ANTIBIOTIC RESISTANCE IN A CHANGING NORTH: UNEARTHING THE EFFECTS OF DISTURBANCE-INDUCED PERMAFROST THAW ON THE ALASKAN SOIL RESISTOME

By

Tracie J. Haan, B.S.

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of

Master of Science
in
Biological Sciences

University of Alaska Fairbanks
August 2021

APPROVED:

Devin M. Drown, Committee Chair
Andrej Podlutsky, Committee Member
Karsten Hueffer, Committee Member
Jason Burkhead, Committee Member
Diane Wagner, Department Chair
Department of Biology and Wildlife
Kinchel C. Doerner, Dean
College of Natural Science and Mathematics
Richard Collins, *Director of the Graduate School*

ABSTRACT

The evolution of antibiotic resistance in pathogenic bacteria is a major threat at the forefront of public health today. By studying soils, one of the ancestral origins of antibiotic production and resistance, we can gain insight into how antibiotic resistance genes (ARGs) from the environment have contributed to the evolution and emergence of resistance in pathogens. These studies are particularly important in soils where polar amplification and human expansion has already impacted the frequency and intensity of soil disturbance events (e.g., wildfires, deglaciation, land-use). In Alaska these disturbances augment permafrost thaw shifting the biogeochemical properties of active layer soils that structure microbial community composition and hypothetically the resistome (i.e., summation of ARGs). Thus, the goal of this thesis was to assess how soil disturbance, and the subsequent shift in community composition, will affect the types, abundance, and mobility of ARGs that comprise the subarctic soil resistome. In the **first chapter** I cultured bacteria from a permafrost thaw gradient in Interior Alaska, tested the isolates for susceptibility to antibiotics, annotated their genomes for ARGs, and compared their resistance profiles to a global database of soil bacteria genomes. I found that phylogenic and ecological factors structured the resistome. Additionally, antibiotic resistance phenotypes and genotypes were widespread in the soil isolates suggesting resistance is an intrinsic component of bacterial evolution. In the **second chapter,** I used long read metagenomics to identify predominant ARGs, ARG host taxa, and the relationship between community composition and ARG abundance. From the long read data, I unearthed major trends in the types of ARGs at our study site and determined ARG abundance had a quadratic relationship with disturbance and negative relationship temporally by year highlighting the complex interplay soil conditions have in structuring the taxa that enrich ARGs in the community. To analyze how individual bacteria contribute to ARGs in the community, I generated metagenome assembled genomes (MAGs) using Hi-C proximity ligation. From the MAGs, I found a significant difference in ARGs per genome between phyla that emphasized how an enrichment of specific bacteria can affect the abundance of ARGs in subarctic soils. I also identified several plasmid-borne ARGs highlighting the potential for horizontal gene transfer. Overall, this thesis provides evidence that ARGs in permafrost-associated soil are structured by disturbance-induced community shifts. Thus, as climate change increases the frequency of disturbance events that shift the microbial communities in active layer soils, One Health can be impacted by alterations to ARGs comprising the resistome.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS

Acronym	Definition
AIC	Akaike information criterion
ARO	antibiotic resistance ontology
ARG	antibiotic resistance gene
BLAST	Basic Local Alignment Search Tool
BVOC	biogenic volatile organic compounds
CARD	Comprehensive Antibiotic Resistance Database
CLSI	Clinical and Laboratory Standards Institute
DNA	deoxyribonucleic acid
FPES	Fairbanks Permafrost Experiment Station
GLM	generalized linear model
HGT	horizontal gene transfer
ISA	Indicator Species Analysis
MAG	Metagenomic assembled genome
MCR	mobilized colistin resistance
MD	Most-disturbed
MGE	mobile genetic element
NCBI	National Center for Biotechnology Information
NMDS	Non-metric multidimensional scaling
ONT	Oxford Nanopore Technologies
RGI	Resistance Gene Identifier
RND	Resistance-nodulation-cell division
SD	Semi-disturbed
TSB	Tryptic Soy Broth
UD	Undisturbed

ACKNOWLEDGEMENTS

Many thanks to members of the Drown lab, my graduate committee, and colleagues at the University of Alaska Fairbanks (UAF) who have provided extensive intellectual, laboratory, and field support throughout the course of my graduate career including Ursel Schütte, Taylor Seitz, Madeline McCarthy, Scout McDougal, Jennie Humphrey, Anne-Lise Ducluzeau, Jeremy Buttler, Shayna Matson, Anna Kardash, and Sara Kline. I would also like to thank the American Society for Microbiology Alaska branch and Environmental Microbiology and Microbial Ecology group members who have provided invaluable feedback. Big thanks to Tom Douglas from the Cold Regions Research and Engineering Laboratory (CRREL) who provided access to our field site, the Fairbanks Permafrost Experiment Station. Thanks to my partner, Chris Smith, for lending a pair of eyes when needed and providing extensive moral support. I also would like to acknowledge the support provided by Henry, my dog, who attended most virtual meetings, presentations, and conferences while being a good boy.

I acknowledge the generous support that made this work possible from the Institute of Arctic Biology, Alaska Institutional Development Award (IDeA) Network of Biomedical Research Excellence (INBRE), and UAF Biomedical Learning and Student Training (BLaST) program. Research reported here was supported by BLaST through the National Institute of General Medical Sciences of the National Institutes of Health under awards UL1GM118991, TL4GM118992, and RL5GM118990. Research reported here was also supported by an IDeA from the National Institute of General Medical Sciences of the National Institutes of Health under grant 2P20GM103395. My graduate work took place in Fairbanks and therefore I would like to acknowledge the Alaska Native nations upon whose ancestral lands our campus resides. In Fairbanks, the Troth Yeddha' Campus is located on the ancestral lands of the Dena people of the lower Tanana River.

INTRODUCTION

Antibiotic resistance is a global threat at the forefront of public health. Genes conferring resistance to commonly used antibiotics have been shown to recurrently emerge, spread, and persist in pathogenic bacteria thereby reducing the clinical efficacy of commonly used antibiotics, increasing treatment cost, hospitalization time, and mortality rates (O'Neill 2016). In 2017, the United States was estimated to have 2,868,700 antibiotic-resistant infections resulting in 35,900 deaths (Center for Disease Control and Prevention 2019). This statistic emphasizes the burden of antibiotic resistance, which can be attributed to a multitude of factors. For example, the rise of globalization, antibiotic misuse and reliance in clinical, agricultural, and veterinary settings, widespread distribution of antibiotic resistance genes (ARGs) in bacteria, and co-selective pressures in the environment such as heavy metal or antibiotic pollution (Aslam et al. 2018). However, it is now evident that selective pressures favoring the maintenance of ARGs extends beyond the clinic to bacteria from environmental biomes predisposed to evolutionary pressures such as competitive inhibition via antibiotic production from competing microbes (Wright et al. 2012).

The soil resistome (i.e., summation of antibiotic resistance genes in soils) has previously been implicated as a risk to human health by comprising a reservoir of diverse and ancient ARGs that can be transferred from soil bacteria to human pathogens via horizontal gene transfer (Forsberg et al. 2012, D'Costa et al. 2011, Davies & Davies 2010). Soils host an abundance of resistant bacteria that have evolved in part as a result of the ancient "arms-shield race" between antibiotic producing and resistant strains of microorganisms (Aminov & Mackie 2007). In fact, most antibiotics used in medicine today are derived from bioactive compounds produced by soil fungi and bacteria such as *Streptomyces*, a genus that produces around two-thirds of all clinical antibiotics including streptomycin, tetracycline, and cephalosporin (de Lima Procópio et al. 2012, Watve et al. 2001). Within soils, these naturally produced antibiotics have been suggested to mediate interspecific competition for limited metabolic resources by inhibiting growth or killing susceptible cells (Westhoff et al. 2021, Lee et al. 2020) while providing a selective advantage for both the antibiotic producer and conspecific non-producing but intrinsically resistant individuals (e.g., individuals with ARGs on chromosomes associated with that specific bacterial genera). Susceptible soil bacteria can thus evolve to evade the effects of antibiotics by acquiring resistance through a random mutation or acquisition of ARGs via horizontal gene transfer mediated by mobile genetic elements (MGEs), such as plasmids, transposons, and lysogenic bacteriophages (Pärnänen et al. 2016).

Along with natural antibiotics potentially selecting for resistant bacteria, humans frequently generate additional selective pressures favoring the maintenance of ARGs (Ager et al. 2010). Activities shown to increase the abundance of ARGs in soils include antibiotic use in livestock (Ghosh et al. 2007, He et al. 2020), agricultural use of biocides (Berg et al. 2005), and introduction of heavy metals from waste management and mining (Xie et al. 2010). These land-use practices introduce compounds (e.g., biocides and heavy metals) that co-select for genes, such as efflux pumps, that confer resistance to both antibiotics and the compound. However, even 30,000-year-old Beringian permafrost sediments (D'Costa et al. 2011) and high-latitude Alaskan soils (Allen et al. 2009) with minimal anthropogenic influence (e.g., pollution, human expansion, agriculture, mining) have been shown to harbor a diverse array of functional ARGs that encode proteins that confer resistance to beta-lactams, tetracycline, and vancomycin. The role of co-selective pressures and presence of ARGs in unpolluted subarctic soils begs the question of how pressures induced by climatic and human driven changes in Alaska will affect the soil resistome in terms of abundance, types, and mobility of ARGs.

Alaskan soils are at the forefront of change. Disturbances to these subarctic soils such as anthropogenic growth, land use, and climate change augment near-surface permafrost thaw rapidly altering physical and chemical properties of the overlying active layer soil (Forsberg et al. 2014, Douglas et al. 2008). These changes may directly impact the conditions that favor the selection of bacterial taxa enriched in ARGs. For example, disturbance events could co-select for bacteria that encode mechanisms (e.g., efflux pumps) that allow them to concurrently cope with both antibiotics and biological stressors released from permafrost like biogenic volatile organic compounds (Ramos et al. 2001, Kramshøj et al. 2018). Changes to niche availability driven by disturbance may also favor cells that are capable of competitive inhibition via antibiotic production therefore directly selecting for bacteria resistant to antibiotics generated by these producers (Hibbing et al. 2010). Moreover, since plasmid-mediated genetic variation allows bacterial populations to respond to environmental challenges, permafrost thaw may increase the abundance of mobile genetic elements, such as plasmids (Djordjevic et al. 2013) and integrons (Stalder et al. 2012), and thus the abundance of ARGs housed on plasmids that pose more risk in terms of dissemination to pathogenic bacteria (Aminov et al. 2009).

As antibiotic resistance continues to emerge and rapidly spread in clinical settings, it is imperative to generate studies that build insight into the ecology of resistance genes in environmental reservoirs that pose a threat to human health. The overarching goal of this thesis is to assess how disturbance to permafrost-associated soils, and the subsequent shift in community composition, will affect the types, abundance, and mobility of ARGs that comprise

the active layer resistome. **Chapter 1,** entitled "Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska" is published in the journal *Microorganisms* in a special issue on *Antimicrobial Resistance: From the Environment to Human Health*. We present a culture-based analysis of 90 bacterial isolates from a disturbance gradient associated with permafrost thaw in Interior Alaska, using both phenotypic antibiotic susceptibility testing and genomic analyses. Our results describe the role of both phylogeny and ecology in shaping the resistome ultimately implying a shift in community composition with disturbance-induced permafrost thaw can drive the diversity and abundance of ARGs in soils. This chapter is confined to culturable bacteria, which is a limited representation of total bacterial diversity in soils (<1%). This thesis chapter's limitation therefore compels the need for further research examining resistance genes from uncultivable bacteria through the lens of metagenomics.

Chapter 2 entitled "Assessing Risks Posed by Alaska's Active Layer Resistome Associated with Permafrost Thaw" addresses this need. In this chapter, I ask how a shift in community composition as a result of disturbance-induced permafrost thaw affects ARG abundance, composition, and mobility in uncultured bacteria. To answer this, I examined resistance genes from shotgun metagenomic long read data along with metagenomic-assembled genomes (MAGs) generated using Hi-C proximity ligation to link plasmids to host genomes. I identified ARGs and bacterial families indicative of soil disturbance in the context of our study site, suggested a relationship between community composition and the abundance of antibiotic resistance genes, and generated genomes from uncultivable bacteria that help us identify plasmid-borne ARGs and determine which species are enriched in ARGs therefore providing context for how a shift in community affects the resistome.

REFERENCES.

Ager, D., Evans, S., Li, H., Lilley, A. K., & Van Der Gast, C. J. (2010). Anthropogenic disturbance affects the structure of bacterial communities. *Environmental microbiology*, *12*(3), 670-678. DOI: 10.1111/j.1462-2920.2009.02107.x

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *The ISME Journal*, *3*(2), 243–251. DOI: 10.1038/ismej.2008.86

Aminov, R. I., & Mackie, R. I. (2007). Evolution and ecology of antibiotic resistance genes. *FEMS Microbiology Letters*, *271*(2), 147–161. DOI: 10.1111/j.1574-6968.2007.00757.x

Aminov, R. I. (2009). The role of antibiotics and antibiotic resistance in nature. *Environmental microbiology*, *11*(12), 2970-2988. DOI: 10.1111/j.1462-2920.2009.01972.x

Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., ... & Baloch, Z. (2018). Antibiotic resistance: a rundown of a global crisis. *Infection and drug resistance*, *11*, 1645. DOI: 10.2147/IDR.S173867

Berg, J., Tom-Petersen, A., & Nybroe, O. (2005). Copper amendment of agricultural soil selects for bacterial antibiotic resistance in the field. *Letters in Applied Microbiology*, *40*(2), 146-151. DOI: 10.1111/j.1472-765X.2004.01650.x.

Center for Disease Control and Prevention (2019). Antibiotic Resistance Threats in the United States, 2019. *Atlanta, GA: U.S. Department of Health and Human Services*.

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews : MMBR*, *74*(3), 417–433. DOI: 10.1128/MMBR.00016-10

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., … Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*. DOI: 10.1038/nature10388

de Lima Procópio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L., & de Araújo, J. M. (2012). Antibiotics produced by Streptomyces. *The Brazilian Journal of infectious diseases*, *16*(5), 466-471. DOI: 10.1016/j.bjid.2012.08.014

Djordjevic, S. P., Stokes, H. W., & Roy Chowdhury, P. (2013). Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Frontiers in microbiology*, *4*, 86. DOI: 10.3389/fmicb.2013.00086

Douglas, T.; Kanevskiy, M.; Romanovsky, V.; Shur, Y.; Yoshikawa, K. (2008). Permafrost Dynamics at the Fairbanks Permafrost Experimental Station Near Fairbanks, Alaska. *Institute of Northern Engineering, University of Alaska Fairbanks*, 1, 373-378. DOI: 10.1.1.419.2936

Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., & Dantas, G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science*, *337*(6098), 1107 LP – 1111. DOI: 10.1126/science.1220761

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., & Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature*, 509(7502), 612-616. DOI: 10.1038/nature13377

Ghosh, S., & LaPara, T. M. (2007). The effects of subtherapeutic antibiotic use in farm animals on the proliferation and persistence of antibiotic resistance among soil bacteria. *The ISME journal*, *1*(3), 191-203. DOI: 10.1038/ismej.2007.31

Haan, T. J., & Drown, D. M. (2021). Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska. *Microorganisms*, *9*(1), 116. DOI: 10.3390/microorganisms9010116

He, Y., Yuan, Q., Mathieu, J., Stadler, L., Senehi, N., Sun, R., & Alvarez, P. J. (2020). Antibiotic resistance genes from livestock waste: occurrence, dissemination, and treatment. *npj Clean Water*, *3*(1), 1-11. DOI: 10.1038/s41545-020-0051-0

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), 15-25. DOI: 10.1038/nrmicro2259

Kramshøj, M., Albers, C. N., Holst, T., Holzinger, R., Elberling, B., & Rinnan, R. (2018). Biogenic volatile release from permafrost thaw is determined by the soil microbial sink. *Nature communications*, *9*(1), 1-9. DOI: 10.1038/s41467-018-05824-y

Lee, N., Kim, W., Chung, J., Lee, Y., Cho, S., Jang, K. S., ... & Cho, B. K. (2020). Iron competition triggers antibiotic biosynthesis in Streptomyces coelicolor during coculture with Myxococcus xanthus. *The ISME Journal*, 1-14. DOI: 10.1038/s41396-020-0594-6

O'Neill, J. (2016). Book review: Tackling drug-resistant infections globally. *Archives of Pharmacy Practice*, *7*(3), 110. DOI: 10.4103/2045-080x.186181

Pärnänen, K., Karkman, A., Tamminen, M., Lyra, C., Hultman, J., Paulin, L., & Virta, M. (2016). Evaluating the mobility potential of antibiotic resistance genes in environmental resistomes without metagenomics. *Scientific reports*, *6*, 35790. DOI: 10.1038/srep35790

Ramos, J. L., Gallegos, M. T., Marqués, S., Ramos-González, M. I., Espinosa-Urgel, M., & Segura, A. (2001). Responses of Gram-negative bacteria to certain environmental stressors. *Current opinion in microbiology*, *4*(2), 166-171. DOI: 10.1016/S1369-5274(00)00183-1

Seitz, T. J., Schütte, U. M., & Drown, D. M. (2021). Soil Disturbance Affects Plant Productivity via Soil Microbial Community Shifts. *Frontiers in microbiology*, *12*, 76. DOI: 10.3389/fmicb.2021.619711

Stalder, T., Barraud, O., Casellas, M., Dagot, C., & Ploy, M. C. (2012). Integron involvement in environmental spread of antibiotic resistance. *Frontiers in microbiology*, *3*, 119. DOI: 10.3389/fmicb.2012.00119

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics are produced by the genus Streptomyces?. *Archives of microbiology*, 176(5), 386-390. DOI: 10.1007/s002030100345

Westhoff, S., Kloosterman, A., van Hoesel, S. F., van Wezel, G. P., & Rozen, D. E. (2021). Competition sensing alters antibiotic production in Streptomyces. *mBio,* 12 (1) e02729-20; DOI: 10.1128/mBio.02729-20

Wright, G. D., & Poinar, H. (2012). Antibiotic resistance is ancient: implications for drug discovery. *Trends in microbiology*, *20*(4), 157-159. DOI: 10.1016/j.tim.2012.01.002

Xie, X., Fu, J., Wang, H., & Liu, J. (2010). Heavy metal resistance by two bacteria strains isolated from a copper mine tailing in China. *African Journal of Biotechnology*, *9*(26), 4056-4066.

CHAPTER 1:

Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska[1]

1.1 Abstract

Monitoring antibiotic resistance genes (ARGs) across ecological niches is critical for assessing the impacts distinct microbial communities have on the global spread of resistance. In permafrost-associated soils, climate and human driven disturbances augment near-surface thaw shifting the predominant bacteria that shape the resistome in overlying active layer soils. This thaw is of concern in Alaska, because 85% of land is underlain by permafrost, making soils especially vulnerable to disturbances. The goal of this study is to assess how soil disturbance, and the subsequent shift in community composition, will affect the types, abundance, and mobility of ARGs that compose the active layer resistome. We address this goal through the following aims: (1) assess resistance phenotypes through antibiotic susceptibility testing, and (2) analyze types, abundance, and mobility of ARGs through whole genome analyses of bacteria isolated from a disturbance-induced thaw gradient in Interior Alaska. We found a high proportion of isolates resistant to at least one of the antibiotics tested with the highest prevalence of resistance to ampicillin. The abundance of ARGs and proportion of resistant isolates increased with disturbance; however, the number of ARGs per isolate was explained more by phylogeny than isolation site. When compared to a global database of soil bacteria, RefSoil+, our isolates from the same genera had distinct ARGs with a higher proportion on plasmids. These results emphasize the hypothesis that both phylogeny and ecology shape the resistome and suggest that a shift in community composition as a result of disturbance-induced thaw will be reflected in the predominant ARGs comprising the active layer resistome.

1.2 Introduction

The rapid evolution and spread of antibiotic resistance is one of the greatest challenges faced in public health today. Antibiotic resistance impedes the successful treatment of bacterial infections by reducing antibiotic efficacy, increasing disease burden, mortality rates, hospitalization time and cost [1]. On an evolutionary time scale, the extensive prevalence of resistant phenotypes in human pathogens is a recent event, driven by the large-scale

---

Article Published as: **Haan, T. J.,** & Drown, D. M. (2021). Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska. *Microorganisms* , *9* (1), 116. PMCID: PMC7825290 DOI: 10.3390/microorganisms9010116

production and widespread use of antibiotics in clinical, agricultural, and veterinary settings [2,3]. Even when antibiotic stewardship is instilled (i.e., antibiotic use is confined to essential needs), antibiotics, or pollutants such as heavy metals that co-select for resistance [4], are dispersed within microbial habitats, thereby generating selective pressures that increase the abundance of resistant strains and their associated antibiotic resistance genes (ARGs). It was originally thought that the genetic variability driving resistance was primarily caused by mutational modification to antibiotic targets, and thus, would remain clonal [5]. However, it is now evident that mutational-driven resistance is a weaker force compared to ARGs acquired via horizontal gene transfer (HGT) [6]. In pathogens, resistance genes can be acquired from diverse microbial habitats and taxa [5,7], including bacteria from pristine environments free of antibiotics introduced via human activities [8]. It is therefore important to assess which bacterial taxa and microbial biospheres are the predominant contributors to the evolution of resistance in pathogens [9].

Soils, one of the most diverse microbial habitats on earth, are a vast repository of both antibiotic-producing and coevolved resistant microbial taxa. Antibiotic production is thought to have originated in soils from 2 Gyr to 40 Myr ago, suggesting that resistance has undergone concomitant evolution over a similar timeframe [10,11]. The evolutionary origin of resistance is also supported by studies that have unveiled soils unpolluted by human activity that harbor diverse resistance mechanisms to modern antibiotics, such as 30,000-year-old Beringian permafrost sediments [8]. Moreover, bacteria such as *Streptomyces*, an Actinomycete genus that produces around two-thirds of clinically used antibiotics, are abundant in soils [12,13]. The presence of these antibiotic-producing genera is thought to promote the evolution, and potential dissemination via HGT, of clinically relevant resistance genes from soils. Recent studies have reported ARGs in soil-borne bacteria identical to those circulating in pathogens, suggesting that HGT has occurred [14,15]. These shared ARGs between soil-borne and pathogenic bacteria emphasize a potential role soils have in the evolution and dissemination of resistance. However, to assess the risks posed by soil-borne ARGs, more attention should be paid to their distribution globally, which environments favor growth of bacteria harboring ARGs, and conditions that promote mobility of ARGs such as plasmid carriage [16]. By examining soils affected by environmental change, we can provide insights into both the global distribution and how change affects resistance determinants that may emerge in pathogens.

Alaskan soils are one environment undergoing unprecedented change, as warming within the arctic is occurring 2.5 times faster than in the rest of the globe [17]. This warming has triggered a rise in the frequency of soil disturbance events, such as wildfires and thermokarst

formation, which are of particular concern for Alaska, since approximately 85% of land is underlain by discontinuous permafrost [18–20]. With both climatic and human-driven soil disturbances, near-surface permafrost thaw is augmented, shifting the physical and chemical properties of the overlying active layer soils [21]. Alterations to pH, moisture, and nutrients in the active layer impact ecosystem functions while shifting microbial community composition [22]. This shift in both biotic and abiotic factors has the potential to enrich ARGs by selecting for their host taxa. For example, the phylum Proteobacteria is significantly enriched with plasmid-borne resistance determinants compared to other bacterial phyla [23] and over-represented in terms of abundance in active layer soils associated with disturbance-induced thaw [24]. This connection between ARGs, phylogeny, and community shifts observed in disturbed soils of Alaska makes it imperative to assess how thaw will affect bacteria from active layer soils as a reservoir of resistance.

The Fairbanks Permafrost Experiment Station (FPES) is a long-term research site established by the Army Corps of Engineers in 1945 containing ice-rich permafrost typical of Alaska [21]. FPES has three levels of increasing soil disturbance with minimal thaw in the undisturbed site and up to 9.8 m in the most-disturbed site. The location's distinctive gradient of disturbance-induced thaw makes it fitting for research on the effect of disturbance on permafrost, vegetation, soils, and microbial communities. Previous culture-independent metagenomic analyses conducted at this site found distinct shifts in microbial community composition, such as an enrichment of Proteobacteria in the disturbed soils [24], necessitating more research into how this site has affected antibiotic resistance.

The goal of this study is to assess how soil disturbance, and the subsequent shift in community composition, will affect the types, abundance, and mobility of ARGs that compose the Alaskan active layer resistome. To explore the effects of both phylogeny and ecology on ARGs from active layer soils of Interior Alaska, we identified ARGs in the whole genome sequences of bacteria cultured from FPES active layer soils. We then put these FPES isolates into a global context by comparing ARGs to those identified in bacteria from a database of global soil bacteria, RefSoil+ [25]. This larger database containing both the whole genomes and plasmid sequences of bacteria cultured from global soil habitats makes it possible to investigate distinguishing features of ARG ecology in Alaskan isolates, such as the plasmid carriage and ARG abundance by taxa. Overall, these analyses will allow us to gain insights into the role both biotic and abiotic factors associated with disturbance-induced permafrost thaw will have on antibiotic resistance. By cataloguing ARGs and their host taxa, this work also contributes to the critical knowledge gap regarding the global distribution of ARGs from the environmental

biosphere that have the potential to spread to clinically significant bacteria and compromise health.

1.3 Materials and Methods

1.3.1. Permafrost Thaw Gradient

The Fairbanks Permafrost Experiment Station (FPES) is an ice-rich permafrost site in Interior Alaska (64.875646° N, 147.668981° W), established by the Army Corps of Engineers as part of the Cold Regions Research and Engineering Lab. The site consists of three 3721 m^2 Linell plots [26] with increasing levels of disturbance, now known as the Un-Disturbed (UD), Semi-Disturbed (SD), and Most-Disturbed (MD) sites (Figure 1.1). In 1946, the three sites were established to simulate soil disturbance events, such as wildfire or anthropogenic disturbance, on permafrost degradation by clearing of vegetation. The UD site was left undisturbed to preserve the subarctic taiga forest, whereas the SD site had the surface vegetation cleared while roots and soil organic matter were left intact and the MD site had both surface vegetation and organic matter removed. In 2007 the UD site, which is now monitored as part of the Circumpolar Active Layer Monitoring Network (CALM), was found to have little to no thaw since 1946 whereas the SD and MD had up to 4.7m and 9.8m respectively [21]. These results suggested that permafrost degradation is dependent on both time and surface vegetation which has major implications when it comes to increasing frequency of disturbance events, like wildfires and thermokarst formation, influenced by climate change.

Vegetation at the FPES is typical of the Alaskan Interior-subarctic taiga forest. The undisturbed site is a relatively open black spruce stand (*Picea mariana*) with an understory of continuous thick moss layer interspersed with low-bush cranberry (*Vaccinium vitis-idaea*) and Labrador tea (*Rhododendron groenlandicum*). The UD site can be classified as mesic with a soil organic layer thickness ranging 2 to 35 cm thick, with little to no thaw during maximal permafrost thaw [27]. The semi-disturbed site is now a mix stand dominated by black spruce, Alaskan paper birch (*Betula neoalaskana*), and willow (*Salix alaxensis*). The understory contains a mixture of Labrador tea (*Rhododendron groenlandicum*), Peltigera lichen, roses, horsetail, cloudberry, and small amounts of grass with little litter cover. The MD site is an open shrub land dominated by willows (*Salix alaxensis*) and a developing over story up to 5 m tall of Alaskan birch and black spruce (*Picea mariana*). The understory contains many grasses, clovers, horse-tail (*Equisetum*), and some bare ground. There is no permafrost within the top 4.7 m in either disturbed sites [21].

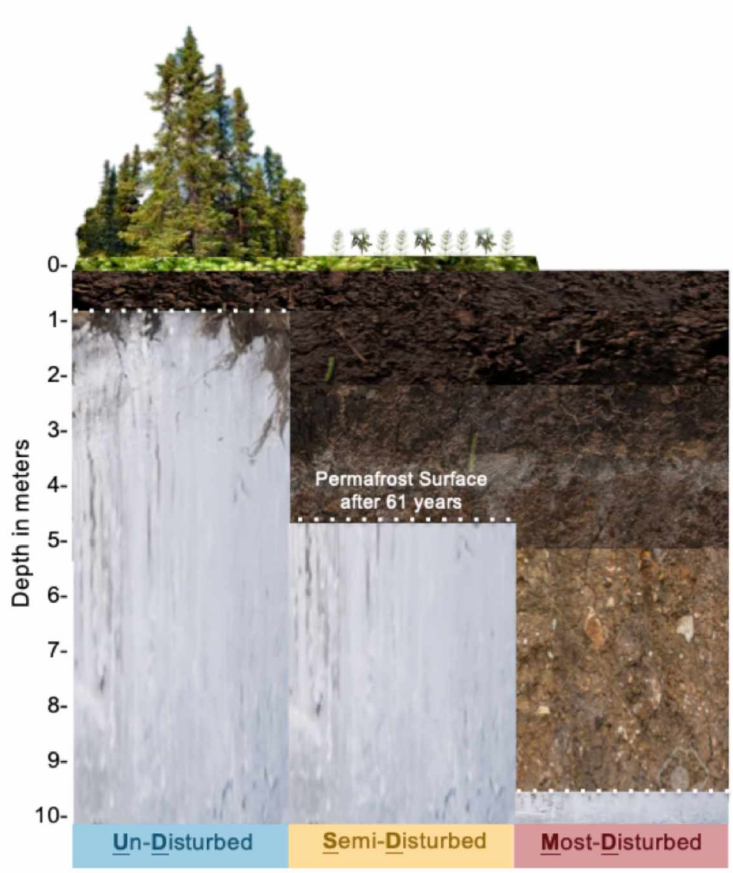


Figure 1.1. Diagram of the disturbance level sites (Undisturbed = UD, Semi-Disturbed = SD, Most Disturbed = MD) at the Fairbanks Permafrost Experiment Station (FPES) and the subsequent depth of permafrost thaw after 61years.

1.3.2. Bacterial Culturing

In the September of 2017, we collected two 10-cm wide by 20-cm deep soil cores from each FPES site with a sterilized soil probe. Prior to coring, the top layer of moss and vegetation were removed. Soil cores were then extracted and immediately stored in a cooler throughout sample collection. Soil cores were moved to a +4 °C fridge until processing 24 h later. To prevent contamination from exogenous cells on the exterior of the soil core, the outer portion of each was removed using a sterile scalpel. The interiors of each core were then sub-sampled using sterile forceps along a depth gradient at intervals of about 2.5 cm for 20 cm to generate a total of 1 g of soil. The 1 g of soil was used to inoculate 100 mL tryptic soy broth (TSB) to produce an enrichment culture. After 48 h at 22 °C, we plated serial dilutions (1:10, 1:100, 1:1000) of the enrichment culture three times for each sample and incubated the plates at three temperatures (+4 °C, +12 °C, and +20 °C), for a total of twelve plates per sample, until distinct colony formation was observed. Ten discrete colonies were chosen at random from each temperature and FPES site by using a random number generator to count colonies across

transects moving across the plate horizontally from the top of the plate to the bottom. Only discrete colonies (i.e., colonies without overlapping colony growth) were counted along the transect lines. This process yielded 90 total colonies, 30 per FPES site. Each colony was isolated and purified using three rounds of streak plate method.

1.3.3. Antibiotic Susceptibility Testing

We screened each isolate for antibiotic resistance using the Kirby-Bauer disk diffusion method [28]. In brief, this method uses paper disks with a fixed concentration of antibiotic that diffuses into agar generating a region where susceptible bacteria cannot grow, called the zone of inhibition. The diameter of this circular zone is then measured and compared to breakpoints established for clinical isolates with values based on isolate taxonomy. In this study, we used five antibiotics, i.e., tetracycline, erythromycin, kanamycin, chloramphenicol, and ampicillin, with each representing a distinct antibiotic class. Standards for antibiotic disks and associated breakpoints were used from the US Clinical and Laboratory Standards Institute M100, 30th ed. Breakpoints established for *Enterobacterales* were used for *Serratia*, *Pantoea*, and *Erwinia* isolates, *Pseudomonas* spp. breakpoints for *Pseudomonas* isolates, and *Enterococcus* spp. breakpoints for *Bacillus* and *Exinguobacterium* isolates to determine if an isolate was susceptible, intermediate, or resistant to each antibiotic tested (Table A1). Isolates without breakpoints to specific antibiotics due to physiological characteristics that render that genus intrinsically resistant, such as in the case of *Pseudomonas* and ampicillin, were removed from subsequent analysis.

1.3.4. Whole Genome Sequencing, Assembly, and Taxonomic Classification

From each purified isolate, we inoculated a liquid culture of TSB, incubated it at 22 °C overnight, and used 1.8 mL of this liquid culture to extract genomic DNA using the Dneasy UltraClean microbial kit (Qiagen, Venlo, The Netherlands) following manufacture protocols. This DNA was then used for sequencing on both Illumina and Oxford Nanopore Technology [ONT] platforms. For Illumina sequencing, we used a Nextera XT library (Illumina, San Diego, CA, USA) prepared by the Genomics Core Lab at the University of Alaska Fairbanks to sequence on an Illumina MiSeq platform with version 3 reagents. We trimmed adapters from Illumina reads with TrimGalore version 0.5.0 [29]. For ONT long reads, a combination of SQK-RBK004 and VSK-VSK002 library preparation (ONT) was employed (Table A2). These libraries were then sequenced on a MinION device (ONT) with r9.4.1 flow cells (FLO-MIN106) for 48–72 h. We base called the raw data using Guppy v3.4.5 (ONT) specifying the high-accuracy model (-c

dna_r9.4.1_450bps_hac.cfg) and default parameters. We de-multiplexed isolate samples using the guppy_barcoder function of Guppy with parameters to discard sequences with middle adapters (–detect_mid_strand_barcodes) and trim barcodes (–trim_barcodes). We used Filtlong v0.2.0 [30] to filter by length (≥50 bp; –min_length 50) and quality (Q) score (≥10; –min_mean_q 90). Flye version 2.7 [31] was used to assemble quality controlled ONT reads specifying nanopore raw reads (–nano-raw) and genome size of 5 mb (–genome-size 5m). The unicycler_polish tool of Unicycler version 0.4.8 [32] was used to polish the flye assemblies with the Illumina reads as input. For isolates TH26, TH81 and TH88, ONT long read assemblies that were previously published in Haan et al. 2019 and Humphrey et al. 2019 respectively were used as inputs for unicycler_polish [33,34].

These assemblies were then annotated with RAST tool kit (RASTtk) in PATRIC v3.6.3 using the Genome Annotation Service [35]. 16S rRNA gene copies for each assembly were aligned using MAFFT v7.450 and consensus sequences were run through blastn version 2.10.0 against the NCBI 16S rRNA database. If the top five hits ranked by bit score were from the same genus, then taxonomy was assigned to an isolate at a genus level (Table 1.1).

Table 1.1. Number of isolates within each phylum and genus by Fairbanks Permafrost Experiment Station (FPES) site.

Taxonomy		FPES Site			
Phylum	**Genus**	**UD**	**SD**	**MD**	**Total**
Firmicute	*Bacillus*	14	11	5	30
Firmicute	*Exiguobacterium*	0	0	1	1
Proteobacteria	*Erwinia*	1	1	3	5
Proteobacteria	*Pantoea*	0	0	2	2
Proteobacteria	*Pseudomonas*	14	18	19	51
Proteobacteria	*Serratia*	1	0	0	1
	Total	30	30	30	90

UD = Undisturbed, **SD** = Semi-Disturbed, **MD** = Most-Disturbed.

