Metagenomics-Based Environmental Monitoring of Antibiotic Resistance: Towards Standardization

Benjamin Cole Davis

Dissertation submitted to the faculty of the Virginia Tech Polytechnic Institute and State University in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

In

Civil Engineering

Amy Pruden, Chair

Marc A. Edwards

Emily D. Garner

Ann Stevens

Peter J. Vikesland

May 10th, 2022

Blacksburg, VA

Keywords: antibiotic resistance, surface water, wastewater, metagenomics, next-generation sequencing, enterococcus, standardization, quantitative metagenomics

Metagenomics-Based Environmental Monitoring of Antibiotic Resistance: Towards Standardization

Benjamin Cole Davis

ABSTRACT

Antibiotic resistance (AR) is a critical and looming threat to human health that requires action across the One Health continuum (humans, animals, environment). Coordinated surveillance within the environmental sector is largely underdeveloped in current National Action Plans to combat the spread of AR, and a lack of effective study approaches and standard analytical methods have led to a dearth of impactful environmental monitoring data on the prevalence and risk of antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARGs) in aquatic environments. In this dissertation, integrated surveillance approaches of surface water and wastewater systems are demonstrated, and efforts are made towards standardizing both metagenomic- and culture-based techniques for globally comparable environmental monitoring.

A field study of differentially-impacted watersheds on the island of Puerto Rico post-Hurricane Maria demonstrated the effectiveness of metagenomics in defining direct impact of anthropogenic stress and human fecal contamination on the proliferation of ARGs in riverine systems. The contribution of treated wastewater effluents to the dissemination of highly mobile and clinically-relevant ARGs and their connection to local clinical settings was also revealed. At the international scale, a transect of conventional activated sludge wastewater treatment plants (WWTPs), representing both US/European and Asian regions, were found to significantly attenuate ARG abundance through the removal of total bacterial load and human fecal indicators, regardless of influent ARG compositions. Strong structural symmetry between microbiome and ARG compositions through successional treatment stages suggested that horizontal gene transfer plays a relatively minor role in actively shaping resistomes during treatment. Risk assessment models, however, indicated high-priority plasmid-borne ARGs in final treated effluents discharged around the world, indicating potentially increased transmission risks in downstream environments.

Advancements were also made toward standardizing methods for the generation of globally representative and comparable metagenomic- and culture-based AR monitoring data via two comprehensive and critical literature reviews. The first review provides guidance in next-generation sequencing (NGS) studies of environmental AR, proposing a framework for experimental controls, adequate sequencing depths, appropriate use of public databases, and the derivation of datatypes that are conducive for risk assessment. The second review focuses on antibiotic-resistant *Enterococcus* spp. as robust monitoring targets and an attractive alternative to more widely adopted Gram-negative organisms, while proposing workflows that generate universally equivalent datatypes.

Finally, quantitative metagenomic (qMeta) techniques were benchmarked using internal reference standards for high-throughput quantification of ARGs with statistical reproducibility.

Metagenomics-Based Environmental Monitoring of Antibiotic Resistance: Towards Standardization

Benjamin Cole Davis

GENERAL AUDIENCE ABSTRACT

Antimicrobials have contributed to the reduction of infectious diseases in humans and animals since the early 20th century, increasing productivity and saving countless lives. However, their industrial-scale application across human, animal, and agricultural sectors over the last several decades, especially the use of antibiotics, have engendered the proliferation of antibiotic resistance (AR). AR occurs when changes in bacteria cause the drugs used to treat infections to become less effective and has become one of the leading public health threats of the 21st century. The global spread of AR through the transmission and evolution of antibiotic resistant bacteria (ARB; known colloquially as "superbugs") and antibiotic resistance genes (ARGs) across the One Health continuum (i.e., humans, animals, and the environment) is resulting in increased hospitalization, length of hospital stays, suffering, death, and overall health-care associated costs globally. This dissertation demonstrates the use of metagenomics, the sequencing of all genetic material (e.g., DNA) recovered from a microbial community, for the comprehensive monitoring of ARB and ARGs in aquatic environments, a key pathway for the dissemination of AR into and out of human populations.

In order to impede the proliferation of AR, surveillance systems are currently in place to track the spread and evolution of ARB and ARGs in humans and livestock, as well as agri-food sectors. However, the surveillance in natural and built environments (i.e., rivers and domestic sewage) has significantly lagged due to the lack of standard monitoring targets and methodologies. It is also a goal of this dissertation to suggest guidance for the collection of metagenomic- and culture-based AR monitoring data to generate universally comparable results that can be included in centralized databases.

Riverine systems are ideal models for tracking input of antibiotic resistance to the natural environment by human activity. After Hurricane-Maria, many of Puerto Rico's wastewater treatment plants (WWTPs) went offline, discharging raw sewage to local surface waters. In a cross-sectional study of watersheds impacted by WWTPs, the abundance of ARGs was directly correlated to increases in local population density. Also, highly mobile and clinically-relevant ARGs were found directly downstream of WWTPs across the island. We found that many of these ARGs corresponded well to forms AR endemic to the region.

WWTPs are the primary engineering controls put in place to curb the spread of human and animal waste streams and can help to reduce AR. An international transect of conventional activated sludge WWTPs representing US/Europe and Asia were sampled to garner a mechanistic understanding of the fate or ARGs through treatment. Although WWTPs remove total bacteria, human fecal indicators, and much of the abundance of ARGs, mobile and clinically-relevant ARGs are discharged around the world in large quantities. Consideration is needed in certain regions of

the world where the managing of human waste streams is the first line of defense against the dissemination of resistance to local communities.

Two comprehensive critical literature reviews were conducted to evaluate the various methodologies for generating and analyzing metagenomic- and culture-based AR monitoring data. These reviews address the need for experimental rigor and disclosure of extensive metadata for inclusion in future, centralized databases. The articles further provide guidance with respect to universally comparable datatypes and efficient workflows that will aid in the scale-up of the collection of environmental monitoring data within a global surveillance framework.

Finally, a study was conducted to benchmark the use of internal DNA reference standards for the absolute quantification of ARGs (i.e., on a ARG copy per volume of sample basis). The statistical framework for ARG detection and its implications for wastewater-based surveillance systems of AR are also discussed.

ACKNOWLEDGEMENTS

I would like to first and foremost express my deepest and most sincere gratitude to my advisor, Dr. Amy Pruden, for providing endless opportunities for personal and professional growth and showing infinite patience and understanding in my pursuit of self-improvement. I can say with certainty that I would not have made it this far without your faith in my abilities, however warranted it may have been. I cannot express in any number of words how thankful I am to have you as an advisor and mentor and friend.

To my committee, I would like to thank Drs. Marc Edwards and Peter Vikesland for your abundant and freely shared wisdom and foresight. Even from passing conversations, many obstacles and pitfalls were avoided, and several novel ideas were generated thanks to you both. To Dr. Ann Stevens, I thank you for my foundational education in microbiology and treating me as a peer from day one. And to Dr. Emily Garner for teaching me the fundamentals of molecular biology from the benchtop to the laptop, including my first introduction to metagenomics and its limitless complexity.

I would like to acknowledge all the financial and professional support that made this dissertation possible. The National Science Foundation; the Water Research Foundation; the Virginia Tech Institute for Critical Technology and Applied Sciences; the Virginia Tech Center for Emerging, Zoonotic, and Arthropod-borne Pathogens; and the Interdisciplinary Graduate Education Program in Infectious Disease. Thank you as well to Advanced Research Computing at Virginia Tech for their technical support and near unlimited access to computational resources that made this research possible.

To the Pruden, Edwards, and Vikesland research group members, past and present, I sincerely thank you. Being around highly motivated, diligent, and fiercely intelligent researchers provided me the roadmap and motivation for becoming the scientist and person I am today. Specifically, I would like to sincerely thank Drs. Ishi Keenum, Suraj Gupta, and (soon to be Dr.) Connor Brown, for wonderfully endless conversations about research. I would not have come up with the ideas and executed the analyses presented in this dissertation without them. I also want to thank the lab groups of Drs. Liqing Zhang and Lenwood Heath whose bioinformatic software made much this research possible.

Lastly, I would like to thank my friends and family. I would have burned out and given up a long time ago if weren't for the very large group of friends and colleagues that I had the pleasure of sharing beers with at tailgates, bar crawls, and trivia night. You all made Blacksburg home from the very beginning. I would like to also thank Amanda Darling for being my absolute rock during the writing of this dissertation and keeping me sane with wonderfully competitive games of Scrabble. And finally, thank you to my family; Mamasan, Papasan, and Andrew for believing that I could even achieve such a feat as a PhD. I'm just as surprised as you are!

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

OVERVIEW AND RESEARCH MOTIVATIONS

The magnitude of the societal and economic costs of antibiotic (AR) and its effect on human wellbeing are still being realized, but current estimates from the European Union (EU) place the burden of resistant bacterial infections to be comparable to influenza, tuberculosis, and HIV/AIDS combined.[1] Antibiotic-resistant bacteria (ARB) and the genetic information which encode for their resistance mechanisms, antibiotic resistance genes (ARGs), are spreading globally among humans, animals, food production systems, and the environment, making AR a true "One Health" problem.[2] According to the World Health Organization (WHO), the increase in AR in human and animal pathogens is among the top ten threats to global health,[3] and some stark estimates predict resistant bacterial infections will become the leading cause of death worldwide by 2050.[4] Under a high-AR impact scenario with no mitigation efforts, an additional 8 million people may fall into extreme poverty, seeing a global reduction in GDP comparable to the 2008-2009 financial crises by as soon as 2030.[5]

In 2015, the WHO put forth a Global Action Plan to help combat the development of AR through "strengthen[ing] the knowledge and evidence base through surveillance and research" in humans, animals, and the environment.[6] In the environmental sector, a growing body of research has demonstrated that water bodies are key recipients, pathways, and sources of AR via anthropic activities such as aquaculture, livestock and agricultural runoff, pharmaceutical manufacturing wastewater, and especially the treatment, or lack thereof, of human waste streams.[7] Globally, it is estimated that nearly 80% of all human excreta is directly discharged into the environment, primarily in low-income and developing countries.[8] In fact, the global burden of disease of AR infection, transmission, and colonization is more attributable to a lack of clean drinking water and poor sanitation infrastructure than regional antibiotic consumption rates.[9,10] Environmental surveillance systems that are aimed at identifying key transmission and evolutionary routes between humans and water environments is therefore critical for understanding and mitigating AR, a point further emphasized in recent EU and US National Action Plans.[11,12] Correspondingly, surface water and wastewater monitoring programs aimed at tracking AR impacts on human and animal populations are currently being developed by the US Environmental Protection Agency[13] and Centers for Disease Control.[14] Much of the focus of these programs have focused on the urban wastewater treatment plant (WWTP) and their influents (i.e., raw sewage) and effluents (i.e., treated sewage being discharged to surface waters), and work is still needed in understanding AR dynamics through conventional biological treatment and the effect of treated sewage on receiving water bodies.[15,16]

Surveillance of surface water and wastewater indicators of AR, through the use of conventional quantitative polymerase chain reaction (qPCR) or culture methods, can provide vital information on the relative prevalence of ARGs and ARB in a sample, establish baselines for prioritizing mitigation efforts, and directly assess attenuation efficiencies put in place by engineering controls (e.g., WWTPs).[17,18] However, AR as an environmental contaminant and phenomenon is a complex and integrated microbiomics problem that warrants sophisticated methodologies for holistic

evaluations of environmental health and AR dynamics. As a result, next-generation nucleic acid (i.e., DNA and RNA) sequencing (NGS) approaches to evaluate water and wastewater quality are rapidly evolving in the fields of environmental science and engineering.[19] Environmental metagenomics, or the use of NGS to study the collection of genetic material recovered directly from an environmental microbial community, has emerged as a powerful tool for the study of environmental AR.[20] Global wastewater surveillance systems have already been proposed[21] and pilot tested using the technique.[22] In order to realize the full potential of metagenomics for integrated surface water and wastewater surveillance programs, extensive research is needed for ensuring the representativeness of sample processing and handling of *in situ* resistomes (i.e., total ARGs carried across a microbial community characteristic of a given environment), universal comparability of metagenomic libraries, and data analytics for accurate and reproducible monitoring data.[23] Ultimately, the field must move towards collaborative insight into the environmental circumstances, human exposures, targeted mitigation strategies, and dose-response considerations for human health risk assessments.[24]

Antibiotic Resistance in the Environment

Samples from 30,000 year old permafrost have revealed that AR is an ancient and naturally occurring phenomenon that well predates the modern antibiotic era.[25] It was not until recently; however, that the industrial-scale application of antimicrobials globally accelerated the evolution of resistance mechanisms via unprecedented and sustained selection pressures, particularly on the enteric bacteriomes of humans and animals.[26,27] These selection pressures have induced the mobilization of a large diversity of ARGs across entire microbiomes, particularly pathogenic taxa that are commensal to gastrointestinal tracts.[28] It is currently understood that most ARGs and virulence factors that occur in modern bacterial pathogens were acquired by horizontal gene transfer (HGT) from commensal or environmental microorganisms.[29,30] In contrast to typical chemical contaminants, once excreted, exogenous ARB and ARGs can persist and even spread in the environment; in other words, they are anthropogenic, self-replicating contaminants, making them distinctly difficult to track and adequately characterize amongst the complexity of environmental microbiomes and the physicochemical dynamics of natural and engineered systems.[31,32] The increased dissemination of AR in the environment is likely due to three principal mechanisms, which occur in tandem: HGT of ARGs within and between phylogenetic groups; genetic mutation and recombination; and the maintenance and enrichment of ARGs and ARB owing to widespread and sustained selection pressures via antimicrobials and other contaminants (e.g., heavy metals and biocides).[33] It has been demonstrated that even sub-minimum inhibitory concentrations (MICs) of antibiotics have the ability to select for ARB and maintain resistance in filter-sterilized wastewater[34,35], warranting new classifications appropriate for environmental settings, such as proposed no effect concentrations (PNECs) of antibiotics.[36]

