673.

Dudek, N.K., Sun, C.L., Burstein, D., Kantor, R.S., Aliaga Goltsman, D.S., Bik, E.M., *et al.* (2017) Novel microbial diversity and functional potential in the marine mammal oral microbiome. *Curr Biol* **27**: 3752–3762.

Duhamel, M. and Edwards, E.A. (2006) Microbial composition of chlorinated ethene-degrading cultures dominated by *Dehalococcoides. FEMS Microbiol Ecol* **58**: 538–549.

Dyksma, S. and Gallert, C. (2019) *Candidatus* Syntrophosphaera thermopropionivorans: a novel player in syntrophic propionate oxidation during anaerobic digestion. *Environ Microbiol Rep* **11**: 558–570.

Eloe-Fadrosh, E.A., Paez-Espino, D., Jarett, J., Dunfield, P.F., Hedlund, B.P., Dekas, A.E., *et al.* (2016) Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat Commun* **7**: 1–10.

Emwas, A.H., Roy, R., McKay, R.T., Tenori, L., Saccenti, E., Nagana Gowda, G.A., *et al.* (2019) NMR spectroscopy for metabolomics research. *Metabolites* **9**: metabo9070123.

Environment and Climate Change Canada (2018) Canadian Environmental Sustainability Indicators- Solid waste diversion and disposal. *Gov Canada*.

Environment and Climate Change Canada (2021) Greenhouse gas sources and sinks: executive summary 2021. *Gov Canada*.

Environment and Climate Change Canada (2022) Reducing methane emissions from Canada's municipal solid waste landfills. *Gov Canada*.

Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015) Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* **2015**: e1319.

Farr, S.B., Touati, D., and Kogoma, T. (1988) Effects of oxygen stress on membrane functions in *Escherichia coli*: Role of HPI catalase. *J Bacteriol* **170**: 1837–1842.

Fiehn, O. (2016) Metabolomics by gas chromatography-mass spectrometry: Combined targeted and untargeted profiling. *Curr Protoc Mol Biol* **2016**: 30.4.1-30.4.32.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., *et al.* (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., and Kjelleberg, S. (2016) Biofilms: An emergent form of bacterial life. *Nat Rev Microbiol* **14**: 563–575.

Flemming, H.C. and Wuertz, S. (2019) Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* **17**: 247–260.

Foulquier, A., Mermillod-Blondin, F., Malard, F., and Gibert, J. (2011) Response of sediment biofilm to increased dissolved organic carbon supply in groundwater artificially recharged with stormwater. *J Soils Sediments* **11**: 382–393.

Francolini, I. and Donelli, G. (2010) Prevention and control of biofilm-based medical-device-related infections. *FEMS Immunol Med Microbiol* **59**: 227–238.

Fridovich, I. (1975) Superoxide dismutases. *Annu Rev Biochem* **44**: 147–159.

Friedlander, R.S., Vlamakis, H., Kim, P., Khan, M., Kolter, R., and Aizenberg, J. (2013) Bacterial flagella explore microscale hummocks and hollows to increase adhesion. *Proc Natl Acad Sci U S A* **110**: 5624–5629.

Galván, A., Urbina, P., and De Castro, F. (2000) Characterization of filamentous microorganisms in rotating biological contactor biofilms of wastewater treatment plants. *Bioprocess Eng* **22**: 257–260.

Garber, A.I., Nealson, K.H., Okamoto, A., McAllister, S.M., Chan, C.S., Barco, R.A., and Merino, N. (2020) FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol* **11**: 37.

Gencic, S. and Grahame, D.A. (2003) Nickel in subunit β of the acetyl-CoA decarbonylase/synthase multienzyme complex in methanogens: Catalytic properties and evidence for a binuclear Ni-Ni site. *J Biol Chem* **278**: 6101–6110.

Gies, E.A., Konwar, K.M., Thomas Beatty, J., and Hallam, S.J. (2014) Illuminating microbial dark matter in meromictic Sakinaw Lake. *Appl Environ Microbiol* **80**: 6807–6818.

Ginestet, C. (2011) ggplot2: Elegant graphics for data analysis. *J R Stat Soc* **174**: 245–246.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., *et al.* (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci* **103**: 425–430.

Gordon, J., Holman, R.A., and McLeod, J.W. (1953) Further observations on the production of hydrogen peroxide by anaerobic bacteria. *J Pathol Bacteriol* **66**: 527–537.

Grostern, A. and Edwards, E.A. (2006) Growth of *Dehalobacter* and *Dehalococcoides* spp. during degradation of chlorinated ethanes. *Appl Environ Microbiol* **72**: 428–436.

Gu, A.Z., Hedlund, B.P., Staley, J.T., Strand, S.E., and Stensel, H.D. (2004) Analysis and comparison of the microbial community structures of two enrichment cultures capable of reductively dechlorinating TCE and *cis*-DCE. *Environ Microbiol* **6**: 45–54.

Guerrero-Cruz, S., Vaksmaa, A., Horn, M.A., Niemann, H., Pijuan, M., and Ho, A. (2021) Methanotrophs: Discoveries, environmental relevance, and a perspective on current and future applications. *Front Microbiol* **12**: 678057.

Han, C., Kotsyurbenko, O., Chertkov, O., Held, B., Lapidus, A., Nolan, M., *et al.* (2012) Complete genome sequence of the sulfur compounds oxidizing chemolithoautotroph *Sulfuricurvum kujiense* type strain (YK-1 T). *Stand Genomic Sci* **6**: 94–103.

Han, Y. and Perner, M. (2015) The globally widespread genus *Sulfurimonas*: Versatile energy metabolisms and adaptations to redox clines. *Front Microbiol* **6**: 989.

Hassa, J., Maus, I., Off, S., Pühler, A., Scherer, P., Klocke, M., and Schlüter, A. (2018) Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions and functional relationships of microbial communities residing in biogas plants. *Appl Microbiol Biotechnol* **102**: 5045–5063.

Havaux, M., Guedeney, G., Hagemann, M., Yeremenko, N., Matthijs, H.C.P., and Jeanjean, R. (2005) The chlorophyll-binding protein IsiA is inducible by high light and protects the cyanobacterium *Synechocystis* PCC6803 from photooxidative stress. *FEBS Lett* **579**: 2289–2293.

He, Q., Dolganov, N., Björkman, O., and Grossman, A.R. (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* **276**: 306–314.

Hernsdorf, A.W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K., *et al.* (2017) Potential for microbial H_2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J* **11**: 1915–1929.

Hess, V., Gallegos, R., Andrew Jones, J., Barquera, B., Malamy, M.H., and Müller, V. (2016) Occurrence of ferredoxin: NAD^+ oxidoreductase activity and its ion specificity in several Gram-positive and Gram-negative bacteria. *PeerJ* **2016**: e1515.

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., and Benndorf, D. (2017) Challenges and perspectives of metaproteomic data analysis. *J Biotechnol* **261**: 24–36.

Hilger, H.H. and Barlaz, M.A. (2006) Anaerobic decomposition of refuse in landfills and methane oxidation in landfill covers. In *Manual of Environmental Microbiology*. pp. 818–842.

Hiller, K., Hangebrauk, J., Jäger, C., Spura, J., Schreiber, K., and Schomburg, D. (2009) Metabolite detector: Comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal Chem* **81**: 3429–3439.

Hölscher, T., Görisch, H., and Adrian, L. (2003) Reductive dehalogenation of chlorobenzene congeners in cell extracts of *Dehalococcoides* sp. Strain CBDB1. *Appl Environ Microbiol* **69**: 2999–3001.

Hoornweg, D. and Bhada-Tata, P. (2012) A global review of solid waste management. *World Bank Urban Dev Ser Knowl Pap* 1–116.

Hoornweg, D., Bhada-Tata, P., and Kennedy, C. (2013) Waste production must peak this century. *Nature* **502**: 615–617.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., *et al.* (2016) A new view of the tree of life. *Nat Microbiol* **1**: 16048.

Hug, L.A., Beiko, R.G., Rowe, A.R., Richardson, R.E., and Edwards, E.A. (2012) Comparative metagenomics of three *Dehalococcoides*-containing enrichment cultures: the role of the non-dechlorinating community. *BMC Genomics* **13**: 327.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Tennessen, K., *et al.* (2016) The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand Genomic Sci* **11**: 1–5.

Hutchison, C.A., Chuang, R.Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science* **351**: aad6253–aad6253.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Imlay, J.A. (2013) The molecular mechanisms and physiological consequences of oxidative stress: Lessons from a model bacterium. *Nat Rev Microbiol* **11**: 443–454.

Imlay, J.A. (2015) Transcription factors that defend bacteria against reactive oxygen species. *Annu Rev Microbiol* **69**: 93–108.

Inagaki, F., Tsunogai, U., Suzuki, M., Kosaka, A., Machiyama, H., Takai, K., *et al.* (2004) Characterization of C1-metabolizing prokaryotic communities in methane seep habitats at the Kuroshima Knoll, southern Ryukyu arc, by analyzing *pmoA*, *mmoX*, *mxaF*, *mcrA*, and 16S rRNA genes. *Appl Environ Microbiol* **70**: 7445–7455.

Inupakutika, M.A., Sengupta, S., Devireddy, A.R., Azad, R.K., and Mittler, R. (2016) The evolution of reactive oxygen species metabolism. *J Exp Bot* **67**: 5933–5943.

Jägevall, S., Rabe, L., and Pedersen, K. (2011) Abundance and diversity of biofilms in natural and artificial aquifers of the Äspö Hard Rock Laboratory, Sweden. *Microb Ecol* **61**: 410–422.

Janda, J.M. and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**: 2761–2764.

Jankowska, E., Duber, A., Chwialkowska, J., Stodolny, M., and Oleskowicz-Popiel, P. (2018) Conversion of organic waste into volatile fatty acids – The influence of process operating parameters. *Chem Eng J* **345**: 395–403.

Johnson, L.A. and Hug, L.A. (2019) Distribution of reactive oxygen species defense mechanisms across domain Bacteria. *Free Radic Biol Med* **140**: 93–102.

Joshi, N. and Fass, J. (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle* 2011.

Joshi, R.V., Gunawan, C., and Mann, R. (2021) We are one: Multispecies metabolism of a biofilm consortium and their treatment strategies. *Front Microbiol* **12**: 80.

Kämpfer, P., Denger, K., Cook, A.M., Lee, S.T., Jäckel, U., Denner, E.B.M., and Busse, H.J. (2006) *Castellaniella* gen. nov., to accommodate the phylogenetic lineage of *Alcaligenes defragrans*, and proposal of *Castellaniella defragrans* gen. nov., comb. nov. and *Castellaniella denitrificans* sp. nov. *Int J Syst Evol Microbiol* **56**: 815–819.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**: 27522v1.

Kang, S.W., Chae, H.Z., Seo, M.S., Kim, K., Baines, I.C., and Rhee, S.G. (1998) Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide generated in response to growth factors and tumor necrosis factor-alpha. *J Biol Chem* **273**: 6297–6302.

Kaster, A.K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., and Thauer, R.K. (2011) More than 200 genes required for methane formation from H_2 and CO_2 and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. *Archaea* **2011**: 973848.

Kasuga, I., Shimazaki, D., and Kunikane, S. (2007) Influence of backwashing on the microbial community in a biofilm developed on biological activated carbon used in a drinking water treatment plant. *Water Sci Technol* **55**: 173–180.

Keele, B.B., McCord, J.M., and Fridovich, I. (1970) Superoxide dismutase from *Escherichia coli* B. A new manganese-containing enzyme. *J Biol Chem* **245**: 6176–6181.

Kelly, R.T., Page, J.S., Luo, Q., Moore, R.J., Orton, D.J., Tang, K., and Smith, R.D. (2006) Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal Chem* **78**: 7796–7801.