1.3.5. Antibiotic Resistance Gene Identification

In this study we identified ARGs by annotating each whole genome assembly with the Comprehensive Antibiotic Resistance Database (CARD) version 3.0.9 using command line tool Resistance Gene Identifier (RGI) version 5.1.0 specifying input type contig (–t contig) with default parameters for BLAST alignment (–a BLAST) and strict and perfect hits only. In order to detect previously unknown homologs using detection models with curated similarity cut-offs

while still ensuring the detected variant is a functional resistance gene rather than spurious partial hits, we used the strict algorithm rather than the algorithms that would select for exclusively perfect or loose hits [36]. Results from RGI were further quality controlled by removing any Antibiotic Resistance Ontology (ARO) hits defined as mutations or ARO hits with less than 50% coverage of the reference sequence unless cutoff on the edge of a contig. To determine if a hit was located on a chromosome or plasmid, contigs containing hits were run through blastn [37] in Geneious Prime version 2019.2.1 against bacteria (taxid = 2) from the RefSeq database. The top hit ranked by bit score was then used to determine if the contig was most similar to a known plasmid or chromosome sequence.

1.3.6. RefSoil+ Comparison

RefSoil+ [25] genomes and plasmids were downloaded from NCBI using accession numbers available on the RefSoil+ github page [38]. Each RefSoil+ sequence was then run through RGI following the same protocol as the FPES analysis of ARGs. Genomes belonging to the matching genera (*Bacillus*, *Erwinia*, *Exinguobacterium*, *Pantoea*, *Pseudomonas*, and *Serratia)* as FPES isolates were used for comparison of the RefSoil+ and FPES resistance genes. The genus *Pseudomonas* contains the largest and most diverse species with eight distinct phylogenomic groups, because the species *Pseudomonas aeruginosa* is distinctive from other groups within our samples *aeruginosa* genomes were removed from FPES versus RefSoil+ analysis.

1.3.7. Data Analyses and Statistics

Tabular outputs from RGI were used to conduct statistical analyses in R Studio version 3.5.7 [39] and visualizations were generated with the R package ggplot2 version 3.2.1 [40]. In order to determine if ARGs were influenced by phylogeny or ecology of FPES sites, we examined statistical differences in the number of ARGs per genome between phyla and across thaw and then compared FPES isolates to correspondent genera from RefSoil+. Kruskal-Wallis one-way analysis of variance was used to test significance of FPES site and phylum as predictors and then a post hoc test using nonparametric Wilcoxon test was used to test for significant differences between groups. We used Akaike Information Criterion model selection to see which Poisson distributed generalized linear model with phylum, FPES site, and both as predictors of the number of resistance genes per isolate has the best-fit. We used a heatmap generated with the R package pheatmap version 1.0.12 [41] to visualize the distribution of ARG hits across phyla and FPES sites. To generate the heatmap, we normalized ARG counts within each group that consisted of phylum and FPES site (e.g., Proteobacteria from MD) by the

number of isolates in that group, and then scaled each normalized count for each resistance gene to generate z-scores. Dendograms grouping each column and row were based on Pearson correlation.

1.4 Results

1.4.1. Assessment of Antibiotic Susceptibility in FPES Isolates

Widespread resistance was observed in the isolates with 91.1% of the 90 total isolates exhibiting at least intermediate resistance to one of the five antibiotics tested, and 45.6% displaying at least intermediate resistance to two or more antibiotics (Figure 1.2a). Ampicillin had the highest prevalence of resistance (82.5%), followed by chloramphenicol (51.1%) and erythromycin (17.5%). Tetracycline had the lowest prevalence of resistance (2.2%). We observed a positive trend in the number of resistant isolates with disturbance-induced thaw for ampicillin, chloramphenicol, and erythromycin that was also observed for intermediate resistance to kanamycin and tetracycline (Figure 1.2b).

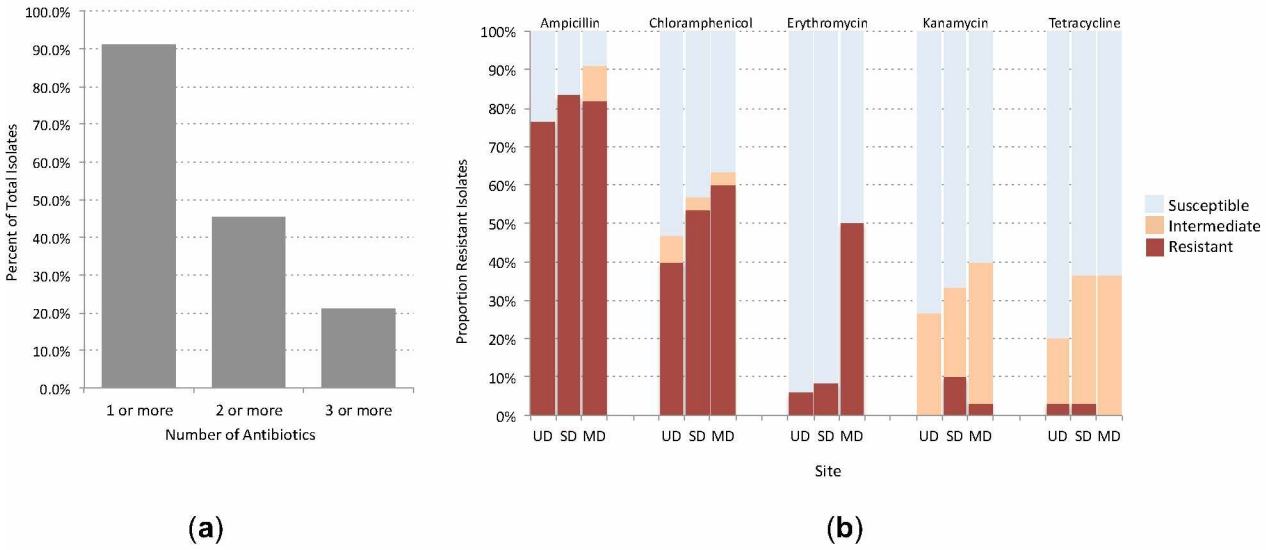


Figure 1.2. (**a**) Percent of total isolates with at least intermediate resistance to one or more, two or more, and three or more of the antibiotics tested. (**b**) Proportion of isolates at each FPES thaw site (UD = Undisturbed, SD= Semi-Disturbed, MD= Most Disturbed) with the associated level of susceptibility (susceptible = light blue, intermediate = orange, resistant = red) to each antibiotic tested based on CLSI breakpoints.

1.4.2. Genome Assembly Statistics

The assembled genomes had a high mean percent completeness (98.91% ± 0.50), low percent contamination (1.16% ± 0.26), and high N50 value (N50 = 3,639,582 ± 279,870 bp, $N50_n$ = 13 ± 5 contigs) for the mean total length of assemblies (6,096,842 ± 946,004 bp). These results, outlined in Table A3, suggest high quality assemblies were produced.

1.4.3. Antibiotic Resistance Genes Identified in FPES Isolates

Across all FPES genomes RGI identified 379 significant hits comprising 27 CARD-based AROs (Table A4). Of these 379 hits, 30 had 100% sequence identity to CARD AROs and another 32 were highly similar (sequence identity >90%). Four genes hits had full length coverage of the reference sequence along with 100% sequence identity, two encoding aminoglycoside inactivating enzymes *AAC(6')-32* and *AAC(6')-Ir* and two encoding *bcrC*, an undecaprenyl pyrophosphate related protein. However, overall mean percent identity across all hits identified was 68.4% ± 19.

Genes encoding proteins for antibiotic efflux, antibiotic inactivation, antibiotic target modification, and antibiotic target protection were observed. The top two most abundant resistance genes identified across isolates were genes encoding proteins for antibiotic efflux. The most abundant efflux pump was *adeF,* a resistance-nodulation-cell division efflux pump that confers multi-drug resistance. There were multiple copies of this gene present in Proteobacteria with 176 chromosomally encoded gene copies distributed across the 100% of Proteobacteria isolates with a high coverage of the gene across assemblies (% length of reference sequence = 99.02 ± 2.902) and variable percent identity (% identity = 52.283 ± 11.822). The second most abundant resistance gene was *AbaQ* (% identity = 72.7 ± 0.427% length of reference sequence = 101.36 ± 0.109), a gene encoding the major facilitator superfamily efflux pump associated with the extrusion of quinolone-type drugs in *Acinetobacter baumannii. AbaQ* was observed across all FPES sites in 78.4% of *Pseudomonas* isolates.

After antibiotic efflux, genes encoding antibiotic inactivating enzymes were the most abundant. These genes were found in isolates from across all FPES sites and genera sampled except *Exinguobacterium*. *FosB* (% identity = 89.055 ± 2.281%, length of reference sequence = 105.012 ± 6.992*)* was the most abundant gene encoding an antibiotic inactivating enzyme with 25 chromosomally encoded gene copies present in 83% of *Bacillus* isolates across all thaw sites. We also observed beta-lactam inactivating genes from four distinct beta-lactamase families (Table A2). The Bc beta-lactamase gene family had the highest abundance (n = 20) all encoding *BcII,* a zinc metallo-beta-lactamase that hydrolyzes a large number of penicillins and

cephalosporins. *BcII* gene copies in our samples were confined to the genus *Bacillus* and found to be both highly similar (% identity = 90.755 ± 0.567) with high gene coverage (% length of reference sequence = 100.39 ± 0) to *BcII* homologs in the CARD database.

Genes encoding target alteration were the least abundant mechanism of resistance (n = 25) and included the genes *armA*, *bcrC*, *MCR-4.1*, *gyrB*, *PmrF*, *sgm*, and *vanJ* which are associated with resistance to aminoglycosides, peptide, glycopeptide, and fluoroquinolone antibiotics. A notable target alteration gene found was the mobilized colistin resistance (MCR) phosphoethanolamine transferase. MCR is a gene superfamily tracked by the Center for Disease Control and Prevention that confers resistance to the last resort antibiotic colistin, a critical antibiotic for treating carbapenem-resistant *Enterobacteriaceae*. We found two significant gene hits for *MCR-4.1*, however they were fragmented on the edge of contigs and thus had low coverage (% length of reference sequence = 10.63 ± 1.174) but 100% sequence identity.

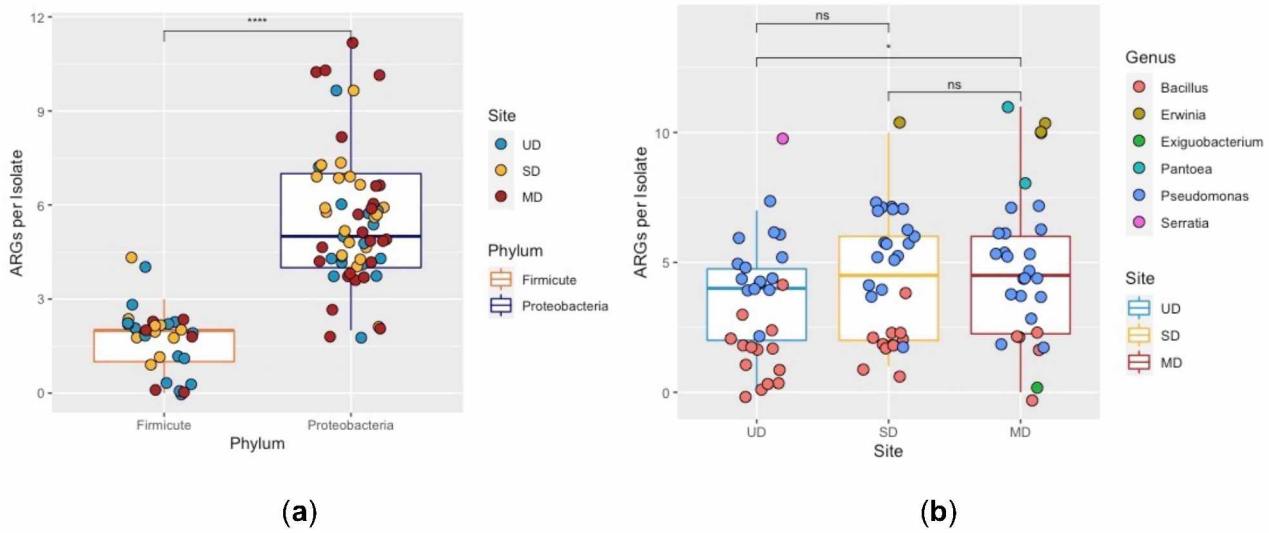


Figure 1.3. (**a**) Boxplot displaying the number of antibiotic resistance genes per isolate by phylum (Kruskal-Wallis $p = 2.8 \times 10^{-13}$) with points representing an isolate color-coded by FPES site. (**b**) Boxplot displaying the number of antibiotic resistance genes per isolate by FPES site (UD = Undisturbed, SD= Semi-Disturbed, MD= Most Disturbed; Kruskal-Wallis $p = 0.083$) with points representing an isolate color-coded by phylum. Wilcoxon test between group significance $p < 0.01$ **, $p < 0.05$ *, ns >0.1.

1.4.4. Influence of Phylogeny and Disturbance-Induced Thaw on ARGs

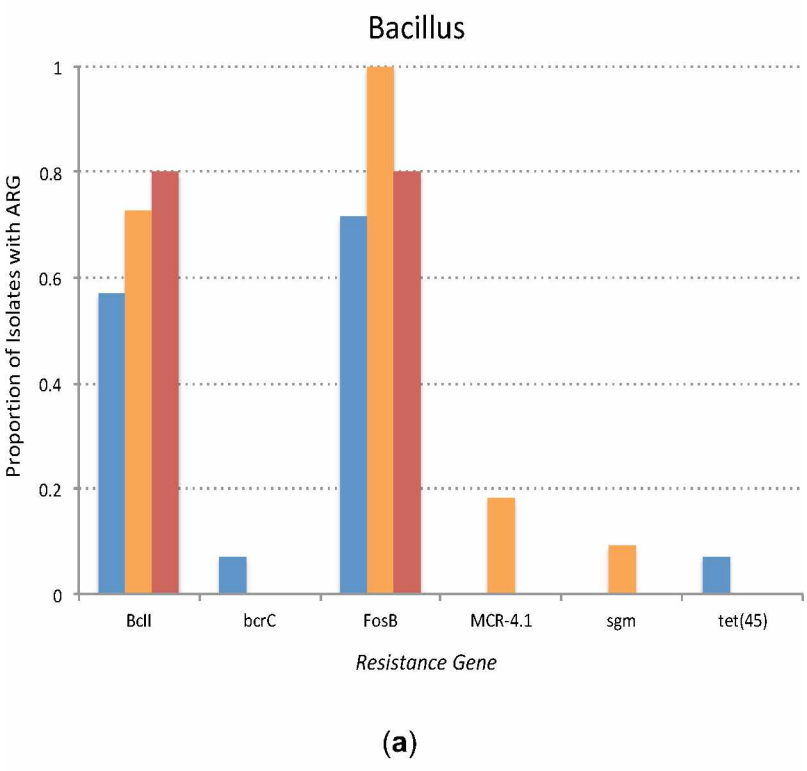
Although the abundance of ARGs increased with FPES disturbance levels (UD = 101, SD = 133, MD = 145), we found the number of CARD hits per isolate was highly significant by phylum (Kruskal-Wallis $p = 2.8 \times 10^{-13}$; Figure 1.3a) rather than FPES site (Kruskal-Wallis p =

0.083; Figure 1.3b). Although FPES site was not significant as a predictor, there was a significant difference between the numbers of ARGs per isolate from the undisturbed to most-disturbed sites (Wilcoxon p = 0.045). When comparing generalized linear models containing FPES site, phylum, and both site and phylum as predictors of ARGs per isolate we found that the model with phylum alone had the best fit (AIC = 349 phylum; 429 FPES site; 351 both).

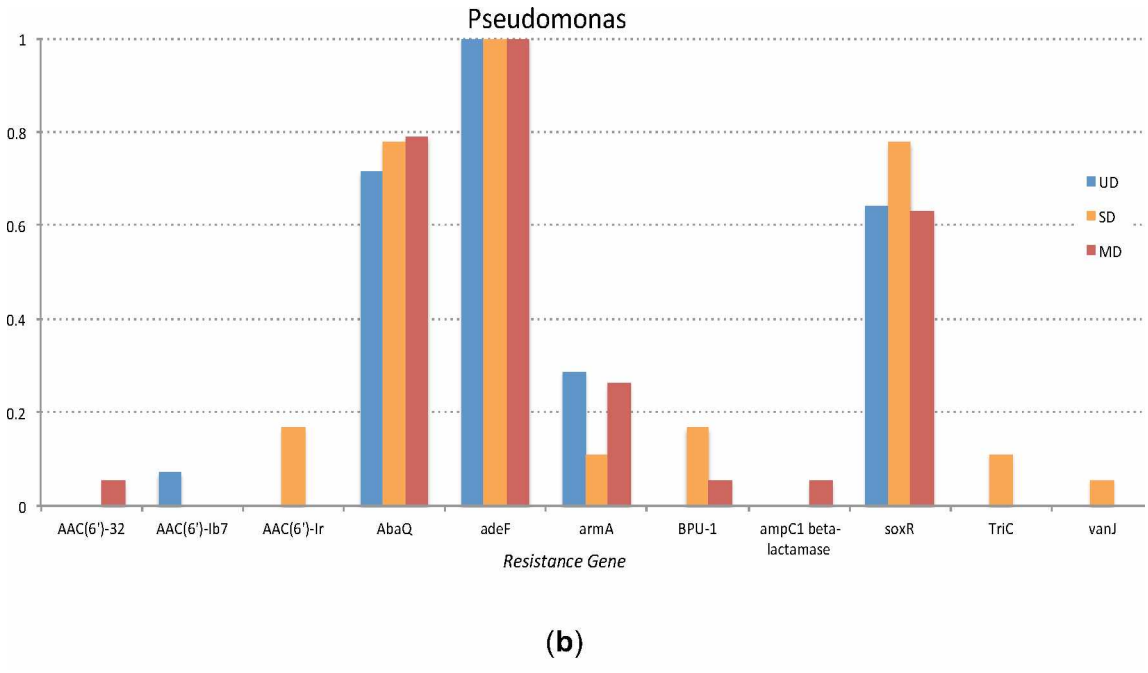
When examining the types of genes more in depth by phylogeny, there is a distinct set of ARGs found within each genus such as FosB, BcII in Bacillus (Figure 1.4a) and adeF, armA, and soxR in *Pseudomonas* (Figure 1.4b). Overall *Bacillus* isolates' core resistance genes were comprised of primarily antibiotic inactivation genes, *Pseudomonas* was antibiotic efflux and target alteration, and *Erwinia* contained antibiotic efflux, target alteration, and inactivation. These sets of core ARGs present in some taxa and absent in the others are what appeared to cause ARGs to cluster more strongly by phylum rather than thaw site in the heatmap (Figure 1.5). Although there was ARGs found across FPES sites in multiple genera (Table A5), some genes were observed exclusively in one taxon and site, such as MCR-4.1 in *Bacillus* from the semi-disturbed site (Figure 1.4a).

1.4.5. Comparison of ARGs in RefSoil+ and FPES Genomes from Corresponding Genera

Across the equivalent genera, RefSoil+ and FPES had 15 similar ARGs variants, 41 unique to RefSoil+ and 12 unique to FPES (Figure 1.6a). The similar ARGs were primarily genes determined to be the more abundant ARGs in FPES isolates whereas the distinct genes were often rare variants (i.e., only one copy across all isolates). When comparing by genus and database we found there were significant differences in number of ARGs per isolate between *Pseudomonas* and *Bacillus* isolates, which was higher in FPES for *Pseudomonas* and higher in RefSoil+ for *Bacillus* (Figure 1.6b). Across the 90 FPES isolates, 4 ARG variants with 6 gene copies were identified in plasmid sequences whereas there were no significant plasmid hits across all 127 RefSoil+ isolates examined from corresponding genera. The plasmid-borne ARGs from FPES included a BES-1 beta lactamase, two antibiotic efflux pumps (*TriC* and *KpnF*), and an undecaprenyl pyrophosphate related protein (*bcrC*) (Table 1.2).



(**a**)



(**b**)

Figure 1.4. The proportion of (**a**) *Bacillus* and (**b**) *Pseudomonas* isolates from each FPES site that contain each ARG.

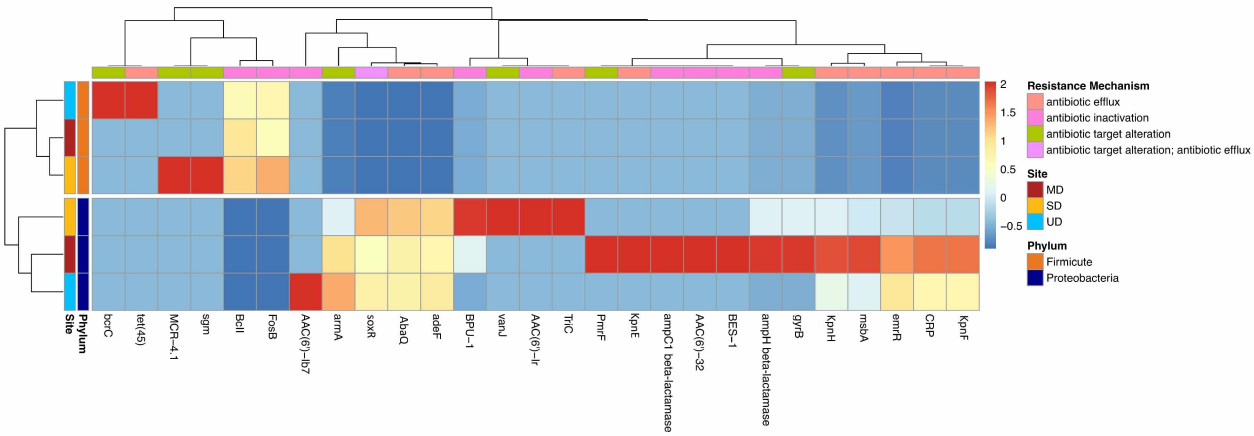
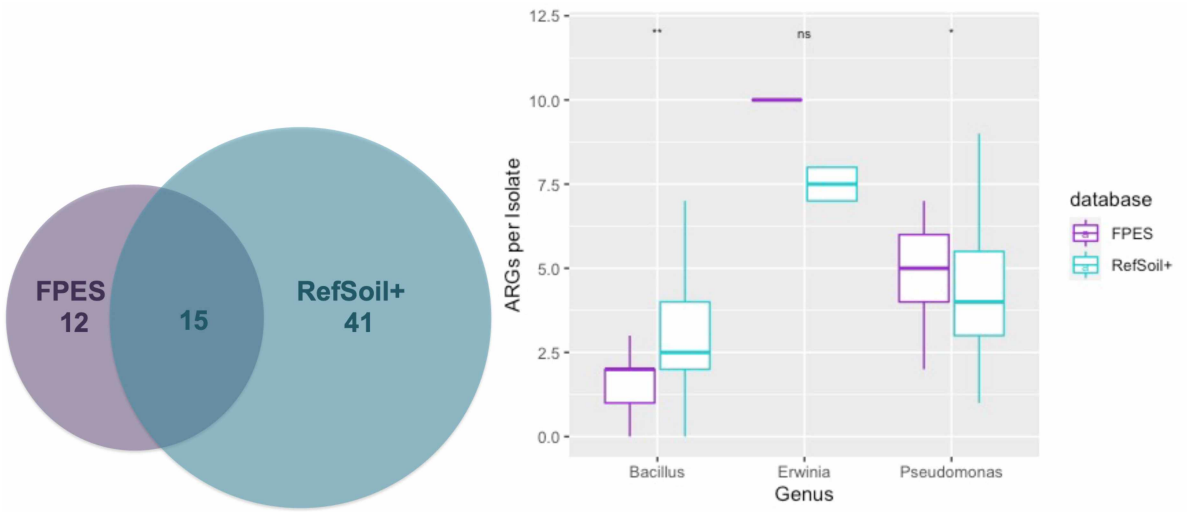


Figure 1.5. Heat map displaying the z-score by column of ARG count normalized by number of isolates in each group (phylum and FPES site). Annotation colors on the side show FPES site and phylum of each group and annotations on top show the resistance mechanism of the associated ARG. Dendograms display clustering based on Pearson correlation.



(**a**) (**b**)

Figure 1.6. (a) Types of ARGs by database (**b**) Boxplot of the number of antibiotic resistance genes per isolate by genus and database (FPES vs RefSoil+). Wilcoxon test significance p<0.01**, <0.05*, >0.1 ns.

Table 1.2. List of ARGs found on plasmids in FPES Isolates along with description of each ARG's resistance mechanism, drug class, gene family, and FPES host taxa.

PLASMID BORNE ARGs IN FPES ISOLATES					
Best Hit ARO	**Resistance Mechanism**	**Drug Class**	**AMR Gene Family**	**Genus origin**	**Count**
BES-1	inactivation	carbapenem; cephalosporin; penam	SIM beta-lactamase	*Pantoea*	1
bcrC	target alteration	peptide antibiotic	udecaprenyl pyrophosphate related proteins	*Bacillus*	2
TriC	efflux	triclosan	RND antibiotic efflux pump	*Pseudomonas*	2
KpnF	efflux	Broad Spectrum	MFS antibiotic efflux pump	*Erwinia*	1

1.5 Discussion

The Alaskan soil bacteria in this study harbored a diverse array of resistance determinants from all major mechanisms of antibiotic resistance, corroborating findings that suggest ARGs are ancient in origin and ubiquitous in soil-dwelling bacterial taxa [8,42]. Although we cannot draw direct functional conclusions from genomic data, such as if a resistance gene will be transcribed in an isolate, we did find that the high abundance of beta-lactam resistance genes in isolates directly corresponds with the high proportion of phenotypic resistance to the beta-lactam antibiotic screened, ampicillin. Of the seven isolates susceptible to ampicillin only two had a hit for a beta-lactamase gene, whereas the prevalence in the resistant isolates was much higher with 24 of the 32 resistant isolates encoding a beta-lactamase gene (Table A5). In terms of the effect of disturbance-induced thaw associated with FPES sites, we observed a positive trend in both the proportion of resistant isolates (Figure 1.2) and abundance of ARGs with disturbance level. There were also significantly more ARG copies per isolate in the MD site compared to the UD (Figure 1.3a). However, this difference is likely a result of the increasing number of randomly sampled Proteobacteria (including *Erwinia*, *Pseudomonas*, *Pantoea,* and *Serratia*) with thaw (n = 15 UD; 19 SD; 24 MD) since isolates from the Proteobacteria phyla had a significantly higher number of ARGs per genome compared to the Firmicutes sampled (*Bacillus* and *Exiguobacterium*) (Figure 1.3a).

Based on AIC model selection, we found that phylum had a stronger effect on the number of ARGs per isolate than FPES site or both FPES site and phylum. The link between host phylogeny and ARG abundance demonstrates how a loss of Firmicutes and enrichment of Proteobacteria as a result of community shifts could increase the abundance of ARGs within a

community. Although we cannot say from this specific dataset of cultured isolates which taxa are enriched across FPES sites, previous analyses conducted on uncultured metagenomic data from across 48 FPES cores in Seitz et al., 2020 identified over-representation of the phylum Proteobacteria in the disturbed cores and of the order *Bacillales* in undisturbed cores [24]. This enrichment in the metagenomic data parallels the enrichment of these taxa in this cultured subset of the community.

Along with the observed association of host taxa and ARG abundance, we found that the types of ARGs clustered by bacterial phylum rather than FPES site (Figure 1.4). Moreover, the predominant mechanisms of resistance (e.g., efflux, inactivation, target protection, and target alteration) were dependent on host taxa. This connection between host taxa and types of resistance determinants means that as microbial community composition shifts in response to permafrost thaw, so can the predominant taxa shaping the types and of ARGs within the resistome. Proteobacteria predominately harbored ARGs encoding efflux pumps (mean = 4.93 per isolate) and very few encoding antibiotic inactivating enzymes (mean = 0.28 per isolate) whereas the most abundant resistance mechanism in *Bacillus* spp. was antibiotic inactivation (mean = 1.5 per isolate) and the one of the lowest abundance mechanisms was ARGs encoding efflux pumps (mean = 0.03 per isolate).

Within each bacterial genus there were both ARGs unique to one genus and site, such as tet(45) in *Bacillus* from the UD site, and a distinct set of core ARGs that were chromosomally encoded and ubiquitous across thaw levels within a genus, such as *adeF* in *Pseudomonas* (Figure 1.4a) and *BcII* in *Bacillus* (Figure 1.4b). The genes unique to one site and taxa, although rare, are more likely accessory determinants that were acquired either through conjugation, transformation, or transduction from other members of the soil community and are therefore more of an interest in terms of clinical risk. The more widespread core genes are likely a result of clonal expansion and less prone to horizontal gene transfer compared to the aforementioned accessory determinants associated with genomic hotspots and mobile genetic elements such as integrons, plasmids, transposons [43,44]. Yet core resistance genes in soil bacteria still pose a risk because they have the potential to be mobilized through transformation or transduction and are widespread within taxa as an intrinsic part of the genome that likely plays a role in both the colonization of the rhizosphere and high-level antibiotic resistance associated with many environmental borne opportunistic pathogens such as *Pseudomonas aeruginosa* [45,46].

Some of the isolates in this study do in fact belong to taxa of known opportunistic human pathogens, such as *Pantonea agglomerans, Bacillus cereus,* and several *Pseudomonas spp.* and were found to carry both chromosomally encoded and plasmid-borne ARGs [47,48].

However, even nonpathogenic soil bacteria regularly interact with waterways, air, and built habitats, such as hospital surfaces generating a potential for HGT from one biosphere to another. When exposed to antibiotics, even the nonpathogenic commensal bacteria carrying resistance determinants acquired from environmental sources can be selected for promoting the clonal expansion and increased risk for spread of ARGs to the pathogenic bacteria the antibiotic is targeting [49]. A study by Hu et al. 2016 analyzed the mobilome of 23,425 bacterial genomes and found that mobile ARGs are mainly present in four bacterial phyla, the top two of which were *Proteobacteria* (399 mobile ARGs) and *Firmicutes* (86 mobile ARGs) [23]. All of the FPES isolates belong to these two phyla and six isolates were shown to carry plasmid-borne ARGs. Although only 1.6% of the ARGs copies from FPES isolates were located on plasmids, presence on these MGEs is telling of the low, but real, potential for ARGs from these soils to be disseminated. Moreover the higher number of plasmid-borne ARGs in FPES isolates compared to the RefSoil+ bacteria from the same genera, suggests the role local soil attributes have in selecting for plasmid-carriage further highlighting the clinical significance of ARGs harbored in Alaskan soils.

We found ARGs with high sequence identity and full-length coverage to those present in the CARD database. This presence of highly homologous ARGs highlights that resistance determinants in soils can be similar to those in clinical settings, rather than just ancient divergent homologs. However, the mean percent identity of ARGs in our isolates (68.4% ± 19.) suggests that many of the ARGs identified in our isolates are novel homologs. The most abundant ARG encoding antibiotic inactivating enzymes, fosB, had a high mean percent identity and full-length gene coverage to fosB genes in the CARD database. This gene encodes fosfomycin thiol transferase that confers resistance to an antibiotic derived from secondary metabolites produced by soil-dwelling bacteria including *Streptomyces* and *pseudomonads* [42]. Both *Streptomyces* and *pseudomonads* have been found to be abundant in FPES metagenomic datasets [24] highlighting the taxonomic potential for the production of fosfomycin at this locale. This potential along with the high abundance of fosB found in this study is suggestive of the selective advantage encoding antibiotic inactivating enzymes may have in competing against antibiotic producing bacterial taxa in the soil community.

In our soil isolates genes encoding efflux pumps were the most abundant with resistance-nodulation-cell division (RND) efflux pumps being the most abundant gene family found in all *Proteobacteria* isolates. RND efflux pumps have been described as a major tolerance mechanism allowing the effective extrusion of organic solvent from the interior of the cell to the exterior environment, this mechanism is found to be especially prevalent in

Pseudomonas [50]. As environmental change increasingly affects the arctic in the form of higher annual ambient air temperature and anthropogenic disturbance of soils, microbial life within the active layer have to cope with the release of biogenic volatile organic compounds (BVOC) amplified by thawing permafrost [51]. Efflux pumps could provide an effective mechanism for coping with toxic substances, such as BVOCs that will increase in concentration within active layer soils with permafrost thaw and antibiotics.

1.5.1. Limitations

There are a variety of published antibiotic resistance gene databases used for the annotation of resistance genes. These resources are often created through the curation of genes identified in the scientific literature, and only contain functional annotations for genes with published experimental data. Results from annotation of these databases can be reflective of the database chosen due to inequality across gene and protein annotation resources [52]. Moreover, since antibiotic resistance is more commonly analyzed in clinical situations, genes in this database can be biased towards clinical phylogenies and are restricted to known resistance determinants thus missing novel resistant determinants in environmental communities that could be identified via functional vector based tests [53].

One of the most widely used ARG databases is the Comprehensive Antibiotic Resistance Database (CARD) [54]. The CARD database provides well-developed and extensive antibiotic resistance ontology (ARO) and monthly curation updates to include the most up to date ARG reference data which is exclusively derived from peer-reviewed publications validated by clinical or experimental data [55]. A 2016 study found that CARD was able to outperform other popular AR databases including ARDB, ResFinder, and CBMAR by correctly identifying down to a variant level for all variants of the two genes tested, bla_{VIM} and bla_{NDM}, and unlike any of the other databases was able to accurately identify the maximum number of resistance genes from the whole genome sequences of 3 strains of methicillin resistant *Staphylococcus aureus* [56]. Based on these findings, we decided to use CARD for the annotation of our isolates in this study.

Another limitation of this study is that we cannot attribute differences between sampling sites specifically to a single factor such as permafrost thaw, vegetation shifts, or soil characteristics. Rather we attributed site level differences as a culmination of these biotic and abiotic factors present at FPES. This study does not provide direct evidence of a horizontal gene transfer event from soil bacteria to pathogenic bacteria; however, we do identify the remarkable abundance and diversity of antibiotic resistance in Alaskan soils. The resistance

genes identified here are a restricted representation of the Alaskan soil community because this study is limited in geography, to the CARD ARG database, and to the culturable nonfastidious aerobic to facultative anaerobic bacteria of the phyla *Proteobacteria* and *Firmicutes*. Despite this restricted scope, the abundance, diversity, and presence of ARGs on plasmids is suggestive of the extent these soils have as a reservoir for antibiotic resistance and for the potential to compromise health. Furthermore, the over-representation of Proteobacteria in disturbed cores and *Bacillales* in undisturbed cores revealed from previous metagenomic analyses of FPES soils parallels the significant difference we observed in the number of ARGs per isolate from MD and SD sites driven by the shifting proportion of Firmicutes and Proteobacteria.

1.6 Conclusions

In this study, we unearthed antibiotic resistance in bacteria from active layer soils of Interior Alaska associated with permafrost thaw. Most bacterial isolates from this locale were phenotypically resistant to clinically significant antibiotics and encoded both highly homologous and divergent homologs to previously identified resistance determinants. We even identified several resistance genes on plasmids, highlighting the risk these soils have in the dissemination of antibiotic resistance. The significant difference in ARG abundance between the most disturbed and undisturbed isolates driven by a difference in the isolate's phylogeny, along with previously identified enrichment in Proteobacteria at this site, highlights how community shifts with thaw may enrich for the taxa that increase the abundance of resistance genes comprising the resistome. When compared to the genomes of soil bacteria from a global database, RefSoil+, there were differences in the number of ARGs per isolate and a higher abundance of plasmid-borne ARGs in our isolates from equivalent genera that emphasized how local biotic and abiotic factors shape fine scale differences in the resistance profiles. Moreover, the high-quality whole genome assemblies generated in this study can be used for future analyses into diverse areas of research, such as the coexistence of virulence factors and antibiotic resistance [57], the genomics of cold adaptation of psychrophilic microorganisms [58], and more in depth analyses into the mobile genetic elements that have the potential to propagate the spread of resistance. As antibiotic resistance continues to emerge and rapidly spread in clinical settings, studies like this will be imperative for building insight into the ecology of environmental resistance genes in order to understand the threat they pose to human health.

