Microbial Diversity and Antimicrobial Resistance Genes in Land Applied Manure

by

Emily M. Crossette

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Environmental Engineering)
in The University of Michigan
2021

Doctoral Committee:

       Professor Lutgarde Raskin, Co-Chair
       Associate Professor Krista R. Wigginton, Co-Chair
       Assistant Professor Melissa B. Duhaime
       Associate Professor Indika Rajapakse

Emily M. Crossette

emcrosse@umich.edu

ORCID iD: 0000-0003-3645-2909

DEDICATION

To the hard-working, talented, and dedicated teachers in the Abington School District.

ACKNOWLEDGEMENTS

iv

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

ABSTRACT

Antimicrobial resistant infections pose a continued threat to human and animal health globally. This work critically evaluated antimicrobial resistance genes (ARGs) present in manure to better quantify and assess the potential human and animal health risks associated with exposure to manure. We first present a systematic review and meta-analysis that identified trends in ARG abundances in cattle manures. We found that the lack of biological replication and insufficient overlap of gene targets between studies prohibited quantitative cross-study comparisons. The meta-analysis identified important methodological gaps and will inform the design of future studies that more effectively and rigorously evaluate ARG abundances in manure.

To address the low-throughput limitation of quantitative PCR (qPCR) approaches to quantify gene concentrations, we developed a novel metagenomic gene quantification approach that is achieved via spike-in standards. We found this approach performed comparably to qPCR when applied to specific ARGs yet facilitates the absolute quantification of all ARG concentrations in metagenomes. We then applied this quantitative metagenomic approach to better identify how ARG abundance and microbial community structure vary in different stored manures. With a cross section of manure collected from dairy farms during land application, we found anaerobic digestion treatments correlated with organism and ARG concentrations. This demonstrated that manure management, including anaerobic digestion, shapes manure microbial

communities. Finally, to address the complexity of microbial interactions in soil after land

application of manures, we introduce a novel mathematical modeling approach. We demonstrate

how the approach can extract microbial network dynamics from time-series observations of

microbial communities and outline how the approach can be used to quantify the impact of

disturbances, like land application, on microbial community dynamics.

This dissertation contributes important new tools to better quantify risks of ARG

pollution and dissemination in the environment and clarifies the role of manure management in

mitigating ARG abundances. Ultimately this work will better inform policy for manure

management and land application to reduce the risk of ARG dissemination in the environment

and consequently reduce the human and veterinary cost of antimicrobial resistance.

Chapter 1

Introduction

1.1 Antimicrobial resistance genes: an emerging contaminant in livestock manures

      Following the discovery and commercialization of antimicrobials, life expectancy at birth increased by 5-10 years in the United States as the leading causes of death shifted from communicable diseases to degenerative diseases such as heart disease and cancer (1). However, as the use of these drugs became widespread, so too did the prevalence of antimicrobial resistance (2). Microorganisms developed resistance to antimicrobials through both genetic mutations and the transfer of genes from proximate microbes.

      In addition to treating microbial infections in humans, antimicrobials are used widely in livestock rearing. Over 11.5 million kilograms of antimicrobials are purchased annually by the American livestock industry (3). The widespread use of antimicrobials by this industry poses risks to human health. For example, while methicillin-resistant *Staphylococcus aureus* (MRSA) was initially thought to only originate in human health care settings, community spread of MRSA has been observed since the 1990s. Soon thereafter, livestock-associated multi-drug resistant strains of MRSA were identified and found to be prevalent in livestock-exposed human populations such as farmers (4). As a result of observations like this, livestock operations have been under scrutiny as potential hot spots for the emergence and spread of antimicrobial resistant bacteria (ARB).

Research on ARB and antimicrobial resistance genes (ARGs) in environmental systems has skyrocketed over the last decade. For example, between the years 2009-2016, the USDA's National Institute of Food and Agriculture (NIFA) Agriculture and Food Research Initiative (AFRI) awarded over $82 million USD in food safety research specifically focused on ARGs and ARB in agricultural environments (5). Despite the extensive number of research projects, a clear consensus on the role of manure management (e.g., anaerobic digestion, storage, and land application) on mitigating ARG pollution remains elusive (6–10). To identify limitations that are confounding the understanding of ARG fate through manure management systems, Chapter 2 of this dissertation qualitatively and quantitatively summarizes literature in a systematic review and meta-analysis of ARGs quantified from field-scale studies.

1.2 Methods for detecting and quantifying ARG phenotypes and genotypes

Methods for studying ARGs and ARB in environmental settings can be broadly classified into phenotypic and genotypic approaches (Figure 1). Phenotypic approaches observe the expression of the resistance trait (phenotype) in an organism, while genotypic approaches identify gene sequences known to confer resistance. Since phenotypic approaches rely on the expression of ARGs, these approaches are culture dependent. Although culture-dependent approaches, such as susceptibility testing, have advanced knowledge of ARG and ARB prevalence in farm settings (11, 12), the approach is limited to culturable ARB. Specifically, susceptibility testing typically targets genera or species of organisms with clinical relevance such as *Campylobacter* or *Escherichia coli* (11).

Genotypic approaches are culture independent and are not biased to only include culturable, targeted organisms. PCR-based tools such as quantitative PCR (qPCR) identify ARGs using primer and probe sequences. qPCR assays have advanced the understanding of ARG fate in the environment; however the approach only targets a few genes per assay. Metagenomic

2

sequencing facilitates simultaneous screening of all documented resistant genotypes but is

limited to providing abundances relative to universal housekeeping genes, including 16S rRNA

genes, rather than gene concentrations in units of gene copies per sample mass or volume.

Chapter 3 of this dissertation presents a novel quantitative, gene-level metagenomic approach

that facilitates quantifying the concentration of all documented ARGs in complex microbiomes

to overcome methodological barriers of current genotypic approaches.
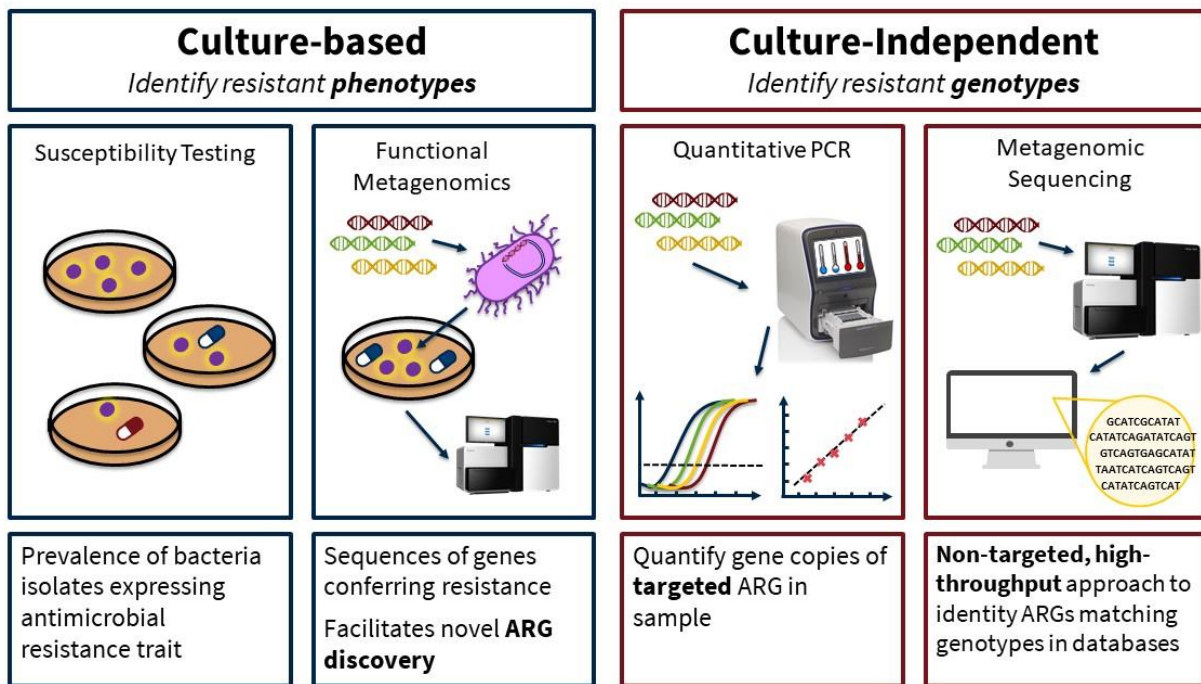
Figure 1-1: *Summary of tools for detecting or quantifying ARG and ARB phenotypes and genotypes*

1.3 Farm management practices shaping ARG occurrence in land applied manure

The USDA estimates that U.S. livestock annually produces over 500 million tons (wet

weight) of manure to be land applied (13). It is therefore critical to understand how different

manure management systems impact the prevalence and abundance of ARGs before the manure

is land applied. Both farm characteristics and manure handling can shape microbial communities

and ARG abundances (14). Some farms employ physical and biological processes such as solid-liquid separation, pasteurization, and anaerobic digestion (15).

Several recent literature reviews have noted the limited number of field-scale studies on anaerobic digestion for ARB and ARG mitigation in manure systems (16–18). One lab-scale study tracked 10 ARGs through anaerobic digesters with qPCR and found gene concentrations were stable (19). Another bench-scale study using metagenomic sequencing found that multi-drug, peptide, tetracycline, and aminoglycoside resistance genes dominated resistance profiles before and after anaerobic digestion. Comparing four biological replicates, the study demonstrated how variability in manure microbiomes between seasons and farms can lead to different outcomes of anaerobic digestion on ARG abundances (20). Wallace *et al.* studied seasonal variation in ARGs through a full-scale advanced anaerobic digestion system with pre-digestion pasteurization and found that the abundances of two tetracycline ARGs did not differ significantly between raw and digested manures, but the study found small reductions in the two sulfonamide ARGs targeted (21). These studies have demonstrated many ARGs persist during anaerobic digestion but demonstrated inconsistent observations in the role of digestion on ARG abundances both within and across studies.

Anaerobic digestion is not the final barrier for manure prior to land application; manure is typically stored in earthen or concrete pits or lagoons. The relative roles of storage and anaerobic digestion in mitigating ARGs are not known. Chapter 4 of this dissertation explores the impact of anaerobic digestion and storage on dairy manure microbial communities and ARGs in a cross-section of dairy farms in the United States.

1.4 Modeling and predicting microbial interactions in manure-amended soils

Following storage, manures are land applied to restore organic carbon, nitrogen, and phosphorous to crop lands. Some field-scale studies have shown that the use of manure and

wastewater biosolids amendments have enriched soil in antimicrobial resistance (22, 23). Other studies, however, have found that over time, soil ARG levels return to pre-land application levels (24, 25). These conflicting observations highlight the complexity of interactions between soil and manure microbiomes.

The public health impacts of ARGs in land applied manures are not well understood (13). There is concern that pathogens residing in soil can acquire ARGs from manure microbiomes through horizontal gene transfer (HGT). The newly resistant pathogens could then contaminate food and water (26–28). Soil harbors a diverse and dynamic microbiome containing naturally occurring ARGs (29). Early modeling work demonstrated that elevated numbers of ARB increase the probability of HGT events (30). However, it is still not possible to monitor or detect HGT in soil environments (31) so the prevalence of HGT in agricultural soil microbial communities is not known.

Quantifying the rate of HGT is critical for assessing risk of gene transfer events in the environment. To combat the challenge of *in situ* HGT detection, a number of systems biology and mathematical modeling concepts have been proposed using mass-action (30, 32), agent-based (33), or game theory frameworks (34). However, these models simulate microbial communities or populations and are not based on empirical evidence, which limits their practical application for risk management. Data-driven models, which infer system dynamics from time series observation of real microbiomes, are needed to assess risks of HGT to understand risk of pathogens acquiring ARGs and inform policies for land application. Chapter 5 of this dissertation outlines a data-driven mathematical framework for quantifying and predicting HGT events.

1.5 Dissertation summary

Quantifying the fate of AMR gene released into the environment is imperative to inform necessary mitigation strategies. The results from this dissertation will provide tools for

understanding the scale of ARGs in different environments, identify remaining knowledge gaps

of manure management's role in ARG mitigation, and better quantify risks after ARGs and ARB

are released via land application.

1.6 References

1. Adedeji WA. 2016. The treasure called antibiotics. Ann Ibadan Postgrad Med 14:56.

2. Clatworthy AE, Pierson E, Hung DT. 2007. Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol 3:541–548.

3. FDA. 2019. Summary Report on Antimicrobials Sold or Distributed for Use in Food-Producing Animals.

4. Chen C, Wu F. 2020. Livestock-associated methicillin-resistant Staphylococcus aureus (LA-MRSA) colonisation and infection among livestock workers and veterinarians: a systematic review and meta-analysis. Occup Environ Med.

5. USDA. 2016. USDA Announces $6 Million in Available Funding for Antimicrobial Resistance Research.

6. Vikesland PJ, Pruden A, Alvarez PJJ, Aga D, Burgmann H, Li X, Manaia CM, Nambi I, Wigginton K, Zhang T, Others. 2017. Toward a comprehensive strategy to mitigate dissemination of environmental sources of antibiotic resistance. ACS Publications.

7. Manaia CM. 2017. Assessing the risk of antibiotic resistance transmission from the environment to humans: non-direct proportionality between abundance and risk. Trends Microbiol 25:173–181.

8. Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F, Bürgmann H, Sørum H, Norström M, Pons M-N, others. 2015. Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol 13:310–317.

9. Pepper IL, Brooks JP, Gerba CP. 2018. Antibiotic resistant bacteria in municipal wastes: is there reason for concern? Environ Sci Technol 52:3949–3959.

10. Huijbers PMC, Blaak H, De Jong MCM, Graat EAM, Vandenbroucke-Grauls CMJE, De Roda Husman AM. 2015. Role of the Environment in the Transmission of Antimicrobial Resistance to Humans: A Review. Environ Sci Technol 49:11993–12004.

11. Young I, Rajić A, Wilhelm BJ, Waddell L, Parker S, McEwen SA. 2009. Comparison of the prevalence of bacterial enteropathogens, potentially zoonotic bacteria and bacterial resistance to antimicrobials in organic and conventional poultry, swine and beef production: a systematic review and meta-analysis. Epidemiol Infect 137:1217–1232.

12. Tran DTQ, Bradbury MI, Ogtrop FFVAN, Bozkurt H, Jones BJ, McCONCHIE R. 2020. Environmental Drivers for Persistence of Escherichia coli and Salmonella in Manure-Amended Soils: A Meta-Analysis. J Food Prot 83:1268–1277.

13. Gollehon NR, Kellogg RL, Moffitt DC. 2016. Estimates of Recoverable and Non-Recoverable Manure Nutrients Based on the Census of Agriculture - 2012 Results.

14. Hurst JJ, Oliver JP, Schueler J, Gooch C, Lansing S, Crossette E, Wigginton K, Raskin L, Aga DS, Sassoubre LM. 2019. Trends in Antimicrobial Resistance Genes in Manure Blend Pits and Long-Term Storage Across Dairy Farms with Comparisons to Antimicrobial Usage and Residual Concentrations. Environ Sci Technol 53:2405–2415.

15. Oliver JP, Gooch CA, Lansing S, Schueler J, Hurst JJ, Sassoubre L, Crossette EM, Aga DS. 2020. Invited review: Fate of antibiotic residues, antibiotic-resistant bacteria, and antibiotic resistance genes in US dairy manure management systems. J Dairy Sci 103:1051–1071.

16. Congilosi JL, Aga D. 2020. Review on the fate of antimicrobials, antimicrobial resistance genes, and other micropollutants in manure during enhanced anaerobic digestion and composting. J Hazard Mater 123634.

17. Gurmessa B, Cardelli V, Cocco S, Corti G, Pedretti EF. 2020. Manure anaerobic digestion effects and the role of pre- and post-treatments on veterinary antibiotics and antibiotic resistance genes removal efficiency. Sci Total Environ 721.

18. Ghirardini A, Grillini V, Verlicchi P. 2020. A review of the occurrence of selected micropollutants and microorganisms in different raw and treated manure - Environmental risk due to antibiotics after application to soil. Sci Total Environ 707.

19. Sun W, Qian X, Gu J, Wang X-J, Duan M-L. 2016. Mechanism and Effect of Temperature on Variations in Antibiotic Resistance Genes during Anaerobic Digestion of Dairy Manure. Sci Rep 6.

20. Flores-Orozco D, Patidar R, Levin DB, Sparling R, Kumar A, Çiçek N. 2020. Effect of mesophilic anaerobic digestion on the resistome profile of dairy manure. Bioresour Technol 315:123889.

21. Wallace JS, Garner E, Pruden A, Aga DS. 2018. Occurrence and transformation of veterinary antibiotics and antibiotic resistance genes in dairy manure treated by advanced anaerobic digestion and conventional treatment methods. Environ Pollut 236:764–772.

22. Ghosh S, Lapara TM. 2007. The effects of subtherapeutic antibiotic use in farm animals on the proliferation and persistence of antibiotic resistance among soil bacteria. ISME J 1:191–203.

23. Udikovic-Kolic N, Wichmann F, Broderick NA, Handelsman J. 2014. Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. Proc Natl Acad Sci U S A 111:15202–7.

24. Marti R, Tien Y-C, Murray R, Scott A, Sabourin L, Topp E. 2014. Safely coupling livestock and crop production systems: how rapidly do antibiotic resistance genes dissipate in soil following a commercial application of swine or dairy manure? Appl Environ Microbiol 80:3258–65.

25. Miller E, Spiehs M, Arthur T, Woodbury B, Cortus E, Chatterjee A, Rahman S, Schmidt JW. 2019. Cropland Amendment with beef cattle manure minimally affects antimicrobial resistance. J Environ Qual 48:1683–1693.

26. Yeom J-R, Yoon S-U, Kim C-G. 2017. Quantification of residual antibiotics in cow manure being spread over agricultural land and assessment of their behavioral effects on antibiotic resistant bacteria. Chemosphere 182:771–780.

27. Li X, Atwill ER, Antaki E, Applegate O, Bergamaschi B, Bond RF, Chase J, Ransom KM, Samuels W, Watanabe N, others. 2015. Fecal indicator and pathogenic bacteria and their antibiotic resistance in alluvial groundwater of an irrigated agricultural region with dairies. J Environ Qual 44:1435–1447.

28. Xie W-Y, Shen Q, Zhao FJ. 2018. Antibiotics and antibiotic resistance from animal manures to soil: a review. Eur J Soil Sci 69:181–195.

29. Forsberg KJ, Reyes A, Wang B, Selleck EM, Morten OA. 2014. The shared antibiotic resistome of soil bacteria and human pathogens 337:1107–1111.

30. Levin BR, Stewart FM. 1980. The population biology of bacterial plasmids: a priori conditions for the existence of mobilizable nonconjugative factors. Genetics 94:425–443.

31. Maheshwari M, Abulreesh H, Khan MS, Ahmad I, Pichtel J. 2017. Horizontal Gene Transfer in Soil and the Rhizosphere: Impact on Ecological Fitness of Bacteria, p. 111–130. In Agriculturally Important Microbes for Sustainable Agriculture. Springer Nature Singapore Pte Ltd, Singapore.

32. Novozhilov AS, Karev GP, Koonin E V. 2005. Mathematical modeling of evolution of horizontally transferred genes. Mol Biol Evol 22:1721–1732.

33. Juhász J, Kertész-Farkas A, Szabó D, Pongor S. 2014. Emergence of collective territorial defense in bacterial communities: Horizontal gene transfer can stabilize microbiomes. PLoS One 9:1–9.

34. Fan Y, Xiao Y, Momeni B, Liu Y-Y. 2018. Horizontal gene transfer can help maintain the equilibrium of microbial communities. J Theor Biol 454:53–59.

Chapter 2

Antimicrobial Resistance Gene Concentrations in Cattle Manure Management Systems: A Systematic Review and Meta-Analysis of Field-Scale Studies

ABSTRACT

We summarize the last 15 years of studies focused on quantifying antimicrobial resistance genes (ARGs) in dairy and beef farm samples and identified common study design features, gene targets, and sample types. ARGs conferring resistance to sulfonamide and tetracycline drug classes were most commonly targeted and those in the beta-lactam drug class were most often found to be below method detection limits. Samples originating from cows treated with antimicrobials had higher concentrations than samples derived from control cows although different patterns were observed between studies and sample types. Similarly, manure-treated soils have higher ARG abundances than control soils but were target and study-specific. Future work on ARG abundances in field samples should include cross-sectional design elements to better quantify within-study variation.

2.1 Introduction

While the emergence and increasing prevalence of antimicrobial resistance bacteria (ARB) in clinical settings has evident implications to human health (1), consensus has not been reached on the public health risks associated with the prevalence of ARB in agricultural environments. Some literature reviews argue that contamination originating from livestock wastes poses a significant threat to resistance in the human population (2–4). Others, however,

9

argue that human health risks associated with land application are trivial relative to other daily life exposures and interventions targeting land application would likely result in insignificant benefits (5). One possible explanation for the lack of consensus across studies is that studies and review papers often compare antimicrobial resistance genes (ARG) abundances between different species of livestock (2–4, 6); different livestock species have distinctly different microbiomes (7) and different types of animals are administered different antimicrobials (8).

In the United States, an estimated 1.23 billion tons (wet weight) of manure is produced annually, with dairy and beef accounting for about half of nitrogen and about 40% of phosphorus produced from all livestock manure (9). Cattle farms, including dairy and beef, tend to be more decentralized than hog and poultry operations (10). As a result, the risks associated with ARGs originating form of cattle manure are more geographically widespread than those of other animals. Moreover, antimicrobials used in the dairy and beef industry in the United States overlap with medically important antibiotics used in human medicine (11).

Meta-analyses are important tools for quantitatively assessing the consensus across studies. For instance, a meta-analysis of ARGs in lake and river systems demonstrated that quantitative PCR (qPCR) was the most common tool for quantifying resistance in these environments and that genes conferring tetracycline and sulfonamides resistances were most frequently targeted (12). Another meta-analysis compared antimicrobial use interventions (13) and found statistically significant correlations between a ban on using glycopeptide antimicrobials (e.g., vancomycin) and *vanA* gene prevalence in human and animals but no effect between beta-lactamase and most tetracycline ARG abundances and interventions with the respective antibiotics (13).

The purpose of this systematic review and meta-analysis was to quantitatively analyze studies measuring abundances of ARGs in bovine manure samples to identify trends in ARG abundances in the beef and dairy industries. The results of the study can inform research directions and study designs and ultimately aid in reaching consensus on the public health risks associated with ARGs in cattle manures.

2.2 Materials and Methods

2.2.1 Systematic Search approach

We first conducted a systematic literature review to collect data on ARGs in manure, fecal, and soil samples from dairy and beef farms. The systematic search involved first identifying ten articles known to contain data relevant to our study objectives (Table A-1). The purpose of the ten studies, we refer to as known relevant articles (KRAs), was to test different sets of search terms in various databases and identify the search criteria that find all the KRAs present in the given database. The KRA list was compiled to identify studies measuring ARG abundances in a variety of bovine sample types including fecal, digester, compost, and stored manures.

We used three databases to collect articles for analyses: PubMed (PM), Scopus (SC), and AGRICOLA (AG). The first iteration of the search included all relevant search terms assigned to each KRA. Then, several search iterations were performed using additional search term synonyms. Final search iterations were considered successful when all KRAs available in the database were returned. These final search iterations and the dates on which they were run are listed in Table A-2.

The articles that resulted from the final search were sorted to remove duplicates and titles and abstracts were reviewed for relevance. Studies were considered relevant if they performed qPCR to quantify ARGs in dairy or beef feedlot environmental samples. The applied exclusion

criteria included: culture-based studies, investigations utilizing a lab-scale application or treatment, studies not utilizing standard-based qPCR techniques, and papers written in a language other than English. If a review of the title and abstract was not adequate to assess inclusion criteria, the entire article was reviewed.

2.2.2 Data Collection

For the purpose of this analysis, a sample from a study was defined as 'unique' if it originated from independent sampling sites, represented different sample types within a farm or site, or was collected at the beginning and end of a longitudinal study. Intermediate data points from longitudinal studies were not recorded. Each unique sample identified in the collected articles was labeled and recorded. Once all unique samples from a study were identified, the gene targets and ARG abundances were recorded. The mean values of biological and technical replicates were recorded when available; in some cases, only median values were available and were extracted for this study. Web Plot Digitizer (https://automeris.io/WebPlotDigitizer/) was used to extract data presented in a graphical format. When we were unable to extract data from a figure, such as when data were presented as a heat map, when sample points overlapped, or when stacked bar charts were illegible, we reached out to the manuscript corresponding authors. Specifically, we emailed a request for raw data and at least one follow-up email was sent when no response was received.

Various metadata were recorded for each unique sample including primer references, extraction kits and manufacturers, qPCR reagents, sample types (soil, feces, etc.), animal types (dairy, beef, unspecified), and the country where the study was conducted. Samples classified as feces were collected before fecal matter interacted with the environment. For the purpose of this study, manure was defined as a mix of feces, urine and bedding that was collected from barn

floors, manure storages, lagoon, manure stockpiles, etc. Different studies employed different names for common gene targets; we recorded the names for each gene target according to the Comprehensive Antibiotic Resistance Database (CARD, (14)). The CARD ontology was also used to leverage relationships between target genes including "ARG Family" and "target drug class" that describe genotypically and phenotypically related targets, respectively (Table A-3). The ARG primer references listed by the studies were not always the original source of the primers. We therefore traced back to the original manuscripts that described primer designs and reference those in our reporting.

2.2.3 Data Analysis

To compare the results from studies addressing similar research questions, each study was assigned at least one study design feature (Table 2-1). "Antibiotic-control" studies compared ARG abundances in samples derived from antibiotic-treated cows to samples derived from non-antibiotic treated cows. "Amendment-control" studies compared soils in fields where manure was applied to control soils without manure amendment. Studies were identified as "Impact-control" studies when the ARG abundances in soils from pastures or barn environments were compared to ARGs in soils from control 'pristine' areas nearby the farm site. "Exploratory/Farm Survey*"* studies compared multiple locations within a manure management system to study the dissipation of ARGs within farms. A single study could include multiple study design features. Studies were classified as "longitudinal" and/or "cross-sectional" depending on the sampling approach. Longitudinal sampling approaches quantified ARGs at multiple time points for specific locations. Cross-sectional studies were defined in this study as studies that compared two or more farms. Several studies analyzed samples from other animals, but those data were not recorded.

Response ratios (RR) between control samples and treatment samples were calculated as the natural log of the ratio of mean ARG abundances reported in the study following Hedges et al. 1999 (15). This approach has been used in environmental ecology meta-analyses to track, for example, the impacts of fertilizers and genes involved in nitrogen cycling (16) and impacts of aquaculture on species diversity (17). Many of the studies reporting ARGs did not consistently report or fully explain measures of variance, such as standard error or standard deviation. Consequently, variances were not used for the RR analyses. Response ratio calculations were performed in R (version 4.0.2) and GGplot2 (Version 3.3.2) was used for data visualization.

2.3 Results and Discussion

2.3.1 Summary of extracted papers

At the conclusion of the systematic review, 43 peer-reviewed studies were identified (Table 2-1) that measured antimicrobial resistance gene abundances with standard qPCR on field-scale samples originating from bovine farms (dairy and beef). Almost half (20) of the studies were conducted in the United States, followed by 7 from Canada, 6 from China, and 3 studies that did not explicitly state the country (Table 2-1). Publication years ranged from 2007-2020, with most papers published between 2014-2018 (24 articles).

Table 2-1: *Studies meeting meta-analysis inclusion criteria organized by study design features.*

Study (Ref.)	Country	Farm	Samples	ARG Targets	Study Designs Assigned, Summary of data Extracted
Antibiotic-Control (ABX-Control) Studies					
Alali2009 (24)	United States	nondairy	feces	*CMY-2*	**Longitudinal, ABX-control.** Compared fecal shedding between different ceftiofur treatment regimens and control over 27-day period.
Beukers2018 (25)	Canada	nondairy, dairy	feces	*CTX-M, ermA, ermB, ermC, ermF, mel, qnrS, sul1, sul2, TEM-1, tet(A), tet(B), tetM, tetO, tetQ, tetW*	**Cross-Section, ABX-control.** Compared fecal composites from different feedlots including conventional and RWA feedlots
Boyer2012 (26)	NR	dairy	feces	*CMY-2*	**ABX-control.** Compared fecal shedding between ceftiofur-treated and untreated dairy cows
Chen2008 (27)	NR	nondairy	feces	*ermA, ermB, ermC, ermF, ermT, ermX, tet(A/C), tet(G), tet(RPP)*	**Cross-Section, ABX-control.** Compared fecal composites from steers with tylosin in feed to control composite fecal samples from control.
Feng2020 (28)	United States	dairy	feces	*tetQ, mel, cfxA*	**Longitudinal, ABX-control.** Compared fecal abundances in dairy cows treated with pirlimycin hydrochloride, ceftiofur, and control, untreated cows over 2 weeks.
Holman2019 (29)	Canada	nondairy	feces	*str, ROB, TEM, ermA, ermX, sul2, tet(C), tet(H), tetM, tetW, tetC, tetH*	**Longitudinal, ABX-control.** Compared fecal shedding between different Oxytetracycline-treated, Tulathromycin-treated, and control cows over 34-day period.
Kanwar2014 (30)	United States	nondairy	feces	*CMY-2, CTX-M, tet(A), tet(B)*	**Longitudinal, ABX-control.** Compared fecal shedding between different ABX treatment regimens and control over 26-day period.
Keijser2019 (31)	Germany	nondairy	feces	*tetM, mel, floR*	**Longitudinal, ABX-control.** Compared fecal shedding between different oxytetracycline (OTC) treatment regimens and control over 42-day period.
Kyselkova2013 (32)	Canada	dairy	feces	*tetQ, tetW, tetZ*	**Microcosm study. ABX-control.** Data extracted from raw manure sample collected from one healthy and one Oxymycine-treated dairy cow

Kyselkova2015 (33)	Czech Republic	dairy	feces, manure, soil, never appl, amended soil, soil unspecified	*tetM, tetQ, tetW, tetY*	**Soil Cross-Section, Amendment Control. ABX-Control. – within-farm study?** Compared manure before and 8-days after ABX treatment, calf samples, and pre-birth heifer samples. Soils compared from manure-impacted and non-impacted control locations.
Peak2007 (34)	United States	cattle unspecified	lagoon	*tetM, tetO, tetQ, tetW, tet(B), tet(L)*	**Cross-Section, ABX-Control.** Longitudinal study, concentrations measured over 6 months, but each ABX-usage level averaged over time.
Sharma2009 (35)	Canada	nondairy	compost	*tet(A/C), tet(G), tet(RPP), ermA, ermB, ermC, ermF, ermT, ermX*	**Longitudinal ABX-Control** Compared genes in composted windrows from treated and control cattle over 11 weeks
Thames2012 (36)	United States		Feces	*ermB, ermF, sul1, sul2, tet(C), tet(G), tetO, tetW, tetX, int1*	**ABX-Control.** Compared gene abundances in calves fed with milk replacers with and without ABX. Measured over 12-weeks but averaged overtime within groups recorded.
Vikram2017 (37)	United States	nondairy	feces	*aadA, CMY-2, CTX-M, KPC-2, ermB, tet(A), tet(B), tetM, AAC(6')-Ie-APH(2")-Ia, mecA*	**ABX-Control** Compared fecal samples collected over a year from conventional and RWA feedlots.
Wichmann2014 (38)	United States	dairy	manure	*bla2, bla3, bla4, tetW, AAC(6')-Ie-APH(2")-Ia, cat*	**ABX-Control/Other** Primarily a functional metagenomic study exploring geno-and phenotypic diversity but also compared pooled manure from treated and control cows.
Xu2016 (39)	Canada	nondairy	compost	*tet(B), tet(C), tet(L), tetM, tetW, ermA, ermB, ermF, ermX, sul1, sul2*	**ABX-control, Longitudinal.** Averaged data was reported, limiting the data extracted

Amendment-Control and Impact-Control Studies of Soil

Bastos2017 (40)	Brazil	dairy	never appl, pasture env	*sul1, qnrA, erm*	**Cross-Section/Impact-Control.** Compared pristine soils to dairy farm soils
Dungan2018 (41)	United States	dairy	lagoon, never appl, amended soil	*CTX-M-1, ermB, int1, sul1, tet(B), tetM, tetX*	**Amendment-control.** Compared soil plots with different volumes of dairy manure amendments and control soils. Measured over time but longitudinal data not reported.
Fahrenfeld2014 (42)	United States	dairy	slurry, hist appl, amended soil	*ermF, sul1, sul2, tet(G), tetO, tetW, vanA*	**Other (Modeling Study) /Amendment-control.** Data extracted from soil before and after dairy manure amendment.

				sul1, sul2, tet(A), tetB(P), tet(C), tet(E), tet(G), tetM, tetO, tetQ, tetS, tetW, tetX	**Amendment Control** Compared abundances in cattle-impacted soil vs. control soil and various samples from other livestock farms
Li2018 (43)	China	cattle unspecified	pasture env, never appl		
Macedo2020 (44)	Netherlands	dairy	manure, never appl, hist appl, amended soil	*sul1, ermB, tetW, TEM*	**Amendment-Control. Longitudinal?** Compared impact of manure on different soil types.
Marti2014 (45)	Canada	dairy	hist appl, amended soil	*ermB, sul1, str(B), int1, repA*	**Longitudinal, Amendment Control.** Control and dairy manure-amended plots sampled over 100-200 days post-application. study repeated over three seasons, two during fall application and one spring.
McKinney2018 (41)	United States	dairy	soil unspecified, amended soil	*CTX-M-1, ermB, int1, sul1, tet(A), tetW, tetX*	**Longitudinal, Amendment Control.** Compared soils with different land application rates over for 4 years. Field plots were amended in the fall and soil samples were collected in spring before planting. Post-harvest sample were also collected in the fall before manure application but these data were not recorded in this study since they were measured at different depths than spring samples.
Miller2019 (46)	United States	nondairy	manure, hist applied, amended soil	*CMY-2, CTX-M, KPC-2, AAC(6')-Ie-APH(2")-Ia, aadA, ermB, tet(A), tet(B), tetM*	**Longitudinal, Amendment Control.** Compared 5 different amendments at field sites in three different sites with samples collected over a year at each site before/after applications and harvest.
Nolvak2016 (47)	Estonia	dairy	slurry, digested, never appl, amended soil	*CTX-M, OXA-2, tet(A), qnrS, sul1, int1, int2*	**Longitudinal, Amendment Control.** Measured abundances in soil over 150 days comparing amendments that were applied three times in 150-day period.
Peng2017 (48)	China	cattle unspecified	raw manure, never appl, amended soil	*tet(G), tet(L), tetZ, tetB(P), tetO, tetW, tetM, sul1, sul2, sul3, ermB, ermF, ermC, CTX-M, TEM, int1, int2*	**Amendment Control.** End of 30-year experiment comparing plots with mineral fertilizer, no fertilizer, piggery manure and cattle manure.
Tien2017 (49)	Canada	dairy	raw manure, digested, SLS, compost, soil unspecified, amended soil	*aadA, OXA-20, ermB, ermF, repA, int1, bla-PSE, APH(6)-Id, APH(3")-Ib, sul1*	**Longitudinal Amendment-Control.** Compared abundances in soils land applied with manure at different stages of treatment over 155-day period
Udikovic-Kolic2014 (50)	United States	dairy	never appl, amended soil, raw manure	*bla-CEP-02, bla-CEP-05, bla-CEP-04, bla-CEP-01, bla-CEP-03*	**Longitudinal Amendment-Control.** Compared beta-lactamase genes over 130-day period.