Kelman, D., Ben-Amotz, A., and Berman-Frank, I. (2009) Carotenoids provide the major antioxidant defence in the globally significant N_2-fixing marine cyanobacterium *Trichodesmium*. *Environ Microbiol* **11**: 1897–1908.

Kim, S. and Pevzner, P.A. (2014) MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**: 1–10.

Kind, T., Wohlgemuth, G., Lee, D.Y., Lu, Y., Palazoglu, M., Shahbaz, S., and Fiehn, O. (2009) FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **81**: 10038–10048.

King, H.C., Khera-Butler, T., James, P., Oakley, B.B., Erenso, G., Aseffa, A., *et al.* (2017) Environmental reservoirs of pathogenic mycobacteria across the Ethiopian biogeographical landscape. *PLoS One* **12**: e0173811.

Kirchman, D.L. (2018) Processes in microbial ecology: Second edition, Oxford University Press.

Kirilovsky, D. (2007) Photoprotection in cyanobacteria: The orange carotenoid protein (OCP)-related non-photochemical-quenching mechanism. *Photosynth Res* **93**: 7–16.

Kirk, M.F., Wilson, B.H., Marquart, K.A., Zeglin, L.H., Vinson, D.S., and Flynn, T.M. (2015) Solute concentrations influence microbial methanogenesis in coal-bearing strata of the Cherokee basin, USA. *Front Microbiol* **6**: 1287.

Kitagawa, T., Dey, A., Lugo-Mas, P., Benedict, J.B., Kaminsky, W., Solomon, E., and Kovacs, J.A. (2006) A functional model for the cysteinate-ligated non-heme iron enzyme superoxide reductase (SOR). *J Am Chem Soc* **128**: 14448–14449.

Korshunov, S. and Imlay, J.A. (2006) Detection and quantification of superoxide formed within the periplasm of *Escherichia coli*. *J Bacteriol* **188**: 6326–6334.

Kuever, J., Rainey, F.A., and Widdel, F. (2015) *Desulfuromonas*. *Bergey's Man Syst Archaea Bact* 1–7.

Kuhns, M., Trifunović, D., Huber, H., and Müller, V. (2020) The Rnf complex is a Na^+ coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. *Commun Biol* **3**: 431.

Kumar, A., Farhana, A., Guidry, L., Saini, V., Hondalus, M., and Steyn, A.J.C. (2011) Redox homeostasis in mycobacteria: the key to tuberculosis control? *Expert Rev Mol Med* **13**: e39.

La Carbona, S., Sauvageot, N., Giard, J.C., Benachour, A., Posteraro, B., Auffray, Y., *et al.* (2007) Comparative study of the physiological roles of three peroxidases (NADH peroxidase, alkyl hydroperoxide reductase and thiol peroxidase) in oxidative stress response, survival inside macrophages and virulence of *Enterococcus faecalis*. *Mol Microbiol* **66**: 1148–1163.

Lam, P. and Kuypers, M.M.M. (2011) Microbial nitrogen cycling processes in oxygen minimum zones. *Ann Rev Mar Sci* **3**: 317–345.

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

Latifi, A., Ruiz, M., and Zhang, C.C. (2009) Oxidative stress in cyanobacteria. *FEMS Microbiol Rev* **33**: 258–278.

Lee, M.D. (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**: 4162–4164.

Lenhart, T.R., Duncan, K.E., Beech, I.B., Sunner, J.A., Smith, W., Bonifay, V., *et al.* (2014) Identification and characterization of microbial biofilm communities associated with corroded oil pipeline surfaces. *Biofouling* **30**: 823–835.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.

Li, K., Whitfield, M., and Vliet, V. (2013) Beating the bugs: Roles of microbial biofilms in corrosion. *Corros Rev* **31**: 73–84.

Liduino, V.S., Cravo-Laureau, C., Noel, C., Carbon, A., Duran, R., Lutterbach, M.T., and Camporese Sérvulo, E.F. (2019) Comparison of flow regimes on biocorrosion of steel pipe weldments: Community composition and diversity of biofilms. *Int Biodeterior Biodegrad* **143**: 104717.

Limam, R.D., Chouari, R., Mazéas, L., Wu, T. Di, Li, T., Grossin-Debattista, J., *et al.* (2014) Members of the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion of cellulose. *Microbiologyopen* **3**: 157–167.

Liu, J., Wu, W., Chen, C., Sun, F., and Chen, Y. (2011) Prokaryotic diversity, composition structure, and phylogenetic analysis of microbial communities in leachate sediment ecosystems. *Appl Microbiol Biotechnol* **91**: 1659–1675.

Liu, R. and Ochman, H. (2007) Stepwise formation of the bacterial flagellar system. *Proc Natl Acad Sci U S A* **104**: 7116–7121.

Liu, X., Huang, M., Bao, S., Tang, W., and Fang, T. (2020) Nitrate removal from low carbon-to-nitrogen ratio wastewater by combining iron-based chemical reduction and autotrophic denitrification. *Bioresour Technol* **301**: 122731.

Lleo, M., Bonato, B., Tafi, M.C., Caburlotto, G., Benedetti, D., and Canepari, P. (2007) Adhesion to medical device materials and biofilm formation capability of some species of enterococci in different physiological states. *FEMS Microbiol Lett* **274**: 232–237.

Loewen, P.C. and Switala, J. (1986) Purification and characterization of catalase HPII from *Escherichia coli* K12. *Biochem Cell Biol* **64**: 638–646.

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

Lozecznik, S., Sparling, R., Clark, S.P., VanGulck, J.F., and Oleszkiewicz, J.A. (2012) Acetate and propionate impact on the methanogenesis of landfill leachate and the reduction of clogging components. *Bioresour Technol* **104**: 37–43.

Madigan, M.T., Martinko, J.M., Stahl, D.A., and Clark, D.P. (2012) Brock Biology of Microorganisms, 13th Edition, Pearson Education Inc.

Manara, S., Asnicar, F., Beghini, F., Bazzani, D., Cumbo, F., Zolfo, M., *et al.* (2019) Microbial genomes from non-human primate gut metagenomes expand the primate-associated bacterial tree of life with over 1000 novel species. *Genome Biol* **20**: 1–16.

Mardanov, A. V., Gruzdev, E. V., Smolyakov, D.D., Rudenko, T.S., Beletsky, A. V., Gureeva, M. V., *et al.* (2020) Genomic and metabolic insights into two novel *Thiothrix* species from enhanced biological phosphorus removal systems. *Microorganisms* **8**: 1–13.

Markley, J.L., Brüschweiler, R., Edison, A.S., Eghbalnia, H.R., Powers, R., Raftery, D., and Wishart, D.S. (2017) The future of NMR-based metabolomics. *Curr Opin Biotechnol* **43**: 34–40.

Masák, J., Čejková, A., Schreiberová, O., and Řezanka, T. (2014) *Pseudomonas* biofilms: Possibilities of their control. *FEMS Microbiol Ecol* **89**: 1–14.

Masoner, J.R., Kolpin, D.W., Furlong, E.T., Cozzarelli, I.M., and Gray, J.L. (2016) Landfill leachate as a mirror of today's disposable society: Pharmaceuticals and other contaminants of emerging concern in final leachate from landfills in the conterminous United States. *Environ Toxicol Chem* **35**: 906–918.

McCord, J.M. and Fridovich, I. (1969) Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *J Biol Chem* **244**: 6049–6055.

McDonald, A.E. and Vanlerberghe, G.C. (2005) Alternative oxidase and plastoquinol terminal oxidase in marine prokaryotes of the Sargasso Sea. *Gene* **349**: 15–24.

McDonald, J.E., Lockhart, R.J., Cox, M.J., Allison, H.E., and McCarthy, A.J. (2008) Detection of novel *Fibrobacter* populations in landfill sites and determination of their relative abundance via quantitative PCR. *Environ Microbiol* **10**: 1310–1319.

McGivern, B.B., Tfaily, M.M., Borton, M.A., Kosina, S.M., Daly, R.A., Nicora, C.D., *et al.* (2021) Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat Commun* **12**: 1–16.

McKay, L.J., Nigro, O.D., Dlakić, M., Luttrell, K.M., Rusch, D.B., Fields, M.W., and Inskeep, W.P. (2022) Sulfur cycling and host-virus interactions in *Aquificales*-dominated biofilms from Yellowstone's hottest ecosystems. *ISME J* **16**: 842–855.

Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A., and Doxey, A.C. (2019) Annotree: Visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* **47**: 4442–4448.

Meyer-Dombard, D.R., Bogner, J.E., and Malas, J. (2020) A review of landfill microbiology and ecology: A call for modernization with 'next generation' technology. *Front Microbiol* **11**: 1127.

Mezzina, M.P. and Pettinari, M.J. (2016) Phasins, multifaceted polyhydroxyalkanoate granule-associated proteins. *Appl Environ Microbiol* **82**: 5060–5067.

Mishra, S. and Imlay, J. (2012) Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Arch Biochem Biophys* **525**: 145–160.

Mission Black Goo: Figuring out what it is and how to tackle it (2022) https://www.waste360.com/special-waste/mission-black-goo-figuring-out-what-it-and-how-tackle-it.

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013) Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* **8**: e53608.

Moestedt, J., Müller, B., Nagavara Nagaraj, Y., and Schnürer, A. (2020) Acetate and lactate production during two-stage anaerobic digestion of food waste driven by *Lactobacillus* and *Aeriscardovia*. *Front Energy Res* **8**: 105.

Monroe, M.E., Shaw, J.L., Daly, D.S., Adkins, J.N., and Smith, R.D. (2008) MASIC: A software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput Biol Chem* **32**: 215–217.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**: W182–W185.

Murphy, M.G. and Condon, S. (1984) Comparison of aerobic and anaerobic growth of *Lactobacillus plantarum* in a glucose medium. *Arch Microbiol* **138**: 49–53.

Mushegian, A.R. and Koonin, E. V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* **93**: 10268–10273.

Nadell, C.D., Drescher, K., and Foster, K.R. (2016) Spatial structure, cooperation and competition in biofilms. *Nat Rev Microbiol* **14**: 589–600.

Nakayasu, E.S., Nicora, C.D., Sims, A.C., Burnum-Johnson, K.E., Kim, Y.-M., Kyle, J.E., *et al.* (2016) MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *mSystems* **1**: e00043-16.

Napolitano, S., Reber, R.J., Rubini, M., and Glockshuber, R. (2019) Functional analyses of ancestral thioredoxins provide insights into their evolutionary history. *J Biol Chem* **294**: 14105–14118.

Ng, V.H., Cox, J.S., Sousa, A.O., MacMicking, J.D., and McKinney, J.D. (2004) Role of KatG catalase-peroxidase in mycobacterial pathogenesis: Countering the phagocyte oxidative burst. *Mol Microbiol* **52**: 1291–1302.

Nguyen, P.M., Do, P.T., Pham, Y.B., Doan, T.O., Nguyen, X.C., Lee, W.K., et al. (2022) Roles, mechanism of action, and potential applications of sulfur-oxidizing bacteria for environmental bioremediation. *Sci Total Environ* **852**: 158203.

Nicholls, P., Fita, I., and Loewen, P.C. (2000) Enzymology and structure of catalases. *Adv Inorg Chem* **51**: 51–106.

Nicoletti, D., Sharma, M., and Gieg, L.M. (2022) Assessing microbial corrosion risk on offshore crude oil production topsides under conditions of nitrate and nitrite treatment for souring. *Microorganisms* **10**: 932.

Niederhoffer, E.C., Naranjo, C.M., Bradley, K.L., and Fee, J.A. (1990) Control of *Escherichia coli* superoxide dismutase (*sodA* and *sodB*) genes by the ferric uptake regulation (*fur*) locus. *J Bacteriol* **172**: 1930–1938.