1.7 Author Contributions: Conceptualization T.H. and D.D.; methodology T.H. and D.D.; software T.H. and D.D.; validation T.H. and D.D.; formal analysis T.H. and D.D.; investigation

T.H. and D.D.; resources T.H. and D.D.; data curation T.H. and D.D.; writing—original draft preparation T.H.; writing—review and editing T.H. and D.D.; visualization T.H.; supervision D.D.; project administration T.H. and D.D.; funding acquisition, T.H. and D.D. All authors have read and agreed to the published version of the manuscript.

1.8 Funding: Research reported here was supported by BLaST through the National Institute of General Medical Sciences of the National Institutes of Health under awards UL1GM118991, TL4GM118992, and RL5GM118990. Research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant P20GM103395.

1.9 Institutional Review Board Statement: Not applicable.

1.10 Informed Consent Statement: Not applicable.

1.11 Data Availability Statement: This genome project is indexed at GenBank under BioProject accession numbers PRJNA684363, PRJNA486198, PRJNA486356, PRJNA525875. These whole genomes have been deposited in GenBank under the accession nos. SAMN17054779-17054865.

1.12 Acknowledgments: Many thanks go to Ursel Schütte, Taylor Seitz, Madeline McCarthy, Scout McDougal, Anne-Lise Ducluzeau, Jennie Humphrey and Sara Kline for laboratory help. Tom Douglas from the Cold Regions Research and Engineering Laboratory (CRREL) Alaska who provided field site access. We acknowledge generous support from the Institute of Arctic Biology, Alaska INBRE, and the BLaST program.

1.13 Conflicts of Interest: The authors declare no conflict of interest.

1.14 References

1. O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*; Review on Antimicrobial Resistance: London, UK, 2016; pp. 1–84.
2. Ghosh, S.; LaPara, T.M. The Effects of Subtherapeutic Antibiotic Use in Farm Animals on the Proliferation and Persistence of Antibiotic Resistance among Soil Bacteria. *ISME J.* **2007**, *1*, 191–203, doi:10.1038/ismej.2007.31.
3. He, Y.; Yuan, Q.; Mathieu, J.; Stadler, L.; Senehi, N.; Sun, R.; Alvarez, P.J.J. Antibiotic Resistance Genes from Livestock Waste: Occurrence, Dissemination, and Treatment. *npj Clean Water* **2020**, *3*, 1–11, doi:10.1038/s41545-020-0051-0.

4. Berg, J.; Tom-Petersen, A.; Nybroe, O. Copper Amendment of Agricultural Soil Selects for Bacterial Antibiotic Resistance in the Field. *Lett. Appl. Microbiol.* **2005**, *40*, 146–151, doi:10.1111/j.1472-765X.2004.01650.x.
5. Djordjevic, S.P.; Stokes, H.W.; Chowdhury, P.R. Mobile Elements, Zoonotic Pathogens and Commensal Bacteria: Conduits for the Delivery of Resistance Genes into Humans, Production Animals and Soil Microbiota. *Front. Microbiol.* **2013**, *4*, 86, doi:10.3389/fmicb.2013.00086.
6. Aminov, R.I. The Role of Antibiotics and Antibiotic Resistance in Nature. *Environ. Microbiol.* **2009**, *11*, 2970–2988, doi:10.1111/j.1462-2920.2009.01972.x.
7. Caruso, G.; Giammanco, A.; Cardamone, C.; Oliveri, G.; Mascarella, C.; Capra, G.; Fasciana, T. Extra-Intestinal Fluoroquinolone-Resistant Escherichia Coli Strains Isolated from Meat. *BioMed. Res. Int.* **2018**, *2018*, e8714975, doi:10.1155/2018/8714975.
8. D'Costa, V.M.; King, C.E.; Kalan, L.; Morar, M.; Sung, W.W.L.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; et al. Antibiotic Resistance Is Ancient. *Nature* **2011**, *477*, 457–461, doi:10.1038/nature10388.
9. Forsberg, K.J.; Reyes, A.; Wang, B.; Selleck, E.M.; Sommer, M.O.A.; Dantas, G. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* **2012**, *337*, 1107–1111, doi:10.1126/science.1220761.
10. Baltz, R.H. Renaissance in Antibacterial Discovery from Actinomycetes. *Curr. Opin. Pharm.* **2008**, *8*, 557–563, doi:10.1016/j.coph.2008.04.008.
11. Hall, B.G.; Barlow, M. Evolution of the Serine Beta-Lactamases: Past, Present and Future. *Drug Resist. Updat.* **2004**, *7*, 111–123, doi:10.1016/j.drup.2004.02.003.
12. de Lima Procópio, R.E.; da Silva, I.R.; Martins, M.K.; de Azevedo, J.L.; de Araújo, J.M. Antibiotics Produced by Streptomyces. *Braz. J. Infect. Dis.* **2012**, *16*, 466–471, doi:10.1016/j.bjid.2012.08.014.
13. Watve, M.G.; Tickoo, R.; Jog, M.M.; Bhole, B.D. How Many Antibiotics Are Produced by the Genus Streptomyces? *Arch. Microbiol.* **2001**, *176*, 386–390, doi:10.1007/s002030100345.
14. Finley, R.L.; Collignon, P.; Larsson, D.G.J.; McEwen, S.A.; Li, X.-Z.; Gaze, W.H.; Reid-Smith, R.; Timinouni, M.; Graham, D.W.; Topp, E. The Scourge of Antibiotic Resistance: The Important Role of the Environment. *Clin. Infect. Dis.* **2013**, *57*, 704–710, doi:10.1093/cid/cit355.
15. Forsberg, K.J.; Patel, S.; Gibson, M.K.; Lauber, C.L.; Knight, R.; Fierer, N.; Dantas, G. Bacterial Phylogeny Structures Soil Resistomes across Habitats. *Nature* **2014**, *509*, 612–616, doi:10.1038/nature13377.

16. Aminov, R.I.; Mackie, R.I. Evolution and Ecology of Antibiotic Resistance Genes. *FEMS Microbiol. Lett.* **2007**, *271*, 147–161, doi:10.1111/j.1574-6968.2007.00757.x.
17. Schuur, E.A.G.; Mack, M.C. Ecological Response to Permafrost Thaw and Consequences for Local and Global Ecosystem Services. *Annu. Rev. Ecol. Evol. Syst.* **2018**, *49*, 279–301, doi:10.1146/annurev-ecolsys-121415-032349.
18. Box, J.E.; Colgan, W.T.; Christensen, T.R.; Schmidt, N.M.; Lund, M.; Parmentier, F.-J. W.; Brown, R.; Bhatt, U.S.; Euskirchen, E.S.; Romanovsky, V.E.; et al. Key Indicators of Arctic Climate Change: 1971–2017. *Environ. Res. Lett.* **2019**, *14*, 045010, doi:10.1088/1748-9326/aafc1b.
19. McGuire, A.D.; Lawrence, D.M.; Koven, C.; Clein, J.S.; Burke, E.; Chen, G.; Jafarov, E.; MacDougall, A.H.; Marchenko, S.; Nicolsky, D.; et al. Dependence of the Evolution of Carbon Dynamics in the Northern Permafrost Region on the Trajectory of Climate Change. *PNAS* **2018**, *115*, 3882–3887, doi:10.1073/pnas.1719903115.
20. Thoman, R.; Walsh, J. Alaska's Changing Environment—International Arctic Research Center https://uaf-iarc.org/our-work/alaskas-changing-environment/ (accessed Dec 16, 2020).
21. Douglas, T.; Kanevskiy, M.; Romanovsky, V.; Shur, Y.; Yoshikawa, K. Permafrost Dynamics at the Fairbanks Permafrost Experimental Station Near Fairbanks, Alaska. *Institute of Northern Engineering, University of Alaska Fairbanks* **2008**, 1, 373-378.
22. Schuur, E.A.G.; Crummer, K.G.; Vogel, J.G.; Mack, M.C. Plant Species Composition and Productivity Following Permafrost Thaw and Thermokarst in Alaskan Tundra. *Ecosystems* **2007**, *10*, 280–292, doi:10.1007/s10021-007-9024-0.
23. Hu, Y.; Yang, X.; Li, J.; Lv, N.; Liu, F.; Wu, J.; Lin, I.Y.C.; Wu, N.; Weimer, B.C.; Gao, G.F.; et al. The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes. *Appl. Environ. Microbiol.* **2016**, *82*, 6672–6681, doi:10.1128/AEM.01802-16.
24. Seitz, T.J.; Schütte, U.M.E.; Drown, D.M. Soil Disturbance Affects Plant Growth via Soil Microbial Community Shifts. *bioRxiv* **2020**, doi:10.1101/2020.10.16.343053.
25. Dunivin, T.K.; Choi, J.; Howe, A.; Shade, A. RefSoil+: A Reference Database for Genes and Traits of Soil Plasmids. *mSystems* **2019**, *4*, e00349-18, doi:10.1128/mSystems.00349-18.
26. Linell, K.A. Long-term effects of vegetation cover on permafrost stability in an area of discontinuous permafrost. In Proceedings of the Permafrost: North American Contribution to the Second Internal Conference, Yakutsk, U.S.S.R., 13–28 July 1973; *National Academy Press: Washington, DC*, USA; pp.688–693.

27. Johnstone, J.F.; Hollingsworth, T.N.; Chapin, F.S. A Key for Predicting Postfire Successional Trajectories in Black Spruce Stands of Interior Alaska. *Gen. Tech. Rep. PNW-GTR-767. Portland OR: U.S. Dep. Agric. For. Serv. Pac. Northwest. Res. Stn. 37 P.* **2008**, *767, 4-8*, doi:10.2737/PNW-GTR-767.
28. Hudzicki, J. Kirby-Bauer Disk Diffusion Susceptibility Test Protocol. Available online: https://www.asmscience.org/content/education/protocol/protocol.3189 (accessed on 16 December 2020).
29. Krueger, F. TrimGalore. Available online: https://github.com/FelixKrueger/TrimGalore (accessed on 16 December 2020).
30. Wick, R. Filtlong. Available online: https://github.com/rrwick/filtlong (accessed on 16 December 2020).
31. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546, doi:10.1038/s41587-019-0072-8.
32. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads. *PLoS Comput. Biol.* **2017**, *13,* e1005595, doi:10.1371/journal.pcbi.1005595.
33. Haan, T.; McDougall, S.; Drown, D.M. Complete Genome Sequence of Bacillus Mycoides TH26, Isolated from a Permafrost Thaw Gradient. *Microbiol. Resour. Announc.* **2019**, *8*, e00507-19, doi:10.1128/MRA.00507-19.
34. Humphrey, J.; Seitz, T.; Haan, T.; Ducluzeau, A.-L.; Drown, D.M. Complete Genome Sequence of Pantoea Agglomerans TH81, Isolated from a Permafrost Thaw Gradient. *Microbiol. Resour. Announc.* **2019**, *8*, e01486-18, doi:10.1128/MRA.01486-18.
35. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A Modular and Extensible Implementation of the RAST Algorithm for Building Custom Annotation Pipelines and Annotating Batches of Genomes. *Sci. Rep.* **2015**, *5*, 8365, doi:10.1038/srep08365.
36. Jia, B.; Raphenya, A.; Alcock, B.; Waglechner, N.; Guo, P.; Tsang, K.; Lago, B.; Dave, B.; Pereira, S.; Sharma, A.; et al. CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2016**, *45*, D566–D573, doi:10.1093/nar/gkw1004.
37. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **2000**, *7*, 203–214, doi:10.1089/10665270050081478.

38. Shade, A. RefSoil_plasmids; Available online: https://github.com/ShadeLab/RefSoil_plasmids (accessed on 16 December 2020).
39. RStudio Team. *RStudio: Integrated Development for R*; RStudio: Boston, MA, USA, 2020.
40. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Use R!; Springer-Verlag: New York, NY, USA, 2009; doi:10.1007/978-0-387-98141-3.
41. Kolde, R. Pheatmap: Pretty Heatmaps. R Package Version 1.0.12. 2018. Available online: http://CRAN.R-project.org/package=pheatmap (accessed on 16 December 2020).
42. Wright, G.D.; Poinar, H. Antibiotic Resistance Is Ancient: Implications for Drug Discovery. *Trends Microbiol.* **2012**, *20*, 157–159, doi:10.1016/j.tim.2012.01.002.
43. Harrison, E.; Brockhurst, M.A. Plasmid-Mediated Horizontal Gene Transfer Is a Coevolutionary Process. *Trends Microbiol.* **2012**, *20*, 262–267, doi:10.1016/j.tim.2012.04.003.
44. Oliveira, P.H.; Touchon, M.; Cury, J.; Rocha, E.P.C. The Chromosomal Organization of Horizontal Gene Transfer in Bacteria. *Nat. Commun.* **2017**, *8*, 841, doi:10.1038/s41467-017-00808-w.
45. Martinez, J.L. The Role of Natural Environments in the Evolution of Resistance Traits in Pathogenic Bacteria. *Proc. Biol. Sci.* **2009**, *276*, 2521–2530, doi:10.1098/rspb.2009.0320.
46. Gallagher, L.A.; Lee, S.A.; Manoil, C. Importance of Core Genome Functions for an Extreme Antibiotic Resistance Trait. *mBio* **2017**, *8*, doi:10.1128/mBio.01655-17.
47. Cruz, A.T.; Cazacu, A.C.; Allen, C.H. Pantoea Agglomerans, a Plant Pathogen Causing Human Disease. *J. Clin. Microbiol.* **2007**, *45*, 1989–1992, doi:10.1128/JCM.00632-07.
48. Kotiranta, A.; Lounatmaa, K.; Haapasalo, M. Epidemiology and Pathogenesis of Bacillus Cereus Infections. *Microbes Infect.* **2000**, *2*, 189–198, doi:10.1016/s1286-4579(00)00269-0.
49. Lax, S.; Gilbert, J.A. Hospital-Associated Microbiota and Implications for Nosocomial Infections. *Trends Mol. Med.* **2015**, *21*, 427–432, doi:10.1016/j.molmed.2015.03.005.
50. Ramos, J.L.; Gallegos, M.T.; Marqués, S.; Ramos-González, M.I.; Espinosa-Urgel, M.; Segura, A. Responses of Gram-Negative Bacteria to Certain Environmental Stressors. *Curr. Opin. Microbiol.* **2001**, *4*, 166–171, doi:10.1016/s1369-5274(00)00183-1.
51. Kramshøj, M.; Albers, C.N.; Holst, T.; Holzinger, R.; Elberling, B.; Rinnan, R. Biogenic Volatile Release from Permafrost Thaw Is Determined by the Soil Microbial Sink. *Nat. Commun.* **2018**, *9*, 3412, doi:10.1038/s41467-018-05824-y.
52. Haynes, W.A.; Tomczak, A.; Khatri, P. Gene Annotation Bias Impedes Biomedical Research. *Sci. Rep.* **2018**, *8*, 1362, doi:10.1038/s41598-018-19333-x.

53. Allen, H.K.; Moe, L.A.; Rodbumrer, J.; Gaarder, A.; Handelsman, J. Functional Metagenomics Reveals Diverse β-Lactamases in a Remote Alaskan Soil. *ISME J.* **2009**, *3*, 243–251, doi:10.1038/ismej.2008.86.
54. McArthur, A.G.; Waglechner, N.; Nizam, F.; Yan, A.; Azad, M.A.; Baylay, A.J.; Bhullar, K.; Canova, M.J.; De Pascale, G.; Ejim, L.; et al. The Comprehensive Antibiotic Resistance Database. *Antimicrob. Agents Chemother.* **2013**, *57*, 3348–3357, doi:10.1128/AAC.00419-13.
55. Alcock, B.P.; Raphenya, A.R.; Lau, T.T.Y.; Tsang, K.K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L. V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2020**, *48*, D517–D525, doi:10.1093/nar/gkz935.
56. Xavier, B.B.; Das, A.J.; Cochrane, G.; De Ganck, S.; Kumar-Singh, S.; Aarestrup, F.M.; Goossens, H.; Malhotra-Kumar, S. Consolidating and Exploring Antibiotic Resistance Gene Data Resources. *J. Clin. Microbiol.* **2016**, *54*, 851–859, doi:10.1128/JCM.02717-15.
57. Fasciana, T.; Gentile, B.; Aquilina, M.; Ciammaruconi, A.; Mascarella, C.; Anselmo, A.; Fortunato, A.; Fillo, S.; Petralito, G.; Lista, F.; et al. Co-Existence of Virulence Factors and Antibiotic Resistance in New Klebsiella Pneumoniae Clones Emerging in South of Italy. *BMC Infect. Dis.* **2019**, *19*, 928, doi:10.1186/s12879-019-4565-3.
58. D'Amico, S.; Collins, T.; Marx, J.-C.; Feller, G.; Gerday, C. Psychrophilic Microorganisms: Challenges for Life. *EMBO Rep.* **2006**, *7*, 385–389, doi:10.1038/sj.embor.7400662.

CHAPTER 2:

Disturbance to Subarctic Soils Shapes the Resistome via Shifts in Microbial Community Composition[2]

2.1 Abstract

The worldwide spread of antibiotic resistance in pathogens is a significant threat to human and animal health. Soils are one of the evolutionary origins of both antibiotic production and resistance. Therefore, examining soil resistomes can help explain the global emergence of antibiotic resistance genes (ARGs) across microbial species and biomes. To assess risks posed by soil-borne ARGs it is imperative to identify the predominant ARG host taxa, types of ARGs that may emerge, and the specific soil conditions that promote bacterial communities enriched with resistant taxa. Previous studies have identified a diverse pool of ARGs in undisturbed subarctic soils, however, conditions in these high latitude biomes are rapidly changing with an increase in the frequency and severity of soil disturbance events such as wildfires and thermokarst formation. Climatic and anthropogenic disturbances to subarctic soils have been shown to augment permafrost thaw and alter microbial communities, however, no study has explored the effect of disturbance-induced thaw on the active layer resistome. In this study we ask how disturbance-induced shifts in active layer community composition affects ARG abundance, composition, and mobility. To do this we collected soil cores from a permafrost thaw gradient in Fairbanks, Alaska and employed long read shotgun metagenomics to examine the relationship between ARGs, disturbance, and community composition. We then used Hi-C proximity ligation to construct metagenomic assembled genomes (MAGs) which unearthed the contribution of individual bacterial genomes to ARGs in the community. We found ARGs from all mechanisms of resistance with genes encoding beta-lactamases as the most abundant across disturbance treatments. The high abundance of genes encoding antibiotic inactivating enzymes paired with the finding that *Streptomycetaceae,* a family of prolific antibiotic producers, was the second most abundant family across soil cores implicates the presence of selective pressures from antibiotic producers in subarctic soils. ARG abundance was found to have a quadratic relationship with disturbance and negative linear relationship with year highlighting the complex interplay soil conditions have in structuring the taxa that enrich ARGs in the community. Our MAGs revealed the microbial mechanism for this relationship by unearthing the

---

[2] To be submitted for publication as: **Haan, T. J.,** & Drown, D. M. (2021). Disturbance to Subarctic Soils Shapes the Resistome via Shifts in Microbial Community Composition. *Environmental Microbiology.*

abundance of ARGs per genome that was significantly different between phyla. For example, in undisturbed soils the phylum Acidobacteria was more abundant and had significantly more ARGs per genome than all other phyla examined except Proteobacteria, which was enriched in disturbed soils. In terms of mobility, a majority of ARGs were chromosomally encoded suggesting resistance has an intrinsic role in bacterial evolution, however, we did identify several plasmids-borne and integron associated ARGs highlighting the potential for horizontal gene transfer. Our findings emphasize how disturbance-induced changes to soils of interior Alaska enrich specific bacterial taxa that structure the resistome thus generating a reservoir of ARGs that has the potential to compromise One Health in the subarctic.

2.2 Introduction

The rise of antibiotic resistance is a global health crisis continually impeding our ability to treat bacterial infections in humans, animals, and plants alike (CDC, 2019). The genes driving this crisis recurrently emerge, disseminate, and persist in bacterial populations thereby reducing the efficacy of antibiotics (O'Neill 2016). However, the lack of additional therapies for treating bacterial infections has led to our dependence on antibiotics. We now know extensive antibiotic use across clinical, veterinary, and agricultural settings generate strong selective pressures favoring resistance. Resistance phenotypes can arise in bacteria as a result of random mutations to drug targets or through horizontal gene transfer of mobile genetic elements (MGEs) carrying antibiotic resistance genes (ARGs) encoding mechanisms such as inactivating enzymes, efflux pumps, or proteins that protect drug targets (Djordjevic et al. 2013). However even microbial habitats with minimal anthropogenic input, such as 30,000-year-old Beringian permafrost sediment, have been identified as reservoirs for bacteria encoding ARGs (D'Costa et al. 2016). The presence of these ancient and diverse ARGs in unpolluted environments, paired with the fact that most antibiotics are derived from secondary metabolites produced by soil microorganisms, has made it evident that human use of antibiotics isn't the sole force driving the evolution of ARGs (Wright et al. 2012). More likely, selective pressures favoring resistance extends to microbial biomes predisposed to evolutionary pressures such as resource competition via competitive inhibition (Nesme et al. 2014).

Soils are one of the richest microbial habitats in terms of both microbial diversity and abundance (Torsvik et al. 2002). Compared to other environmental habitats, soils host a high prevalence of antibiotic producing and resistant bacterial taxa (Surette & Wright 2017). In fact, most clinically significant antibiotics today are derived from bioactive compounds synthesized by soil-dwelling bacteria and fungi, namely *Streptomyces*, a bacterial genus that produces around

two thirds of antibiotics (Watve et al. 2001). Thus, the higher prevalence of ARGs in soils compared to non-soil biomes has been repeatedly attributed to the hypothesized co-evolutionary 'arms-shields race' between producers and resistant bacteria (Nesme et al. 2014). Although hypothetically probable, it has been difficult to disentangle the exact selective forces driving the evolution and dissemination of soil-borne ARGs. This is due to soil's biological complexity and technical challenges in measuring antibiotic concentrations. Despite these technical challenges, advancements in next-generation sequencing technologies have revealed that soil-borne ARG's have a worldwide distribution (Nesme et al. 2014), can confer resistance (Allen et al. 2009), and can be highly similar to those circulating in human pathogens (Forsberg et al. 2012). These findings implicate soils as a risk to global health by acting as a conduit of ARGs between disparate groups of bacteria, such as non-pathogenic soil saprophytes to human pathogens (Finley et al. 2013, Raphael & Riley 2017).

Despite the apparent risk posed by soils, factors influencing ARG composition and abundance in soils remain largely unknown. Large-scale metagenomic studies have previously unveiled bacterial community composition as the primary determinant of soil ARG content (Forsberg et al. 2014) and have revealed potential land-use practices that enrich ARG abundance in soils. For example, ARGs were significantly enriched in agricultural soils amended with manure (Udikovic-Kolic et al. 2014), copper (Berg et al. 2005), and ammonium nitrogen (Sun et al. 2014) along with soils polluted with heavy metals because of waste management and mining (Xie et al. 2010). However, factors that at a global level alter the microbial communities that structure ARGs in soils, such as climate change, are largely understudied. Soils from high latitude ecosystems are already experiencing the effects from climate change because of polar amplification. This phenomenon has contributed to temperatures 1.5 to 4.5 times higher in arctic and subarctic ecosystems than the global average over the last half-century (Holland & Bitz 2003). As a result, soils at high latitudes may provide critical insight into how factors associated with climactic driven disturbances will affect ARG abundance, selection, and dissemination. Such factors include soil disturbance events like wildfires (Whitman et al. 2019), flooding (Pérez-Valdespino et al. 2021), and permafrost thaw.

Since Alaskan soils are one high latitude microbial biosphere at the forefront of change, in this study we explore how disturbance to Alaskan soils affects the evolution of antibiotic resistance. In Alaska human expansion, land use, and climate change increase the frequency and intensity of soil disturbance events, like wildfires and thermokarst formation (Schuur & Mack 2018). With approximately 85% of Alaska's landmass underlain by permafrost (McGuire et al. 2018), disturbance to Alaskan vegetation and surface soils are of great global concern. These

disturbances augment permafrost thaw releasing ancient carbon that is then decomposed by microbes and released to atmosphere in the form of CH_4 and CO_2 further amplifying climate change (Schuur & Abbott 2011). Additionally, disturbances shift biogeochemical properties of active layer soils that overlay permafrost (Douglas et al. 2008). Changing physical and chemical properties (e.g., nutrient availability, pH, rooting space, and soil moisture) shift microbial communities negatively affecting plant health (Seitz et al. 2021, Schütte et al. 2019) and ecosystem functions (Whitman et al. 2019). Moreover, these shifts in microbial communities and soil properties may impact the conditions that favor the selection of bacterial taxa enriched in ARGs (Haan & Drown 2021). For instance, wildfire in the Canadian boreal forest was shown to shift soil niches enriching the prolific antibiotic producing microbial taxa *Penicillium* and *Streptomyces* (Whitman et al. 2019). Thus, disturbance in this system could breed a co-evolutionary arms-shield race that directly selects for resistant bacteria by enriching antibiotic producers through shifting niche availability (Hibbing et al. 2010). Disturbance to subarctic soils typical of Interior Alaska could also indirectly select for ARGs by favoring mechanisms that allow bacteria to concurrently cope with antibiotics and biological stressors released from permafrost, such as biogenic volatile organic compounds (Kramshøj et al. 2018). For example, efflux pumps are proteins encoded by bacteria that have been shown to confer multidrug resistance and promote environmental stress tolerance (Ramos et al. 2001). Thus, disturbance could promote multidrug resistant bacteria encoding efflux pumps in soils. Lastly, environmental stress may stimulate intra-genomic (e.g., integrons) and inter-species gene transfer (Velkov 1999). As a result, disturbance may increase the rate of ARG dissemination and the abundance of ARGs housed on conjugative plasmids (Djordjevic et al. 2013)

Previously at our study site, a disturbance-induced permafrost thaw gradient in Interior Alaska, we found that around 90% of cultured isolates were resistance to at least one of the five antibiotics tested with over 45% displaying multidrug resistance (Haan & Drown 2021). Through whole genome sequencing of these isolates, we revealed that while disturbance treatment had significant roles in determining the abundance and composition of ARGs, phylogeny had a more significant effect. Proteobacteria encoded significantly more ARGs than Firmicutes. That analysis, paired with significant community shifts at our site in which Proteobacteria were enriched in the disturbed treatment (Seitz et al. 2021), led us to hypothesize how community shifts with disturbance could impact the abundance of ARGs in subarctic soils. However, Haan & Drown (2021) was limited to cultivable bacteria that represent less than 1% of total bacterial diversity in soils warranting further metagenomics. Thus, this study addresses this limitation by employing both long read shotgun metagenomics and Hi-C proximity ligation to unearth the

relationship between disturbance and the resistome at a community level. We hypothesize that a disturbance-induced shift in the subarctic soil community will structure the composition, abundance, and mobility of ARGs comprising the resistome by selecting for bacterial taxa enriched with ARGs (e.g., Proteobacteria).

Long read sequencing technologies (e.g., Nanopore and PacBio) have massively improved over the last decade and are especially promising for the analysis of environmental resistance. These technologies can be more costly than other methods, like qPCR, for estimating a gene's abundance from environmental samples. However, long read shotgun sequencing of total genomic DNA provides major advantages by avoiding amplification and primer biases while increasing the power to detect either highly similar or divergent ARG homologs (Liu et al. 2019). Integron gene cassettes, a major MGE associated with the dissemination of resistance, are important for revealing the mobility of ARGs in the environment (Stalder et al. 2012). Unlike short reads that often require assembly to capture ARG-MGE associations, long reads sequence data can be directly annotated with a higher chance of capturing a full-length gene and flanking MGEs (Huson et al. 2018).

Here, we use long read shotgun sequencing to identify differences in the predominant ARG and ARG host taxa with disturbance to unveil how disturbance-induced community shifts in Interior Alaskan soils structure the resistome. We then determine how individual bacterial taxa contribute to the resistome by examining genome-level ARG composition and abundance by reconstructing metagenomic-assembled genomes (MAG) using Hi-C proximity ligation. Previous studies have demonstrated the ability of Hi-C proximity ligation with culture-independent de novo deconvolution to produce high quality MAGs that reveal plasmid-host associations reconstructed from microbial communities without the need for prior information (e.g., reference genomes; Press et al. 2017, Stadler et al. 2019). These studies were conducted on microbial communities much less complex than soils. In complex communities, like soils, traditional binning methods often produce incomplete MAGs, especially from rare taxa. In this study, we evaluate the ability of Hi-C to construct individual genomes that reveal ARG profiles from complex communities. Together our results unveil how disturbance to Alaskan soils may act as a selective force driving subarctic soil microbial community composition that structures the types and abundance of ARGs. These findings have implications in terms of how soil disturbance, such as climatic-driven events like wildfires, affect the evolution of ARGs that may disseminate and emerge in pathogenic taxa as permafrost thaws and human expansion into the subarctic increases.

2.3 Methods

2.3.1 Study Site.

Soil samples were collected from a permafrost thaw gradient in Interior Alaska known as the Fairbanks Permafrost Experiment Station (FPES) (64.875646°N, 147.668981°W). FPES is an ice-rich permafrost site on a gentle southward slope six kilometers outside of the city of Fairbanks in Interior Alaska. The Army Corps of Engineers established the site as part of the Cold Regions Research and Engineering Lab to simulate soil disturbance events, such as wildfire and land-use, so that the long-term ecological effects of disturbance-induced permafrost thaw could be studied. FPES consists of three 3,721 m^2 experimental plots with increasing levels of disturbance to vegetation; Undisturbed (UD), Semi-Disturbed (SD), and Most-Disturbed (MD). When these treatments were generated in 1946, the UD treatment was left pristine, SD was cleared of trees while roots and soil organic matter were left intact, and MD was stripped of all vegetation and surface soil organic matter down to the mineral soil. After 25 years, the pristine plot (UD) was found to have little to no thaw during periods of maximal thaw, the semi-disturbed treatment had up to 4.5 meters of thaw with a stabilized permafrost table, and the MD treatment had up to 9.8 meters of thaw with continual downward migration of permafrost. This evidence supported previous work suggesting disturbance to boreal vegetation significantly affects the long-term stability of permafrost (Douglas et al. 2008).

Vegetation at the FPES is typical of the Alaskan Interior-subarctic taiga forest. Soils from the UD treatment can be classified as mesic (Johnstone et al. 2008) consisting of tan silt and windblown loess with a variable organic horizon thickness (2-35 cm). The UD treatment is a relatively open black spruce stand (*Picea mariana*) with an understory of continuous thick moss layer interspersed with low-bush cranberry (*Vaccinium vitis-idaea*) and Labrador tea (*Rhododendron groenlandicum*). The semi-disturbed treatment has reestablished the vegetation of a dense boreal forest after 25 years. It is now a mixed stand dominated by black spruce (*Picea mariana*), Alaskan paper birch (*Betula neoalaskana*), and willow (*Salix alaxensis*). The understory contains a mixture of Labrador tea (*Rhododendron groenlandicum*), Peltigera lichen, roses, horsetail, cloudberry, and small amounts of grass with little litter cover. The MD treatment is continually evolving from its current state of a shrub, birch, and willow forest to one with a higher density of spruce trees and moss. It is currently an open shrub land dominated by willows (*Salix alaxensis*) and a developing over story up to 5 m tall of Alaskan birch (*Betula neoalaskana*) and black spruce (*Picea mariana*). The understory contains many grasses, clovers, horse-tail (*Equisetum*), and some bare ground. There is no permafrost within the top 4.7 m in either disturbed treatments (Douglas et al. 2008).

2.3.2 Soil Collection for Long Read Metagenomics.

We collected 48 paired active layer soil cores from FPES in May 2018 (Seitz et al. 2021) and May 2019 yielding a total of 96 cores with 16 cores per FPES disturbance treatment each year. The 4.5 cm wide by 10 cm deep cores were sampled using a sterilized corer along an established grid layout with four 1 m by 1 m quadrants at each FPES treatment. The corners of each quadrant were sampled allowing us to capture fine-scale heterogeneity present within each disturbance treatment. Prior to coring the top layer of moss and vegetation was removed. After extracting the soil core from the corer, soil was immediately stored in a cooler throughout sample collection and transportation process. Upon arriving at the lab, soil cores were stored at 4° C to be used between 7 to 45 days later for DNA extractions. For DNA extraction, soil cores had the outer portion of each core removed using a sterile scalpel to prevent contamination from any exogenous cells. The interiors of each core were then sub-sampled along a transect using sterile forceps along a depth gradient at intervals of about 2.5 cm for 10 cm to yield a total of 0.25 g of soil.

2.3.3 DNA extraction, library preparation, and sequencing.

With the 0.25 g of soil from each of the 96 cores we extracted total genomic DNA using DNeasy PowerSoil kti (Qiagen; Germany) following manufacturer's instructions. Extracted DNA quality and yield were assessed using a NanoDrop One spectrophotometer (Thermo Scientific; USA) and a Qubit (Thermo Scientific; USA) respectively. DNA extractions were then randomly divided into eight sequencing runs. Libraries were prepared using Oxford Nanopore Technologies Ligation Kit (SQK-LSK109, ONT; UK) with Native barcodes (EXP-NBD104) in order to multiplex 12 samples per run. DNA was diluted to 400 ng as input and prepared following manufactures instructions. Multiplexed libraries were then run on a MinION Mk1b with R9.4.1 and R 9.5 flow cells (FLO-MIN106) (Table A6). The raw data was base called using Guppy v3.6.1 (ONT) specifying the high accuracy model. Base called runs were then de-multiplexed with the Guppy barcoder specifying parameters to discard sequences with middle adapters and to trim barcodes. We used Filtlong v0.2.0 (https://github.com/rrwick/Filtlong) to control for length (≥ 50 bp) and quality (Q) score (≥ 10). The summary of base calling specifications and sequencing statistics for each sample can be found in Table A6.

2.3.4 Annotating ARGs in Long Reads and Assigning Taxonomy.

A subset of quality controlled Nanopore reads for each soil core was selected using seqTK (https://github.com/lh3/seqtk) specifying 43,263 reads per sample based on the lowest

yield sequencing run. This subset of reads was then annotated using the Comprehensive Antibiotic Resistance Database (CARD) version 3.0.9 and its associated command line tool Resistance Gene Identifier (RGI) Main version 5.1.0 specifying input type contig (-t contig) with default parameters for alignment (-a BLAST). Initial RGI hits were further quality controlled by removing any ARO hits defined as mutations along with ARO hits with less than 25% coverage of the reference sequence to further ensure hits are ARG homologs rather than false positives or spurious hits. Long reads containing significant ARG hits were then run through Kraken 2 (v.2.0.9-beta; Wood et al. 2019) against a reference database built using all the standard Kraken 2 microbial databases with domains for bacteria, archaea, and viruses along with UniVec_Core library (Seitz et al. 2021).