Pollution is a direct driver of AR in the environment, primarily via human and animal excreta, which can contain high concentrations of ARB, ARGs, and antimicrobials. Although environmental concentrations of antibiotics are often well below MIC and PNEC levels, untreated municipal and hospital wastewater, as well as surface waters impacted by pharmaceutical manufacturing wastes, routinely contain antibiotics at selective concentrations.[37] Metals and biocides can also co-select for AR via co-resistance (i.e., genetic linkage through neighboring genes) and cross-resistance (i.e., the same gene with multiple functions).[38,39] In the environment,

antibiotics and ARG abundance regularly correlate. However, these observations are often accompanied with parallel levels of fecal pollution, which is a source of both, confounding the ability to discern the contributions of *in situ* selection pressure to observed ARG occurrence patterns.[40] It is also often difficult to distinguish whether changes in ARB and ARG abundance are attributed to direct selective pressure imposed by antimicrobials in the environmental matrix, or merely the result of shifts in taxonomic structure due to ambient physicochemical changes.[41] The relative rates of induced HGT versus direct selection or co-selection by naturally-occurring or anthropogenic stressors are also unknown, making HGT events exceedingly difficult to quantify and assign causal effects *in situ*.[42] This is problematic because there is great concern that novel resistance mechanisms to critically important antibiotics are being recruited into pathogens from the vast and diverse reservoir of environmental microbiomes that is being induced by anthropogenic pollution.[28]

In addition to evolutionary concerns of new resistant strains, the environment provides a route for already-resistant bacteria to colonize or infect humans. There is a great deal of literature on the transmission of resistant bacteria that occurs in food production environments (e.g., *Salmonella enterica*) and especially within the hospital "environment" (e.g., methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus* spp.).[43] Exposures to surface waters polluted with fecal contamination can also lead to various antibiotic-resistant infections.[44] It has been demonstrated that recreational swimmers utilizing surface water impacted by treated wastewater discharge were more likely to be colonized by cephalosporin-resistant *E. coli* than control groups.[45,46] Transmission and colonization events are further highlighted in areas with poor sanitation and hygiene infrastructure[47], but also in regions of developed countries with decentralized (e.g., on-site septic systems) or aging wastewater treatment systems. Drinking water, stormwater, and wastewater infrastructure in the US were recently assigned grades "C-", "D", and "D+", respectively by the American Society of Civil Engineers, with increasing prevalence of sanitary sewer overflow events that will inevitably be impacted by changing climates.[48] In the US, we are plagued by a tremendously "leaky" and aging system that is reaching its designed lifespan, which will further exacerbate the ability to control AR outputs. Now is a critical moment to begin to consider how our infrastructure can be improved in a manner that reduces the evolution and dissemination of AR. Establishing appropriate, globally comparable, AR monitoring programs is a critical need towards achieving this goal.

RESEARCH OBJECTIVES

The specific objectives of the research described herein were to:

1. Investigate the effect of wastewater effluent discharge and anthropogenic stress on the proliferation of AR in impacted watersheds,
2. Investigate the dynamics of conventional biological wastewater treatment on the attenuation or proliferation of AR across a global transect of WWTPs and assess risk for mobility and dissemination of ARGs to receiving environments,
3. Critically review and assess current culture- and metagenomic-based AR monitoring techniques, and provide guidance towards standard protocols that will produce globally comparable data to support environmental monitoring of AR,
4. Explore quantitative metagenomic techniques with internal reference standards for high-throughput quantification of ARGs in wastewater.

ANNOTATED DISSERTATION OUTLINE

Each chapter in this dissertation is formatted according to the scientific journal to which it was submitted or is intended to be submitted.

Chapter 1: *Introduction*

This chapter introduces the general themes of the research described herein and provides context for the specific research objectives addressed in this dissertation.

Chapter 2: *Demonstrating an integrated antibiotic resistance gene surveillance approach in Puerto Rican watersheds post-hurricane Maria*

Chapter 2 addresses objectives (1) and (2) by investigating the relationship between qPCR indicators of antibiotic resistance, human fecal markers, and resistome dynamics in three differentially-impacted watersheds on the island of Puerto Rico after a natural disaster. The manuscript highlights the applicability of metagenomics for in-depth surveillance of clinically-relevant and mobile ARGs and their relationship to treated wastewater inputs across anthropogenically-impacted watersheds. It further makes direct connections between resistance prevalence throughout the surface water and wastewater samples to historical AR prevalence in local clinical settings.

This manuscript has been published:

> Davis, B. C., Riquelme, M. V., Ramirez-Toro, G., Bandaragoda, C., Garner, E., Rhoads, W. J., ... & Pruden, A. (2020). Demonstrating an integrated antibiotic resistance gene surveillance approach in Puerto Rican watersheds Post-Hurricane Maria. *Environmental Science & Technology*, 54(23), 15108-15119. https://dx.doi.org/10.1021/acs.est.0c05567

Attributions: I conducted surface water and sediment sampling, analysis of samples, analyzed data, and led the writing of the manuscript. Virginia Riquelme collected wastewater treatment plant samples. I, Graciela Ramirez-Toro, William Rhoads, Virginia Riquelme, and Amy Pruden designed the sampling scheme. Christina Bandaragoda conducted ArcGIS land use analysis of watersheds. Emily Garner provided guidance on the qPCR and metagenomic analysis of the samples and assisted in manuscript preparation. Manuscript revisions and presentation was primarily done by Amy Pruden.

Chapter 3: *Evaluation of resistome risk reduction through secondary wastewater treatment across an international transect*

Chapter 3 addresses Objective (2) more thoroughly via the systematic assessment of an international transect of wastewater samples that highlight resistome behavior through conventional biological treatment trains. The manuscript explores the convergent aspects of wastewater treatment process on overall microbiome and resistome dynamics across geographically-distinct WWTPs located in six countries and makes critical assessments on the potential for ARG mobilization. It finally reports a comprehensive relative resistome risk assessment, derived from metagenomic-based estimates of ARG and MGE prevalence in human

pathogens, and identifies clinically-relevant ARGs that are commonly being emitted to surface water around the world.

This manuscript is in preparation for *Microbiome*.

Attributions: Samples were collected, processed, and sequenced by Virginia Riquelme, Jake Metch, Emily Garner, Matt Blair, Joyce Zhu, Haniyyah Majeed, Greg House, Marjorie Willner, Gustavo Arango-Argoty, Ayella Maile-Moskowitz, Suraj Gupta, Ishi Keenum, Kris Mapili, me, and many others. I conducted the data analysis and led the writing of the manuscript. Peter Vikesland, Emily Garner, and Amy Pruden contributed to the data interpretation and well as manuscript preparation and review.

Chapter 4: *Critical review of metagenomic workflows for monitoring antibiotic resistance in water and wastewater*

Chapter 4 address Objective (3) through a systematic review of 95 workflows for the metagenomic investigation of surface water, wastewater, and recycled water resistomes. The manuscript proposes critical AR monitoring objectives for which metagenomics as a technique is particularly well suited for addressing. Guidance is provided for the appropriate design and implementation of metagenomic-based monitoring programs, including the use of DNA extraction techniques, process controls, the curation of ARG databases, and selection of bioinformatic software and parameters for universal comparability and reproducibility across studies. It concludes with a view of the future of metagenomics as a tool for the comprehensive monitoring of AR in aquatic matrices and posits the integration of such techniques into comprehensive risk assessment frameworks.

This manuscript will be submitted to *Critical Reviews in Environmental Science and Technology*.

Attributions: The literature review was conducted by me, Erin Milligan, Jeanette Calarco, Krista Liguori, and Ishi Keenum. I conducted the manual extraction and analysis of data from the identified articles. Ishi Keenum and I wrote the first draft of the manuscript. Amy Pruden and Ishi Keenum conducted in-depth reviews and revisions of the manuscript and helped frame the article to get it to its current state.

Chapter 5: *Towards the standardization of Enterococcus culture methods for waterborne antibiotic resistance monitoring: a critical review and analysis of environmental trends*

Chapter 5 addresses Objective (3) through a critical review and meta-analysis of over 100 articles that report enumeration of resistant *Enterococcus* spp. in surface water, wastewater, and recycled water matrices. The manuscript proposes a standardized framework for the isolation and characterization of total and resistant *Enterococcus* from these environments, considering the universal applicability of datatypes derived from culture-based methods and standard techniques already in place for water quality monitoring. Methods that allow phenotyping, genotyping, and cross-isolate comparisons for epidemiological source tracking and human-health risk assessments are critically reviewed, with an emphasis on vancomycin resistance among the genus. Through the systematic collection of species and phenotypic frequency across studies and water matrices, the manuscript further provides overarching trends in antibiotic-resistant *Enterococcus* prevalence in various impacted water matrices to inform future studies.

This manuscript has been submitted to *Water Research*.

Attributions: The literature review was conducted by me, Erin Milligan, Jeanette Calarco, Krista Liguori, and Ishi Keenum. I conducted the manual extraction and analysis of data from the identified articles and writing of the manuscript. Valerie Harwood and Amy Pruden conducted in-depth reviews and revisions of the manuscript and lead the framing of the article.

Chapter 6: *Benchmarking quantitative metagenomics for wastewater-based surveillance of antibiotic resistance*

Chapter 6 addresses Objective (4) through a proof-of-concept experimental validation of internal reference standards for absolute quantification of ARGs in wastewater samples. The study applies deep metagenomic sequencing of replicate influent, activated sludge, and secondary effluent samples to identify the limits of quantification and detection of the analytical technique with direct comparison to qPCR. The study provides a discussion on the implications of non-targeted sequencing approaches for the wastewater-based surveillance systems and the statistical thresholds necessary for consistent ARG detection at extremely low abundances.

This manuscript is in preparation for *Applied Environmental Microbiology*.

Attributions: I came up with and designed the experiment myself. I also conducted the field sampling, sample processing, and coordination of all sequencing services. The qPCR data was generated by Gabriel Moldonado Rivera. I analyzed the data and led the writing of the manuscript. Amy Pruden and Peter Vikesland provided critical feedback in the writing of the first draft of the article.

Chapter 7: *Conclusions and recommendations for future work*

This final chapter briefly synthesizes the body work presented herein and summarizes its contributions to the field environmental AR research. Future research directions are also presented.

Published: In addition to the five manuscripts that are included in this dissertation, described above, several related collaborative works have recently been published:

1. Liguori, K., Keenum, I., **Davis, B. C.**, Calarco, J., Milligan, E., Harwood, V. J., Pruden, A. (2022) Antimicrobial resistance monitoring of water environments: A call for standardized methods and quality control. *Environmental Science & Technology* (in press) https://doi.org/10.1021/acs.est.1c08918
2. Keenum, I., Liguori, K., Calarco, J., **Davis, B. C.**, Milligan, E., Harwood, V. J., & Pruden, A. (2021). A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. *Critical Reviews in Environmental Science and Technology*, 1-25. https://doi.org/10.1080/10643389.2021.2024739
3. Pruden, A., Vikesland, P., **Davis, B. C.**, de Roda Husman, A.M. (2021). Seizing the moment: now is the time for global surveillance of antimicrobial resistance in wastewater environments. *Current Opinion in Microbiology*, 64, 91-99. https://doi.org/10.1016/j.mib.2021.09.013
4. Majeed, H. J., Riquelme, M. V., **Davis, B. C.**, Gupta, S., Angeles, L., Aga, D. S., Garner, E., Pruden, A., & Vikesland, P. J. (2021). Evaluation of metagenomic-enabled antibiotic

resistance surveillance at a conventional wastewater treatment plant. *Frontiers in Microbiology*, 12, 1048. https://doi.org/10.3389/fmicb.2021.657954

5. Garner, E., **Davis, B. C.**, Milligan, E., Blair, M. F., Keenum, I., Maile-Moskowitz, A., Pan, J., Gnegy, M., Liguori, K., Gupta, S., Prussin II, A.J., Marr, L.C., Heath, L.S., Vikesland, P.J., & Pruden, A. (2021). Next generation sequencing approaches to evaluate water and wastewater quality. *Water Research*, 116907. https://doi.org/10.1016/j.watres.2021.116907

REFERENCES

1. Jonas, O., Irwin, A., Berthe, F., Le Gall, F. & Marquez, P. *Drug-Resistant Infections: A Threat to Our Economic Future.* **2**, (Washington D.C.: World Bank Group, 2017).

2. Lebov, J. *et al.* A framework for One Health research. *One Heal.* **3**, 44–50 (2017).

3. World Health Organization, World Organisation for Animal Health & Food and Agriculture Organization of the United Nations. *Monitoring and evaluation of the global action plan on antimicrobial resistance: Framework and recommended indicators*. (2019).

4. O'Neill, J. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*. **4**, (2014).

5. The World Bank Group. *Pulling Together to Beat Superbugs: Knowledge and Implementation Gaps in Addressing Antimicrobial Resistance*. (2019). doi:10.1596/32552

6. World Health Organization. *Global Action Plan on Antimicrobial Resistance*. (2015).

7. Berendonk, T. U. *et al.* Tackling antibiotic resistance: The environmental framework. *Nat. Rev. Microbiol.* **13**, 310–317 (2015).

8. JPIAMR. *Strategic Research and Innovation Agenda on Antimicrobial Resistance*. *Strategic Research and Innovation Agenda on Antimicrobial Resistance* (2019).