Nieto, L.M.R., Mehaffy, C., Creissen, E., Troudt, J.L., Troy, A., Bielefeldt-Ohmann, H., *et al.* (2016) Virulence of *Mycobacterium tuberculosis* after acquisition of Isoniazid resistance: Individual nature of *katG* mutants and the possible role of AhpC. *PLoS One* **11**: e0166807.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017) MetaSPAdes: A new versatile metagenomic assembler. *Genome Res* **27**: 824–834.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**: 29–34.

Oh, E., McMullen, L., and Jeon, B. (2015) Impact of oxidative stress defense on bacterial survival and morphological change in *Campylobacter jejuni* under aerobic conditions. *Front Microbiol* **6**: 295.

Omar, H. and Rohani, S. (2015) Treatment of landfill waste, leachate and landfill gas: A review. *Front Chem Sci Eng* **9**: 15–32.

Von Ossowski, I., Mulvey, M.R., Leco, P.A., Borys, A., and Loewen, P.C. (1991) Nucleotide sequence of *Escherichia coli katE*, which encodes catalase HPII. *J Bacteriol* **173**: 514–520.

Pal, D., Bhardwaj, A., Sudan, S.K., Kaur, N., Kumari, M., Bisht, B., *et al.* (2018) *Thauera propionica* sp. nov., isolated from downstream sediment sample of the river Ganges, Kanpur, India. *Int J Syst Evol Microbiol* **68**: 341–346.

Papadatou, M., Robson, S.C., Dobretsov, S., Watts, J.E.M., Longyear, J., and Salta, M. (2021) Marine biofilms on different fouling control coating types reveal differences in microbial community composition and abundance. *Microbiologyopen* **10**: e1231.

Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020) A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**: 1079–1086.

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

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., *et al.* (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**: 1533–1542.

Patel, M.P. and Blanchard, J.S. (1999) Expression, purification, and characterization of *Mycobacterium tuberculosis* mycothione reductase. *Biochemistry* **38**: 11827–11833.

Paul, B.D. and Snyder, S.H. (2010) The unusual amino acid L-ergothioneine is a physiologic cytoprotectant. *Cell Death Differ* **17**: 1134–1140.

Pelletier, E., Kreimeyer, A., Bocs, S., Rouy, Z., Gyapay, G., Chouari, R., *et al.* (2008) "*Candidatus* Cloacamonas acidaminovorans": Genome sequence reconstruction provides a first glimpse of a new bacterial division. *J Bacteriol* **190**: 2572–2579.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Pérez-De-Mora, A., Zila, A., McMaster, M.L., and Edwards, E.A. (2014) Bioremediation of chlorinated ethenes in fractured bedrock and associated changes in dechlorinating and nondechlorinating microbial populations. *Environ Sci Technol* **48**: 5770–5779.

Perkins, A., Nelson, K.J., Parsonage, D., Poole, L.B., and Karplus, P.A. (2015) Peroxiredoxins: Guardians against oxidative stress and modulators of peroxide signaling. *Trends Biochem Sci* **40**: 435–445.

Pesci, E.C., Cottle, D.L., and Pickett, C.L. (1994) Genetic, enzymatic, and pathogenic studies of the iron superoxide dismutase of *Campylobacter jejuni*. *Infect Immun* **62**: 2687–2694.

Piddington, D.L., Fang, F.C., Laessig, T., Cooper, A.M., Orme, I.M., and Buchmeier, N.A. (2001) Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. *Infect Immun* **69**: 4980–4987.

Poudel, S., Dunham, E.C., Lindsay, M.R., Amenabar, M.J., Fones, E.M., Colman, D.R., and Boyd, E.S. (2018) Origin and evolution of flavin-based electron bifurcating enzymes. *Front Microbiol* **9**: 1762.

Pourcher, A.-M., Sutra, L., Hébé, I., Moguedet, G., Bollet, C., Simoneau, P., and Gardan, L. (2006) Enumeration and characterization of cellulolytic bacteria from refuse of a landfill. *FEMS Microbiol Ecol* **34**: 229–241.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Qin, X., Wu, X., Li, L., Li, C., Zhang, Z., and Zhang, X. (2018) The advanced anaerobic expanded granular sludge bed (AnaEG) possessed temporally and spatially stable treatment performance and microbial community in treating starch processing wastewater. *Front Microbiol* **9**: 589.

Ramaswamy, S. V., Reich, R., Dou, S.-J., Jasperse, L., Pan, X., Wanger, A., *et al.* (2003) Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **47**: 1241–1250.

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

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**: D624–D632.

Reiss, R., Ihssen, J., Richter, M., Eichhorn, E., Schilling, B., and Thöny-Meyer, L. (2013) Laccase versus laccase-like multi-copper oxidase: a comparative study of similar enzymes with diverse substrate spectra. *PLoS One* **8**: e65633.

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

Renou, S., Givaudan, J.G., Poulain, S., Dirassouyan, F., and Moulin, P. (2008) Landfill leachate treatment: Review and opportunity. *J Hazard Mater* **150**: 468–493.

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

Rosenzweig, A.C. and O'Halloran, T. V. (2000) Structure and chemistry of the copper chaperone proteins. *Curr Opin Chem Biol* **4**: 140–147.

Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C., *et al.* (2015) Comparative integrated omics: Identification of key functionalities in microbial community-wide metabolic networks. *npj Biofilms Microbiomes* **1**: 1–11.

Rumbaugh, K.P. and Sauer, K. (2020) Biofilm dispersion. *Nat Rev Microbiol* **18**: 571–586.

Russell, D.G., Barry, C.E., and Flynn, J.L. (2010) Tuberculosis: What we don't know can, and does, hurt us. *Science* **328**: 852–856.

Saha, S., Jeon, B.H., Kurade, M.B., Chatterjee, P.K., Chang, S.W., Markkandan, K., *et al.* (2019) Microbial acclimatization to lipidic-waste facilitates the efficacy of acidogenic fermentation. *Chem Eng J* **358**: 188–196.

Sauk, A.H. and Hug, L.A. (2022) Substrate-restricted methanogenesis and limited volatile organic compound degradation in highly diverse and heterogeneous municipal landfill microbial communities. *ISME Commun* **2**: 1–14.

Schubert, H., Matthijs, H.C.P., and Mur, L.R. (1995) In vivo assay of P700 redox changes in the cyanobacterium *Fremyella diplosiphon* and the role of cytochrome *c* oxidase in regulation of photosynthetic electron transfer. *Photosynthetica* **31**: 517–527.

Schulze-Röbbecke, R. (1993) Mycobacteria in the environment. *Immun Infekt* **21**: 126–31.

Schwermer, C.U., Lavik, G., Abed, R.M.M., Dunsmore, B., Ferdelman, T.G., Stoodley, P., *et al.* (2008) Impact of nitrate on the structure and function of bacterial biofilm communities in pipelines used for injection of seawater into oil fields. *Appl Environ Microbiol* **74**: 2841–2851.

Sciutteri, V., Smedile, F., Vizzini, S., Mazzola, A., and Vetriani, C. (2022) Microbial biofilms along a geochemical gradient at the shallow-water hydrothermal system of Vulcano Island, Mediterranean Sea. *Front Microbiol* **13**: 1–13.

Seaver, L.C. and Imlay, J.A. (2001) Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in *Escherichia coli*. *J Bacteriol* **183**: 7173–7181.

Semrau, J.D. (2011) Current knowledge of microbial community structures in landfills and its cover soils. *Appl Microbiol Biotechnol* **89**: 961–969.

Serra, D.O., Richter, A.M., Klauck, G., Mika, F., and Hengge, R. (2013) Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. *MBio* **4**: e00103-13.

Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., *et al.* (2020) DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**: 8883–8900.

Shakeri Yekta, S., Liu, T., Axelsson Bjerg, M., Šafarič, L., Karlsson, A., Björn, A., and Schnürer, A. (2019) Sulfide level in municipal sludge digesters affects microbial community response to long-chain fatty acid loads. *Biotechnol Biofuels* **12**: 259.

Sharon, I. and Banfield, J.F. (2013) Genomes from metagenomics. *Science* **342**: 1057–1058.

Sheng, Y., Abreu, I.A., Cabelli, D.E., Maroney, M.J., Miller, A.F., Teixeira, M., and Valentine, J.S. (2014) Superoxide dismutases and superoxide reductases. *Chem Rev* **114**: 3854–3918.

Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**: 836–843.

Sigalevich, P., Baev, M. V., Teske, A., and Cohen, Y. (2000) Sulfate reduction and possible aerobic metabolism of the sulfate-reducing bacterium *Desulfovibrio oxyclinae* in a chemostat coculture with *Marinobacter* sp. strain MB under exposure to increasing oxygen concentrations. *Appl Environ Microbiol* **66**: 5013–5018.

Slack, R.J., Gronow, J.R., and Voulvoulis, N. (2005) Household hazardous waste in municipal landfills: Contaminants in leachate. *Sci Total Environ* **337**: 119–137.

Slesak, I., Slesak, H., and Kruk, J. (2012) Oxygen and hydrogen peroxide in the early evolution of life on earth: *In silico* comparative analysis of biochemical pathways. *Astrobiology* **12**: 775–784.

Solli, L., Håvelsrud, O.E., Horn, S.J., and Rike, A.G. (2014) A metagenomic study of the microbial communities in four parallel biogas reactors. *Biotechnol Biofuels* **7**: 146.

Song, L., Wang, Y., Tang, W., and Lei, Y. (2015a) Archaeal community diversity in municipal waste landfill sites. *Appl Microbiol Biotechnol* **99**: 6125–6137.

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

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

Spector, A., Yan, G.Z., Huang, R.R., McDermott, M.J., Gascoyne, P.R., and Pigiet, V. (1988) The effect of H_2O_2 upon thioredoxin-enriched lens epithelial cells. *J Biol Chem* **263**: 4984–90.

Spokas, K., Bogner, J., Chanton, J.P., Morcet, M., Aran, C., Graff, C., *et al.* (2006) Methane mass balance at three landfill sites: What is the efficiency of capture by gas collection systems? *Waste Manag* **26**: 516–525.

Staley, B.F., de los Reyes, F.L., Wang, L., and Barlaz, M.A. (2018) Microbial ecological succession during municipal solid waste decomposition. *Appl Microbiol Biotechnol* **102**: 5731–5740.

Staley, B.F., Xu, F., Cowie, S.J., Barlaz, M.A., and Hater, G.R. (2006) Release of trace organic compounds during the decomposition of municipal solid waste components. *Environ Sci Technol* **40**: 5984–5991.

Stamps, B.W., Lyles, C.N., Suflita, J.M., Masoner, J.R., Cozzarelli, I.M., Kolpin, D.W., and Stevenson, B.S. (2016) Municipal solid waste landfills harbor distinct microbiomes. *Front Microbiol* **7**: 534.

Stokke, R., Dahle, H., Roalkvam, I., Wissuwa, J., Daae, F.L., Tooming-Klunderud, A., *et al.* (2015) Functional interactions among filamentous *Epsilonproteobacteria* and *Bacteroidetes* in a deep-sea hydrothermal vent biofilm. *Environ Microbiol* **17**: 4063–4077.

Stolze, Y., Bremges, A., Maus, I., Pühler, A., Sczyrba, A., and Schlüter, A. (2018) Targeted *in situ* metatranscriptomics for selected taxa from mesophilic and thermophilic biogas plants. *Microb Biotechnol* **11**: 667–679.

Stolze, Y., Bremges, A., Rumming, M., Henke, C., Maus, I., Pühler, A., *et al.* (2016) Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. *Biotechnol Biofuels* **9**: 156.