2.3.5 Community Metrics.

To assess community composition, long reads from each core were assigned taxonomy using Kraken 2 (v.2.0.9-beta; Wood et al. 2019). To estimate family-level abundance, Kraken 2 reports were used as inputs for Bracken (Lu et al. 2017) specifying a read length of 500 bp and a threshold of 10 at the family level. Alpha diversity metrics were calculated for each core using Shannon–Wiener index (diversity; VEGAN package v2.5-6). Differences in bacterial community composition across FPES treatments were visualized using non-metric multidimensional scaling (NMDS) with Bray-Curtis dissimilarity (Johnson & Wichern, 1988, VEGAN package v2.5-6, Rstudio v 1.2.5019).

2.3.6 Statistical Analyses and Visualizations.

All statistical analyses were completed in R version 3.6.1 (R Development Core Team, 2020). An indicator species analysis (indicspecies R package, Cáceres and Legendre 2009, version 1.7.9) was used to identify indicators of disturbance from the (1) ARG hits and (2) ARG bacterial-host families. The reported values from this indicator analysis included an A indicator value which in this context is the mean abundance of an ARG in each FPES treatment divided by the sum of the mean ARG abundance values over all FPES treatments, a B indicator value which is the relative frequency of occurrence (presence–absence) of the ARG inside the target FPES treatment, and a p-value from a significance test with 999 permutations. Analysis of Similarity (ANOSIM; VEGAN package v2.5-6) was used to test significance of groupings by FPES treatment and soil core collection year. To determine if a shift in community by year or disturbance treatment significantly affected the abundance of ARGs, NMDS axes were used as a covariate in generalized linear models specifying Poisson distribution with ARG abundance

per soil core as the response. These models were then visualized against scatterplots of NMDS axes versus ARG abundance per core using ggplot2 (version 3.2.1; Wickham 2016). To determine if alpha diversity significantly affected the abundance of ARGs, alpha diversity was used in a generalized linear model specifying Poisson distribution with ARG abundance per soil core as the response. To test for significant differences in ARG abundance by FPES disturbance treatments and soil collection year without taking Bray-Curtis distances into account, we used non-parametric Wilcoxon signed rank test to compare ARGs per metagenome between disturbance treatments (UD, SD, MD) and soil collection year (2018, 2019) with Kruskal-Wallis one-way ANOVA to evaluate significance of FPES treatment on abundance of ARGs. This analysis was visualized using the R package ggplot2. Sankey visualization displaying relationship of ARG hits by FPES treatment, bacterial phyla, type of antibiotic associated with ARGs, and mechanism of resistance were generated using SankeyDiagram with flipPlots R package (version 1.3.2; https://github.com/Displayr/flipPlots/).

2.3.7 Soil Collection, DNA extraction, library preparation, and sequencing for Hi-C MAGs.

In June of 2020, we collected two 10 cm wide by 10 cm deep active layer soil cores from the far corners of each FPES treatment. Sterile technique and transportation as described in the metagenomics method section were employed. The interiors of the six cores were then sub-sample using sterile forceps along a depth gradient at intervals of about 2.5 cm for 10 cm to yield a total of 0.25 g of soil per sample for shotgun metagenomic libraries and 2g of soil per sample for Hi-C libraries. Shotgun libraries for each soil core were constructed by extracting total genomic DNA from subsampled soil using the DNeasy PowerSoil kit (Qiagen; Germany) followed by library prep using Nextera DNA Flex Library Prep (Illumina, San Diego, CA, USA) following manufacture protocols. We also used the total genomic DNA from these soil cores to prepare an SQK-LSK109 library (ONT; UK) that were subsequently sequenced using a MinION device with R9.4.1 flow cells, however, sequencing yields were too low (<10 Gb per sample) producing inadequate assembly depth for downstream analyses. Hi-C libraries were prepared by following ProxiMeta Hi-C preparation kit v3.0 (Phase Genomics, Seattle, WA, USA). Hi-C works by linking DNA-associated proteins located in proximity within a cell during DNA extraction, ligating the links during library prep, and then bioinformatically mapping ligated DNA libraries back to traditional shotgun assemblies. This method allows for the de-convolution of like tetranucleotide frequency or read depth (Stadler et al. 2019), therefore binning plasmid with host genomes despite a plasmid's potentially distinctive genomic characteristics retained from the donor bacteria. Nextera and Hi-C libraries were pooled such that 2/3 of the total DNA from

each soil core was from the Nextera library and 1/3 was the Hi-C library. GENEWIZ (South Plainfield, NJ, USA) sequenced the pooled libraries across six 2 x 150 bp HiSeq lanes.

2.3.8 Assembly and Binning via ProxiMeta Deconvolution for Hi-C MAGs.

Quality controlled metagenomic paired-end reads for each soil sample were assembled on the University of Alaska Fairbank's Research Computing System's high performance cluster using Megahit specifying default parameters (Li et al. 2015). The Megahit contigs were then filtered by size to select for contigs greater than 1500 bp in length. Size-selected contigs were used as input for alignment of Hi-C reads and binning following ProxiMeta workflow as previously described in Press et al. 2017. In brief, Hi-C reads were aligned to the Megahit contigs via Burrows-Wheeler alignment tool BWA-MEM (Li 2013). All Hi-C reads that were unmapped, not mapped uniquely, had a MAPQ score <20, or pairs mapping to the same contig were removed from the analysis since they are not useful for deconvolution of contigs. Contigs were then deconvoluted via ProxiMeta workflow by clustering based on Hi-C linkages using a proprietary Markov Chain Monte Carlo algorithm.

2.3.9 Phylogenetic analysis of Hi-C MAGs.

The quality of all metagenome-assembled genomes (MAGs) was assessed using CheckM which estimates phylogenetic placements along with genome quality metrics, such as completeness and contamination, based on single-copy core marker genes (Parks et al. 2015). A high-quality draft MAG was defined using the standards suggested by Bowers *et al.* (2017) as having >90% completeness and <5% contamination, medium quality was >50% completeness and <10% contamination, and a low quality draft was <50% completeness and <10% contamination. To place MAGs into a more meaningful phylogenetic context, we created a phylogenetic tree using the Hi-C MAGs and genomes from RefSoil+ (Dunivin et al. 2019), a database of over 922 RefSeq genomes from soil microbes. To do this, we predicted coding sequences in our MAGs and the RefSoil genomes using PRODIGAL44 v.2.6.3 (parameters: -m -p meta), and then searched the sequences for 16 ribosomal proteins including L2, L3, L4, L5, L6, L14, L16, L18, L22, L24, S3, S8, S10, S17 and S19. These ribosomal proteins have been used for comprehensive phylogenetic assessments in Bacteria, Archaea, and Eukaryota (Lavrinienko et al. 2020). The 16 phylogenetic markers from our MAGs and RefSoil genomes were found using HMMSEARCH with HMMER v.3.1b2 (parameters: -E 1e-5) using the Hidden Markov Models for the 16 ribosomal proteins downloaded from the Pfam database (https://pfam.xfam.org). Genomes that lacked half or more of the phylogenetic markers were not

used in the phylogenetic tree construction or downstream analyses of ARGs. Sequences found for each marker were aligned using MUSCLE v.3.8.31 (parameters: --maxiters 16), trimmed using TRIMAL v.1.2rev59 (parameters: -automated1), and then concatenated using the CONCAT script implemented by BinSanity. We constructed the phylogenetic tree for the resulting 194 Hi-C MAGs and 985 RefSoil genomes using FASTTREE v2.1.9 (parameters: -gamma -lg). The tree was then viewed and annotated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

2.3.10 Identifying ARGs and Mobile Genetic Elements in MAGs.

We annotated all ProxiMeta MAGs that were placed in the phylogenetic tree with CARD version 3.0.9 (Alcock et al. 2020) using command line tool RGI Main version 5.1.0 specifying input type contig (-t contig) with default parameters for BLAST alignment (-a BLAST) for strict and perfect hits only. RGI results were further quality controlled by removing any ARO hits defined as mutations along with ARO hits with less than 50% coverage of the reference sequence unless the hit was cutoff on the edge of a contig to further ensure hits are ARG homologs rather than spurious hits. To determine if a hit was located on a plasmid, we first screened all contigs containing CARD hits against the PLSDB database (version 2020_06_29; Galata et al. 2019) using Mash screen specifying max p-value of 0.01 (Valentina et al. 2018). Contigs initially screened as plasmids in mash were then verified using blastn version 2.11.0 so that hits with >70% nucleotide identity were classified as plasmids for downstream analyses. We used the command line tool Integron Finder, specifying default settings, to search sequence data for *attC* sites using covariance models, (Cury et al. 2016).

2.3.11 Community Composition from Nextera Library Illumina Reads.

To assess community composition the Illumina paired-end shotgun reads from the six soil cores were run through Kraken 2 (Wood et al. 2019) specify –paired for paired end reads against kraken2_drown_microbial. Kraken reports were then run through Bracken specifying a read length of 150 bp and a threshold of 10 at the family level (Lu et al. 2017). Visualization of the Bracken estimated abundances were then generated in R using the package pheatmap (v 1.0.12) make heatmaps and the package ggplot2 (version 3.2.1; Wickham 2016) to generate a relative abundance plot.

2.3.12 Statistical Analyses and Visualizations.

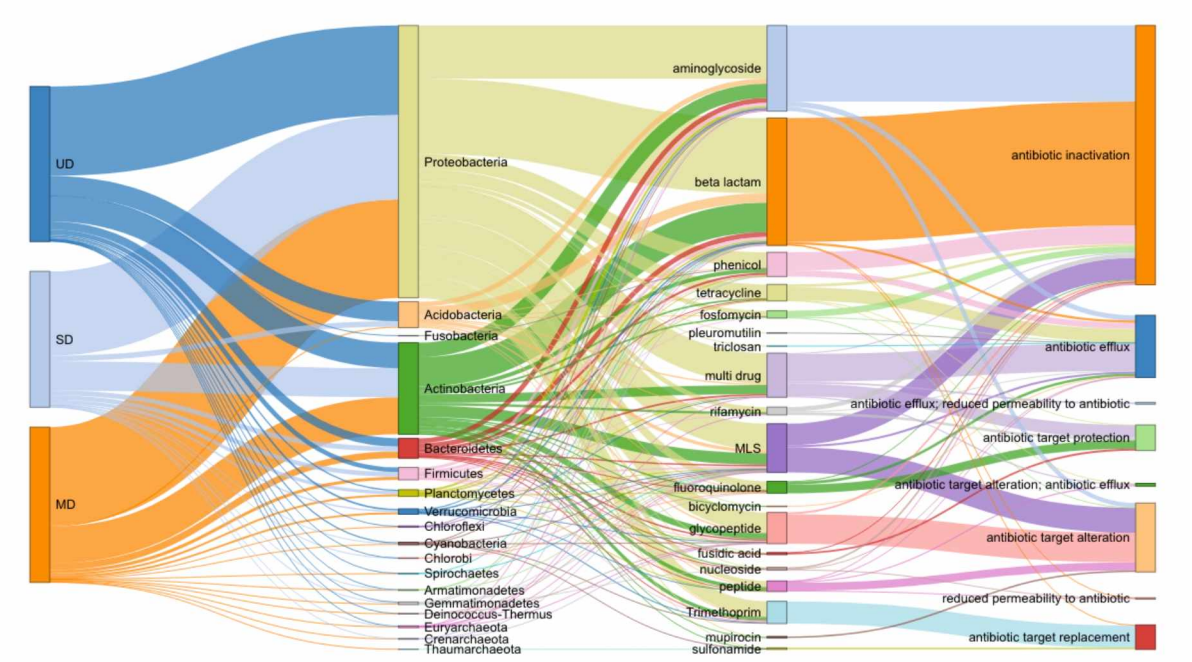
To test for significant differences in the number of ARGs per genome across phyla, we used the non-parametric Wilcoxon signed rank test to compare ARGs per MAG between the four most abundant bacterial phyla in the MAGs with Kruskal-Wallis one-way ANOVA to evaluate the effect of taxonomy. This analysis was visualized using the R package ggplot2 (version 3.2.1; Wickham 2016). To put the diversity of ARGs identified in MAGs into context with the larger FPES long read metagenomic dataset and the RefSoil+ global soil bacteria database, we compared how many of the ARG families identified in each dataset were unique or shared. A Venn diagram depicting this comparison was generated using meta-chart (https://www.meta-chart.com/venn).

2.4 Results

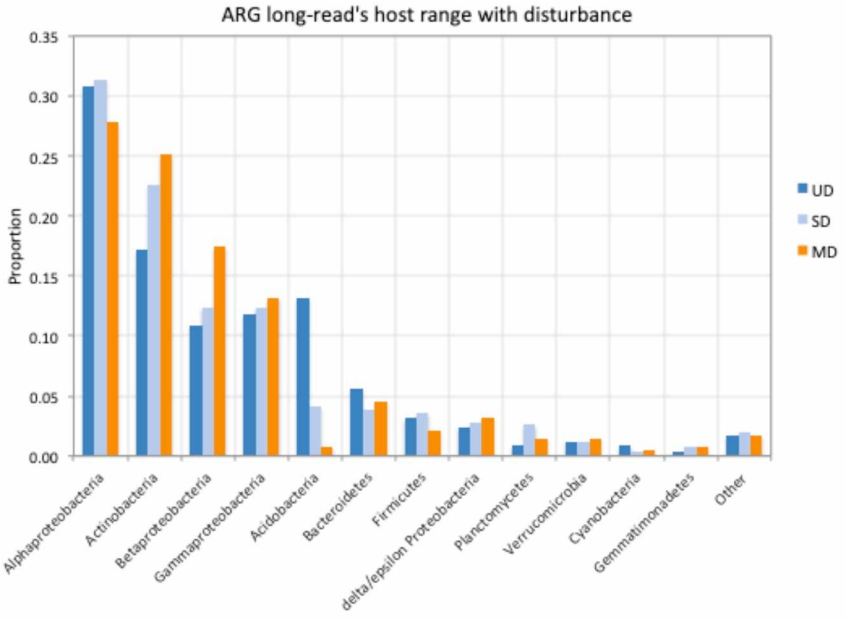
2.4.1 Community-level ARG profile and ARGs as predictors of soil disturbance.

Across the 96 soil samples, we detected 2339 significant gene hits for 984 unique ARGs encoding all mechanisms of resistance including antibiotic efflux, inactivation, target alteration, target protection, and reduction of permeability (Figure 2.1a). Across all treatments, the most abundant mechanism of resistance encoded by all ARGs was antibiotic inactivation (n= 1380) followed by antibiotic efflux and target alteration (n=339 and n=360 respectively). ARGs encoding β-lactamases were the predominant type of ARGs for inactivating enzymes with OXA β-lactamases (n= 203) as the most abundant β-lactamase gene family. ARGs encoding aminoglycoside inactivating enzymes were the second most abundant inactivation gene with AAC(6') (n=145) as the predominant gene family. The most abundant antibiotic efflux gene family was resistance-nodulation cell division (RND) antibiotic efflux pumps (n = 151), one of the most important determinants of multidrug resistance in Gram-negative bacteria.

We examined homology of ARGs detected at this locale to those present in the Comprehensive Antibiotic Resistance Database (CARD), which includes various ARGs from clinical pathogens. We found that 43 genes had full-length coverage and 100% sequence identity with 55 gene hits having high gene coverage (>90% length of reference sequence) and high identity (>90% identity). These high identity genes encoded a mix of antibiotic inactivating enzymes (n = 26), efflux pumps (n = 11), target replacement (n = 11), target alteration (n = 6), and target protection (n = 1) genes.



(**a**)



(**b**)

Figure 2.1. a) Sankey diagram of ARG hits in long reads with nodes representing disturbance treatment (UD= Undisturbed, SD= Semi-Disturbed, and MD= Most-Disturbed), the Kraken2 classified phyla, resistance to antibiotics classes, and mechanisms of resistance identified in the reads from across the 96 soil cores. b) Comparison of the proportions of ARG-carrying reads from a given bacterial phylum [except for Proteobacteria at a class level] by total reads per disturbance treatment.

Some of these ARGs were found in non-pathogenic soil saprophytes but were highly homologous to pathogenic soil microbes. For example, FosB, which encodes a fosfomycin inactivating enzyme, was identified on a single long read belonging to the genus *Massillia*, a major group of rhizosphere and root colonizing bacteria. However, this ARG hit was highly homologous to the FosB gene from *Bacillus anthracis*, the causal agent for anthrax that is commonly found in soil. OXA-229 cephalosporinase was another example of an ARG hit highly similar to a pathogen identified on a long read assigned to a non-pathogenic soil taxon. This long read from the UD treatment was identified as *Granulicella mallensis*, which is a dominant soil Acidobacteria that is abundant at low temperatures and nutrient limiting conditions. OXA-229 was found to be identical the homolog from *Acinetobacter spp.*, such as the pathogen *Acinetobacter baumannii* that is known to cause a broad range of nosocomial infections.

Using indicator species analysis (ISA), we identified which of the 984 ARG homologs were indicative of soil disturbance (Table 2.1). Between the MD and UD soils there was one shared indicator, mphN, encoding a macrolide phosphotransferase enzyme. Ten ARGs were identified as indicators for a single disturbance treatment (Table 2.1). In the MD treatment the indicator ARGs encoded aminoglycoside inactivating enzymes (n=3), beta-lactamases (n=1), efflux pumps (n=1), and target alteration for a glycopeptide resistance gene cluster (n=1). In the SD treatment there was only one indicator for a beta-lactamase gene, VMB-1, and in the UD there were three indicator genes encoding fosfomycin inactivating enzymes (n=1), efflux pumps (n=1), and target alteration (n=1).

2.4.2 Identifying predominant ARG host taxa, taxonomic indicators of disturbance, and effect of community changes on ARG abundance.

Using Kraken 2, 74.1% of long reads encoding at least one ARG were classified to a phylum level, 61.9% to a family level, and 56.5% to a genus level. ARGs on classified reads were found across diverse phylogenetic groups within 18 distinct bacterial phyla. 45.2% of ARGs were encoded on reads assigned to the phylum Proteobacteria (ARG n = 1058) making this phylum the predominant ARG host taxa across our soil samples followed by Actinobacteria (ARG n = 355), Acidobacteria (ARG n = 101), Bacteroidetes (ARG n= 77), and then Firmicutes (ARG n = 48; Figure 2.1a). We also identified major trends in ARG host phyla in the classified reads with increasing disturbance (Figure 1b). For example, there was decreasing proportion of Acidobacteria-associated resistance genes with disturbance (UD = 0.13, SD= 0.04, MD = 0.01) and an increasing proportion of Actinobacteria (UD = 0.17, SD= 0.23, MD = 0.25) and Betaproteobacteria-associated ARGs (UD = 0.11, SD= 0.12, MD = 0.17, Figure 2.1b).

Without taking between sample differences in bacterial community (i.e., beta-diversity), we did not find a significant difference in ARG abundance with disturbance treatment (Kruskal-Wallis P = 0.33; Figure A1) but did observe a temporal shift (Kruskal-Wallis P = 0.0082) with a significantly lower number of ARGs per metagenome in 2018 compared to 2019 (Figure 1b). When taking metrics of community composition into account, we observed statistically significant shifts both with disturbance (ANOSIM statistic R = 0.619, p-value = 0.001) and temporally (ANOSIM statistic R= 0.0944, p-value=0.001) (Figure 2.2a, Figure A7). Alpha diversity was found to be significantly higher in the disturbed soil treatments (MD and SD) compared to the undisturbed treatment (Figure 2.2b). There was no significant relationship between microbial alpha diversity and the abundance of ARGs (Figure A8) Therefore, to determine if the observed shift in community composition with disturbance and soil collection year had a significant effect on ARG abundance, we used the NMDS axes as covariates and found that disturbance-induced community shifts had a quadratic relationship with ARG abundance (ANOVA p = 0.017, Figure 2.2C) and temporal community shifts had a negative linear relationship with ARG abundance (ANOVA p = 4.3e-07, Figure 2.2D). The temporal relationship suggests that taxa from the 2018 soil communities are enriched in ARGs compared to the 2019 soils. The relationship with disturbance and ARG abundance suggests that both most disturbed and undisturbed soils select for community members enriched in ARGs compared to the semi-disturbed taxa.

At a family-level the top ten most dominant ARG hosts were *Bradyrhizobiaceae* (n= 189 ARGs), *Streptomycetaceae* (n= 100 ARGs), *Acidobacteriaceae* (n= 91 ARGs), *Burkholderiaceae* (n= 87 ARGs), *Pseudomonadaceae* (n= 72 ARGs), *Mycobacteriaceae* (n= 49 ARGs), *Rhizobiaceae* (n= 46 ARGs), *Comamonadaceae* (n= 36 ARGs), *Phyllobacteriaceae* (n= 35 ARGs), and *Sphingomonadaceae* (n= 30 ARGs). By conducting an indicator species analysis on all 161 ARG host families identified, we were able to discriminate which families could be driving the relationship between ARG abundance and disturbance (Table 2.1). *Myxococcaceae, Chitinophagaceae*, and *Streptosporangiaceae* were identified as significant ARG host indicators for the most-disturbed treatment (Table 2.1). *Comamonadaceae*, which was the 8th most common host family at FPES, was a strong shared indicator for disturbed soil treatments (SD and MD; p = 0.009) with ARGs on reads assigned to *Comamonadaceae* largely confined to disturbed cores (A= 0.86) with at least one *Comamonadaceae-associated* ARG in 37% of disturbed cores (B = 0.37). Whereas *Acidobacteriaceae,* which is the 3rd most common host family, was found to be a strong shared indicator for less disturbed cores (UD &SD; p =

0.001) with ARGs on reads assigned to this family mostly confined to less disturbed soils (A=0.9434) and found across 50% of UD and SD cores (B = 0.50).

Table 2.1. Indicator analysis of ARGs and ARG-host families from long reads across FPES treatment treatments with p-value indicating significance association between treatment and ARG. (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$)

Indicator Species Analysis of ARGs

MD # sps. 6	A	B	stat	p.value	
AAC(2')-IIb	0.81	0.13	0.32	0.03	*
AAC(3)-Xa	1.00	0.10	0.31	0.03	*
AAC(6')-Ia	1.00	0.10	0.31	0.04	*
LAP-1	1.00	0.10	0.31	0.04	*
tetU	1.00	0.10	0.31	0.03	*
vanKI	1.00	0.10	0.31	0.02	*
Group SD #sps. 1	A	B	stat	p.value	
VMB-1	1.00	0.13	0.35	0.03	*
Group UD #sps. 3	A	B	stat	p.value	
FosB	1.00	0.16	0.40	0.01	*
AbaF	1.00	0.13	0.35	0.04	*
vanRL	1.00	0.13	0.35	0.03	*
Group MD+UD #sps 1	A	B	stat	p.value	
mphN	1.00	0.14	0.38	0.05	*

Indicator Species Analysis of ARG Host Families

Group MD #sps. 2	A	B	stat	p.value	
Chitinophagaceae (B)	0.71	0.23	0.40	0.02	*
Myxococcaceae (P)	0.81	0.13	0.32	0.04	*
Streptosporangiaceae (Ac)	0.72	0.16	0.34	0.049	*
Group MD+SD #sps. 1	A	B	stat	p.value	
Comamonadaceae (P)	0.86	0.37	0.56	0.01	**
Group SD+UD #sps. 1	A	B	stat	p.value	
Acidobacteriaceae (A)	0.94	0.50	0.69	0.001	***

A= Acidobacteria, **Ac=** Actinobacteria, **B=** Bacteroidetes, **P=** Proteobacteria

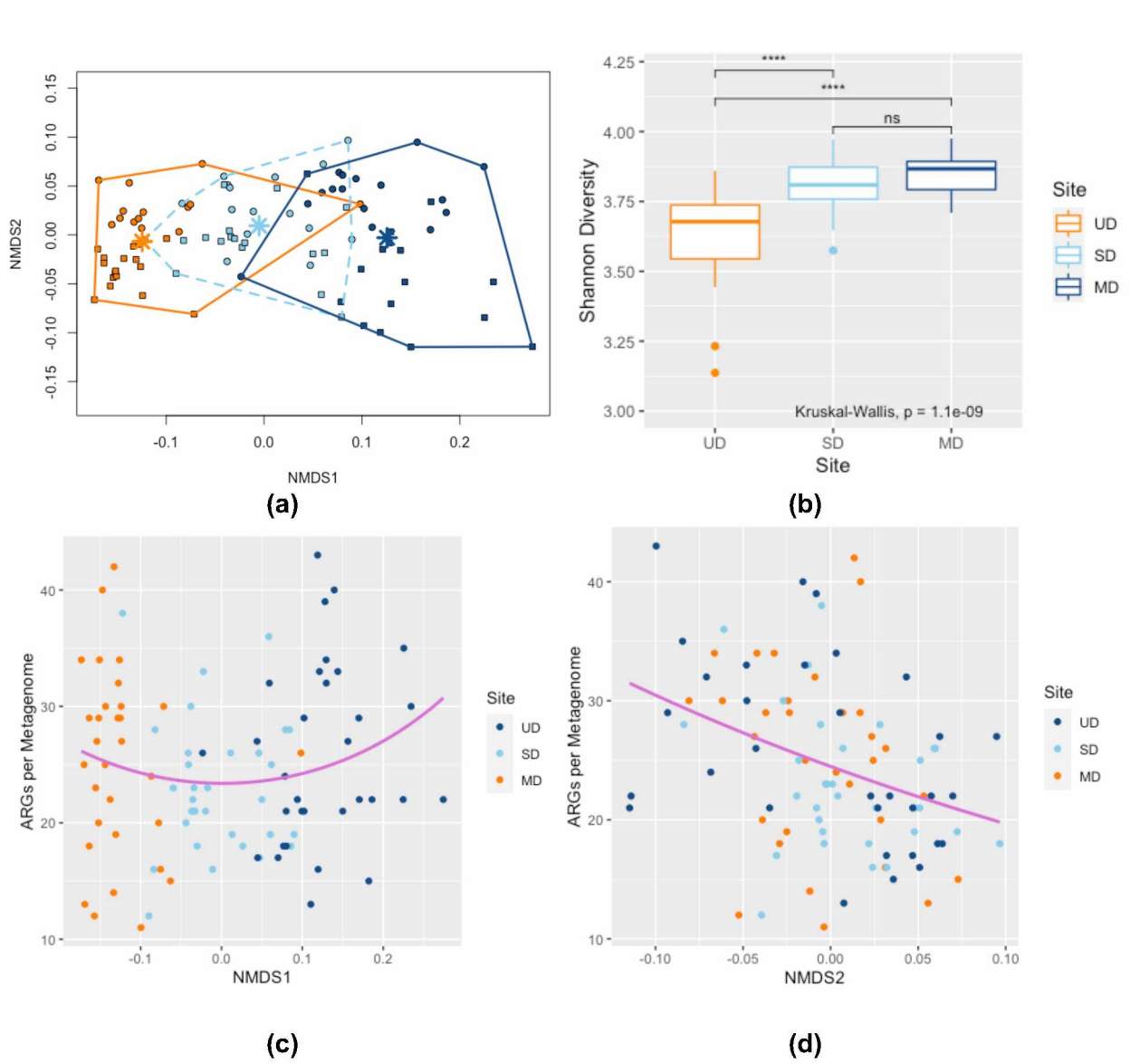


Figure 2.2. A) NMDS based on Bray-Curtis distance showing a shift in bacterial community composition between FPES treatments (UD=orange, SD=light blue, and MD=dark blue) and year (2018= square, 2019= circles) with convex hulls and group centroids (*) highlighting group differences**. B)** Shannon-Weiner diversity by FPES treatment. Kruskal-Wallis *P* value is shown in addition to Wilcoxon test results between means. Plots showing the effects of changes in bacterial community composition by treatment (NMDS1, **C**) and year (NMDS2, **D**) ARG abundance with pink GLM regression lines.

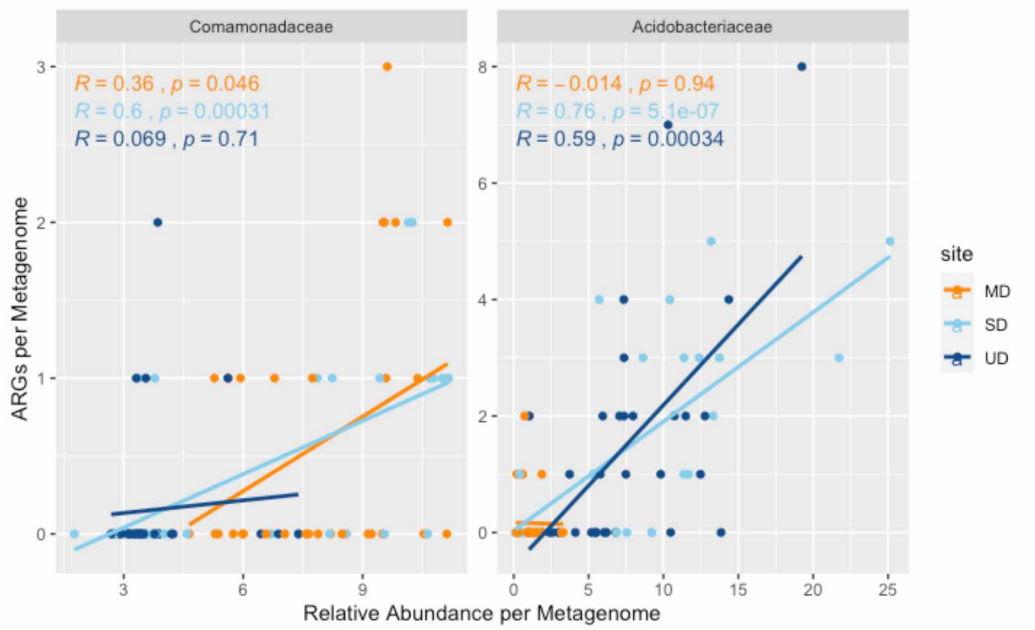


Figure 2.3. Scatterplot of Bracken estimated relative abundance for bacterial families *Comamonadaceae* (left) and *Acidobacteriaeceae* (right) with points representing a soil core color coded by FPES treatment versus the abundance of resistance genes from that family for each core. Lines show regression for each trend color coded by FPES treatment and top left shows Pearson R and p values associated with each trend color coded by treatment.

We conducted a correlation analysis to test the association of these indicator families with the detected ARGs on reads assigned to those families. We found a significant positive correlation between the relative abundance of *Comamonadaceae* in disturbed treatments and the number of *Comamonadaceae-associated* ARGs (Figure 2.3a). This same positive correlation was observed for *Acidobacteriaceae* in the less disturbed treatments (UD and SD) (Figure 2.3b). Moreover, trends between disturbance and the number of ARGs on reads assigned to these indicator families, *Acidobacteriaceae* (UD = 71, SD = 17, MD = 3 ARGs) and *Comamonadaceae* (UD = 7, SD = 5, MD = 24 ARGs), as well as the two most abundant host families, Bradyrhizobiaceae (UD = 79, SD = 65, MD = 45 ARGs) and Streptomycetaceae (UD = 29, SD = 33, MD = 38 ARGs), correspond with the relative abundance of these families with disturbance (Figure 2.4).

2.4.3 Quality of library reads, assemblies, and MAGs.

Across the six FPES soil cores (two per disturbance treatment), we generated an average of 1.65×10^8 paired-end reads per Hi-C library and 2.35×10^8 paired-end reads per Nextera library with mean Q-scores of 37.97 and 37.92 respectively (Table A8). The Nextera reads assembled via Megahit and size selected for contigs greater than 1,500 bp resulted in a

total of 3,053,190 contigs with an average length of 3,339 bp and the largest contig of 693,351 bp in length. Using ProxiMeta Deconvolution to generate MAGs from the size-selected contigs and Hi-C reads, we generated 921 genome clusters of which 59 were high quality (6.4%), 172 were medium quality (18%), and 690 were low quality drafts (74%) (Figure 2.5a). The number of MAGs per soil core was highly associated with Megahit total assembly length (R^2=0.75) rather than Hi-C library sequencing yield (R^2=0.18) (Figure A3). For the 921 Hi-C MAGs mean percent completeness was 33.5 % ± 34.3, mean percent contamination was 2.24 % ± 4.27, mean N_{50} per MAG was 29,972 bp ± 44,015, and mean novelty score was 73.9 ± 39.4 (Figure 2.5b).

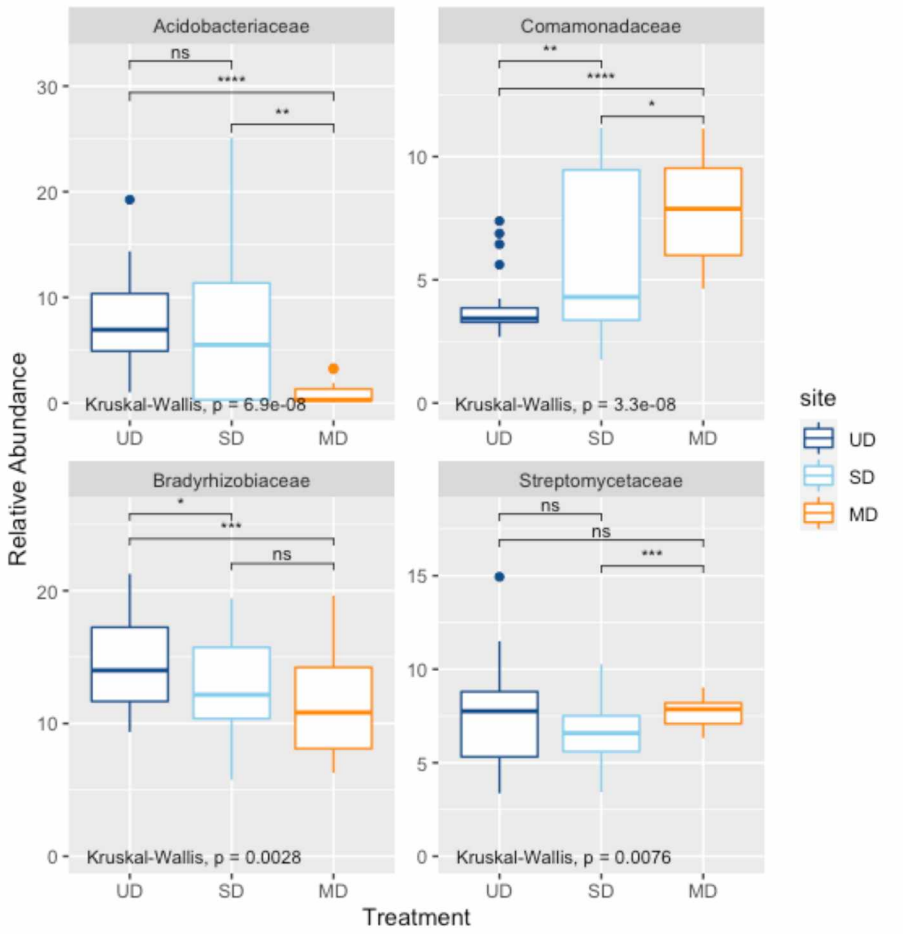


Figure 2.4. Boxplot depicting relative abundance by FPES treatments (UD= Undisturbed; SD= Semi-Disturbed, MD= Most-Disturbed) of two indicator ARG host families (top row) and top two most prevalent ARG host families (bottom row). Kruskal-Wallis *P* value is shown in addition to Wilcoxon test results between groups (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$).

2.4.4 MAGs phylogenetic placement.

To place MAGs in a more robust phylogenetic context of a tree with all RefSoil+ genomes for context, we included MAGs with at least half of the sixteen ribosomal marker proteins. This cutoff resulted in the placement of 194 MAGs from across the six soil cores (Figure 2.6) that were mainly high and medium quality (mean completeness 80.7 % ± 21.8; mean contamination 5.54 % ± 6.75; and mean N50 29,373 bp ± 35,379; Figure 2.5C). This tree corresponded with marker lineages assigned using MASH resulting in MAGs (n = 194) that were taxonomically inferred to be from 8 bacterial phyla including Acidobacteria (n = 72 MAGs), Proteobacteria (n = 69), Actinobacteria (n = 19), Bacteroidetes (n=19), Verrucomicrobia (n = 7), Tenericutes (n = 4), Gemmatimonadetes (n = 3), and Planctomycetes (n=1) (Figure A6).