Study	Country	Type	Samples	ARGs	Description
Wepking2017 (51)	United States	dairy	never appl, pasture env	*tetO, tetW, ampC, ermB*	**Cross-section, Impact-Control** Compared abundances in soils in control or farm-impacted environments across 11 sites across the United States
Xi2015 (52)	China	cattle unspecified	feces, pasture env, never appl	*tetO, tet(L), tet(RPP), rpoB, sul2, sulA, floR, mdtL, mphA, ermC*	**Other/Impact-Control** Primarily a source tracking study but measured several relevant field samples

Exploratory/Farm Survey

Wallace2018 (53)	NR	dairy	raw manure, pasteurized, digested, lagoon, SLS	*tetO, tetW, sul1, sul2*	**Exploratory/Farm Survey** Compared ARG abundances throughout treatment process at two farms and over winter and spring.
Yang2010 (54)	United States	nondairy, dairy	compost, feces, pasture env, slurry, raw manure	*tet(B), tet(C), tetO, tetW, CMY-2*	**Exploratory/Farm Survey** Compared samples from different locations on a beef and dairy farm and compared to urban and pristine sites.
Zhang2013 (55)	United States	nondairy	lagoon	*sul1, sul2, tetO, tetQ, tetX*	**Exploratory/Farm Survey** Measured abundances in different storage a at a cattle farm, including solid and liquid fractions. Compared to a swine farm.

Other Study Design Features

Katakweba2015 (56)	Tanzania	cattle unspecified	feces	*sul2, tetW, sulII*	**Cross-Section.** Zebu cattle feces compared to other wild animals.
Hurst2019 (57)	United States	dairy	lagoon, raw manure	*int1, OXA-1, sul1, tetO*	**Longitudinal Cross-Section of manure, Cross-section of storages.** Raw manure samples collected over two-year period.
Ji2012 (58)	China	dairy, cattle unspecified	raw manure, amended soil	*sul1, sul2, sul3, sulA, tet(B)P, tetM, tetO, tetW*	**Cross Section of manures and soils.** Compared cattle manure and manured soil to other farms.
McKinney2010 (59)	United States	dairy, nondairy	lagoon	*tetO, tetW, sul1, sul2*	**Cross-Section** Lagoon samples collected over a year from dairies and beef feedlots. Yearly samples were averaged so not classified as longitudinal.
Mu2015 (60)	China	nondairy	feces, pasture env	*tetM, tetO, tetQ, tetW, sul1, sul2, qnrS, oqxB, ermB, ermC*	**Fecal Sample Cross-Section** Abundances in 7 fecal samples and 4 samples from nearby soils.
Munir2011 (61)	United States	dairy	lagoon, soil unspecified, amended soil	*tetW, tetO, sul1*	**Cross-Section.** Manures (3 farms) and pre- and post-amended soils (2 farms) sampled longitudinally over 4 months but

					abundances averaged. Results compared to biosolids and biosolid-amended soils.
Chen2015 (62)	China	cattle unspecified	slurry	*acrA, acrB, aac, aph, ANT(4)-Ia, ermB, ermC, oqxB, qnrD, qnrS, sul1, sul2, sul3, tet(A), tet(C), tet(E), tet(G), tetM, tetO, tetQ, tetT, tetW*	**Other/Cross-Section.** Compared wastewaters from farms and municipalities and surface water. Data collected here were averages of 4 cattle farms.
Netthisinghe2018 (63)	United States	nondairy	pasture env	*sul1, ermB, int1*	**Longitudinal.** Measured mitigation of chemical and ARGs over time at abandoned beef feedlots.
Ruuskanen2015 (64)	Finland	dairy	raw manure, lagoon, soil unspecified, amended soil	*OXA-58, sul1, tetM*	**Longitudinal, Cross-Section.** Compared fresh, stored and pre-/post-fertilized soils. Longitudinal samples taken two weeks post-fertilization. Compared to swine samples.
Storteboom2007 (65)	United States	dairy, nondairy	compost	*tetW, tetO*	**Longitudinal.** Compared different-level of intensity composting on mitigation of ARGs at beef and dairy farms. Measured over 180 days

Manure, fecal samples, and soil samples were the most probed matrices (Table 2-2). In total, 79 different ARGs were measured across all studies, with 40 of those genes targeted in only one study each. *sul1, tetW* and *tetO* were the most commonly targeted genes, measured in 23, 22, and 18 of the 43 studies, respectively. *ErmB* was targeted in 17 studies and *sul2* and *tetM* were targeted in 15 studies, each. Collectively, genes conferring resistance to tetracyclines were targeted in 36 out of the 43 studies (Table 2-1).

Genes targeted in more than one study were often targeted using several different primer sets. For instance, *ermC* was targeted in seven studies using six different primer sets. Other cases where all but one study used different primers for the same targets included *CTX-M*, *tet(C)*, and *tetX. tetM* had the most primer sets referenced, with 11 different sets used in 15 studies. *ermB, sul1 and tetW* were targeted in 17, 23, and 22 studies respectively with seven unique primer sets each for these targets. *tetM,* therefore, is an outlier in the number of different primer sets used for a single gene target.

The most common study design feature was the "Antibiotic-control" (16 studies), which involved comparing samples from groups of cows that were either treated or not treated with antibiotics (Table 2-1). The next most common study design features were "Amendment-control" (10 studies) and "Impact-control" (4 studies). Three studies which compared abundances of antimicrobial resistance genes in different stages of manure treatment were assigned the "Exploratory/Farm Survey" design label. Most studies incorporated longitudinal study designs (Table 2-1).

After summarizing the studies and common sample types, we sought to identify trends in gene abundances which may help identify where the field has come to a consensus on the role of farm management in ARG prevalence. First, we analyzed gene abundances in the two of the

most targeted samples types, feces and manure to determine ranges of abundances detected

across studies. Then, to evaluate the robustness of observations of antibiotic use and land

application, we compared results across studies assigned to the Antibiotic-control and

Amendment-control study designs, respectively.

Table 2-2: *Frequently targeted genes grouped by CARD drug class in different sample types*

	Manure[1]	Soil[2]	Feces	Compost	Digester
	Sample Type				
tetracycline	18	16	14	4	2
sulfonamide	14	16	6	2	3
beta-lactam	11	10	8	2	2
MLS[3]	7	13	7	3	1
aminoglycoside	4	3	2	1	1
fluoroquinolone	2	3	2	0	1
other	5	6	2	1	2

[1]Manure includes samples designated as 'lagoon', raw manure, wastewater, slurry, and
manure
[2]Soil includes amended and impacted soils as well as various control soil types
[3]MLS = macrolides, lincosamides, and streptogramins

*2.3.2 Reported concentrations in ARG abundance within fecal and manure samples range over
10 orders of magnitude between studies.*

To determine if there is a consensus in trends of ARGs present in fecal and manure samples from

dairy and beef farms, we observed ranges and patterns across studies of the most common

resistance gene targets. The most common genes targeted in samples of manure and feces

conferred resistance to tetracycline, followed by sulfonamides, beta-lactams and the macrolides,

lincosamides, and streptogramins (MLS) group (Table 2-2). Gene abundances were most often

reported as copies of ARGs per copies of 16S rRNA genes. Comparing abundances reported in

both mass-standardized and 16S rRNA gene-standardized found that while mass normalized

abundances ranged over greater order of magnitude (Figure A-1), 16S rRNA gene-standardized

had higher coefficient of variation meaning the abundances were more dispersed (Figure A-2).

This suggests that 16S rRNA gene-standardized abundances may introduce more between-study and between-observation variability.

2.3.2.1 ARG abundances in fecal samples

13 studies measured gene abundances in fecal samples and reported abundances in ARG copies per 16S rRNA gene copies (Figure 2-1). The abundances of genes related to the tetracycline, beta-lactam, MLS, and sulfonamide groups varied as many as 10 orders of magnitude, excluding one paper which reported tetracycline genes from 0.06-13.5 log10 copies/16S rRNA gene copies (Thames *et al.* 2012). The large range in tetracycline resistance genes were driven in part by the diversity of the genotypes of tetracycline resistance targeted across studies. 14 different tetracycline resistance genes were targeted between studies, with eight targets in major facilitator superfamily (MFS) antibiotic efflux pump, five from the tetracycline-resistant ribosomal protection protein gene family, and one gene in the tetracycline inactivation enzyme gene family. Genes in the (MFS) antibiotic efflux pump and tetracycline-resistant ribosomal protection protein gene families both ranged 6.5 orders of magnitude.

Eight studies reported non-detected genes. Genes in the beta-lactam drug class were the most frequently undetected; four studies reported undetected beta-lactamase genes including *CMY-2, CTX-M, mecA, ROB,* and *TEM*. Three studies reported undetected genes in the tetracycline drug class including *tet(C), tet(H), tetM,* and *tetY*.
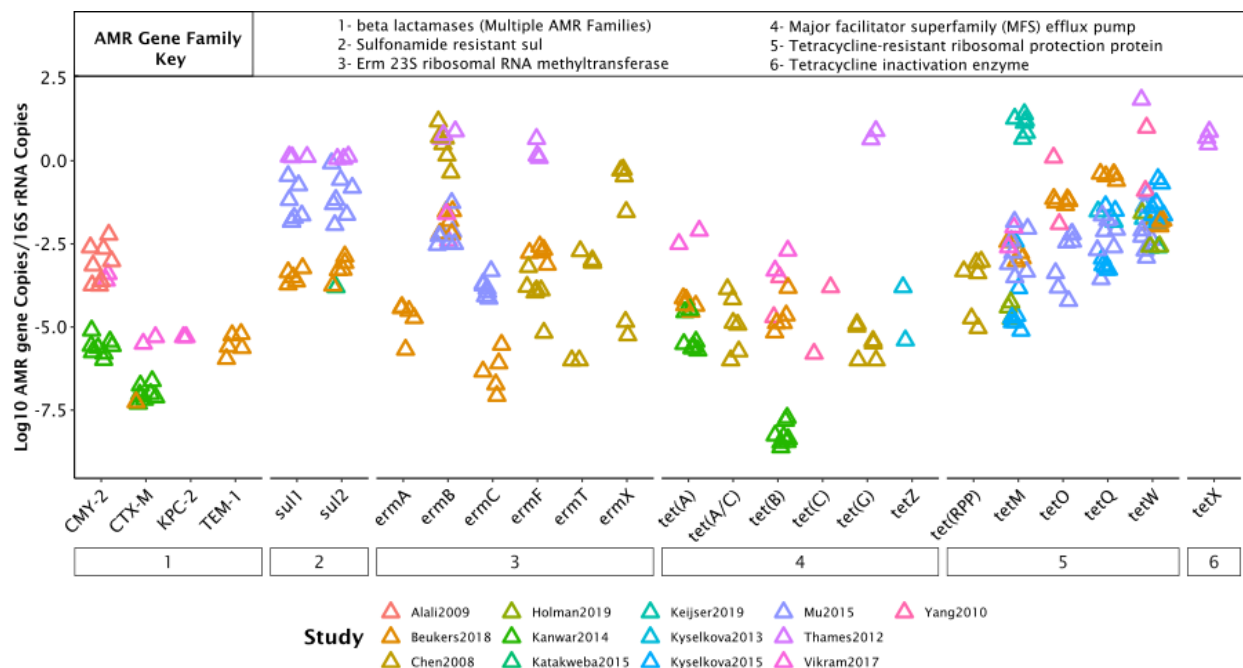
Figure 2-1: *Antimicrobial resistance gene abundances (log10(copies /16S rRNA gene copies)) in the six most targeted ARG families across the 13 studies which reported copies/16S rRNA gene copies from fecal samples.*
Y-axis limits cut-off tet(G), tetW, and tetO targeted by Thames et al. 2012 which were measured at abundances ranging 4-13.5 log10(copies/16S rRNA gene copies).

2.3.2.2 ARG abundances in manure samples

20 studies collectively targeted 57 different resistance genes in samples that we classified as manure and reported results as copies/16S rRNA gene copies. Some studies specifically reported the locations where manure samples were collected (i.e. barn floor, manure stockpile, lagoon, earthen pit) while other studies described manure samples as a "slurry" or as "dairy/beef wastewater." We anticipated a greater range in ARG abundances in manures compared to feces due to the greater number of manure samples extracted in the systematic review as well as the variety of sample types characterized as manure with different water contents, storage times and storage conditions. In this set of studies, one was an apparent outlier, reporting abundances of beta lactamase genes, *tetW,* and *AAC(6')-Ie-APH(2″)-Ia* as low as -22 $\log_{10}$ gene copies/ 16S rRNA gene copies (Wichmann 2014). Excluding this study, gene abundances in units of

copies/16S rRNA gene copies ranged four, three, seven and seven orders of magnitude in drug

class groups aminoglycoside, beta-lactam, tetracycline, and sulfonamides, respectively (Figure 2-

2).

A greater number of ARGs were targeted in manure samples than feces samples, likely

due to the greater number of studies quantifying gene abundances in manure samples. Like fecal

samples, genes in the tetracycline-resistant ribosomal protection protein gene family were the

most commonly targeted, with all but one study targeting at least one tetracycline resistance

gene. However, genes in the sulfonamide resistant sul gene family were more commonly

targeted in manure (14 out of 20 studies) than fecal samples (4 out of 13 studies) with a total of

88 unique observations. Tetracycline genes were detected in almost every sample, with only one

study reporting a manure sample with undetected levels of *tetB(P)*. Similar to the fecal samples,

beta lactamases were most frequently not detected; six studies reported nine different genes in

the beta-lactam drug class undetected or not-quantifiable. The most commonly targeted genes

were *sul1* (14 studies, 88 unique samples), *tetW* (12 studies, 50 unique samples), and *tetO* (11

studies, 74 unique samples), (Figure 2-2). *TetM* was also frequently targeted (8 studies, 20

samples), and almost every study used a unique primer set.
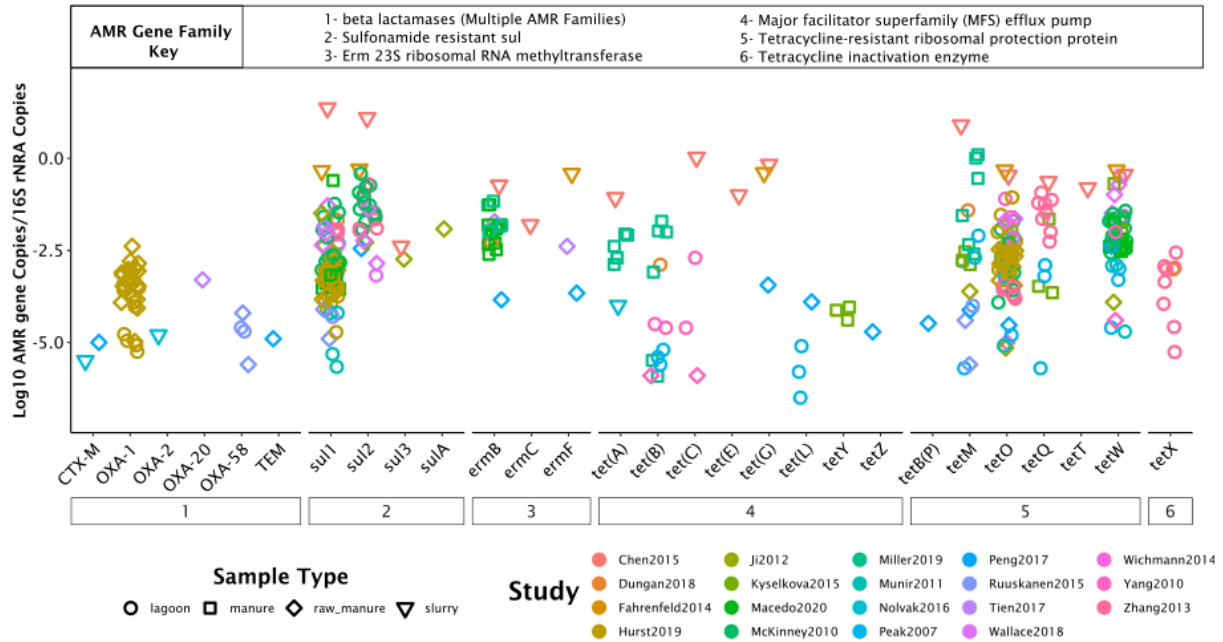
Figure 2-2: *Antimicrobial resistance gene abundances (log10(copies /16S rRNA gene copies) in the top six most targeted ARG Families across the 19 studies which reported copies/16S rRNA gene copies from manure samples.*
Y-axis limits cut-off tetW targeted by Wichmann et al. 2014 and tetL targeted by Peak et al. 2017 which were measured at abundances ranging -6.5-14 log10(copies/16S rRNA gene copies).
Beta-lactamase gene targets with custom primer sets based on sample isolates targeted by Wichmann et al. 2014 and Udikovic-Kolic et al. 2014 in addition to bla-PSE gene were excluded since they were not measured in other studies.

2.3.3 Antibiotic use on ARG abundances in manure samples

To study the effect of antibiotic use on ARG concentrations in samples, the response ratio was calculated for all pairs of samples originating from cows treated with antibiotics and control samples. Inclusion criteria for response ratio calculation required that control samples were collected at the same time as the samples from treated cows. 14 of the 16 studies that were classified as "Antibiotic-control" studies included samples that met this criterion for calculating response ratios. Sample types analyzed from treated and control cows included fecal samples, compost, and manure samples.

For longitudinal studies in which more than three samples were collected over the experimental period, only the final sample points for the treatment and control samples were

collected. The most administered antibiotics in these studies were in the tetracycline class, specifically oxytetracycline and chlortetracycline, which were administered in 7 of 13 studies. Three of the studies did not specify which antibiotics were administered to the cattle but instead described the antibiotic treatment as "conventional" and compared to samples from these animals to animals that were raised without antibiotics. 34 observations of gene targets out of a total of 196 meeting the inclusion criteria were not quantified in either the control or treatment sample. Only one observation detected the target, *CTX-M*, in the treated cow sample but not in the control. Five studies reported at least one non-detect of 14 different gene targets from aminoglycoside, beta-lactamase, MLS, fluoroquinolone, sulfonamide and tetracycline antibiotic classes.

Overall, the median response ratio of all ARG observations in these "Antibiotic-control" studies was 0.43. This equates to a 1.5-fold higher abundance of resistance genes observed in samples originating from animals treated with antibiotics than in manure samples from animals not treated with antibiotics. Peak *et al.* 2007, which measured abundances in stored manure samples from conventional and raised-without-antibiotic farms, had a median response ratio of 3.7, a clear outlier in the analysis. This may be due to the sample types in the study—this was the only study that compared the abundances of genes in manure stored in lagoons. More studies of stored manure would clarify if the observation of higher fold-differences between treated and untreated herds are indeed higher than fold-differences between treated and untreated fecal samples and identify why stored manure from treated cows has higher abundances of ARGs.
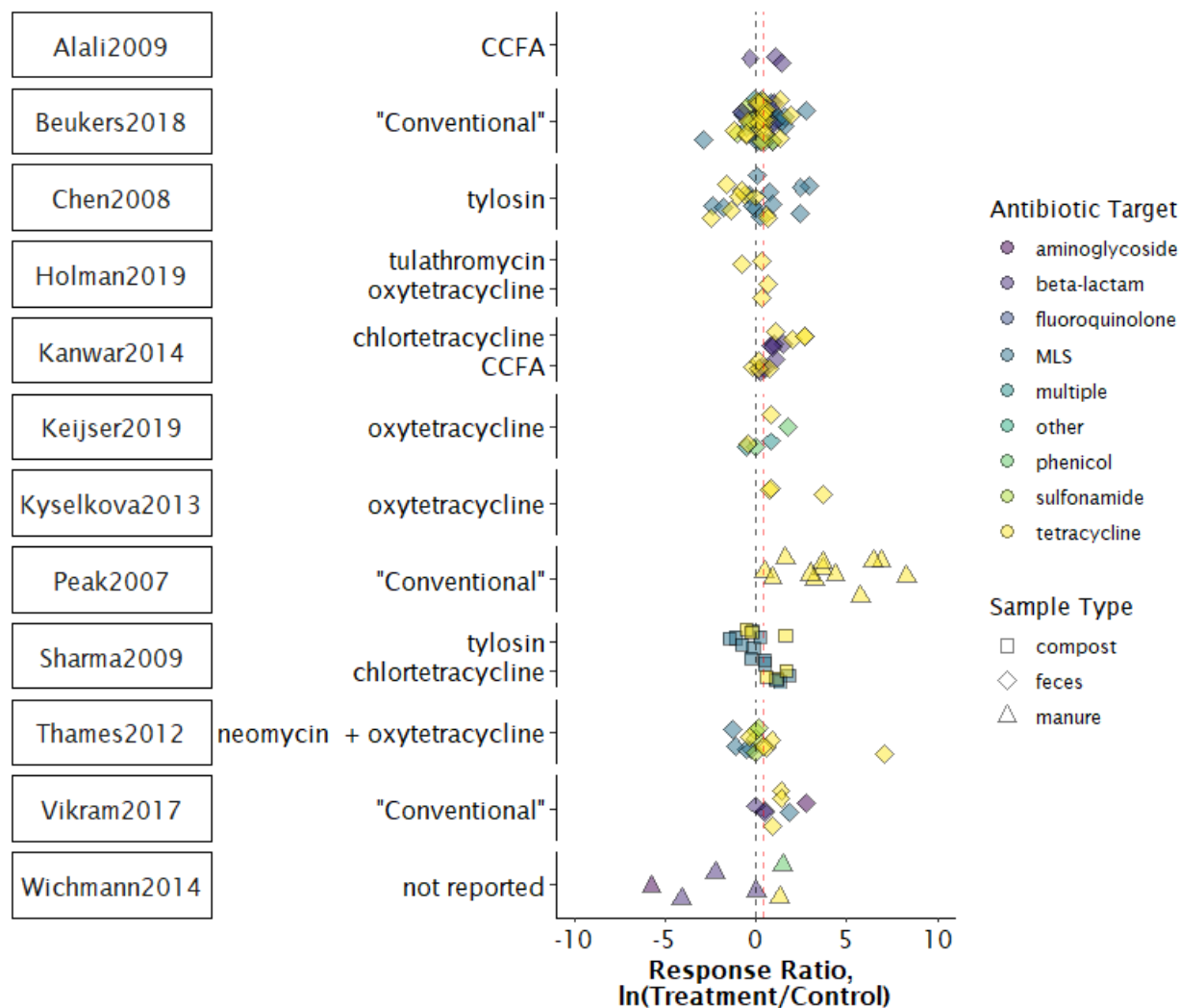
Figure 2-3: *Response ratios organized by study comparing unique samples (fecal, compost, manure) originating from cows or herds treated with antibiotics and control cows or herds. Y-axis lists the antibiotics used in the treated cattle. CCFA = Ceftiofur crystalline-free acid. "Conventional" farms did not list the specific antibiotics and were compared to farms raised without antibiotics. Red dashed line at x=0.43 indicates median RR across all studies and samples. Dashed black line at x=0 indicates where there is no observed difference in treatment and control concentrations. Outliers from Thames et al. 2012 are cut off to enhance visibility of results. Only 16S rRNA gene-standardized gene abundances used for response ratio computation.*

2.3.4 Land application of manure increased gene abundances of some gene targets.

Another common study design feature across the literature assess the impact of land application

on ARG abundances in soil. To evaluate the impact of bovine manure amendment on soil

antimicrobial resistance gene abundances, response ratios comparing amended soils to control soils were computed from amendment control-studies, which included a total of 10 studies. Inclusion criteria for response ratio calculation required that a control plot was sampled at the same time as amended samples. This excluded two papers from this analysis (Fahrenfeld *et al.* 2014, Macedo *et al.* 2020). Lastly, Tien *et al.* 2017 was excluded because only mass-standardized (copies/mass) abundances were reported.

Of the relevant studies identified through the systematic review, seven studies contained relevant data and met the inclusion criteria for this analysis. Response ratios were computed as the natural log of 16S rRNA gene-standardized concentrations for each unique sample-control pair. In some cases, multiple controls were performed. For instance, Miller *et al.* 2019 compared manured soil to control plots with no amendments and to control plots with mineral amendments. In these types of studies, two response ratios were calculated, one comparing treatment to the mineral amendment control and one to the no-amendment control. From the seven studies, 422 total unique observations were made. Miller *et al.* 2019 had the most observations, comparing sites in three different states and measuring gene abundances during three growing seasons. McKinney *et al. 2018* also had several unique observations, comparing three different amendment rates ranging from 17-52 million gallons per hectare over three growing seasons (Table A-5).

Over half the observations, response ratios could not be calculated for 262 amendment-control pairs because the gene was not detected in both the control and treatment sample. Dungan *et al.* 2018 and McKinney *et al.* 2018 both targeted *tetX* and *CTX-M-1* and did not detect or infrequently detected genes in both treatment and control samples. *ermB* was also not detected in samples from Miller *et al.* 2019 and McKinney *et al.* 2018. Beta-lactamases had the fewest

observations because it was not commonly targeted in soil samples (Table 2-1, Table 2-2) and when targeted, it was frequently not detected (Table A-5). In the studies meeting the inclusion criteria for this analysis, beta-lactamases from the CTX-M, OXA, TEM, KPC, and CMY gene families were targeted but only TEM and CTX-M classes were observed. Additionally, unclassified beta-lactamases identified in soil isolates (*bla-CEP*) were used as targets and were detected in control and amended soil by Udikovic-Kolic *et al.* 2014.

Response ratios spanned from -1 to 5.6 with a median response ratio of 0.46 (Figure 2-4, Table A-5). Tetracycline and aminoglycoside resistance genes in Miller *et al.* 2019 exhibited little spread and were close to zero, meaning there was not a large difference in 16S-rRNA gene-standardized abundances in treatment and control soils. *ermB* genes, however, tended to be higher in the treated soils across all studies targeting MLS genes. Sulfonamide, integrase, and tetracycline genes were also observed in higher abundances in treated soils, specifically in the McKinney *et al.* 2018 samples. This meta-analysis of gene abundance in control and manured soils demonstrated that gene abundances are higher in manure soils for some gene classes but depend in part on the gene target and study.
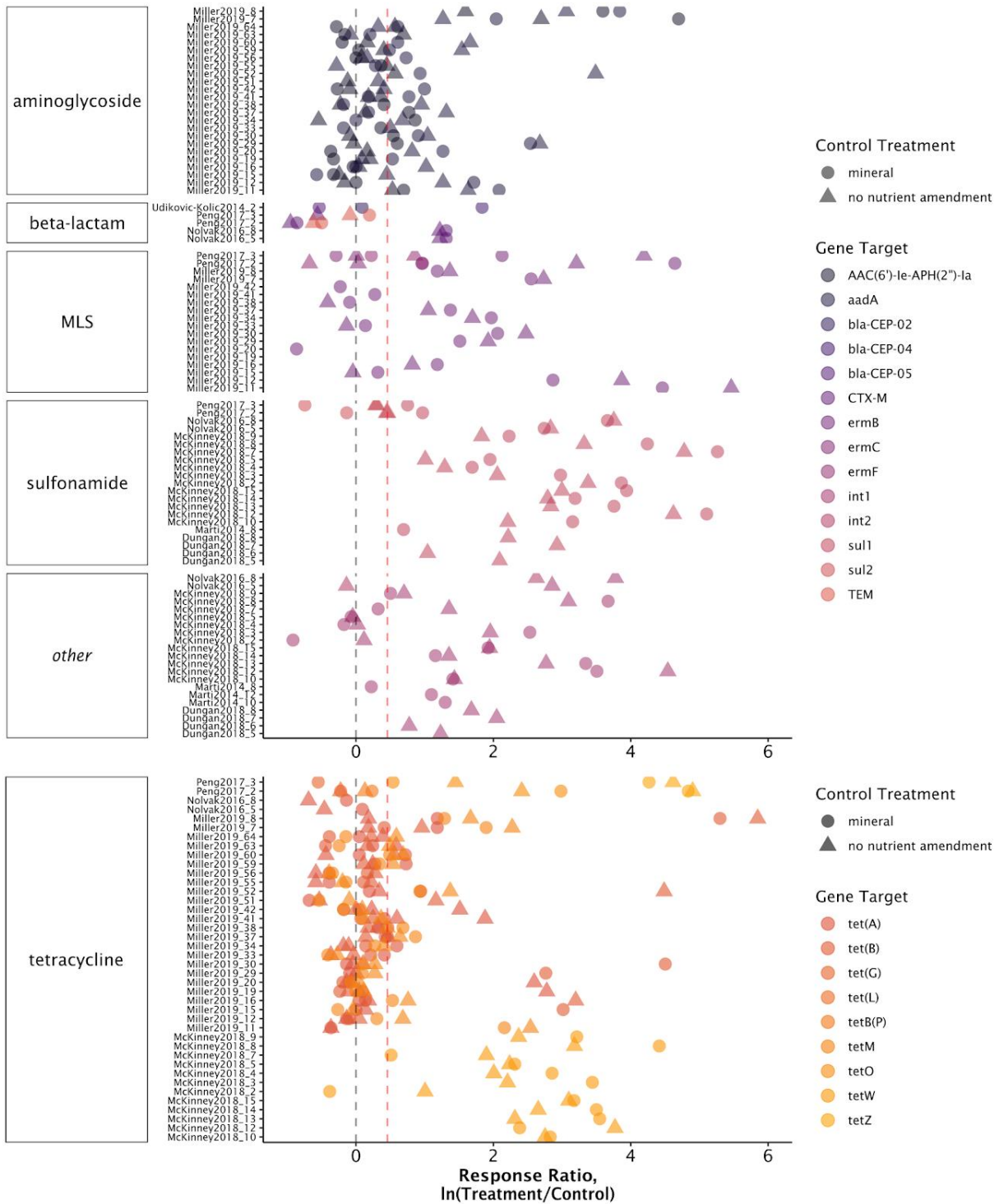
Figure 2-4: *Response ratios (RR) grouped by drug classes assigned to ARGs comparing soil samples in control plots and manure-amended plots.*
Dashed black line at x=0 indicates where there is no observed difference in treatment and control concentrations. Dashed red line at x=0.46 indicates median RR across all observations.

2.4 Conclusions

This systematic review and meta-analysis synthesized peer-reviewed journal articles reporting qPCR-quantified ARG abundances measured in samples collected from dairy and beef farms around the world. Metadata mined from papers identified the most common gene targets, primer sets, and study design features employed over the last decade and a half to determine the impacts of antibiotic use and manure management and investigated risks of land application of manure. 79 different genes were targeted across the 43 studies but only half the targets were quantified in more than one study (Table 2-1). We found that studies typically probe manure, feces, and soil (Table 2-2). Fold-differences in treated and control soil (median = 0.46, 7 studies, Table A-5) and fold-difference observed in fecal samples between treated and non-treated cows (median RR=.43, 12 studies, Table A-4) were quantified across studies.

Differences in gene targets and primer references employed across studies makes comparisons across studies difficult. Future meta-analysis can evaluate differences in high-throughput PCR or metagenomics to identify overlapping detected genes across broader gene families. Throughout the data extraction process, there were various inconsistencies encountered between studies which made data extraction and comparison difficult. Inconsistencies in sample terminology or vague sample descriptions also limited between-study comparisons. Lack of biological replicates within studies and inconsistency in reported variances in observations (standard error, standard deviation) prohibited extraction of these values for response ratio analysis which prohibited calculating the statistical significance of response ratios. Future work to evaluate risks posed by ARGs in cattle or other livestock settings should employ cross-sectional design elements features to maximize biological replicates, observe within-study variance, and compute statistical significance.

2.5 References

1.   Frieden T. 2013. Antibiotic resistance threats. Cdc 22–50.

2.   Heuer H, Schmitt H, Smalla K. 2011. Antibiotic resistance gene spread due to manure application on agricultural fields. Curr Opin Microbiol 14:236–243.

3.   Chee-Sanford JC, Aminov RI, Koike S, Krapac IG, Lin Y-F, Mackie RI, Maxwell S, Yannarell AC. 2009. Fate and Transport of Antibiotic Residues and Antibiotic Resistance Genes following Land Application of Manure Waste. J Environ Qual 38:1086–1108.

4.   Xie W-Y, Shen Q, Zhao FJ. 2018. Antibiotics and antibiotic resistance from animal manures to soil: a review. Eur J Soil Sci 69:181–195.

5.   Pepper IL, Brooks JP, Gerba CP. 2018. Antibiotic Resistant Bacteria in Municipal Wastes: Is There Reason for Concern? Environ Sci Technol 52:3949–3959.

6.   Zhang Y-J, Hu H-W, Gou M, Wang J-T, Chen D, He J-Z. 2017. Temporal succession of soil antibiotic resistance genes following application of swine, cattle and poultry manures spiked with or without antibiotics. Environ Pollut 231:1621–1632.

7.   Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, Ley RE, Farnleitner AH. 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nat Commun 10:1–15.

8.   Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, Teillant A, Laxminarayan R. 2015. Global trends in antimicrobial use in food animals. Proc Natl Acad Sci 112:5649–5654.

9.   Gollehon NR, Kellogg RL, Moffitt DC. 2016. Estimates of Recoverable and Non-Recoverable Manure Nutrients Based on the Census of Agriculture - 2012 Results.

10.   USDA. 2019. Animal Products.

11.   Oliver JP, Gooch CA, Lansing S, Schueler J, Hurst JJ, Sassoubre L, Crossette EM, Aga DS. 2020. Invited review: Fate of antibiotic residues, antibiotic-resistant bacteria, and antibiotic resistance genes in US dairy manure management systems. J Dairy Sci 103:1051–1071.

12.   Yang Y, Song W, Lin H, Wang W, Du L, Xing W. 2018. Antibiotics and antibiotic resistance genes in global lakes: A review and meta-analysis. Environ Int 116:60–73.

13.   Nobrega DB, Tang KL, Caffrey NP, De Buck J, Cork SC, Ronksley PE, Polachek AJ, Ganshorn H, Sharma N, Kastelic JP, others. 2020. Prevalence of antimicrobial resistance genes and its association with restricted antimicrobial use in food-producing animals: a systematic review and meta-analysis. J Antimicrob Chemother.

14.   Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-L V, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar G V, McArthur AG.

2019. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:517–525.

15. Hedges L V, Gurevitch J, Curtis PS. 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80:1150–1156.

16. Ouyang Y, Evans SE, Friesen ML, Tiemann LK. 2018. Effect of nitrogen fertilization on the abundance of nitrogen cycling genes in agricultural soils: A meta-analysis of field studies. Soil Biol Biochem 127:71–78.

17. Barrett LT, Swearer SE, Dempster T. 2019. Impacts of marine and freshwater aquaculture on wildlife: a global meta-analysis. Rev Aquac 11:1022–1044.

18. Alali WQ, Scott HM, Norby B, Gebreyes W, Loneragan GH. 2009. Quantification of the bla(CMY-2) in feces from beef feedlot cattle administered three different doses of ceftiofur in a longitudinal controlled field trial. Foodborne Pathog Dis 6:917–924.

19. Beukers AG, Zaheer R, Cook SR, Chaves A V, Ward MP, Tymensen L, Morley PS, Hannon S, Booker CW, Read RR, McAllister TA. 2018. Comparison of antimicrobial resistance genes in feedlots and urban wastewater. Can J Vet Res 82:24–38.

20. Boyer TC, Singer RS. 2012. Quantitative measurement of blaCMY-2 in a longitudinal observational study of dairy cattle treated with ceftiofur. Foodborne Pathog Dis 9:1022–1027.

21. Chen J, Fluharty FL, St-Pierre N, Morrison M, Yu Z. 2008. Technical note: Occurrence in fecal microbiota of genes conferring resistance to both macrolide-lincosamide-streptogramin B and tetracyclines concomitant with feeding of beef cattle with tylosin. J Anim Sci 86:2385–2391.

22. Feng X, Chambers LR, Knowlton KF. 2020. Antibiotic resistance genes in the faeces of dairy cows following short-term therapeutic and prophylactic antibiotic administration. J Appl Anim Res 48:34–37.

23. Holman DB, Yang W, Alexander TW. 2019. Antibiotic treatment in feedlot cattle: a longitudinal study of the effect of oxytetracycline and tulathromycin on the fecal and nasopharyngeal microbiota. Microbiome 7:86.

24. Kanwar N, Scott HM, Norby B, Loneragan GH, Vinasco J, Cottell JL, Chalmers G, Chengappa MM, Bai J, Boerlin P. 2014. Impact of treatment strategies on cephalosporin and tetracycline resistance gene quantities in the bovine fecal metagenome. Sci Rep 4:5100.

25. Keijser BJF, Agamennone V, van den Broek TJ, Caspers M, van de Braak A, Bomers R, Havekes M, Schoen E, van Baak M, Mioch D, Bomers L, Montijn RC. 2019. Dose-dependent impact of oxytetracycline on the veal calf microbiome and resistome. BMC Genomics 20:65.