9. Bürgmann, H. *et al.* Water and sanitation: An essential battlefront in the war on antimicrobial resistance. *FEMS Microbiol. Ecol.* **94**, 1–14 (2018).

10. Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *The Lancet Planetary Health* **2**, e398–e405 (2018).

11. United States. *National Action Plan for Combating Anitibiotic-Resistant Bacteria 2020-2025*. (2020).

12. European Commission. *A European One Health Action Plan against Antimicrobial Resistance (AMR)*. (2017).

13. Garland, J. L. *et al.* Geospatial Distribution of Antimicrobial Resistance Genes in US Rivers and Streams United States Environmental Protection Agency. (2019).

14. Kirby, A. Antibiotic Resistance Monitoring Using Extended Spectrum Beta-Lactamase-Producing E. coli as an Indicator Organism. in *AWWA International Symposium on Potable Reuse* (2020).

15. Manaia, C. M. *et al.* Antibiotic resistance in wastewater treatment plants : Tackling the black box. *Environ. Int.* **115**, 312–324 (2018).

16. Miłobedzka, A. *et al.* Monitoring antibiotic resistance genes in wastewater environments: the challenges of filling a gap in the One-Health cycle. *J. Hazard. Mater.* 127407 (2021). doi:10.1016/j.jhazmat.2021.127407

17. Davis, B. C. *et al.* Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. *Environ. Sci. Technol* (2020). doi:10.1021/acs.est.0c05567

18. Quintela-Baluja, M. *et al.* Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. *Water Res.* **162**, 347–357 (2019).

19. Garner, E. *et al.* Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res.* **194**, 116907 (2021).

20. Bengtsson-Palme, J., Larsson, D. G. J. & Kristiansson, E. Using metagenomics to investigate human and environmental resistomes. *J. Antimicrob. Chemother.* 2690–2703 (2017). doi:10.1093/jac/dkx199

21. Pruden, A., Vikesland, P. J., Davis, B. C., Maria, A. & Husman, D. R. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Curr. Opin. Microbiol.* **64**, 91–99 (2021).

22. Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on human sewage. *Nat. Commun.* (2019). doi:10.1038/s41467-019-08853-3

23. Larsson, D. G. J. *et al.* Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environ. Int.* **117**, 132–138 (2018).

24. Liguori, K. *et al.* Antimicrobial Resistance Monitoring of Water Environments: A Call for Standardized Methods and Quality Control. *Environ. Sci. Technol.* (2022).

25. Costa, V. M. D. *et al.* Antibiotic resistance is ancient. *Nature* 1–5 (2011). doi:10.1038/nature10388

26. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, (2015).

27. Davies, J. & Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417–433 (2010).

28. Larsson, D. G. J. & Flach, C.-F. Antibiotic resistance in the environment. *Nat. Rev. Microbiol.* **0123456789**, (2021).

29. Martínez, J. L. Ecology and evolution of chromosomal gene transfer between environmental microorganisms and pathogens. *Microb. Transm.* 139–160 (2017). doi:10.1128/9781555819743.ch8

30. Forsberg, K. J. *et al.* The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science (80-.).* **337**, 1107–1111 (2012).

31. Pruden, A., Pei, R., Storteboom, H. & Carlson, K. H. Antibiotic resistance genes as emerging contaminants: Studies in northern Colorado. *Environ. Sci. Technol.* **40**, 7445–7450 (2006).

32. Pruden, A. *et al.* An Environmental Science and Engineering Framework for Combating Antimicrobial Resistance. *Environ. Eng. Sci.* **35**, 1005–1011 (2018).

33. Lupo, A., Coyne, S. & Berendonk, T. U. Origin and evolution of antibiotic resistance : the common mechanisms of emergence and spread in water bodies. *Front. Microbiol.* **3**, 1–13 (2012).

34. Chow, L. K. M., Ghaly, T. M. & Gillings, M. R. A survey of sub-inhibitory concentrations of antibiotics in the environment. *J. Environ. Sci. (China)* **99**, 21–27 (2021).

35. Kraupner, N. *et al.* Evidence for selection of multi-resistant E. coli by hospital effluent. *Environ. Int.* **150**, 106436 (2021).

36. Bengtsson-Palme, J. & Larsson, D. G. J. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. (2016). doi:10.1016/j.envint.2015.10.015

37. Kraemer, S. A., Ramachandran, A. & Perron, G. G. Antibiotic pollution in the environment: From microbial ecology to public policy. *Microorganisms* **7**, 1–24 (2019).

38. Wales, A. D. & Davies, R. H. Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics* **4**, 567–604 (2015).

39. Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E. & Larsson, D. G. J. BacMet: Antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* **42**, 737–743 (2014).

40. Karkman, A., Pärnänen, K. & Larsson, D. G. J. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat. Commun.* 1–8 (2019). doi:10.1038/s41467-018-07992-3

41. Bengtsson-Palme, J. *et al.* Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Sci. Total Environ.* **572**, 697–712 (2016).

42. Moralez, J., Szenkiel, K., Hamilton, K., Pruden, A. & Lopatkin, A. J. Quantitative analysis of horizontal gene transfer in complex systems. *Curr. Opin. Microbiol.* **62**, 103–109 (2021).

43. Dancer, S. J. Controlling hospital-acquired infection: Focus on the role of the environment and new technologies for decontamination. *Clin. Microbiol. Rev.* **27**, 665–690 (2014).

44. Weber, D. J., Anderson, D. & Rutala, W. A. The role of the surface environment in healthcare-associated infections. *Curr. Opin. Infect. Dis.* **26**, 338–344 (2013).

45. Leonard, A. F. C., Zhang, L., Balfour, A. J., Garside, R. & Gaze, W. H. Human recreational exposure to antibiotic resistant bacteria in coastal bathing waters. *Environ. Int.* **82**, 92–100 (2015).

46. Leonard, A. F. C. *et al.* Exposure to and colonisation by antibiotic-resistant E. coli in UK

coastal water users: Environmental surveillance, exposure assessment, and epidemiological study (Beach Bum Survey). *Environ. Int.* **114**, 326–333 (2018).

47. Nadimpalli, M. L. *et al.* Urban informal settlements as hotspots of antimicrobial resistance and the need to curb environmental transmission. *Nat. Microbiol.* **5**, 787–795 (2020).

48. ASCE. *A Comprehensive Assessment of America's Infrastructure: 2021 Report Card*. (2021).

49. Loewenstein, Y. *et al.* Protein function annotation by homology-based inference. *Genome Biol.* **10**, 207 (2009).

50. Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems* **3**, 1–9 (2018).

51. Gweon, H. S. *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ. Microbiomes* **14**, 1–15 (2019).

52. Garner, E. *et al.* Towards risk assessment for antibiotic resistant pathogens in recycled water: A systematic review and summary of research needs. *Environ. Microbiol.* **00**, (2021).

CHAPTER 2: DEMONSTRATING AN INTEGRATED ANTIBIOTIC RESISTANCE GENE SURVEILLANCE APPROACH IN PUERTO RICAN WATERSHEDS POST-HURRICANE MARIA

Benjamin C. Davis[1], Maria Virginia Riquelme[1,2,*], Graciela Ramirez-Toro[3], Christina Bandaragoda[4], Emily Garner[1,5], William J. Rhoads[1], Peter Vikesland[1], Amy Pruden[1,*]

1 Virginia Tech, Department of Civil & Environmental Engineering, Blacksburg, VA 24060, United States

2 Current Affiliation: Diversigen, Inc., Houston, TX 77021, United States

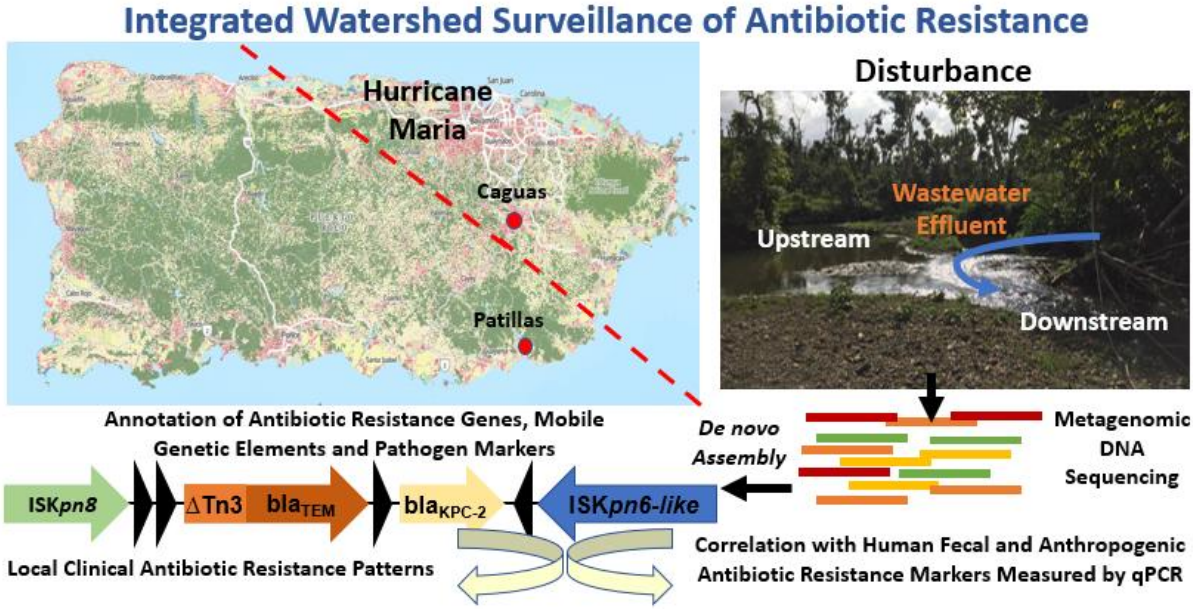
3 Inter American University, Center for Environmental Education, Conservation and Research, San Germán, PR 00683, United States

4 Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, 98195, United States

5 Current Affiliation: West Virginia University, Department of Civil & Environmental Engineering, Morgantown, WV 26506, United States

*Corresponding Author

TOC Graphic:



ABSTRACT

Comprehensive surveillance approaches are needed to assess sources, clinical relevance, and mobility of antibiotic resistance genes (ARGs) in watersheds. Here, we examined metrics derived from shotgun metagenomic sequencing and relationship to human fecal markers (HFMs; crAssphage, enterococci) and anthropogenic antibiotic resistance markers (AARMs; *int*I1, *sul*1) in three distinct Puerto Rican watersheds as a function of adjacent land-use and wastewater treatment plant (WWTP) input six months after Hurricane Maria, a Category V storm. Relative abundance and diversity of total ARGs increased markedly downstream of WWTP inputs, with ARGs unique to WWTP and WWTP-Impacted river samples predominantly belonging to the aminoglycoside and β-lactam resistance classes. WWTP and other anthropogenic inputs were similarly associated with elevated resistome risk scores and mobility incidence (M%). Contig analysis indicated a wide variety of mobile β-lactam ARGs associated with pathogens downstream of WWTP discharge that were consistent with regional clinical concern, e.g., *Klebsiella pneumoniae* contigs containing KPC-2 within an *ISKpn6*-like transposase. HFMs and AARMs correlated strongly with the absolute abundance of total ARGs, but AARMs better predicted the majority of ARGs in general (85.4% vs <2%) and β-lactam ARGs in particular. This study reveals sensitive, quantitative, mobile, clinically-relevant, and comprehensive targets for antibiotic resistance surveillance in watersheds.

INTRODUCTION

The need for environmental surveillance to better understand the development, spread, and circulation of antibiotic resistance between and among humans, animals, food, and water networks is increasingly being recognized by the World Health Organization (WHO) and others (Aarestrup & Woolhouse, 2020; United Nations, 2017; World Health Organization, 2015). Surface waters and sediments are an environmental monitoring point of interest (Qiao, Ying, Singer, & Zhu, 2018; J. Xu et al., 2015; Zhu et al., 2017), as they are recipients, reservoirs, and pathways for the transport of antibiotics, antibiotic resistance genes (ARGs), and antibiotic resistant bacteria (ARB) from a variety of point and non-point sources, including wastewater treatment plants (WWTPs) and livestock operations (Amos, Zhang, Hawkey, Gaze, & Wellington, 2014; Knapp et al., 2012; Luo et al., 2010; Marti, Variatza, & Balcazar, 2014; Pruden, Arabi, & Storteboom, 2012). Correspondingly, watershed monitoring programs aimed at tracking antibiotic resistance impacts are currently being explored by the U.S. Environmental Protection Agency (EPA) (Garland et al., 2019; Nappier, Ichida, Jaglo, Haugland, & Jones, 2019) and Centers for Disease Control (Kirby, 2020).

Surveillance of various indicators of and contributing factors to antibiotic resistance in watersheds can help identify input sources, understand microbial ecological processes contributing to ARG amplification or attenuation, establish a baseline for targeting and prioritizing mitigation efforts (Berendonk et al., 2015; Crofts, Gasparrini, & Dantas, 2017), and assess impacts of major disruptions, such as storms (Yu et al., 2018). Of particular interest is whether ARGs in rivers and streams originate from human fecal material, versus other sources, and the extent to which horizontal gene transfer and selection pressure drive *in situ* propagation and amplification of ARGs. In a recent *in silico* study of ~500 publicly-available metagenomes, it was found that, in the majority of cases, total ARGs in WWTP-influenced environments strongly correlated with the highly-specific human fecal marker (HFM), crAssphage (Karkman, Pärnänen, & Larsson, 2019). Given that the exceptions to this case were rivers in India with

extremely high concentrations of antibiotics, the authors proposed a monitoring scheme in which deviations from correlation between ARGs and HFMs are flagged as presumable hot spots for environmental selection pressure and gene exchange. While such an approach is promising, improvements can be made by increasing resolution with respect to which ARGs are best predicted by HFMs versus coming from other sources, in addition to characterizing these ARGs in terms of their mobility, likelihood of being carried by human pathogens, and correspondence to endemic antibiotic resistance in clinical settings (Aarestrup & Woolhouse, 2020; Karkman et al., 2019; Pruden et al., 2018).