Based on the Bracken estimation of abundance at a phylum level from both Nextera libraries and long read libraries, all phyla assigned to the MAGs were from the top ten most abundant bacterial phyla across treatments, except for Tenericutes which were the eighteenth most abundant bacterial phylum (Figure A4).

2.4.5 Antibiotic Resistance in Hi-C MAGs Across FPES Treatments and Phylogeny

We did not observe a significant effect of soil disturbance on the mean number of ARGs per MAG (Kruskal-Wallis P = 0.25; Figure A2) but rather a strong effect of bacterial phylum (Kruskal-Wallis P = 0.00054; Figure 2.7), especially in taxa that were affected by disturbance induced shifts. For example, Proteobacteria, which were enriched in terms of abundance in MD soils, and Acidobacteria, which were enriched in terms of abundance in UD soils, were found to have significantly more ARGs per MAG than Bacteroidetes (Wilcoxon P = 0.0031 & 0.0215) and Actinobacteria (Wilcoxon P = 0.0016 & 0.0064). The four Tenericutes MAGs and one Planctomycetes MAG had no ARG hits. Unlike the long read metagenomic data which looks at community level rather than individual genome level differences, the most abundant mechanism of resistance in MAGs was antibiotic efflux (n = 238), notably the multidrug efflux gene adeF which was found in MAGs from all phyla but Acidobacteria, Tenericutes and Planctomycetes. The next most abundant mechanism of resistance in MAGs were genes encoding antibiotic inactivation (n= 34), with β-lactamase genes being the most frequent type of inactivating enzyme. All 31 unique ARG families identified in the MAGs were also ARG families identified in the long read metagenomic data's 223 unique ARG families (Figure 2.9). When compared to the RefSoil+ database of cultured soil bacteria, there are 24 ARG families distinct to that of the long read metagenomic data and four found in FPES MAGs not found in the RefSoil+ genomes (Figure 2.9). In terms of types of genes across phylogenies, we found genes

encoding tetracycline specific efflux pumps, including *tet*(D),*tet*(E),*tet*(45), *tet*A(58) along with vancomycin inactivation, were confined to the genomes of Proteobacteria and Acidobacteria.

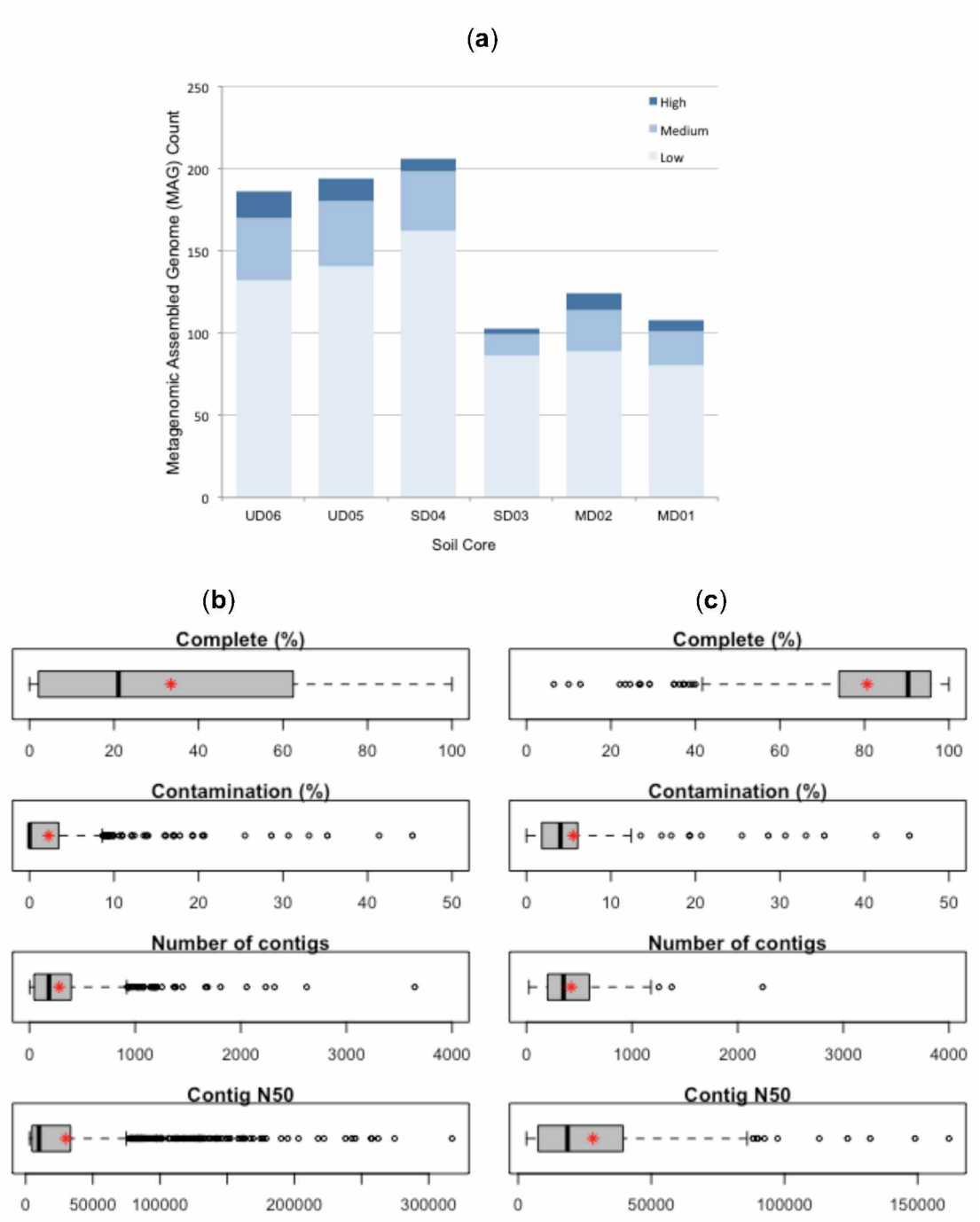


Figure 2.5. Bin quality by soil core based on CheckM metrics with high quality draft MAGs defined as >90% complete and <5% contamination, medium quality as > 50% complete and <10% contamination, and low-quality as <50% complete. **B & C)** Boxplots of genomic features including percent completeness, contamination, number of contigs and N50 for **(B)** all MAGs and for **(C)** MAGs used as inputs for the phylogenetic tree (red asterisk indicates mean).

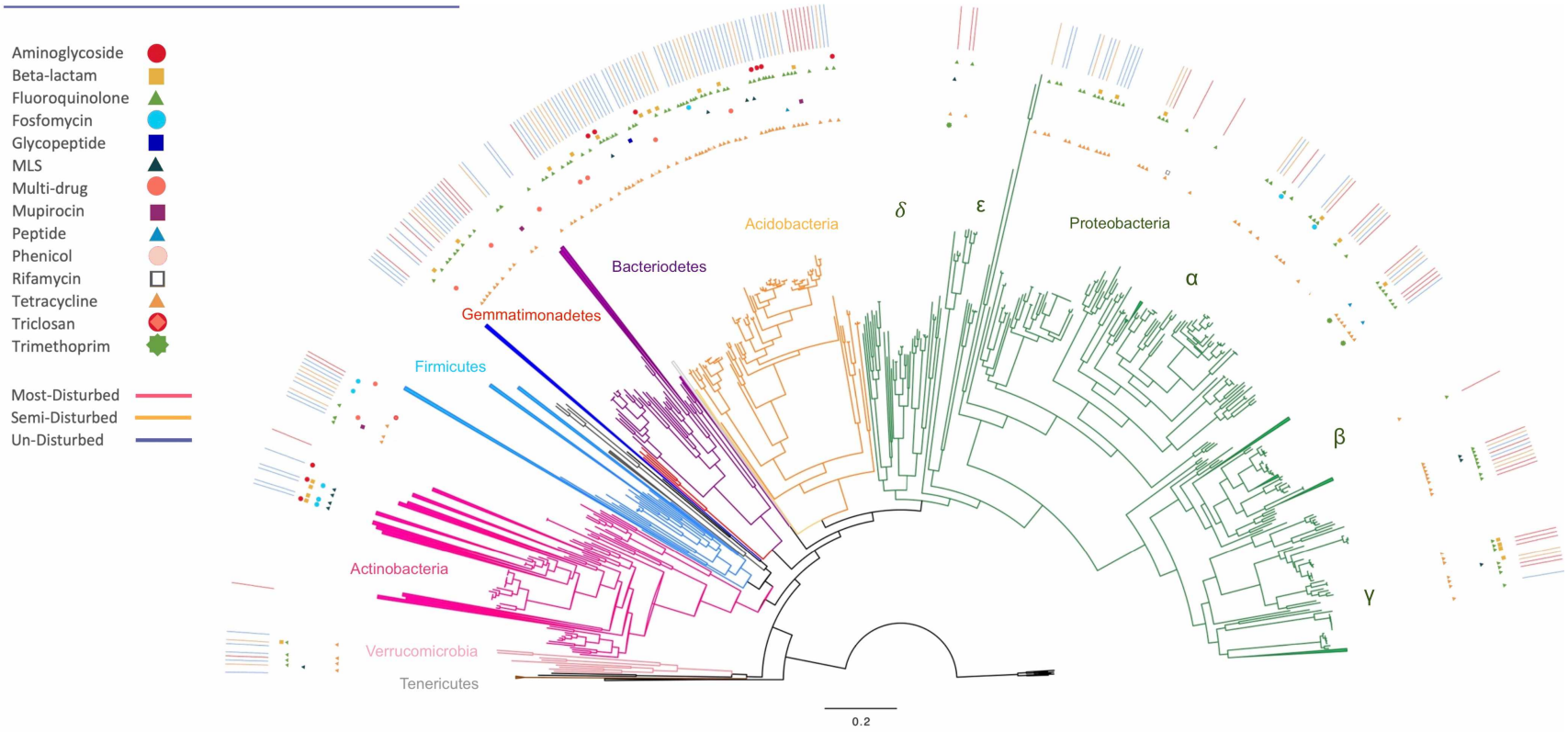


Figure 2.6. Phylogenetic tree of ProxiMeta metagenome assembled genomes (MAGs) from across FPES thaw treatments and reference genomes of Bacteria and Archaea from the RefSoil+ database. Presence of a Hi-C MAGs is highlighted by the colored lines of the outer ring with blue lines representing the undisturbed plot, orange representing the semi-disturbed treatment, and red representing the most-disturbed treatment. Inner ring shows antibiotic associated with ARGs in each MAG. Maximum likelihood tree constructed using concatenated alignments of 16 ribosomal proteins including only genomes with at least half of these proteins.

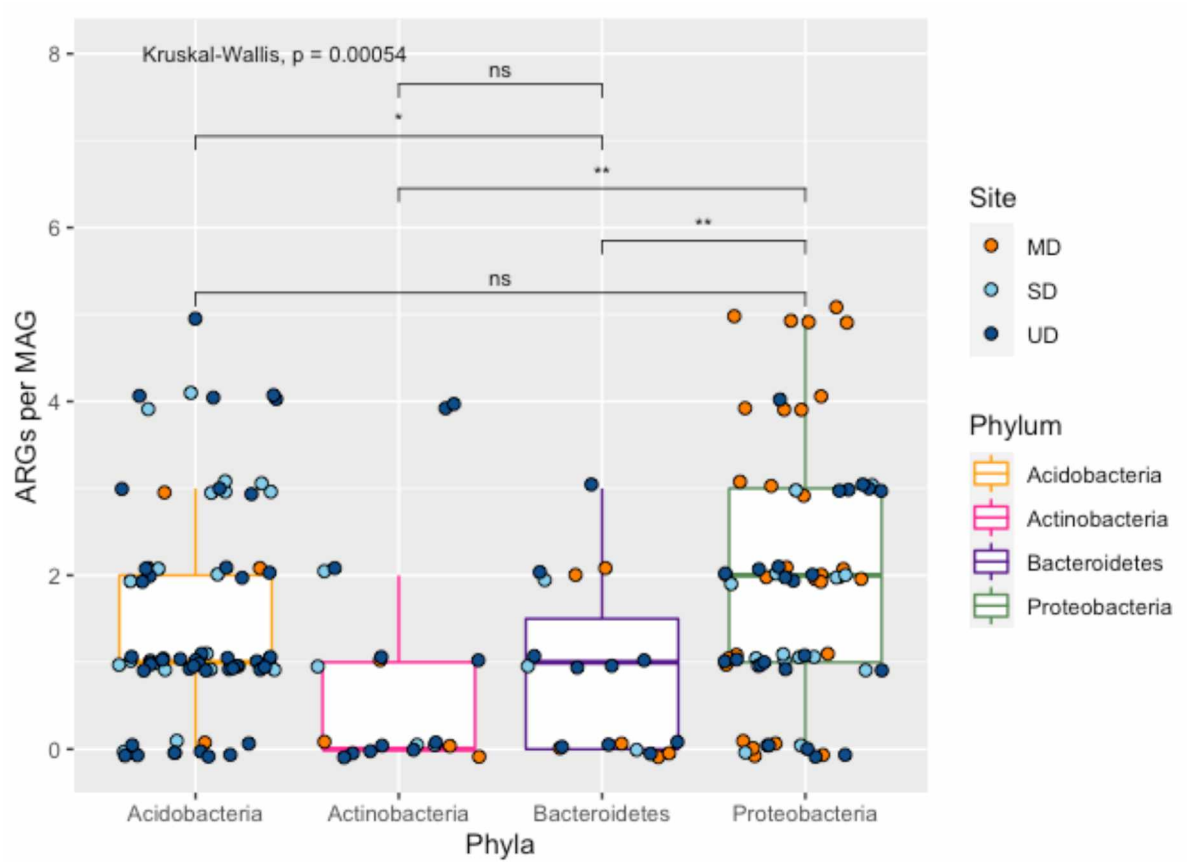


Figure 2.7. Boxplot of number of ARGs per Hi-C MAG from the phylogenetic tree shows significant differences in ARGs count between distinct phyla. Points are vertically jittered as to prevent overlap; points are colored by FPES treatment. Kruskal-Wallis *P* value is shown in addition to Wilcoxon test results between groups (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$).

The phyla associated with tetracycline specific efflux pumps genes in these MAGs were consistent with our long read dataset; however, vancomycin inactivation was identified in a greater diversity of phyla such as Actinobacteria, Verrucomicrobia, Firmicutes, and Bacteriodetes in the long read dataset. Aminoglycoside-specific inactivation genes from the gene families AAC(2'), AAC(3), APH(2"), APH(9) were confined to Acidobacteria and Actinobacteria in our MAGs but not in the long read dataset. The gene adeF, a membrane fusion protein of the multidrug efflux complex adeFGH, was found across all phyla except Actinobacteria. β-lactamases from the OXA family were found across phyla, which corresponds with the finding that OXA β-lactamases were the most abundant β-lactamase across FPES identified using long read metagenomics.

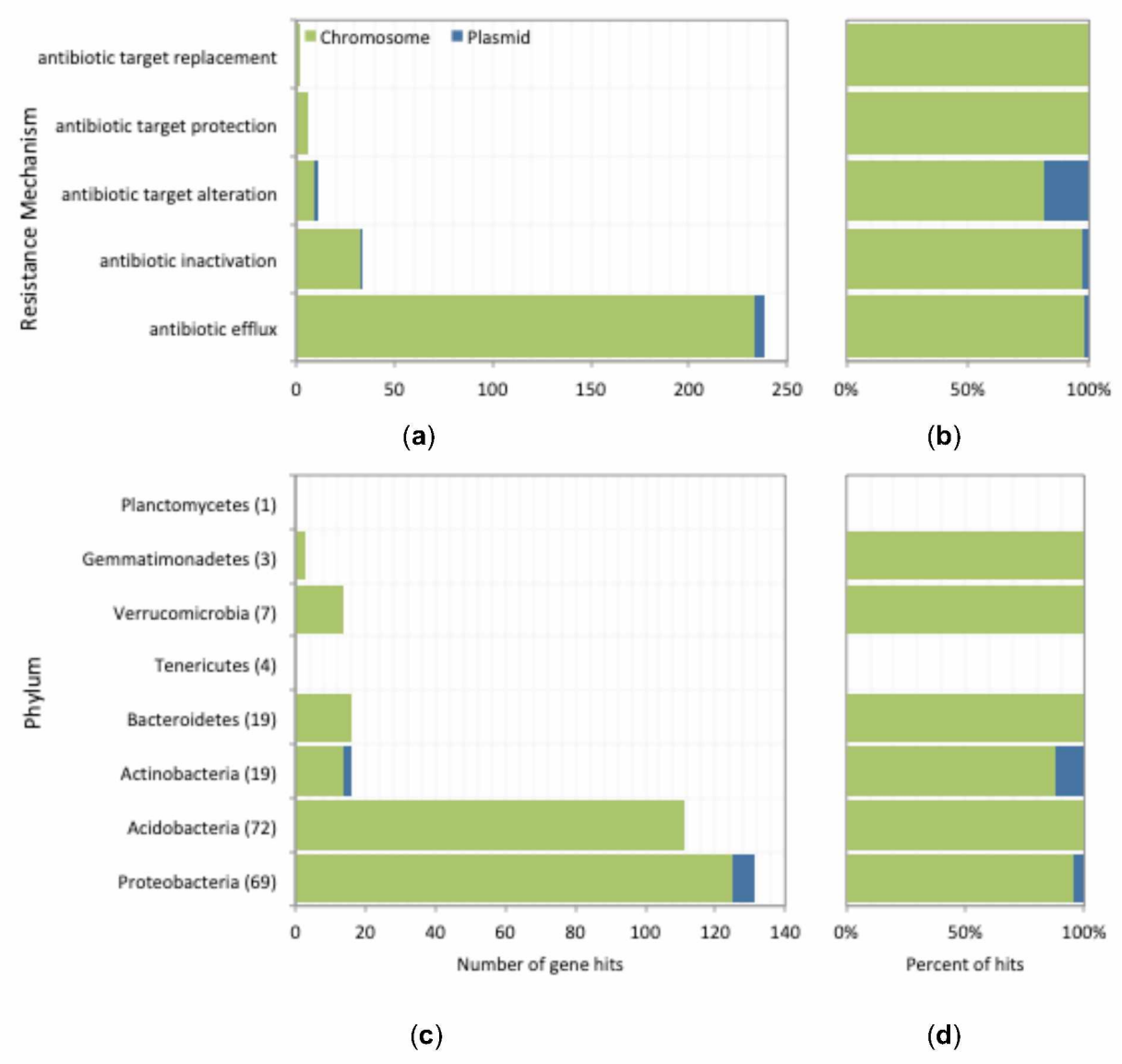


Figure 2.8. From MAGs used in the phylogenetic analysis, raw ARG count (A) and percent (B) of ARGs by resistance mechanism in MAGs on plasmids (blue) or chromosomes (green). The raw ARG count (C) and percent (D) of ARGs by phyla on plasmids (blue) or chromosomes (green).

Among the 194 MAGs placed in the phylogenetic tree (Figure 6), we identified eight plasmid-borne ARGs encoding antibiotic inactivation, efflux, and target alteration (Figure 2.8; Table 2.3). These plasmid-borne ARGs were confined to Proteobacteria and Actinobacteria MAGs and included homologs for an *adeF* efflux pumps, a *tet(50)* inactivating enzyme, and two *murA transferase* genes associated with resistance to fosfomycin. Three plasmids-borne ARGs were found in MAGs from the MD and SD treatments, and two were found in UD MAGs. The proportion of ARG hits encoded on plasmids versus chromosomes was higher in the FPES

Proteobacteria MAGs than in RefSoil+ Proteobacteria genomes (0.048 > 0.040) and for FPES Actinobacteria MAGs versus RefSoil Actinobacteria genomes (0.143 > 0.009). Using the Integron Finder tool, we did not detect integrons on any contigs in the Hi-C MAGs containing ARGs, however, we did detect at least one type 1 integron in 17 of the 194 MAGs. In contrast, in the long read metagenomic dataset, we identified five type 1 integrons on long reads containing ARGs, four reads with ARGs encoding antibiotic inactivating enzymes and one for target protection (Table 2.2).

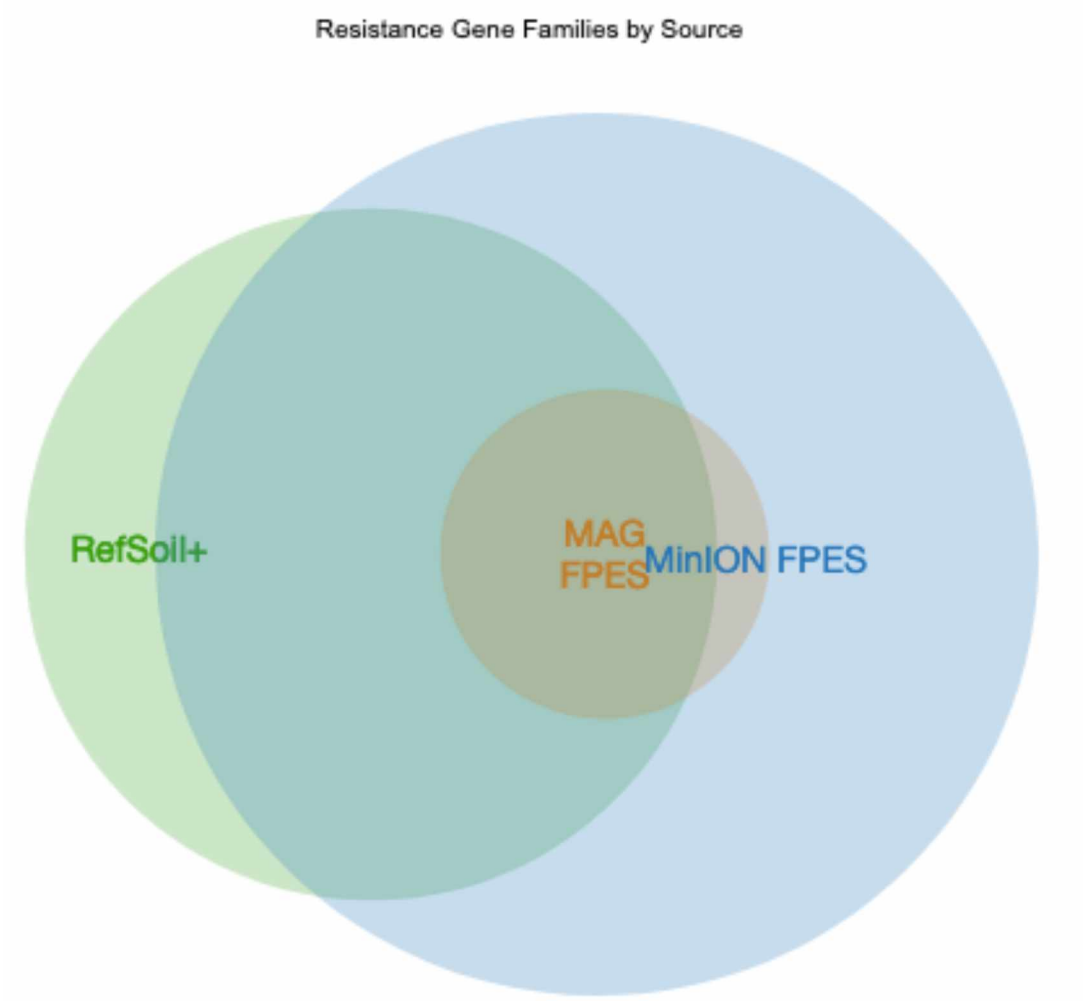


Figure 2.9. Venn diagram depicting resistance gene families in FPES MAGs, FPES MinION metagenomic samples, and RefSoil+ genomes.

Table 2.2. Description of ARGs identified on long reads containing integrons.

ARGs IDENTIFIED ON LONG READS WITH INTEGRONS								
FPES Sample	Year	Mechanism of Resistance	ARG	ARG Family	Antibiotic class	% Length of Reference	% Identity	Integron Type
MD_10.4	2019	inactivation	AAC(3)-IIc	AAC(3)	aminoglycoside	40.6	100	1
MD_11.1	2019	inactivation	DHA-15	DHA beta-lactamase	cephalosporin; cephamycin	40.1	100	1
MD_11.2	2018	inactivation	ERP-1	ERP beta-lactamase	penam	35.5	100	1
UD_10.1	2019	target protection	QnrD1	quinolone resistance protein	fluoroquinolone	32.7	100	1
UD_11.3	2019	inactivation	mphA	macrolide phosphotransferase	macrolide antibiotic	25.2	100	1

Table 2.3. Description of ARGs identified on plasmids in Hi-C MAGs.

ARGs IDENTIFIED ON PLASMIDS IN MAGs								
Hi-C MAG	Phylum	Family	Mechanism of Resistance	ARG	ARG Family	Antibiotic class	% Length of Reference	% Identity
MD02 bin_23	Proteobacteria	Pseudomonadaceae	efflux	adeF	RND antibiotic efflux pump	fluoroquinolone ; tetracycline	100.38	60.11
MD02 bin_4	Proteobacteria	Pseudomonadaceae	efflux	adeF	RND antibiotic efflux pump	fluoroquinolone ; tetracycline	145.14	72.12
MD02 bin_69	Proteobacteria	Rhodospirillaceae	efflux	adeF	RND antibiotic efflux pump	fluoroquinolone ; tetracycline	81.78	50.69
SD03 bin_13	Proteobacteria	Rhodospirillaceae	efflux	adeF	RND antibiotic efflux pump	fluoroquinolone ; tetracycline	96.22	52.91
SD04 bin_65	Proteobacteria	Rhodospirillaceae	efflux	adeF	RND antibiotic efflux pump	fluoroquinolone ; tetracycline	82.63	48.35
SD04 bin_65	Proteobacteria	Mycobacteriaceae	inactivation	tet(50)	tetracycline inactivation enzyme	tetracycline	19.33	100
UD05 bin_34	Actinobacteria	Mycobacteriaceae	target alteration	murA	antibiotic-resistant murA transferase	fosfomycin	98.56	94.9
UD06 bin_12	Actinobacteria	Mycobacteriaceae	target alteration	murA	antibiotic-resistant murA transferase	fosfomycin	98.56	94.9

2.5 Discussion

The overarching goal of this study was to determine how disturbance-induced shifts in active layer microbial communities affect the types, abundance, and mobility of antibiotic resistance genes (ARGs) that comprise the resistome in permafrost-associated soils. To answer this question, we collected culture-independent shotgun metagenomics and reconstructed MAGs to examine ARG composition, host range, and abundance's relationship with community composition from a permafrost thaw gradient in Fairbanks, Alaska. We found evidence that implicates antibiotic resistance as an intrinsic component of bacterial evolution in soils and unveils the role of disturbance in subarctic soils for shifting community composition that structures the predominant ARGs comprising the resistome. These efforts reveal permafrost-associated soils as a potential risk to One Health in the north by identifying soils typical of Interior Alaska as a dynamic reservoir of diverse and abundant ARGs that can shift with disturbance-induced community changes augmented by climate change.

Using long read metagenomics, we identified major trends in the predominant ARGs and host taxa in our subarctic soil samples (Figure 2.1a-b). The most prevalent mechanism of resistance detected from our long read sequences was antibiotic inactivation with 59% of the ARGs identified encoding an antibiotic inactivating enzyme of which 47.5% were beta-lactamases followed by aminoglycoside inactivating enzymes (29.8%). A previous study examining beta-lactamases in Alaskan soils found an incredibly diverse set of beta-lactamases that could confer high level resistance phenotypes in *E. coli* even without manipulating gene expression machinery (Allen et al. 2009). This study's findings that provide evidence soil-borne ARGs are functional, paired with the abundance and diversity of beta-lactamases identified in our Alaskan soil samples, demonstrates the potential for ARGs in subarctic soils to compromise human health by acting as a reservoir of genes that are effective against clinically significant antibiotics. Moreover, given minimal presence of people at our remote study site, antibiotic pollution from anthropogenic sources is likely insignificant. Rather the prevalence of genes encoding antibiotic inactivating enzymes in our soils implies the presence of selective pressures favoring these genes, such as pressure that would be imposed by the biosynthesis of antibiotic compounds by soil microbes.

Indeed, we identified Actinobacteria, a phylum of prolific antibiotic producers that account for 45% of all bioactive microbial metabolites discovered (Anandan et al. 2013), as the second most abundant taxa at our site (Figure A4). More specifically the bacterial family *Streptomycetaceae,* a prolific producer of clinically significant antibiotics, was the second most abundant taxa across disturbance treatments further highlighting competitive inhibition as a

potential selection force for favoring the maintenance of ARGs in subarctic soil bacteria (Figure 2.4). Previous work in Fairbanks, Alaska identified several cultured soil bacteria with antibiotic activity against ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) pathogen relatives (Haan et al. 2020). The ESKAPE pathogens are a major public health threat as nosocomial *pathogens* that exhibit multidrug resistance and virulence. Consequently, our findings of diverse ARGs paired with Haan et al. (2020) implicates subarctic soils as a potential source of antibiotic biosynthesis and segue into the discovery of novel antibiotics that can exhibit activity against major nosocomial pathogens. However, the diversity of ARGs unveiled from the subarctic soils presented here also suggests that if novel antibiotics were to be discovered, there may already be resistance determinants against those compounds circulating in soil-dwelling bacterial communities.

In our long read sequences, we detected a high diversity of significant ARG hits (Figure 2.1) with 32.3% of the CARD-based ARGs identified at our study site. This is comparable to results from soils presented in a large-scale metagenomic study analyzing antibiotic resistance from various environmental sources including fecal, ocean, lake, sediment, and soil samples (Nesme et al. 2014). Nesme et al. (2014) found that the environmental source with the highest diversity of ARGs was soils, which were found on average to have 28.8% ± 3.14% of the database's ARG sequences present. In terms of diversity of ARGs in human associated biomes, a study examining antimicrobial resistance across global urban microbiomes found that high ARG diversity varied across cities and taxonomic differences were not related to human population density. Interestingly, Fairbanks, which is a city with a population of 0.052 M people, had a high distribution of ARGs across samples with more ARGs per sample on average than New York City, a city with a population of 8.623 M people (Danko et al. 2021). With a low human population density and high abundance of ARGs across urban surfaces in Fairbanks, it is likely that non-human associated environmental bacteria, such as the taxa described from our Fairbanks soil samples, had some role in contributing to the high abundance of ARGs detected.

In terms of ARG host range, our long read metagenomics revealed a majority of the ARGs (45.2%) were encoded by the phylum Proteobacteria with Alphaproteobacteria as the most abundant ARG host at a class level. Although Proteobacteria had a relatively even abundance of ARGs across disturbance treatments, ARGs from several other bacterial taxa varied by disturbance. For example, in the disturbed treatment Actinobacteria were enriched as an ARG host versus the undisturbed treatment where Acidobacteria were enriched as a host (Figure 2.1b). By examining the association between the abundance of ARGs and relative

abundance of specific bacterial taxa at a family level, we revealed that enrichment of specific families could affect the aggregate abundance of ARGs in the community. For instance, the bacterial family *Comamonadaceae* was found to be a significant ARG host indicator for the disturbed treatments (MD & SD) and had a higher relative abundance in these soils. Consistent with this host family, *Acidobacteriaceae* was found to be an ARG host indicator for less disturbed soils (UD & SD) and had a significantly higher relative abundance in corresponding soil treatments (Figure 2.3-2.4). A similar association between enriched abundance of bacterial families and ARGs was observed for the *Bradyrhizobiaceae,* a common plant-associated bacteria and most abundant family at FPES that was significantly enriched in the UD soils encoding a higher proportion of ARGs detected at this treatment (Figure 2.4).

Interestingly the families *Comamonadaceae* and *Acidobacteriaceae*, which are indicator ARG host families in this study, were previously found to have significant associations with primary productivity in plants. Seitz et al. (2021) found that plants inoculated with SD and MD soils had an enrichment of *Comamonadaceae* that was associated with a significant decrease in plant growth whereas plants inoculated with UD soils with a significant enrichment of *Acidobacteriaceae* had an increase in plant growth. In other high-latitude soils, soil disturbance as a result of deglaciation resulted in community shifts that paralleled changes observed at FPES in which *Comamonadaceae* was a dominant clade in the young soils, decreased in intermediate age soil, and was absent from long established soils and where *Acidobacteriaceae* had an opposite trend (Nemergut et al. 2007, Seitz et al. 2021). In terms of bacterial diversity at FPES, the MD and SD treatments were found to have significantly higher alpha diversity than undisturbed soils (Figure 2.2b). This trend in microbial diversity corresponds to other studies that found a decrease in rhizosphere community diversity with vegetation successional development similar to that observed at our study site (Hu et al. 2020). Hu et al. (2020) suggest this observed shift in community assemblages with plant growth shows that unvegetated rhizosphere microbiomes harbor more pioneer assemblages of species with random resource overlap that shifts to climax communities at later stages with more antagonistic and stress tolerant bacteria (Hu et al. 2020). Efflux pumps, one microbial mechanism of resistance encoded by ARGs in our soils, have been described as a major tolerance mechanism in bacteria allowing the effective extrusion of organic solvent from the interior of the cell to the exterior environment, thus the presence of stress tolerant species in soils with mature vegetation at our site (UD) could influence the observed higher proportion of ARGs encoding efflux pumps (UD = 0.147, SD = 0.145, MD = 0.142). The corresponding shifts in soil microbial communities observed across studies exploring various forms of soil disturbance stresses how

an increase in the frequency of disturbance events with climate change can negatively affect global One Health in more ways than one.

Our results show how soil disturbance affects One Health by highlighting the relationship between disturbance-induced shifts in microbial composition and antibiotic resistance. We identified a significant decrease in ARG abundance with shifting communities from 2018 to 2019 (Figure 2.2d) and a quadratic relationship in ARG abundance with soil disturbance (Figure 2.2c). These relationships show a complex interplay between soil biogeochemical conditions over time, disturbance-induced differences in soil community, and ARG abundance. In this study, long read shotgun sequences were down sampled to the lowest sequencing yield and then annotated to estimate ARG abundance (Table A6). Thus, this measure of ARG abundance doesn't take into account absolute microbial abundance but rather represents the proportion of reads from each sample that encode an ARG (i.e., number of ARGs per 43,263 reads). Previous studies have defined ARG abundance in terms of copy number estimated through qPCR using primer libraries targeted to specific ARGs (Zhu et al. 2021). Unlike shotgun metagenomics that use databases to annotate the total genomic DNA from a sample, qPCR-based studies are limited in their power of detection based on the primers designed by the researcher. These primer libraries can lead to significant biases that miss divergent homologs and genes not targeted (Gaby & Buckley 2017). Moreover, a standard practice for normalizing gene copy numbers estimated with qPCR is by using estimates of absolute bacterial abundance with 16S rRNA qPCR. This metric of absolute bacterial abundance comes with inherent biases of its own. For example, there is significant intragenomic 16S rRNA gene copy number heterogeneity that can bias samples based on community composition (Sun et al. 2013). For example, at our study site Acidobacteria were enriched in SD and UD treatments while almost absent from MD whereas Proteobacteria were more abundant in MD soils (Figure A4b, Figure A7). Sun et al. (2013) found significantly higher 16S rRNA gene copies in Proteobacteria (3.94 ±2.62) compared to Acidobacteria (1.3 ±0.4 16S). Therefore, at our study site 16S rRNA qPCR would overestimate absolute bacterial abundance in disturbed soils enriched in Proteobacteria (Seitz et al. 2021). For these reasons, along with the fact that long read sequences may increase the likelihood of capturing full length ARGs and flanking MGEs (Huson et al. 2018), we used long read metagenomics to unearth the diversity, abundance, and types of ARGs in our subarctic soils.