26. Kyselkova M, Jirout J, Chronakova A, Vrchotova N, Bradley R, Schmitt H, Elhottova D. 2013. Cow excrements enhance the occurrence of tetracycline resistance genes in soil regardless of their oxytetracycline content. Chemosphere 93:2413–2418.

27. Kyselková M, Jirout J, Vrchotová N, Schmitt H, Elhottová D. 2015. Spread of tetracycline resistance genes at a conventional dairy farm. Front Microbiol 6:1–14.

28. Peak N, Knapp CW, Yang RK, Hanfelt MM, Smith MS, Aga DS, Graham DW. 2007. Abundance of six tetracycline resistance genes in wastewater lagoons at cattle feedlots with different antibiotic use strategies. Environ Microbiol 9:143–151.

29. Sharma R, Larney FJ, Chen J, Yanke LJ, Morrison M, Topp E, McAllister TA, Yu Z. 2009. Selected antimicrobial resistance during composting of manure from cattle administered sub-therapeutic antimicrobials. J Environ Qual 38:567–575.

30. Thames CH, Pruden A, James RE, Ray PP, Knowlton KF. 2012. Excretion of antibiotic resistance genes by dairy calves fed milk replacers with varying doses of antibiotics. Front Microbiol 3:139.

31. Vikram A, Rovira P, Agga GE, Arthur TM, Bosilevac JM, Wheeler TL, Morley PS, Belk KE, Schmidt JW. 2017. Impact of "Raised without Antibiotics" Beef Cattle Production Practices on Occurrences of Antimicrobial Resistance. Appl Environ Microbiol 83.

32. Wichmann F, Udikovic-Kolic N, Andrew S, Handelsman J. 2014. Diverse antibiotic resistance genes in dairy cow manure. MBio 5:e01017.

33. Xu S, Sura S, Zaheer R, Wang G, Smith A, Cook S, Olson AF, Cessna AJ, Larney FJ, McAllister TA. 2016. Dissipation of Antimicrobial Resistance Determinants in Composted and Stockpiled Beef Cattle Manure. J Environ Qual 45:528–536.

34. Camotti Bastos M, dos Santos DR, Aubertheau E, Labanowski J, de Castro Lima JAM, Caner L, Mondamert L, Guet T Le. Antibiotics and microbial resistance in Brazilian soils under manure application. L Degrad Dev 29:2472–2484.

35. Dungan RS, McKinney CW, Leytem AB. 2018. Tracking antibiotic resistance genes in soil irrigated with dairy wastewater. Sci Total Environ 635:1477–1483.

36. Fahrenfeld N, Knowlton K, Krometis LA, Hession WC, Xia K, Lipscomb E, Libuit K, Green BL, Pruden A. 2014. Effect of manure application on abundance of antibiotic resistance genes and their attenuation rates in soil: field-scale mass balance approach. Environ Sci Technol 48:2643–2650.

37. Li C, Jiang C, Wu Z, Cheng B, An X, Wang H, Sun Y, Huang M, Chen X, Wang J. 2018. Diversity of antibiotic resistance genes and encoding ribosomal protection proteins gene in livestock waste polluted environment. J Environ Sci Health B 53:423–433.

38. Macedo G, Hernandez-Leal L, van der Maas P, Heederik D, Mevius D, Schmitt H. 2020. The impact of manure and soil texture on antimicrobial resistance gene levels in farmlands

and adjacent ditches. Sci Total Environ 737:139563.

39. Marti R, Tien Y-C, Murray R, Scott A, Sabourin L, Topp E. 2014. Safely coupling livestock and crop production systems: how rapidly do antibiotic resistance genes dissipate in soil following a commercial application of swine or dairy manure? Appl Environ Microbiol 80:3258–3265.

40. Miller E, Spiehs M, Arthur T, Woodbury B, Cortus E, Chatterjee A, Rahman S, Schmidt JW. 2019. Cropland Amendment with beef cattle manure minimally affects antimicrobial resistance. J Environ Qual 48:1683–1693.

41. Nolvak H, Truu M, Kanger K, Tampere M, Espenberg M, Loit E, Raave H, Truu J. 2016. Inorganic and organic fertilizers impact the abundance and proportion of antibiotic resistance and integron-integrase genes in agricultural grassland soil. Sci Total Environ 562:678–689.

42. Peng S, Feng Y, Wang Y, Guo X, Chu H, Lin X. 2017. Prevalence of antibiotic resistance genes in soils after continually applied with different manure for 30 years. J Hazard Mater 340:16–25.

43. Tien Y-C, Li B, Zhang T, Scott A, Murray R, Sabourin L, Marti R, Topp E. 2017. Impact of dairy manure pre-application treatment on manure composition, soil dynamics of antibiotic resistance genes, and abundance of antibiotic-resistance genes on vegetables at harvest. Sci Total Environ 581–582:32–39.

44. Udikovic-Kolic N, Wichmann F, Broderick NA, Handelsman J. 2014. Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. Proc Natl Acad Sci U S A 111:15202–15207.

45. Wepking C, Avera B, Badgley B, Barrett JE, Franklin J, Knowlton KF, Ray PP, Smitherman C, Strickland MS. 2017. Exposure to dairy manure leads to greater antibiotic resistance and increased mass-specific respiration in soil microbial communities. Proceedings Biol Sci 284.

46. Xi X, Zhang J, Kwok L, Huo D, Feng S, Zhang H, Sun T. 2015. Microbial Pollution Tracking of Dairy Farm with a Combined PCR-DGGE and qPCR Approach. Curr Microbiol 71:678–686.

47. Wallace JS, Garner E, Pruden A, Aga DS. 2018. Occurrence and transformation of veterinary antibiotics and antibiotic resistance genes in dairy manure treated by advanced anaerobic digestion and conventional treatment methods. Environ Pollut 236:764–772.

48. Yang H, Byelashov OA, Geornaras I, Goodridge LD, Nightingale KK, Belk KE, Smith GC, Sofos JN. 2010. Presence of antibiotic-resistant commensal bacteria in samples from agricultural, city, and national park environments evaluated by standard culture and real-time PCR methods. Can J Microbiol 56:761–770.

49. Zhang Y, Zhang C, Parker DB, Snow DD, Zhou Z, Li X. 2013. Occurrence of

antimicrobials and antimicrobial resistance genes in beef cattle storage ponds and swine treatment lagoons. Sci Total Environ 463–464:631–638.

50. Katakweba AAS, Moller KS, Muumba J, Muhairwa AP, Damborg P, Rosenkrantz JT, Minga UM, Mtambo MMA, Olsen JE. 2015. Antimicrobial resistance in faecal samples from buffalo, wildebeest and zebra grazing together with and without cattle in Tanzania. J Appl Microbiol 118:966–975.

51. Hurst JJ, Oliver JP, Schueler J, Gooch C, Lansing S, Crossette E, Wigginton K, Raskin L, Aga DS, Sassoubre LM. 2019. Trends in Antimicrobial Resistance Genes in Manure Blend Pits and Long-Term Storage Across Dairy Farms with Comparisons to Antimicrobial Usage and Residual Concentrations. Environ Sci Technol 53:2405–2415.

52. Ji X, Shen Q, Liu F, Ma J, Xu G, Wang Y, Wu M. 2012. Antibiotic resistance gene abundances associated with antibiotics and heavy metals in animal manures and agricultural soils adjacent to feedlots in Shanghai; China. J Hazard Mater 235–236:178–185.

53. McKinney CW, Loftin KA, Meyer MT, Davis JG, Pruden A. 2010. tet and sul antibiotic resistance genes in livestock lagoons of various operation type, configuration, and antibiotic occurrence. Environ Sci Technol 44:6102–6109.

54. Mu Q, Li J, Sun Y, Mao D, Wang Q, Luo Y. 2015. Occurrence of sulfonamide-, tetracycline-, plasmid-mediated quinolone- and macrolide-resistance genes in livestock feedlots in Northern China. Environ Sci Pollut Res Int 22:6932–6940.

55. Munir M, Xagoraraki I. 2011. Levels of antibiotic resistance genes in manure, biosolids, and fertilized soil. J Environ Qual 40:248–255.

56. Chen B, Hao L, Guo X, Wang N, Ye B. 2015. Prevalence of antibiotic resistance genes of wastewater and surface water in livestock farms of Jiangsu Province, China. Environ Sci Pollut Res Int 22:13950–13959.

57. Netthisinghe AMP, Cook KL, Gilfillen RA, Woosley PB, Kingery T, Sistani KR. 2018. Managing beef backgrounding residual soil contaminants by alum and biochar amendments. J Environ Qual 47:1275–1283.

58. Ruuskanen M, Muurinen J, Meierjohan A, Parnanen K, Tamminen M, Lyra C, Kronberg L, Virta M. 2016. Fertilizing with Animal Manure Disseminates Antibiotic Resistance Genes to the Farm Environment. J Environ Qual 45:488–493.

59. Storteboom HN, Kim S-C, Doesken KC, Carlson KH, Davis JG, Pruden A. 2007. Response of antibiotics and resistance genes to high-intensity and low-intensity manure management. J Environ Qual 36:1695–1703.

Chapter 3

Metagenomic Quantification of Genes with Internal Standards

ABSTRACT

We demonstrate that an assembly-independent and spike-in facilitated metagenomic

quantification approach can be used to screen and quantify over 2,000 genes simultaneously,

while delivering absolute gene concentrations comparable to quantitative PCR (qPCR). DNA

extracted from dairy manure slurry, digestate, and compost was spiked with genomic DNA from

a marine bacterium and sequenced using the Illumina HiSeq4000. We compared gene copy

concentrations, in gene copies per mass of sample, of five ARGs generated with (i) our

quantitative metagenomics approach, (ii) targeted qPCR, and (iii) a hybrid quantification

approach involving metagenomics and qPCR-based 16S rRNA gene quantification. Although

qPCR achieved lower quantification limits, the metagenomic method avoided biases caused by

primer specificity inherent to qPCR-based methods and was able to detect orders of magnitude

more genes than is possible with qPCR assays. We used the approach to simultaneously quantify

ARGs in the Comprehensive Antimicrobial Resistance Database (CARD). We observed that the

total abundance of tetracycline resistance genes was consistent across different stages of manure

treatment on three farms, but different samples were dominated by different tetracycline

resistance gene families.

IMPORTANCE

qPCR and metagenomics are central molecular techniques that have offered insights into biological processes for decades, from monitoring spatial and temporal gene dynamics to tracking ARGs or pathogens. Still needed is a tool that can quantify thousands of relevant genes in a sample as gene copies per sample mass or volume. We compare a quantitative metagenomic approach with traditional qPCR approaches in the quantification of ARG targets in dairy manure samples. By leveraging the benefits of non-targeted community genomics, we demonstrate high-throughput absolute gene quantification of all known ARG sequences in environmental samples.

3.1 Introduction

A high-throughput, quantitative, gene-level screening tool is needed for studying dynamic, complex, and diverse microbial communities and the biological processes they perform. Quantitative PCR (qPCR) is widely used to measure the absolute concentrations of short segments of nucleic acid sequences, which serve as proxies of organisms or genes. This approach has been critical in a wide range of applications including assessing the relative roles of different microorganisms in nitrification and denitrification in wastewater treatment,[1] the abundances of viruses in wastewater following outbreaks,[2] and the impact of antibiotic use on antimicrobial resistance gene (ARG) abundances in livestock manure.[3] However, qPCR is only capable of targeting limited sequences at a time and primer bias, sensitivity, and specificity can confound results.[4,5] These aspects of qPCR limit our ability to compare, between samples and studies, the composition of microorganisms and genes in their community context.

Metagenomic sequencing has emerged as a powerful tool to study the structure and functional capacity of microbial communities. Metabolic gene databases, such as Carbohydrate-Active enZYmes Database (CAZy), have facilitated gene classification from metagenomes for diverse applications ranging from evaluating the gut microbiome colonization in infants[6] to studying enrichment of cellulases in bioreactors for bioenergy production.[7] Virulence gene

38

databases, such as the Virulence Factor Database (VFDB), have enabled the development of metagenomic pathogen screening tools applied in a variety of settings from food safety to wastewater.[8,9] Although metagenomic analyses provide a comprehensive inventory of the genes and organisms that are present in samples, the data are compositional and results are typically reported as relative abundances. As a result, studies that perform metagenomic sequencing alone cannot report absolute gene abundances, which are essential in many studies; particularly those evaluating changes in a pathogen marker gene or ARG concentrations through food, water, waste, and air treatment processes. Instead, metagenomic studies are limited to reporting relative changes in community diversity or the enrichment of certain genes between samples by normalizing to total sequence reads[10], 16S rRNA gene reads,[11] or single copy gene reads.[12] In some cases, hybrid approaches convert relative abundance data from metagenomic analyses to absolute abundances by relying on ancillary analyses such as the number of cells measured by flow cytometry[13] or the number of 16S rRNA genes measured by qPCR.[14,15] These additional analyses require method optimization and can introduce biases. A more direct option for obtaining the absolute abundance of genes from metagenomic data involves spiking nucleic acid internal standards into samples before extraction or sequencing.[16]

Incorporating internal standard spike-ins, as commonly used in analytical chemistry, can establish a ratio of metagenomic read abundance to gene copy concentration. Internal standard protocols were first applied to sequencing methods in transcriptomics experiments (RNA-seq) to quantify gene expression, identify protocol-dependent biases, and compare method sensitivity and reproducibility.[17] Since then, protocols have been developed for 16S rRNA gene-amplicon[18] metagenome,[19] and metatranscriptome[16] sequencing. Previous quantitative metagenomic spike-in studies have performed metagenome assemblies, then mapped short metagenomic reads to the

assembled contigs[20]. Such assembly-dependent methods are time-intensive and can fail to assemble genomes that harbor ARGs, particularly those of viruses[21] or plasmids and within genomic islands,[22,23] thus increasing false negative detection rates. Additionally, assemblies can introduce bias towards highly abundant organisms, which are more likely to be assembled correctly.[24]

In this study, we applied an assembly-independent, spike-in metagenomic approach for quantifying gene concentrations in environmental samples. We first quantified the recovery of the spike-in genes across different concentrations, %G+C contents, and gene sizes. We then compared the spike-in quantitative metagenomic approach with traditional gene quantification by qPCR and with a hybrid, spike-independent metagenomic method. Finally, we applied the approach to quantify ARG concentrations in dairy farm samples and demonstrated the benefit of quantifying broader groups of genes than is possible with targeted qPCR methods. Ultimately, we envision this high-throughput, quantitative, gene-targeted method will improve exposure and risk assessment modeling, optimize treatment processes for water, waste, and air, enhance microbiome-driven resource recovery or bioenergy production, and quantify the roles of microbes in host health and global nutrient and carbon cycling.

3.2 Results
3.2.1 Equation for assembly-independent, absolute gene quantification using spike-in normalization.

Genomic DNA of a marine bacterium, *Marinobacter hydrocarbonoclasticus* (ATCC® 700491™) was spiked into DNA extracted from environmental samples to determine the relationship between read counts and gene copy concentrations (Fig. B-1). We used genomic DNA from *M. hydrocarbonoclasticus* as our spike-in DNA because it is a marine microbe foreign to our samples. In our study, DNA was spiked after extraction to ensure that differences

between spike-in and sample DNA recoveries were limited to sequencing and read-mapping

biases only, rather than to biases introduced during any prior sample processing steps. Prior to

performing the DNA extraction and spike-ins on the samples compared in this study, we

assessed extraction recoveries and bias across different manure matrices using a gram-positive

and a gram-negative bacterium (Supplemental Text B-1, Fig. B-2). Mean recoveries of spike-in

gram-positive and gram-negative organisms ranged from 75-110% and did not differ

significantly (p-value = 0.27, Fig. B-2).

The average ratio of the known spike-in gene copy concentration to gene length-

normalized counts of mapped reads was calculated. This ratio was defined as the spike-in

normalization factor, η:

$$\eta = \frac{1}{n}\sum_{i}^{n}\frac{c_{s,i}}{z_{s,i}/_{L_{s,i}}}$$ Eqn. 1

Where *n*, is the total number of genes in the *M. hydrocarbonoclasticus* genome, $c_{s,i}$, is

the known spike-in gene copy concentration for each gene *i* in the *M. hydrocarbonoclasticus*

genome (gene copies/µL DNA extract) and $\frac{z_{s,i}}{L_{s,i}}$ is the length-normalized read count

(reads/basepair) for gene *i*. In this approach, we assume the relationships between gene copy

concentrations and length-normalized read counts are consistent between the target genes and

spike-in genes. We confirmed the gene recovery was robust across gene lengths and %G+C

contents and different spike-in gene abundances, by observing read-mapping rates using different

tools and settings (Supplemental Text B-2, Fig B-3, Fig B-4).

By multiplying η by the target gene's length-normalized read counts ($\frac{z_t}{L_t}$, reads/basepair), we can predict the unknown concentration of our target gene ($\widehat{c}_t$, gene copies/volume of DNA extract):

$$\widehat{c}_t = \eta \cdot \frac{z_t}{L_t}$$

Eqn. 2

However, we ultimately aimed to determine the number of copies of the target gene per mass or volume of sample. For this, the target gene concentration was multiplied by the volume eluted during DNA extraction (V_{eluted}, μL) to obtain the total copies of the target gene extracted, which was then divided by the mass (or volume) of the sample extracted:

$$\frac{copies,target}{Sample\ mass} = \widehat{c}_t \cdot \frac{V_{eluted}}{Sample\ mass}$$

Eqn. 3

Here, it is assumed that spike-in genes are recovered at the same rate as the target genes in the sample. The spike-in facilitated approach establishes a relationship between read abundances and gene concentrations; we are therefore able to directly compare gene abundances between samples without corrections for average genome sizes or single copy gene concentrations.

Lastly, we found that the dynamic range of the relationship between gene concentration and read abundance spanned over three orders of magnitude and was consistent over different sequencing depths by spiking aliquots of a sample with different concentrations of the internal standard (Fig. B-4). We found that the limit of detection corresponded to about 3×10^4 gene copies/mg sample (Supplemental Text B-2).

3.2.2 Agreement between sequencing- and spike-independent approaches validates our method.

We compared gene quantities measured with the spike-in quantitative metagenomic approach to those measured with qPCR and a hybrid, spike-independent metagenomic quantification approach. We used six manure samples from different farms and treatment stages

(untreated, composted, or digested). Five target genes, *tetM, tetG, sul1, sul2,* and *ErmB,* were chosen because they have been frequently quantified in environmental samples and primer sets are available.[25–29] In the quantitative metagenomics approach, reads were assigned to ARGs in the Comprehensive Antimicrobial Resistance Database (CARD) using Graphing Resistance Out Of meTagenomes or "GROOT" (Figure 3-1A).[30] Additionally, read abundances were assigned to resistance genes in the MEGARes database using AMR++ (Fig. B-5).[31]
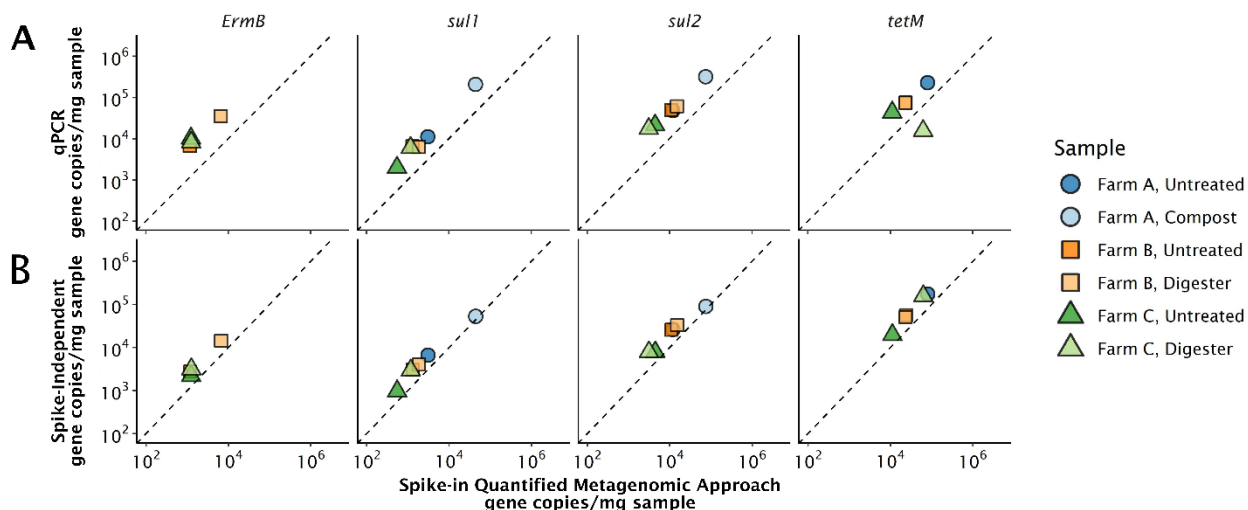
Figure 3-1: *Comparisons of the gene quantification approaches using GROOT for assigning reads to resistance genes. The dotted line is a 1:1 line that represents theoretical perfect correlation between approaches.*
A. *Spike-in-quantified metagenomic absolute abundance approach vs. qPCR;* ***B.*** *Spike-in-quantified metagenomic absolute abundance approach vs. spike-independent, 16S rRNA gene-based metagenomic approach. ErmB was not detected in the Farm A Samples with the quantitative metagenomic approach, but was detected in the Farm A Compost sample with qPCR. TetG was detected with qPCR in all samples, but not with the quantitative metagenome approach.*

In the hybrid, spike-independent metagenomic quantification approach, 16S rRNA gene concentrations are measured in each sample using qPCR. Then, target read counts are divided by 16S rRNA gene read counts.[14,15] This approach assumes that the target gene/16S rRNA gene quotient is equivalent for metagenomic sequencing and qPCR and is computed as follows:

$$\frac{target\ gene\ copies}{sample\ mass} = \frac{z_{s,i}}{z_{16S\ rRNA}} \cdot \frac{16S\ rRNA\ gene\ copies}{sample\ mass}$$

where $z_{s,i}$ is the number of reads mapping to the target gene, $z_{16S\ rRNA}$ is the number of reads

mapping to a 16S rRNA sequence, and the 16S rRNA gene copies/sample mass ratio is

determined by qPCR.

TetG was not detected using the metagenomics approach in any of the samples but was

detected with qPCR with abundances ranging from 1,000-2,400 copies/mg of sample,

corresponding to about 2,000-6,000 copies/µL extract. *ErmB* was not detected with the

metagenomic approach in the Farm A samples, but was detected in the Farm A Untreated sample

with qPCR at 900 copies/mg sample. In these samples, we are approaching the method detection

limit of the quantitative metagenomic approach ($\sim 3 \times 10^4$ gene copies/mg sample), but not that of

qPCR (4-24 gene copies/µL mg sample, Supplemental Text B-2, Table B-1).

On average, qPCR quantities were 22% greater that of those using the quantitative

metagenomic approach, with *tetM* as a visible outlier. Specifically, the spike-in quantitative

metagenomic approach predicted a four-fold greater concentration of *tetM* than qPCR in the

Farm C Digester. This discrepancy between approaches for *tetM* could result in major

differences in study conclusions. For example, if the qPCR assay was used to measure how *tetM*

concentrations changed between the Farm C untreated and digester manures, one would observe

a 95% decrease in *tetM* concentration. Using the spike-in quantitative metagenomic approach,

however, one would observe a 140% higher concentration of *tetM* in the digester sample than the

untreated sample (Figure 3-1A) These patterns were also observed with AMR++ using the

MEGARes database (Fig. B-5).

GROOT and AMR++ tools use different approaches to reduce ambiguous mapping to

resistance genes. Specifically, AMR++ employs the "ResistomeAnalyzer" algorithm which

removes sparse alignments at thresholds that can be set by the user. GROOT uses variation graph

representation of a user-specified database that stores shared gene sequences as graphical nodes,

reducing ambiguous mapping to homologous gene regions. We therefore hypothesized that the

incongruence observed for the *tetM* gene was due to qPCR primers failing to capture the diverse

tetM genes in the digester manure sample. To test this hypothesis, we evaluated the read

mapping patterns to *tetM* and *sul1* reference genes using Bowtie2[32] and the Integrative Genome

Viewer software.[33] We included *sul1* in the analysis because it is a highly conserved gene

sequence; we therefore expected mapped reads to perfectly match the reference gene sequence.

Six single nucleotide variants (SNVs) were observed in the reads mapping to the 19 basepair

tetM forward primer sequence. When the allele frequencies at the primer binding sites were

quantified (Table B-2), 99% of *tetM*-mapped reads from the Farm C Digester had a mismatch at

five of the six SNVs in the primer sequence. In the other samples, between 60 and 80% of the

mapped reads had a mismatch at the same primer SNVs (Table B-2). Incongruencies in primer

binding sites and metagenomic reads were not observed for the *sul1* primer set. This analysis

demonstrates that the *tetM* primers likely did not capture the diversity of this gene. As a result,

the *tetM* qPCR assay underestimated *tetM* concentrations, especially in the Farm C Digester

sample. In other words, the spike-in quantitative metagenomic method resulted in more reliable

absolute *tetM* abundances than qPCR because it did not rely on primer design and *a priori*

knowledge of sequence diversity.

To further assess the reproducibility of the spike-in quantitative metagenomic method, we

compared the estimated absolute concentrations to a hybrid, spike-independent, 16S rRNA gene-

based quantitative metagenomic approach (Figure 3-1B). The spike-independent approach has

been used previously for absolute quantification of ARGs in a river system[34] and markers of

opportunistic pathogens in a drinking water distribution system.[15] Percent differences between

the hybrid, rRNA gene-based approach and the spike-dependent quantitative approach are a

function of the two normalization factors, since the number of reads mapping to each target are

the same for each approach. The limits of detection are the same for these approaches since they

are both determined by read counts. Percent differences ranged from 13% to 21% between the

spike-independent and spike-dependent approach, except for the Farm A compost sample, which

had a 4% difference (Figure 3-1B).

Cross-validating our approach to sequencing-independent qPCR assays and a hybrid,

spike-independent metagenomic approach for the five ARG targets validated that our method

generates values comparable to established gene quantification tools. Although the spike-in

metagenomic approach had higher detection limits than qPCR, it overcomes biases caused by

primer specificity.

3.2.3 Spike-in metagenomic approach facilitates quantitative screening of diverse gene families.

In six dairy farm manure samples, the spike-in metagenomic approach enabled the

quantification of all genes in CARD. This is not feasible using traditional qPCR since each gene

would require a validated set of primers and standard curves. Out of the 2,617 genes in CARD,

411 genes were detected in the six dairy manure samples using GROOT. The total number of

different ARGs detected in each sample ranged from 62 to 361.
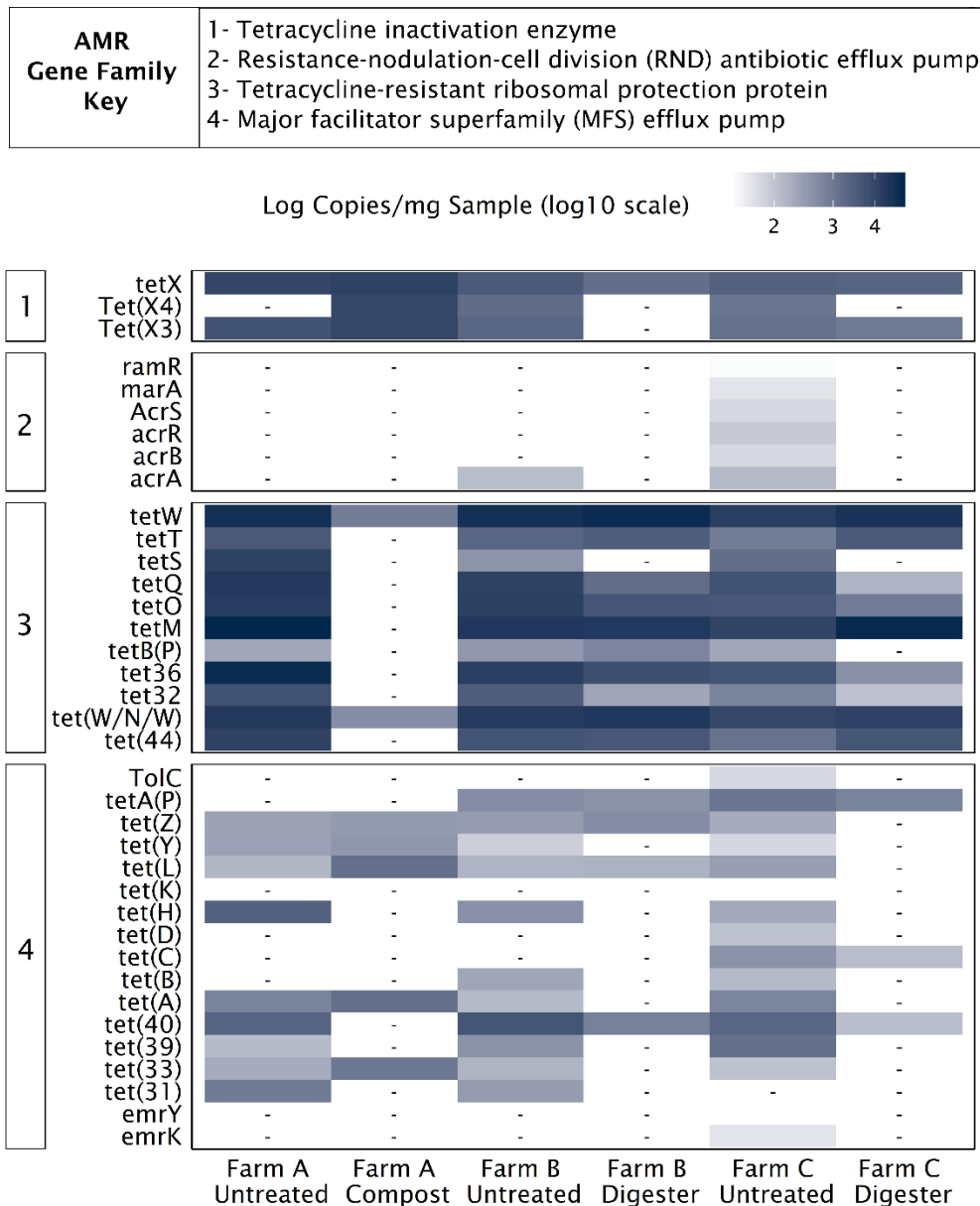
Figure 3-2*: Absolute concentrations of all genes that confer resistance to tetracycline in six different dairy farm samples.*
Dashes ("-") indicate non-detects (no reads mapping to target).

To illustrate the diversity of genes that can be detected and quantified with a single approach, we leveraged the "confers_resistance_to_antibiotic" relation in the CARD ontology to extract genes within the tetracycline drug class. We then assessed the diversity and absolute

abundance of these tetracycline resistance genes across broad gene families in each sample

(Figure 3-2). Of the 93 tetracycline resistance genes matching the

"confers_resistance_to_antibiotic" ontology, 37 were detected in our samples with genes

detected from four of the seven resistance gene families (Figure 3-2). Genes were not detected

from the small multidrug resistance antibiotic efflux pump, ATP-binding cassette ribosomal

protection protein, and ATP-binding cassette efflux pump families in the manure samples.

Interestingly, the sums of all the tetracycline resistance gene concentrations in each sample were

all within one order of magnitude, ranging from $3.8x10^4$ - $3x10^5$ copies per mg sample (Table 3-

1). However, different resistance gene families dominated the tetracycline resistance gene

concentrations within different sample groups. Tetracycline resistance ribosomal protection

proteins were the most abundant gene family, comprising 82-97% of the total tetracycline

resistance genes in all samples except the compost sample from Farm A. In this sample,

tetracycline inactivation enzymes dominated the resistance profile, comprising 85% of the total

tetracycline resistance gene concentrations (Figure 3-2; Table 3-1). These data demonstrate that

no single ARG could have been selected to represent the total tetracycline resistance abundances.

For example, targeting just one or two tetracycline resistance genes with qPCR would have

inadequately assessed the impact of residual concentrations of tetracyclines or how effective a

manure treatment strategy had been at reducing the quantity of resistance genes within a drug

class.

Table 3-1: *Total abundance of tetracycline resistance genes, organized by CARD Gene Family, in six different dairy farm samples.*

AMR Gene Family	Total abundance of tetracycline resistance genes (copies/mg sample)					
	Farm A, Untreated	Farm A, Compost	Farm B, Untreated	Farm B, Digester	Farm C, Untreated	Farm C, Digester
major facilitator superfamily (MFS) antibiotic efflux pump	$7.3x10^3$	$4.5x10^3$	$7.3x10^3$	$1.7x10^3$	$6.8x10^3$	9.1E+02

resistance-nodulation-cell division (RND) antibiotic efflux pump	-	-	1.2×10^{2}	-	4.4×10^{2}	-
tetracycline-resistant ribosomal protection protein	2.7×10^{5}	1.3×10^{5}	1.4×10^{5}	1.2×10^{5}	5.6×10^{4}	1.2×10^{5}
tetracycline inactivation enzyme	1.6×10^{4}	3.2×10^{4}	7.3×10^{3}	1.5×10^{3}	5.1×10^{3}	3.2×10^{3}
Total gene abundance	**3.0E+05**	**3.8E+04**	**1.5E+05**	**1.3E+05**	**6.8E+04**	**1.2E+05**

RND = resistance-nodulation-cell division, MFS = major facilitator superfamily, RPP = ribosomal protection protein

2.3.4 Changes in spike-in-based absolute abundances and relative abundances through treatment.

Our samples comprised untreated and treated manure samples from three dairy farms; two treated samples were collected from anaerobic digesters and one treated sample consisted of compost. Thus, this study provided an opportunity to evaluate the degree to which between-sample relationships in gene levels compared between the relative and absolute quantities.

The number of metagenomic reads mapping to a target gene is used to determine both relative and absolute abundances of that gene in the sample. However, the normalization parameter is different between approaches. The simplest normalization parameter is the library size, or total number of reads generated in a sequencing run, though the approach poorly resolves log-fold changes between samples.[19] Another normalization parameter is the total number of reads mapping to 16S rRNA or single copy genes.[35] These relative abundance approaches approximate the abundance of reads relative to bacterial and archaeal biomass. In contrast, the normalization factor in the absolute abundances spike-in approach derives from the relationship between gene concentration and read abundance established by a spike-in standard. The normalization parameter is the product of all terms in Equation 3, save the number of reads

mapping to the target. Between-sample comparisons in ARG abundances quantified by both relative and absolute normalization demonstrated that different approaches can predict conflicting directionality of change for genes for which the change in abundance is small (Figure 3-3). In the sample pairs from farms A, B, and C, there were 42, 71, and 49 total genes detected in both samples, respectively. Between the untreated and treated sample pairs on farms A, B, and C, conflicting directionally of the change in abundance was observed in three, 17, and seven of those ARGs; specifically, decreases in absolute abundance were observed, whereas increases in 16S rRNA gene-normalized relative abundances were observed. The 1:1 correlation between observed changes in absolute abundance and relative abundance demonstrates that both approaches are functions of the reads recruited per target sequence. The intercept depends on the normalization parameter values for each sample. Plotting log-fold changes in absolute gene abundances versus the library size normalized relative abundances (Figure 3-3, blue triangles) resulted in a y-intercept greater than that of the absolute abundances versus the rRNA gene read normalized abundances (Figure 3-3, red squares). A greater intercept means more ARGs fell within the quadrant II for the library-size normalization approach, therefore more ARGs were observed to increase in relative abundances while decreasing in absolute abundance. For those examples where the direction of the log2-fold changes disagreed between relative and absolute abundances approaches, all changes were less than four-fold. When observed between-sample differences in abundance are small (log2-fold changes < 2), distinguishing between abundance increases and decreases becomes noisier and more challenging to resolve.