The purpose of this study was to pilot-test an integrated antibiotic resistance monitoring scheme targeting three distinct watersheds on the island of Puerto Rico representing a gradient of anthropogenic inputs, having experienced the same massive natural disturbance as a result of Hurricane Maria, a Category V storm that made landfall in September 2017. Following Maria, over a third of the 55 WWTPs operated by the Puerto Rico Aqueduct and Sewer Authority were inoperable due to widespread flooding and power outages, allowing untreated raw sewage to be discharged into streams, rivers, and coastal waters (EPA, 2017). Urban flooding and hurricanes are known to significantly disturb the microbial composition and quality of surface waters leading to elevated levels of fecal indicator organisms and pathogenic bacteria (Kapoor, Gupta, Pasha, & Phan, 2018; Schwab et al., 2007; Sinigalliano et al., 2007; ten Veldhuis, Clemens, Sterk, & Berends, 2010). Inputs of untreated human waste streams have also been directly linked to the emergence of carbapenem- and cephalosporin-resistant *Enterobacteriaceae* in urban sediments (Marathe et al., 2017). Thus, analysis of Puerto Rican watersheds post-Hurricane Maria could further provide insight into the extent to which antibiotic resistance is dispersed by major storms.

The integrated surveillance scheme evaluated herein combined shotgun metagenomic DNA sequencing for high resolution comparison of resistomes (i.e., total ARGs) along with more technologically-accessible quantitative polymerase chain reaction (qPCR) targets. Specifically, analysis of a relatively pristine rural low impact (RLI) watershed served to distinguish native/background ARGs from those found in periurban medium-impact (PMI) and urban high-impact (UHI) watersheds. Correlations with crAssphage and enterococci (HFMs) (Stachler, Akyon, Carvalho, Ference, & Bibby, 2018) versus the class 1 integron integrase gene, *int*I1 (Gillings et al., 2014), and associated sulfonamide ARG, *sul*1 (Gillings, 2014) [anthropogenic antibiotic resistance markers (AARMs)], aided in identifying ARG classes and mechanisms specifically associated with human fecal versus more generalized anthropogenic pollution, respectively. Read matching and assembly strategies were applied to gain insight into the roles of WWTP discharges and non-point sources on ARG occurrence patterns, mobility, and association with putative pathogens of concern regionally in Puerto Rico. The approach demonstrated herein can be applied to other watersheds, globally, with varying degrees of anthropogenic and other impacts, to assess effects of disruptive events such as major storms and begin to identify and prioritize mitigation efforts and interventions to protect public health.

MATERIALS AND METHODS

Sample collection and processing

Location and characteristics of the sampling sites are shown in Figure 2-1. According to ArcGIS analysis (USA NLCD Land Cover), watersheds RLI and PMI are sparsely populated,

rural systems with less than 1% cropland and 99.5% and 96% of land cover attributed to evergreen forest, respectively. A single WWTP treating 4000 cubic meters per day (CMD at the outlet of the PMI catchment services 20,000 residents in the city of Patillas (Table S2-1). There were no identified wastewater point-sources to the RLI catchment. The UHI watershed is a highly developed urban area with a 2200 CMD WWTP servicing 28,000 residents in the city of Aguas Buenas to the west and a 150,000 CMD plant servicing 130,000 residents in the city of Caguas to the east. All WWTPs were in operation at the time of sampling. Moderate commercial and industrial development was noted along the length of UHI's riverine system. Sixteen sampling sites were chosen to capture within-river and across-watershed anthropogenic gradients, sampling as far upstream as possible for pristine samples and directly upstream and downstream of three WWTPs. The influent and final effluent of the Patillas WWTP were sampled as combined sewage resistome of the local population (Hutinel et al., 2019).

Surface water (2 L) and sediment (~5 g) grab samples were collected in April of 2017 (6 months after Hurricane Maria), immediately placed on ice, and transported to El Centro de Educación, Conservación e Interpretación Ambiental (CECIA) for same-day processing. Water samples were taken from the top 0.5 m of the surface of the center of the flow in autoclaved polypropylene bottles. One liter of water was collected in an acid-washed HDPE bottle for inorganics analysis (Table S2-11). Sediment samples were taken in 50 mL tubes from the top 2-3 cm of sediment. The water samples for DNA extraction were homogenized and filter-concentrated in duplicate onto 0.22-µm mixed-cellulose ester filters (Millipore, USA) until clogging. The volume that could be filtered was recorded, as filtering was limited by the extent of clogging for each sample (15-1000 mL) (Table S2-2). Filters were transferred to sterile 2-mL tubes and fixed in 50% ethanol. For sediments, samples were homogenized, 0.5 g were aliquoted in duplicate and then fixed in 100% ethanol. All samples were stored at -20°C before being shipped on ice to Virginia Tech, Blacksburg, VA, for further processing. One liter of autoclaved deionized water was prepared at Virginia Tech prior to the trip and accompanied all sampling events. It was then shipped back unopened, and DNA extracted to serve as a "trip blank".

DNA extraction and qPCR

The ethanol-fixed filters were fragmented (~1 cm^2) with flame-sterilized tweezers and transferred to 2-mL DNA extraction tubes. The remaining ethanol solution was centrifuged at 5,000 x g for 10 minutes and the pellet resuspended in FastDNA Spin Kit for Soil sodium phosphate buffer (MPBio, USA) and added to the extraction tube. Both water and sediment samples were homogenized via bead-beating (40 seconds at 6 m/s) with the FastPrep-24™ 5G (MPBio, USA) and then further extracted according to manufacturer instructions. One liter of autoclaved deionized water was also filtered and extracted as a filter bank to ensure no contamination during DNA extraction. Using qPCR, the abundance of total bacteria (16S rRNA genes) (Suzuki & Taylor, 2000), *sul1* (Pei, Kim, Carlson, & Pruden, 2006), *int*I1 (Hardwick, Stokes, Findlay, Taylor, & Gillings, 2008), crAssphage-056 (Stachler et al., 2017), and enterococci (EPA, 2012) were quantified in analytical triplicate. Each assay included autoclaved deionized water as no template control. Dilution series were carried out for the 16S rRNA gene as a representative assay to select a dilution for each sample type (i.e., bulk water, sediment, and wastewater) that minimized effects of PCR inhibitors (Table S2-2). Standard curves were generated from double-stranded gBlock™ gene fragments (IDT, USA), resuspended according to manufacturer specifications, and quantified via dsDNA High Sensitivity Assay kit on a Qubit® Fluorometer (Invitrogen, USA). The minimum accepted qPCR standard curve

efficiencies and R^2 values were 80% and 0.980, respectively. The limit of quantification was set as the lowest standard that amplified at least in duplicate.

Metagenomic Sequencing and Bioinformatics Pipeline

One-hundred nanograms of DNA (260/280 ratio 1.8-2.1; Table S2-2) from a subset of 24 samples were subjected to library preparation using a NexteraXT DNA Library Prep Kit and were sequenced via 2 x 150 bp paired-end shotgun metagenomic sequencing on an Illumina NovaSeq 6000 by Diversigen, Inc. (Houston, Texas). All metagenomic reads were deposited to the Sequence Read Archive under accession number PRJNA626373. For quality trimming, gene annotation, and contig assembly, all samples were analyzed via MetaStorm (G. Arango-Argoty et al., 2016). For functional analysis, short reads were annotated against the Comprehensive Antibiotic Resistance Database (McArthur et al., 2013) (CARD, v2.0.1), the ACLAME (Leplae, Lima-Mendez, & Toussaint, 2009) database for plasmid-associated protein alignment (v0.4), and MetaStorm's mobile genetic elements (MGEs) dataset, which is a collection of all NCBI non-redundant genes related to any of the following keywords — transposase, transposon, integrase, integron, and recombinase to identify indicators of mobility (Forsberg et al., 2014a) (MetaStorm MGEs, v1.0) using MetaStorm's default parameters (minimum length alignment = 25 aa, aa identity ≥ 80%, e-value cutoff 1e-10). To estimate relative abundance, functional genes were normalized to 16S rRNA abundance as enumerated by alignment to the Greengenes Database (Larsen et al., 2006). Absolute abundances were calculated by multiplying the relative abundances of the functional genes by total abundance of 16S rRNA quantified by qPCR (Garner et al., 2018). Reads were assembled via MetaStorm's assembly pipeline, which utilizes the IDBA-UD *de novo* assembler (Peng, Leung, Yiu, & Chin, 2012), and the resulting contigs were analyzed via NanoARG (G. A. Arango-Argoty et al., 2019) for functional annotation, taxonomic assignment using the Centrifuge engine (Kim, Song, Breitwieser, & Salzberg, 2016), and gene contextualization. NanoARG outputs were filtered for Centrifuge scores ≥300 (hit length ≥ 31) and Diamond (Buchfink, Xie, & Huson, 2014) alignments with bitscores ≥ 50 and e-values ≤ 1e-15. Contigs were then comprehensively searched for WHO Global Priority List of Antibiotic-Resistant Bacteria (Tacconelli et al., 2018; WHO, 2017) ARG-taxonomy co-locations. Contigs were also analyzed via the MetaCompare pipeline (Oh et al., 2018) for the determination of each metagenome's relative "resistome risk" (Martínez, Coque, & Baquero, 2015; Oh et al., 2018) by projecting each sample into a 3-dimensional hazard space, normalizing the co-occurrence of (1) ARGs, (2) ARGs and MGEs, and (3) ARGs, MGEs, and human bacterial pathogens to the total contig library size.

Systematic Review of Local Clinical Data

A systematic literature review was carried out to identify antibiotic-resistant bacterial infections of concern locally in Puerto Rico. Boolean keyword searches such as "antibiotic" or "antimicrobial" and "resist*" and "clinic" and "Puerto Rico" were carried out in the Web of Science and PubMed. Search terms and relevant results are summarized in Table S2-4.

Statistical Analysis and Visualizations

Data were analyzed using R (v3.6.1) with a significance cutoff of $\alpha < 0.05$. For analysis of significant differences in abundances and Shannon diversities between groups, the non-parametric Wilcox Rank Sum test was used. Shannon diversities, non-metric multidimensional scaling (NMDS) plots, and analysis of ordination similarities were generated in the 'vegan' (Oksanen et al., 2009) (v2.5-6) package with functions 'diversity', 'isoMDS', 'anosim', 'envfit', and 'protest'. Boxplots, non-metric multidimensional scaling coordinates, and stacked bar charts

were visualized with the 'ggplot2' (v3.2.1) package. Heat maps and Venn diagrams were generated using the 'pheatmap' (v1.0.12) and 'VennDiagram' (v1.6.20) packages, respectively.

RESULTS AND DISCUSSION

Defining Pristine, Core, Background, and Anthropogenically-Sensitive Resistomes

Metagenomic sequence quality and detection of ARGs are summarized in SI Section 2-1. Notably, very strong correlations were found between metagenomic versus qPCR enumeration of both *sul*1 and *int*I1 (Pearson R^2 >0.9), supporting quantitative analysis of metagenomic data (Figure S2-1). To gain insight into which ARGs were anthropogenically-sensitive, we characterized the Pristine, Core, and Background resistomes. We considered ARGs found in samples furthest upstream in each catchment to represent Pristine conditions, while the Core resistome was defined as the subset of ARGs common to every sample. The Background resistome, i.e., the portion of the resistome that is least sensitive to anthropogenic influence, was further defined as ARGs found in the Core + Pristine samples. Remarkably, 282 of the 816 ARG types observed were shared among all four sample groupings (WWTP, WWTP-Impacted, Impacted, Background), representing 93.01% of the total abundance (Figure 2-2). The Core resistome shared across all samples consisted of 63 ARGs and was primarily composed of multidrug (75.2%), peptide (8.8%), and macrolide-lincosamide-streptogramin (MLS) (5.2%) ARGs. The Background resistome consisted of 394 ARGs and was dominated by the same three classes (53.5%, 5.1%, 12.5%, respectively).

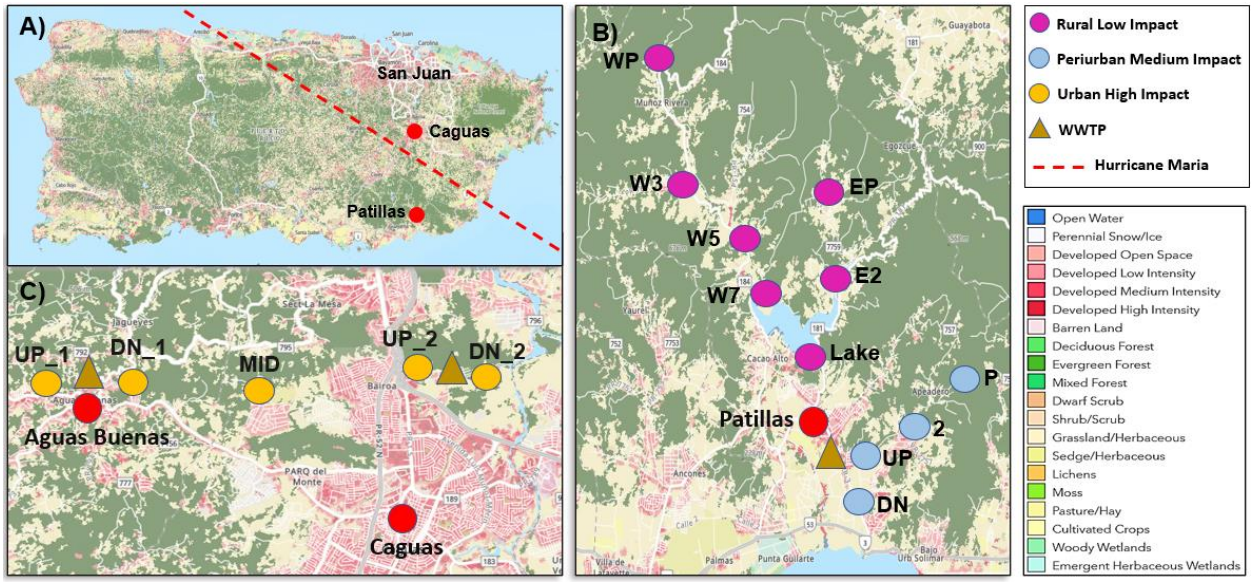


Figure 2-1: Sampling scheme for surface water and wastewater samples across Puerto Rico. Geographical distribution of sampling sites and corresponding municipalities (A). The dashed red line indicates the path of Hurricane Maria. (B) Sampling points for the watersheds Rural Low Impact (RLI) and Periurban Medium Impact (PMI). (C) Sampling points for the Urban High Impact (UHI) watershed.