Another fruitful method for exploring ARGs in soils is through the analysis of individual bacterial genomes. In previous work at this site, we cultured and sequenced the genomes of bacterial isolates and found Proteobacteria had significantly more ARGs per genome than

Firmicutes (Haan et al. 2021). Although this culture-based study revealed key functional differences in high-quality genome assemblies, such as presence of plasmid-borne ARGs, it was limited to two bacterial phyla as the result of the limitations imposed by culture-based methods. Soil cultures are estimated to capture less than 1% of total bacterial diversity thus requiring metagenomics to capture a higher proportion of the soil microbial community. Metagenomic-assembly and binning can be used without requiring cultures to reconstruct individual genomes of soil bacteria, however, genetic elements like plasmids are often not binned with the host genome since they can have distinct characters retained from the donor bacteria (e.g., read depth and tetra nucleotide frequency). To try and overcome these binning issues, we reconstructed individual genomes using Hi-C proximity ligation. Previous studies have demonstrated the ability of Hi-C proximity ligation with culture-independent de novo deconvolution to reveal plasmid-host associations in high quality bacterial genomes reconstructed from mixed samples without the need for prior information (e.g., reference genomes; Press et al. 2017, Stadler et al. 2019). Therefore, genomes generated with Hi-C may provide critical insight into the functional potential of individual genomes along with novel host-MGE associations from a greater diversity of bacteria than culture-independent genomics alone. However, the aforementioned studies examined fecal and wastewater samples which are several orders of magnitude less rich than soil microbial communities (Torsvik et al. 2002) potentially making it difficult to capture complete genomes without extremely deep high throughput sequencing.

In this study, we used Hi-C to generate bacterial genomes, theoretically containing plasmids, from six soil cores using Illumina ONT based assemblies as the scaffold for Hi-C libraries. We initially included long read data, but we did not have a deep enough sequencing depth (<10 Gb per core) to generate adequate assemblies, thus only Illumina sequence (64 - 85 Gb per core) assemblies were used in this study resulting in the deconvolution of 921 genome clusters using Hi-C proximity ligation and ProxiMeta deconvolution. The number of clusters generated strongly correlated with the sequencing yield of the Nextera library rather than the Hi-C library (Figure A3) emphasizing the importance of sequencing depth for the initial assembly step even when Hi-C is employed for deconvolution. Of the 921 genome clusters, a majority were low quality (74%) based on single copy core gene metrics of completeness and contamination (<50% complete and/or >10% contaminated) (Figure 2.5a-b). To select higher quality genomes for downstream analysis, we only used genomes that contained eight of sixteen ribosomal marker genes that resulted in an improvement of single copy core gene metrics (Figure 5c) and the generation of 194 MAGs. These genomes were used for

downstream analysis into the types, abundance, and mobility of ARGs in individual bacterial genomes (Figure 6). MAGs were generated from across 8 bacterial phyla. Some of these genomes belong to clades of bacteria that are difficult to culture under standard conditions but still dominant members of soil communities such as Acidobacteria (n = 72 genomes), Verrucomicrobia (n = 7), and Tenericutes (n = 4). Thus, genomes produced in this study not only provide insight into the types and abundance of ARGs, but also generate a wealth of genomic information that can be used to distinguish other functional traits from difficult to culture bacterial phylogenies. These functional traits within previously unidentified genomes may have implications for ecosystem resilience and health that can be uncovered with future research.

In terms of similarities across datasets in this study, all of the 31 types of ARG families identified in our Hi-C MAGs were also present in the long read metagenomic data (n = 223 ARG families) despite using different sequencing platforms (ONT versus Illumina) (Figure 2.9). Unlike our metagenomic dataset, which found genes encoding antibiotic inactivating enzymes to be the most abundant, ARGs encoding efflux pumps were the most abundant across Hi-C MAGs. This difference is reflective of the distinct nature behind metagenomic assembly and binning versus direct annotation of long reads using metagenomics. In MAGs you concatenate contigs that represent a singular microbial genome that is then annotated for ARGs. Therefore, we would expect ARGs in MAGs to be similar to what has been identified in whole genomes from culturable bacteria. Indeed, in bacterial isolates from our study site we found adeF, which is the resistance-nodulation-cell division efflux pump most abundant in MAGs, was also most abundant in the isolate's genomes (Haan et al. 2021). On the other hand, using metagenomics we would detect ARGs that are most dominant in the community as a whole that may not be the most common ARG in MAGs if a dominant community member only represents a single MAG.

If we compare both metagenomic and MAG datasets from this study to a global database of cultured soil bacterial genomes, RefSoil+ (Dunivin et al. 2019), we find that the greatest diversity of ARG families is from our long read metagenomic dataset. Moreover, there were fewer ARG families in our 194 MAGs compared to the 922 RefSoil+ genomes with 24 ARG families found in RefSoil+ genomes not found in either of our datasets highlighting how location can affect types of ARGs detected. The fact that we captured a greater diversity of ARG families than the RefSoil+ dataset that relied on culture based methods, emphasizes the importance of analyzing soil bacteria using direct sequencing methods in order to capture the difficult to culture majority. By analyzing total genomic DNA from soils in this study, we detected ARGs in soil bacteria that may pose a threat to human health if located on MGEs that may have not been captured with culture-based techniques.

In terms of MGEs, the majority of ARGs described throughout both datasets were chromosomally encoded suggesting ARGs are more commonly spread through vertical rather than lateral means. Yet we did detect eight plasmid-borne ARGs in the MAGs of bacteria from the phyla Proteobacteria and Actinobacteria highlighting a potential for HGT (Figure 2.8; Table 2.3). This result isn't surprising given previous research that analyzed the genomes of 23,425 bacteria and found mobile ARGs were mainly present in four phyla (i.e. *Proteobacteria, Firmicutes, Bacteroidetes*, and *Actinobacteria*) and were significantly enriched in *Proteobacteria* (Hu et al. 2016). In addition to plasmids, integrons are another MGE often embedded in promiscuous plasmids and transposons that have played a key role in the worldwide dissemination of ARGs. Especially type 1 integrons, which were detected in our samples and have been shown to facilitate lateral transfer of diverse ARGs across divergent bacterial taxa (Gillings et al. 2017). In our MAGs, no integrons were detected on contigs containing ARGs, however, we did detect type 1 integrons in MAGs and on long reads from metagenomic dataset (Table 2.2). This detection of ARG-integron associations on long reads and lack of associations detected on contigs in MAGs highlights the advantage of long read technologies for capturing full-length genes with flanking MGEs and importance of initial assembly quality for producing more contiguous contigs in MAGs (Huson et al. 2018).

Some of these ARGs identified in our soil samples from non-pathogenic soil saprophytes were highly homologous to pathogenic bacteria. For example, we found a full-length gene for an OXA-229 cephalosporinase on a long read belonging to the Acidobacteria species *Granulicella mallensis*, which is a dominant member of soils at low temperatures and nutrient limiting conditions. This gene was identical to that of an OXA-229 from *Acinetobacter baumannii* which is a virulent pathogen that causes a broad range of nosocomial infections. The high sequence identity between genes from these two disparate taxa suggests some ARGs between human pathogens and soil bacteria could be the result of conserved homologs or integrated into genomes from past horizontal gene transfer events thus suggesting evidence of lateral exchange. The fact subarctic soils have a widespread distribution of ARGs of which many are highly homologous to pathogenic bacteria, including MGE-associated genes, emphasizes the clinical importance of resistance in the soils described here.

2.6 Conclusions

The presence of highly diverse ARGs, including those encoded on MGEs, across bacterial taxa in our samples implicates active layer Alaskan soils as a reservoir of resistance with the potential to compromise human health. This study, paired with previous culture-based

susceptibility tests (Haan and Drown. 2021), affirms subarctic soils are a reservoir of bacteria resistant to clinically significant antibiotics and abundant ARGs that may arise as a clinical problem as human expansion into subarctic ecosystems continues. We detected ARGs in dominant non-pathogenic soil bacteria highly homologous to human pathogens. Moreover, the high abundance of genes encoding antibiotic inactivating enzymes suggests the presence of selective pressures for the maintenance of these ARGs. Given the remoteness of our study site, these selective pressures are likely the result of antibiotic producing soil microbes rather than anthropogenic antibiotic pollution. Indeed, *Streptomycetaceae*, a prolific producer of antibiotics, was the second most abundant bacterial family at FPES. The significant relationship between shifts in community composition and ARG abundance with disturbance unearths the role climate change, which increases the frequency and intensity of soil disturbance events, may have on ARG evolution in soils. The individual genomes provide a wealth of genomic information for future researchers to unveil the functional role of specific Alaskan soil bacterial phylogenies. For example, these genomes may help with the discovery of gene clusters that can reveal novel bioactive compounds for future antibiotic therapies or functional traits that affect ecosystem resilience. Overall, our findings emphasize how the enrichment of specific bacterial taxa with disturbance-induced thaw can shape the resistome and identified subarctic soils as a reservoir of ARGs potentially available to pathogens via horizontal gene transfer. Thus, the evidence presented here implicates antibiotic resistance as an intrinsic component of bacterial evolution in soils. These findings also highlight the need for novel therapeutics paired with continued antibiotic stewardship to conserve the arsenal of drugs we already have for treating bacterial infections.

2.7 Acknowledgments.

Thanks to Tom Douglas from the Cold Regions Research and Engineering Laboratory (CRREL) Alaska who provided access to the field site. Many thanks to Taylor Seitz, Jennie Humphrey, and Scout McDougal for 2018 soil core collection and Anne-Lise Duclezeau for MinION expertise. We acknowledge generous support from the Institute of Arctic Biology, Alaska INBRE, and the BLaST program. Thanks to the Institute of Arctic Biology Genomics Core lab where lab work and shared instruments were accessed. Research reported here was supported by BLaST through the National Institute of General Medical Sciences of the National Institutes of Health under awards UL1GM118991, TL4GM118992, and RL5GM118990. Research reported in this publication was supported by an Institutional Development Award (IDeA) from

the National Institute of General Medical Sciences of the National Institutes of Health under grant 2P20GM103395.

2.8 References.

Alcock, B.P., Raphenya, A.R., Lau T.T., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.L., Cheng, A.A., Liu S., Min, S.Y. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research*, *48*(D1), D517-D525. DOI: 10.1093/nar/gkz935

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *The ISME Journal*, *3*(2), 243–251. DOI: 10.1038/ismej.2008.86

Anandan, R., Dharumadurai, D., & Manogaran, G. P. (2016). An introduction to actinobacteria. In *Actinobacteria-Basics and Biotechnological Applications*. Intechopen. DOI: 10.5772/62329

Berg, J., Tom-Petersen, A., & Nybroe, O. (2005) Copper amendment of agricultural soil selects for bacterial antibiotic resistance in the field. *Letters in Applied Microbiology*, 40(2), 146-151. DOI: 10.1111/j.1472-765X.2004.01650.x.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A. and Tringe, S.G. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature biotechnology*, 35(8), 725-731. DOI: 10.1038/nbt.3893

Cáceres, M.D. and Legendre, P. (2009), Associations between species and groups of sites: indices and statistical inference. *Ecology*, 90: 3566-3574. DOI: 10.1890/08-1823.1

Center for Disease Control and Prevention (2019). Antibiotic Resistance Threats in the United States, 2019. *Atlanta, GA: U.S. Department of Health and Human Services*.

Cury, J., Jové, T., Touchon, M., Néron, B., & Rocha, E. P. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic acids research*, *44*(10), 4539–4550. DOI: 10.1093/nar/gkw319

Danko, D., Bezdan, D., Afshin, E. E., Ahsanuddin, S., Bhattacharya, C., Butler, D. J., ... & Bittner, L. (2021). A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell*. DOI: 10.1016/j.cell.2021.05.002

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*. DOI: 10.1038/nature10388

Djordjevic, S. P., Stokes, H. W., & Roy Chowdhury, P. (2013). Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Frontiers in microbiology*, *4*, 86. DOI: 10.3389/fmicb.2013.00086

Douglas, T.; Kanevskiy, M.; Romanovsky, V.; Shur, Y.; Yoshikawa, K. (2008). Permafrost Dynamics at the Fairbanks Permafrost Experimental Station Near Fairbanks, Alaska. *Institute of Northern Engineering, University of Alaska Fairbanks*, 1, 373-378. DOI: 10.1.1.419.2936

Dunivin, T. K., Choi, J., Howe, A., & Shade, A. (2019). RefSoil+: a Reference Database for Genes and Traits of Soil Plasmids. *mSystems*, 4(1). DOI: 10.1128/mSystems.00349-18

Finley, R. L., Collignon, P., Larsson, D. J., McEwen, S. A., Li, X. Z., Gaze, W. H., & Topp, E. (2013). The scourge of antibiotic resistance: the important role of the environment. *Clinical infectious diseases*, 57(5), 704-710. DOI: 10.1093/cid/cit355

Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., & Dantas, G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science*, *337*(6098), 1107 LP – 1111. DOI: 10.1126/science.1220761

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., & Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature*, 509(7502), 612-616. DOI: 10.1038/nature13377

Gaby, J. C., & Buckley, D. H. (2017). The use of degenerate primers in qPCR analysis of functional genes can cause dramatic quantification bias as revealed by investigation of nifH primer performance. *Microbial ecology*, *74*(3), 701-708. DOI: 10.1007/s00248-017-0968-0

Galata, V., Fehlmann, T., Backes, C., & Keller, A. (2019). PLSDB: a resource of complete bacterial plasmids. *Nucleic acids research*, *47*(D1), D195-D202. DOI: 10.1093/nar/gky1050

Gillings, M. R., Paulsen, I. T., & Tetu, S. G. (2017). Genomics and the evolution of antibiotic resistance. *Annals of the New York Academy of Sciences*, *1388*(1), 92-107. DOI: 10.1111/nyas.13268

Haan, T. J., & Drown, D. M. (2021). Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska. *Microorganisms*, 9(1), 116. DOI: 10.3390/microorganisms9010116

Haan, T., Seitz, T. J., Francisco, A., Glinter, K., Gloger, A., Kardash, A., Matsui, N., Reast, E., Rosander, K., Sonnek, C., Wellman, R., & Drown, D. M. (2020). Complete Genome Sequences of Seven Strains of *Pseudomonas* spp. Isolated from Boreal Forest Soil in Interior Alaska. *Microbiology resource announcements*, *9*(25), e00511-20. DOI: 10.1128/MRA.00511-20

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology*, *8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Holland, M. M., & Bitz, C. M. (2003). Polar amplification of climate change in coupled models. *Climate Dynamics*, *21*(3-4), 221-232. DOI 10.1007/s00382-003-0332-6

Hu, J., Wei, Z., Kowalchuk, G. A., Xu, Y., Shen, Q., & Jousset, A. (2020). Rhizosphere microbiome functional diversity and pathogen invasion resistance build up during plant development. *Environmental Microbiology*, *22*(12), 5005-5018. DOI: 10.1111/1462-2920.15097

Hu, Y., Yang, X., Li, J., Lv, N., Liu, F., Wu, J., ... & Zhu, B. (2016). The bacterial mobile resistome transfer network connecting the animal and human microbiomes. *Applied and environmental microbiology*, *82*(22), 6672-6681. DOI: 10.1128/AEM.01802-16

Huson, D. H., Albrecht, B., Bağcı, C., Bessarab, I., Gorska, A., Jolic, D., & Williams, R. B. (2018). MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biology direct*, *13*(1), 6.

Johnstone, J. F., Hollingsworth, T. N., & Chapin, F. S. (2008). A key for predicting postfire successional trajectories in black spruce stands of interior Alaska. Gen. Tech. Rep. PNW-GTR-767. Portland, OR: US Department of Agriculture, Forest Service, *Pacific Northwest Research Station*. 37 p., 767. DOI: 10.2737/PNW-GTR-767

Kramshøj, M., Albers, C. N., Holst, T., Holzinger, R., Elberling, B., & Rinnan, R. (2018). Biogenic volatile release from permafrost thaw is determined by the soil microbial sink. *Nature communications*, 9(1), 1-9. DOI: 10.1038/s41467-018-05824-y

Lavrinienko, A., Tukalenko, E., Mousseau, T.A., Thompson, L.R., Knight, R., Mappes, T., Watts, P.C. (2020). Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. *Sci Data* 7(312). DOI: 10.1038/s41597-020-00656-2

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*.

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674-1676.

Liu, X., Xiao, P., Guo, Y., Liu, L., & Yang, J. (2019). The impacts of different high-throughput profiling approaches on the understanding of bacterial antibiotic resistance genes in a freshwater reservoir. *Science of The Total Environment*, *693*, 133585.

Lu J, Breitwieser FP, Thielen P, Salzberg SL. (2017) Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* 3:e104, DOI: 10.7717/peerj-cs.104

McGuire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., Chen, G., ... & Zhuang, Q. (2018). Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change. *Proceedings of the National Academy of Sciences*, *115*(15), 3882-3887. DOI: 10.1073/pnas.1719903115

Nemergut, D. R., Anderson, S. P., Cleveland, C. C., Martin, A. P., Miller, A. E., Seimon, A., & Schmidt, S. K. (2007). Microbial community succession in an unvegetated, recently deglaciated soil. *Microbial ecology*, *53*(1), 110-122. DOI: 10.1007/s00248-006-9144-7

Nesme, J., Cécillon, S., Delmont, T. O., Monier, J. M., Vogel, T. M., & Simonet, P. (2014). Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current biology*, *24*(10), 1096-1100. DOI: 10.1016/j.cub.2014.03.036

O'Neill, J. (2016). Book review: Tackling drug-resistant infections globally. *Archives of Pharmacy Practice*, *7*(3), 110. DOI: 10.4103/2045-080x.186181

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, *25*(7), 1043-1055. DOI: 10.1101/gr.186072.114

Pérez-Valdespino, A., Pircher, R., Pérez-Domínguez, C. Y., & Mendoza-Sanchez, I. (2021). Impact of flooding on urban soils: Changes in antibiotic resistance and bacterial community after Hurricane Harvey. *Science of the Total Environment*, *766*, 142643. DOI: 10.1016/j.scitotenv.2020.142643

Press, M. O., Wiser, A. H., Kronenberg, Z. N., Langford, K. W., Shakya, M., Lo, C. C., ... & Liachko, I. (2017). Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *biorxiv*, 198713. DOI: 10.1101/198713

Ramos, J. L., Gallegos, M. T., Marqués, S., Ramos-González, M. I., Espinosa-Urgel, M., & Segura, A. (2001). Responses of Gram-negative bacteria to certain environmental stressors. *Current opinion in microbiology*, *4*(2), 166-171. DOI: 10.1016/s1369-5274(00)00183-1

Raphael, E., & Riley, L. W. (2017). Infections Caused by Antimicrobial Drug-Resistant Saprophytic Gram-Negative Bacteria in the Environment. *Frontiers in medicine*, *4*, 183. DOI: 10.3389/fmed.2017.00183

Schütte, U. M., Henning, J. A., Ye, Y., Bowling, A., Ford, J., Genet, H., ... & Bever, J. D. (2019). Effect of permafrost thaw on plant and soil fungal community in a boreal forest: Does fungal community change mediate plant productivity response?. *Journal of Ecology*, 107(4), 1737-1752. DOI: 10.1111/1365-2745.13139

Schuur, E. A., & Abbott, B. (2011). High risk of permafrost thaw. *Nature*, 480(7375), 32-33. DOI: 10.1038/480032a

Schuur, E. A., & Mack, M. C. (2018). Ecological response to permafrost thaw and consequences for local and global ecosystem services. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 279-301. DOI: 10.1146/annurev-ecolsys-121415-032349

Seitz, T. J., Schütte, U. M., & Drown, D. M. (2021). Soil Disturbance Affects Plant Productivity via Soil Microbial Community Shifts. *Frontiers in microbiology*, *12*, 76. DOI: 10.3389/fmicb.2021.619711

Stalder, T., Barraud, O., Casellas, M., Dagot, C., & Ploy, M. C. (2012). Integron involvement in environmental spread of antibiotic resistance. *Frontiers in microbiology*, *3*, 119. DOI: 10.3389/fmicb.2012.00119

Sun, S., Lu, C., Liu, J., Williams, M. A., Yang, Z., Gao, Y., & Hu, X. (2020). Antibiotic resistance gene abundance and bacterial community structure in soils altered by Ammonium and Nitrate Concentrations. *Soil Biology and Biochemistry*, *149*, 107965. DOI: 10.1016/j.soilbio.2020.107965

Sun, D. L., Jiang, X., Wu, Q. L., & Zhou, N. Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Applied and environmental microbiology*, *79*(19), 5962-5969. DOI: 10.1128/AEM.01282-13

Surette, M. D., & Wright, G. D. (2017). Lessons from the environmental antibiotic resistome. *Annual review of microbiology*, *71*, 309-329. DOI: 10.1146/annurev-micro-090816-093420

Torsvik V, Øvreås L, Thingstad TF. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science*. 2002 May 10;296(5570):1064-6. DOI: 10.1126/science.1071698

Udikovic-Kolic, N., Wichmann, F., Broderick, N. A., & Handelsman, J. (2014). Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proceedings of the National Academy of Sciences*, *111*(42), 15202-15207. DOI: 10.1073/pnas.1409836111

Velkov, V. V. (1999). How environmental factors regulate mutagenesis and gene transfer in microorganisms. *Journal of Biosciences*, *24*(4), 529-559.

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics are produced by the genus Streptomyces? *Archives of microbiology*, *176*(5), 386-390. DOI: 10.1007/s002030100345.

Whitman, T., Whitman, E., Woolet, J., Flannigan, M. D., Thompson, D. K., & Parisien, M. A. (2019). Soil bacterial and fungal response to wildfires in the Canadian boreal forest across a burn severity gradient. *Soil Biology and Biochemistry*, 138, 107571. DOI: 10.1016/j.soilbio.2019.107571

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer.* ISBN: 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome biology*, 20(1), 1-13.DOI: 10.1186/s13059-019-1891-0

Wright, G. D., & Poinar, H. (2012). Antibiotic resistance is ancient: implications for drug discovery. *Trends in microbiology*, *20*(4), 157-159. DOI: 10.1016/j.tim.2012.01.002

Xie, X., Fu, J., Wang, H., & Liu, J. (2010). Heavy metal resistance by two bacteria strains isolated from a copper mine tailing in China. *African Journal of Biotechnology*, *9*(26), 4056-4066.

Zhu, G., Wang, X., Yang, T., Su, J., Qin, Y., Wang, S., ... & Zhu, Y. G. (2021). Air pollution could drive global dissemination of antibiotic resistance genes. *The ISME Journal*, *15*(1), 270-281. DOI: 10.1038/s41396-020-00780-2

OVERALL CONCLUSION

As antibiotic resistance continues to emerge and rapidly spread in clinical settings, it is imperative to generate studies that build insight into the ecology and evolution of antibiotic resistance genes (ARGs) in environmental reservoirs. Soils are one of the richest microbial habitats both in terms of microbial abundance and diversity (Torsvik et al. 2002). Soils from biomes across the world have been found to host an abundance of antibiotic producing fungi and bacteria, namely *Streptomyces*, a soil-dwelling bacterial genus that produces bioactive compounds in which ⅔ of clinical antibiotics are derived from (Watve et al. 2001). Thus, the higher prevalence of ARGs detected in soils compared to other non-soil environments has frequently been attributed to the hypothesized co-evolutionary 'arms-shield race' between antibiotic producers and co-evolved resistant bacterial taxa (Nesme & Simonet 2015). Although hypothetically probable, it has been difficult to disentangle the exact selective forces driving the evolution and dissemination of soil-borne ARGs due to soil's biological complexity and technical challenges in measuring antibiotic concentrations. Despite this limitation metagenomic and culture-based studies have revealed the worldwide distribution of antibiotic resistant soil bacteria that encode a diverse range of genes recognized to confer resistance (Nesme et al. 2014, Dunivin et al. 2019). The apparent risk posed by soils makes understanding the factors influencing ARG composition and abundance critical.

Large-scale metagenomic studies have previously unveiled bacterial community composition as a primary determinant of soil ARG content (Forsberg et al. 2014). Land use practices and soil disturbance events have also been found to effect ARG abundance in soils, such as amendment of copper in agricultural soils (Berg et al. 2005) and flooding (Pérez-Valdespino et al. 2021). However, there is very little research on how climate change-induced disturbances (e.g., wildfires, thermokarst formation, etc.) will affect the microbial communities that structure the types and abundance of ARGs in soils. Climate change is increasingly affecting soil biogeochemical properties (Zepp et al. 2007), especially in Alaska where polar amplification has already led to an increase in the frequency and severity of soil disturbance events (Wendler & Shulski 2009, Holland & Bitz 2003). In Alaska, soil disturbance events, and anthropogenic land use practices, remove vegetation in return amplifying permafrost thaw and shifting microbial community composition in active layer soils (Seitz et al. 2021). The overarching goal of this thesis was to assess how disturbance to permafrost-associated soils in Alaska, and the subsequent shift in bacterial community composition, affects the types, abundance, and mobility of ARGs that comprise the active layer resistome.

In Chapter 1 I presented a culture-based analysis of 90 bacterial isolates from our study site, the Fairbanks Permafrost Experiment Station (FPES), and found that 90% of the isolates were resistant to at least one of the antibiotics tested with over 40% of the isolates displaying multidrug resistance. Using whole genome analysis, I identified a diverse array of resistance determinants from all major mechanisms of antibiotic resistance including antibiotic efflux, inactivation, target protection, and target alteration. Although direct functional conclusions from our genomic dataset cannot be drawn, the high abundance of beta-lactam resistance genes directly corresponds with the high proportion of phenotypic resistance to the beta-lactam antibiotic screened, ampicillin. This emphasizes the functionality of Alaskan-borne ARGs and their potential to compromise health if disseminated via horizontal gene transfer. Moreover, our high-proportion of isolates resistant to ampicillin corroborated previous findings in which beta-lactamases from Alaskan soils conferred high level resistance in *E. coli* vectors even without manipulating gene expression machinery (Allen et al. 2009).

Some of the isolates presented in chapter 1 do in fact belong to taxa of known opportunistic human pathogens, such as *Pantonea agglomerans* (Cruz et al. 2007), *Bacillus cereus (*Kotiranta et al. 2000), and several other Pseudomonas species. Several of these potential opportunistic pathogens were found to carry both chromosomally encoded and plasmid-borne ARGs. However, even non-pathogenic soil bacteria regularly interact with waterways, air, and built habitats such as hospital surfaces generating a potential for HGT from one biosphere to another (Woolhouse et al. 2015). When exposed to antibiotics, nonpathogenic commensal bacteria that acquired an ARG from environmental sources can be selected for promoting clonal expansion and increased risk for ARG spread to pathogenic bacteria (Lax et al. 2015). These microbial interactions across biospheres may become more frequent in Alaska with climate change. For example, as climate change increases the frequency of wildfires, microbes, including pathogenic fungi and bacteria, can more frequently be aerosolized and transported by wildland fire smoke to downstream human populations (Kobziar et al. 2018). With thawing permafrost there are changes to hydrologic conditions, including alterations in soil moisture, connectivity of inland waters, streamflow seasonality (Walvoord & Kurylyk 2016). These changes may mobilize active layer microbes, and microbes previously frozen in permafrost sediment, into waterways that humans and animals interact with. Lastly, Alaska's population has seen a 3.3% growth over the past decade (U.S. Census Bureau 2020). This expansion into Alaska may further disturb permafrost soils augmenting thaw and intensifying the potential for the spread of invasive species that carry vector-borne diseases. For example,

migration into Alaska can stimulate the chance for invasion of insects and arachnids that can colonize forests bringing vector-borne infectious disease into Alaska (El-Sayed & Kamel 2020).

The culture-independent analyses conducted in chapter 2 capture more microbial diversity than chapter 1 by directly sequencing total genomic DNA from soils rather than relying on cultures. In this chapter I presented evidence of highly diverse and abundant ARGs, including those encoded on MGEs, across bacterial taxa in our samples. Thus, the evidence presented in chapter 2 implicated antibiotic resistance in soils as an intrinsic component of bacterial evolution while highlighting the need for novel therapeutics and continued antibiotic stewardship to conserve the arsenal of drugs we already have for treating bacterial infections. From the long read metagenomic data, I identified genes encoding all mechanisms of resistance similarly to the findings of chapter 1. However, in this metagenomic dataset there was a much higher diversity of ARGs detected highlighting the importance of culture-independent microbial analyses for capturing more variation in soils, which are dominated by uncultivable taxa. The high abundance of beta-lactamase genes detected from our metagenomic dataset suggests the presence of selective pressures for the maintenance of these ARGs. Given the remoteness of our study site, these selective pressures are likely the result of antibiotic producing soil microbes rather than anthropogenic antibiotic pollution. Indeed, we found that Actinobacteria, a phylum of prolific antibiotic producers that account for 45% of all bioactive microbial metabolites discovered (Anandan et al. 2013), were the second most abundant phylum with *Streptomycetaceae,* a prolific producer of clinically significant antibiotics, being the second most abundant family in our soils. This high abundance of antibiotic producing taxa highlights the potential co-evolutionary 'arms-shield race' occurring in Alaskan soils.

The significant relationship between shifts in community composition and ARG abundance both temporally and with soil disturbance, unearths the complex role climatic-induced changes to permafrost soil conditions have on ARG evolution. I found that there was a quadratic relationship between disturbance and ARG abundance such that undisturbed soil communities with intact permafrost and late successional stage vegetation were enriched in ARGs and more disturbed soil communities with high-level permafrost thaw and early successional stage vegetation were enriched in ARGs. On the other hand, semi-disturbed soils, which had a stable permafrost table but a mid successional stage forest stand, had the lowest abundance of ARGs. The shift in taxa observed with soil disturbance at FPES directly corresponded with microbial succession observed in other high-latitude soils disturbed via deglaciation (Nemergut et al. 2007) and wildfire (Whitman et al. 2019). The de-glaciation study found that *Comamonadaceae* was a dominant clade in the young soils, decreased in

intermediate age soil, and was absent from long established soils which corresponds to the trend observed with *Comamonadaceae* across our disturbance gradient (Nemergut et al. 2007, Seitz et al. 2021). Although *Comamonadaceae* is a phenotypically diverse group of organisms, isolates from this clade are capable of metabolizing carbon from limited resources in carbon poor environments, such as those released from thawing permafrost and glacial ice cores (Sheridan et al 2003). In studies examining environmental resistomes, such as wastewater, *Comamonadaceae* was a major reservoir of ARGs (Narciso-da-Rocha et al. 2018). Analogously to our findings, in wastewater an increasing abundance of *Comamonadaceae* was significantly correlated with the prevalence of resistance (Gerzova et al. 2014). Another clade at our site where we observed consistent trends with other undisturbed soils was Acidobacteria (Seitz et al. 2021). Acidobacteria utilizes complex carbon substrates that are present in older established soils (Kielak et al. 2016). Moreover, previous findings also showed Acidobacteria were a primary source of ARGs in undisturbed Antarctic soils (Van Goethem et al. 2018). These findings, along with studies that have shown enrichment of soil nutrients like N increase the prevalence of ARGs (Fierer et al. 2012), highlight the significance of disturbance induced shifts in soil characteristics and the subsequent shift in bacterial taxa that shape the resistome.

Another fruitful method for exploring ARGs in soils is through the analysis of individual bacterial genomes. Whole genomes can reveal functional potential of individual taxa for determining the abundance of ARGs detected in soil communities. In the second thesis chapter we reconstructed individual genomes without cultures using Hi-C proximity ligation. Previous studies have demonstrated the ability of Hi-C proximity ligation with culture-independent de novo deconvolution to reveal plasmid-host associations in high quality bacterial genomes reconstructed from mixed samples without the need for prior information (e.g., reference genomes; Press et al. 2017, Stadler et al. 2019). Therefore, genomes generated with Hi-C may provide critical insight into host-MGE associations from a greater diversity of bacteria than culture-dependent genomics alone. Our Hi-C MAGs revealed that specific taxa had more ARGs per genome than others, including Proteobacteria and Acidobacteria. Previous findings showed that Acidobacteria were significantly more abundant in UD FPES soils and Proteobacteria were more abundant in MD soils (Seitz et al. 2021). These findings, paired with the MAGs produced in chapter 2 and the genomes produced in chapter 1 both emphasized how the enrichment of specific bacterial taxa would result in an increased abundance of ARGs at a community level (e.g., the quadratic relationship with disturbance-induced community shifts and ARG abundance).

In terms of the mechanism behind why disturbance may impact ARGs, a previous study found wildfire in the Canadian boreal forest enriched prolific antibiotic producing microbial taxa *Penicillium* and *Streptomyces* (Whitman et al. 2019). Our findings found *Streptomyces* was abundant taxa across disturbance treatments. The high abundance of this taxa in Alaskan soils suggests a co-evolutionary arms-shield race that directly selects for resistant bacteria by enriching antibiotic producers through shifting niche availability (Hibbing et al. 2010). Disturbance to Alaskan soils could also indirectly select for ARGs by favoring mechanisms that allow bacteria to concurrently cope with antibiotics and biological stressors released from permafrost, such as biogenic volatile organic compounds (Kramshøj et al. 2018). For example, efflux pumps are proteins in a variety of environmental bacteria that have been shown to confer multidrug resistance and promote stress tolerance (Ramos et al. 2001) and were detected in most Hi-C MAGs generated in chapter 2. Environmental stress has also been suggested to stimulate intra-genomic (e.g., integrons) and inter-species gene transfer (Velkov 1999). Thus, as climate change progressively generates stressors that affect soil bacteria, there may be an increase the rate of ARG dissemination and the abundance of ARGs housed on conjugative plasmids (Djordjevic et al. 2013).

Future studies could use the metagenomic data produced in this thesis to examine the prevalence of soil-borne plant, human, and animal pathogens in subarctic soils along with how disturbance-induced permafrost thaw shifts the abundance and types of pathogenic bacterial, fungal, and viral taxa (e.g., *Pseudomonas aeruginosa, Bacillus anthracis, Clostridium tetani, Francisella tularensis, Aspergillus fumigatus* and *Sporothrix schenckii*) (Nieder et al. 2018). Moreover, by identifying biogeochemical properties of soils at our study site we could disentangle how these abiotic factors structure predominant pathogenic microbes and antibiotic resistant taxa in soil. The MAGs generated using Hi-C proximity ligation with short-read assemblies can be refined with additional long read sequence data in order to produce completely circularized genomes of bacteria. Near-complete genomes are much easier to annotate and identify biosynthetic gene clusters that can shed light on novel secondary metabolites in Alaskan soils. Thus, a refinement of MAGs would complement findings in this study by providing insight into the potential antibiotic producing taxa driving this high prevalence of ARGs.

To recap, the presence of diverse and abundant ARGs, including ARGs encoded on mobile elements, in both chapter 1 and 2 implicates active layer soils as a reservoir of resistant bacteria. This reservoir has the potential to compromise human, animal, and plant health by acting as a source of bacteria that can transfer ARGs to pathogens via horizontal gene transfer.