Relative Normalization Approach ■ 16S rRNA Gene Reads ▲ Reads per Million Reads
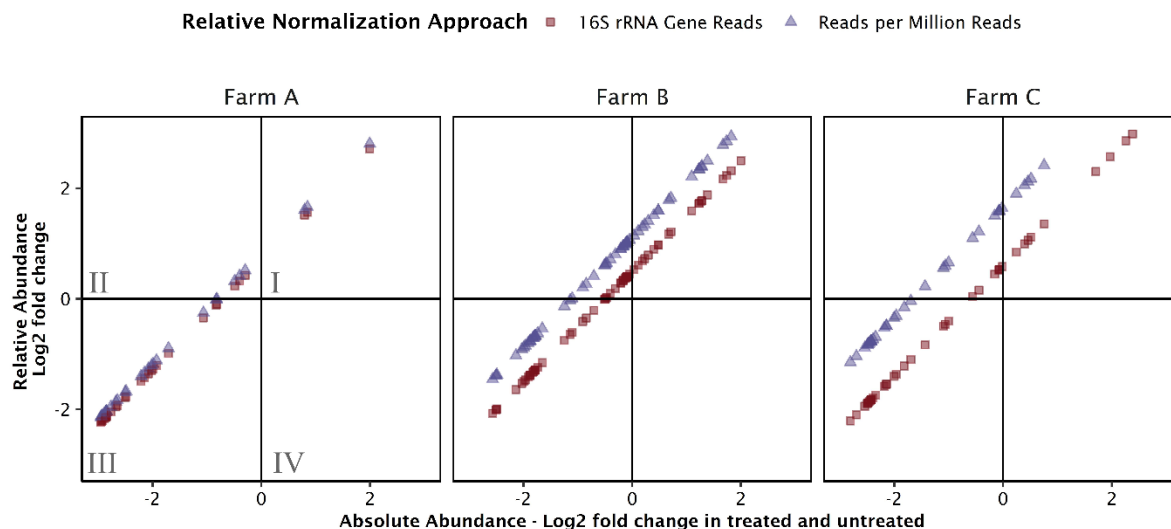
Figure 3-3: Comparing relative vs. absolute abundances of resistance genes between treated and untreated manure samples detected with GROOT.
Each point represents the log2 fold change in abundance of a gene between untreated and treated samples. Red squares compare fold changes determined by relative abundances normalized by 16S rRNA gene reads to our fold changes in absolute abundance using the spike-in quantitative metagenomic approach. Blue triangles compare fold changes in relative abundance normalized by library size to fold changes in absolute abundances. If a point falls in quadrants II (quadrants labeled in Farm A plot), a positive log2-fold change was observed with relative abundance and a negative log2 fold change was observed with absolute abundance.

3.3 Discussion

We demonstrated that assembly-independent, gene-targeted metagenome quantification with a genomic spike-in internal standard resulted in absolute gene quantities for thousands of genes simultaneously and at levels comparable to those obtained with qPCR and a spike-independent hybrid metagenomic approach. The spike-in approach circumvents the low throughput and primer design challenges of qPCR and the bias-prone ancillary molecular methods required for hybrid approaches. Our approach requires only a spike-in internal DNA standard and a relevant gene database, such as CARD, CAZy, or a custom gene set of interest.

Previous quantitative sequencing studies have spiked DNA into samples prior to extraction,[18,20] into samples after cell lysis,[36] and into nucleic acid extracts.[19] Cellular internal standard spike-ins enable estimation of recovery from cell lysis to sequencing. However, each

organism has a unique cellular morphology and susceptibility to lysis, so one or a handful of cellular spike-ins may not reflect extraction efficiencies for the diverse organisms in a sample.[37] We spiked in *M. hydrocarbonoclasticus* genomic DNA into the DNA extracted from environmental samples prior to sequencing to directly observe spike-in recovery in the absence of these complex biases and compared extraction efficiently in separate experiments (Supplemental Text B-1). We note that the same nucleic acid extraction efficiency issues are present with any gene quantification method.

The number of documented and catalogued genes will continue to grow as new samples are sequenced, as isolation and culturing of environmental strains reveals more diversity, and as organisms evolve. qPCR primers are designed for gene targets using the gene sequences available in a database at a specific point in time. The primers are often applied in future studies. Consequently, without constant primer redesign and evaluation, qPCR quantification methods lag novel gene diversity discovery. Spike-in metagenomic quantification, on the other hand, can provide absolute quantification comparable to qPCR while using the most up-to-date gene databases. Furthermore, archived metagenomes can be reanalyzed to quantify newly discovered genes as databases expand.

The number of genes that can be measured simultaneously in a sample using spike-in metagenomic quantification is limited by database completeness. In our study, we were able to simultaneously screen the 2,617 genes in CARD. qPCR assays, on the other hand, are limited by the number of targets they can include. ARG studies that employ qPCR, for example, commonly target between five to 20 genes per sample.[26,38–40] Although qPCR arrays have increased the throughput of qPCR and provided valuable insights on ARG profiles,[28,41] qPCR arrays without standard curves do not deliver absolute concentrations, are subject to the same primer specificity

limitations as traditional qPCR, and are similarly limited by available knowledge on ARGs at the time of qPCR array design.

Determining detection limits and establishing detection ranges are critical for all quantitative methods. Here, spiking at different internal standard concentrations revealed that the linear range of detection spanned at least three orders of magnitude (Supplemental Text B-2). Spiking a standard corresponding to 0.1% of the total DNA in our sample revealed that we were reaching limits of detection for several genes at the order of 10^4 gene copies/mg sample. The single genome spike-in used in this study meant that nearly all internal standard genes were present at the same gene copies concentration in a single sample. Staggered spike-in standards, which contain different sequences over a range of concentrations within one spike-in, can better characterize the quantification range in individual samples.[18,19,36] Limits of detection, which are trumped by qPCR assays, are a primary limitation of this quantitative approach.

Our spike-in genome presented non-specific mapping in regions of the *M. hydrocarbonoclasticus* genome that were homologous to genes of interest in the sample. For example, a gene in the *M. hydrocarbonoclasticus* genome shared 75% homology with an efflux-pump encoding gene in CARD. Interestingly, *M. hydrocarbonoclasticus* also shared >70% homology with three genes in the bovine genome. Synthetic DNA internal standards, as opposed to genomic standards, can eliminate non-specific mapping.[19] Similarly, assigning metagenomic reads to target genes of interest within databases can also result in false positive and false negative assignments. We emphasize that our approach is intended for high-throughput screening and is not appropriate for exploring and quantifying potential novel resistance genotypes that are not yet represented in databases. Other tools that leverage machine learning and functional

metagenomics[42] or build models for specific genes[43] would be more appropriate in these applications.

This assembly-independent, spike-in-facilitated gene quantification is a fast, effective, and non-targeted approach to quantify known genes from microbial communities. This approach will be valuable when qPCR throughput and primer design limit the conclusions that can be drawn and when quantifying genes at low abundances is not required. The approach is especially useful in ARG research, where absolute quantification of diverse genes is imperative for evaluating technologies to reduce ARG abundances and informing models focused on antimicrobial resistance risk assessment.

3.4 Materials and Methods

3.4.1 Sample Collection

One-hour composite samples were collected in June of 2016 from dairy manure at three farms in New York following a protocol described previously.[28] Samples included an untreated manure sample from blend pits at each farm and a post-treatment sample, either anaerobic digester effluent or compost. Samples were aliquoted into 15 mL centrifuge tubes, frozen at -80 °C, and shipped overnight on dry ice to the University of Michigan.

3.4.2 DNA Extraction, Internal Standard Spike-in, and Sequencing

DNA was extracted from approximately 250 mg (wet weight) of each sample in duplicate reactions using the QiaAMP PowerFecal kit (QIAGEN, Germantown, MD). Extracted DNA was eluted in 100 µL of elution buffer following the kit's protocol. Duplicate extractions were pooled. DNA concentrations were measured with a Qubit 2.0 Fluorometer. The pooled DNA extracts were spiked with genomic DNA from *M. hydrocarbonoclasticus* (ATCC strain 700491D5™, GenBank AN: CP000514), obtained from ATCC (Manassas, VA) at 1% total DNA by mass. *M. hydrocarbonoclasticus* was resuspended following ATCC recommendations

in molecular grade water and the concentration was confirmed using the Nanodrop1000 instrument. This marine bacterium was selected due to its typical bacterial genome size (4,326,849 basepairs) and GC content (57%) and because it was unlikely to be present in manure samples. All six pooled and spiked DNA samples were sequenced with paired-end, Illumina (HiSeq4000) technology at the University of Michigan DNA Sequencing Core using PCR-free library preparation with a read length of 250 basepairs and an insert size of 450 basepairs. Total post-QC reads per sample ranged from $5.1\text{-}6.3 \times 10^8$. To establish the linear quantification range of genes, replicates of one of the samples were spiked with different ratios of internal standard DNA to total community DNA (0.1%, 1% and 10%, by mass).

3.4.3 Bioinformatic Approaches

Reads were trimmed and checked for quality with BBDuk from the BBTools Package.[44] The Comprehensive Antibiotic Resistance Gene Database (CARD),[45] the *M. hydrocarbonoclasticus* gene multifasta files from NCBI, and 16S rRNA gene-specific small subunit (SSU) SILVA database,[46] and the bovine genome[47] were downloaded on December 20, 2019, October 24, 2018, May 24, 2019, and November 2, 2018, respectively. Read mapping approaches were first evaluated by comparing observed to expected recoveries of *M. hydrocarbonoclasticus* genes. Specifically, Bowtie2[32] (version 2.3.4.3) was run with default parameters and kallisto[47] (version 0.46.0) was run with the "--single overhang" setting which counts reads only partially mapping to the end of reads and the bias-correction setting "--bias" that can reduce the bias from larger reference genes where the effects of overhanging reads are less impactful. These tools were selected because they represent two common algorithms for fast short-read mapping, pseudoalignment and Borrows-Wheeler transform-based read assignments. Read assignments were performed with both paired reads and unpaired reads. Average recovery across read lengths was assessed by clustering genes into 20 bins using quantile binning with the

Hmisc package in R. One-way analysis of variance (ANOVA) was performed in R to compare distribution of average spike-in gene recoveries. After validating read mapping performance with the *M. hydrocarbonoclasticus* genes, kallisto in an unpaired mode with the "--single overhang" and "--bias" setting was selected to map reads to CARD and 16S rRNA gene-specific small subunit (SSU) SILVA database. Reads were mapped to the bovine genome with bbmap from the BBTools Package in the paired setting to assess host contamination. IGV software (version 2.7.2) was used to visual read pile-ups to assess nucleotide variants at primer binding sites.

Two additional tools were used to assign reads to ARGs: (i) ARG-specific read assignment tools for antimicrobial resistance genes, AMR++[31] and GROOT.[30] Both tools apply unique approaches to reduce non-specific mapping of reads to ARGs. GROOT was run with default parameters using CARD2020 as a reference except that the 50% reporting threshold was used. AMR++ (version 2.0.0) was performed using the singularity container with the MEGARes database as a reference and default parameters (--threshold 80 --min 1 --max 100 --samples 5 --skip 5).

3.4.4 qPCR Primer Selection and Design

ARG targets were chosen based on initial metagenomic results to capture a range of ARG concentrations. The ARG primer sets were selected based on their use for measuring ARGs in environmental samples[25,48–50]. The primer sets were verified for specificity using NCBI Primer-Blast and archaea, virus, viroid, and eukaryote databases. Details of the qPCR assays, including primer sequences and annealing temperatures are provided in Table B-1. qPCR reactions were carried out on an Eppendorf MasterCyler ep realplex[2] using Fast EvaGreen Fast Master Mix (Biotium, Fremont, CA). The 20 µL reactions were performed following the manufacturer's recommendations, with 0.4 µM of forward and reverse primers, 0.625 mg/mL of Ultrapure BSA (Invitrogen), and 2 µL of diluted DNA extracts. Plates were centrifuged for 2 minutes at 500

RPM at 4°C before thermocycling. Unpooled sample DNA extracts, with total DNA

concentrations ranging from 20-50 ng/µL, were diluted 10-fold and 100-fold to detect inhibition

from the sample matrix. Inhibition was not observed and the gene concentrations from both

dilutions were averaged. Two technical replicates were performed per diluted sample and two

no-template controls were conducted per plate. After amplification, melt curves were performed

to confirm the specificity of the reactions. The template for the standard curve consisted of

Gblock Fragments (IDT, Skokie, Illinois) with the inserted target sequences taken from a

sequence from CARD or NCBI if the primers did not hit the CARD reference gene (Table B-1).

The qPCR assay limit of detection and limit of quantification were evaluated for each of the

ARGs (Table B-1) following the European Network of GMO Laboratory Guidelines.[51]

3.5 References

(1)   Harms, G.; Layton, A. C.; Dionisi, H. M.; Gregory, I. R.; Garrett, V. M.; Hawkins, S. A.; Robinson, K. G.; Sayler, G. S. Real-Time PCR Quantification of Nitrifying Bacteria in a Municipal Wastewater Treatment Plant. *Environ. Sci. Technol.* **2003**, *37* (2), 343–351. https://doi.org/10.1021/es0257164.

(2)   Katayama, H.; Haramoto, E.; Oguma, K.; Yamashita, H.; Tajima, A.; Nakajima, H.; Ohgaki, S. One-Year Monthly Quantitative Survey of Noroviruses, Enteroviruses, and Adenoviruses in Wastewater Collected from Six Plants in Japan. *Water Res.* **2008**, *42* (6–7), 1441–1448. https://doi.org/10.1016/j.watres.2007.10.029.

(3)   Peak, N.; Knapp, C. W.; Yang, R. K.; Hanfelt, M. M.; Smith, M. S.; Aga, D. S.; Graham, D. W. Abundance of Six Tetracycline Resistance Genes in Wastewater Lagoons at Cattle Feedlots with Different Antibiotic Use Strategies. *Environ. Microbiol.* **2007**, *9* (1), 143–151. https://doi.org/10.1111/j.1462-2920.2006.01123.x.

(4)   Ouyang, Y.; Evans, S. E.; Friesen, M. L.; Tiemann, L. K. Effect of Nitrogen Fertilization on the Abundance of Nitrogen Cycling Genes in Agricultural Soils: A Meta-Analysis of Field Studies. *Soil Biol. Biochem.* **2018**, *127*, 71–78.

(5)   Ye, L.; Zhang, T.; Wang, T.; Fang, Z. Microbial Structures, Functions, and Metabolic Pathways in Wastewater Treatment Bioreactors Revealed Using High-Throughput Sequencing. *Environ. Sci. Technol.* **2012**, *46* (24), 13244–13252.

(6)   Bäckhed, F.; Roswall, J.; Peng, Y.; Feng, Q.; Jia, H.; Kovatcheva-Datchary, P.; Li, Y.; Xia, Y.; Xie, H.; Zhong, H.; et al. Dynamics and Stabilization of the Human Gut

Microbiome during the First Year of Life. *Cell Host Microbe* **2015**, *17* (5), 690–703. https://doi.org/10.1016/j.chom.2015.04.004.

(7)   Kanokratana, P.; Wongwilaiwalin, S.; Mhuantong, W.; Tangphatsornruang, S.; Eurwilaichitr, L.; Champreda, V. Characterization of Cellulolytic Microbial Consortium Enriched on Napier Grass Using Metagenomic Approaches. *J. Biosci. Bioeng.* **2018**, *125* (4), 439–447.

(8)   Cai, L.; Zhang, T. Detecting Human Bacterial Pathogens in Wastewater Treatment Plants by a High-Throughput Shotgun Sequencing Technique. *Environ. Sci. Technol.* **2013**, *47* (10), 5433–5441. https://doi.org/10.1021/es400275r.

(9)   Grützke, J.; Malorny, B.; Hammerl, J. A.; Busch, A.; Tausch, S. H.; Tomaso, H.; Deneke, C. Fishing in the Soup – Pathogen Detection in Food Safety Using Metabarcoding and Metagenomic Sequencing. *Front. Microbiol.* **2019**, *10* (AUG), 1–15. https://doi.org/10.3389/fmicb.2019.01805.

(10)   Cheng, H.; Guan, Q.; Villalobos, L. F.; Peinemann, K. V.; Pain, A.; Hong, P. Y. Understanding the Antifouling Mechanisms Related to Copper Oxide and Zinc Oxide Nanoparticles in Anaerobic Membrane Bioreactors. *Environ. Sci. Nano* **2019**, *6* (11), 3467–3479. https://doi.org/10.1039/c9en00872a.

(11)   Ng, C.; Tan, B.; Jiang, X. T.; Gu, X.; Chen, H.; Schmitz, B. W.; Haller, L.; Charles, F. R.; Zhang, T.; Gin, K. Metagenomic and Resistome Analysis of a Full-Scale Municipal Wastewater Treatment Plant in Singapore Containing Membrane Bioreactors. *Front. Microbiol.* **2019**, *10* (FEB), 1–13. https://doi.org/10.3389/fmicb.2019.00172.

(12)   Chu, B. T. T.; Petrovich, M. L.; Chaudhary, A.; Wright, D.; Murphy, B.; Wells, G.; Poretsky, R. Metagenomic Analysis Reveals the Impact of Wastewater Treatment Plants on the Dispersal of Microorganisms and Genes in Aquatic Sediments. *Appl. Environ. Microbiol.* **2017**, *84* (December), online first. https://doi.org/10.1128/AEM.02168-17.

(13)   Props, R.; Kerckhof, F. M.; Rubbens, P.; Vrieze, J. De; Sanabria, E. H.; Waegeman, W.; Monsieurs, P.; Hammes, F.; Boon, N. Absolute Quantification of Microbial Taxon Abundances. *ISME J.* **2017**, *11* (2), 584–587. https://doi.org/10.1038/ismej.2016.117.

(14)   Garner, E.; Wallace, J. S.; Argoty, G. A.; Wilkinson, C.; Fahrenfeld, N.; Heath, L. S.; Zhang, L.; Arabi, M.; Aga, D. S.; Pruden, A. Metagenomic Profiling of Historic Colorado Front Range Flood Impact on Distribution of Riverine Antibiotic Resistance Genes. *Sci. Rep.* **2016**, *6*, 38432.

(15)   Haig, S. J.; Kotlarz, N.; Lipuma, J. J.; Raskin, L. A High-Throughput Approach for Identification of Nontuberculous Mycobacteria in Drinking Water Reveals Relationship between Water Age and Mycobacterium Avium. *MBio* **2018**, *9* (1), 1–13. https://doi.org/10.1128/mBio.02354-17.

(16)   Satinsky, B. M.; Gifford, S. M.; Crump, B. C.; Moran, M. A. *Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis*, 1st ed.; Elsevier Inc.,

2013; Vol. 531. https://doi.org/10.1016/B978-0-12-407863-5.00012-5.

(17)   Jiang, L.; Schlesinger, F.; Davis, C. A.; Zhang, Y.; Li, R.; Salit, M.; Gingeras, T. R.; Oliver, B. Synthetic Spike-in Standards for RNA-Seq Experiments. *Genome Res.* **2011**, *21* (9), 1543–1551.

(18)   Stämmler, F.; Gläsner, J.; Hiergeist, A.; Holler, E.; Weber, D.; Oefner, P. J.; Gessner, A.; Spang, R. Adjusting Microbiome Profiles for Differences in Microbial Load by Spike-in Bacteria. *Microbiome* **2016**, *4*, 1–13. https://doi.org/10.1186/s40168-016-0175-0.

(19)   Hardwick, S. A.; Chen, W. Y.; Wong, T.; Kanakamedala, B. S.; Deveson, I. W.; Ongley, S. E.; Santini, N. S.; Marcellin, E.; Smith, M. A.; Nielsen, L. K.; et al. Synthetic Microbe Communities Provide Internal Reference Standards for Metagenome Sequencing and Analysis. *Nat. Commun.* **2018**, *9* (1), 1–10.

(20)   Satinsky, B. M.; Crump, B. C.; Smith, C. B.; Sharma, S.; Zielinski, B. L.; Doherty, M.; Meng, J.; Sun, S.; Medeiros, P. M.; Paul, J. H.; et al. Microspatial Gene Expression Patterns in the Amazon River Plume. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (30), 11085–11090. https://doi.org/10.1073/pnas.1402782111.

(21)   Beaulaurier, J.; Luo, E.; Eppley, J. M.; Den Uyl, P.; Dai, X.; Burger, A.; Turner, D. J.; Pendleton, M.; Juul, S.; Harrington, E.; et al. Assembly-Free Single-Molecule Sequencing Recovers Complete Virus Genomes from Natural Microbial Communities. *Genome Res.* **2020**, gr--251686.

(22)   Sheppard, A.; Stoesser, N.; Wilson, D. J.; Sebra, R.; Kasarskis, A.; Anson, L. W.; Giess, A.; Pankhurst, L. J.; Vaughan, A.; Grim, C. J.; et al. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene Bla KPC. **2016**, *60* (6), 3767–3778. https://doi.org/10.1128/AAC.00464-16.Address.

(23)   Clausen, P. T. L. C.; Zankari, E.; Aarestrup, F. M.; Lund, O. Benchmarking of Methods for Identification of Antimicrobial Resistance Genes in Bacterial Whole Genome Data. *J. Antimicrob. Chemother.* **2016**, *71* (9), 2484–2488.

(24)   Papudeshi, B.; Haggerty, J. M.; Doane, M.; Morris, M. M.; Walsh, K.; Beattie, D. T.; Pande, D.; Zaeri, P.; Silva, G. G. Z.; Thompson, F.; et al. Optimizing and Evaluating the Reconstruction of Metagenome-Assembled Microbial Genomes. *BMC Genomics* **2017**, *18* (1), 915.

(25)   Peng, S.; Feng, Y.; Wang, Y.; Guo, X.; Chu, H.; Lin, X. Prevalence of Antibiotic Resistance Genes in Soils after Continually Applied with Different Manure for 30 Years. *J. Hazard. Mater.* **2017**, *340*, 16–25. https://doi.org/10.1016/j.jhazmat.2017.06.059.

(26)   Jang, H. M.; Lee, J.; Kim, Y. B.; Jeon, J. H.; Shin, J.; Park, M.-R.; Kim, Y. M. Fate of Antibiotic Resistance Genes and Metal Resistance Genes during Thermophilic Aerobic Digestion of Sewage Sludge. *Bioresour. Technol.* **2018**, *249*, 635–643.

(27)   Pei, R.; Kim, S.-C.; Carlson, K. H.; Pruden, A. Effect of River Landscape on the Sediment

Concentrations of Antibiotics and Corresponding Antibiotic Resistance Genes (ARG). *Water Res.* **2006**, *40* (12), 2427–2435.

(28)   Hurst, J. J.; Oliver, J. P.; Schueler, J.; Gooch, C.; Lansing, S.; Crossette, E.; Wigginton, K.; Raskin, L.; Aga, D. S.; Sassoubre, L. M. Trends in Antimicrobial Resistance Genes in Manure Blend Pits and Long-Term Storage Across Dairy Farms with Comparisons to Antimicrobial Usage and Residual Concentrations. *Environ. Sci. Technol.* **2019**, *53* (5), 2405–2415. https://doi.org/10.1021/acs.est.8b05702.

(29)   Makowska, N.; Koczura, R.; Mokracka, J. Class 1 Integrase, Sulfonamide and Tetracycline Resistance Genes in Wastewater Treatment Plant and Surface Water. *Chemosphere* **2016**, *144*, 1665–1673.

(30)   Rowe, W. P. M.; Winn, M. D. Indexed Variation Graphs for Efficient and Accurate Resistome Profiling. *Bioinformatics* **2018**, *34* (21), 3601–3608.

(31)   Doster, E.; Lakin, S. M.; Dean, C. J.; Wolfe, C.; Young, J. G.; Boucher, C.; Belk, K. E.; Noyes, N. R.; Morley, P. S. MEGARes 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data. *Nucleic Acids Res.* **2020**, *48* (D1), D561--D569.

(32)   Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat Methods* **2012**, *9* (4), 357–359. https://doi.org/10.1038/nmeth.1923.

(33)   Thorvaldsdóttir, H.; Robinson, J. T.; Mesirov, J. P. Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* **2013**, *14* (2), 178–192.

(34)   Garner, E.; Wallace, J. S.; Argoty, G. A.; Wilkinson, C.; Fahrenfeld, N.; Heath, L. S.; Zhang, L.; Arabi, M.; Aga, D. S.; Pruden, A. Metagenomic Profiling of Historic Colorado Front Range Flood Impact on Distribution of Riverine Antibiotic Resistance Genes. *Sci. Rep.* **2016**, *6* (November), 1–10. https://doi.org/10.1038/srep38432.

(35)   Yin, X.; Jiang, X.-T.; Chai, B.; Li, L.; Yang, Y.; Cole, J. R.; Tiedje, J. M.; Zhang, T. ARGs-OAP v2. 0 with an Expanded SARG Database and Hidden Markov Models for Enhancement Characterization and Quantification of Antibiotic Resistance Genes in Environmental Metagenomes. *Bioinformatics* **2018**, *34* (13), 2263–2270.

(36)   Tourlousse, D. M.; Yoshiike, S.; Ohashi, A.; Matsukura, S.; Noda, N.; Sekiguchi, Y. Synthetic Spike-in Standards for High-Throughput 16S RRNA Gene Amplicon Sequencing. *Nucleic Acids Res.* **2017**, *45* (4), e23--e23.

(37)   Albertsen, M.; Hansen, L. B. S.; Saunders, A. M.; Nielsen, P. H.; Nielsen, K. L. A Metagenome of a Full-Scale Microbial Community Carrying out Enhanced Biological Phosphorus Removal. *ISME J.* **2012**, *6* (6), 1094–1106. https://doi.org/10.1038/ismej.2011.176.

(38)   Burch, T. R.; Sadowsky, M. J.; LaPara, T. M. Modeling the Fate of Antibiotic Resistance

Genes and Class 1 Integrons during Thermophilic Anaerobic Digestion of Municipal Wastewater Solids. *Appl. Microbiol. Biotechnol.* **2016**, *100* (3), 1437–1444. https://doi.org/10.1007/s00253-015-7043-x.

(39) Zhang, J.; Liu, J.; Lu, T.; Shen, P.; Zhong, H.; Tong, J.; Wei, Y. Fate of Antibiotic Resistance Genes during Anaerobic Digestion of Sewage Sludge: Role of Solids Retention Times in Different Configurations. *Bioresour. Technol.* **2018**, *274* (December 2018), 488–495. https://doi.org/10.1016/j.biortech.2018.12.008.

(40) Zarei-Baygi, A.; Harb, M.; Wang, P.; Stadler, L. B.; Smith, A. L. Evaluating Antibiotic Resistance Gene Correlations with Antibiotic Exposure Conditions in Anaerobic Membrane Bioreactors. *Environ. Sci. Technol.* **2019**, *53* (7), 3599–3609. https://doi.org/10.1021/acs.est.9b00798.

(41) Zhu, Y.-G.; Johnson, T. A.; Su, J.-Q.; Qiao, M.; Guo, G.-X.; Stedtfeld, R. D.; Hashsham, S. A.; Tiedje, J. M. Diverse and Abundant Antibiotic Resistance Genes in Chinese Swine Farms. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (9), 3435–3440. https://doi.org/10.1073/pnas.1222743110.

(42) Gibson, M. K.; Forsberg, K. J.; Dantas, G. Improved Annotation of Antibiotic Resistance Determinants Reveals Microbial Resistomes Cluster by Ecology. *ISME J.* **2014**, *9* (1), 1–10. https://doi.org/10.1038/ismej.2014.106.

(43) Orellana, L. H.; Rodriguez-R, L. M.; Konstantinidis, K. T. ROCker: Accurate Detection and Quantification of Target Genes in Short-Read Metagenomic Data Sets by Modeling Sliding-Window Bitscores. *Nucleic Acids Res.* **2017**, *45* (3), e14--e14.

(44) Bushnell, B. BBTools Software Package. **2017**.

(45) 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.* **2019**, *48* (October 2019), 517–525. https://doi.org/10.1093/nar/gkz935.

(46) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2013**, *41* (D1), 590–596. https://doi.org/10.1093/nar/gks1219.

(47) Elsik, C. G.; Unni, D. R.; Diesh, C. M.; Tayal, A.; Emery, M. L.; Nguyen, H. N.; Hagen, D. E. Bovine Genome Database: New Tools for Gleaning Function from the Bos Taurus Genome. *Nucleic Acids Res.* **2016**, *44* (D1), D834--D839.

(48) Luo, Y.; Mao, D.; Rysz, M.; Zhou, Q.; Zhang, H.; Xu, L.; Alvarez, P. J. J. Trends in Antibiotic Resistance Genes Occurrence in the Haihe River, China. *Environ. Sci. Technol.* **2010**, *44* (19), 7220–7225. https://doi.org/10.1021/es100233w.

(49) Fahrenfeld, N.; Knowlton, K.; Krometis, L. A.; Hession, W. C.; Xia, K.; Lipscomb, E.;

Libuit, K.; Green, B. L.; Pruden, A. Effect of Manure Application on Abundance of Antibiotic Resistance Genes and Their Attenuation Rates in Soil: Field-Scale Mass Balance Approach. *Environ. Sci. Technol.* **2014**, *48* (5), 2643–2650. https://doi.org/10.1021/es404988k.

(50) Ji, X.; Shen, Q.; Liu, F.; Ma, J.; Xu, G.; Wang, Y.; Wu, M. Antibiotic Resistance Gene Abundances Associated with Antibiotics and Heavy Metals in Animal Manures and Agricultural Soils Adjacent to Feedlots in Shanghai; China. *J. Hazard. Mater.* **2012**, *235–236*, 178–185. https://doi.org/10.1016/j.jhazmat.2012.07.040.

(51) Hougs, L.; Gatto, F.; Goerlich, O.; Grohmann, L.; Lieske, K.; Mazzara, M.; Narendja, F.; Ovesna, J.; Papazova, N.; Scholtens, I. M. J.; et al. *Verification of Analytical Methods for GMO Testing When Implementing Interlaboratory Validated Methods*; 2019. https://doi.org/10.1201/9781315178592-8.

Chapter 4

Anaerobic Digestion Shapes Land-Applied Dairy Manure Microbial Community Structure and Antimicrobial Resistance Gene Profiles

ABSTRACT

The relative roles of dairy manure storage and anaerobic digestion (AD) in shaping community structure and antimicrobial resistance gene (ARG) concentrations in land-applied manure is not well understood. We perform a cross-sectional study of land-applied manures from 12 dairy farms to evaluate drivers of microbial community structure and ARG abundances. Land-applied manure samples were collected during spring and fall storage draw-down events. Half of the samples were collected from farms employing AD and the remainder from farms without AD. We found that stored, anaerobically digested manures were less diverse and less variable in community and ARG composition than stored manure without prior anaerobic digestion. Fall samples were also more diverse than samples collected in the spring and diversity increased with decreasing percent total solids. We show that AD is a major driver of microbial community structure in stored dairy manure, which differed from storage-only land-applied manures by higher concentrations of *Synergistota* and lower concentrations of *Spirochaetota* and *Proteobacteria*. ARG profiles across the farms were predominantly comprised of genes in the beta-lactams, aminoglycosides, macrolides, lincosamides, and streptrogramin A and B (MLS) drug classes.

4.1 Introduction

Land application of manure restores organic carbon and nutrients into agricultural soils, but also introduces antimicrobial resistant bacteria (ARB) and antimicrobial resistance genes (ARGs) into the environment where they present human and veterinary health risks (Heuer, Schmitt, and Smalla 2011; Xie, Shen, and Zhao 2018). Understanding the major factors that shape the microbial community structure and ARG concentrations in land-applied manure is critical for designing waste management best practices that mitigate risks associated with antimicrobial resistance. Anaerobic digestion (AD) is a biological process employed by some farms to reduce pathogen prevalence, stabilize waste, and generate biogas. Compared to lower-temperature digestion systems, thermophilic digestion has been shown to have a greater impact on community structure (W Sun et al. 2016), levels of viable antimicrobial resistant pathogens (Beneragama et al. 2013; Youngquist, Mitchell, and Cogger 2016), ARG concentrations (copies/mass dry solids) (Ma et al. 2011; W Sun et al. 2016), and ARG relative abundances (reads/total metagenome reads) (Zhang, Yang, and Pruden 2015). However, these studies have been largely limited to laboratory scales.

Additionally, AD is typically not the final stage of manure management on farms before land application. After anaerobic digestion, treated manure is often stored for several months. One of the few studies on dairy manure storage showed that microbial communities became less diverse through storage following anaerobic treatment, aerobic treatment, and no treatment (McGarvey et al. 2007). Tetracycline resistance genes have been found at high abundances in stored manure. Abundances of six different tetracycline ARGs ranged from 10^1-10^6 copies per mg and 10^1-10^6 copies/16S rRNA gene copies across lagoons from 4 different cattle feedlots (Peak et al. 2007) and measured *tetO* abundances ranged from 10^{-3}-10^{-2} copies/16S rRNA gene copies in stored manures from 11 different dairies (Hurst et al. 2019). One study found higher

16S rRNA gene-standardized abundances of three ARGs (*sul1, tetM*, and *bla_{OXA-58}*) in stored dairy manure compared to fresh manure (Ruuskanen et al. 2016). A longitudinal cross section of livestock operations, including a poultry, beef, swine and dairy farms, found 16S rRNA gene-standardized abundances of sulfonamide resistance genes were stable through storage while tetracycline resistance gene abundances declined over time (McKinney et al. 2010). With limited sample sizes from the same farms (i.e. dairy vs. swine), unrepresentative grab sampling, and limited number of genes targeted per study, it is difficult to draw conclusions on the impact of storage on ARGs and community composition.

In this study, we compare a cross-section of 15 dairy storages from 12 farms across different regions in the United States using high throughput DNA sequencing. Stored manure samples were collected during land application events to obtain representative samples of stored manure. The dairy cohort is intended to serve as a cross-section of dairy farms from different regions, different manure management systems, and herd sizes. The aim of this study was to (1) determine the variability in farm characteristics, microbial community structure, and ARG composition of stored dairy manure as they are applied to land, (2) evaluate the factors that shape the microbial community structure, and (3) evaluate co-occurrence patterns in concentrations of ARG and microbial populations across farms.

4.2 Results

4.2.1 Dairy cohort

Twelve dairy farms participated in the study representing the Upper Midwest (MW), Northeast (NE), and Mid-Atlantic (MA) regions of the Unites States and ranged in herd size from 80-4,200 cows. Samples were collected while stored manure was agitated, pumped and hauled for land application. Fall samples were collected in September and October, and spring samples were collected in April and May. Three farms were sampled twice, either in different

seasons or from different storages, resulting in a total of 15 samples collected from manure

storages during land application events. All five farms using AD prior to storage ("AD+storage")

were located in the NE and MA regions and tended to have larger herd sizes (Table 1).

Concentrations of nitrogen, ammonia, and phosphorus are reported in Supplemental Table 1.

Total solids varied over three orders of magnitude in the storage-only manures, ranging from 0.2-

15.4% but only ranged from 1.3-6.7% in AD+storage manures (Table 4-1). DNA yield (ng

DNA/mg sample) was also more variable in the storage only samples (Table 4-1). Every manure

storage in the cohort received manure from both healthy cows and sick cows that were treated

with antibiotics except manure storage from Farm MW2. At this farm, there was separate storage

for manure from sick cows.