The types of ARGs detected in high abundance in the Core resistome across watersheds were consistent with the assumption that they were not anthropogenically- or clinically-relevant. For example, the dominant functional categories were almost exclusively intrinsic ATP-binding cassette, major facilitator superfamily, and resistance-nodulation-cell division antibiotic efflux pump systems common in both Gram-negative and Gram-positive bacteria (Table S2-6). The relative proportions of each ARG class were noted to be distinct across the four sample groupings (Figure 2-2, S3-2). For example, ARGs that were unique to the WWTP-Impacted and WWTP compartments, and their overlap, were proportionally dominated (67.8%) by aminoglycoside and β-lactam ARGs.

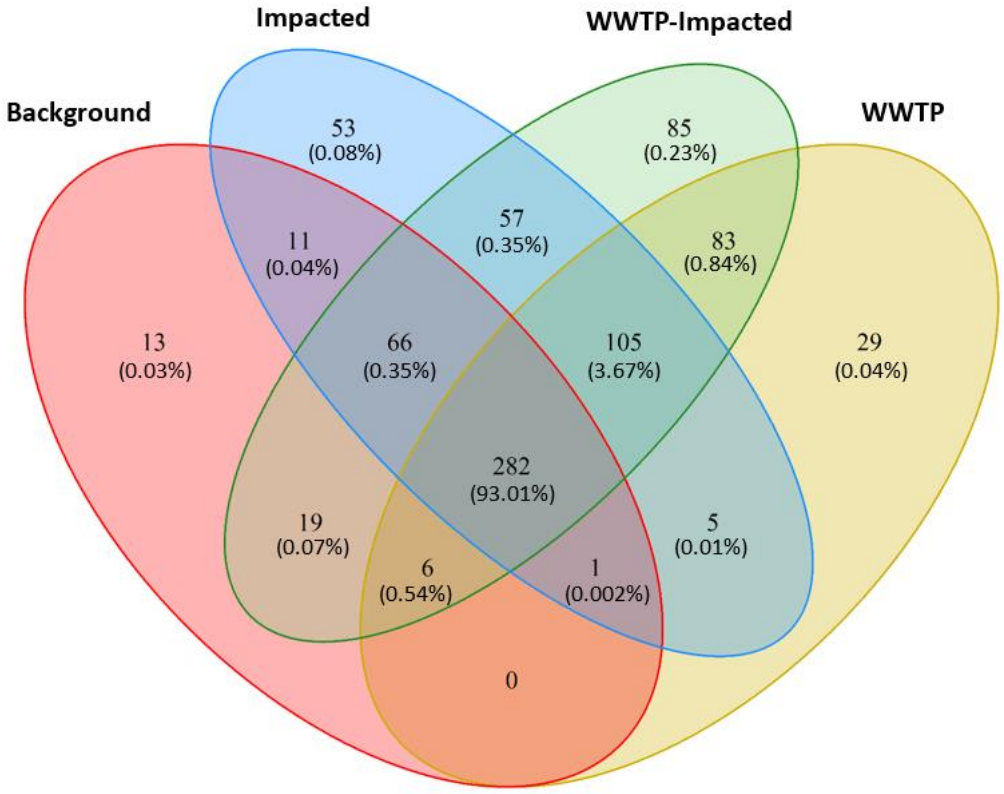


Figure 2-2: Venn Diagram of overlapping unique ARGs across all watersheds and WWTP samples through metagenomics. ["Background" ARGs are the Core ARGs common to all 24 samples plus ARGs identified in the pristine samples furthest upstream in each catchment, "Impacted" ARGs are those from sites influenced by adjacent land-use, "WWTP-Impacted" are those from any sites where their bulk water consists of some portion of wastewater effluent, and WWTP are those from the influent and effluent of the 4,000 CMD WWTP in the PMI watershed]. Values in parentheses correspond to the percent relative abundance of each ARG category relative to the total abundance of ARGs detected across all samples.

Trends in Abundance and Diversity of ARGs across Watersheds with Varying Land-Use and WWTP Inputs

A striking trend of increased total ARG relative abundance with increased anthropogenic impact was noted across watersheds, with the WWTP samples representing the theoretical maximum (Figure 2-3A). Total ARG relative abundance was significantly higher in WWTP and UHI watershed samples than those collected from the PMI or RLI watersheds (Wilcox, $p<0.05$). Hierarchical clustering of Bray-Curtis dissimilarities of total ARG absolute abundances generated two distinct clusters (Figure S2-3). The first cluster was comprised of bulk water samples taken directly downstream of WWTP discharge, urban sediments, and the raw and treated wastewater, while the second cluster consisted of rural and lesser impacted riverine sites.

Remarkably, diversity of total ARGs followed a similar trend as total ARG relative abundance (Figure 2-3B). The most diverse classes of ARGs across all sites were multidrug and β-lactam, with average Shannon diversities of 3.52 and 3.49, respectively, highlighting the array of β-lactam ARGs inherent to these environments. The Shannon diversity was significantly higher in UHI and WWTP samples than in RLI or PMI samples (Wilcox, $p<0.05$). While WWTP samples were not significantly higher than UHI samples (likely because of low statistical power of two WWTP samples), they did trend highest when compared to all other samples. The compositions of the resistomes were also measurably different across the watersheds (ANOSIM, R^2 0.1963, $p < 0.05$) (Figure S2-4A).

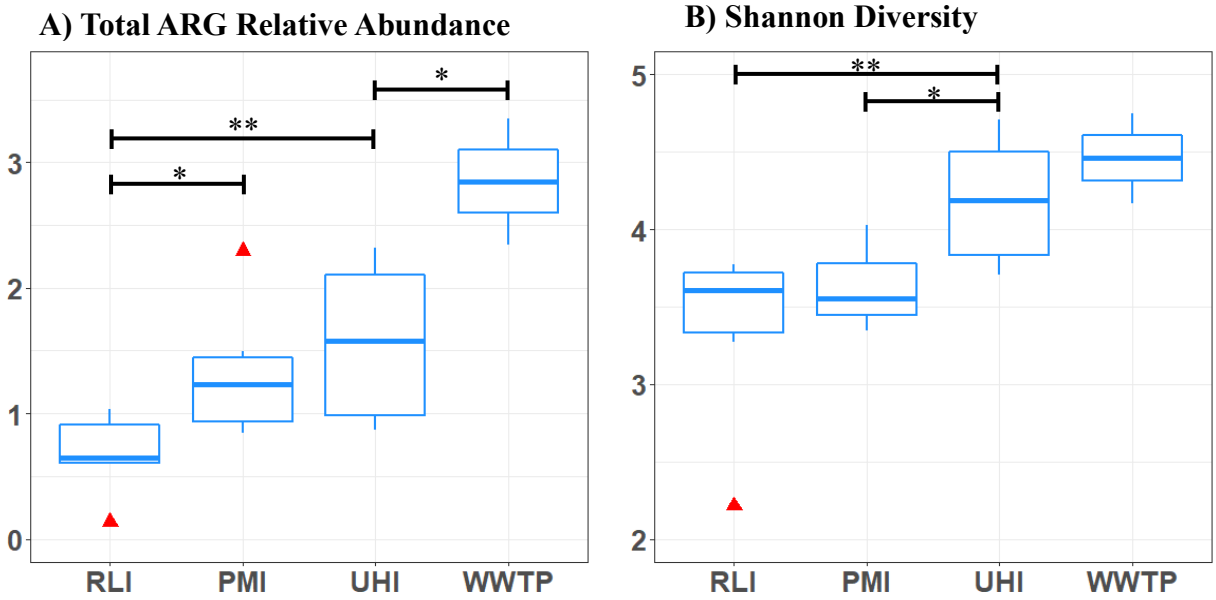


Figure 2-3: Resistome abundance and diversity across watersheds. (A) Total ARG relative abundance (ARGs/16S rRNA) identified by metagenomics read matching to CARD (v 2.0.1) by watershed: Rural Low Impact (RLI, n=7), Periurban Medium Impact (PMI, n=6), and Urban High Impact (UHI, n=9) and a 4,000 CMD WWTP (Influent and Effluent) serving watershed PMI. (B) Shannon diversity of detected ARGs by watershed. Midlines represent the median and the box represents the upper and lower quartiles (25th and 75th percentiles). Outliers are indicated

by red triangles. Pairwise differences were determined by Wilcoxon Rank Sum test. Significance cutoffs: * 0.05, ** 0.01, *** 0.001.

WWTP Effluents as a Key Driver of Riverine Resistomes

WWTP effluent was associated with a significant increase in the absolute and relative abundance of ARGs in downstream versus upstream riverine samples collected across all three WWTP discharge sites (Wilcox, $p < 0.001$) (Figures S3-5, S3-6). Further, the composition of the resistome also shifted from upstream to downstream (ANOSIM, R^2 0.2465, $p < 0.001$) (Figure S2-4A). Certain ARG classes; including aminoglycoside, β-lactam, fluoroquinolone, MLS, phenicol, tetracycline, and trimethoprim, were higher in both relative and absolute abundance in bulk water of sites with any portion of flow consisting of effluent (n=5), when compared to sites unimpacted by wastewater (n=11) (Table S2-7) (Wilcox, $p < 0.05$). Aminoglycoside, β-lactam, and tetracycline classes also increased in richness (i.e., total number of unique ARG types observed) (Table S2-7), although inherent variance in ARG class richness limited the ability to compare across classes (e.g., there are only 4 known *sul* gene variants, but over 1,500 gene variants in the β-lactam class in CARD). Still, the increase in β-lactam ARG richness from 46.7 ± 14.0 to 85.6 ± 27.2 from upstream to downstream of WWTPs was particularly striking.

Diversity of β-lactam ARGs and their Taxonomic Associations Across Watersheds

As a class of resistance with critical clinical significance globally and substantial diversity and apparent anthropogenic influence in their occurrence among the samples collected in this study, the β-lactam ARGs were subject to further analysis. Across all samples, β-lactam ARGs represented 5.18% of the total ARG abundance, with 267 unique β-lactam ARGs detected spanning 78 different families, conferring resistance to all major β-lactam antibiotics. The greatest number of recognized gene variants within a single β-lactam ARG family was noted among several carbapenamase-encoding ARGs, including IMP (20 variants), KPC (4 variants), OXA (51 variants), and VIM (1 variant). Other notable β-lactamase families detected were GES (5 variants), CMY (10 variants), CTX-M (12 variants), SHV (9 variants), TEM (9 variants), and MOX (9 variants), which are all plasmid-mediated and common to Gram-negatives and *Enterobacteriaceae*. These ten clinically-relevant gene families were predominantly found in high abundance in the UHI watershed downstream of WWTP discharge, with the greatest absolute abundance in sediments but greatest relative abundance in the water column, as well as the influent and final effluent of the WWTP itself (Figure 2-4, Table S2-7, S3-8).

Using the WHO Global Priority List of Antibiotic-Resistant Bacteria(Tacconelli et al., 2018; WHO, 2017) as guidance, assembled contig libraries were comprehensively searched for clinically-relevant β-lactam ARG-taxonomy associations. Characteristics of the assembled contigs across samples and relationship to sequencing coverage are summarized in SI Section 3-2. Of interest were contigs corresponding to carbapenem and third-generation cephalosporin resistant *Acinetobacter baumanii, Pseudomonas aeruginosa,* and members of the family *Enterobacteriaceae* (including *Escherichia coli* and *Klebsiella pneumoniae*), which have been ranked as Priority 1: Critical. The most common β-lactam-taxa associations detected were the class D β-lactamases, OXA (oxacillinase), where at least a single association was found in every sample, co-located with 35 different taxonomic families representing 51 different genera. Notably, the most clinically-relevant co-locations among *A. baumanii* (OXA-35, 97, 420), *P. aeruginosa* (OXA-10), *K. pneumoniae* (OXA-163), and *E. coli* (OXA-10) were found almost

exclusively in the UHI watershed (Table S2-10). The second most common WHO-relevant ARG with the widest taxonomic distribution was the class B metallo-β-lactamase IMP (imipenemase) gene, which was found associated with 15 different genera across 15 taxonomic families, the most notable of which was *E. coli (IMP-8, 11, 47)*. By contrast, KPC (*Klebsiella pneumoniae carbapenamase*) contigs exhibited a narrow taxonomic and geographic distribution. KPC-2 was the only corresponding variant found in the contig dataset and was exclusively associated with *A. hydrophila*, *Proteus mirabilis*, and its genetic originator, *K. pneumoniae.* It was found in the water column downstream of WWTP discharge across all three plants, in the influent and final effluent of the WWTP, and in one water sample in the UHI watershed (UHI_Wat_UP_2; Table S2-10).