The findings in both chapters affirm subarctic soils contain bacteria resistant to clinically significant antibiotics highlighting the intrinsic nature of resistance in soil. By monitoring which ARGs are present in soils at this locale, we can predict which ARGs may arise as clinical problems as human expansion into subarctic ecosystems continues. The high abundance of ARGs encoding antibiotic inactivating enzymes implicates selective pressures that maintain these genes, most likely driven by the presence of antibiotic producing taxa. The individual genomes provide a wealth of genomic information for future researchers to unveil the functional role of specific Alaska soil bacterial phylogenies in ecosystem health and resilience. Moving forward these genomes can be refined further and used as a resource to mine for biosynthetic gene clusters that encode novel bioactive compounds that have the potential to aid in therapeutics that could target drug resistant pathogens in a clinical setting. Overall, this thesis provides insight into the ARGs comprising the subarctic soil resistome that may emerge clinically and how the resistome is shaped by disturbance-induced shifts in soil communities.

REFERENCES.

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A., & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *The ISME Journal*, *3*(2), 243–251. DOI: 10.1038/ismej.2008.86

Anandan, R., Dharumadurai, D., & Manogaran, G. P. (2016). An introduction to actinobacteria. In *Actinobacteria-Basics and Biotechnological Applications*. Intechopen. DOI: 10.5772/62329

Berg, J., Tom-Petersen, A., & Nybroe, O. (2005) Copper amendment of agricultural soil selects for bacterial antibiotic resistance in the field. *Letters in Applied Microbiology*, 40(2), 146-151. DOI: 10.1111/j.1472-765X.2004.01650.x.

Cruz, A. T., Cazacu, A. C., & Allen, C. H. (2007). Pantoea agglomerans, a plant pathogen causing human disease. *Journal of clinical microbiology*, *45*(6), 1989-1992. DOI: 10.1128/JCM.00632-07

Djordjevic, S. P., Stokes, H. W., & Roy Chowdhury, P. (2013). Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Frontiers in microbiology*, *4*, 86. DOI: 10.3389/fmicb.2013.00086

Dunivin, T. K., Choi, J., Howe, A., & Shade, A. (2019). RefSoil+: a Reference Database for Genes and Traits of Soil Plasmids. *mSystems*, 4(1). DOI: 10.1128/mSystems.00349-18

El-Sayed, A., & Kamel, M. (2020). Climatic changes and their role in emergence and re-emergence of diseases. *Environmental Science and Pollution Research*, *27*(18), 22336-22352. DOI: 10.1007/s11356-020-08896-w

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., ... & Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, *109*(52), 21390-21395. DOI: 10.1073/pnas.1215210110

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., & Dantas, G. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature*, 509(7502), 612-616. DOI: 10.1038/nature13377

Gerzova, L., Videnska, P., Faldynova, M., Sedlar, K., Provaznik, I., Cizek, A., & Rychlik, I. (2014). Characterization of microbiota composition and presence of selected antibiotic resistance genes in carriage water of ornamental fish. *PLoS One*, *9*(8), e103865. DOI: 10.1371/journal.pone.0103865

Haan, T. J., & Drown, D. M. (2021). Unearthing Antibiotic Resistance Associated with Disturbance-Induced Permafrost Thaw in Interior Alaska. *Microorganisms*, 9(1), 116. DOI: 10.3390/microorganisms9010116

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology*, *8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Holland, M. M., & Bitz, C. M. (2003). Polar amplification of climate change in coupled models. *Climate Dynamics*, *21*(3-4), 221-232. DOI 10.1007/s00382-003-0332-6

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., & Kuramae, E. E. (2016). The ecology of Acidobacteria: moving beyond genes and genomes. *Frontiers in microbiology*, 7, 744. DOI: 10.3389/fmicb.2016.00744

Kobziar, L. N., Pingree, M. R., Larson, H., Dreaden, T. J., Green, S., & Smith, J. A. (2018). Pyroaerobiology: the aerosolization and transport of viable microbial life by wildland fire. *Ecosphere*, 9(11), e02507. DOI: 10.1002/ecs2.2507

Kotiranta, A., Lounatmaa, K., & Haapasalo, M. (2000). Epidemiology and pathogenesis of Bacillus cereus infections. *Microbes and infection*, *2*(2), 189-198. DOI: 10.1016/s1286-4579(00)00269-0

Kramshøj, M., Albers, C. N., Holst, T., Holzinger, R., Elberling, B., & Rinnan, R. (2018). Biogenic volatile release from permafrost thaw is determined by the soil microbial sink. *Nature communications*, 9(1), 1-9. DOI: 10.1038/s41467-018-05824-y

Lax, S., & Gilbert, J. A. (2015). Hospital-associated microbiota and implications for nosocomial infections. *Trends in molecular medicine*, *21*(7), 427-432. DOI: 10.1016/j.molmed.2015.03.005

Narciso-da-Rocha, C., Rocha, J., Vaz-Moreira, I., Lira, F., Tamames, J., Henriques, I., ... & Manaia, C. M. (2018). Bacterial lineages putatively associated with the dissemination of antibiotic resistance genes in a full-scale urban wastewater treatment plant. *Environment international*, *118*, 179-188. DOI: 10.1016/j.envint.2018.05.040

Nemergut, D. R., Anderson, S. P., Cleveland, C. C., Martin, A. P., Miller, A. E., Seimon, A., & Schmidt, S. K. (2007). Microbial community succession in an unvegetated, recently deglaciated soil. *Microbial ecology*, *53*(1), 110-122. DOI: 10.1007/s00248-006-9144-7

Nesme, J., Cécillon, S., Delmont, T. O., Monier, J. M., Vogel, T. M., & Simonet, P. (2014). Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current biology*, *24*(10), 1096-1100. DOI: 10.1016/j.cub.2014.03.036

Nesme, J., & Simonet, P. (2015). The soil resistome: a critical review on antibiotic resistance origins, ecology and dissemination potential in telluric bacteria. *Environmental microbiology*, *17*(4), 913-930. DOI: 10.1111/1462-2920.12631

Nieder, R., Benbi, D. K., & Reichl, F. X. (2018). Soil as a Transmitter of Human Pathogens. In *Soil Components and Human Health* (pp. 723-827). Springer, Dordrecht. DOI: 10.1007/978-94-024-1222-2_13

Pérez-Valdespino, A., Pircher, R., Pérez-Domínguez, C. Y., & Mendoza-Sanchez, I. (2021). Impact of flooding on urban soils: Changes in antibiotic resistance and bacterial community after Hurricane Harvey. *Science of the Total Environment*, *766*, 142643. DOI: 10.1016/j.scitotenv.2020.142643

Press, M. O., Wiser, A. H., Kronenberg, Z. N., Langford, K. W., Shakya, M., Lo, C. C., ... & Liachko, I. (2017). Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *biorxiv*, 198713. DOI: 10.1101/198713

Ramos, J. L., Gallegos, M. T., Marqués, S., Ramos-González, M. I., Espinosa-Urgel, M., & Segura, A. (2001). Responses of Gram-negative bacteria to certain environmental stressors. *Current opinion in microbiology*, *4*(2), 166-171. DOI: 10.1016/s1369-5274(00)00183-1

Seitz, T. J., Schütte, U. M., & Drown, D. M. (2021). Soil Disturbance Affects Plant Productivity via Soil Microbial Community Shifts. *Frontiers in microbiology*, *12*, 76. DOI: 10.3389/fmicb.2021.619711

Sheridan, P. P., Miteva, V. I., & Brenchley, J. E. (2003). Phylogenetic analysis of anaerobic psychrophilic enrichment cultures obtained from a Greenland glacier ice core. *Applied and Environmental Microbiology*, *69*(4), 2153-2160. DOI: 10.1128/AEM.69.4.2153-2160.2003

Torsvik V, Øvreås L, Thingstad TF. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science*. 2002 May 10;296(5570):1064-6. DOI: 10.1126/science.1071698

Van Goethem, M. W., Pierneef, R., Bezuidt, O. K., Van De Peer, Y., Cowan, D. A., & Makhalanyane, T. P. (2018). A reservoir of 'historical'antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome*, *6*(1), 1-12. DOI: 10.1186/s40168-018-0424-5

Velkov, V. V. (1999). How environmental factors regulate mutagenesis and gene transfer in microorganisms. *Journal of Biosciences*, *24*(4), 529-559.

Walvoord, M. A., & Kurylyk, B. L. (2016). Hydrologic impacts of thawing permafrost—A review. *Vadose Zone Journal*, *15*(6). DOI: 10.2136/vzj2016.01.0010

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics are produced by the genus Streptomyces?. *Archives of microbiology*, *176*(5), 386-390. DOI: 10.1007/s002030100345.

Wendler, G., & Shulski, M. (2009). A Century of Climate Change for Fairbanks, Alaska. *Arctic, 62*(3), 295-300. Retrieved April 30, 2021, from http://www.jstor.org/stable/40513307

Whitman, T., Whitman, E., Woolet, J., Flannigan, M. D., Thompson, D. K., & Parisien, M. A. (2019). Soil bacterial and fungal response to wildfires in the Canadian boreal forest across a burn severity gradient. *Soil Biology and Biochemistry*, 138, 107571. DOI: 10.1016/j.soilbio.2019.107571

Woolhouse, M., Ward, M., van Bunnik, B., & Farrar, J. (2015). Antimicrobial resistance in humans, livestock and the wider environment. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 370(1670), 20140083.

Zepp, R. G., Erickson Iii, D. J., Paul, N. D., & Sulzberger, B. (2007). Interactive effects of solar UV radiation and climate change on biogeochemical cycling. *Photochemical & Photobiological Sciences*, *6*(3), 286-300.

APPENDIX

Table A1. Zone of inhibition breakpoints (mm) from the US Clinical and Laboratory Standards Institute M100, 30th ed. based on the genus of FPES isolates against Ampicillin, Chloramphenicol, Erythromycin, Kanamycin, and Tetracycline with R = resistant, I = intermediate, and S = susceptible.

CLSI Breakpoint	FPES Isolate	Ampicillin			Chloramphenicol			Erythromycin			Kanamycin			Tetracycline		
Class	Genus	R	I	S	R	I	S	R	I	S	R	I	S	R	I	S
Pseudomonas spp	*Pseudomonas*	–	–	–	12	13–17	18	–	–	-	12	13–14	15	11	12–14	15
Enterobacterales	*Erwinia*	13	14–16	17	12	13–17	18	12	–	13	13	14–17	18	11	12–14	15
Enterobacterales	*Pantonea*	13	14–16	17	12	13–17	18	12	–	13	13	14–17	18	11	12–14	15
Enterobacterales	*Serratia*	13	14–16	17	12	13–17	18	12	–	13	13	14–17	18	11	12–14	15
Enterococcus spp	*Bacillus*	16	–	17	12	13–17	18	13	14–22	23	–	–	–	14	15–18	19
Enterococcus spp	*Exiguobacterium*	16	–	17	12	13–17	18	13	14–22	23	–	–	–	14	15–18	19

Table A2. ONT library prep and sequencing run by isolate

ONT Libarary Prep	Sequencing Date	Run Script	RunTime	Samples Sequenced												
SQK-RBK004	20180724A	RBK004	48hr	TH01	TH10	TH13	TH21	TH31	TH33	TH45	TH52	TH61	TH64	TH72	TH80	
	20180724B	RBK004	48hr	TH04	TH12	TH15	TH30	TH34	TH35	TH41	TH56	TH65	TH66	TH74	TH86	
	20180724C	RBK004	48hr	TH03	TH18	TH19	TH29	TH39	TH40	TH47	TH60	TH69	TH76	TH77	TH90	
	20200203	RBK004	72hr	TH02	TH11	TH22	TH32	TH42	TH51	TH62	TH73	TH82	TH05	TH43	TH83	
	20200207	RBK004	72hr	TH08	TH58	TH50	TH14	TH16								
	20200214	RBK004	72hr	TH63	TH36	TH06	TH75	TH44	TH84	TH53	TH23	TH67	TH54			
	20200221	RBK004	72hr	TH07	TH17	TH24	TH25	TH37	TH46	TH55	TH68	TH71	TH85	TH87		
	20200224	RBK004	72hr	TH70	TH38	TH10	TH78	TH48	TH20	TH86	TH57	TH27	TH15	TH49		
VSK-VSK002	20180505	VSK002	48hr	TH81												
	20180515	VSK002	48hr	TH59												
	20190228	VSK002	48hr	TH26												
VSK- VMK002	20191113	VMK002	72hr	TH28												

Table A3. Assembly statistics and checkM metrics of quality for each isolate's assembly.

Isolate	length	Contigs	Mean contig length	largest contig	N50	N50,n	GC %	GC std	Coding density:	# genes:	% Complete:	% Contam.	marker lineage
TH01	7289415	1	7289415	7289415	7289415	1	0.62	0.00	0.89	6521	99.67	0.87	'o__Pseudomonadales'
TH02	6428555	1	6428555	6428555	6428555	1	0.59	0.00	0.87	5729	99.93	0.05	'g__Pseudomonas'
TH03	6292986	158	39829	230652	56073	39	0.59	0.02	0.88	5752	84.78	1.69	'o__Pseudomonadales'
TH04	6270276	2	3135138	6269627	6269627	1	0.60	0.00	0.89	5593	99.86	0.52	'g__Pseudomonas'
TH05	7101889	70	101456	491789	165760	13	0.62	0.02	0.89	6507	98.59	0.87	'o__Pseudomonadales'
TH06	6565344	4	1641336	5329916	5329916	1	0.59	0.00	0.88	5759	99.89	0.34	'g__Pseudomonas'
TH07	7282377	2	3641189	7270375	7270375	1	0.62	0.00	0.89	6530	100.00	0.87	'o__Pseudomonadales'
TH08	6565193	6	1094199	2115574	1462546	2	0.59	0.00	0.88	5865	99.93	0.34	'g__Pseudomonas'
TH09	6300396	72	87506	1275846	563232	4	0.60	0.04	0.89	5737	100.00	0.19	'o__Pseudomonadales'
TH10	4977657	179	27808	105294	35424	52	0.59	0.02	0.87	5195	75.11	0.76	'g__Pseudomonas'
TH11	3239959	138	23478	116995	30084	34	0.36	0.02	0.81	4199	80.39	0.27	'o__Bacillales'
TH12	5819177	6	969863	5081309	5081309	1	0.35	0.03	0.82	5863	98.90	0.50	'g__Bacillus'
TH13	6132576	25	245303	4565714	4565714	1	0.35	0.04	0.83	6375	99.34	0.76	'g__Bacillus'
TH14	6352696	3	2117565	5830573	5830573	1	0.38	0.00	0.78	6056	98.63	3.44	'f__Bacillaceae'
TH15	6252465	2	3126233	5280705	5280705	1	0.59	0.00	0.87	5674	99.67	0.11	'o__Pseudomonadales'
TH16	6232267	12	519356	2999140	2358609	2	0.38	0.03	0.78	5927	98.63	3.72	'f__Bacillaceae'
TH17	5747064	6	957844	5159847	5159847	1	0.35	0.03	0.82	5826	98.02	0.77	'g__Bacillus'
TH18	7152072	1	7152072	7152072	7152072	1	0.60	0.00	0.88	6271	100.00	0.54	'o__Pseudomonadales'
TH19	5767003	16	360438	2590843	2009945	2	0.35	0.03	0.82	5822	98.90	0.29	'g__Bacillus'
TH20	6987576	201	34764	277831	70357	26	0.59	0.02	0.88	6383	98.53	0.24	'o__Pseudomonadales'
TH21	6760171	2	3380086	6753006	6753006	1	0.62	0.11	0.89	6000	99.57	0.41	'g__Pseudomonas'
TH22	5822214	3	1940738	5423614	5423614	1	0.35	0.02	0.82	5874	98.33	1.55	'g__Bacillus'
TH23	5806530	7	829504	3853346	3853346	1	0.35	0.03	0.82	5907	98.49	1.55	'g__Bacillus'
TH24	6188561	5	1237712	5688067	5688067	1	0.38	0.02	0.78	5938	98.09	3.17	'f__Bacillaceae'
TH25	5862716	39	150326	1409789	711156	3	0.35	0.02	0.82	5916	97.84	0.95	'g__Bacillus'
TH26	5790936	2	2895468	5307118	5307118	1	0.35	0.01	0.82	5779	98.68	0.15	'g__Bacillus'
TH27	6276694	4	1569174	5739858	5739858	1	0.38	0.02	0.78	5984	98.63	3.17	'f__Bacillaceae'

TH28	5123563	239	21438	170590	39071	37	0.55	0.04	0.88	4859	98.09	0.89	'f__Enterobacteriaceae'
TH29	3154365	1477	2136	11893	2306	442	0.54	0.04	0.87	4077	71.97	0.55	'f__Enterobacteriaceae'
TH30	5897091	6	982849	5270848	5270848	1	0.35	0.04	0.82	5863	98.90	0.21	'g__Bacillus'
TH31	7196786	38	189389	950737	452287	6	0.59	0.03	0.86	7112	98.42	1.69	'g__Pseudomonas'
TH32	7208438	4	1802110	6375311	6375311	1	0.61	0.03	0.89	6670	99.89	0.27	'g__Pseudomonas'
TH33	6929409	1	6929409	6929409	6929409	1	0.61	0.00	0.89	6258	99.86	0.83	'g__Pseudomonas'
TH34	7008149	35	200233	752473	447628	6	0.60	0.02	0.89	6412	99.83	0.21	'g__Pseudomonas'
TH35	6185610	409	15124	421554	141290	14	0.61	0.05	0.88	5996	97.72	0.57	'o__Pseudomonadales'
TH36	6757750	1	6757750	6757750	6757750	1	0.58	0.00	0.89	6241	100.00	0.62	'o__Pseudomonadales'
TH37	6297323	357	17640	345118	170805	13	0.61	0.05	0.88	6060	100.00	0.57	'o__Pseudomonadales'
TH38	6881557	56	122885	870092	341859	6	0.59	0.02	0.89	6255	99.12	0.75	'g__Pseudomonas'
TH39	7179220	8	897403	2756366	1664454	2	0.59	0.01	0.88	6562	99.93	1.08	'g__Pseudomonas'
TH40	6136442	221	27767	225211	73707	29	0.61	0.04	0.87	6090	97.20	0.68	'o__Pseudomonadales'
TH41	6782636	9	753626	3093317	1191987	2	0.59	0.00	0.87	6095	99.35	1.17	'g__Pseudomonas'
TH42	6581347	1	6581347	6581347	6581347	1	0.60	0.00	0.87	5801	99.93	0.97	'g__Pseudomonas'
TH43	6500574	5	1300115	2489632	1546032	2	0.59	0.00	0.87	5811	99.93	0.11	'g__Pseudomonas'
TH44	5938642	59	100655	1212980	347271	5	0.35	0.03	0.81	6211	96.69	1.55	'g__Bacillus'
TH45	5862293	15	390820	4124331	4124331	1	0.35	0.05	0.82	5984	97.89	1.86	'g__Bacillus'
TH46	6582152	1	6582152	6582152	6582152	1	0.60	0.00	0.87	5804	99.93	0.97	'g__Pseudomonas'
TH47	6580544	1	6580544	6580544	6580544	1	0.60	0.00	0.87	5800	99.93	0.97	'g__Pseudomonas'
TH48	6751879	1	6751879	6751879	6751879	1	0.62	0.00	0.89	5973	99.93	0.14	'g__Pseudomonas'
TH49	6641420	2	3320710	3792784	3792784	1	0.60	0.00	0.87	5906	99.93	0.40	'g__Pseudomonas'
TH50	5932849	8	741606	3087957	3087957	1	0.35	0.04	0.82	6012	98.38	1.86	'g__Bacillus'
TH51	5966177	6	994363	3086880	3086880	1	0.35	0.04	0.82	6163	98.24	1.55	'g__Bacillus'
TH52	5965749	15	397717	1979436	1731552	2	0.35	0.02	0.82	6074	98.68	0.61	'g__Bacillus'
TH53	5954056	5	1190811	5462858	5462858	1	0.35	0.04	0.82	6117	98.46	1.53	'g__Bacillus'
TH54	5941371	4	1485343	5495595	5495595	1	0.35	0.08	0.82	6055	98.33	1.55	'g__Bacillus'
TH55	4877094	637	7656	159431	12721	90	0.36	0.02	0.85	5058	93.87	0.38	'g__Bacillus'
TH56	4853631	1	4853631	4853631	4853631	1	0.55	0.00	0.88	4388	100.00	0.00	'f__Enterobacteriaceae'
TH57	5203616	617	8434	188393	15558	81	0.36	0.02	0.85	5407	96.57	1.48	'g__Bacillus'
TH58	5992248	9	665805	2344683	2197132	2	0.35	0.03	0.82	6107	98.55	1.86	'g__Bacillus'

TH59	4906365	790	6211	80706	10625	111	0.36	0.02	0.86	5137	94.34	0.73	'g__Bacillus'
TH60	6581323	1	6581323	6581323	6581323	1	0.60	0.00	0.87	5795	99.93	0.97	'g__Pseudomonas'
TH61	6408248	1	6408248	6408248	6408248	1	0.59	0.00	0.88	5691	99.93	0.05	'g__Pseudomonas'
TH62	6683507	3	2227836	6638173	6638173	1	0.59	0.06	0.88	6108	100.00	0.28	'o__Pseudomonadales'
TH63	6157383	1	6157383	6157383	6157383	1	0.59	0.00	0.87	5566	99.39	0.79	'g__Pseudomonas'
TH64	6733090	2	3366545	6465401	6465401	1	0.59	0.05	0.88	6231	100.00	0.30	'o__Pseudomonadales'
TH65	6397841	1	6397841	6397841	6397841	1	0.60	0.00	0.88	5882	99.73	0.66	'g__Pseudomonas'
TH66	6131349	144	42579	201108	70924	28	0.59	0.02	0.88	5639	98.25	14.04	'k__Bacteria'
TH67	6682590	3	2227530	6637397	6637397	1	0.59	0.06	0.88	6102	100.00	0.28	'o__Pseudomonadales'
TH68	7157743	3	2385914	4088561	4088561	1	0.59	0.03	0.87	6529	99.68	0.21	'o__Pseudomonadales'
TH69	5914792	1	5914792	5914792	5914792	1	0.59	0.00	0.88	5243	99.66	0.08	'g__Pseudomonas'
TH70	7078794	3	2359598	6999391	6999391	1	0.60	0.01	0.88	6367	100.00	1.09	'o__Pseudomonadales'
TH71	6270423	38	165011	1061543	465089	5	0.59	0.03	0.88	5540	96.84	0.15	'g__Pseudomonas'
TH72	7744191	23	336704	3545502	1844773	2	0.59	0.05	0.87	6963	98.76	0.67	'g__Pseudomonas'
TH73	3185552	5	637110	3152340	3152340	1	0.48	0.09	0.89	3267	99.34	0.33	'c__Bacilli'
TH74	6766338	6	1127723	2599615	2167177	2	0.60	0.03	0.87	6131	99.79	1.07	'g__Pseudomonas'
TH75	6765294	3	2255098	6708000	6708000	1	0.60	0.04	0.87	6128	99.79	1.07	'g__Pseudomonas'
TH76	5908038	1	5908038	5908038	5908038	1	0.63	0.00	0.89	5248	99.84	0.14	'o__Pseudomonadales'
TH77	7361083	8	920135	6195251	6195251	1	0.60	0.05	0.88	6784	99.86	0.87	'g__Pseudomonas'
TH78	5908405	1	5908405	5908405	5908405	1	0.63	0.00	0.89	5221	100.00	0.14	'o__Pseudomonadales'
TH79	4583151	71	64551	283231	114121	14	0.56	0.03	0.88	4203	99.64	0.24	'f__Enterobacteriaceae'
TH80	6539415	7	934202	3216008	1092864	2	0.59	0.01	0.88	5775	99.66	0.15	'g__Pseudomonas'
TH81	4984468	4	1246117	4128817	4128817	1	0.55	0.03	0.85	4539	100.00	1.50	f__Enterobacteriaceae'
TH82	4647952	2	2323976	4554121	4554121	1	0.56	0.01	0.88	4269	99.96	0.24	'f__Enterobacteriaceae'
TH83	3297868	60	54964	369722	114945	8	0.66	0.21	0.84	3993	89.34	18.46	'k__Bacteria'
TH84	4930048	54	91297	520953	141253	11	0.55	0.03	0.85	4540	99.54	1.80	'f__Enterobacteriaceae'
TH85	5873192	6	978865	5302587	5302587	1	0.35	0.03	0.82	5896	99.01	0.32	'g__Bacillus'
TH86	5476598	280	19559	240496	45743	29	0.35	0.03	0.84	5642	97.91	0.77	'g__Bacillus'
TH87	5967729	13	459056	2640400	2484126	2	0.35	0.03	0.82	6009	98.90	0.48	'g__Bacillus'
TH88	4903410	2	2451705	4824998	4824998	1	0.55	0.01	0.88	4444	100.00	0.05	'f__Enterobacteriaceae'

TH89	5889526	7	841361	5293100	5293100	1	0.35	0.03	0.82	6099	98.90	0.04	'g__Bacillus'
TH90	6751295	1	6751295	6751295	6751295	1	0.62	0.00	0.88	6311	99.35	0.23	'g__Pseudomonas'
MEAN	**6096842**	**75**	**2102452**	**3839076**	**3639582**	**13**	**0.52**	**0.03**	**0.86**	**5772**	**98.01**	**1.16**	
STDER	**96902**	**22**	**251424**	**261837**	**279870**	**5**	**0.01**	**0.00**	**0.00**	**73**	**0.50**	**0.26**	

Table A4. Description of CARD gene hits from all isolates in terms of resistance mechanism, best hit gene, drug class associated with that gene, gene family, count of gene copies for each hit across the 90 isolates, along with the mean and standard deviation of percent length of reference sequence and percent identity as defined by RGI.

Resistance Mechanism	*Best_Hit_ARO*	*Drug Class*	*AMR Gene Family*	*Count*	*% length of reference sequence*	*% identity*
antibiotic efflux	Acinetobacter baumannii AbaQ	fluoroquinolone antibiotic	major facilitator superfamily (MFS) antibiotic efflux pump	40	101.363 ± 0.109	72.675 ± 0.427
	adeF	fluoroquinolone antibiotic; tetracycline antibiotic	resistance-nodulation-cell division (RND) antibiotic efflux pump	176	99.02 ± 2.902	52.283 ± 11.822
	CRP	macrolide antibiotic; fluoroquinolone antibiotic; penam	RND antibiotic efflux pump	8	100 ± 0	98.87 ± 0.248
	emrR	fluoroquinolone antibiotic	MFS antibiotic efflux pump	7	99.757 ± 0.643	82.049 ± 1.001
	Klebsiella pneumoniae KpnE	Broad Spectrum	MFS antibiotic efflux pump	2	96.67 ± 0	66.95 ± 0
	Klebsiella pneumoniae KpnF	Broad Spectrum	MFS antibiotic efflux pump	8	100 ± 0	73.389 ± 2.903
	Klebsiella pneumoniae KpnH	Broad Spectrum	MFS antibiotic efflux pump	6	99.8 ± 0	86.705 ± 1.38
	msbA	nitroimidazole antibiotic	ABC antibiotic efflux pump	7	100 ± 0	87.827 ± 1.059
	tet(45)	tetracycline antibiotic	MFS antibiotic efflux pump	1	100 ± NA	89.5 ± NA
	TriC	triclosan	RND antibiotic efflux pump	2	2.76 ± 0	100 ± 0
antibiotic inactivation	AAC(6')-32	aminoglycoside antibiotic	AAC(6')	1	122.83 ± NA	100 ± NA
	AAC(6')-Ib7	aminoglycoside antibiotic	AAC(6')	1	65.13 ± NA	100 ± NA
	AAC(6')-Ir	aminoglycoside antibiotic	AAC(6')	3	272.6 ± 0	100 ± 0
	BcII	cephalosporin; penam	Bc beta-lactamase	20	100.39 ± 0	90.755 ± 0.567
	BES-1	penam	BES Beta-lactamase	1	27.05 ± NA	100 ± NA
	BPU-1	penam	BPU Beta-lactamase	4	51.15 ± 0	100 ± 0
	Escherichia coli ampC1 beta-lactamase	cephalosporin; penam	ampC-type beta-lactamase	1	94.7 ± NA	100 ± NA
	Escherichia coli ampH beta-lactamase	cephalosporin; penam	ampC-type beta-lactamase	6	101.733 ± 1.17	70.063 ± 2.367
	FosB	fosfomycin	fosfomycin thiol transferase	25	105.012 ± 6.992	89.055 ± 2.281
antibiotic target alteration	armA	aminoglycoside antibiotic	16S rRNA methyltransferase (G1405)	11	81.214 ± 0.182	100 ± 0
	bcrC	peptide antibiotic	undecaprenyl pyrophosphate related proteins	2	166.995 ± 100.317	100 ± 0
	MCR-4.1	peptide antibiotic	MCR phosphoethanolamine transferase	2	10.63 ± 1.174	100 ± 0
	Morganella morganii gyrB	fluoroquinolone antibiotic	fluoroquinolone resistant gyrB	6	99.75 ± 0	80.43 ± 0.279
	PmrF	peptide antibiotic	pmr phosphoethanolamine transferase	2	100.93 ± 0	80.5 ± 0
	sgm	aminoglycoside antibiotic	16S rRNA methyltransferase	1	16.42 ± NA	100 ± NA
	vanJ	glycopeptide antibiotic	vanJ membrane protein	1	11.52 ± NA	100 ± NA
antibiotic target alteration; antibiotic efflux	Pseudomonas aeruginosa soxR	Broad Spectrum	ABC, MFS, RND antibiotic efflux pump	35	96.793 ± 1.854	68.494 ± 1.961

Table A5. Number of copies of each gene hit in the sequence data by isolate and phenotypic results with zone of inhibition measurements (mm) and CLSI breakpoint interpretation (R-resistant; I-Intermediate; S-susceptible; NA- not applicable for due to intrinsic resistance) for Kirby-Bauer disk diffusion assay.

ISOLATE INFO			Antibiotic Inactivation									Antibiotic Efflux											Antibiotic Target Alteration							Phenotypic Results									
			aminoglycoside			[illegible]					fosfomycin	fluoroquinolone, tetracycline, macrolide, penam				Broad Spectrum				Nitroimidazole	Tetracycline	Triclosan	aminoglycoside		Peptide antibiotic			fluoroquinolone	glycopeptide	Ampicillan		Chloramphenicol		Erythromycin		Kanamycin		Tetracycline	
COLONY	ID	GENUS	[illegible]	[illegible]	[illegible]	ampC1 beta-lactamase	ampH beta-lactamase	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CRP	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TriC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Inhibition zone (mm)	Level of Resistance	Inhibition zone (mm)	Level of Resistance	Inhibition zone (mm)	Level of Resistance	Inhibition zone (mm)	Level of Resistance	Inhibition zone (mm)	Level of Resistance
UD+4 colony 1	TH01	Pseudomonas	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	10	R	0	NA	17	I	16	I
UD+4 colony 2	TH02	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	19	S	14	I
UD+4 colony 3	TH03	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	18	S
UD+4 colony 4	TH04	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	10	S	16	I
UD+4 colony 5	TH05	Pseudomonas	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	16	I	15	I
UD+4 colony 6	TH06	Pseudomonas	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	8	R	0	NA	18	I	19	S
UD+4 colony 7	TH07	Pseudomonas	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	15	I	11	I
UD+4 colony 8	TH08	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	17	I	21	S
UD+4 colony 9	TH09	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	11	R	0	NA	20	S	20	S
UD+4 colony 10	TH10	Pseudomonas	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	18	I	18	S
UD+12 colony 1	TH11	Bacillus	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	S	35	S	35	S	26	S	33	S
UD+12 colony 2	TH12	Bacillus	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	30	S	32	S	25	S	26	S
UD+12 colony 3	TH13	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	R	23	S	30	S	20	S	25	S
UD+12 colony 4	TH14	Bacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	S	32	S	30	S	27	S	29	S
UD+12 colony 5	TH15	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	13	I	0	NA	14	I	21	S
UD+12 colony 6	TH16	Bacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	S	30	S	33	S	27	S	32	S
UD+12 colony 7	TH17	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	40	S	32	S	27	S	25	S
UD+12 colony 8	TH18	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	22	S	20	S
UD+12 colony 9	TH19	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	30	S	32	S	23	S	24	S
UD+12 colony 10	TH20	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	NA	12	I	0	NA	24	S	25	S
UD+20 colony 1	TH21	Erwinia	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	27	S	22	S	28	S
UD+20 colony 2	TH22	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	40	S	40	S	32	S	38	S
UD+20 colony 3	TH23	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	36	S	30	S	23	S	30	S
UD+20 colony 4	TH24	Bacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	37	S	32	S	25	S	29	S
UD+20 colony 5	TH25	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	R	35	S	28	S	24	S	27	S
UD+20 colony 6	TH26	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	R	34	S	32	S	24	S	29	S
UD+20 colony 7	TH27	Bacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	37	S	30	S	24	S	30	S
UD+20 colony 8	TH28	Serratia	0	0	0	0	0	0	0	0	0	0	5	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	25	S	36	S	37	S	26	S	32	S
UD+20 colony 9	TH29	Erwinia	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	27	S	22	S	25	S
UD+20 colony 10	TH30	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	0	R	0	R	14	I	14	R
SD+4 colony 1	TH31	Pseudomonas	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	23	S	22	S
SD+4 colony 2	TH32	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	19	S	16	I
SD+4 colony 3	TH33	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	16	I	18	S
SD+4 colony 4	TH34	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	11	R	17	I
SD+4 colony 5	TH35	Pseudomonas	0	0	1	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	NA	0	R	0	NA	24	S	19	S
SD+4 colony 6	TH36	Pseudomonas	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	16	I
SD+4 colony 7	TH37	Pseudomonas	0	0	1	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	20	S
SD+4 colony 8	TH38	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	20	S
SD+4 colony 9	TH39	Pseudomonas	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	23	S	20	S
SD+4 colony 10	TH40	Pseudomonas	0	0	1	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	20	S
SD+12 colony 1	TH41	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	19	S	11	NA	9	R	14	I
SD+12 colony 2	TH42	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	9	R	14	I
SD+12 colony 3	TH43	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	18	I	16	I
SD+12 colony 4	TH44	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	28	S	24	S	21	S	26	S
SD+12 colony 5	TH45	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	28	S	24	S	20	S	24	S
SD+12 colony 6	TH46	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	15	I	15	I
SD+12 colony 7	TH47	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	15	I	14	I
SD+12 colony 8	TH48	Pseudomonas	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	17	I	11	R
SD+12 colony 9	TH49	Pseudomonas	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	11	R	0	NA	16	I	21	S
SD+12 colony 10	TH50	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	30	S	23	S	25	S
SD+20 colony 1	TH51	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	R	31	S	29	S	28	S	30	S
SD+20 colony 2	TH52	Bacillus	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	24	S	22	S	22	S
SD+20 colony 3	TH53	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	29	S	30	S	20	S	22	S
SD+20 colony 4	TH54	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	S	33	S	34	S	27	S	30	S
SD+20 colony 5	TH55	Bacillus	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	R	32	S	28	S	24	S	15	I
SD+20 colony 6	TH56	Erwinia	0	0	0	0	1	0	0	0	0	0	3	1	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	R	29	S	30	S	24	S	24	S
SD+20 colony 7	TH57	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	23	S	30	S	35	S	24	I	30	S
SD+20 colony 8	TH58	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	14	I	0	R	16	S	16	I
SD+20 colony 9	TH59	Bacillus	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	27	S	22	S	24	S
SD+20 colony 10	TH60	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	32	S	30	NA	25	S	25	S
MD+4 colony 1	TH61	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	12	R	19	S
MD+4 colony 2	TH62	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	18	I	16	I
MD+4 colony 3	TH63	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	18	S
MD+4 colony 4	TH64	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	19	S	15	I
MD+4 colony 5	TH65	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	21	S
MD+4 colony 6	TH66	Pseudomonas	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	20	S
MD+4 colony 7	TH67	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	16	I	15	I
MD+4 colony 8	TH68	Pseudomonas	1	0	0	0	0	0	0	1	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	22	S	25	S
MD+4 colony 9	TH69	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	15	I	15	I
MD+4 colony 10	TH70	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	20	S	16	I
MD+12 colony 1	TH71	Pseudomonas	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	8	R	0	NA	19	S	12	I
MD+12 colony 2	TH72	Pseudomonas	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	22	S	18	S
MD+12 colony 3	TH73	Exiguobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	S	27	S	26	S	21	S	31	S
MD+12 colony 4	TH74	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	11	R	0	NA	21	S	21	S
MD+12 colony 5	TH75	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	9	R	0	NA	20	S	19	S
MD+12 colony 6	TH76	Pseudomonas	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	17	I	14	I
MD+12 colony 7	TH77	Pseudomonas	0	0	0	1	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	14	I	18	S
MD+12 colony 8	TH78	Pseudomonas	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	0	R	0	NA	16	I	13	I
MD+12 colony 9	TH79	Erwinia	0	0	0	0	1	0	0	0	0	0	2	1	1	0	1	0	1	1	0	0	0	0	0	0	1	1	0	9	R	18	I	9	R	16	I	15	I
MD+12 colony 10	TH80	Pseudomonas	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	NA	0	R	0	NA	19	S	15	I
MD+20 colony 1	TH81	Pantonea	0	0	0	0	1	0	1	0	0	0	2	1	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	15	I	22	S	10	R	20	S	21	S
MD+20 colony 2	TH82	Erwinia	0	0	0	0	1	0	0	0	0	0	2	1	1	0	1	0	1	1	0	0	0	0	0	0	1	1	0	8	R	19	S	9	R	17	I	13	I
MD+20 colony 3	TH83	Bacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	29	S	29	S	19	S	19	S
MD+20 colony 4	TH84	Pantonea	0	0	0	0	1	0	0	0	0	0	2	1	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	12	R	23	S	11	R	20	S	23	S
MD+20 colony 5	TH85	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	31	S	30	S	24	S	23	S
MD+20 colony 6	TH86	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	29	S	27	S	16	I	19	S
MD+20 colony 7	TH87	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	29	S	34	S	26	S	21	S
MD+20 colony 8	TH88	Erwinia	0	0	0	0	1	0	0	0	0	0	3	1	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	R	20	S	11	R	14	I	19	S
MD+20 colony 9	TH89	Bacillus	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R	28	S	25	S	18	I	23	S
MD+20 colony 10	TH90	Pseudomonas	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	NA	30	S	25	NA	19	S	18	S
		Sum	1	1	3	1	6	20	1	4	25	40	176	8	7	2	8	35	6	7	1	2	11	1	2	2	2	6	1										

Table A6. Guppy base calling specifications and summary of sequencing statistics per soil sample post quality control.