Sun, L., Pope, P.B., Eijsink, V.G.H., and Schnürer, A. (2015) Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure. *Microb Biotechnol* **8**: 815–827.

Sun, W., Sun, M., and Barlaz, M.A. (2016) Characterizing the biotransformation of sulfur-containing wastes in simulated landfill reactors. *Waste Manag* **53**: 82–91.

Suominen, S., Dombrowski, N., Sinninghe Damsté, J.S., and Villanueva, L. (2021) A diverse uncultivated microbial community is responsible for organic matter degradation in the Black Sea sulphidic zone. *Environ Microbiol* **23**: 2709–2728.

Takenaka, S., Koshiya, J., Okugawa, S., Takata, A., Murakami, S., and Aoki, K. (2011) Fe-superoxide dismutase and 2-hydroxy-1,4-benzoquinone reductase preclude the auto-oxidation step in 4-aminophenol metabolism by *Burkholderia* sp. strain AK-5. *Biodegradation* **22**: 1–11.

Tang, W., Wang, Y., Lei, Y., and Song, L. (2016) Methanogen communities in a municipal landfill complex in China. *FEMS Microbiol Lett* **363**: 75.

Tarnovetskii, I.Y., Merkel, A.Y., and Pimenov, N. V. (2019) Analysis of cultured methanogenic Archaea from the Tarkhankut Peninsula coastal methane seeps. *Microbiology* **88**: 681–688.

Tatusov, R.L. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Theuerl, S., Klang, J., Heiermann, M., and De Vrieze, J. (2018) Marker microbiome clusters are determined by operational parameters and specific key taxa combinations in anaerobic digestion. *Bioresour Technol* **263**: 128–135.

Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* **2**: 3.

Toth, C.R.A. and Gieg, L.M. (2018) Time course-dependent methanogenic crude oil biodegradation: Dynamics of fumarate addition metabolites, biodegradative genes, and microbial community composition. *Front Microbiol* **8**: 2610.

Touchon, M., Cury, J., Yoon, E.J., Krizova, L., Cerqueira, G.C., Murphy, C., *et al.* (2014) The genomic diversification of the whole *Acinetobacter* genus: Origins, mechanisms, and consequences. *Genome Biol Evol* **6**: 2866–2882.

Townsend, T.G., Powell, J., Jain, P., Xu, Q., Tolaymat, T., and Reinhart, D. (2015) Sustainable practices for landfill design and operation, New York, NY: Springer, New York, NY.

Triggs-Raine, B.L., Doble, B.W., Mulvey, M.R., Sorby, P.A., and Loewen, P.C. (1988) Nucleotide sequence of *katG*, encoding catalase HPI of *Escherichia coli*. *J Bacteriol* **170**: 4415–4419.

Tully, B.J., Wheat, C.G., Glazer, B.T., and Huber, J.A. (2018) A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *ISME J* **12**: 1–16.

Tyagi, P., Dharmaraja, A.T., Bhaskar, A., Chakrapani, H., and Singh, A. (2015) *Mycobacterium tuberculosis* has diminished capacity to counteract redox stress induced by elevated levels of endogenous superoxide. *Free Radic Biol Med* **84**: 344–354.

Umezawa, K., Kojima, H., Kato, Y., and Fukui, M. (2021) *Dissulfurispira thermophila* gen. nov., sp. nov., a thermophilic chemolithoautotroph growing by sulfur disproportionation, and proposal of novel taxa in the phylum *Nitrospirota* to reclassify the genus *Thermodesulfovibrio*. *Syst Appl Microbiol* **44**: 126184.

US Environmental Protection Agency, A. (2021) Inventory of U.S. Greenhouse gas emissions and sinks: 1990-2019. *Fed Regist* **79**: 10143–10144.

US Environmental Protection Agency, C. (2017) LFG Energy Project Development Handbook, Chapter 1: Landfill Gas Energy Basics. 1–15.

Valette, O., Tran, T.T.T., Cavazza, C., Caudeville, E., Brasseur, G., Dolla, A., *et al.* (2017) Biochemical function, molecular structure and evolution of an atypical thioredoxin reductase from *Desulfovibrio vulgaris*. *Front Microbiol* **8**: 1855.

Vido, K., Diemer, H., Van Dorsselaer, A., Leize, E., Juillard, V., Gruss, A., and Gaudu, P. (2005) Roles of thioredoxin reductase during the aerobic life of *Lactococcus lactis*. *J Bacteriol* **187**: 601–610.

Vigneron, A., Head, I.M., and Tsesmetzis, N. (2018) Damage to offshore production facilities by corrosive microbial biofilms. *Appl Microbiol Biotechnol* **102**: 2525–2533.

Viljakainen, V.R. and Hug, L.A. (2021) New approaches for the characterization of plastic-associated microbial communities and the discovery of plastic-degrading microorganisms and enzymes. *Comput Struct Biotechnol J* **19**: 6191–6200.

Villanueva, L., von Meijenfeldt, F.A.B., Westbye, A.B., Yadav, S., Hopmans, E.C., Dutilh, B.E., and Damsté, J.S.S. (2021) Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *ISME J* **15**: 168–182.

Van Vliet, A.H. and Ketley, J.M. (2001) Pathogenesis of enteric *Campylobacter infection*. *J Appl Microbiol* **90**: 45S-56S.

Van Vliet, A.H.M., Ketley, J.M., Park, S.F., and Penn, C.W. (2002) The role of iron in *Campylobacter* gene regulation, metabolism and oxidative stress defense. *FEMS Microbiol Rev* **26**: 173–186.

Waite, D.W., Chuvochina, M., Pelikan, C., Parks, D.H., Yilmaz, P., Wagner, M., *et al.* (2020) Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia,* and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int J Syst Evol Microbiol* **70**: 5972–6016.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., *et al.* (2016) Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**: e00009-15.

Warnecke, F. and Hugenholtz, P. (2007) Building on basic metagenomics with complementary technologies. *Genome Biol* **8**: 231.

Wasserfallen, A., Nölling, J., Pfister, P., Reeve, J., and De Macario, E.C. (2000) Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. *Int J Syst Evol Microbiol* **50**: 43–53.

Webster, N.S., Smith, L.D., Heyward, A.J., Watts, J.E.M., Webb, R.I., Blackall, L.L., and Negri, A.P. (2004) Metamorphosis of a scleractinian coral in response to microbial biofilms. *Appl Environ Microbiol* **70**: 1213–1221.

Westerholm, M., Crauwels, S., Houtmeyers, S., Meerbergen, K., Van Geel, M., Lievens, B., and Appels, L. (2016) Microbial community dynamics linked to enhanced substrate availability and biogas production of electrokinetically pre-treated waste activated sludge. *Bioresour Technol* **218**: 761–770.

Whalan, S. and Webster, N.S. (2014) Sponge larval settlement cues: The role of microbial biofilms in a warming ocean. *Sci Rep* **4**: 1–5.

Wilkins, L.G.E., Ettinger, C.L., Jospin, G., and Eisen, J.A. (2019) Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci Rep* **9**: 1–15.

Williams, T.J., Allen, M.A., Berengut, J.F., and Cavicchioli, R. (2021) Shedding light on microbial "dark matter": Insights into novel Cloacimonadota and Omnitrophota from an Antarctic lake. *Front Microbiol* **12**: 741077.

Wilson, A., Boulay, C., Wilde, A., Kerfeld, C.A., and Kirilovsky, D. (2007) Light-induced energy dissipation in iron-starved Cyanobacteria: roles of OCP and IsiA proteins. *Plant Cell* **19**: 656–672.

Wu, Y., Cai, P., Jing, X., Niu, X., Ji, D., Ashry, N.M., *et al.* (2019) Soil biofilm formation enhances microbial community diversity and metabolic activity. *Environ Int* **132**: 105116.

Wu, Y.W., Simmons, B.A., and Singer, S.W. (2016) MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607.

Xia, Y., Yang, C., and Zhang, T. (2018) Microbial effects of part-stream low-frequency ultrasonic pretreatment on sludge anaerobic digestion as revealed by high-throughput sequencing-based metagenomics and metatranscriptomics. *Biotechnol Biofuels* **11**: 47.

Yang, C.H., Su, P.W., Moi, S.H., and Chuang, L.Y. (2019) Biofilm formation in *Acinetobacter baumannii*: Genotype-phenotype correlation. *Molecules* **24**: 1849.

Yao, D., Kieselbach, T., Komenda, J., Promnares, K., Hernández Prieto, M.A., Tichy, M., *et al.*

(2007) Localization of the small CAB-like proteins in photosystem II. *J Biol Chem* **282**: 267–276.

Yoshikawa, M., Zhang, M., and Toyota, K. (2017) Integrated anaerobic-aerobic biodegradation of multiple contaminants including chlorinated ethylenes, benzene, toluene, and dichloromethane. *Water Air Soil Pollut* **228**: 1–13.

Yuan, K., Li, S., and Zhong, F. (2020) Treatment of coking wastewater in biofilm-based bioaugmentation process: Biofilm formation and microbial community analysis. *J Hazard Mater* **400**: 123117.

Zaborowska, M., Wyszkowska, J., and Borowik, A. (2020) Soil microbiome response to contamination with bisphenol A, bisphenol F and bisphenol S. *Int J Mol Sci* **21**: 3529.

Zapata, C., Paillavil, B., Chávez, R., Álamos, P., and Levicán, G. (2017) Cytochrome *c* peroxidase (CcP) is a molecular determinant of the oxidative stress response in the extreme acidophilic *Leptospirillum* sp. CF-1. *FEMS Microbiol Ecol* **93**: fix001.

Zeller, T. and Klug, G. (2006) Thioredoxins in bacteria: Functions in oxidative stress response and regulation of thioredoxin genes. *Naturwissenschaften* **93**: 259–266.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., *et al.* (2018) DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95–W101.

Zhang, W., Ding, W., Li, Y.X., Tam, C., Bougouffa, S., Wang, R., *et al.* (2019) Marine biofilms constitute a bank of hidden microbial diversity and functional potential. *Nat Commun* **10**: 1–10.

Zhang, W., Ding, W., Yang, B., Tian, R., Gu, S., Luo, H., and Qian, P.Y. (2016) Genomic and transcriptomic evidence for carbohydrate consumption among microorganisms in a cold seep brine pool. *Front Microbiol* **7**: 1825.

Zhang, Y., Heym, B., Allen, B., Young, D., and Cole, S. (1992) The catalase—peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* **358**: 591–593.

Zhao, S., Ye, Z., and Stanton, R. (2020) Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* **26**: 903–909.