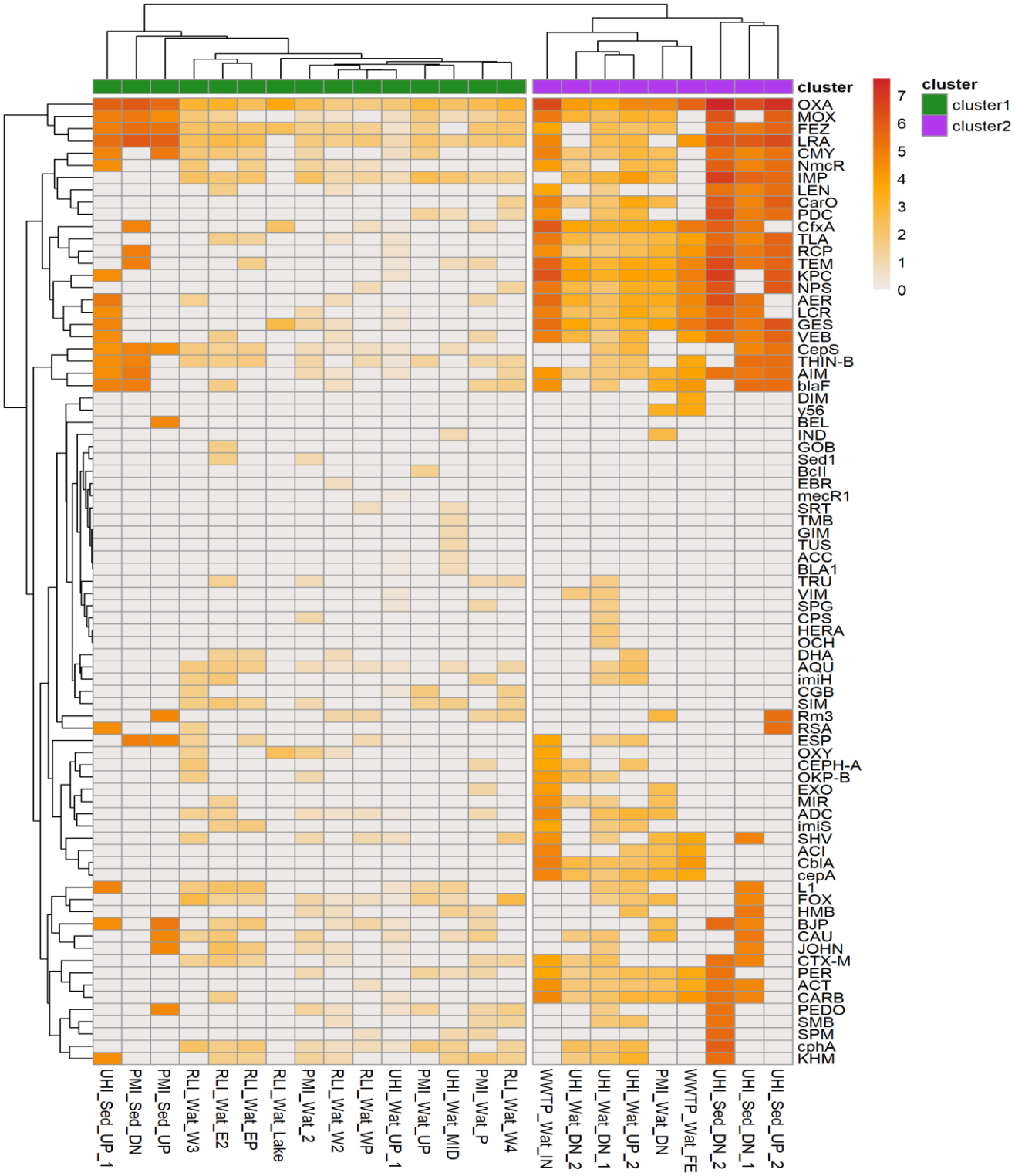


Figure 2-4: Heat map of absolute abundance of all β-lactam ARG types. Each gene/protein is represented as the sum of all detected variants (right axis). Color gradient represents $\log_{10}$ transformed absolute abundances. Complete-linkage clustering was used to hierarchically cluster both samples (columns) and ARGs (rows). Columns have been separated between the two main sample clusters. Sample coding convention is as described in

Relative Resistome Risk and Mobility

Assembled scaffolds were further analyzed to characterize relative "resistome risk", an empirical comparative analysis of the relative degree of co-localization of ARGs, MGEs and human pathogen taxonomies (Martinez, Coque, & Baquero, 2015; Oh et al., 2018). Consistent with the above observed trends in total ARG relative abundance and diversity, the relative resistome risk scores were significantly higher in the UHI and PMI watersheds than in the RLI watershed (Wilcox, $p < 0.05$) (Figure 2-5A). Interestingly, while WWTP and WWTP-impacted samples trended towards higher resistome risk scores, the two samples with the highest scores were impacted riverine samples, not samples of the WWTP itself (UHI_Sed_UP_2 and UHI_Wat_UP_2 samples upstream of the 150,000 CMD Caguas WWTP, but downstream of the most densely developed urban area). This suggests that mixtures of WWTP effluent and other anthropogenic stresses with natural riverine microbiomes elevate the potential for a mobile resistome and for pathogens to acquire new resistance genotypes. However, it is acknowledged that, although sequencing depth was uniform across this study, the percent of reads successfully assembled varied (<1% - 64%). In general, more pristine samples yielded greater sequence diversities and assembled less effectively. Although resistome risk scores were normalized to the contig library size, their comparability cannot be guaranteed. It is also noted that ARGs are especially difficult to accurately assemble from short reads, especially in complex environmental samples, because they occur in multiple contexts (Ayling, Clark, & Leggett, 2020; Ghurye, Cepeda-Espinoza, & Pop, 2016). Limitations of assembly and how they were addressed in this study are discussed further in SI Section S2-3.

Co-localization of indicators of mobility (e.g., plasmids, transposons, integrons) (Forsberg et al., 2014b; Ju et al., 2019) with ARGs on assembled contigs was also examined as a proxy for potential of ARGs to be transferred among bacteria and to provide information about their genetic history (Forsberg et al., 2014b; Ju et al., 2019). Analysis of variance indicated that WWTP samples and those directly impacted by wastewater (3.97%) had a greater average mobility incidence (M%)(Ju et al., 2019) than pristine environments (1.18%) ($p = 0.00157$). Within the ARG contig library (225,342 contigs), 83,193 contigs were also co-located with an indicator of mobility (36.92%). The average M% of individual ARGs detected in this study was 4.67%, 2.93%, and 2.35% for wastewater, sediment, and the bulk water samples, respectively, but this varied substantially by ARG class. Specifically, the β-lactam and aminoglycoside ARGs displayed disproportionally high M% due to many of the individual ARGs occurring at low gene abundances, but with a high likelihood of being co-located with an MGE indicator (Figure 2-5B). For example, there were 13 instances of a β-lactam ARG (ACT-28, ADC-23, CMY-99, DHA-22, IMP-33, OXA-228, OXA-398, OXA-420, OXA-97, OXY-1-2, PDC-90, TEM-194, TEM-33) occurring only on one contig that was also co-located with an indicator of mobility.

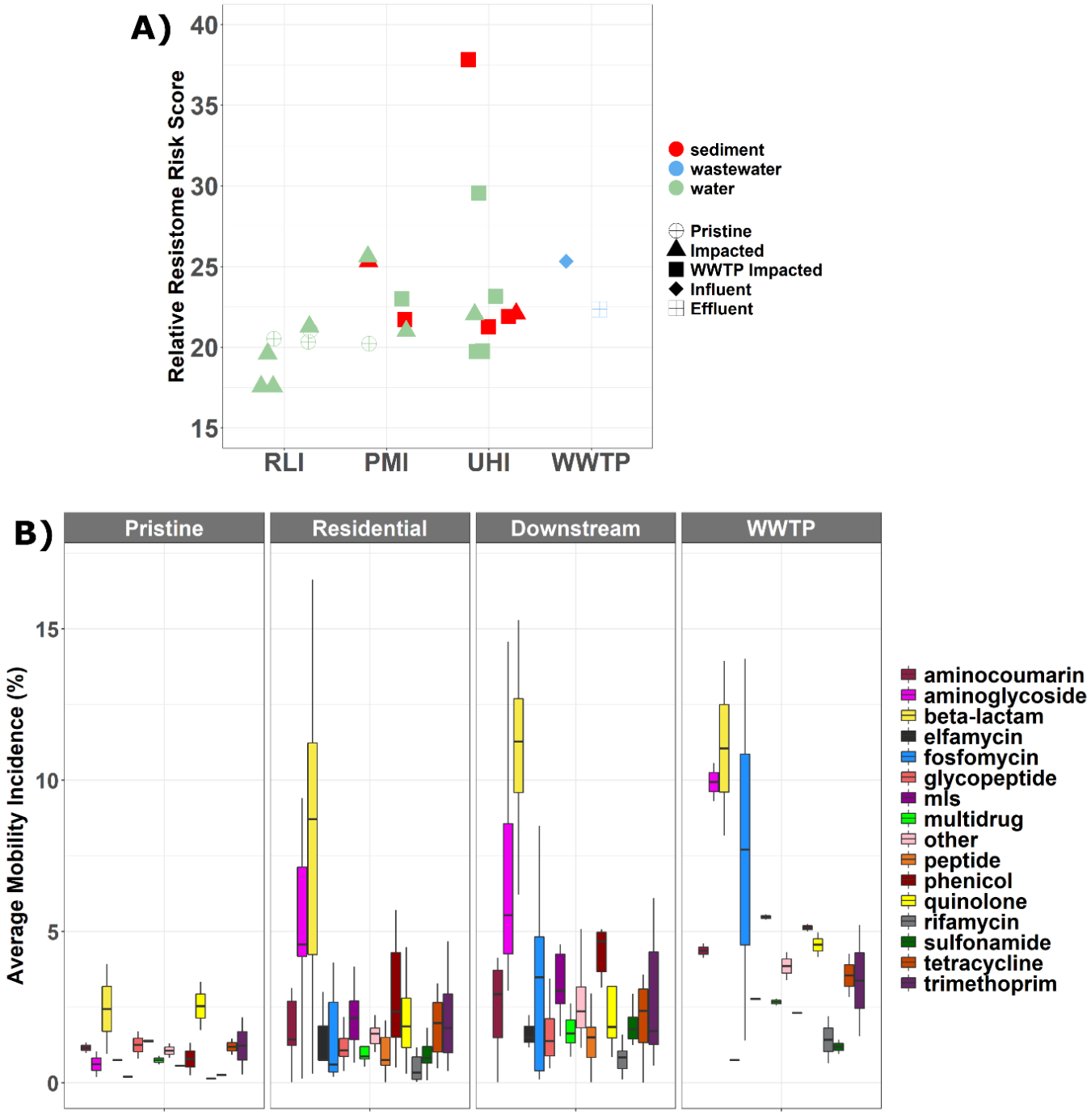


Figure 2-5: Relative resistome risk and ARG mobility potential across anthropogenic gradient. (A) MetaCompare relative resistome risk scores by watershed: Rural Low Impact (RLI, n=7), Periurban Medium Impact (PMI, n=6), and Urban High Impact (UHI, n=9) and a 4,000 CMD (Influent and Effluent) serving watershed PMI. Significance difference determined by Wilcoxon Rank Sum test. * = $p < 0.05$ **(B)** Mobility incidence (M%) compared across samples with varying degrees of anthropogenic stress ["Pristine" samples are sites at the headwaters of catchments within watersheds, "Impacted" ARGs are those from sites influenced

by adjacent land-use, "WWTP-Impacted" are those from any sites where their bulk water consists of some portion of wastewater effluent]. M% is defined as the percentage of ARG contigs co-occurring with an indicator of mobility (i.e., contigs where MGEs were annotated). ANOVA, $p = 0.00157$. Post-hoc TukeyHSD, WWTP-Impacted > Pristine; WWTP > Pristine ($p < 0.05$).

Comparing Clinically-Relevant ARG-MGE-Taxonomy Annotations with Local Clinical Resistance Prevalence

The systematic literature review identified clinical surveillance reports highlighting KPC and IMP in clinical isolates of *P. aeruginosa* (Wolter et al., 2009), *A. baumanii* (Robledo et al., 2010), and *K. pneumoniae* (Gregory et al., 2010) across Puerto Rico 2008-2012 (Table S2-4). Notably, KPC has been increasingly reported in hospitals in Latin America, including Brazil, Argentina, Uruguay, and Cuba (Belder et al., 2017). In an island-wide survey, 10,507 clinical Gram-negative bacilli isolates were investigated from 17 hospitals, representing primarily nosocomial-, but also environmentally-acquired infections and colonizations (Robledo, Aquino, & Vázquez, 2011). Of the 1,239 multi-β-lactam-resistant isolates, 534 (5.1% of all isolates) were KPC-positive (Table S2-4). The source of rapid dissemination of KPC is still unknown. The authors hypothesized that the small size of Puerto Rico (3,435 square miles), high population density, global connectivity of the San Juan airport, ease of ground-travel, and frequent patient transport between hospitals are contributing factors. Additionally, broad-spectrum antibiotics are widely used due to the already high number of extended-spectrum β-lactamase infections, while carriage of KPC on MGEs facilitates mobilization and co-selection (Robledo et al., 2011).