Table 4-1: *Dairy storage cohort farm and manure characteristics*

Sample	% Total solids	ng DNA/mg Sample	Herd size	Season
Storage only Samples				
MW1	2.3	13.6	150	Fall
MW2	15.4	22.4	600	Spring
MW3	0.7	2.6	400	Spring
MW4 - Storage 1	0.8	5.1	330	Spring
MW4 - Storage 2	1.0	14.9	330	Spring
MW5	14.1	15.7	680	Spring
MA1	0.2	1.7	80	Fall
NE1	4.2	15.1	1350	Fall
Anaerobic Digestion + Storage Samples				
NE2 - Fall Storage	6.8	23.3	4300	Fall

NE2 - Spring Storage	4.2	18.3	4300	Spring
NE3 - Fall Storage	6.7	19.1	2050	Fall
NE3 - Spring storage	3.1	18.2	2050	Spring
NE4	1.3	12.5	1820	Spring
MA2	2.8	11.9	2700	Fall
MA3	4.6	14.8	690	Spring

MW = Midwest, NE = Northeast, MA = Mid-Atlantic

4.2.2 Diversity in community compositions differed by AD status, total solids concentration, and season.

phyloFlash extracts and assigns metagenomic reads to 16S rRNA gene sequences in the SILVA database using BBMap (Gruber-Vodicka, Seah, and Pruesse 2019). Between 1.6×10^5 and 5.5×10^5 reads were identified as 16S rRNA genes from each of the metagenomes in this study, comprising between 0.1 and 0.2% of total metagenomic reads (Table C-2). Across all samples, 14 phyla accounted for 83-99% total read assignments (Figure 4-1). All samples but the Mid-Atlantic 1 (MA1) sample were dominated by organisms from the *Firmicutes* and *Bacteroidetes* phyla; which comprised 15-59% and 10-37% of 16S rRNA reads, respectively. Across all samples *Clostridia* and *Bacilli* were the dominant classes in *Firmicutes* and *Bacteroidia* was the dominant class in *Bacteroides* (Fig. C-2). *Proteobacteria* was the dominant phylum in the Mid-Atlantic 1 (MA1) sample. These patterns of dominant phyla were visible when comparing organism concentrations (16S rRNA gene copies/mg wet weight, Fig. C-1). *Halobacterota*, a phylum of archaea including several genera of methanogens, and *Verrucomicrobiota,* a phylum of bacteria prevalent in the soil (Bergmann et al. 2011) were also abundant across most samples (Figure 4-1).
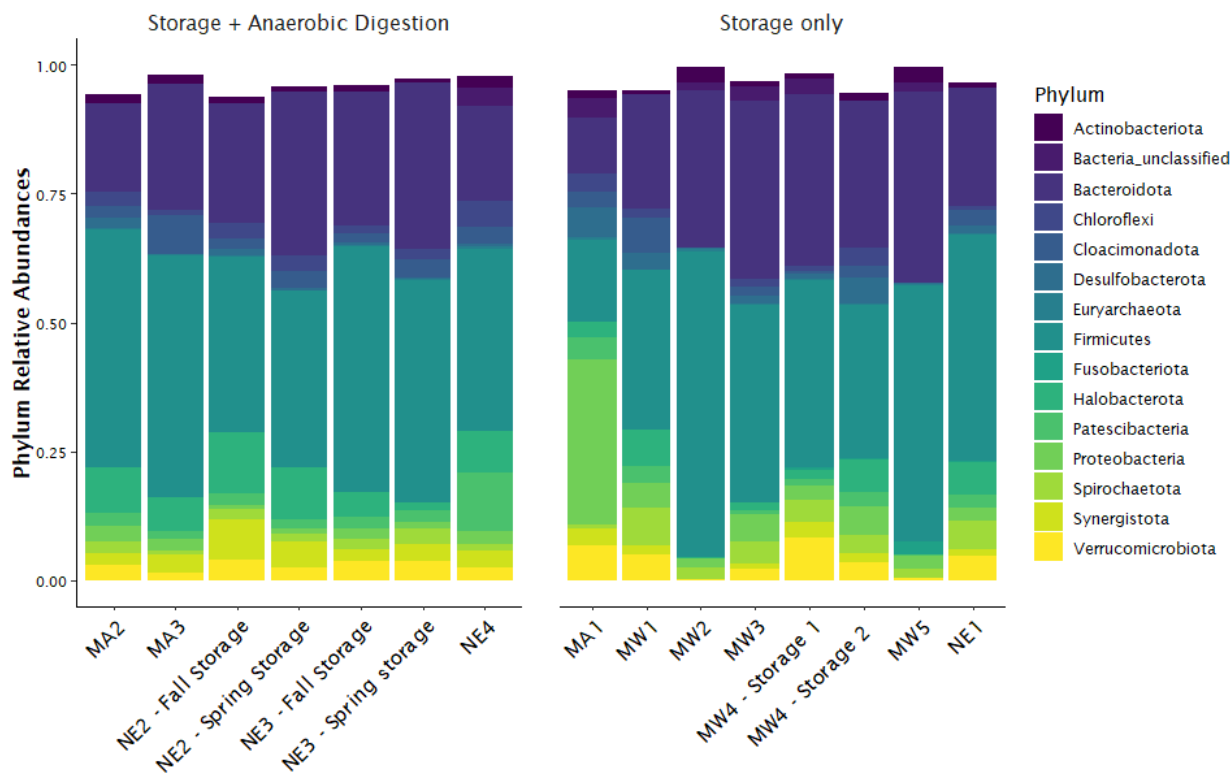
Figure 4-1: *Relative abundances of phyla in manure samples.*

Reads were assigned to ARG sequences using AMRPlusPlus (AMR++) using the MEGARes database (Doster et al. 2020) and relative abundances (reads/total ARG-assigned reads, Figure 4-2) and gene concentrations (copies/mg wet weight, Fig C-3) were computed. AMR++ assigned between 5.1×10^5 and 2.1×10^6 reads to ARGs, corresponding to 0.4-1.1% of total metagenomic reads (Table C-2). The most prevalent drug classes represented by the ARGs across all samples were beta-lactams, aminoglycosides, macrolides, lincosamides, and streptrogramin A and B (MLS) drugs, collectively comprising between 60% and 80% of the drug classes detected. ARGs in tetracyclines, elfamycins, and sulfonamides drug classes were prevalent but present at lower relative abundances (Figure 4-2). Relative abundance profiles were similar across all samples but NE4 and NE3-Fall samples exhibited higher prevalence of beta-lactams than other samples in the cohort (Figure 4-2).
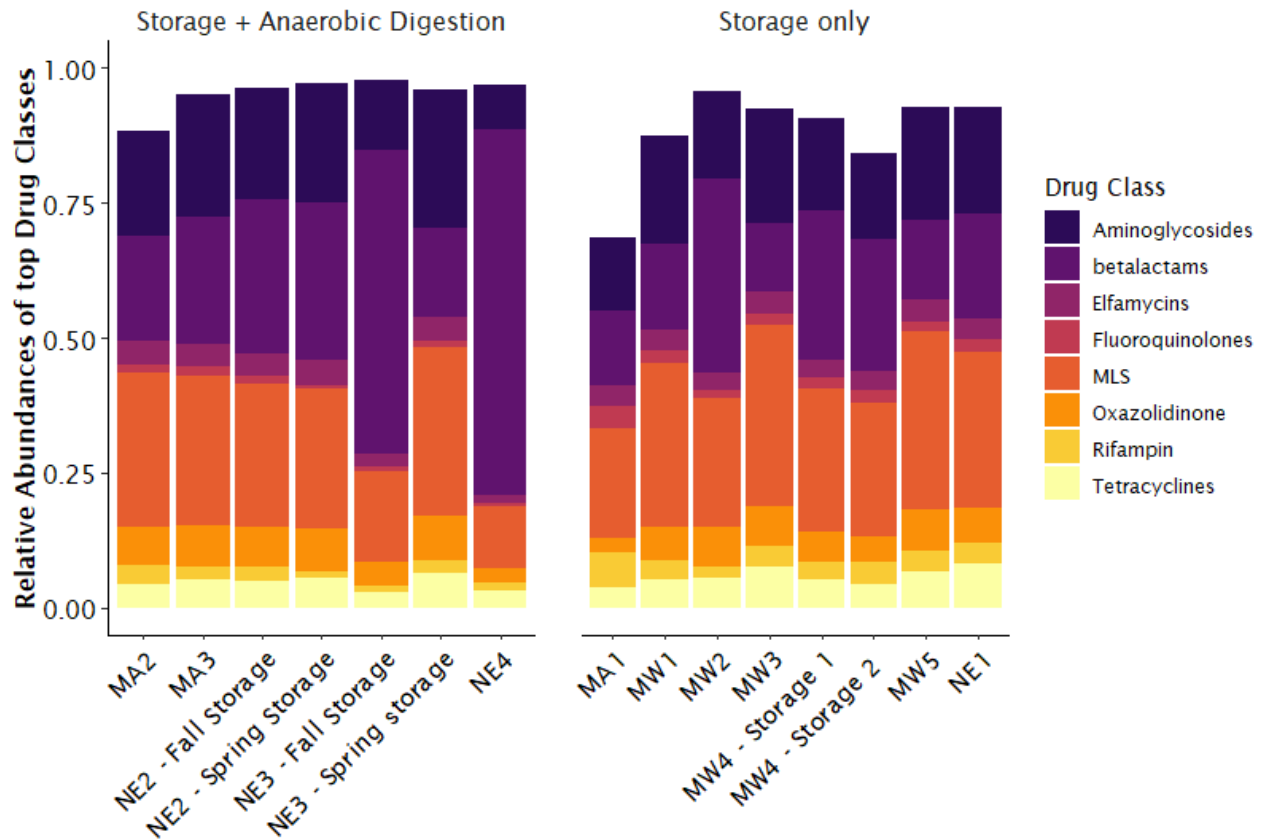
Figure 4-2: *AMR++ Relative ARG abundances grouped by drug classes.*

To assess if observed ARG read assignments were consistent across different assignment approaches and databases, Graphing Resistance Out Of meTagenomes (GROOT) was used to obtain resistance gene profiles using the Comprehensive Antimicrobial Resistance Database as a reference (Alcock et al. 2019). However, the relative abundance profiles were dominated by TEM genes which are a family of beta-lactamases. Specifically, 99% of assigned reads in each sample were assigned to TEM genes. TEM were also dominant in the resistance profiles using AMR++ with the MEGARes database but only 12-67% of ARG reads mapped to TEM using AMR++.

Compared to storage-only farm samples, AD+storage samples were less diverse in ARGs (Figure C-4) as well as in family and genera-level community composition, although these differences were not statistically different (Figure 4-3A). This pattern was observed with alpha diversity metrics measuring species richness (Chao1) and metrics quantified from species richness and relative abundances (Shannon, Simpson's Index). AD+storage samples also had lower inter-quartile ranges indicating that the variability in diversity was lower than that of storage-only farms. At higher taxonomic levels (order and class), however, median alpha diversity from AD+storage samples surpassed those of storage-only samples.

Alpha diversity differed with other manure characteristics in addition to AD treatment. Samples collected in the fall were more diverse than samples collected in the spring (Figure 4-3B). Low total solids (TS) samples (0.23-2.82%) were more diverse than high TS samples (3.05-15.4%, Figure 4-3C). The observed differences in alpha diversity with solids content and season were significantly different and consistent at higher taxonomic levels—in contrast to what was observed between AD+storage and storage-only groups (Figure 4-3).

Figure 4-3: *Alpha diversity summarized at different taxonomic levels between (A) storage-only and storage+AD samples, (B) samples collected in the fall and spring, and (B) high and low total solids. (t-test, ns: p > 0.05, *: p ≤ 0.05, **: p ≤ 0.01, ***: p ≤ 0.001)*

4.2.3 AD status correlates with microbial community structure and ARG profiles across farms

We performed canonical correlation analysis (CONCOR) to identify relationships between the organism concentrations (16S rRNA gene copies/mg, wet weight) and ARG concentrations (gene copies/mg, wet weight) with the farm management and manure metadata. CONCOR identifies independent factors within dominant principal components of the microbiome and metadata that are maximally correlated with each.

We found that AD was the most dominant factor explaining the differences in community structure (Figure 4-4B). Herd size and longitude were also important factors, but they are highly correlated with digestion status, making it difficult to determine which of these three parameters drives microbial community composition (Figure 4-4B). Since CONCOR first involves linear dimensional reduction, co-correlated factors, such as herd size and longitude, are linearly combined and therefore the effects are not double counted. Storage-only manure microbial communities were differentiated from AD+storage manure microbial communities by higher concentrations of *Spirochaetota* and *Proteobacteria* phyla and lower concentrations of *Synergistota* (Figure 4-4A). Class, order, family, and genus-level data were also correlated with farm metadata (Fig. C-5). *Methanosarcinaceae* (*Halobacterota* phylum) and *Caldicoprobacteraceae* (*Firmicutes* phylum) families were more prevalent in AD+storage manures prior to storage but not in storage-only manures. Storage-only community profiles were differentiated AD+storage by the higher concentrations and prevalence of *Bacteroidaceae* and *Lachnospiraceae* families (Fig. C-5).

AD was also a driver of ARG structure as demonstrated by the separation between AD+storage and storage-only samples along the AD axis (Figure 4-4D). However, MA2 sample clustered closely with storage-only samples meaning ARG concentration profiles were more

72

similar to those of the MW farms (Figure 4-4C). TS percent and latitude also correlated with

ARG concentrations. Northeast farms were higher in concentrations of ERMF genes while

Midwest farms were higher in Macrolide-resistant 23S rRNA mutations (MLS23S, Figure 4-4C).
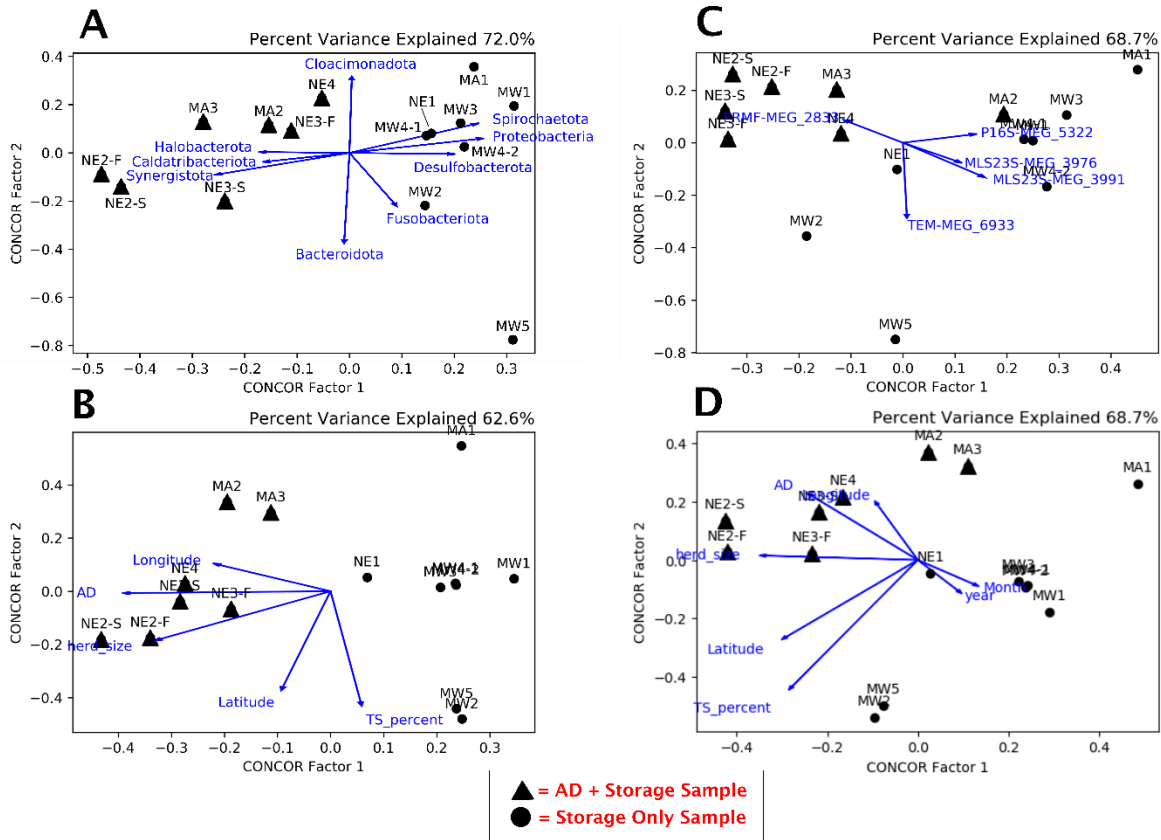
Figure 4-4: *Canonical Correlation (CONCOR) analysis.*
Co-correlated sample Phylum-level organism concentrations (A) and meta-data (B). Co-

correlated sample (C) ARG concentrations and (D) meta-data.

4.2.4 Anaerobically digested manures were more similar to each other in community structure
and ARG composition

The ARG and family-level taxonomic compositions (read counts) were compared

between farms using the Procrustes algorithm with multi-dimensional scaling (MDS) of

Pairwise-Hellinger distances (Figure 4-5). We found that AD+storage samples were more similar

to each other than they were to samples from storage-only farms (Figure 4-5). Samples from

storage-only farms were less similar to each other as they did not cluster as tightly as samples

from the AD-farms. Samples collected from the same farm were not always more similar to each

other than to other farms, suggesting that even on-farm variability in microbial community and

ARG composition is considerable (Figure. 4-5). The observation of tighter clustering of

AD+storage compared to storage-only was consistent using Bray-Curtis distances instead ($M^2 =$

0.37, Fig. C-6).
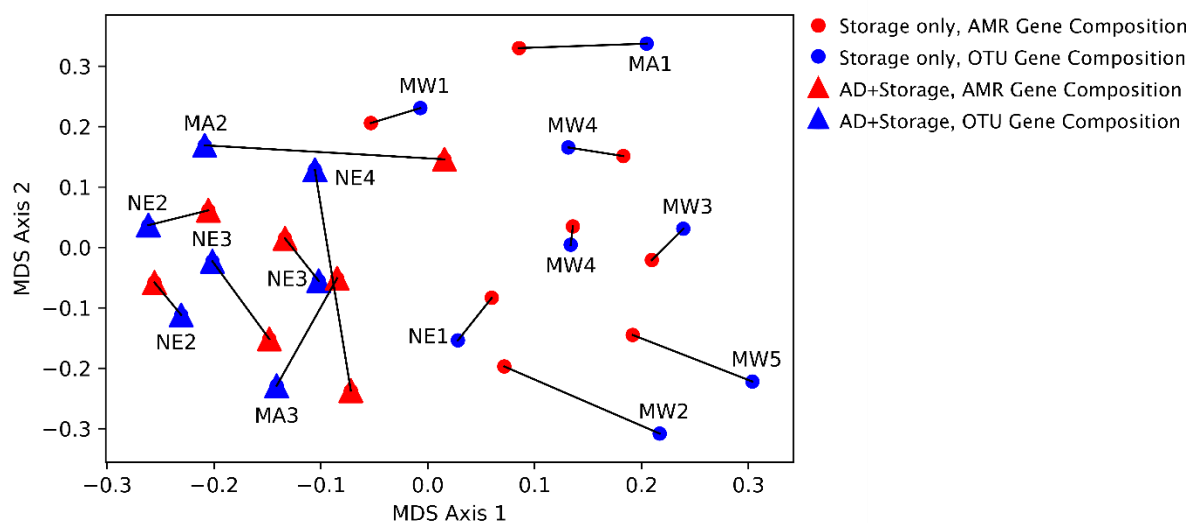
Figure 4-5: *Procrustes Analysis demonstrating relationship between sample ARG and OTU composition across farm cohort*
M^2 = 0.34. Pairwise-Hellinger distance computed between read count profiles.

4.3 Discussion

We computed the concentrations and relative abundances of microbial taxa (Fig. C-1,

Figure 4-1) and ARGs (Fig. C-3, Figure 4-2) in dairy manure storage systems as the stored

manure was agitated, pumped, and hauled for land application. Microbial communities in stored

dairy manure were distinct between AD+storage and storage-only systems (Figure 4-4). The

manures that were digested prior to storage (AD+storage) were less diverse at finer taxonomic

74

scales and less variable in diversity and community composition storage-only manures (Figure 4-3). Across storage-only and AD+storage samples, *Firmicutes* and *Bacteroidetes* were the dominant phyla. Concentrations of *Synergistota, Spirochaetota* and *Proteobacteria* phyla differentiated the communities, with *Spirochaetota* and *Proteobacteria* more prevalent in storage-only manure and *Synergistota* more prevalent in the AD-treated manure (Figure 4-4). *Firmicutes* and *Bacteroidetes* were also prevalent phyla in a cross-section of feces from lactating dairy cows and pre-weaned calves in Pennsylvania dairies (Haley et al. 2020) and lactating cattle in California dairies (Hagey et al. 2019). *Firmicutes* was dominant in a full-scale anaerobic digester treating manure (De Vrieze et al. 2016).

Anaerobic digestion is uncommon on dairy farms, less than 200 of the dairies in the United States, or less that 1% of US dairies use AD for manure management (Oliver et al. 2020). Consequently, there are limited full-scale studies on how AD shapes manure microbiomes. One bench-top study found temperature-dependent effects on microbial community succession through digestion with the most extreme changes in community structure observed during thermophilic digestion. Specifically, digestion shifted the dominant community members from *Actinobacteria* and *Proteobacteria* to *Firmicutes* and *Chloroflexi* (Wei Sun et al. 2016). In this study, AD use correlated with farm size and region (Table 4-1); consequently, it was not possible to isolate the effect AD specifically had on shaping microbial communities (Figure 4-4B). A larger samples size with more AD farms across different regions and herd sizes would be needed to address this limitation in future research.

Minimizing the release of pathogenic organisms via land-application is critical to reducing risks in animal and human populations. Determining clinical relevance of observed differences in composition between AD+storage and storage-only manures is not possible due to

the lack of species and sub-species level resolution of metagenomic approaches. *Salmonella spp.,*

Campylobacter spp., Listeria monocytogenes, Yersinia enterocolitica, and *Escherichia coli* are

clinically relevant pathogens that are prevalent in cattle manure (Pachepsky et al. 2006). Each of

these pathogens, except *Listeria monocytogenes,* are in the *Proteobacteria* phylum, which was

more dominant in storage only manures. However, without isolation of pathogenic species, it is

not possible to infer if pathogen prevalence was higher in the storage-only land-applied manures

in this study.

To combat the complexity and growing diversity of ARGs identified and annotated, there

has been an explosion of tools and databases for identifying and quantifying antimicrobial

resistance from microbial communities (Boolchandani, DeSouza, and Dantas 2019). Databases

use different ontological structures for organizing and classifying resistance genes by function,

sequence variation, and target antibiotic. This makes comparisons between studies more

challenging as they often employ different databases and read-assignment algorithms. Further,

functional metagenomics, an approach where DNA fragments found to express a resistance

phenotype are sequenced allowing discovery of novel ARG sequences, has revealed limitations

of databases. For instance, a functional metagenomic study on manure found the majority of

resistance genes identified shared between 30-70% identity to closest matches in GenBank

(Wichmann et al. 2014). A study on agricultural soils found that only 0.5% of genes conferring

resistance had >99% nucleotide identity to NCBI nucleotide sequences (Forsberg et al. 2014).

Here we used two different open-source tools and databases to assign metagenomic reads

to ARG sequences. We selected these two tools because they have been programmed to

differentiate reads between homologous ARGs. Additionally, AMR++ was previously used to

characterize resistance genes in bovine fecal systems (Zaheer et al. 2019) and calves (Liu et al.

2019). In a previous study, we demonstrated that both AMR++ GROOT performed comparably to qPCR for quantifying gene concentrations from dairy manure (Crossette *et al*. 2020, *in review*). In this study, however, GROOT with CARD as a reference detected only TEM beta lactamase gene, suggesting a potential bias in CARD and/or the GROOT read assignment algorithm. The total number of reads mapping to ARGs using both approaches was the same or higher than the total number of reads mapping to 16S rRNA gene sequences (Table C-2), suggesting that the cut-off thresholds for ARG read assignment might need to be set higher to reduce possible ambiguous mapping.

In this cross-section of land-applied dairy manures, beta-lactams, specifically *bla-TEM* dominated the resistance gene profiles (Figure 4-2). Previous studies of fecal microbial communities using qPCR have found high prevalence and abundances of tetracycline resistance genes in fresh (Kyselková et al. 2015) and stored manure (Peak et al. 2007). Further, metagenomic studies of cattle manure have confirmed that tetracycline resistance genes are dominant resistance genes in ARG profiles in manures from beef cattle herds (Noyes et al. 2016; Zaheer et al. 2019) and lactating dairy cows (Haley et al. 2020; Noyes et al. 2016). One study reported tetracycline genes were more abundant in beef manures compared to dairy resistomes (Rovira et al. 2019). This suggests that over time in storage, the core gut microbial community members harboring tetracycline resistance gene decline relative to other populations in the storages.

We found that both community profiles and ARG profiles were shaped by similar factors (Figure 4-4, Fig. C-5) and indeed were correlated using the Procrustes method (Figure 4-5, Fig. C-6). This observation of correlated abundances profiles has been observed in soil (Forsberg et al. 2014; Han et al. 2018), wastewater treatment plants (Ju et al. 2019), and full-scale biogas

reactors (Luo et al. 2017). In other words, the differences in community structure correspond to differences in resistance gene profiles in the storage systems. In these previous studies, the coupling between ARG abundances and community structure was interpreted as evidence that horizontal gene transfer is not a dominant driver of gene abundances (Forsberg et al. 2014; Han et al. 2018; Ju et al. 2019; Luo et al. 2017).

This pilot-scale cross-sectional study demonstrated that stored manures vary in microbial community compositions and total solids. This variability in manures between farms highlights how studies observing community structure and ARG profiles in samples collected from on one or two farms are likely not generalizable, especially if the farms do not use AD. Despite the observed variability, we found that land-applied manure microbial communities were distinct between stored manure with prior AD and storage-only manure. Therefore, AD is an important driver of microbial community structure and ARG abundances in stored dairy manure.

4.4 Materials and Methods

4.4.1 Manure storage Sampling and Nutrient Analysis.

Manure storage samples were collected during drawdown events, when storages were mixed and hauled for land application. Samples were collected as well distributed during the hauling as possible to best capture a representative sample of the stored manure that was land-applied. Samples for molecular analysis were collected in sterile 15 or 50mL Falcon Tubes and stored on ice until returning to the lab where it was frozen at -80C until DNA extraction. Midwest stored manure samples were sent to the University of Wisconsin Soil and Forage Analysis Lab for nutrient analyses. Northeast and Mid-Atlantic stored manure sample nutrient analysis were performed in house as described previously (Lansing et al. 2019). Metadata about the farm was collected during the hauling events including herd size, bedding type, and whether or not sick cow bedding and manure were fed to the sampled storage.

4.4.2 DNA Extraction and Extraction Recovery Calculation.

Manure samples were thawed on ice and samples from the same drawdown event were pooled in equal proportions and well-mixed. Between 150-200 mg of each drawdown composite sample was allocated into QIAAMP PowerFecal DNA extraction kit bead tubes (QIAGEN, Germantown, MD). Triplicate extractions were performed, one triplicate was spiked with 10 µL of ZymoBIOMICS Spike-in Control I (ZYMO Research, Irving CA), which contains a gram-positive and gram-negative organism. DNA was extracted following the manufacture's guidelines (Supplemental Text C-1).

4.4.3 DNA Sequencing

Before DNA sequencing, unspiked DNA extracts from each drawdown were pooled then spiked with Sequins (Hardwick et al. 2018) at 1.7% DNA by mass then sequenced on an Illumina NovSeq (300 cycle) instrument at the University of Michigan Advanced Genomics Core Facilities. PCR-Free library preparation was used with an insert size of 175 basepairs (bp) Approximately 200 million reads were generated per sample.

4.4.4 Bioinformatics

Quality control was performed using Trimmomatic (version 0.39). Performance of quality control steps were assessed with FastQC (version 0.11.8) and visualized with MultiQC (version 1.8). Taxonomic counts for bacteria and archaea were generated using phyloFlash (version 3.3b3) with default settings but specifying the read length (150 bp) and insert size (175 bp). Reads were assigned to antimicrobial resistance genes using GROOT and AMR++ tools. GROOT was run with default parameters, except a 50% reporting threshold was used and the read length was set to 150 bp. CARD2020 was used as the reference database. AMR++ (version 2.0.0) was performed using the singularity container with the MEGARes database as a reference and default parameters (--threshold 80 --min 1 --max 100 --samples 5 --skip 5). Read

abundances assigned to target antimicrobial resistance genes using GROOT and AMR++, and read abundances assigned to NTUs using phyloflash, were converted to absolute abundances, (copies per mass of sample) following *Crossette et al. 2020 in prep.* In short, the *Sequins* spiked after DNA extraction and before sequencing establish a ratio of read abundances to copy concentration. This ratio was then used to convert read counts to gene copies in each sample. Abundance profiles were plotted using GGplot in R.

4.4.5 Statistical Methods

Alpha diversity metrics were computed using the alpha diversity package from skbio and plotted with GGplot in R. Canonical correlation analysis (CONCOR) was used to correlate the abundances of NTUs with metadata and again to correlate ARG abundances with sample metadata. However, since the number of observed NTUs and ARGs exceed the number of samples, principal component analysis was first performed to reduce the NTU and ARG abundances into linear combinations. CONCOR is an unsupervised approach that finds linear combinations of two data matrices that correlate highly with each other. Percent variance explained by each 2-D visualization was determined as the product of the explained variation in the dominant principal components used in the correlation and the square of the CONCOR correlation coefficient. CONCOR was implemented and plotted using custom python scripts. Dimensional reduction and Procrustes analysis were also performed in python using custom scripts and plotted with matplotlib.

4.5 References

Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, Anna-Lisa V Nguyen, Annie A. Cheng, Sihan Liu, Sally Y. Min, Anatoly Miroshnichenko, Hiu-Ki Tran, Rafik E. Werfalli, Jalees A. Nasir, Martins Oloni, David J. Speicher, Alexandra Florescu, Bhavya Singh, Mateusz Faltyn, Anastasia Hernandez-Koutoucheva, Arjun N. Sharma, Emily Bordeleau, Andrew C. Pawlowski, Haley L. Zubyk, Damion Dooley, Emma Griffiths, Finlay Maguire, Geoff L. Winsor, Robert G. Beiko, Fiona S. L. Brinkman, William W. L. Hsiao, Gary V Domselaar, and Andrew G. McArthur. 2019. "CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 48(October 2019):517–25.

Beneragama, Nilmini, Masahiro Iwasaki, Suraju A. Lateef, Takaki Yamashiro, Ikko Ihara, and Kazutaka Umetsu. 2013. "The Survival of Multidrug-Resistant Bacteria in Thermophilic and Mesophilic Anaerobic Co-Digestion of Dairy Manure and Waste Milk." *Animal Science Journal* 84(5):426–33.

Bergmann, Gaddy T., Scott T. Bates, Kathryn G. Eilers, Christian L. Lauber, J. Gregory Caporaso, William A. Walters, Rob Knight, and Noah Fierer. 2011. "The Under-Recognized Dominance of Verrucomicrobia in Soil Bacterial Communities." *Soil Biology and Biochemistry* 43(7):1450–55.

Boolchandani, Manish, Alaric W. DeSouza, and Gautam Dantas. 2019. "Sequencing-Based Methods and Resources to Study Antimicrobial Resistance." *Nature Reviews Genetics* 1.

Doster, Enrique, Steven M. Lakin, Christopher J. Dean, Cory Wolfe, Jared G. Young, Christina Boucher, Keith E. Belk, Noelle R. Noyes, and Paul S. Morley. 2020. "MEGARes 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data." *Nucleic Acids Research* 48(D1):D561--D569.

Forsberg, Kevin J., Sanket Patel, Molly K. Gibson, Christian L. Lauber, Rob Knight, Noah Fierer, and Gautam Dantas. 2014. "Bacterial Phylogeny Structures Soil Resistomes across Habitats." *Nature* 509(7502):612–16.

Gruber-Vodicka, Harald R., Brandon K. B. Seah, and Elmar Pruesse. 2019. "PhyloFlash--Rapid SSU RRNA Profiling and Targeted Assembly from Metagenomes." *BioRxiv* 521922.

Hagey, Jill V, Srijak Bhatnagar, Jennifer M. Heguy, Betsy M. Karle, Patricia L. Price, Deanne Meyer, and Elizabeth A. Maga. 2019. "Fecal Microbial Communities in a Large Representative Cohort of California Dairy Cows." *Frontiers in Microbiology* 10:1093.

Haley, Bradd J., Seon-Woo Kim, Serajus Salaheen, Ernest Hovingh, and Jo Ann S. Van Kessel. 2020. "Differences in the Microbial Community and Resistome Structures of Feces from Preweaned Calves and Lactating Dairy Cows in Commercial Dairy Herds." *Foodborne*

Pathogens & Disease (2400447067; IND606862932).

Han, Xue-Mei, Hu Hang-Wei, Li He-Lian, Yang Li-Yuan, Chen Qing-Lin, Li Xiang-Zhen, Ma Yi-Bing, and Zhu Yong-Guan. 2018. "Antibiotic Resistance Genes and Associated Bacterial Communities in Agricultural Soils Amended with Different Sources of Animal Manures." *Soil Biology & Biochemistry* 126(2153611980; IND606130768):91–102.

Hardwick, Simon A., Wendy Y. Chen, Ted Wong, Bindu S. Kanakamedala, Ira W. Deveson, Sarah E. Ongley, Nadia S. Santini, Esteban Marcellin, Martin A. Smith, Lars K. Nielsen, Catherine E. Lovelock, Brett A. Neilan, and Tim R. Mercer. 2018. "Synthetic Microbe Communities Provide Internal Reference Standards for Metagenome Sequencing and Analysis." *Nature Communications* 9(1):1–10.

Heuer, Holger, Heike Schmitt, and Kornelia Smalla. 2011. "Antibiotic Resistance Gene Spread Due to Manure Application on Agricultural Fields." *Current Opinion in Microbiology* 14(3):236–43.

Hurst, Jerod J., Jason P. Oliver, Jenna Schueler, Curt Gooch, Stephanie Lansing, Emily Crossette, Krista Wigginton, Lutgarde Raskin, Diana S. Aga, and Lauren M. Sassoubre. 2019. "Trends in Antimicrobial Resistance Genes in Manure Blend Pits and Long-Term Storage Across Dairy Farms with Comparisons to Antimicrobial Usage and Residual Concentrations." *Environmental Science & Technology* 53(5):2405–15.

Ju, Feng, Karin Beck, Xiaole Yin, Andreas Maccagnan, Christa S. McArdell, Heinz P. Singer, David R. Johnson, Tong Zhang, and Helmut Bürgmann. 2019. "Wastewater Treatment Plant Resistomes Are Shaped by Bacterial Composition, Genetic Exchange, and Upregulated Expression in the Effluent Microbiomes." *The ISME Journal* 13(2):346–60.

Kyselková, Martina, Jirí Jirout, Nadežda Vrchotová, Heike Schmitt, and Dana Elhottová. 2015. "Spread of Tetracycline Resistance Genes at a Conventional Dairy Farm." *Frontiers in Microbiology* 6(MAY):1–14.

Lansing, Stephanie, Benedikt Hülsemann, Abhinav Choudhury, Jenna Schueler, Maria Sol Lisboa, and Hans Oechsner. 2019. "Food Waste Co-Digestion in Germany and the United States: From Lab to Full-Scale Systems." *Resources, Conservation and Recycling* 148:104–13.

Liu, J., D. H. Taft, M. X. Maldonado-Gomez, D. Johnson, M. L. Treiber, D. G. Lemay, E. J. DePeters, and D. A. Mills. 2019. "The Fecal Resistome of Dairy Cattle Is Associated with Diet during Nursing." *Nature Communications* 10(1).

Luo, Gang, Bing Li, Li Li-Guan, Tong Zhang, and Angelidaki Irini. 2017. "Antibiotic Resistance Genes and Correlations with Microbial Community and Metal Resistance Genes in Full-Scale Biogas Reactors As Revealed by Metagenomic Analysis." *Environmental Science & Technology* 51(7):4069–80.

Ma, Yanjun, Christopher A. Wilson, John T. Novak, Rumana Riffat, Sebnem Aynur, Sudhir Murthy, and Amy Pruden. 2011. "Effect of Various Sludge Digestion Conditions on

Sulfonamide, Macrolide, and Tetracycline Resistance Genes and Class I Integrons." *Environmental Science & Technology* 45(18):7855–61.

McGarvey, Jeffery A., William G. Miller, Ruihong Zhang, Yanguo Ma, and Frank Mitloehner. 2007. "Bacterial Population Dynamics in Dairy Waste during Aerobic and Anaerobic Treatment and Subsequent Storage." *Applied and Environmental Microbiology* 73(1):193–202.

McKinney, Chad W., Keith A. Loftin, Michael T. Meyer, Jessica G. Davis, and Amy Pruden. 2010. "Tet and Sul Antibiotic Resistance Genes in Livestock Lagoons of Various Operation Type, Configuration, and Antibiotic Occurrence." *Environmental Science & Technology* 44(16):6102–9.