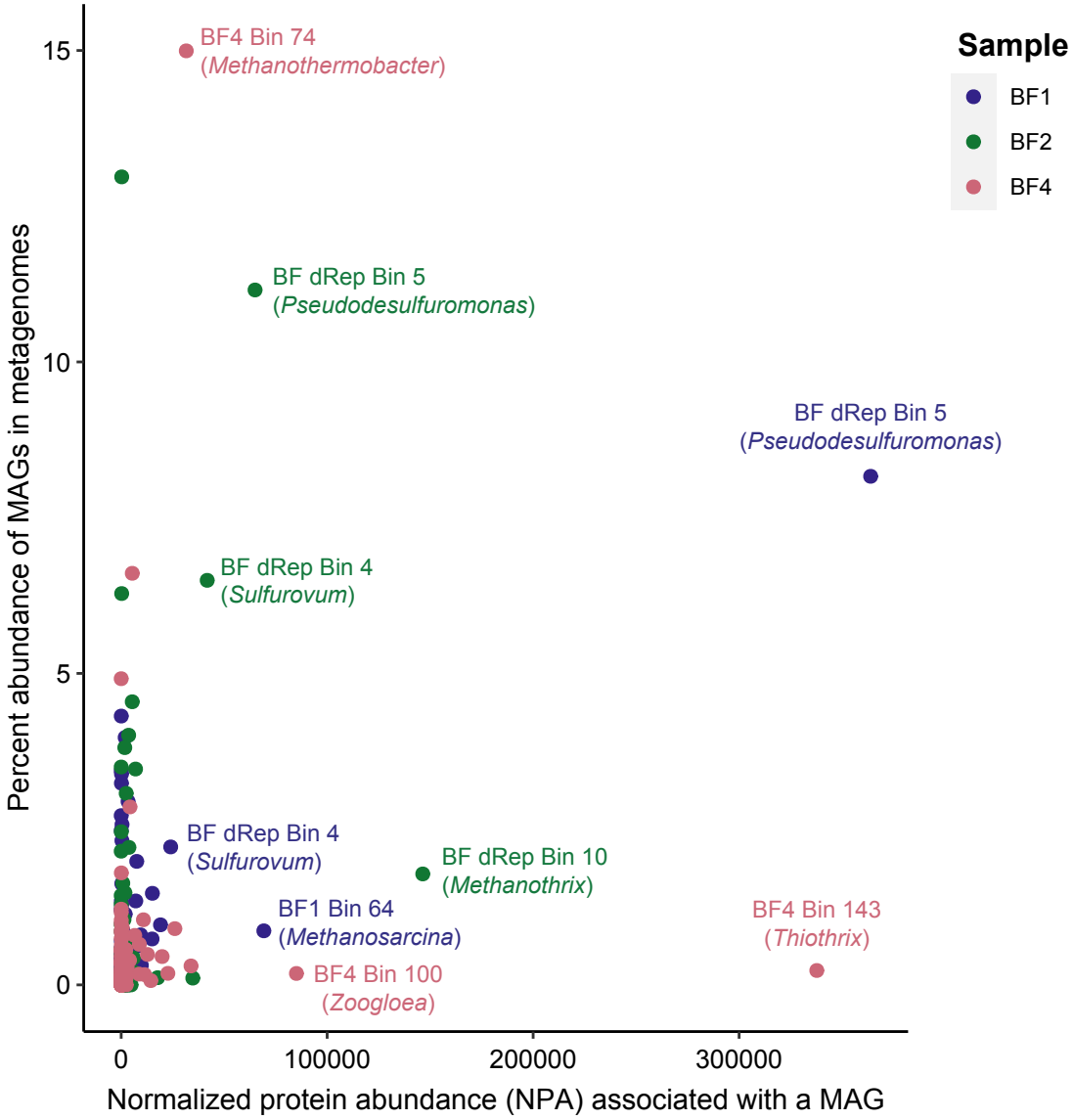
Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., *et al.* (2022) METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**: 1–22.

Zotta, T., Parente, E., and Ricciardi, A. (2017) Aerobic metabolism in the genus *Lactobacillus*: impact on stress response and potential applications in the food industry. *J Appl Microbiol* **122**: 857–869.

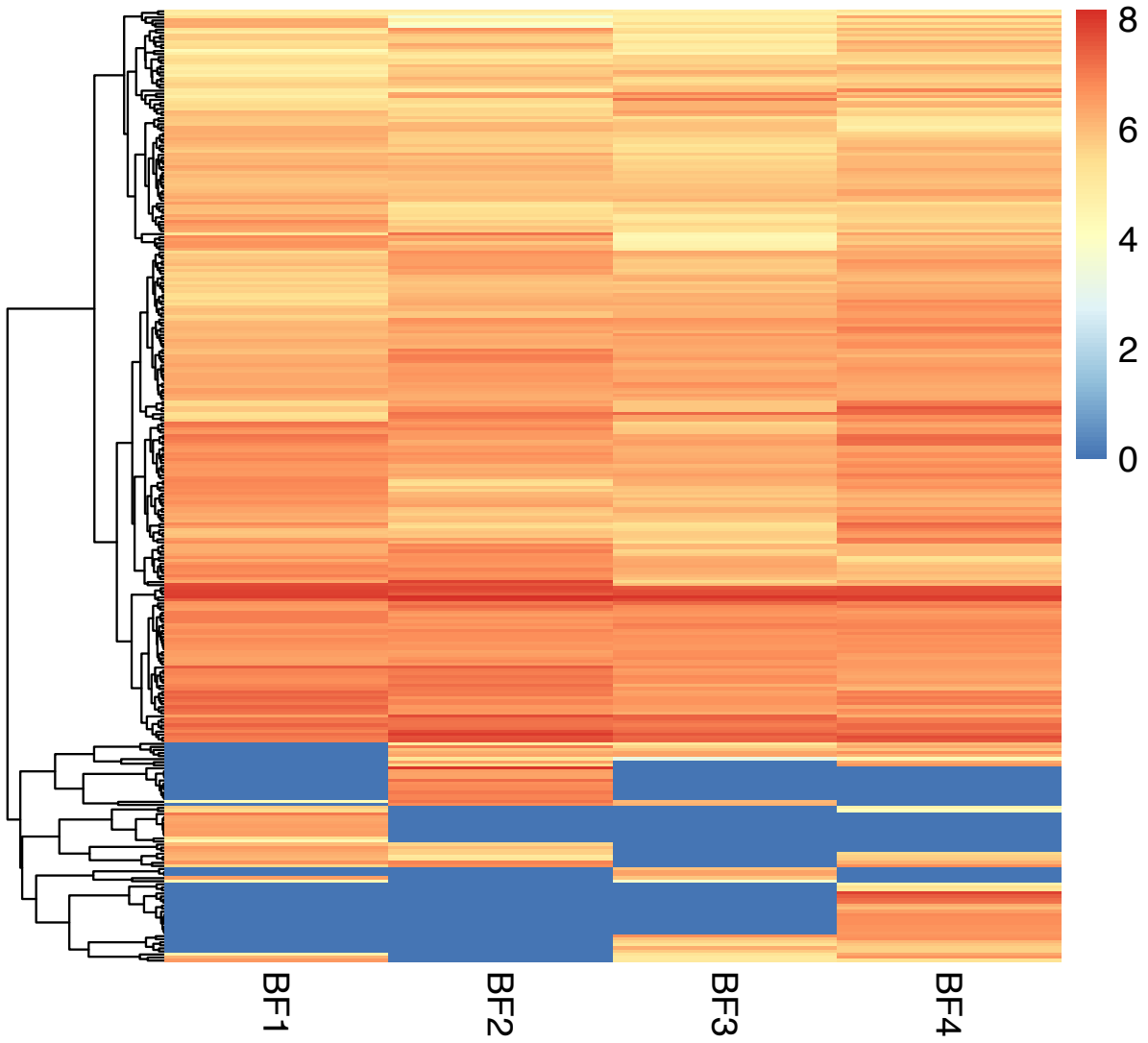
Zotta, T., Ricciardi, A., Ianniello, R.G., Parente, E., Reale, A., Rossi, F., *et al.* (2014) Assessment of aerobic and respiratory growth in the *Lactobacillus casei* group. *PLoS One* **9**: e99189.

Appendix A

Supplementary Material – Chapter 2



Supplementary Figure A1: Relationship between relative abundance of MAGs with normalized protein abundance (NPA) associated with MAGs. Each dot represents a MAG, and dots are colored to denote biofilm sample. MAGs that contained the highest proportion of the NPA are labelled with the MAG ID along with genus-level classification in brackets.



Supplementary Figure A2: Heatmap depicting the level of detection for only unknown metabolites. Abundance values were log10 transformed, and hierarchical clustering and visualization were completed with the "pheatmap" package in R.

Supplemental Table A1: Quality statistics for metagenome-assembled genomes (MAGs) from biofilm metagenomes. MAGs that were dereplicated have all genomes that belonged to the group listed with the genome being used as a representative listed first and denoted with a star.

MAG ID	MAGs combined with DRep	Phylum	Genome size (bp)	Scaffolds	N50 (scaffolds)	Completeness (%)	Contamination (%)
BF dRep Bin 1	BF1 Bin 12*	Bipolaricaulota	1372617	97	22337	88.1	3.0
	BF4 Bin 46						
BF dRep Bin 10	BF1 Bin 79*	Halobacteriota	2446556	253	14575	91.8	1.7
	BF2 Bin 20						
BF dRep Bin 11	BF1 Bin 80*	Desulfobacterota	2878564	41	106200	99.4	0.0
	BF4 Bin 4						
BF dRep Bin 12	BF1 Bin 82*	Cloacimonadota	2812189	98	61416	97.8	2.8
	BF4 Bin 35						
BF dRep Bin 13	BF1 Bin 95*	Firmicutes	1682454	231	8885	84.7	0.0
	BF4 Bin 21						
BF dRep Bin 14	BF2 Bin 1*	Halobacteriota	1605377	216	9530	85.8	0.7
	BF1 Bin 62						
BF dRep Bin 15	BF2 Bin 11*	Campylobacterota	1788694	39	96273	98.8	0.6
	BF1 Bin 3						
BF dRep Bin 16	BF2 Bin 21*	Thermotogota	1622400	191	12339	87.6	0.0
	BF1 Bin 41						
BF dRep Bin 17	BF2 Bin 23*	Bacteroidota	3295802	86	67409	95.2	2.2
	BF4 Bin 11						
BF dRep Bin 18	BF2 Bin 30*	Campylobacterota	2408291	142	22917	97.1	0.4
	BF1 Bin 66						

BF dRep Bin 19	BF2 Bin 41* BF4 Bin 90	Atribacterota	993323	180	6154	84.8	0.0
BF dRep Bin 2	BF1 Bin 17* BF4 Bin 48	Halobacteriota	6536225	743	13941	99.5	8.7
BF dRep Bin 20	BF2 Bin 49* BF1 Bin 106	Desulfobacterota	2098889	300	8601	81.7	1.9
BF dRep Bin 21	BF2 Bin 6* BF1 Bin 98	Bacteroidota	2105722	41	97754	95.1	1.7
BF dRep Bin 22	BF4 Bin 102* BF1 Bin 72 BF2 Bin 18	Cloacimonadota	2020177	129	26764	98.0	1.1
BF dRep Bin 23	BF4 Bin 108* BF1 Bin 96	Spirochaetota	1842184	162	16243	89.6	0.0
BF dRep Bin 24	BF4 Bin 109* BF1 Bin 51	Desulfobacterota	3016419	59	78193	96.8	1.3
BF dRep Bin 25	BF4 Bin 111* BF1 Bin 33	Planctomycetota	4973829	184	51055	100.0	2.3
BF dRep Bin 26	BF4 Bin 113* BF1 Bin 103	Desulfobacterota	3215630	71	86124	98.8	0.7
BF dRep Bin 27	BF4 Bin 116* BF1 Bin 23	Desulfobacterota	3248180	148	34935	97.4	2.6
BF dRep Bin 28	BF4 Bin 146* BF4 Bin 146	Cloacimonadota	2359966	107	36893	95.5	1.1
BF dRep Bin 29	BF4 Bin 159* BF1 Bin 77 BF2 Bin 12	Bacteroidota	1923930	36	85017	93.0	0.8