It is noteworthy that some β-lactam ARGs, including KPC, OXA, and IMP known to be locally problematic in the clinic, were found in the Background resistome in this study, highlighting the endemic nature of these ARGs. But such β-lactam ARGs also independently stood out in terms of both non-point-source adjacent land-use and WWTP inputs, elevating their numbers and diversity in affected rivers and streams (Table S2-9). KPC is generally located in Tn*4401* transposons and other Tn3-like genetic elements found on conjugative plasmids, flanked by ISK*pn6* and/or ISK*pn8*-like insertion sequences (Belder et al., 2017). Here, KPC-2 was found on six distinct *K. pneumoniae* contigs, co-located with the insertion sequence *ISKpn6*, two contigs of which were also co-located with TEM-17. These contigs were found downstream of the three WWTPs and in sample UHI_Wat_UP_2, consistent with detection of these ARGs in human feces and environments contaminated with human fecal material (Bengtsson-palme, Kristiansson, & Larsson, 2018). These findings demonstrate that clinically-relevant β-lactam ARGs could serve as sensitive surveillance targets (Huijbers, Flach, & Larsson, 2019) and also inform risk assessment models for individuals coming into direct contact with contaminated water, which would especially be a concern in a scenario of improper sewage control such as that instigated by Hurricane Maria.

Predicting Resistomes with Human Fecal and Anthropogenic Antibiotic Resistance Markers

To advance an integrated surveillance scheme, we cross-compared the above resistome characterization via metagenomics with direct enumeration of the highly-specific HFM, crAssphage 056 (Stachler et al., 2017) and the large ribosomal RNA gene common to *Enterococcus* spp. (EPA, 2012), via qPCR. CrAssphage and enterococci were detected in all 24 samples. The absolute abundance of HFMs were significantly elevated in the UHI watershed samples (UHI > RLI) and a clear ranking was seen in samples impacted by wastewater (WWTP-

Impacted > Impacted > Pristine) (Wilcox, $p < 0.001$). Further HFM performance metrics can be found in the **SI Section S2-7**.

Of particular interest were ARGs successfully predicted by the HFMs versus the more generalized AARM indicators, *sul1* and *intI1*. All four markers correlated well (Pearson, $R^2 > 0.7$, $p < 0.0001$) with the absolute abundance of total ARGs ($\log_{10}$ Total ARG copies), with *sul1* performing the best ($R^2 = 0.901$) (Figure S2-8). Remarkably, a clear divergence was observed in terms of the individual ARGs that were successfully predicted by the HFMs versus the AARMs (i.e., Pearson $R^2 \geq 0.7$, $p < 0.0001$) (Figure 2-6). Of the 816 unique ARGs detected, 368 were successfully predicted by the AARMs, representing 93.4% of the total ARG abundance, whereas the HFMs predicted 140 ARGs, representing less than 2%. The AARMs successfully predicted > 85% of the abundance of the aminoglycoside, macrolide-lincosamide-streptogramin, peptide, fluoroquinolone, glycopeptide, and streptogramin ARGs and 100% of the quinolone, sulfonamide, and elfamycin. They also predicted 62/63 core ARGs (all except *AAC(6')-32*) and 94.7% of the Background resistome abundance. The HFMs correlated with less than 3% of the abundance of any individual gene class, except the β-lactams. The clinically-relevant β-lactams were more consistently predicted by the AARMs than the HFMs, although KPC-2 was correlated strongly with both HFMs (Table S2-9). There was not a single instance of an HFM and an AARM successfully predicting the same ARG, suggesting that HFMs and AARMs predict distinct and complimentary attributes of the resistome.

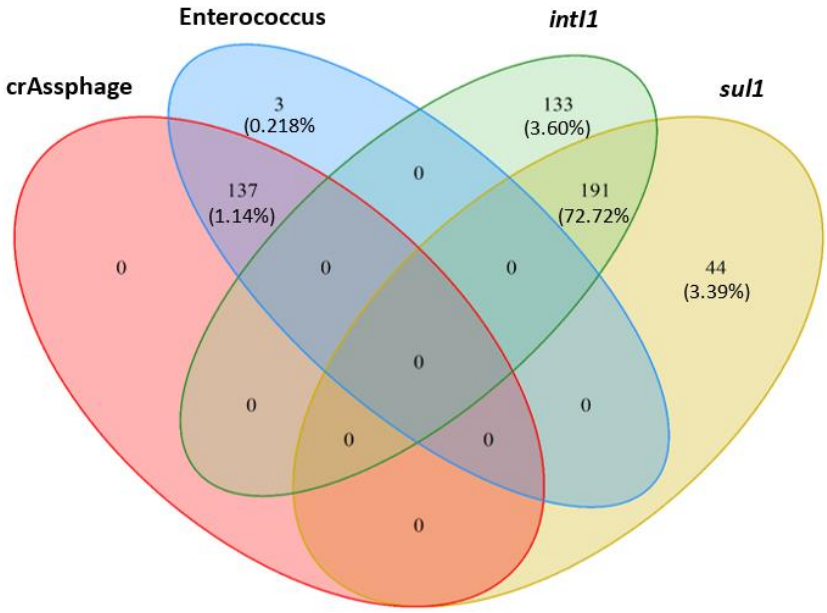


Figure 2-6: Venn Diagram of accurately predicted ARG types (measured by metagenomic sequencing) by each anthropogenic marker (measured by qPCR). Accurate predictions defined as cutoff for strong Pearson correlations, i.e., $R^2 > 0.7$ with p-values < 0.0001. Correlations were performed on absolute abundances of individual ARGs with the corresponding qPCR assay (excluding WWTP samples). Values in parentheses correspond to the abundance of the successfully predicted ARG types as a percent of the total ARGs detected by metagenomic sequencing for all non-WWTP samples.

An Integrated Surveillance Approach for ARGs in the Environment

Here we demonstrated a comprehensive ARG monitoring approach, integrating metagenomics with qPCR enumeration of HFMs and AARMs, by characterizing three distinct Puerto Rican watersheds in the wake of massive disruption by Hurricane Maria. Metagenomic-based approaches are particularly attractive, as they have the potential to capture the full resistome of a given environment, as well as indicators of ARG mobility (e.g., plasmids, integrons, transposons, and modulators of genetic mobility) and host organisms. Consistent with studies of other riverine environments (Koczura, Mokracka, Taraszewska, & Łopacinska, 2016; Rodriguez-Mozaz et al., 2015; Xiang, Chen, Zhu, An, & Yang, 2018), a clear ranking of ARG abundance and diversity metrics was observed as environments shifted in both land use characteristics and population density (Fresia et al., 2019; Y. Xu et al., 2016; Zheng et al., 2018).

Prior studies have reported ARGs that correlate well with anthropogenic activity in watersheds (He et al., 2014; Pruden et al., 2012; Storteboom, Arabi, Davis, Crimi, & Pruden, 2010). The present study employs metagenomics to expand this list of ARGs, while also combining with qPCR to further gain insight into their sources. Metagenomic approaches remain limited by cost and specialized expertise for analysis, with a lack of consensus on protocols (Hendriksen, Bortolaia, et al., 2019), while also facing inherent limitations in terms of high detection limits. Recently it has been pointed out that the majority of ARGs can simply be predicted by HFMs, especially the highly-specific crAssphage (Karkman et al., 2019). qPCR assays are attractive for monitoring of environmental sources of resistance, as they are much more broadly accessible, while also yielding sensitive quantitative information. This study takes a step towards validating such qPCR-based monitoring approaches, providing insight with respect to the specific ARGs predicted by crAssphage and other proposed monitoring targets and the extent of their clinical-relevance and mobility.

A striking finding of the present study is not only confirmation of a strong correlation between crAssphage and total ARG abundance (Karkman et al., 2019), but the fact that the AARMs, *sul*1 and *int*I1, exhibit stronger correlations and predict an entirely different and broader range of ARGs. The clinically-derived class 1 integron integrase gene, *int*I1, has been established as a highly sensitive stressor-responsive, xenogenetic, generic marker that has the ability to accumulate gene cassettes from the environment and integrate them into both commensal and pathogenic bacteria (Gillings et al., 2014). The prevalence of *int*I1 has been used as a direct bioindicator of these evolutionary pressures in riverine environments at the full-watershed scale (Barrón, Merlin, Guilloteau, Montargès-Pelletier, & Bellanger, 2018; Koczura et al., 2016; Lehmann et al., 2016). Still, accurately predicting which ARGs emanate from human fecal contamination versus other sources, and their relative proportions, remains a challenge (Li, Yin, & Zhang, 2018). In this study, many of the same environments were likely affected both by human fecal contamination and generalized anthropogenic pollution. In such situations, we expect that the AARMs reflect, in part, enrichment of ARGs represented in the Core resistome, by selection or gene exchange, while ARGs sourced directly from human fecal-related taxa would be more subject to attenuation in the environment. This is consistent with the observation in this study that ARGs that correlated strongly with HFMs, although clinically-relevant, made up a small fraction of the total resistome.

β-lactam and aminoglycoside ARGs were disproportionately diverse, co-located with indicators of mobility, and associated with increased anthropogenic stress in these sub-tropical catchments, suggesting that they are largely allochthonous. Studies have shown highly diverse β-

lactam ARGs in aquatic environments (Bai, Jing, Teng, Chen, & Chen, 2018; Zhang et al., 2019), but these prior studies of arguably more polluted riverine environments in China and India did not detect the extent of the diversity of 3rd-Generation cephalosporinases or carbapenamases that were observed in the present study. Interestingly, compared to mainland United States and Canada in recent decades, there has been an observed increase in β-lactam resistant *Enterobacteriaceae, A. baumanii,* and *P. aeruginosa* nosocomial infections in Puerto Rico which also displayed increased levels of aminoglycoside and quinolone resistance (Doern, Jones, Pfaller, Erwin, & Ramirez-rhonda, 1998; Vazquez et al., 2003). The clear clustering of these enzymatic signatures surrounding urban developments and WWTPs, compounded with the observed increase in co-occurrence of these genes with indicators of mobility, highlights the need for increased surveillance approaches appropriate for these clinically-relevant ARGs.

Future surveillance efforts targeting antimicrobial resistance in watersheds would ideally be centered around inputs from WWTPs and other sources of human, as well as livestock/animal, fecal bacteria introduced to aquatic environments. It is also worth considering industrial inputs, e.g., Puerto Rico has a substantial drug production industry. Surveillance efforts can be implemented using lower cost and more technologically-accessible approaches, such as culturing of extended-spectrum beta-lactamase (ESBL)-producing *E. coli* (Kirby, 2020) and qPCR assays for select ARGs, ideal candidates being 3rd-Generation cephalosporinases and carbapenamase-producing genes/bacteria already prominent in the healthcare system (e.g. KPC-2 and IMP variants). These methods, alongside routine screening of indicators, such as crAssphage and *intI1,* would aid in identifying input sources, assessing potential for ARGs to amplify/mobilize, and establishing baselines for assessing trends with time and effects of disruptions, including storms or intentional interventions intended to mitigate the spread of antibiotic resistance. Recently, it has been suggested that metagenomic sequencing of urban sewage could be an informative tool for assessing the resistome composition of local populations. The authors found that the resistomes of individual sewage samples were mainly influenced by local factors and tended to be representative of broader patterns observed across a given country (Hendriksen, Munk, et al., 2019). Here, a profile of the local resistome was revealed via shotgun metagenomics of the Patillas WWTP, but it would be of interest to compare to the more densely populated San Juan metropolitan areas as well.

In an era of increasing intensity and frequency of major storms (Levin, 2019; Rahmani et al., 2016; Roque-malo & Kumar, 2017), surveillance can also help address concerns that such events may exacerbate the spread of ARB and ARGs. WWTPs are of particular interest, as they serve as central nodes for the processing of human waste, acting as both a barrier and a source for ARGs and ARB entering the aquatic environment (Bürgmann et al., 2018). During storms, loss of WWTP functionality or discharge of untreated wastewater when WWTPs become inundated could elevate their potential role in disseminating antibiotic resistance and increasing risk of infection in human populations. The downstream environment of the WWTPs that were the subject of this study certainly presented concerns with respect to the clinically-relevant ARGs detected. Thus, as surveillance methodologies and the information yielded are established, future studies are merited to more intentionally assess the impacts of major storms on the spread of antibiotic resistance.

ACKNOWLEDGEMENTS

This study was supported by NSF RAPID (PI Bandaragoda) Award 1810886 and NSF PIRE (PI Vikesland) Award 1545756. The authors acknowledge the Advanced Research Computing at

Virginia Tech for providing computational resources and technical support as well as NSF NNCI Award 1542100 for equipment/facilities, CECIA- Inter American University of Puerto Rico and the Gabriella and Paul Rosenbaum Foundation for their support with the field activities and sampling. This work was also partially supported by the United States Department of Agriculture National Institute of Food and Agriculture Award 2017-68003-26498. The authors also thank other members of the sampling team: Matthew Blair, Benjamin Davis, William Rhoads, Gregory House, Melitza Crespo Medina, Taína Rodriguez Curet, Yira Rodríguez, Antonio Gonzalez, and Javier Chévere.