Sample	Year	Run	Barcode	Flow cell + kit model	Total Yield (Mbp)	Total Read Count	Average Read Length (bp)
MD10.1	2018	FPES_20180719A	8	dna_r9.4.1_450bps_hac	336	107,848	3,111
MD10.2	2018	FPES_20180611	10	dna_r9.5_450bps	237	66,156	3,590
MD10.3	2018	FPES_20180719B	7	dna_r9.4.1_450bps_hac	266	95,477	2,789
MD10.4	2018	FPES_20180719C	6	dna_r9.4.1_450bps_hac	528	199,447	2,650
MD11.1	2018	FPES_20180611	11	dna_r9.5_450bps	340	141,202	2,406
MD11.2	2018	FPES_20180719A	7	dna_r9.4.1_450bps_hac	320	151,923	2,105
MD11.3	2018	FPES_20180719B	5	dna_r9.4.1_450bps_hac	245	87,798	2,785
MD11.4	2018	FPES_20180719C	3	dna_r9.4.1_450bps_hac	465	152,761	3,044
MD12.1	2018	FPES_20180719A	1	dna_r9.4.1_450bps_hac	242	104,932	2,310
MD12.2	2018	FPES_20180611	12	dna_r9.5_450bps	187	87,766	2,133
MD12.3	2018	FPES_20180719B	4	dna_r9.4.1_450bps_hac	237	134,994	1,759
MD12.4	2018	FPES_20180719C	2	dna_r9.4.1_450bps_hac	627	198,765	3,154
MD9.1	2018	FPES_20180719A	5	dna_r9.4.1_450bps_hac	400	194,937	2,053
MD9.2	2018	FPES_20180719B	6	dna_r9.4.1_450bps_hac	309	107,312	2,883
MD9.3	2018	FPES_20180611	9	dna_r9.5_450bps	203	93,380	2,179
MD9.4	2018	FPES_20180719C	5	dna_r9.4.1_450bps_hac	374	140,143	2,667
SD10.1	2018	FPES_20180719A	3	dna_r9.4.1_450bps_hac	435	203,659	2,138
SD10.2	2018	FPES_20180719B	3	dna_r9.4.1_450bps_hac	309	136,862	2,254
SD10.3	2018	FPES_20180611	6	dna_r9.5_450bps	305	101,986	2,986
SD10.4	2018	FPES_20180719C	12	dna_r9.4.1_450bps_hac	465	140,711	3,308
SD11.1	2018	FPES_20180719A	2	dna_r9.4.1_450bps_hac	270	129,469	2,083
SD11.2	2018	FPES_20180611	7	dna_r9.5_450bps	183	53,543	3,418
SD11.3	2018	FPES_20180719B	12	dna_r9.4.1_450bps_hac	264	100,301	2,633
SD11.4	2018	FPES_20180719C	10	dna_r9.4.1_450bps_hac	652	220,446	2,957
SD12.1	2018	FPES_20180719A	6	dna_r9.4.1_450bps_hac	523	263,679	1,984
SD12.2	2018	FPES_20180611	8	dna_r9.5_450bps	346	169,926	2,036
SD12.3	2018	FPES_20180719B	2	dna_r9.4.1_450bps_hac	487	246,027	1,979
SD12.4	2018	FPES_20180719C	8	dna_r9.4.1_450bps_hac	700	184,567	3,792
SD9.1	2018	FPES_20180719A	4	dna_r9.4.1_450bps_hac	684	392,010	1,744
SD9.2	2018	FPES_20180719B	1	dna_r9.4.1_450bps_hac	400	173,176	2,311
SD9.3	2018	FPES_20180719C	7	dna_r9.4.1_450bps_hac	737	299,103	2,464
SD9.4	2018	FPES_20180611	5	dna_r9.5_450bps	204	77,269	2,635
UD10.1	2018	FPES_20180719A	10	dna_r9.4.1_450bps_hac	385	124,052	3,106
UD10.2	2018	FPES_20180611	2	dna_r9.5_450bps	276	78,460	3,516
UD10.3	2018	FPES_20180719B	8	dna_r9.4.1_450bps_hac	297	106,709	2,782
UD10.4	2018	FPES_20180719C	4	dna_r9.4.1_450bps_hac	556	178,135	3,122
UD11.1	2018	FPES_20180611	3	dna_r9.5_450bps	295	97,759	3,020
UD11.2	2018	FPES_20180719A	11	dna_r9.4.1_450bps_hac	345	116,453	2,961

UD11.3	2018	FPES_20180719B	11	dna_r9.4.1_450bps_hac	207	101,380	2,038
UD11.4	2018	FPES_20180719C	11	dna_r9.4.1_450bps_hac	516	135,427	3,811
UD12.1	2018	FPES_20180719A	12	dna_r9.4.1_450bps_hac	272	108,655	2,504
UD12.2	2018	FPES_20180611	4	dna_r9.5_450bps	317	90,290	3,506
UD12.3	2018	FPES_20180719B	10	dna_r9.4.1_450bps_hac	294	121,689	2,413
UD12.4	2018	FPES_20180719C	9	dna_r9.4.1_450bps_hac	546	207,151	2,638
UD9.1	2018	FPES_20180611	1	dna_r9.5_450bps	235	80,943	2,904
UD9.2	2018	FPES_20180719A	9	dna_r9.4.1_450bps_hac	336	122,957	2,731
UD9.3	2018	FPES_20180719B	9	dna_r9.4.1_450bps_hac	316	134,144	2,359
UD9.4	2018	FPES_20180719C	1	dna_r9.4.1_450bps_hac	451	147,992	3,045
MD10.1	2019	FPES_20190606A	1	dna_r9.4.1_450bps_hac	267	87,093	3,062
MD10.2	2019	FPES_20190606A	4	dna_r9.4.1_450bps_hac	264	141,507	1,863
MD10.3	2019	FPES_20190606A	10	dna_r9.4.1_450bps_hac	391	168,414	2,323
MD10.4	2019	FPES_20190605B	2	dna_r9.4.1_450bps_hac	932	403,239	2,311
MD11.1	2019	FPES_20190606B	6	dna_r9.4.1_450bps_hac	303	74,470	4,071
MD11.2	2019	FPES_20190606A	6	dna_r9.4.1_450bps_hac	368	143,732	2,563
MD11.3	2019	FPES_20190605B	7	dna_r9.4.1_450bps_hac	702	264,676	2,652
MD11.4	2019	FPES_20190605A	2	dna_r9.4.1_450bps_hac	314	155,202	2,022
MD12.1	2019	FPES_20190606B	3	dna_r9.4.1_450bps_hac	251	94,902	2,643
MD12.2	2019	FPES_20190605A	5	dna_r9.4.1_450bps_hac	339	130,163	2,606
MD12.3	2019	FPES_20190605A	11	dna_r9.4.1_450bps_hac	12	5,116	2,398
MD12.4	2019	FPES_20190605A	8	dna_r9.4.1_450bps_hac	263	114,914	2,290
MD9.1	2019	FPES_20190605B	3	dna_r9.4.1_450bps_hac	657	243,294	2,700
MD9.2	2019	FPES_20190606B	1	dna_r9.4.1_450bps_hac	278	97,341	2,856
MD9.3	2019	FPES_20190606B	11	dna_r9.4.1_450bps_hac	326	101,212	3,223
MD9.4	2019	FPES_20190605B	10	dna_r9.4.1_450bps_hac	321	126,031	2,546
SD10.1	2019	FPES_20190605B	11	dna_r9.4.1_450bps_hac	393	204,242	1,925
SD10.2	2019	FPES_20190606A	8	dna_r9.4.1_450bps_hac	330	157,270	2,099
SD10.3	2019	FPES_20190606B	7	dna_r9.4.1_450bps_hac	215	79,168	2,719
SD10.4	2019	FPES_20190606B	2	dna_r9.4.1_450bps_hac	333	186,932	1,782
SD11.1	2019	FPES_20190605A	9	dna_r9.4.1_450bps_hac	408	198,591	2,054
SD11.2	2019	FPES_20190606A	2	dna_r9.4.1_450bps_hac	378	192,900	1,961
SD11.3	2019	FPES_20190605A	3	dna_r9.4.1_450bps_hac	405	194,850	2,079
SD11.4	2019	FPES_20190606B	9	dna_r9.4.1_450bps_hac	175	74,331	2,353
SD12.1	2019	FPES_20190605A	10	dna_r9.4.1_450bps_hac	924	445,339	2,075
SD12.2	2019	FPES_20190606A	5	dna_r9.4.1_450bps_hac	278	132,892	2,089
SD12.3	2019	FPES_20190605B	5	dna_r9.4.1_450bps_hac	622	352,394	1,766
SD12.4	2019	FPES_20190606B	5	dna_r9.4.1_450bps_hac	337	171,481	1,962
SD9.1	2019	FPES_20190606B	8	dna_r9.4.1_450bps_hac	233	89,440	2,608
SD9.2	2019	FPES_20190605A	6	dna_r9.4.1_450bps_hac	105	51,196	2,043
SD9.3	2019	FPES_20190606A	11	dna_r9.4.1_450bps_hac	186	60,142	3,089
SD9.4	2019	FPES_20190605B	1	dna_r9.4.1_450bps_hac	711	376,037	1,891
UD10.1	2019	FPES_20190605B	4	dna_r9.4.1_450bps_hac	633	257,532	2,457

UD10.2	2019	FPES_20190606A	3	dna_r9.4.1_450bps_hac	134	49,394	2,721
UD10.3	2019	FPES_20190605A	1	dna_r9.4.1_450bps_hac	471	231,151	2,037
UD10.4	2019	FPES_20190606A	7	dna_r9.4.1_450bps_hac	119	43,263	2,748
UD11.1	2019	FPES_20190606B	12	dna_r9.4.1_450bps_hac	127	68,557	1,858
UD11.2	2019	FPES_20190606A	9	dna_r9.4.1_450bps_hac	202	117,338	1,721
UD11.3	2019	FPES_20190605B	6	dna_r9.4.1_450bps_hac	641	509,983	1,256
UD11.4	2019	FPES_20190606A	12	dna_r9.4.1_450bps_hac	175	75,728	2,315
UD12.1	2019	FPES_20190605B	12	dna_r9.4.1_450bps_hac	429	224,621	1,909
UD12.2	2019	FPES_20190606B	4	dna_r9.4.1_450bps_hac	260	100,894	2,574
UD12.3	2019	FPES_20190606B	10	dna_r9.4.1_450bps_hac	239	122,560	1,950
UD12.4	2019	FPES_20190605A	12	dna_r9.4.1_450bps_hac	209	99,385	2,099
UD9.1	2019	FPES_20190605A	7	dna_r9.4.1_450bps_hac	481	183,056	2,627
UD9.2	2019	FPES_20190605A	4	dna_r9.4.1_450bps_hac	669	320,815	2,085
UD9.3	2019	FPES_20190605B	9	dna_r9.4.1_450bps_hac	329	174,409	1,886
UD9.4	2019	FPES_20190605B	8	dna_r9.4.1_450bps_hac	443	255,906	1,731

Table A7. Illumina HiSeq sample barcodes and sequencing statistics per soil core for de-multiplexed HiC and Nextera samples

Sample ID	Barcode Sequence	Total Read Count	Yield (Mbp)	Mean Quality Score	% Bases >= 30
hic01	ATCCACGA+GGCATACT	118,219,457	35,466	37.16	87.94
hic02	GAGCGCCA+GGAGAGTT	207,701,849	62,311	38.33	92.03
hic03	GCGCGGTG+GGTCGGGT	129,125,923	38,737	37.84	90.31
hic04	ACGACAGA+ACCAGACT	195,349,364	58,604	38.23	91.58
hic05	TAATGATG+TTCGATAC	173,253,822	51,977	38.2	91.41
hic06	GTCCTAAG+GAGCAGTA	168,486,855	50,547	38.06	90.96
nextera01	CGAGGCTG+TATGCAGT	240,648,488	72,194	37.95	90.66
nextera02	GCTCATGA+CTCCTTAC	216,182,280	64,855	37.71	89.78
nextera03	AAGAGGCA+CTCCTTAC	199,818,148	59,946	37.54	89.21
nextera04	CTCTCTAC+TCTTACGC	209,549,167	62,864	37.71	89.81
nextera05	GGACTCCT+AGAGGATA	284,952,290	85,486	38.44	92.42
nextera06	TAGGCATG+TATGCAGT	256,584,426	76,975	38.17	91.46
	Sum	***2,399,872,069***	***719,962***		

Mean	*# reads*	*yield*	*Q score*	*% Bases >= 30*
HiC	1.65E+08	49,607	37.97	90.71
Nextera	2.35E+08	70,387	37.92	90.56

Table A8. Assembly statistics for Megahit assemblies of Nextera reads size selected for contigs greater than 1,500 bp.

	assembly sum (bp)	sum, contig #	ave contig length (bp)	largest contig (bp)	N50	N50, contig #
MD01	1519381163	511921	2968	417435	2957	128223
MD02	1251434727	422855	2959.49	367604	2876	102380
SD03	1345612131	432225	3113.22	328434	3071	97256
SD04	1869385081	486892	3839.42	693351	4533	82876
UD06	2263788991	630802	3588.75	605266	4005	118792
UD06	2025514159	568495	3562.94	331885	3969	107649
Mean	**1712519375**	**508865**	**3338.63**	**457329.16**	**3568.5**	**106196**
stdev	402738903	80234.56	372.98	154650.55	690.24	16026.64

Table A9. Number of integrons and integron cassette types identified in Hi-C MAGs used for tree construction.

MAG	Phylum	Number of contigs with integrons	Integron types
MD01_bin_12	Proteobacteria	2	1
MD02_bin_13	Proteobacteria	2	1
MD02_bin_14	Proteobacteria	5	1
MD02_bin_15	Proteobacteria	2	1
MD02_bin_20	Proteobacteria	5	1
MD02_bin_21	Verrucomicrobia	2	1
MD02_bin_23	Proteobacteria	1	1
MD02_bin_29	Proteobacteria	1	1
MD02_bin_3	Proteobacteria	3	1
MD02_bin_37	Proteobacteria	5	1
MD02_bin_39	Proteobacteria	1	1
SD04_bin_20	Verrucomicrobia	11	1
SD04_bin_44	Proteobacteria	2	1
SD04_bin_88	Proteobacteria	3	1
UD06_bin_57	Proteobacteria	1	1
UD06_bin_70	Verrucomicrobia	6	1
UD06_bin_78	Acidobacteria	1	1

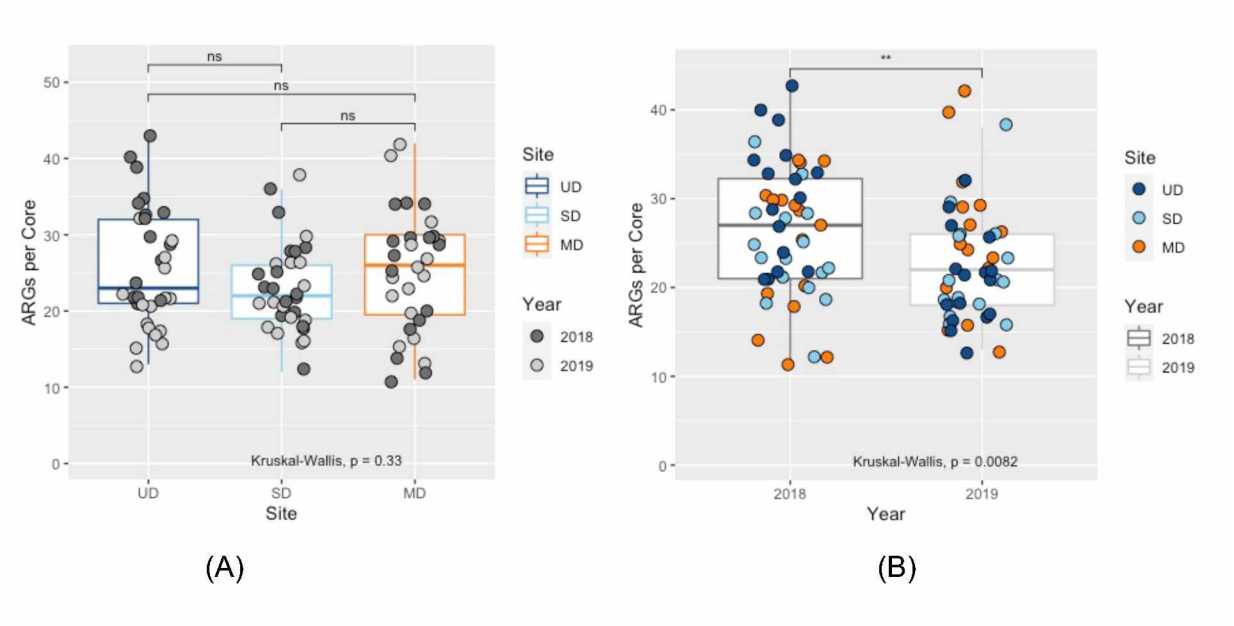


Figure A1. A) Boxplot of the number of ARGs per soil core. Kruskal-Wallis *P* value is shown in addition to Wilcoxon test results between means (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$). B) Boxplot displaying the number of ARGs per metagenome year collected (Kruskal-Wallis $p = 0.0082$) with points representing a soil core color-coded by FPES treatment. Wilcoxon test between group significance $p < 0.01$ **, $p < 0.05$ *, ns >0.1.

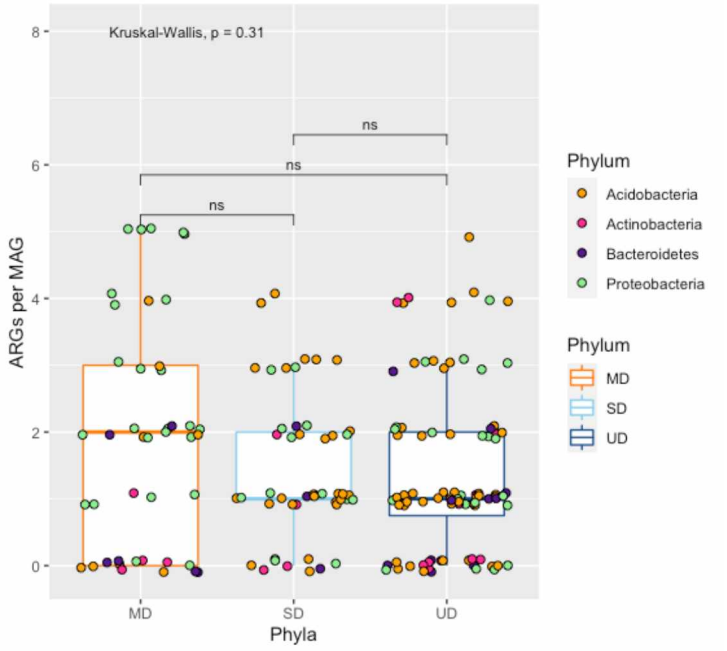


Figure A2. Number of ARGs per MAG used in phylogenetic tree not significantly different between sites. Boxplot of ARG per MAG by site. Points are colored by bacterial phyla assigned to MAG. Kruskal-Wallis *P* value is shown in addition to Wilcoxon test results between means (ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$).

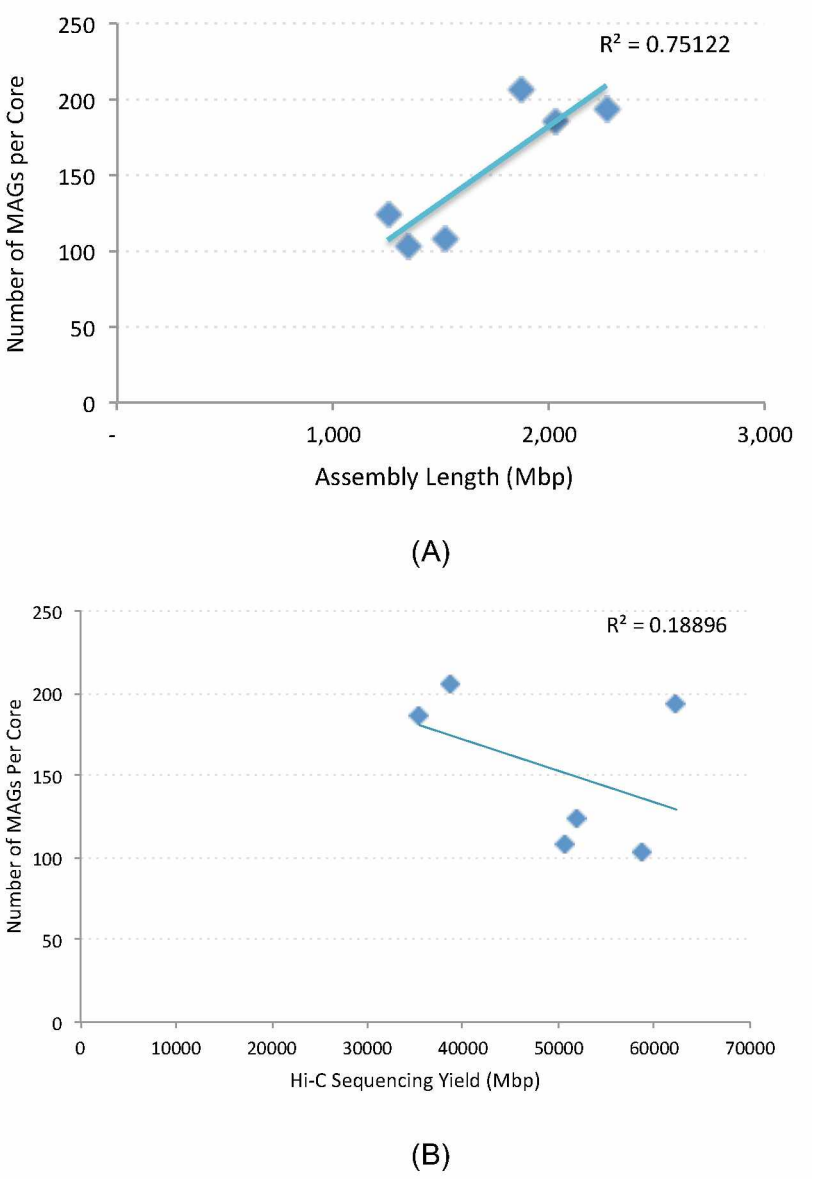


Figure A3. Scatterplot depicting the relationship between A) Megahit assembly length (Mbp) and number of MAGs per each soil core and B) Hi-C library sequencing yield (Mbp) and number of MAGs per soil core. R-squared value is displayed on the top right corner of the plot.

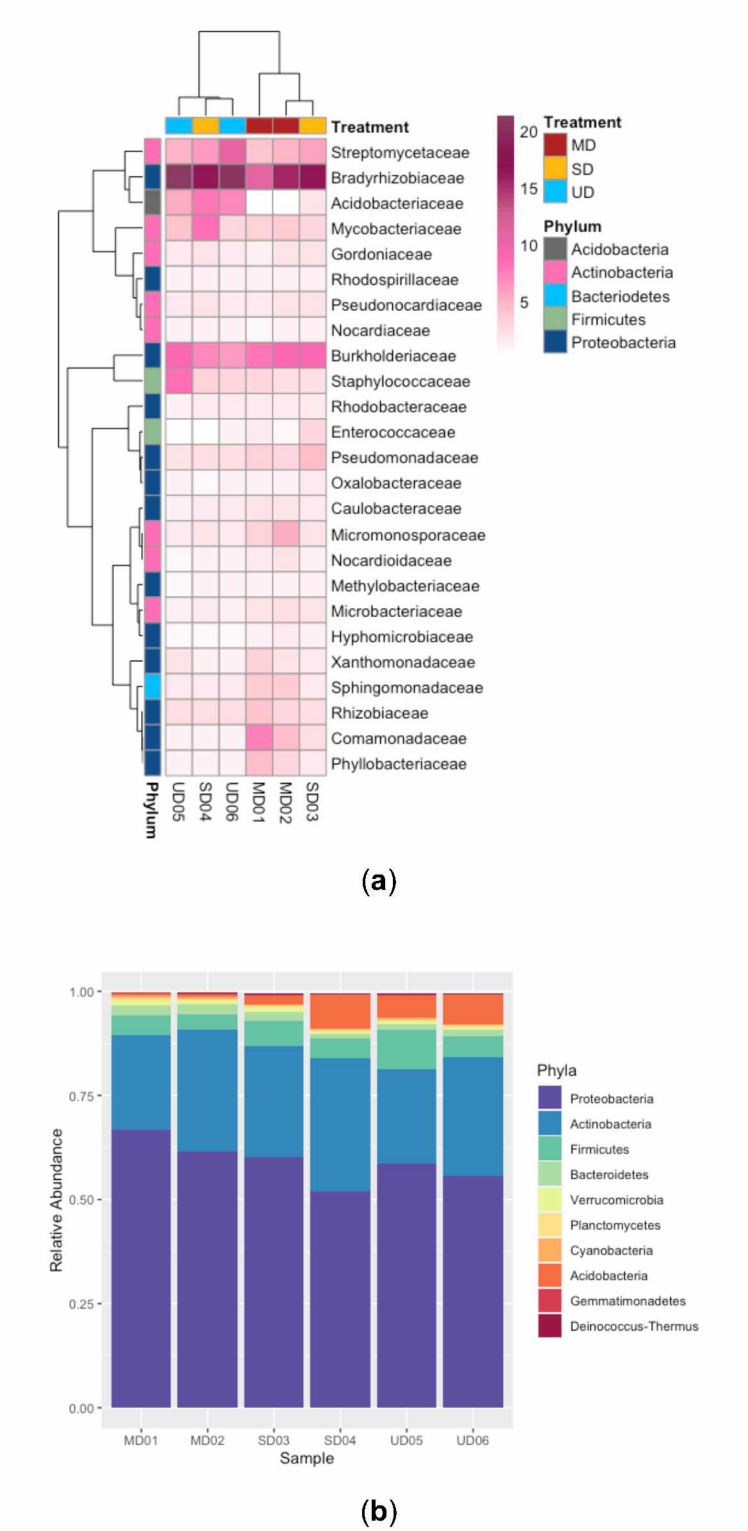


Figure A4. A) Bracken heatmap of top 25 most abundant families across the six soil cores using the 150 bp Illumina shotgun reads as input. Colors on rows indicate which FPES treatment each soil core is from and colors on columns indicate which phylum each family belong to. B) Relative abundance of the top ten most abundant bacterial phyla across soil cores based on bracken estimation

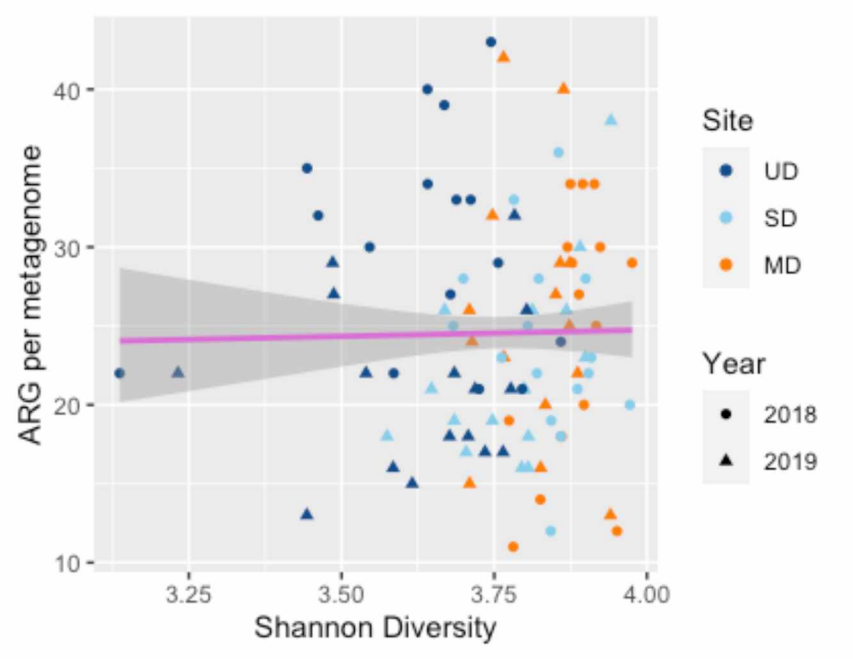


Figure A5. Plots showing the effects of Shannon-Weiner diversity on ARG abundance with pink GLM regression lines, points colored by FPES treatment, and points shape based on soil collection year.

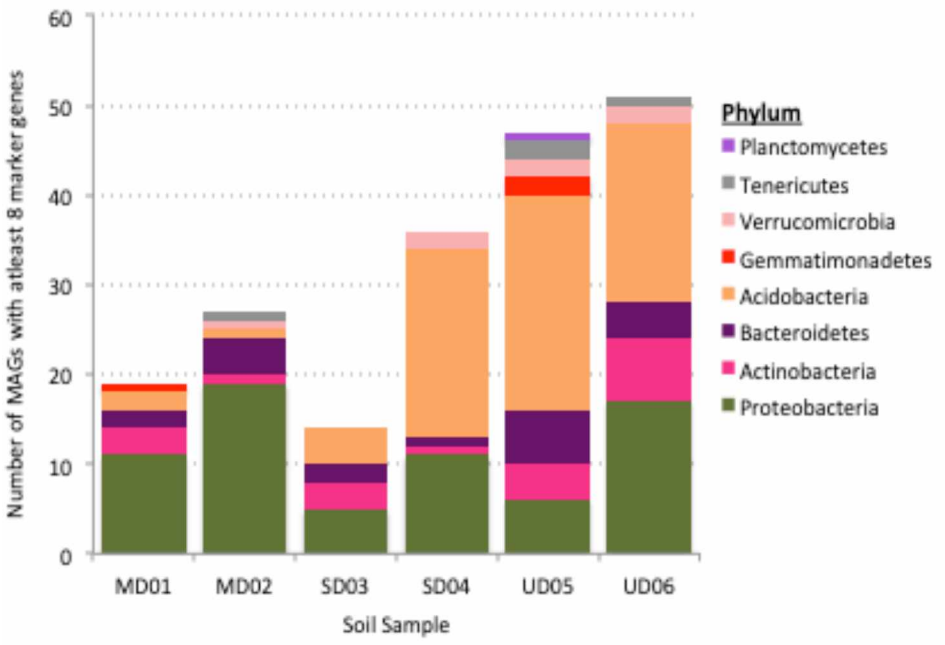


Figure A6. Number of MAGs in phylogenetic tree by phyla across each soil core.

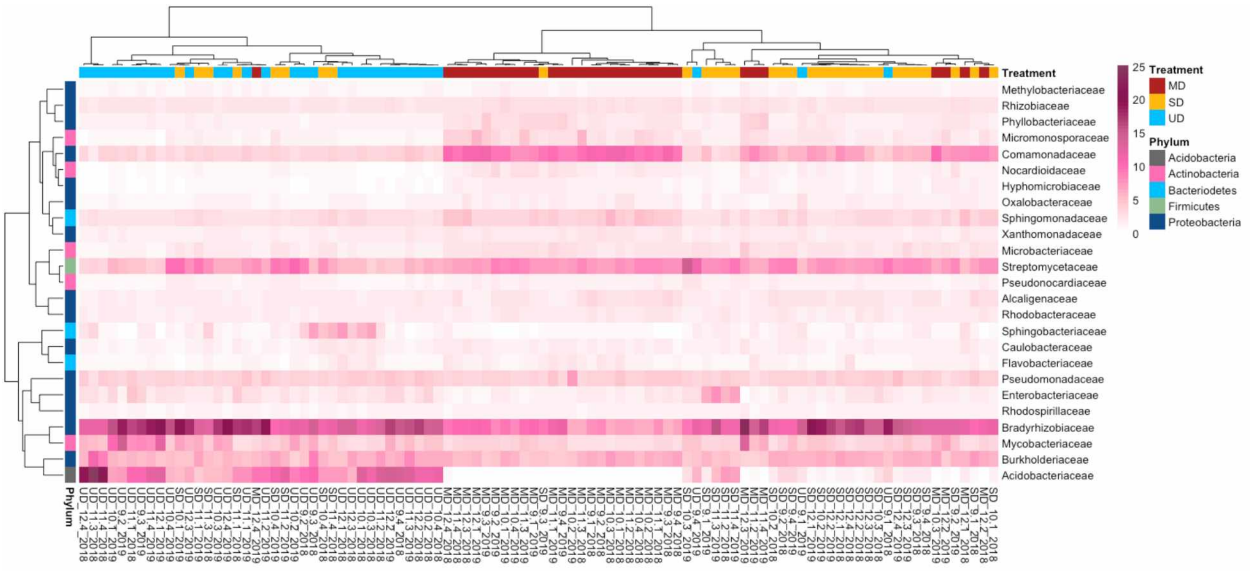


Figure A7. Bracken estimated abundance at a family level of the top 25 families in the MinION metagenomic dataset with groupings of soil cores and bacterial families based on Pearson correlation.