BF dRep Bin 3	BF1 Bin 38*	Desulfobacterota	2560658	41	125506	98.1	0.7
	BF2 Bin 46						
BF dRep Bin 30	BF4 Bin 161*	Bacteroidota	3740389	235	24794	93.1	0.8
	BF1 Bin 59						
	BF2 Bin 38						
BF dRep Bin 31	BF4 Bin 163*	Bacteroidota	3742656	184	31086	95.7	1.3
	BF1 Bin 15						
BF dRep Bin 32	BF4 Bin 182*	Acidobacteriota	2883445	44	106051	94.7	5.1
	BF2 Bin 29						
BF dRep Bin 33	BF4 Bin 186*	Firmicutes	1125556	107	15651	80.7	1.3
	BF1 Bin 50						
BF dRep Bin 34	BF4 Bin 37*	Cloacimonadota	2096142	151	21415	94.0	1.1
	BF1 Bin 29						
BF dRep Bin 35	BF4 Bin 40*	Firmicutes	1562975	134	19604	88.7	1.8
	BF1 Bin 34						
BF dRep Bin 36	BF4 Bin 42*	Campylobacterota	2320125	51	107918	99.2	0.8
	BF1 Bin 69						
BF dRep Bin 37	BF4 Bin 43*	Firmicutes	1422459	181	10439	89.8	3.8
	BF1 Bin 28						
BF dRep Bin 38	BF4 Bin 44*	Firmicutes	1434980	128	16277	94.7	1.3
	BF1 Bin 16						
BF dRep Bin 39	BF4 Bin 45*	Cloacimonadota	2167308	89	53453	98.8	2.2
	BF1 Bin 89						
BF dRep Bin 4	BF1 Bin 45*	Campylobacterota	2219040	31	99439	99.6	2.5
	BF2 Bin 10						
BF dRep Bin 40	BF4 Bin 47*	Cloacimonadota	2591264	157	29866	98.4	5.6

	BF1 Bin 92						
BF dRep Bin 41	BF4 Bin 7*	Patescibacteria	897815	88	13514	75.9	0.0
	BF1 Bin 37						
BF dRep Bin 42	BF4 Bin 89*	Bacteroidota	1644182	69	33527	91.3	1.5
	BF1 Bin 22						
	BF2 Bin 45						
BF dRep Bin 5	BF1 Bin 53*	Desulfobacterota	2261941	50	79672	88.4	1.3
	BF2 Bin 35						
BF dRep Bin 6	BF1 Bin 55*	Firmicutes	1083512	30	91610	98.9	0.0
	BF2 Bin 2						
	BF4 Bin 39						
BF dRep Bin 7	BF1 Bin 60*	Firmicutes	1251962	49	37558	100.0	0.0
	BF2 Bin 28						
	BF4 Bin 41						
BF dRep Bin 8	BF1 Bin 73*	Bacteroidota	2165684	115	27583	96.4	0.7
	BF2 Bin 40						
BF dRep Bin 9	BF1 Bin 75*	Proteobacteria	3036842	159	47069	97.0	0.6
	BF2 Bin 19						
BF1 Bin 1		Bacteroidota	3121686	160	28898	96.2	1.8
BF1 Bin 10		Desulfobacterota	3724931	104	60920	99.4	0.4
BF1 Bin 100		Desulfobacterota	3424068	285	21913	95.7	7.2
BF1 Bin 101		Spirochaetota	1736552	263	9192	81.4	2.9
BF1 Bin 102		Firmicutes	761295	146	5669	83.9	0.7
BF1 Bin 104		Bacteroidota	3057550	303	14287	97.3	3.0
BF1 Bin 105		Actinobacteriota	2316003	241	14549	85.4	4.3
BF1 Bin 11		Proteobacteria	3165859	45	174472	98.3	0.9

BF1 Bin 13	Firmicutes	2227771	313	8513	88.1	3.1
BF1 Bin 14	Firmicutes	1110881	147	9030	91.6	3.9
BF1 Bin 18	Campylobacterota	2256946	210	16913	98.0	4.1
BF1 Bin 19	Desulfobacterota	2202883	204	16015	94.8	0.7
BF1 Bin 2	Thermoplasmatota	1268126	55	50665	98.8	0.0
BF1 Bin 20	Proteobacteria	2413714	217	16567	95.9	1.7
BF1 Bin 21	Spirochaetota	4046635	96	65714	93.3	3.5
BF1 Bin 24	Bacteroidota	3314849	280	19023	95.7	9.1
BF1 Bin 25	Bacteroidota	4545005	458	14736	98.3	4.9
BF1 Bin 26	Desulfobacterota	2532776	135	52239	92.9	2.6
BF1 Bin 27	Firmicutes	2937486	95	85796	99.3	2.3
BF1 Bin 30	Bacteroidota	3285121	247	20491	95.3	6.4
BF1 Bin 31	Desulfobacterota	4752372	271	37586	98.7	7.4
BF1 Bin 32	Desulfobacterota	2735731	169	33238	96.1	1.9
BF1 Bin 35	Spirochaetota	2085268	356	6572	87.4	2.3
BF1 Bin 36	Desulfobacterota	3023142	71	88839	99.0	1.3
BF1 Bin 39	Firmicutes	1208762	200	7011	84.7	2.9
BF1 Bin 4	Campylobacterota	2621000	265	13048	93.4	6.0
BF1 Bin 40	Cloacimonadota	3027221	228	24443	96.7	1.7
BF1 Bin 42	Proteobacteria	3031906	203	24666	96.2	1.9
BF1 Bin 43	Bacteroidota	1602125	248	7739	80.4	2.6
BF1 Bin 44	Firmicutes	2570172	46	126307	100.0	4.9
BF1 Bin 46	Firmicutes	1996376	75	86789	93.3	5.3
BF1 Bin 47	Firmicutes	2835217	402	9061	88.7	6.0
BF1 Bin 49	Acidobacteriota	1620313	298	6074	76.9	2.1
BF1 Bin 5	Desulfobacterota	3603415	96	63020	96.8	1.4

BF1 Bin 52	Synergistota	1630752	236	7972	94.9	0.0
BF1 Bin 54	Firmicutes	1502812	49	350035	98.9	2.3
BF1 Bin 56	Thermoplasmatota	1341953	62	47127	96.4	1.6
BF1 Bin 57	Proteobacteria	3321613	61	111525	97.7	1.8
BF1 Bin 58	Halobacteriota	1690072	134	18468	90.5	2.0
BF1 Bin 6	Proteobacteria	2640633	331	9993	83.5	0.6
BF1 Bin 61	Campylobacterota	1610514	201	11285	89.5	1.7
BF1 Bin 63	Desulfobacterota	3566720	146	48052	94.5	1.2
BF1 Bin 64	Halobacteriota	3530267	359	14685	91.5	2.0
BF1 Bin 65	Desulfobacterota	3500553	44	171143	98.8	0.8
BF1 Bin 67	Campylobacterota	2010641	185	15398	92.5	3.3
BF1 Bin 68	Desulfobacterota	2892246	52	169334	99.4	4.4
BF1 Bin 7	Bacteroidota	2862570	186	22635	92.6	4.0
BF1 Bin 70	Thermoplasmatota	1432563	178	10639	88.9	1.8
BF1 Bin 71	Campylobacterota	1706702	95	29249	96.4	7.2
BF1 Bin 74	Campylobacterota	1965470	23	199257	99.2	0.2
BF1 Bin 76	Campylobacterota	2510936	101	51188	98.4	1.3
BF1 Bin 78	Cloacimonadota	2641007	141	50456	95.0	6.0
BF1 Bin 8	Desulfobacterota	3048658	173	29239	96.1	8.2
BF1 Bin 81	Desulfobacterota	3127451	82	66525	97.4	1.3
BF1 Bin 83	Bacteroidota	2965205	285	17545	96.2	1.8
BF1 Bin 84	Firmicutes	1683505	111	26464	98.7	7.3
BF1 Bin 85	Bacteroidota	3807482	203	34468	99.1	4.3
BF1 Bin 86	Desulfobacterota	2306123	83	67315	87.7	0.0
BF1 Bin 87	Desulfobacterota	3012050	30	276757	99.4	1.2
BF1 Bin 88	Cloacimonadota	3035819	154	39069	92.0	5.0

BF1 Bin 9	Bacteroidota	3681070	208	42482	96.7	7.1
BF1 Bin 90	Firmicutes	2197453	359	6697	81.6	5.8
BF1 Bin 91	Bacteroidota	5139265	374	25345	99.5	5.8
BF1 Bin 93	Desulfobacterota	3844761	121	75505	99.4	1.6
BF1 Bin 94	Desulfobacterota	2638986	31	173525	99.4	0.0
BF1 Bin 97	Verrucomicrobiota	3059299	273	15540	87.6	2.1
BF1 Bin 99	Proteobacteria	3839739	254	32313	93.1	5.9
BF2 Bin 13	Bacteroidota	1860588	214	13523	87.1	3.3
BF2 Bin 14	Proteobacteria	4998778	118	72346	96.7	0.8
BF2 Bin 15	Bacteroidota	2011771	414	5264	71.2	1.1
BF2 Bin 16	Chloroflexota	2036483	329	7167	82.1	3.9
BF2 Bin 17	Firmicutes	639836	139	4973	75.6	0.0
BF2 Bin 22	Atribacterota	1438749	309	4565	76.3	0.0
BF2 Bin 24	Proteobacteria	2635263	139	31628	98.1	1.5
BF2 Bin 25	Campylobacterota	2200321	92	41876	98.0	1.8
BF2 Bin 26	Campylobacterota	1912760	199	13721	95.7	2.5
BF2 Bin 27	Campylobacterota	2626796	117	41122	98.2	4.7
BF2 Bin 3	Firmicutes	1129392	57	53540	92.1	1.5
BF2 Bin 31	Cloacimonadota	2757324	360	9202	87.9	2.2
BF2 Bin 32	Proteobacteria	2066704	283	8978	88.9	2.2
BF2 Bin 33	Bacteroidota	2432597	302	10315	92.0	1.3
BF2 Bin 34	Firmicutes	1287619	79	26345	98.0	0.1
BF2 Bin 36	Campylobacterota	2119891	43	114119	99.2	1.0
BF2 Bin 37	Actinobacteriota	2315821	344	7852	88.9	0.7
BF2 Bin 39	Halobacteriota	2798024	531	5718	71.2	1.3
BF2 Bin 4	Cloacimonadota	2645439	164	25460	96.7	2.2

BF2 Bin 42	Verrucomicrobiota	2461822	218	15043	86.7	1.4
BF2 Bin 43	Synergistota	1339273	245	6126	84.0	0.0
BF2 Bin 44	Cloacimonadota	1593826	100	27264	94.5	0.6
BF2 Bin 47	Firmicutes	623690	110	7079	72.5	2.3
BF2 Bin 48	Verrucomicrobiota	2473619	303	9781	83.8	1.4
BF2 Bin 5	Firmicutes	1331462	184	8494	80.7	0.0
BF2 Bin 50	Planctomycetota	8687633	432	35541	97.7	2.9
BF2 Bin 51	Actinobacteriota	979248	183	5563	74.8	0.9
BF2 Bin 7	Firmicutes	888203	161	5914	75.9	0.0
BF2 Bin 8	Proteobacteria	1698590	382	4509	70.5	1.2
BF2 Bin 9	Campylobacterota	1940249	314	6812	76.3	4.7
BF4 Bin 1	Bacteroidota	4118373	256	32000	96.0	4.8
BF4 Bin 10	Bacteroidota	3116076	197	27540	94.4	0.5
BF4 Bin 100	Proteobacteria	4785374	211	43121	96.2	2.4
BF4 Bin 101	Firmicutes	1377741	83	26469	93.3	1.3
BF4 Bin 103	Bacteroidota	3025295	470	7271	82.7	5.6
BF4 Bin 104	Bacteroidota	2599527	100	51243	90.3	1.7
BF4 Bin 105	Proteobacteria	3364882	157	35384	92.0	2.5
BF4 Bin 106	Bacteroidota	4769188	142	75529	96.2	2.7
BF4 Bin 107	Bacteroidota	3504506	117	56028	97.1	2.9
BF4 Bin 110	Proteobacteria	4069588	138	50911	96.1	3.2
BF4 Bin 112	Proteobacteria	2067235	13	314035	95.4	0.9
BF4 Bin 114	Actinobacteriota	1782256	145	18064	95.2	4.7
BF4 Bin 115	Moduliflexota	6574552	610	14091	86.8	5.1
BF4 Bin 117	Patescibacteria	837040	75	15785	70.9	0.2
BF4 Bin 118	Bacteroidota	1821615	149	14798	93.3	0.5