Noyes, Noelle R., Xiang Yang, Lyndsey M. Linke, Roberta J. Magnuson, Shaun R. Cook, Rahat Zaheer, Hua Yang, Dale R. Woerner, Ifigenia Geornaras, Jessica A. McArt, Sheryl P. Gow, Jaime Ruiz, Kenneth L. Jones, Christina A. Boucher, Tim A. McAllister, Keith E. Belk, and Paul S. Morley. 2016. "Characterization of the Resistome in Manure, Soil and Wastewater from Dairy and Beef Production Systems." *Scientific Reports* 6(April):1–12.

Oliver, J. P., C. A. Gooch, S. Lansing, J. Schueler, J. J. Hurst, L. Sassoubre, E. M. Crossette, and D. S. Aga. 2020. "Invited Review: Fate of Antibiotic Residues, Antibiotic-Resistant Bacteria, and Antibiotic Resistance Genes in US Dairy Manure Management Systems." *Journal of Dairy Science* 103(2):1051–71.

Pachepsky, Ya A., A. M. Sadeghi, S. A. Bradford, D. R. Shelton, A. K. Guber, and Thanh Dao. 2006. "Transport and Fate of Manure-Borne Pathogens: Modeling Perspective." *Agricultural Water Management* 86(1–2):81–92.

Peak, Nicholas, Charles W. Knapp, Richard K. Yang, Margery M. Hanfelt, Marilyn S. Smith, Diana S. Aga, and David W. Graham. 2007. "Abundance of Six Tetracycline Resistance Genes in Wastewater Lagoons at Cattle Feedlots with Different Antibiotic Use Strategies." *Environmental Microbiology* 9(1):143–51.

Rovira, Pablo, Tim McAllister, Steven M. Lakin, Shaun R. Cook, Enrique Doster, Noelle R. Noyes, Maggie D. Weinroth, Xiang Yang, Jennifer K. Parker, Christina Boucher, Calvin W. Booker, Dale R. Woerner, Keith E. Belk, and Paul S. Morley. 2019. "Characterization of the Microbial Resistome in Conventional and 'Raised Without Antibiotics' Beef and Dairy Production Systems." *Frontiers in Microbiology* 10:1980.

Ruuskanen, Matti, Johanna Muurinen, Axel Meierjohan, Katariina Parnanen, Manu Tamminen, Christina Lyra, Leif Kronberg, and Marko Virta. 2016. "Fertilizing with Animal Manure Disseminates Antibiotic Resistance Genes to the Farm Environment." *Journal of Environmental Quality* 45(2):488–93.

Sun, W, X. Qian, J. Gu, X. J. Wang, and M. L. Duan. 2016. "Mechanism and Effect of Temperature on Variations in Antibiotic Resistance Genes during Anaerobic Digestion of Dairy Manure." *Scientific Reports* 6.

Sun, Wei, Xun Qian, Jie Gu, Xiao-Juan Wang, and Man-Li Duan. 2016. "Mechanism and Effect of Temperature on Variations in Antibiotic Resistance Genes during Anaerobic Digestion of Dairy Manure." *Scientific Reports* 6(1):30237.

De Vrieze, Jo, Linde Raport, Hugo Roume, Ramiro Vilchez-Vargas, Ruy Jáuregui, Dietmar H. Pieper, and Nico Boon. 2016. "The Full-Scale Anaerobic Digestion Microbiome Is Represented by Specific Marker Populations." *Water Research* 104:101–10.

Wichmann, Fabienne, Nikolina Udikovic-Kolic, Sheila Andrew, and Jo Handelsman. 2014. "Diverse Antibiotic Resistance Genes in Dairy Cow Manure." *MBio* 5(2):e01017.

Xie, W. Y., Q. Shen, and F. J. Zhao. 2018. "Antibiotics and Antibiotic Resistance from Animal Manures to Soil: A Review." *European Journal of Soil Science* 69(1):181–95.

Youngquist, Caitlin P., Shannon M. Mitchell, and Craig G. Cogger. 2016. "Fate of Antibiotics and Antibiotic Resistance during Digestion and Composting: A Review." *Journal of Environment Quality* 45(2):537.

Zaheer, R., S. M. Lakin, R. O. Polo, S. R. Cook, F. J. Larney, P. S. Morley, C. W. Booker, S. J. Hannon, G. Van Domselaar, R. R. Read, and T. A. McAllister. 2019. "Comparative Diversity of Microbiomes and Resistomes in Beef Feedlots, Downstream Environments and Urban Sewage Influent." *BMC Microbiology* 19(1).

Zhang, Tong, Ying Yang, and Amy Pruden. 2015. "Effect of Temperature on Removal of Antibiotic Resistance Genes by Anaerobic Digestion of Activated Sludge Revealed by Metagenomic Approach." *Applied Microbiology and Biotechnology* 99(18):7771–79.

Chapter 5

Perspective: Mathematics of Horizontal Gene Transfer

SUMMARY

We present a simple mathematical framework to describe horizontal gene transfer (HGT) in microbial communities. This framework facilitates quantification and monitoring of HGT and dynamics of microbial communities. Our mathematical framework will ultimately facilitate quantification of antimicrobial resistance gene dissemination risks in different microbial communities under environmental stressors. First, we outline concepts in linear algebra central to the proposed framework. We then describe the dynamics of cell-cell interactions — *Morphogenesis*—introduced by Alan Turing. We adapt Turing's system to describe HGT. Next, we describe how we can quantify the impact of HGT on the dynamics of microbial communities from time-series data using dynamic mode decomposition. Finally, we propose how this framework may be tested in lab-scale experiments with synthetic microbial communities.

5.1 Introduction

Microbial communities are complex biological entities comprised of diverse interacting microorganisms including viruses, fungi and bacteria. Some organisms can share genetic material with other organisms in their community via horizontal gene transfer (HGT). Most notoriously, multi-drug resistant organisms have emerging in hospitals (Sheppard et al., 2016)

and veterinary settings (Liu et al., 2016) as the result of horizontally transferred antimicrobial resistance genes has increased mortality from microbial infections (Cassini et al., 19 2019).

Antimicrobial resistance genes are natural defense mechanisms microbes have developed over years competing with other microbes. Soil microbial communities are a vast reservoir of antimicrobial resistance genes (Gibson et al., 2015; Forsberg et al., 2014). Surveys of sequenced bacterial genomes in GenBank have shown that over many hundreds of thousands of years of microbial evolution, mobilized genes have spread across diverse populations of bacteria (Popa et al., 2011; Smillie et al., 2011).

While we have evidence that use of antibiotics can increase the prevalence of antimicrobial resistance in clinical settings (Sheppard et al., 2016), we have little understanding of the complexity of selection in the environment. Specifically, the livestock industry has been under increased scrutiny as novel drug-resistant pathogens have emerged with use of antibiotics (Liu et al., 2016). Emergence of resistant pathogens during antibiotic administration is not the only concern, but also the increase in diversity of resistance genes in non-pathogenic organisms residing in soil and surface water which could be acquired by pathogens long after antibiotic administration (Pruden et al., 2013). Quantify the risk of pathogens acquiring antimicrobial resistance genes in the environment due to anthropogenic pollution remains a challenge.

Several studies have developed models to predict HGT frequency. Mass-action model, for instance, are based on the theory that the transfer of genetic material occurs at a rate proportional to the density of a microbial population. One of the most classic of these mass-action models quantified the stability of plasmids within a population as a function of the trade-off between the metabolic maintenance cost and selective pressure (Levin and Stewart, 1980). This model was used to model synthetic fatty acids concentrations that could reduce conjugation

rates and cure plasmids from a population in a chemostat system (Getino et al., 2015).

Novozhilov et al. extended Stewart and Levin's horizontal gene transfer modeling to multiple

populations to evaluate parameters that lead to fixation of a novel gene in a community

(Novozhilov et al., 2005).

While these models have provided valuable insights into evolution theory and population

dynamics, a data-driven modeling approach capable of quantifying risks of horizontal gene

transfer is still needed to breakdown the complexity of gene transfer in environmental microbial

systems. We aim to develop a testable, mathematical framework to quantify horizontal gene

transfer's impact on microbial communities and risks of HGT.

5.2 Dynamics of horizontal gene transfer

Our mathematical framework for HGT is inspired by Alan Turing's *The Chemical Basis of*

Morphogenesis, which describes the impact of two cells coupling and exchanging proteins on the

dynamics of two cells (Turing, 1952). Complex microbial systems are highly dynamic

ecosystems prone to environmental and genetic disturbances. Chemical disturbance from

antibiotics can kill cells in the system, reducing redundancy and diversity. Changes in nutrient

concentrations conversely, can cause blooms in bacterial populations that can adapt to

metabolize the chemical. Genetic disturbances, such as mutations or gene transfer change the

fitness of organisms which can unbalance a system. Despite these disturbances, microbial

communities are able to maintain stability in a range of systems from biological wastewater

treatment to the human gut. The ability to maintain stability despite disturbances is the definition

of robustness (Ellens and Kooij, 2013). The following sections outline the underlying

mathematics of linear and nonlinear systems and how we propose to quantify HGT in microbial

systems.

5.2.1 Linear dynamics: Turing system

87

Applying dynamical systems theory to biological data has facilitated the discovery of cell and genome dynamics (Rajapakse and Smale, 2017b) and described fundamental phenomena such as exponential growth or decay. The most simple dynamical systems are linear and defined as:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}x \tag{1}$$

where A is a matrix of constants. For example, a two-component system of x_1 and x_2 which interact by constants $a,b,c,$ and d, we can write as:

$$\frac{dx_1}{dt} = \dot{x}_1 = ax_1 + bx_2 \tag{2}$$

$$\frac{dx_2}{dt} = \dot{x}_2 = cx_1 + dx_2 \tag{3}$$

Which simplifies to:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4}$$

The matrix of coefficient values is known as the adjacency matrix, denoted as A. This system, or network, can be also be represented as a graph, where each variable, x_1 and x_2, is represented as a node and its interaction as an edge. We can study the dynamics of these systems and evaluate important features of our network. Firstly, one can compute the Laplacian Matrix (**L**) as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where **D** is the degree matrix. Computing the eigenvalues of both **A** and **L** then can reveal spectral properties of the network. Specifically, if the real parts of the eigenvalues of the Laplacian matrix are negative, the system dynamics are stable.

In *The Chemical Basis of Morphogenesis,* Turing illustrated the breakdown of stability of two identical cells as they interact through chemical morphogens *X* and *Y*. When cells are

uncoupled, chemical reactions alone in each cell drive concentrations of morphogens (Turing, 1952). These reactions can be modeled with a system of linear equations:

$$\frac{dX}{dt} = \dot{X} = 5X - 6Y + 1 \tag{5}$$

$$\frac{dY}{dt} = \dot{Y} = 6X - 7Y + 1 \tag{6}$$

We can summarize these within-cell reaction dynamics with the following:

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 6 & -7 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{7}$$

The 2×2 matrix of reaction coefficients is the $\mathbf{A}$ matrix (Rajapakse and Smale, 2017a). The eigenvalues of $\mathbf{A}$ are $\lambda_1, \lambda_2 = \frac{trace(A) \pm \sqrt{trace(A)^2 - 4det(A)}}{2}$. Since the trace($\mathbf{A}$)$= -2$, and the det($\mathbf{A}$) $= 1$, the eigenvalues of $\mathbf{A} = -1$, with a multiplicity of 2. Since $\lambda_{1,2} < 0$, the within-cell, uncoupled dynamics are stable for each cell.

Now, if the cells couple, morphogens X and Y can diffuse between cells. Turing assigned diffusion coefficients from cell 1 to cell 2 of 0.5 for morphogen X and 4.5 for morphogen Y. Now the dynamics of the morphogens can be described as follows, where the subscript of each term represents the concentration within cell 1 or cell 2 of the coupling pair (Chua, 2005):

$$\begin{bmatrix} \dot{X_1} \\ \dot{Y_1} \\ \dot{X_2} \\ \dot{Y_2} \end{bmatrix} = \begin{bmatrix} 5 & -6 & 0 & 0 \\ 6 & -7 & 0 & 0 \\ 0 & 0 & 5 & -6 \\ 0 & 0 & 6 & -7 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{bmatrix} - \begin{bmatrix} 0.5 & 0 & -0.5 & 0 \\ 0 & 4.5 & 0 & -4.5 \\ -0.5 & 0 & 0.5 & 0 \\ 0 & -4.5 & 0 & 4.5 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{bmatrix} \tag{8}$$

We can think of the diffusion matrix as the Laplacian matrix (Rajapakse and Smale, 2017a). The eigenvalues of the diffusion/Laplacian matrix, $\mathbf{L}$, are $[-9, -1, 0, 0]$, so the diffusion matrix is also stable. We can also simplify the equation above in terms of the cell dynamics, $\mathbf{A}$, and the diffusion dynamics, $\mathbf{L}$, as $\frac{dx}{dt} = (\mathbf{A} - \mathbf{L})x$. The final set of matrix values are as follows:

$$
\begin{bmatrix} \dot{X_1} \\ \dot{Y_1} \\ \dot{X_2} \\ \dot{Y_2} \end{bmatrix} = \begin{bmatrix} 4.5 & -6 & 0.5 & 0 \\ 6 & -11.5 & 0 & 4.5 \\ 0.5 & 0 & 4.5 & -6 \\ 0 & 4.5 & 6 & -11.5 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{bmatrix} \tag{9}
$$

Now, one of the eigenvalues of the system describing the within cell dynamics and between cell

diffusion system is positive, meaning the system is no longer stable. This two-cell coupling

resulting in a change in system dynamics is analogous to horizontal gene transfer.

5.2.2 Nonlinear dynamics

Many systems in nature are nonlinear, classic Monod growth kinetics of monocultures in nutrient

limited environments are nonlinear as are competitive and generalized Lokta-Volterra models for

interacting species. Nonlinear systems are defined as:

$$
\frac{d\mathbf{x}}{dt} = F(\mathbf{x}) \tag{10}
$$

where, as opposed to Equation 1, $F(\mathrm{x})$ is a function that is nonlinear, so the derivative of

$F(\mathbf{x})$, $\frac{d}{dx}F(\mathbf{x})$, does not equal a constant. However, we can still explore the dynamics using the

properties of linear systems by linearizing the system and observing the dynamics in the

neighborhood of equilibrium points, which occur where $\frac{d}{dx}F(\mathbf{x}) = 0$. The Jacobian matrix, J, is

a matrix of partial derivatives at the equilibrium defined as:

$$
J = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \frac{\partial F_i}{\partial x_i} & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} \end{pmatrix} \tag{11}
$$

If any of the eigenvalues of the Jacobian solved at the equilibrium point have a positive real part,

the system is unstable as it approaches that equilibrium point. Computing eigenvalues of the

Jacobian Matrix can also reveal bifurcations in a dynamical system. Bifurcations occur when a

parameter change causes the stability of an equilibrium point to emerge or break down. The Hopf

bifurcation, for example, is notable bifurcation observed in population dynamics, which occurs

when the parameter, μ, called the bifurcation parameter, is introduced causing a stable system to

destabilize (Fussmann et al., 2000). In the following section we illustrate this concept in the

context of microbial systems.

5.2.3 Horizontal gene transfer in robust microbial networks

Many different structural features can contribute to biological robustness (Kitano, 2004). In

microbial communities, functional redundancy and diversity have been observed to contribute to

stability of the system (Shade et al., 2012). We can visualize the stability of a microbial

community using stability landscapes in Figure 5-1. Basins within a landscape represent stable

states in a system. Disturbances can push a system out of a stability basin which can cause the

system to stabilize in a new state or basin or remain unstable.
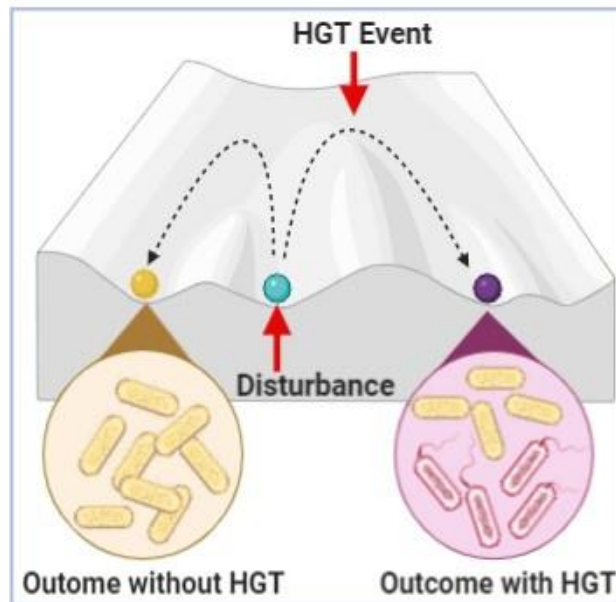
Figure 5-1: *Microbial communities are prone to environmental and genetic disturbances that can impact the fate of the community's structure.*

Horizontal gene transfer enables functional plasticity via near instantaneous adaptation to environmental pressures or conditions. Therefore, we hypothesize that HGT is a 'fail-safe' mechanism, a feature of robust networks. Other modeling work has demonstrated that HGT makes systems more robust (Juhasz et al., 2014; Fan et al., 2018). Specifically, Fan *et al.* applied a game theory-based modeling framework to show that HGT enabled a wider range of initial conditions enabling stable species co-existence (Fan et al., 2018). Juhasz *et al.* found that HGT in simulated microbial communities made the communities more robust to invading organisms using agent-based modeling (Juhasz et al., 2014). We now extend the HGT coupling to a microbial community consisting of three different organisms, x_1, x_2 and x_3. Each organism interacts with weights defined as $w_{i,j}$ which reflect the net mutualistic, commensal, competitive or predatory interactions between two organisms. This is analogous to the generalized Lokta-Volterra equations used to describe ecological networks including microbial communities. The dynamics of the system can be summarized as:

$$F(\mathbf{x}) = \mathbf{A}\mathbf{x} \tag{12}$$

Where **A** is the matrix of edge weights:

$$\mathbf{A} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \tag{13}$$

We then define horizontal gene transfer as a coupling mechanism that changes edge dynamics:

$$F(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{L}\mathbf{x} \tag{14}$$

Here, **L** is a matrix that alters the weight of a species interaction due to horizontal gene transfer.

5.3 Learning from data

One important benefit of this framework is the ability to derive the dynamics from data. Here we outline how dynamics of microbial communities can be quantified from data by applying dynamic mode decomposition (DMD) to longitudinal sampling of an infant's microbiome. Next, we outline how the approach can be leveraged for evaluating a community's susceptibility to horizontal gene transfer under different thresholds of environmental stress.

5.3.1 Dynamic mode decomposition

Before we can assess the stability and robustness of a system, we need to determine the underlying system dynamics, or specifically the adjacency matrix. Other approaches have been used to infer microbial community dynamics, based on multiple linear regression techniques which can inform generalized Lokta-Volterra models (Ruan et al., 2006; Faust and Raes, 2012; Stein et al., 2013; Bucci et al., 2016). Here we propose applying DMD to compute dynamic modes from time series observations.

The benefit of extracting network dynamics from real data using DMD facilitates quantification of the network's robustness. Metrics which quantify features of robust communities enable comparison of different microbial communities or between community states. Robustness measures can enable quantification of the role of HGT in community stability before and after HGT events, between different communities with different gene transfer frequencies, or before and after environmental stressors.

Using the eigenvalues computed from DMD, a couple different metrics of robustness can be determined. The algebraic connectivity of a network is the second smallest eigenvalue and has been used to infer the robustness of protein-protein networks within bacteria (Guimaraes et al., 2018). Effective resistance is another metric of robustness which is a function of Laplacian eigenvalues. Comparing effective resistance between microbial networks has demonstrated that fungi contribute to microbial stability in human microbiomes (Tipton et al., 2018).

To illustrate the steps to compute system dynamics from data, we first denote the sequence of snapshots collected by the following description:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{m-1}] \in^{n \times m-1} \tag{15}$$

$$\mathbf{X}' = [\mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_m] \in^{n \times m-1} \tag{16}$$

where, m is total number of snapshots and $\mathbf{X}'$ is the time-shifted snapshot matrix of $\mathbf{X}$. Let us suppose that there is an unknown linear operator $\mathbf{A}$ such that

$$\mathbf{X}' = \mathbf{A}\mathbf{X}. \tag{17}$$

The dynamic mode decomposition of the data matrix pair $\mathbf{X}$ and $\mathbf{X}'$ is given by the eigendecomposition of $\mathbf{A}$. We may think of $\mathbf{A}$ as describing a high dimensional linear regression of the nonlinear dynamics which relate $\mathbf{X}$ to $\mathbf{X}'$. To obtain an approximation of $\mathbf{A}$, one approach is to use the singular value decomposition (SVD) of the data matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$ to compute its pseudoinverse: $\mathbf{A} \approx \mathbf{X}'\mathbf{X}^\dagger$. The DMD modes are the eigenvectors of the DMD approximated adjacency matrix and represent different components of the dynamics. The eigenvalues, as explained in the *Dynamics of Horizontal Gene Transfer* section reveal the behavior of the system.

In summary, DMD enables data-driven estimation of a system's dynamics and can be applied to study the dynamics of microbial communities to quantify stability of the community. We now illustrate how DMD can be applied on a real microbiome dataset then describe how the approach can be used to quantify the impact of horizontal gene transfer on the stability of microbes facing environmental stress.

5.3.2 Example: Infant gut longitudinal cohort

To illustrate how DMD can be used to infer the dynamics of a system based on time series observations, we explore the dynamics of the gut microbiome from a subject in a longitudinal

cohort examining infant gut microbiome development. Data from the Early Childhood

Antibiotics and the Microbiome (ECAM) study were downloaded from "MicrobiomeDB," an

online platform containing microbial community composition data originating from thousands of

samples (Oliveira et al., 2018). Microbiome data at the "class" taxonomic level from the subject

with the most sample points was downloaded.

Every month over the first year of the infant's life and every two months in the second

year, a fecal sample was collected. DNA was extracted from the samples and 16S rRNA genes

were amplified and sequenced to identify and estimate relative abundances of taxa (Bokulich et

al., 2016). At each point in time, we then obtained a vector, x_t, which contains the relative

abundance of each class at time, t. We can combine these vectors into an $n \times m$ matrix where m

is the number of observations and n is the number of microbial classes observed in the dataset.

$$subject_1 = \begin{bmatrix} | & | & & | \\ x_{t_1} & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix}_{n \times m} \tag{18}$$

Next, we can perform DMD to estimate the dynamics of the microbial communities.

Eigenvalues within the unit circle represent decaying portions of the dynamics, values on the unit

circle represent oscillatory behavior and outside the unit circle are unstable dynamics. We found

that the dynamic modes from the subject are negative, meaning they are stabilizing (Figure 5-2).

This agrees with the findings in the original manuscript that as the infant ages, the microbial

composition stabilizes (Bokulich et al., 2016). Now that we have introduced DMD and

demonstrated the process on real microbiome data from a longitudinal cohort, we can explore

how the approach can reveal the relationship between HGT and microbial community

robustness.
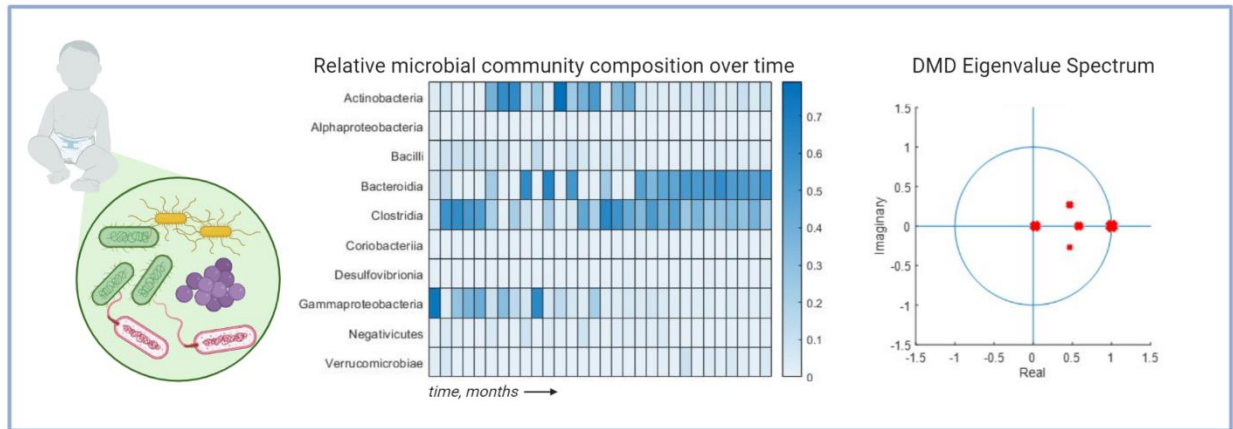
***Figure 5-2:** DMD modes computed from subject's microbial community profile measured over time.*
Time series data depicted by the heatmap. DMD eigenvalues within the unit circle demonstrate that the system dynamics are stabilizing over time.

5.3.3 Robustness and stress thresholds: Quantifying HGT in model microbiomes

It is near impossible to track real-time gene transfer events within subjects or in

environmental microbiomes such as those in agricultural soils because *in situ* gene transfer is

difficult to observe (Sorensen *et al.* 2005). However, synthetic microbial communities offer an

exciting opportunity for experimenting with drivers and outcomes of gene transfer. For example,

to quantify the impact of HGT on community stability, one could take snapshots of a synthetic

microbial community over time and use tools like susceptibility testing or genotyping to evaluate

if an HGT event has occurred in the given time window. The observations of the community

structure over time can be bisected into pre-HGT and post-HGT and DMD performed on each

subset to quantify the impact of the gene transfer event.

The risk of pathogenic organisms acquiring antimicrobial resistance genes through HGT

is also a critical research question this framework, coupled with synthetic microbial

communities, can address. By performing experiments subjecting synthetic microbiomes to

different environmental stressors, like antibiotic loading, this modeling framework can be used to

quantify thresholds of stress that lead to HGT events. Paired experiments involving communities

with and without HGT potential can quantitatively compare the robustness of communities

(Figure 5-3).
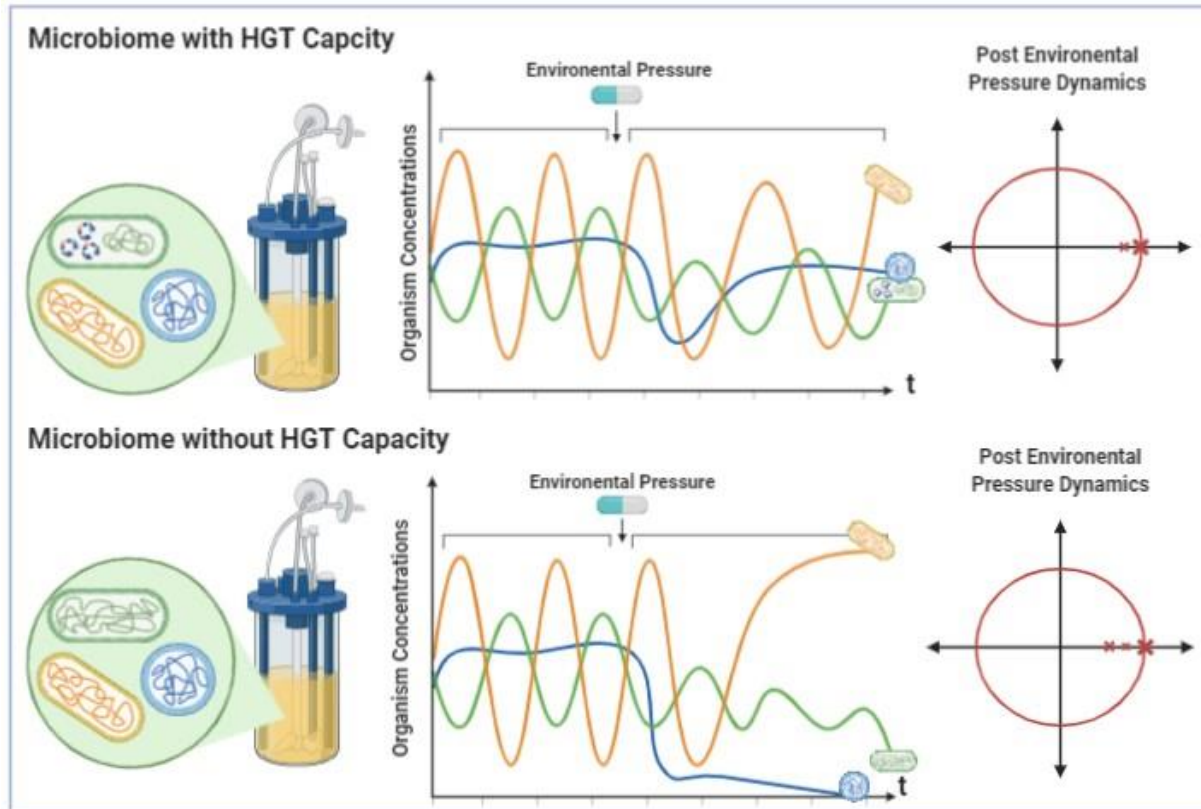
Figure 5-3: *Example experimental set up to quantitatively compare robustness of communities with and without HGT potential.*
One microbial community is capable of gene exchange (top) and one is not capable of HGT (bottom). The systems capable of HGT in this illustration rebounds under environmental stress while the system without HGT capacity is not robust to the environmental stress. These differences in robustness can be by quantified from the eigenvalues represented in the unit circle (right).

5.4 Importance

In this manuscript we have introduced a novel framework with which to explore the impact HGT

has in disrupting or maintaining stability within microbial communities. We demonstrated how

principles in linear and non-linear algebra enable inference of the microbial community networks

and identify features of robust microbial communities. Lastly, we proposed a testable framework

for evaluating the role of HGT that can be applied to answer different scientific questions.

5.5 References

Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science translational medicine* 8, 219 343ra82–343ra82

Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., et al. (2016). Mdsine: Microbial 221 dynamical systems inference engine for microbiome time-series analyses. *Genome biology* 17, 1–17

Cassini, A., Hogberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., et al. (2019).¨Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant 224 bacteria in the eu and the european economic area in 2015: a population-level modelling analysis. *The* 225 *Lancet infectious diseases* 19, 56–66

Chua, L. O. (2005). Local activity is the origin of complexity. *International journal of bifurcation and* 227 *chaos* 15, 3435–3456

Ellens, W. and Kooij, R. E. (2013). Graph measures and network robustness. *arXiv preprint* arXiv:1311.5064

Fan, Y., Xiao, Y., Momeni, B., and Liu, Y.-Y. (2018). Horizontal gene transfer can help maintain the 231 equilibrium of microbial communities. *Journal of theoretical biology* 454, 53–59

Faust, K. and Raes, J. (2012). Microbial interactions: from networks to models. *Nature Reviews* 233 *Microbiology* 10, 538–550

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., et al. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature* 509, 612–616

Fussmann, G. F., Ellner, S. P., Shertzer, K. W., and Hairston Jr, N. G. (2000). Crossing the hopf bifurcation in a live predator-prey system. *Science* 290, 1358–1360

Getino, M., Sanabria-R´ıos, D. J., Fernandez-Lopez, R., Campos-Gomez, J., Sanchez-Lopez, J. M., Fernandez, A., et al. (2015). Synthetic fatty acids prevent plasmid-mediated horizontal gene transfer. MBio 6

Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME journal* 9, 207–216

Guimaraes, P. R., Pires, M. M., Cantor, M., and Coltri, P. P. (2018). Interaction paths promote module~ 244 integration and network-level robustness of spliceosome to cascading effects. *Scientific reports* 8, 1–11

Juhasz, J., Kertesz-Farkas, A., Szabo, D., and Pongor, S. (2014). Emergence of collective territorial defense in bacterial communities: horizontal gene transfer can stabilize microbiomes. *PloS one* 9

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics* 5, 826–837

Levin, B. R. and Stewart, F. M. (1980). The population biology of bacterial plasmids: a priori conditions for the existence of mobilizable nonconjugative factors. *Genetics* 94, 425–443

Liu, Y.-Y., Wang, Y., Walsh, T. R., Yi, L.-X., Zhang, R., Spencer, J., et al. (2016). Emergence of plasmid- mediated colistin resistance mechanism mcr-1 in animals and human beings in china: a microbiological and molecular biological study. *The Lancet infectious diseases* 16, 161–168

Novozhilov, A. S., Karev, G. P., and Koonin, E. V. (2005). Mathematical modeling of evolution of horizontally transferred genes. *Molecular Biology and Evolution* 22, 1721–1732

Oliveira, F. S., Brestelli, J., Cade, S., Zheng, J., Iodice, J., Fischer, S., et al. (2018). Microbiomedb: a systems biology platform for integrating, mining and analyzing microbiome experiments. *Nucleic acids research* 46, D684–D691

Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome research* 21, 599–609

Pruden, A., Larsson, D. J., Amezquita, A., Collignon, P., Brandt, K. K., Graham, D. W., et al. (2013). Management options for reducing the release of antibiotics and antibiotic resistance genes to the environment. *Environmental health perspectives* 121, 878–885

Rajapakse, I. and Smale, S. (2017a). Emergence of function from coordinated cells in a tissue. *Proceedings of the National Academy of Sciences* 114, 1462–1467

Rajapakse, I. and Smale, S. (2017b). Mathematics of the genome. *Foundations of Computational Mathematics* 17, 1195–1217

Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A., and Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22, 2532–2538

Shade, A., Peter, H., Allison, S. D., Baho, D., Berga, M., Burgmann, H., et al. (2012). Fundamentals of˙microbial community resistance and resilience. *Frontiers in microbiology* 3, 417

Sheppard, A. E., Stoesser, N., Wilson, D. J., Sebra, R., Kasarskis, A., Anson, L. W., et al. (2016). Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blakpc. *Antimicrobial agents and chemotherapy* 60, 3767–3778

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244

Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N., and Wuertz, S. (2005). Studying plasmid horizontal 279 transfer in situ: a critical review. *Nature Reviews Microbiology* 3, 700–710

Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Ratsch, G., Pamer, E. G., et al. (2013). Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol* 9, e1003388

Tipton, L., Muller, C. L., Kurtz, Z. D., Huang, L., Kleerup, E., Morris, A., et al. (2018). Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome* 6, 12

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal* Society of London 237, 37–72

Chapter 6

Conclusions and Future Work

This work has resulted in an improved understanding of the role that manure management plays in mitigating antimicrobial resistance gene (ARG) abundances and brought the field closer to achieving a risk-assessment framework for ARGs in the environment. Previous research on anthropogenic releases of ARGs into the environment have been largely exploratory. Interventions effective in mitigating ARG abundances and the relative importance of different sources of ARGs in the environment is still unknown. To approach a consensus on the role of the agricultural sector in ARG pollution and to enable risk assessment frameworks capable of informing policy, we applied research principles in epidemiology to environmental engineering systems and developed novel quantitative molecular and mathematical tools.

6.1: Applying environmental epidemiology in environmental engineering studies

In Chapters 2 and 4, principles in environmental epidemiology study design were applied through a meta-analysis and cross-sectional study, respectively. Meta-analyses are important in the public health and clinical fields to assess the generalizability of findings between studies (1). While literature reviews published over the past years have qualitatively described publications documenting ARG abundances in livestock environment (2, 3), these reviews were broadly performed across different livestock species and have concluded more research is needed to inform policy to evaluate the risk of pathogens airing ARGs in the environment. The systematic review and quantitative meta-analysis of field-scale research in cattle livestock operations

101

performed in Chapter 2 demonstrated cross-study variability in observed impacts of antibiotic use and land application on environmental ARG abundances. Studies often lacked enough biological replication, preventing the computation of statistical variation and significance. These observations highlighted the importance of the cross-sectional study designs for evaluating drivers of ARG abundances.

Cross-sectional studies are an important tool in environmental epidemiology for quantifying relationships between exposures and health outcomes (4). Cross-sectional studies in ARG research have demonstrated relationships between agricultural exposure and carriage of Methicillin-resistant Staphylococcus aureus (5) and prevalence resistant organisms in conventional farms compared to raised-without-antibiotic farms (6). In Chapter 4 we applied a pilot-scale cross-section study of manures from 12 dairy farms as manures were applied to land. This study further demonstrated the value of cross-sectional studies designs for identifying relationships in farm management and microbial ecology. We were able to demonstrate that anaerobic digestion (AD) indeed shaped the microbial ecology of stored manures. Using unsupervised statistical approaches, we could extract populations of bacteria and antimicrobial resistance genes prevalent in stored manures with and without AD prior to storage. We observed that the AD was a stronger driver of microbial community structure than the storage season, but a larger population of dairies needs to be surveyed to determine how geographic region and herd size shape the communities and ARG abundances.