BF4 Bin 119	Firmicutes	1408119	103	24113	96.2	0.4
BF4 Bin 12	Proteobacteria	3001139	75	78470	96.8	0.6
BF4 Bin 120	Proteobacteria	3185725	500	7245	71.4	8.5
BF4 Bin 121	Desulfobacterota	4486996	893	5379	79.2	5.6
BF4 Bin 122	Firmicutes	1015589	166	6649	86.7	1.5
BF4 Bin 123	Myxococcota	7406059	336	39436	86.9	4.1
BF4 Bin 124	Proteobacteria	3472832	193	27654	93.8	2.0
BF4 Bin 125	Proteobacteria	3775319	173	45343	90.1	1.2
BF4 Bin 126	Bacteroidota	3138883	152	31598	92.2	0.6
BF4 Bin 127	Krumholzibacteriota	3727882	52	211552	98.9	2.2
BF4 Bin 128	Desulfobacterota	3006217	253	17172	89.3	4.5
BF4 Bin 129	Bacteroidota	3770279	165	38155	93.7	2.4
BF4 Bin 13	Hydrogenedentota	4939003	600	10986	98.9	3.3
BF4 Bin 130	Proteobacteria	2503519	248	13037	95.2	0.7
BF4 Bin 131	Methanobacteriota	1413707	66	30665	91.5	2.7
BF4 Bin 132	Proteobacteria	2894558	301	12348	93.0	3.3
BF4 Bin 133	Proteobacteria	2960146	381	9545	88.9	4.6
BF4 Bin 134	Myxococcota	6933659	848	10370	83.0	6.2
BF4 Bin 135	Chloroflexota	2667557	442	7039	74.1	0.5
BF4 Bin 136	Firmicutes	1127427	160	9859	84.0	1.6
BF4 Bin 137	Verrucomicrobiota	4984879	871	6547	74.6	4.8
BF4 Bin 138	Bacteroidota	1729642	289	6681	74.3	1.1
BF4 Bin 139	Proteobacteria	2650485	144	33354	86.9	3.5
BF4 Bin 14	Proteobacteria	3531316	83	70444	98.2	2.1
BF4 Bin 140	Chloroflexota	1061173	174	6338	86.6	0.5
BF4 Bin 141	Proteobacteria	1941590	338	6601	77.6	3.7

BF4 Bin 142	Patescibacteria	1195214	9	673799	73.3	2.0
BF4 Bin 143	Proteobacteria	3504809	18	277498	99.5	0.6
BF4 Bin 144	Hydrogenedentota	5893695	906	7666	89.7	7.7
BF4 Bin 145	Delongbacteria	2496241	54	116418	97.8	0.0
BF4 Bin 147	Chloroflexota	3567577	185	27985	88.0	7.5
BF4 Bin 148	Cloacimonadota	2396250	176	21452	96.6	3.3
BF4 Bin 149	Proteobacteria	3322560	123	40155	98.8	1.8
BF4 Bin 15	Spirochaetota	3340094	375	13606	95.3	1.1
BF4 Bin 150	Campylobacterota	1888727	175	14937	96.9	1.9
BF4 Bin 151	Cloacimonadota	2103950	67	48930	98.9	1.2
BF4 Bin 152	Proteobacteria	1896897	210	12462	87.7	2.4
BF4 Bin 153	Spirochaetota	2172386	322	7656	88.6	1.7
BF4 Bin 154	Planctomycetota	3311368	41	148477	96.6	0.0
BF4 Bin 155	Chloroflexota	3094014	28	146269	90.9	1.3
BF4 Bin 156	Bacteroidota	3255711	278	19011	89.9	7.8
BF4 Bin 157	Desulfobacterota	3703186	353	16623	94.1	1.5
BF4 Bin 158	Spirochaetota	2428697	85	65246	97.7	1.1
BF4 Bin 16	Bacteroidota	2556660	534	4959	77.3	4.1
BF4 Bin 160	Bacteroidota	2275454	107	49634	94.8	3.9
BF4 Bin 162	Bacteroidota	2728482	409	8174	85.1	2.7
BF4 Bin 164	Desulfobacterota	3600384	170	41872	98.1	0.7
BF4 Bin 165	Proteobacteria	2452836	329	10155	71.2	2.9
BF4 Bin 166	Chloroflexota	4974554	679	9096	80.9	3.3
BF4 Bin 167	Bacteroidota	3792091	252	21616	91.5	6.2
BF4 Bin 168	Spirochaetota	3137834	501	7236	80.5	2.7
BF4 Bin 169	Firmicutes	2001707	123	31211	97.6	1.7

BF4 Bin 17	Campylobacterota	2245655	237	13210	94.1	2.7
BF4 Bin 170	Bacteroidota	3045563	82	59582	95.0	2.1
BF4 Bin 171	Desulfobacterota	2377201	18	160190	92.9	1.5
BF4 Bin 172	Bacteroidota	2974990	209	22729	88.8	6.1
BF4 Bin 173	Bacteroidota	3441552	77	120773	98.9	1.9
BF4 Bin 174	Spirochaetota	2863052	105	46047	96.6	1.2
BF4 Bin 175	Patescibacteria	1070801	12	272827	74.3	1.0
BF4 Bin 176	Campylobacterota	2522514	72	67519	99.2	0.6
BF4 Bin 177	Patescibacteria	1295649	121	20593	82.6	0.2
BF4 Bin 178	Bacteroidota	1825809	305	6657	84.8	1.5
BF4 Bin 179	Bacteroidota	2819518	431	8035	78.3	1.4
BF4 Bin 18	Patescibacteria	1027073	21	165992	84.3	2.3
BF4 Bin 180	Bacteroidota	3257398	485	8133	77.3	4.8
BF4 Bin 181	Spirochaetota	2185625	287	9463	88.0	5.9
BF4 Bin 183	Bacteroidota	3241807	64	63681	98.9	1.1
BF4 Bin 184	Proteobacteria	2642213	374	8648	88.1	1.0
BF4 Bin 185	Firmicutes	1748931	33	72971	93.3	0.0
BF4 Bin 187	Firmicutes	2497813	179	20843	90.3	2.9
BF4 Bin 19	Zixibacteria	3728517	35	197852	98.4	3.3
BF4 Bin 2	Planctomycetota	5985068	67	134022	98.9	3.4
BF4 Bin 20	Proteobacteria	3966795	463	12377	98.8	2.7
BF4 Bin 22	Proteobacteria	3937580	489	10142	95.6	5.0
BF4 Bin 23	Cloacimonadota	1899649	336	6161	81.0	1.1
BF4 Bin 24	Bacteroidota	4524349	92	95998	98.4	4.0
BF4 Bin 25	Verrucomicrobiota	3732651	215	25282	92.6	0.7
BF4 Bin 26	Bacteroidota	3598903	280	19568	95.2	3.7

BF4 Bin 27	Chloroflexota	3317553	504	7683	86.3	7.2
BF4 Bin 28	Chloroflexota	3636872	144	39357	90.2	6.4
BF4 Bin 29	Proteobacteria	3921602	87	109387	93.5	3.8
BF4 Bin 3	Desulfobacterota	2983975	551	6138	85.6	9.1
BF4 Bin 30	Proteobacteria	3603534	57	115169	99.7	0.2
BF4 Bin 31	Campylobacterota	2156342	139	87847	98.4	6.9
BF4 Bin 32	Desulfobacterota	2401181	465	5718	74.1	0.9
BF4 Bin 33	Bacteroidota	2937436	495	6497	78.7	5.7
BF4 Bin 34	Desulfobacterota	2258153	509	4553	75.2	6.8
BF4 Bin 36	Bacteroidota	3036205	130	34241	96.5	2.5
BF4 Bin 38	Desulfobacterota	3067374	642	5130	74.4	5.0
BF4 Bin 49	Spirochaetota	2886697	315	13575	97.2	1.1
BF4 Bin 5	Spirochaetota	2570057	420	7303	90.1	7.1
BF4 Bin 50	Desulfobacterota	4623853	46	213408	97.4	0.0
BF4 Bin 51	Firmicutes	1688131	77	34671	98.7	1.3
BF4 Bin 52	Bacteroidota	4458495	92	82586	99.5	2.6
BF4 Bin 53	Bacteroidota	3781614	241	27533	94.8	1.7
BF4 Bin 54	Bacteroidota	2542833	404	7388	85.2	1.9
BF4 Bin 55	Bacteroidota	3117096	565	6096	84.1	8.7
BF4 Bin 56	Spirochaetota	2239850	377	7190	92.6	0.6
BF4 Bin 57	Bacteroidota	2620635	261	14351	84.4	5.4
BF4 Bin 58	Spirochaetota	2811256	273	14719	98.4	5.6
BF4 Bin 59	Chloroflexota	1672267	188	13356	94.2	1.0
BF4 Bin 6	Marinisomatota	1604666	229	8886	70.2	0.1
BF4 Bin 60	Marinisomatota	2054740	261	9912	81.5	0.0
BF4 Bin 61	UBP6	3205908	233	19871	86.3	5.2

BF4 Bin 62	Campylobacterota	1915205	293	8070	78.7	1.4
BF4 Bin 63	Desulfobacterota	1789084	368	5037	77.7	6.0
BF4 Bin 64	Bacteroidota	2006634	98	63071	90.6	0.7
BF4 Bin 65	Spirochaetota	2165121	189	21064	94.2	2.3
BF4 Bin 66	Chloroflexota	5870008	317	30063	95.5	7.5
BF4 Bin 67	Verrucomicrobiota	4227477	610	8660	77.0	4.3
BF4 Bin 68	Patescibacteria	820211	19	99007	71.3	1.0
BF4 Bin 69	Proteobacteria	2020015	36	134053	89.1	2.2
BF4 Bin 70	Patescibacteria	818390	42	38324	76.3	0.0
BF4 Bin 71	Bacteroidota	3475963	155	51274	95.2	4.3
BF4 Bin 72	Aureabacteria	3608127	26	453619	98.9	2.2
BF4 Bin 73	Firmicutes	1704057	298	6507	76.7	8.7
BF4 Bin 74	Methanobacteriota	1509987	48	47793	100.0	0.8
BF4 Bin 75	Bacteroidota	4001767	421	13158	90.6	5.8
BF4 Bin 76	Bacteroidota	2968779	237	20555	94.2	3.9
BF4 Bin 77	Proteobacteria	2794494	82	75973	97.3	6.0
BF4 Bin 78	Campylobacterota	1477833	330	4776	70.2	2.4
BF4 Bin 79	Nitrospirota	2013132	6	1113409	99.0	0.9
BF4 Bin 8	Desulfobacterota	3073854	205	31329	96.6	3.2
BF4 Bin 80	Proteobacteria	3719484	243	24730	95.5	3.1
BF4 Bin 81	Proteobacteria	2682084	156	26092	89.8	2.3
BF4 Bin 82	Bacteroidota	1894486	249	9467	96.2	2.8
BF4 Bin 83	Proteobacteria	3109429	148	33007	89.5	6.5
BF4 Bin 84	Bacteroidota	3205579	206	29025	85.8	2.4
BF4 Bin 85	Proteobacteria	2608681	240	16540	82.8	2.4
BF4 Bin 86	Proteobacteria	2183240	269	10575	80.9	2.3

BF4 Bin 87	Acidobacteriota	4679414	312	26638	95.3	4.3
BF4 Bin 88	Bacteroidota	3473027	21	278267	97.2	0.4
BF4 Bin 9	Desulfobacterota	2161278	93	38286	89.8	2.5
BF4 Bin 91	Campylobacterota	2472549	437	6488	86.0	7.3
BF4 Bin 92	Proteobacteria	3399247	196	30974	87.1	1.5
BF4 Bin 93	Myxococcota	7466586	703	14881	84.0	4.9
BF4 Bin 94	OLB16	4227372	676	7294	85.9	2.9
BF4 Bin 95	Bacteroidota	3075521	424	9254	86.3	1.6
BF4 Bin 96	Desulfobacterota	3103881	233	22078	89.7	0.7
BF4 Bin 97	Bacteroidota	3698673	130	43937	95.1	3.3
BF4 Bin 98	Proteobacteria	1991996	297	8393	72.2	4.1
BF4 Bin 99	Bdellovibrionota	2380300	436	5854	76.6	1.3

*Genome used as a representative for dereplicated genomes that were grouped at 95% average nucleotide identity (ANI).

Appendix B

Supplementary Material – Chapter 3

Supplementary Figure B1. Presence or absence of functional pathways encoded in Cloacimonadota genomes generated from the Distilled and Refined Annotation of Metabolism (DRAM) annotation pipeline. Left-most facets are heat maps depicting carbon metabolism and energy pathways, with partial or complete presence of a pathway or complex coded by increasing intensity of blues. Right hand facets depict specific metabolic capacities as presence/absence, with green indicating presence of the trait and blue indicating absence. The file is available online at doi: https://doi.org/10.1111/1758-2229.13061.

Supplementary Table B1. Characteristics of Cloacimonadota genomes from the landfill site and public databases (as of July 2019), and reference genomes used in analyses. All publicly available genomes for members of the phylum Cloacimonadota were downloaded from the Joint Genome Institute's Integrated Microbial Genomes and Microbiomes (http://img.jgi.doe.gov) or the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) (July 2019). Three genomes for members of the Gemmatimonadota, five genomes for members of the Fibrobacterota and six genomes for members of the Zixibacteria were also downloaded as phylogenetically related yet distinct outgroups for phylogenetic trees. Genome quality of landfill MAGs and publicly available genomes were evaluated with CheckM, and genomes that did not meet the >70% completion and < 10% contamination criteria were not included in metabolic analyses. The file is available online at doi: https://doi.org/10.1111/1758-2229.13061.

Supplementary Table B2. Presence or absence of metabolic genes of interest identified in Cloacimonadota genomes from the KEGG annotations. Genomes were evaluated using the KEGG Automatic Annotation Server (KAAS) (**http://www.genome.jp/kegg/kaas/**) (Moriya *et al.*, 2007). KAAS was used with bi-directional best hit information based on sequence similarities to provide KEGG orthology identifiers for genes within each genome. Enzyme Commission (EC) numbers or KO numbers were used to cross reference against metabolic functions of interest to ascertain presence or absence. The file is available online at doi: https://doi.org/10.1111/1758-2229.13061.

Supplementary Table B3. Annotations for each Cloacimonadota genome obtained with DRAM. The file is available online at doi: https://doi.org/10.1111/1758-2229.13061.

Appendix C

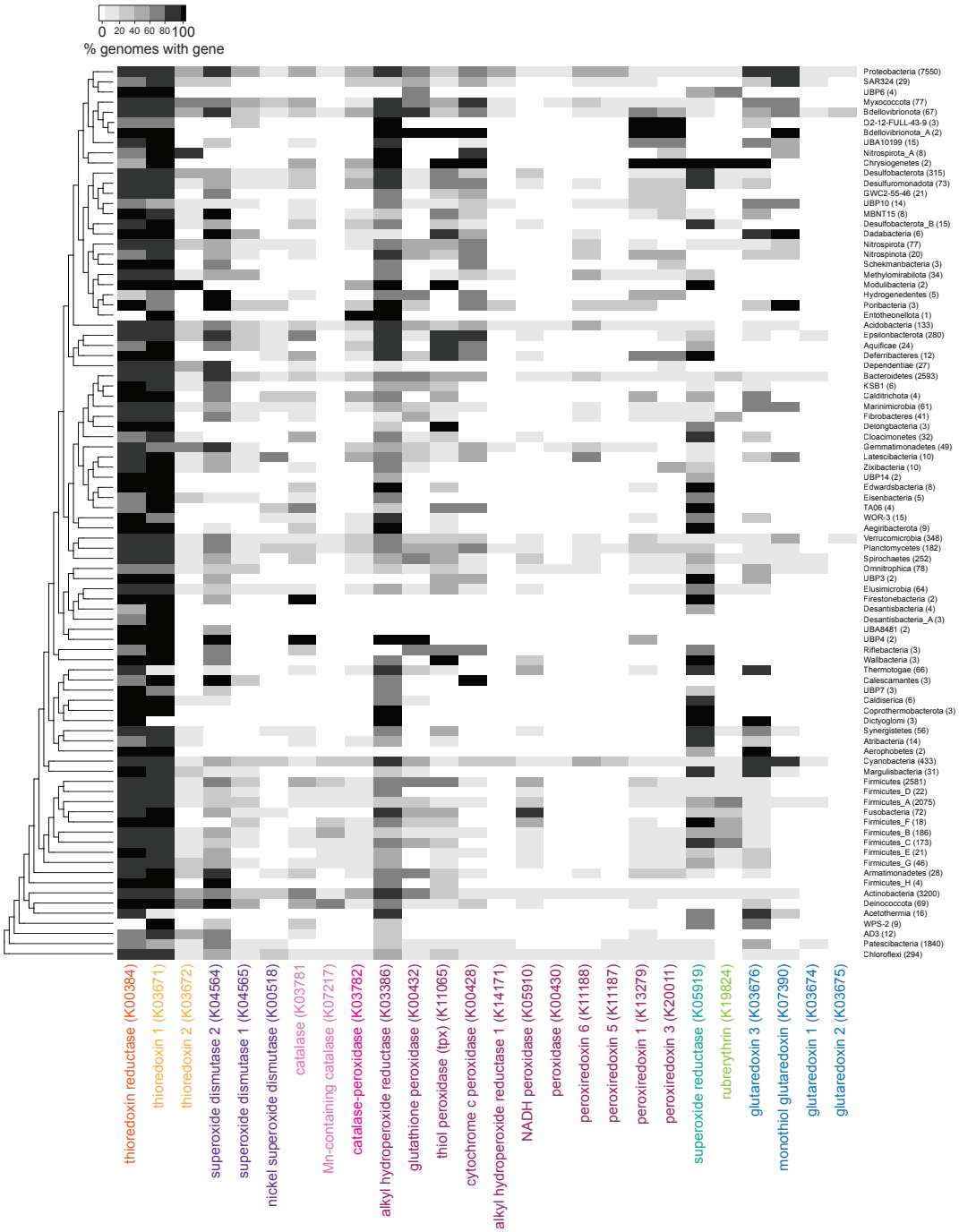
Supplementary Material – Chapter 4

Supplementary Figure C1. ROS scavenging enzymes' distribution across the domain Bacteria. Heat map depicts prevalence of a given ROS enzyme (columns) across bacterial phyla (rows), where black indicates 100% of sequenced organisms in that lineage contain the gene of interest, white indicates zero genomes per clade, and the greys are 20% increments in a gradient from light (1- 20%) to dark (80-99%). Enzyme names are colored by function as in Figure 4.2. Columns are ordered by type of ROS defense (*e.g.*, superoxide dismutases), with the members ranked by prevalence. Phyla are ordered based on their phylogenetic relationships as depicted in the GTDB/AnnoTree phylogeny (Parks *et al.* 2018), as opposed to Figure 4.1 which is ordered by clustering of gene presence. Gene distributions were determined from 23,913 bacterial genomes in the Genome Taxonomy Database (Parks *et al.* 2018) using the AnnoTree software tool (Mendler *et al.* 2019).

References

Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A., and Doxey, A.C. 2019. Annotree: Visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 47(9): 4442–4448. doi:10.1093/nar/gkz246.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36(10): 996. doi:10.1038/nbt.4229.


0 20 40 60 80 100
% genomes with gene
Proteobacteria (7550)
SAR324 (29)
UBP6 (4)
Myxococcota (77)
Bdellovibrionota (67)
O2-12-FULL-43-9 (3)
Bdellovibrionota_A (2)
UBA10199 (15)
Nitrospirota_A (8)
Chrysiogenetes (2)
Desulfobacterota (315)
Desulfuromonadota (73)
GWC2-55-46 (21)
UBP10 (14)
MBNT15 (8)
Desulfobacterota_B (15)
Dadabacteria (6)
Nitrospirota (77)
Nitrospinota (20)
Schekmanbacteria (3)
Methylomirabilota (34)
Modulibacteria (2)
Hydrogenedentes (5)
Poribacteria (3)
Entotheonellota (1)
Acidobacteria (133)
Epsilonbacterota (280)
Aquificae (24)
Deferribacteres (12)
Dependentiae (27)
Bacteroidetes (2593)
KSB1 (6)
Calditrichota (4)
Marinimicrobia (61)
Fibrobacteres (41)
Delongbacteria (3)
Cloacimonetes (32)
Gemmatimonadetes (49)
Latescibacteria (10)
Zixibacteria (10)
UBP14 (2)
Edwardsbacteria (8)
Eisenbacteria (5)
TA06 (4)
WOR-3 (15)
Aegiribacteota (9)
Verrucomicrobia (348)
Planctomycetes (182)
Spirochaetes (252)
Omnitrophica (78)
UBP3 (2)
Elusimicrobia (64)
Firestonebacteria (2)
Desantisbacteria (4)
Desantisbacteria_A (3)
UBA8481 (2)
UBP4 (2)
Riflebacteria (3)
Wallbacteria (3)
Thermotogae (66)
Calescamantes (3)
UBP7 (3)
Caldiserica (6)
Coprothermobacterota (3)
Dictyoglomi (3)
Synergistetes (56)
Atribacteria (14)
Aerophobetes (2)
Cyanobacteria (433)
Margulisbacteria (31)
Firmicutes (2581)
Firmicutes_D (22)
Firmicutes_A (2075)
Fusobacteria (72)
Firmicutes_F (18)
Firmicutes_B (186)
Firmicutes_C (173)
Firmicutes_E (21)
Firmicutes_G (46)
Armatimonadetes (28)
Firmicutes_H (4)
Actinobacteria (3200)
Deinococcota (69)
Acetothermia (16)
WPS-2 (9)
AD3 (12)
Patescibacteria (1840)
Chloroflexi (294)
thioredoxin reductase (K00384)
thioredoxin 1 (K03671)
thioredoxin 2 (K03672)
superoxide dismutase 2 (K04564)
superoxide dismutase 1 (K04565)
nickel superoxide dismutase (K00518)
catalase (K03781)
Mn-containing catalase (K07217)
catalase-peroxidase (K03782)
alkyl hydroperoxide reductase (K03386)
glutathione peroxidase (K00432)
thiol peroxidase (tpx) (K11065)
cytochrome c peroxidase (K00428)
alkyl hydroperoxide reductase 1 (K14171)
NADH peroxidase (K05910)
peroxidase (K00430)
peroxiredoxin 6 (K11188)
peroxiredoxin 5 (K11187)
peroxiredoxin 1 (K13279)
peroxiredoxin 3 (K20011)
superoxide reductase (K05919)
rubrerythrin (K19824)
glutaredoxin 3 (K03676)
monothiol glutaredoxin (K07390)
glutaredoxin 1 (K03674)
glutaredoxin 2 (K03675)