The meta-analysis and cross-sectional study demonstrated that exploratory study designs for environmental ARG research are obsolete. Well-designed studies are needed for evaluating field-scale interventions for mitigating ARGs and reducing risk.

6.2: Novel tools to approach risk assessment framework

Chapters 3 and 5 introduced novel approaches to study ARGs in complex environmental matrices. In Chapter 3 we introduce a quantitative metagenomic approach which enabled non-targeted screening of ARGs using standard spike-in DNA standards. This quantitative metagenomic approach filled a critical need for absolute quantification of genes (gene copies per sample mass or volume) in complex environmental systems. We demonstrate the value of this novel approach in Chapter 4, in which we apply the method in a cross-section study of stored manures. Using a high-throughput quantitative metagenomic tool, we could screen the entire resistome and better compare the impact of anaerobic digestion on stored manure ARG concentrations.

Lastly, in Chapter 5, we proposed a simple, data-driven mathematical framework for quantifying risks of pathogen acquisition of ARGs via horizontal gene transfer (HGT); this had been a critical limitation of previous approaches identified in several review papers (7, 8). HGT involves the coupling of microbial cells and has been implicated in the dissemination of ARGs in clinical environments (9). With microbial community composition data over time, we outline how Dynamic Mode Decomposition (DMD) can extract the dynamics of a microbial network. Coupled with synthetic microbiomes, like "the hitchhikers of the rhizosphere" system (10) we describe experimentation that can be used to compute robustness of microbial communities with and without HGT potential and subject to environmental stress. After experimental validation of this approach, we envision it can be used to predict if residual antibiotics and ARGs in land applied manure destabilize agricultural soils, increasing the risk of HGT of ARGs to pathogenic organisms.

Both the quantitative metagenomic approach and mathematical framework introduced tools required for assessing risks of ARGs in environmental systems. However, the tools are both

more broadly applicable for a wide range of research questions in microbial ecology. For instance, the quantitative metagenomic approach could assess performance of engineered biological processes. The data-driven modeling framework proposed offers a promising approach to quantify microbial community robustness which can inform operation of engineered systems for more stable performance and advance in personalized medicine.

6.3 References

1.   Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, others. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4:1.

2.   Heuer H, Schmitt H, Smalla K. 2011. Antibiotic resistance gene spread due to manure application on agricultural fields. Curr Opin Microbiol 14:236–243.

3.   Xie W-Y, Shen Q, Zhao FJ. 2018. Antibiotics and antibiotic resistance from animal manures to soil: a review. Eur J Soil Sci 69:181–195.

4.   CDC. 2019. Principles of Epidemiology in Public Health Practice, Third Edition An Introduction to Applied Epidemiology and Biostatistics.

5.   Wang XL, Li L, Li SM, Huang JY, Fan YP, Yao ZJ, Ye XH, Chen SD. 2017. Phenotypic and molecular characteristics of Staphylococcus aureus and methicillin-resistant Staphylococcus aureus in slaughterhouse pig-related workers and control workers in Guangdong Province, China. Epidemiol Infect 145:1843–1851.

6.   Rinsky JL, Nadimpalli M, Wing S, Hall D, Baron D, Price LB, Larsen J, Stegger M, Stewart J, Heaney CD. 2013. Livestock-associated methicillin and multidrug resistant Staphylococcus aureus is present among industrial, not antibiotic-free livestock operation workers in North Carolina. PLoS One 8:e67641.

7.   Pruden A, Alcalde RE, Alvarez PJJ, Ashbolt N, Bischel H, Capiro NL, Crossette E, Frigon D, Grimes K, Haas CN, others. 2018. An environmental science and engineering framework for combating antimicrobial resistance. Environ Eng Sci 35:1005–1011.

8.   Vikesland PJ, Pruden A, Alvarez PJJ, Aga D, Burgmann H, Li X, Manaia CM, Nambi I, Wigginton K, Zhang T, Others. 2017. Toward a comprehensive strategy to mitigate dissemination of environmental sources of antibiotic resistance. ACS Publications.

9.   Sheppard A, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, Giess A, Pankhurst LJ, Vaughan A, Grim CJ, Cox HL, Yeh AJ, Sifri CD, Walker AS, Peto TE, Crook DW, Mathers AJ. 2016. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene bla KPC 60:3767–3778.

10.   Lozano GL, Bravo JI, Diago MFG, Park HB, Hurley A, Peterson SB, Stabb E V, Crawford JM, Broderick NA, Handelsman J. 2019. Introducing THOR, a model microbiome for genetic dissection of community behavior. MBio 10.

APPENDICES

Supplementary Material for Chapter 2
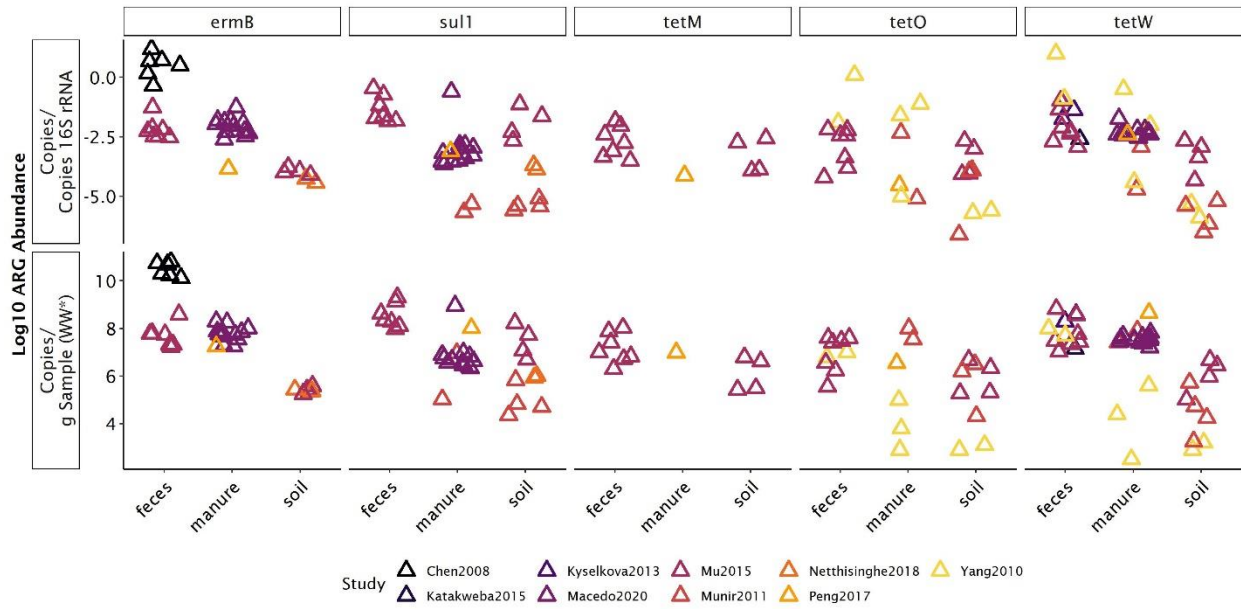
Figure A-1: *Studies reporting both ARG copies in units of gene copies/sample mass and gene copies/copies 16S rRNA gene copies.*
WW= wet weight
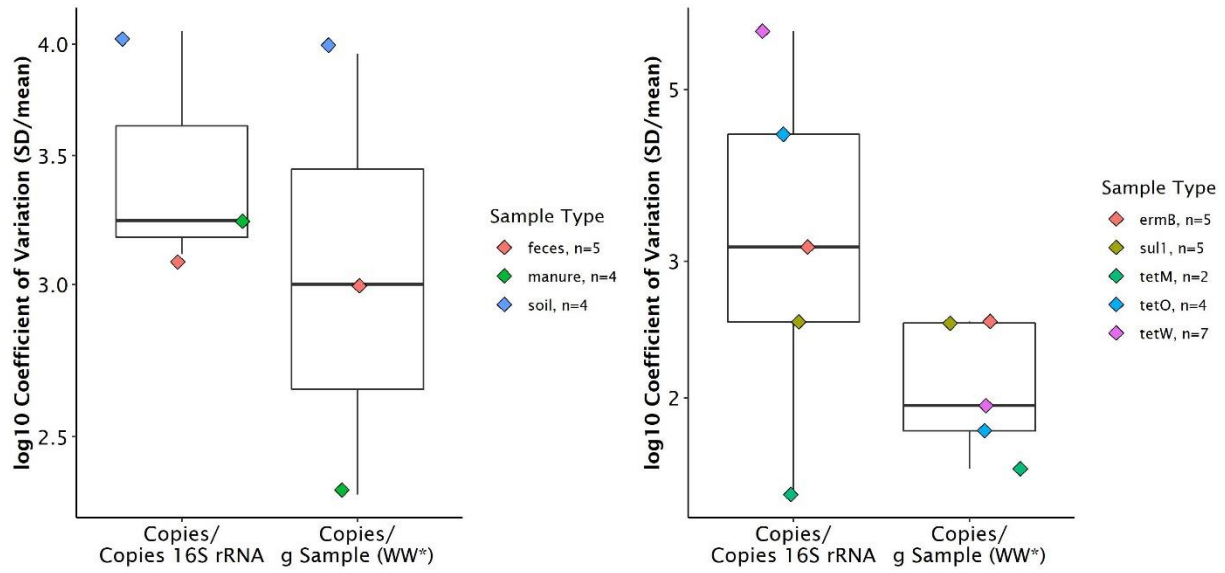
Figure A-2: *Coefficient of variation between different reported units of gene abundances*

Table A-1: *Known Relevant articles (KRAs) Used to identify key search terms and assess search completeness.*

Article	Available in Database		
	PM	**AG**	**SC**
Dungan, R. S.; Mckinney, C. W.; Leytem, A. B. Tracking Antibiotic Resistance Genes in Soil Irrigated with Dairy Wastewater. Sci. Total Environ. **2018**, 635, 1477–1483.	X	X	X
Fahrenfeld, N.; Knowlton, K.; Krometis, L. A.; Hession, W. C.; Xia, K.; Lipscomb, E.; Libuit, K.; Green, B. L.; Pruden, A. Effect of Manure Application on Abundance of Antibiotic Resistance Genes and Their Attenuation Rates in Soil: Field-Scale Mass Balance Approach. Environ. Sci. Technol. **2014**, 48 (5), 2643–2650.	X	X	X
Marti, R.; Tien, Y.-C.; Murray, R.; Scott, A.; Sabourin, L.; Topp, E. Safely Coupling Livestock and Crop Production Systems: How Rapidly Do Antibiotic Resistance Genes Dissipate in Soil Following a Commercial Application of Swine or Dairy Manure? Appl. Environ. Microbiol. **2014**, 80 (10), 3258–3265.	X		X
Munir, M.; Xagoraraki, I. Levels of Antibiotic Resistance Genes in Manure, Biosolids, and Fertilized Soil. J. Environ. Qual. **2011**, 40 (1), 248–255.	X	X	X
McKinney, C. W.; Dungan, R. S.; Moore, A.; Leytem, A. B. Occurrence and Abundance of Antibiotic Resistance Genes in Agricultural Soil Receiving Dairy Manure. FEMS Microbiol. Ecol. **2018**, 94 (3), 1–10.	X		X
Wallace, J. S.; Garner, E.; Pruden, A.; Aga, D. S. Occurrence and Transformation of Veterinary Antibiotics and Antibiotic Resistance Genes in Dairy Manure Treated by Advanced Anaerobic Digestion and Conventional Treatment Methods. Environ. Pollut. **2018**, 236, 764–772.	X	X	X
Vikram, A.; Rovira, P.; Agga, G. E.; Arthur, T. M.; Bosilevac, J. M.; Wheeler, T. L.; Morley, P. S.; Belk, K. E.; Schmidt, J. W. Impact of "Raised without Antibiotics" Beef Cattle Production Practices on Occurrences of Antimicrobial Resistance. Appl. Environ. Microbiol. **2017**, 83 (22).	X		X
Tien, Y. C.; Li, B.; Zhang, T.; Scott, A.; Murray, R.; Sabourin, L.; Marti, R.; Topp, E. Impact of Dairy Manure Pre-Application Treatment on Manure Composition, Soil Dynamics of Antibiotic Resistance Genes, and Abundance of Antibiotic-Resistance Genes on Vegetables at Harvest. Sci. Total Environ. **2017**, 581–582, 32–39.	X	X	X
Chen, J.; Fluharty, F. L.; St-Pierre, N.; Morrison, M.; Yu, Z. Technical Note: Occurrence in Fecal Microbiota of Genes Conferring Resistance to Both Macrolide-Lincosamide-Streptogramin B and Tetracyclines Concomitant with Feeding of Beef Cattle with Tylosin. J. Anim. Sci. **2008**, 86 (9), 2385–2391.	X	X	X
Li, J.; Hu, H.; Li, H.; Chen, J.; Yan, J.; Zhang, Y.; Xin, Z. Long-Term Manure Application Increased the Levels of Antibiotics and Antibiotic Resistance Genes in a Greenhouse Soil. Appl. soil Ecol. 121 (2010224060; IND605844576), 193–200. **2017**		X	X

PM = Pubmed, AG = AGRICOLA, SC = Scopus.

Table A-2: *Summary of searches and date conducted.*

Database	Search	Date	Results
PubMed (PM)	(((((((((cattle[Title/Abstract]) OR (dairy[Title/Abstract])) OR (bovine[Title/Abstract])) OR (cow[Title/Abstract])) OR (beef[Title/Abstract])) OR (cattle[MeSH Terms])) OR (dairying[MeSH Terms])) AND ((((manure[Title/Abstract]) OR (feces[Title/Abstract])) OR (waste[Title/Abstract])) OR (fecal[Title/Abstract]))) AND ((((drug resistance, microbial[MeSH Terms]) OR (drug resistance, bacterial[MeSH Terms])) OR (antimicrobial resistance[Title/Abstract])) OR (antibacterial resistance[Title/Abstract]))) NOT (((isolated[Title]) OR (salmonella[Title])) OR (escherichia coli[Title]))	6 July 2020	262
Scopus (SC)	(TITLE-ABS-KEY (("antimicrobial resistan*" AND gene) OR "antibiotic resistan* gene") AND TITLE-ABS-KEY (manure OR feces OR waste) AND TITLE-ABS-KEY ("polymerase chain reaction" OR "PCR" OR "QPCR" OR "quantitative real time PCR" OR "quantitative polymerase chain reaction" OR "quantitative PCR" OR "abundance" OR "relative abundance" OR "gene transfer") AND TITLE-ABS-KEY (dairy OR beef OR cattle OR bovine OR cow) AND NOT TITLE-ABS-KEY ("isolate"))	26 May 2020	166
Agricola (AG)	(su("anaerobic digesters" OR "feces" OR "anaerobic digestion" OR "animal manure management" OR "cattle manure" OR "composted manure" OR "Effluents" OR "Lagoons" OR "land application" OR "management systems" OR "manure spreading" OR "manure storage" OR "Ponds" OR "waste lagoons" OR "wastewater irrigation" OR "soil amendments" OR "land application" OR "manure spreading" OR "slurries" OR "aeration") OR ti("anaerobic digesters" OR "feces" OR "anaerobic digestion" OR "animal manure management" OR "cattle manure" OR "composted manure" OR "Effluents" OR "Lagoons" OR "land application" OR "management systems" OR "manure spreading" OR "manure storage" OR "Ponds" OR "waste lagoons" OR "wastewater irrigation" OR "soil amendments" OR "land application" OR "manure spreading" OR "slurries" OR "aeration") OR ab("anaerobic digesters" OR "feces" OR "anaerobic digestion" OR "animal manure management" OR "cattle manure" OR "composted manure" OR "Effluents" OR "Lagoons" OR "land application" OR "management systems" OR "manure spreading" OR "manure storage" OR "Ponds" OR "waste lagoons" OR "wastewater irrigation" OR "soil amendments" OR "land application" OR "manure spreading" OR "slurries" OR "aeration")) AND (su("cattle" OR "dairy manure" OR "cattle manure" OR "manure" OR "dairy farming" OR "dairy cattle" OR "beef cattle" OR "dairy cows") OR ti("cattle" OR "dairy manure" OR "cattle manure" OR "manure" OR "dairy farming" OR "dairy cattle" OR "beef cattle" OR "dairy cows") OR ab("cattle" OR "dairy manure" OR "cattle manure" OR "manure" OR "dairy farming" OR "dairy cattle" OR "beef cattle" OR "dairy cows")) AND (su(antibiotic resistance genes OR antibiotic resistance) OR ti("antibiotic resistance genes" OR "antimicrobial resistance genes") OR ab("antibiotic resistance genes" OR "antimicrobial resistance genes")) NOT (ti("isolate*") NOT (su(isolation OR culturing))	7 July 2020	390

Table A-3: *CARD Ontology and ARG Targets identified in 43 papers included in meta-analysis*

CARD Drug Class	Figure Drug Class	CARD AMR Gene Family	Gene Target (CARD name)
tetracycline antibiotic	tetracycline	major facilitator superfamily (MFS) antibiotic efflux pump	*tet(A), tet(A/C), tet(B), tet(C), tet(E), tet(G), tet(H), tet(L), tetY, tetZ*
		tetracycline-resistant ribosomal protection protein	*tet(RPP), tet(W), tetB(P), tetM, tetO, tetQ, tetS, tetT, tetW*
		tetracycline inactivation enzyme	*tetX*
sulfonamide antibiotic	sulfonamide	sulfonamide resistant sul	*sul1, sul2, sul3, sulA*
beta-lactam antibiotic	beta-lactam	CfxA beta-lactamase	*cfxA*
		class C beta-lactamase	*ampC*
		CMY beta-lactamase	*CMY-2*
		CTX-M beta-lactamase	*CTX-M, CTX-M-1*
		KPC beta-lactamase	*KPC-2*
		methicillin resistant PBP2	*mecA*
		other beta-lactamase	*bla2, bla3, bla4, bla-CEP-01, bla-CEP-02, bla-CEP-03, bla-CEP-04, bla-CEP-05, bla-PSE*
		OXA beta-lactamase	*OXA-1, OXA-2, OXA-20, OXA-58*
		ROB beta-lactamase	*ROB*
		TEM beta-lactamase	*TEM, TEM-1*
streptogramin antibiotic, macrolide antibiotic, lincosamide antibiotic	MLS	Erm 23S ribosomal RNA methyltransferase	*ermA, ermB, ermC, ermF, ermT, ermX*
		other Erm	*erm*
aminoglycoside antibiotic	aminoglycoside	ambiguous aminoglycoside antibiotic	*str*
		aminoglycoside acetyltransferase (AAC)	*aac*
		aminoglycoside phosphotransferase (APH)	*aph*
		ANT(3")	*aadA*

		ANT(4)	*ANT(4)-Ia*
		APH(2"), AAC(6')	*AAC(6')-Ie-APH(2")-Ia*
		APH(3")	*APH(3")-Ib*
		APH(6)	*APH(6)-Id, str(B)*
fluoroquinolone antibiotic	fluoroquinolone	quinolone resistance protein (qnr)	*qnrA, qnrD, qnrS*
other	other	integrase (not in card)	*int1, int2*
		major facilitator superfamily (MFS) antibiotic efflux pump	*mdtL*
		replication initiation protein RepA	*repA*
phenicol antibiotic	phenicol	chloramphenicol acetyltransferase (CAT)	*cat*
		major facilitator superfamily (MFS) antibiotic efflux pump	*floR*
rifamycin antibiotic	rifamycin	rifamycin-resistant beta-subunit of RNA polymerase (rpoB)	*rpoB*
glycopeptide antibiotic	glycopeptide	van ligase, glycopeptide resistance gene cluster	*vanA*
macrolide antibiotic	macrolide	macrolide phosphotransferase (MPH)	*mphA*
tetracycline antibiotic, glycylcycline, rifamycin antibiotic, cephalosporin, phenicol antibiotic, penam, triclosan, fluoroquinolone antibiotic	multiple	resistance-nodulation-cell division (RND) antibiotic efflux pump	*acrB*
tetracycline antibiotic, lincosamide antibiotic, phenicol antibiotic, pleuromutilin antibiotic, oxazolidinone antibiotic, streptogramin antibiotic, macrolide antibiotic	multiple	ABC-F ATP-binding cassette ribosomal protection protein	*mel*

tetracycline antibiotic, nitrofuran antibiotic, fluoroquinolone antibiotic, glycylcycline, diaminopyrimidine antibiotic	multiple	resistance-nodulation-cell division (RND) antibiotic efflux pump	*oqxB*
tetracycline antibiotic, penam, rifamycin antibiotic, glycylcycline, cephalosporin, phenicol antibiotic, triclosan, fluoroquinolone antibiotic	multiple	resistance-nodulation-cell division (RND) antibiotic efflux pump	*acrA*

Table A-4: *Antibiotic-Control Response Ratio (RR) Summary statistics by meta-data grouping*

	# Samples	# Studies	# ND*	Mean RR	Median RR
Antibiotic Administered					
"Conventional"	7	3	10	0.99	0.46
CCFA	5	2	0	0.47	0.31
chlortetracycline	3	2	0	1.24	1.11
neomycin + oxytetracycline	2	1	4	0.07	-0.03
not reported	1	1	0	-1.55	-1.11
oxytetracycline	4	3	8	0.80	0.71
tulathromycin	1	1	8	-0.24	-0.24
tylosin	4	2	6	-0.13	-0.18
Sample Type					
compost	2	1	0	0.33	0.37
feces	21	9	36	0.40	0.36
manure	3	2	0	2.17	2.30
Study					
Alali2009	3	-	0	0.72	1.09
Beukers2018	4	-	8	0.32	0.35
Chen2008	3	-	6	-0.06	0.00
Holman2019	2	-	16	0.14	0.33
Kanwar2014	3	-	0	0.98	0.85
Keijser2019	2	-	0	0.42	0.41
Kyselkova2013	1	-	0	1.75	0.85
Peak2007	2	-	0	4.03	3.68
Sharma2009	2	-	0	0.33	0.37
Thames2012	2	-	4	0.07	-0.03
Vikram2017	1	-	2	1.15	1.15
Wichmann2014	1	-	0	-1.55	-1.11
Gene Target Antibiotic					
tetracycline	23	11	6	0.98	0.49
beta-lactam	14	6	9	0.20	0.46
MLS	14	6	10	0.31	0.24
sulfonamide	8	3	2	0.06	0.01
fluoroquinolone	4	1	4	--	--
aminoglycoside	4	3	3	-1.50	-1.50
phenicol	3	2	0	1.09	1.49
multiple	6	2	0	0.02	0.07
other	2	1	2	--	--
Farm					
cattle_unspecified	2	1	0	4.03	3.68
dairy	8	4	8	-0.02	0.03
non_dairy	20	8	28	0.45	0.45
Overall	**26**	**12**	**36**	**0.57**	**0.43**

*ND = non-detect reported

Table A-5: *Amendment-Control Response Ratio (RR) summary statistics by meta-data grouping*

	# Sample Pairs	# Studies	# ND*	Mean RR	Median RR
Gene Target Antibiotic					
beta-lactam	45	6	178	0.14	-0.09
MLS	45	5	51	1.37	1.12
sulfonamide	23	5	2	2.41	2.74
tetracycline	44	5	69	0.70	0.23
aminoglycoside	27	2	3	0.59	0.39
fluoroquinolone	2	1	4	--	--
other	21	4	7	1.54	1.35
Control (Non-Manure) Treatment					
mineral	44	6	154	0.89	0.41
no nutrient amendment	44	5	160	1.02	0.51
Study					
Dungan2018	4		20	1.75	1.86
Marti2014	3		11	0.83	0.90
McKinney2018	12		96	2.39	2.46
Miller2019	24		164	0.59	0.27
Nolvak2016	2		12	1.62	1.31
Peng2017	2		9	0.52	0.20
Udikovic-Kolic2014	1		2	0.46	0.09
Farm					
cattle_unspecified	2	1	9	0.52	0.20
dairy	22	5	141	2.11	2.21
non_dairy	24	1	164	0.59	0.27
Overall	**48**	**7**	**314**	**0.95**	**0.46**

*ND = non-detect reported

Appendix B

Supplementary Material for Chapter 3

1. Sample		The mass or volume of sample extracted is known and we aim to determine the concentration of gene targets in this sample. **Known:** $Sample\ mass\ (or\ Volume)$
2. DNA Extraction		During extraction, DNA is concentrated and purified. **The volume in which the DNA is eluted determines the concentration of target genes in the DNA extract.** The following ratio converts a gene copy concentration in an extract to a sample concentration (copies/sample mass): **Known:** $\dfrac{V_{eluted}}{Sample\ mass}$
3. Spike-in		Spike-in DNA must be foreign to the sample and can be genomic DNA extracted from an organism not present in the sample or synthetic DNA. Concentration of genes in spike-in genome or synthetic gene fragments (c_s) and lengths (L_s, basepairs) of spike-in genes are known. **Known:** c_S, L_S
4. DNA Sequencing	GCATCGCATATTTATA ACGCATATCAGATATCAGT CATATCAGTCATATCAGTCAT ACGCATATCAGATATCAGT TTAATCATCAGTCAGT GTCAGTGAGCATATCAGT	After sequencing, the number of sequence reads mapping to spike-in sequences (z_s) are obtained. The spike-in normalization factor, η, can be determined as the average ratio of the spike-in sequence concentration (c_s) to the length-normalized read counts ($\frac{z_s}{L_s}$). **Compute:** $\eta = avg\left(\dfrac{c_s}{z_s/L_s}\right)$
5. Gene Quantification		Next, the number of sequence reads mapping to target gene (z_t) are obtained using a database containing genes of interest. The target gene length (L_t) is the number of basepairs in the target sequence. Our key assumption is that the relationship between gene copy concentration and length-normalized read counts are consistent between the target genes and spike-in genes. We convert z_t/L_t to a sample concentration (copies/sample mass) using the relationships in Box 2 and Box 4: **Compute:** $\dfrac{copies, target}{Sample\ mass} = \eta \cdot \dfrac{z_t}{L_t} \cdot \dfrac{V_{eluted}}{Sample\ mass}$

Figure B-1: *Spike-in experimental and bioinformatic approach*

117

Figure B-2: *Extraction recovery observed for raw manure and land-applied manure slurry.*
TEXT B-1. *Describes details of the approach to evaluate extraction recovery.*
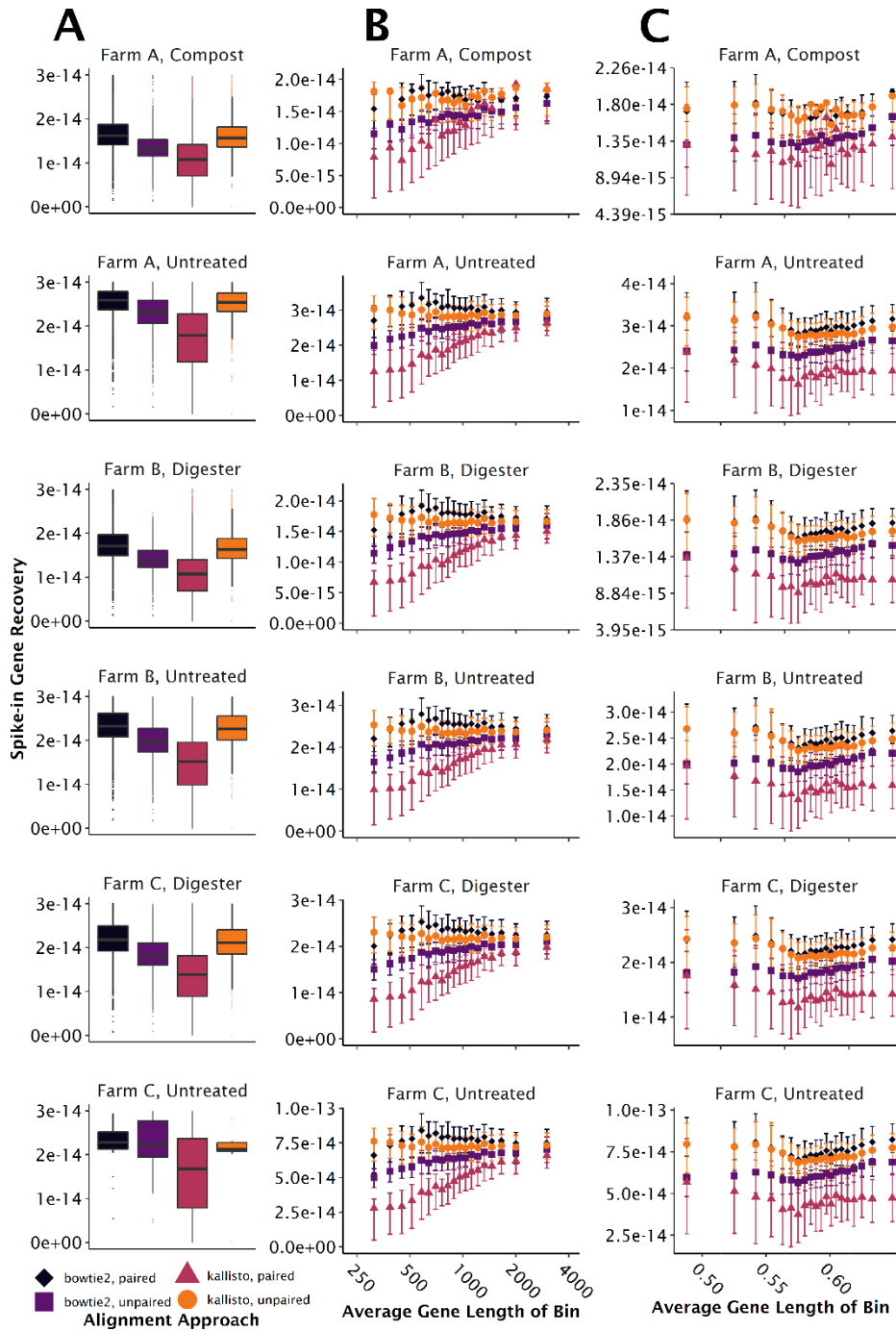
Figure B-3: *Variation in spike-in gene recoveries*

A. *spike-in gene recoveries ($\frac{z_{s,i}}{z_{tot}L_{s,i}}/c_{s,i}$) using four mapping approaches. Units of the y-axis ar*

$1/\frac{gene\ copies}{\mu L*basepairs}$ *where the gene copies/μL are the known copies per DNA extract volume and basepairs are based on the length of the gene.*

B and C. *Spike-in gene recovery of M. hydrocarbonoclasticus genes from the metagenome across (B) gene lengths and (C) %G+C-contents. Genes are binned into 20 quantiles and extended lines represent the interquartile range for each bin.*

119
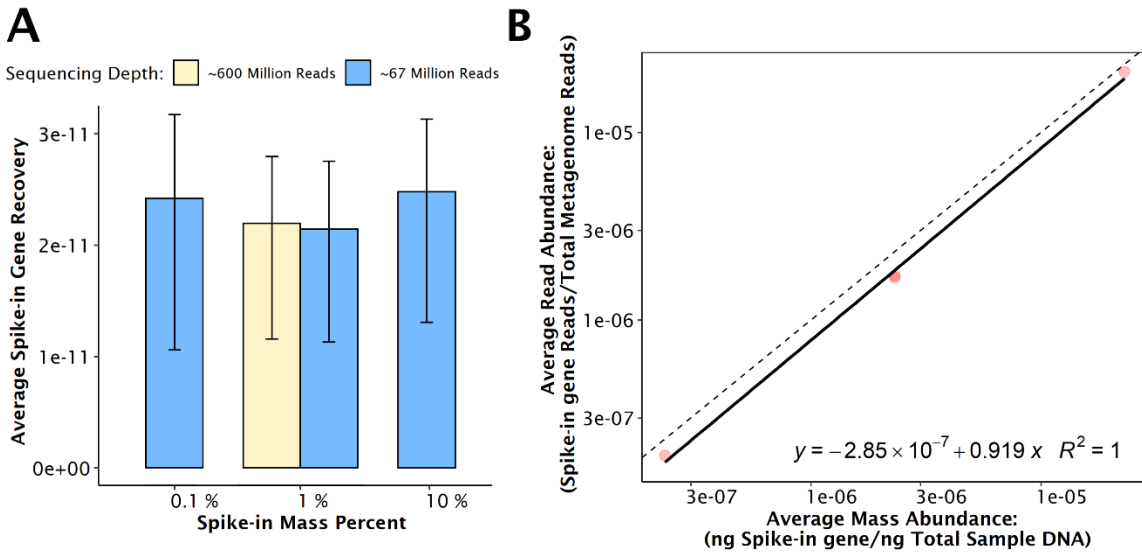
Figure B-4: *Spike-in gene recovery of internal standards across different concentrations*
A. Spike-in gene recovery of internal standards across different concentrations of spike-in.
B. Correlation of relative number of reads mapped to spike-in genes (z_i/z_{tot}) to spike-in gene
mass abundance (mass of Spike-in DNA/Total DNA) for the four same spike-ins in Fig 2A.
__Dotted line__: Theoretical 1:1 relationship, __Solid line__ linear regression. Analysis performed using
the Farm C Digester Sample.
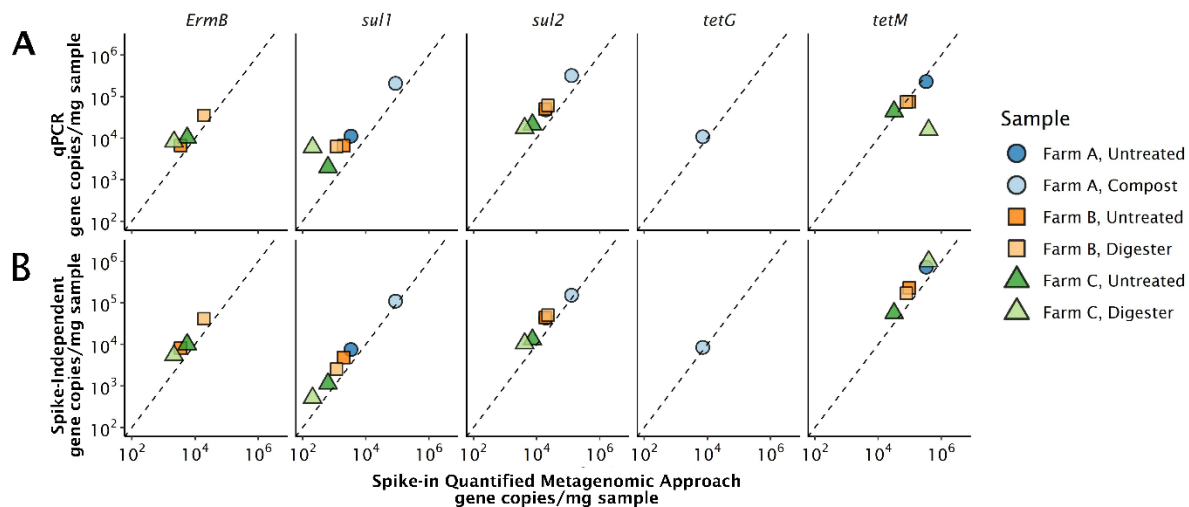
Figure B-5: *Comparisons of the gene quantification approaches using AMR++ tool for assigning reads to ARG reference sequences.*
*The dotted line is a 1:1 line that represents theoretical perfect correlation between approaches. **A.** Spike-in-quantified metagenomic absolute abundance approach vs. qPCR; **B.** Spike-in-quantified metagenomic absolute abundance approach vs. spike-independent, 16S rRNA gene-based metagenomic approach. ErmB was not detected in the Farm A, Samples with AMR++ but was detected in the Farm A Compost sample with qPCR. TetG was detected with qPCR in all samples but not the quantitative metagenome approach.*

SUPPLEMENTAL TEXT B-1: Extraction Recovery Analysis

Spike-in and Extraction

We validated that our extraction methods provided adequate recovery of DNA and minimal extraction bias between gram-positive and gram-negative bacteria in various dairy manure sample types. 10 µL of ZymoBIOMICS Spike-in Control I (ZYMO Research, Irvine, California), which contains a gram-negative bacterium, *Imtechella halotolerans* (LMG 26483), and a gram-positive bacterium, *Allobacillus halotolerans* (LMG 24826), was spiked into 50-250 mg of land-applied manure slurry and manure stockpile samples. DNA was extracted from the Zymo-containing spikes with the QiaAMP PowerFecal kit (QIAGEN, Germantown, MD).

Extraction efficiency – qPCR Protocol

Custom qPCR primers were designed to target a sequence unique to each organism (Table B-1) and percent recovery was calculated as the (Expected Gene Copies – Recovered Gene copies) / Expected Gene Copies * 100. qPCR reactions were carried as described in the main manuscript with the following differences. Extracts were diluted 10-fold and 100-fold prior to qPCR to identify potential inhibition. G-Block fragments (IDT, Skokie, Illinois) containing amplicon sequences (reference in Table B-1) were used for the qPCR standard curve. Assuming one sequence per CFU or Zymo-reported cell numbers, the expected recovery was the number of organisms spiked into the sample. Recoveries of the *Imtechella halotolerans* and *Allobacillus halotolerans* from the land applied and raw manure samples ranged from 75-110% and did not differ significantly between the gram-positive and gram-negative organisms spiked into the samples (p-value = 0.27, Fig. B-2). In both sets of extraction experiments, the spiked organisms were not detected in unspiked samples.

SUPPLEMENTAL TEXT B-2: Read-mapping validation

Spike-in gene recoveries are dependent on read mapping method, gene size, and %G+C content.

We examined the impact of %G+C content and gene length on the recovery of the *M. hydrocarbonoclasticus* genes spiked into DNA extracted from dairy manure samples and sequenced using an Illumina HiSeq4000. This quantitative metagenomic approach uses a known ratio of read abundance to gene concentration to convert read abundances of target sequences to target gene concentrations. Therefore, it is critical to assess factors that can impact the read recovery rate such as gene length and %G+C content. Recovery was defined as the ratio of observed read abundances to the known input concentration of each spike-in gene. Specifically, the ratio consisted of the length-normalized number of reads that mapped to each of the *M. hydrocarbonoclasticus* spike-in genes divided by the total number of metagenomic reads in the sample ($\frac{z_{s,i}}{z_{tot}L_{s,i}}$) over the known concentration of that gene in the sequenced extract ($c_{s,i}$, gene copies/µL of DNA extract). Herein, this recovery ratio ($\frac{z_{s,i}}{z_{tot}L_{s,i}}/c_{s,i}$) is referred to as the "spike-in gene recovery" and reflects recoveries through sequencing and read annotation. A uniform spike-in gene recovery across gene %G+C content and gene length would demonstrate that the sequencing and mapping approach is not biased by these factors.

We mapped reads to the *M. hydrocarbonoclasticus* genes using two mapping tools, kallisto and Bowtie2. Kallisto and Bowtie2 were used since they are common tools for read alignment in metagenomic samples[1–3] and employ two different algorithms for read assignment. Kallisto is a pseudo-alignment tool that uses k-mer hashing to find exact matches between reads and references, and Bowtie2 is a Burrows-Wheeler transform-based alignment. The stringency of read-mapping parameters was relaxed by mapping reads as individual reads (unpaired) to

observe the impact of paired and unpaired read-mapping on recovery of the spike-in genes. The four mapping approaches resulted in significantly different average gene recoveries for all samples (Fig. B-3 A, ANOVA, $F < 2 \times 10^{-16}$ for all samples). Kallisto with paired mapping resulted in the lowest recoveries and is likely too strict of a read assignment approach.

Gene length and %G+C-dependent biases in gene recovery within each read mapping approach were next identified. Recoveries decreased as gene lengths decreased for all the approaches except for kallisto in the unpaired setting (Figure B-3B). The spread of recoveries, measured as the inter-quartile ranges of recoveries across gene bins, decreased with increasing gene size (Figure B-3B). The recoveries leveled-out at gene sizes above 600 basepairs with the Bowtie2-paired mapping approach, but never leveled out for the kallisto-paired and Bowtie2-unpaired mapping approaches. As gene lengths approach the library insert size (here, 450 basepairs), read pairs are less likely to map since a critical fraction of the read pair may extend beyond the target gene reference sequence. Depending on the mapping approach used, this limits read mapping to the spike-in reference gene sequence and thus biases quantification.

Patterns of spike-in gene recovery across %G+C contents were consistent for all mapping approaches (Figure B-3C). Average spike-in gene recoveries decreased as average %G+C content increased. A minimum average spike-in gene recovery was reached around 55% G+C content and then increased again as %G+C contents increased. Large biases caused by %G+C were not expected in this study, since previous studies using Illumina platforms found PCR-free library preparation minimized this bias.[4,5] Minimum and maximum spike-in gene recoveries in the %G+C content bins within mapping tools differed at most by 10% with kallisto in the paired setting in all samples. For the other three tools, the average percent difference in minimum and maximum gene recovery across bins was between 4-5%.

It is important to note that in complex environmental samples, some genes in the *M. hydrocarbonoclasticus* genome may share sequence similarity with genes of microorganisms present in the samples, whereby false-positives could artificially inflate read mapping rates. To assess if outliers with high read mapping drive the recovery ratio, we compared the mean and median spike-in gene recoveries. The mean and median values differed by less than 8% and, therefore, it is unlikely that false-positives have a major impact on the recovery. However, to completely avoid incorrect read mapping to the spike-in genome, synthetic, non-coding DNA can be used as an internal standard spike-in.[4] The kallisto-unpaired read mapping approach was selected for the sequencing efficiency calculation because it consistently demonstrated the lowest spike-in gene recovery bias for the spike-in genes across gene lengths.

Spike-in gene read abundances correlated with spike-in gene mass abundances over three orders of magnitude

To examine the impact of the spike-in gene concentration on measured read abundances (z_i/z_{tot}) and to ensure the spike-in quantitative approach was valid across a range of gene concentrations, we compared the spike-in gene recoveries at spike-in concentrations that spanned three orders of magnitude. Specifically, one sample extract was separately spiked with the *M. hydrocarbonoclasticus* such that the gene concentrations were 8×10^4, 8×10^5, and 8×10^6 copies per µL (which equates to 0.1%, 1%, and 10% of total sample DNA mass as the spike-in genome based on fluorometric quantification (Qubit, Thermo Scientific)). Gene recoveries were reproducible when run at the same spike-in abundance but sequenced at different depths (Figure B-4A). Spike-in gene recoveries were consistent across the different mass spikes (Figure B-4A) and read abundances of internal standard genes were proportional to the gene concentration of spike-in internal standard DNA (linear regression $R^2 = 1$; Figure B-4B).

In addition to confirming that read abundances increase proportionally with the known gene concentrations, these results indicate the general range of gene detection limits. At our sequencing effort in this experiment (50 million reads per sample), no reads mapped to 95 of the 4272 genes in the sample with the lowest spike-in gene concentrations (8×10^4 copies of each gene per µL of DNA extract). Therefore, we can estimate the detection limit of detection range of 10^4 gene copies/µL extract, or about 3.2×10^4 gene copies/mg sample. Despite all genes being present at the same gene copy concentration, the mass concentrations in the extract ranged from 9×10^{-7} to 6×10^{-5} ng DNA/µL, with a mean of 8×10^{-6} ng DNA/µL. The concentrations of the 95 undetected *M. hydrocarbonoclasticus* genes were on the lower end of these ranges, from 9×10^{-7} to 1×10^{-5} ng DNA/µL, with a mean of 3×10^{-6} ng DNA/µL. For comparison, our qPCR limits of detection were as low as 5 gene copies/µL DNA extract and our limits of quantification ranged from 10-60 gene copies/µL DNA extract, depending on the assay (Table B-1). Based on this assessment, qPCR remains the more appropriate method for quantifying genes at low abundances. However, as the cost of sequencing declines, the ability to capture more sequences from a sample will facilitate lower detection limits. To achieve detection limits approaching those of qPCR in this study, we would have needed a 1000-fold higher sequencing depth (equivalent to 50 billion reads).

(1)   Nayfach, S.; Pollard, K. S. Toward Accurate and Quantitative Comparative Metagenomics. *Cell* **2016**, *166* (5), 1103–1116. https://doi.org/10.1016/j.cell.2016.08.007.

(2)   McCall, C.; Xagoraraki, I. Comparative Study of Sequence Aligners for Detecting Antibiotic Resistance in Bacterial Metagenomes. *Lett. Appl. Microbiol.* **2018**, *66* (3), 162–168.

(3)   Schaeffer, L.; Pimentel, H.; Bray, N.; Melsted, P.; Pachter, L. Pseudoalignment for Metagenomic Read Assignment. *Bioinformatics* **2017**, *33* (14), 2082–2088.

(4)   Hardwick, S. A.; Chen, W. Y.; Wong, T.; Kanakamedala, B. S.; Deveson, I. W.; Ongley, S. E.; Santini, N. S.; Marcellin, E.; Smith, M. A.; Nielsen, L. K.; et al. Synthetic Microbe

Communities Provide Internal Reference Standards for Metagenome Sequencing and Analysis. *Nat. Commun.* **2018**, *9* (1), 1–10. https://doi.org/10.1038/s41467-018-05555-0.

(5)   Browne, P. D.; Nielsen, T. K.; Kot, W.; Aggerholm, A.; Gilbert, M. T. P.; Puetz, L.; Rasmussen, M.; Zervas, A.; Hansen, L. H. GC Bias Affects Genomic and Metagenomic Reconstructions, Underrepresenting GC-Poor Organisms. *Gigascience* **2020**, *9* (2), giaa008.

(6)   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.* **2019**, *48* (October 2019), 517–525. https://doi.org/10.1093/nar/gkz935.

Table B-1: *qPCR primers used in this study with annealing temperatures and results of LOD, LOQ, and efficiency assessments.*

Target	Primer Set (5`-3`)	Amplicon Accession #	Ref	LOD copies/µL	LOQ	LOD[a] copies/mg	AT[b]	η[c]	r²
sul1	CGCACCGGAAACATCGCTGCAC GAAGTTCCGCCGCAAGGCTCG	gb|JF969163|+|1054-1893|sul	1	20	60	8	62	98%	0.997
sul2	TCCGGTGGAGGCCGGTATCTG CGGGAATGCCATCTGCCTTGAG	NG_048113.1	1	10	20	4	58	88%	0.999
TetG	GCAGAGCAGGTCGCTGG CCYGCAAGAGAAGCCAGAAG	gb|AF133139|+|1-1176|tet(G)	2	5	20	2	53	92%	0.999
TetM	CCGTTGGGAAGTGGAATGC TCCGAAAATCTGCTGGGGTA	NC_004116.1:c929374-927455	3	5	10	2	59	88%	0.999
ErmB	AAAACTTACCCGCCATACCA TTTGGCGTGTTTCATTGCTT	gb|AF242872|+|2132-2878|ErmB	4	5	20	4	50	91%	0.999
16S rRNA	ACTCCTACGGGAGGCAG ATTACCGCGGCTGCTGG	CP026677.1:1353823-1354031	5	-	-	-	54	88%	1.000
Imtechella halotolerans	TTACCCGCCGACAGATTAGC TTGCGTCCTATGGGGCTTTCT	LMG 26483	TS[d]	-	-	-	60	82%	0.997
Allobacillus halotolerans	TCGCTCCAAACCAGTCCATC ACACCAGGGTAAGTGACTGC	LMG 24826	TS	-	-	-	60	79%	0.998

*a. LOD in sample mass = gene copies/µL * 100 µL DNA extract / Sample mass (0.250 g), b. Annealing Temperature, c. Efficiency from standard curve, d. TS = this study*

Primer References:

1. Pei, R.; Kim, S.-C.; Carlson, K. H.; Pruden, A. Effect of River Landscape on the Sediment Concentrations of Antibiotics and Corresponding Antibiotic Resistance Genes (ARG). *Water Res*. 2006, 40 (12), 2427–2435.

2. Aminov, R.; Chee-Sanford, J.; Garrigues, N.; Teferedegne, B.; Krapac, I.; White, B.; Mackie, R. I. Development, Validation, and Application of PCR Primers for Detection of Tetracycline Efflux Genes of Gram-Negative Bacteria. *Appl. Environ. Microbiol*. 2002, 68 (4), 1786–1793. https://doi.org/10.1128/AEM.68.4.1786.

3. Luo, Y.; Mao, D.; Rysz, M.; Zhou, Q.; Zhang, H.; Xu, L.; Alvarez, P. J. J. Trends in Antibiotic Resistance Genes Occurrence in the Haihe River, China. *Environ. Sci. Technol*. 2010, 44 (19), 7220–7225. https://doi.org/10.1021/es100233w.

4. Knapp, C. W.; Dolfing, J.; Ehlert, P. A. I.; Graham, D. W. Evidence of Increasing Antibiotic Resistance Gene Abundances in Archived Soils since 1940. *Environ. Sci. Technol*. 2010, 44 (2), 580–587. https://doi.org/10.1021/es901221x.

5. Fierer, N.; Jackson, J. A.; Vilgalys, R.; Jackson, R. B. Assessment of Soil Microbial Community Structure by Use of Taxon-Specific Quantitative PCR Assays. *Appl. Environ. Microbiol*. 2005, 71 (7), 4117–4120. https://doi.org/10.1128/AEM.71.7.4117-4120.2005.

Table B-2: *Allele frequencies of single nucleotide variants (SNV) compared to the tetM forward primer sequence aligned using Bowtie2. The base in the forward primer and position is provided as the column headers. Depth refers to the average read abundance.*

Sample	Primer Position, Base (5'-3')						
	3, G	6, G	9, A	12, T	16, A	18, G	Depth
Farm A, Untreated	79%, A*	79% A	77% G	77% C	77% C	76% A	~2600
Farm A, Compost	71% A	71% A	29% G	71% C	71% G	71% A	7
Farm B, Untreated	79% A	78% A	21% G	78% C	78% G	77% A	~1150
Farm B, Digester	73% A	73% A	27% G	72% C	71% G	71% A	~450
Farm C, Untreated	63% A	62% A	38% G	60% C	60% G	56% A	~650
Farm C, Digester	99% A	99% A	01% G	99% C	99% G	99% A	~4000

*79% of reads that mapped to the former primer position 3 contained an "A" rather than a "G"

Appendix C

Supplementary Material for Chapter 4

Table C-1: *Stored manure sample data and nutrient concentrations*

Sample	Month	Year	Total N[a]	NH4-N[b]	Total P[c]	% Total Solids	[d]Latitude	[d]Longitude	Herd size
Storage only Samples									
MW1	Sept	2019	1.671	1.007	0.357	2.3	42.27805	-83.723844	150
MW2	April	2019	4.136	2.144	0.812	15.4	42.27805	-83.723844	600
MW3	April	2019	0.539	0.376	0.052	0.7	42.27805	-83.723844	400
MW4 - Storage 1	May	2019	0.995	0.780	0.129	0.8	42.27805	-83.723844	330
MW4 - Storage 2	May	2019	0.935	0.497	0.130	1.0	42.27805	-83.723844	330
MW5	April	2019	4.208	1.963	0.884	14.1	42.27805	-83.723844	680
MA1	Oct	2017	0.110	0.038	0.026	0.2	38.98619	-76.942587	80
NE1	Oct	2016	2.160	0.944	0.288	4.2	42.44206	-76.502839	1350
Anaerobic Digestion + Storage Samples									
NE2 - Fall Storage	Oct	2016	3.055	0.953	0.561	6.8	42.44206	-76.502839	4300
NE2 - Spring Storage	May	2017	NA	NA	NA	4.2	42.44206	-76.502839	4300
NE3 - Fall Storage	Oct	2016	2.678	1.156	0.658	6.7	42.44206	-76.502839	2050
NE3 - Spring storage	May	2017	NA	NA	NA	3.1	42.44206	-76.502839	2050
NE4	May	2017	2.614	1.055	0.427	1.3	42.44206	-76.502839	1820
MA2	Sept	2016	2.100	0.906	0.295	2.8	38.98619	-76.942587	2700
MA3	May	2017	2.625	1.720	0.481	4.6	38.98619	-76.942587	690

Nutrient data reported as kg/L. MW = Midwest, NE = Northeast, MA = Mid-Atlantic, NA = Not available.

[a]Total Nitrogen as N. [b]Ammonia-Nitrogen. [c]Total Phosphorus as P. [d]Coordinates used are centroid of sampled region

Table C-2: *Metagenomic read abundances*

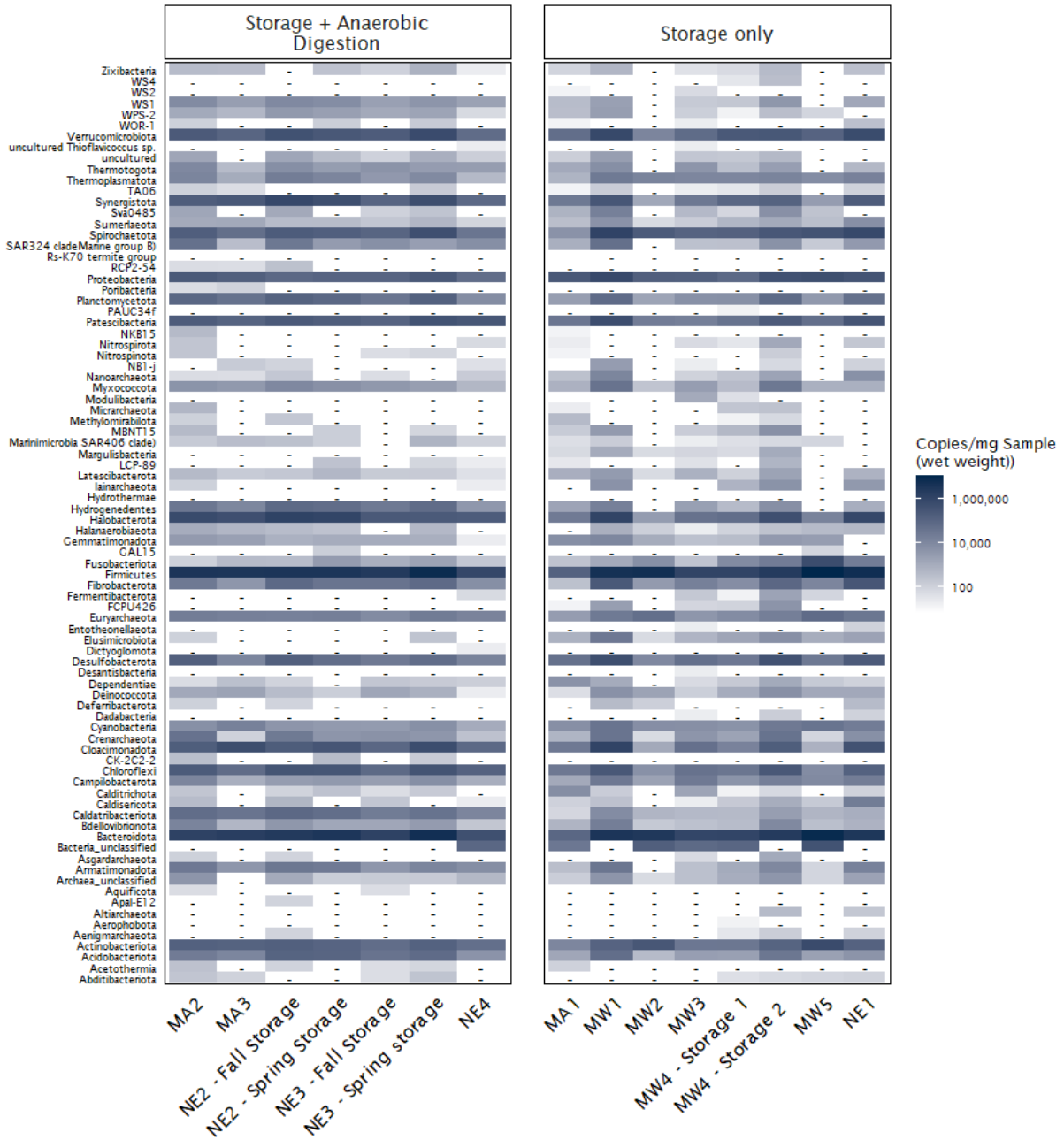| Sample	16S rRNA gene reads phyloFlash	ARG reads MEGARes	ARG reads GROOT	Total post-QC read
MW1	2.38×10^5	8.88×10^5	9.80×10^5	3.17×10^8
MW2	3.34×10^5	1.60×10^6	3.44×10^6	2.89×10^8
MW3	2.92×10^5	1.06×10^6	8.01×10^5	3.31×10^8
MW4 - Storage 1	4.17×10^5	1.93×10^6	3.15×10^6	4.86×10^8
MW4 - Storage 2	2.41×10^5	1.13×10^6	1.64×10^6	3.62×10^8
MW5	5.54×10^5	2.10×10^6	1.71×10^6	4.70×10^8
MA1	2.49×10^5	1.63×10^6	1.30×10^6	4.81×10^8
NE1	2.85×10^5	1.10×10^6	1.39×10^6	4.00×10^8
NE2 - Fall storage	2.26×10^5	7.56×10^5	1.45×10^6	3.38×10^8
NE2 - Spring storage	1.64×10^5	5.11×10^5	9.42×10^5	2.64×10^8
NE3 - Fall storage	2.07×10^5	1.27×10^6	4.48×10^6	2.74×10^8
NE3 - Spring storage	4.43×10^5	1.41×10^6	1.48×10^6	4.66×10^8
NE4	2.49×10^5	2.04×10^6	8.26×10^6	3.67×10^8
MA2	2.93×10^5	1.16×10^6	1.46×10^6	4.54×10^8
MA3	2.97×10^5	1.08×10^6	1.74×10^6	4.44×10^8

Figure C-1*: Organism concentrations summarized at the Phylum level. "-" indicates no reads mapped to references sequences in a phylum.*

Figure C-2: *Relative abundances of classes in Firmicutes and Bacteroidetes phyla Classes in Firmicutes (top) and classes in Bacteroidetes (bottom)*
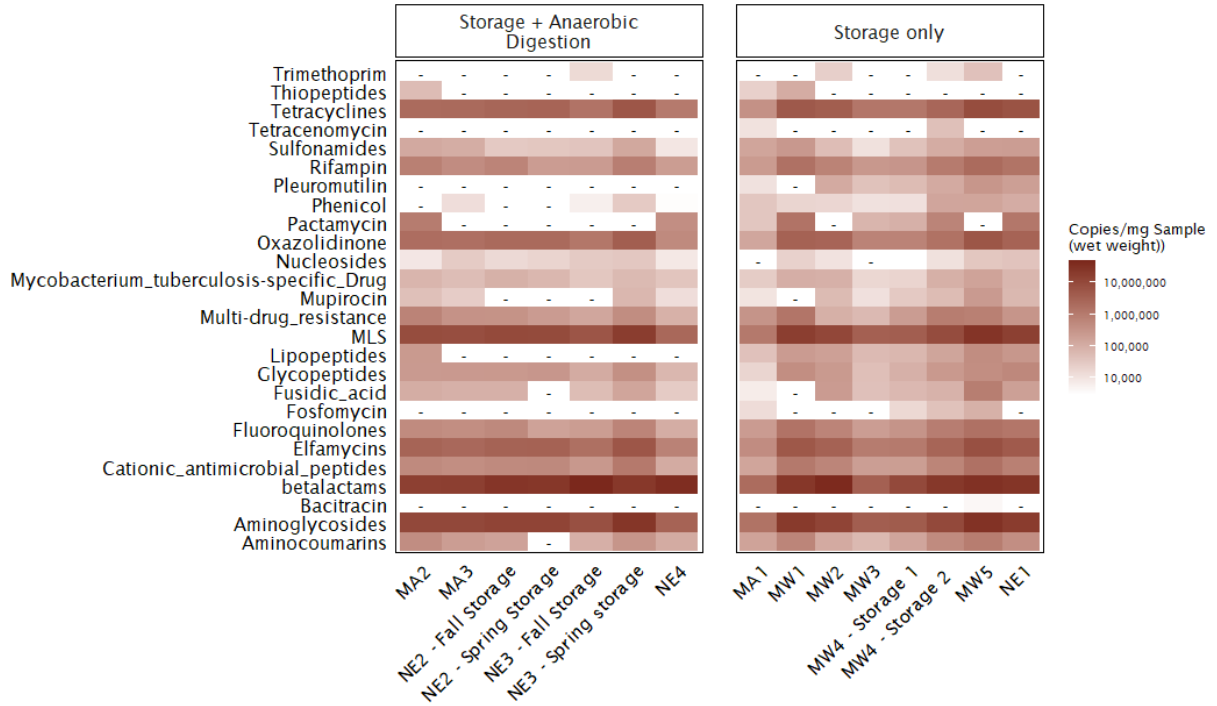
Figure C-3: *Concentrations of AMR++ ARGs grouped by Drug class.*
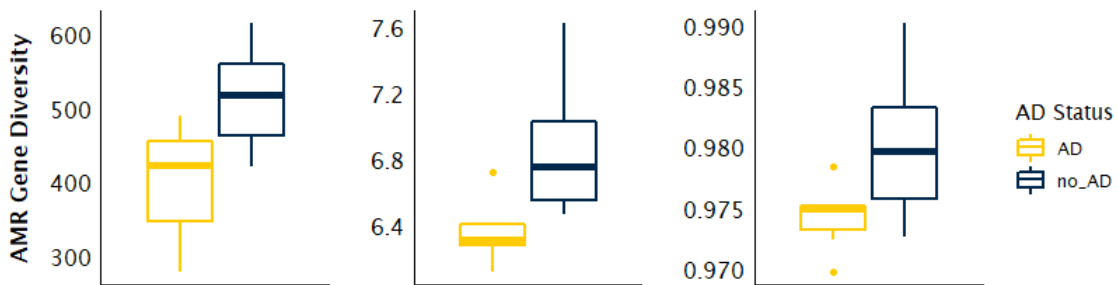*"-" indicates no reads mapped to references sequences in a phylum.*

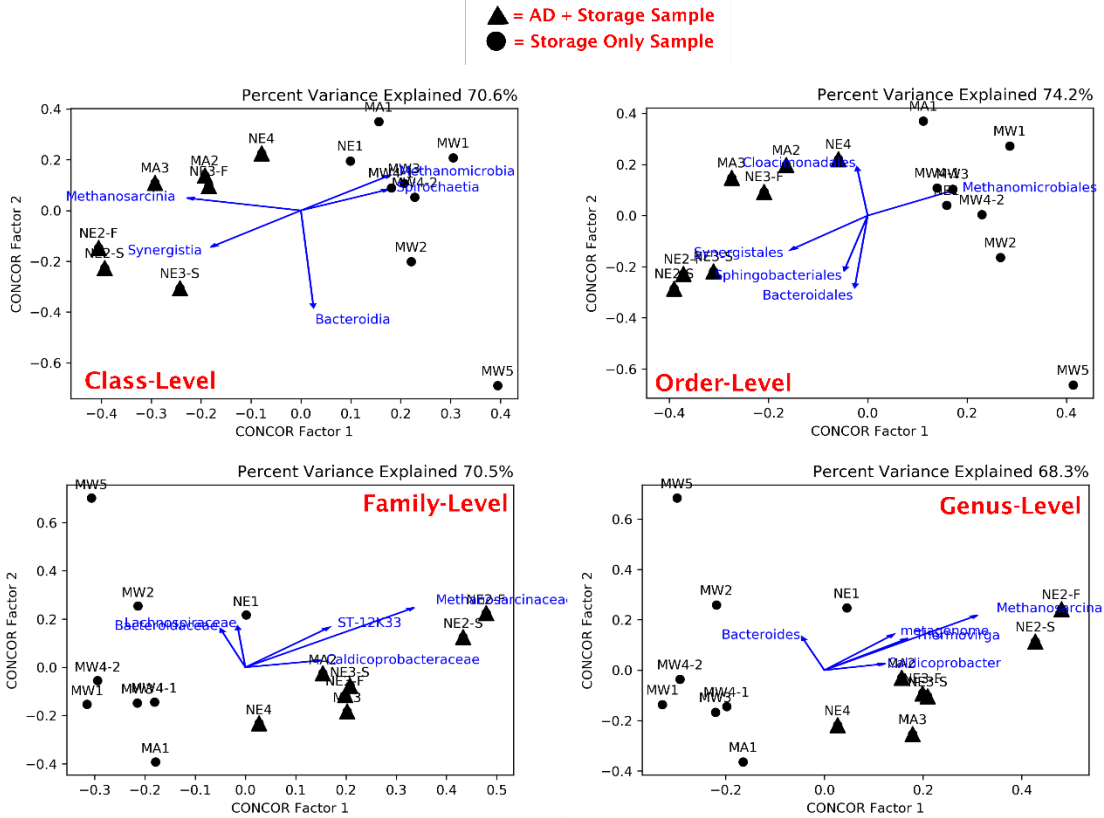***Figure C-4:****Diversity in microbial community and ARGs between manure storages with AD and storage only*

Figure C-5: *Differential signatures of AD and storage only visible at different taxonomic levels*
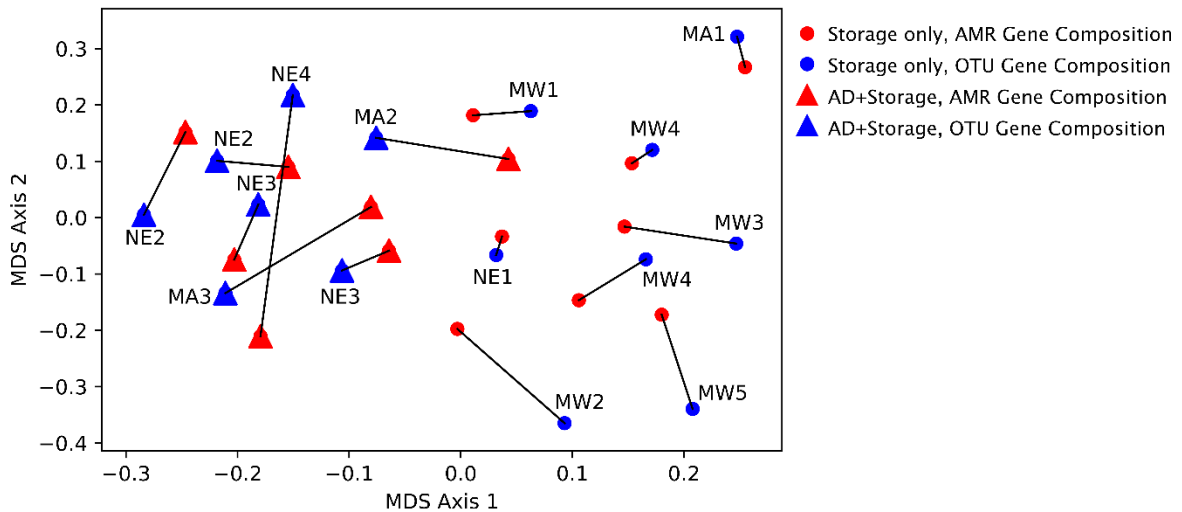

Figure C-6: *Procrustes analysis on Bray-Curtis distances demonstrating relationship between sample ARG and OTU composition similarity across farm cohort*
$M^2 = 0.35.$

Supplemental Text C-1: DNA Extraction Efficiency

Extraction recoveries were then assessed by quantifying recovery of gene targets from the

Imtechella halotolerans and *Allobacillus halotolerans* genomes. Primers were designed using

PrimerBlast and checked for specificity against the virus, eukaryote, and bacteria nonredundant

nucleotide and RefSeq databases. The concentration of cells (and gene targets, assuming one

target per organism) in the DNA extract was computed as follows:

$$Extract\ Gene\ Copies\ Concentration = \frac{V_{Zymo\ spike} * 10^7 \frac{Cells}{uL}}{V_{eluted}}$$

Where the $V_{Zymo\ Spike}$ is the amount of ZymoBIOMICS Spike-in Control I spiked into the sample,

10^7 is the concentration of each cell according to the manufacturer, and V_{eluted} is the volume in

which the DNA was eluted using elution buffer or molecular grade water. qPCR reactions were

then carried out on an Eppendorf MasterCyler ep realplex[2] using Fast EvaGreen Fast Master Mix

(Biotium, Fremont, CA). Details of the qPCR assays, including primer sequences and annealing

temperatures are provided in Table C-3. 20 µL reactions were performed following the

manufacturer's recommendations, with 0.4 µM of forward and reverse primers, 0.625 mg/mL of

Ultrapure BSA (Invitrogen), and 2 µL of diluted DNA extracts. DNA extracts were diluted both

10- and 100-fold. Plates were centrifuged for 2 minutes at 500 RPM at 4°C before

thermocycling. qPCR reactions were performed in triplicate. After amplification, melt curves

were performed to confirm the specificity of the reactions. qPCR was also performed on pooled,

unspiked extracts to ensure no background signal was present. The template for the standard

curve consisted of Gblock Fragments (IDT, Skokie, Illinois; Table C-3).

Next, the extraction efficiency was computed as the following:

$$Extraction\ efficiency = \frac{Measured\ Gene\ Copy\ Concentration}{Extract\ Gene\ Copies\ Concentration}$$

Where the numerator is the concentration measured by qPCR.

Table C-3: *qPCR primers and standard curve used in this study*

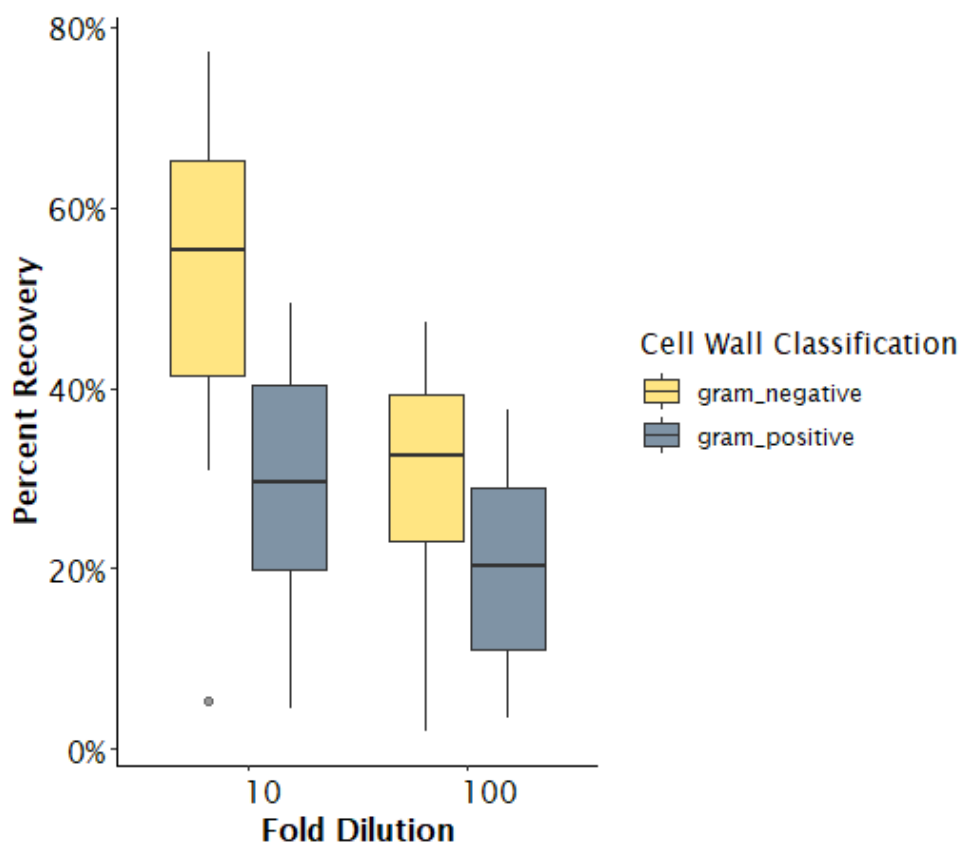| Target	Primer Set (5`-3`)	Amplicon Accession #	Annealing Temp. ºC	η[a]	r²
Imtechella halotolerans	TTACCCGCCGACAGATTAGC TTGCGTCCTATGGGGCTTTCT	LMG 26483	60	86%	0.996
Allobacillus halotolerans	TCGCTCCAAACCAGTCCATC ACACCAGGGTAAGTGACTGC	LMG 24826	60	95%	0.990

a. Efficiency from standard curve

Figure C-7: *DNA extraction efficiency from stored manure extraction*