REFERENCES

Aarestrup, F. M., & Woolhouse, M. E. J. (2020). Using sewage for surveillance of antimicrobial resistance. *Science*, *367*(6478), 630–632. https://doi.org/10.1126/science.aba3432

Amos, G. C. A., Zhang, L., Hawkey, P. M., Gaze, W. H., & Wellington, E. M. (2014). Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. *Veterinary Microbiology*, *171*(3–4), 441–447. https://doi.org/10.1016/j.vetmic.2014.02.017

Arango-Argoty, G. A., Dai, D., Pruden, A., Vikesland, P., Heath, L. S., & Zhang, L. (2019). NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. *Microbiome*, *7*(1), 88. https://doi.org/10.1186/s40168-019-0703-9

Arango-Argoty, G., Singh, G., Heath, L. S., Pruden, A., Xiao, W., & Zhang, L. (2016). MetaStorm: A public resource for customizable metagenomics annotation. *PLoS ONE*, *11*(9), 1–13. https://doi.org/10.1371/journal.pone.0162442

Ayling, M., Clark, M. D., & Leggett, R. M. (2020). New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*, *21*(2), 584–594. https://doi.org/10.1093/bib/bbz020

Bai, X., Jing, L., Teng, Y., Chen, R., & Chen, H. (2018). Characterization of antibiotic resistance genes in the sediments of an urban river revealed by comparative metagenomics analysis. *Science of The Total Environment*, *653*, 1513–1521. https://doi.org/10.1016/j.scitotenv.2018.11.052

Barrón, M. D. la C., Merlin, C., Guilloteau, H., Montargès-Pelletier, E., & Bellanger, X. (2018). Suspended materials in river waters differentially enrich class 1 integron- and IncP-1 plasmid-carrying bacteria in sediments. *Frontiers in Microbiology*, *9*(JUL), 1–10. https://doi.org/10.3389/fmicb.2018.01443

Belder, D. De, Lucero, C., Rapoport, M., Rosato, A., Faccone, D., Petroni, A., … Gomez, S. A. (2017). Genetic Diversity of KPC-Producing Escherichia coli, Klebsiella oxytoca, Serratia marcescenes, and Citrobacter freundii isolates from Argentina. *Microbial Drug Resistance*, *00*(00). https://doi.org/10.1089/mdr.2017.0213

Bengtsson-palme, J., Kristiansson, E., & Larsson, D. G. J. (2018). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiology Reviews*, (July 2017), 68–80. https://doi.org/10.1093/femsre/fux053

Berendonk, T. U., Manaia, C. M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., …

Martinez, J. L. (2015). Tackling antibiotic resistance: The environmental framework. *Nature Reviews Microbiology*, *13*(5), 310–317. https://doi.org/10.1038/nrmicro3439

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bürgmann, H., Frigon, D., Gaze, W. H., Manaia, C. M., Pruden, A., Singer, A. C., … Zhang, T. (2018). Water and sanitation: An essential battlefront in the war on antimicrobial resistance. *FEMS Microbiology Ecology*, *94*(9), 1–14. https://doi.org/10.1093/femsec/fiy101

Crofts, T. S., Gasparrini, A. J., & Dantas, G. (2017). Next-generation approaches to understand and combat the antibiotic resistome. *Nature Reviews Microbiology*, *15*(7), 422–434. https://doi.org/10.1038/nrmicro.2017.28

Doern, G. V, Jones, R. N., Pfaller, M. A., Erwin, M., & Ramirez-rhonda, C. (1998). Multicenter Evaluation of the In Vitro Activity of Six Broad-Spectrum ␤ -Lactam Antimicrobial Agents in Puerto Rico The Puerto Rico Antimicrobial Resistance Study Group ,. *Diagn Microbiol Infect Dis*, *8893*(97), 113–119.

EPA. (2012). *Method 1611 : Enterococci in Water by TaqMan ® Quantitative Polymerase Chain Reaction (qPCR) Assay*.

EPA. (2017). EPA Hurricane Maria Update for Wednesday, October 4, 2017. Retrieved from https://www.epa.gov/newsreleases/epa-hurricane-maria-update-wednesday-october-4-2017

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

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

Fresia, P., Antelo, V., Salazar, C., Giménez, M., D'Alessandro, B., Afshinnekoo, E., … Iraola, G. (2019). Urban metagenomics uncover antibiotic resistance reservoirs in coastal beach and sewage waters. *Microbiome*, 35. https://doi.org/10.1186/s40168-019-0648-z

Garland, J. L., Ph, D., Brinkman, N., Ph, D., Jahne, M., & Ph, D. (2019). Geospatial Distribution of Antimicrobial Resistance Genes in US Rivers and Streams United States Environmental Protection Agency.

Garner, E., Chen, C., Xia, K., Bowers, J., Engelthaler, D. M., Mclain, J., … Pruden, A. (2018). Metagenomic Characterization of Antibiotic Resistance Genes in Full- Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. *Environmental Science & Technology*, *52*, 6113–6125. research-article. https://doi.org/10.1021/acs.est.7b05419

Ghurye, J. S., Cepeda-Espinoza, V., & Pop, M. (2016). Metagenomic assembly: Overview, challenges and applications. *Yale Journal of Biology and Medicine*, *89*(3), 353–362.

Gillings, M. R. (2014). Integrons : Past , Present , and Future Structure of Integrons, *78*(2), 257–277. https://doi.org/10.1128/MMBR.00056-13

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M., & Ku, J. (2014). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *The ISME Journal*, *9*(6), 1269–1279. https://doi.org/10.1038/ismej.2014.226

Gregory, C. J., Llata, E., Stine, N., Gould, C., Santiago, L. M., Vazquez, G. J., … Tomashek, K. M. (2010). Outbreak of Carbapenem-Resistant Klebsiella pneumoniae in Puerto Rico Associated with a Novel Carbapenemase Variant. *Infection Control & Hospital Epidemiology*, *31*(05), 476–484. https://doi.org/10.1086/651670

Hardwick, S. A., Stokes, H. W., Findlay, S., Taylor, M., & Gillings, M. R. (2008). Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. *FEMS Microbiology Letters*, *278*(2), 207–212. https://doi.org/10.1111/j.1574-6968.2007.00992.x

He, L. Y., Liu, Y. S., Su, H. C., Zhao, J. L., Liu, S. S., Chen, J., … Ying, G. G. (2014). Dissemination of antibiotic resistance genes in representative broiler feedlots environments: Identification of indicator ARGs and correlations with environmental variables. *Environmental Science and Technology*, *48*(22), 13120–13129. https://doi.org/10.1021/es5041267

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., & Mcdermott, P. F. (2019). Using Genomics to Track Global Antimicrobial Resistance. *Frontiers in Public Health*, *7*(September). https://doi.org/10.3389/fpubh.2019.00242

Hendriksen, R. S., Munk, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., … Aarestrup, F. M. (2019). Global monitoring of antimicrobial resistance based on human sewage. *Nature Communications*. https://doi.org/10.1038/s41467-019-08853-3

Huijbers, P. M. C., Flach, C.-F., & Larsson, D. G. J. (2019). A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environment International*, *130*(May), 104880. https://doi.org/10.1016/j.envint.2019.05.074

Hutinel, M., Huijbers, P. M. C., Fick, J., Åhrén, C., Larsson, D. G. J., & Flach, C. F. (2019). Population-level surveillance of antibiotic resistance in Escherichia coli through sewage analysis. *Eurosurveillance*, *24*(37), 1–11. https://doi.org/10.2807/1560-7917.ES.2019.24.37.1800497

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C. S., Singer, H. P., … Bürgmann, H. (2019). Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME Journal*, *13*(2), 346–360. https://doi.org/10.1038/s41396-018-0277-8

Kapoor, V., Gupta, I., Pasha, A. B. M. T., & Phan, D. (2018). Real-Time Quantitative PCR Measurements of Fecal Indicator Bacteria and Human-Associated Source Tracking Markers in a Texas River following Hurricane Harvey, 2–8. https://doi.org/10.1021/acs.estlett.8b00237

Karkman, A., Pärnänen, K., & Larsson, D. G. J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nature Communications*, (2019), 1–8. https://doi.org/10.1038/s41467-018-07992-3

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and accurate classification of metagenomic sequences. *BioRxiv*, *26*(12), 054965. https://doi.org/10.1101/054965

Kirby, A. (2020). Antibiotic Resistance Monitoring Using Extended Spectrum Beta-Lactamase-Producing E. coli as an Indicator Organism. In *AWWA International Symposium on Potable Reuse*. Atalanta, GA.

Knapp, C. W., Lima, L., Olivares-Rieumont, S., Bowen, E., Werner, D., & Graham, D. W. (2012). Seasonal variations in antibiotic resistance gene transport in the Almendares River, Havana, Cuba. *Frontiers in Microbiology*, *3*(NOV), 1–11. https://doi.org/10.3389/fmicb.2012.00396

Koczura, R., Mokracka, J., Taraszewska, A., & Łopacinska, N. (2016). Abundance of Class 1 Integron-Integrase and Sulfonamide Resistance Genes in River Water and Sediment Is Affected by Anthropogenic Pressure and Environmental Factors. *Microbial Ecology*, *72*(4), 909–916. https://doi.org/10.1007/s00248-016-0843-4

Larsen, N., Brodie, E. L., Andersen, G. L., Huber, T., DeSantis, T. Z., Keller, K., … Hu, P. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/aem.03006-05

Lehmann, K., Bell, T., Bowes, M. J., Amos, G. C. A., Gaze, W. H., Wellington, E. M. H., & Singer, A. C. (2016). Trace levels of sewage effluent are sufficient to increase class 1 integron prevalence in freshwater biofilms without changing the core community. *Water Research*, *106*, 163–170. https://doi.org/10.1016/j.watres.2016.09.035

Leplae, R., Lima-Mendez, G., & Toussaint, A. (2009). ACLAME: A CLAssification of mobile genetic elements, update 2010. *Nucleic Acids Research*, *38*(SUPPL.1), 57–61. https://doi.org/10.1093/nar/gkp938

Levin, E. L. (2019). Impact of Anthropogenic Climate Change on United States Major Hurricane Landfall Frequency, (October 2005).

Li, L., Yin, X., & Zhang, T. (2018). Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. *Microbiome*, 1–12.

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

Marathe, N. P., Pal, C., Gaikwad, S. S., Jonsson, V., Kristiansson, E., & Larsson, D. G. J. (2017). Untreated urban waste contaminates Indian river sediments with resistance genes to last resort antibiotics. *Water Research*, *124*, 388–397. https://doi.org/10.1016/j.watres.2017.07.060

Marti, E., Variatza, E., & Balcazar, J. L. (2014). The role of aquatic ecosystems as reservoirs of antibiotic resistance. *Trends in Microbiology*, *22*(1), 36–41. https://doi.org/10.1016/j.tim.2013.11.001

Martinez, J., Coque, T., & Baquero, F. (2015). What is a resistance gene? Ranking risk in

resistomes. *Nature Reviews Microbiology*.

Martínez, J. L., Coque, T. M., & Baquero, F. (2015). What is a resistance gene? Ranking risk in resistomes. *Nature Reviews Microbiology*, *13*(2), 116–123. https://doi.org/10.1038/nrmicro3399

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., … Wright, G. D. (2013). The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3348–3357. https://doi.org/10.1128/aac.00419-13

Nappier, S. P., Ichida, A., Jaglo, K., Haugland, R., & Jones, K. R. (2019). Advancements in mitigating interference in quantitative polymerase chain reaction (qPCR) for microbial water quality monitoring. *Science of the Total Environment*, *671*, 732–740. https://doi.org/10.1016/j.scitotenv.2019.03.242

Oh, M., Pruden, A., Chen, C., Heath, L. S., Xia, K., & Zhang, L. (2018). MetaCompare : a computational pipeline for prioritizing environmental resistome risk. *FEMS Microbiology Ecology*, (April), 1–9. https://doi.org/10.1093/femsec/fiy079

Oksanen, A. J., Kindt, R., Legendre, P., Hara, B. O., Simpson, G. L., Stevens, M. H. H., & Wagner, H. (2009). The vegan Package, (May 2014).

Pei, R., Kim, S.-C., Carlson, K. H., & Pruden, A. (2006). Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Research*, *40*(12), 2427–2435. https://doi.org/10.1016/j.watres.2006.04.017

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Pruden, A., Alcalde, R. E., Alvarez, P. J. J., Ashbolt, N., Bischel, H., Capiro, N. L., … Zhou, Z. (2018). An Environmental Science and Engineering Framework for Combating Antimicrobial Resistance. *Environmental Engineering Science*, *35*(10), 1005–1011. https://doi.org/10.1089/ees.2017.0520

Pruden, A., Arabi, M., & Storteboom, H. N. (2012). Correlation between upstream human activities and riverine antibiotic resistance genes. *Environmental Science and Technology*, *46*(21), 11541–11549. https://doi.org/10.1021/es302657r

Qiao, M., Ying, G. G., Singer, A. C., & Zhu, Y. G. (2018). Review of antibiotic resistance in China and its environment. *Environment International*, *110*(October 2017), 160–172. https://doi.org/10.1016/j.envint.2017.10.016

Rahmani, V., Hutchinson, S. L., Jr, A. H., Hutchinson, J. M. S., Scott, F., & Francis, S. (2016). Analysis of frequency and magnitude of extreme rainfall events with potential impacts on flooding : a case study from the central United States, *3587*(December 2015), 3578–3587. https://doi.org/10.1002/joc.4577

Robledo, I. E., Aquino, E. E., Santé, M. I., Santana, J. L., Otero, D. M., León, C. F., & Vázquez, G. J. (2010). Detection of KPC in Acinetobacter spp. in Puerto Rico. *Antimicrobial Agents and Chemotherapy*, *54*(3), 1354–1357. https://doi.org/10.1128/AAC.00899-09

Robledo, I. E., Aquino, E. E., & Vázquez, G. J. (2011). Detection of the KPC gene in Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, and Acinetobacter baumannii during a PCR-based nosocomial surveillance study in Puerto Rico. *Antimicrobial Agents and Chemotherapy*, *55*(6), 2968–2970. https://doi.org/10.1128/AAC.01633-10

Rodriguez-Mozaz, S., Chamorro, S., Marti, E., Huerta, B., Gros, M., anchez-Melsi, A. S., … Luis Balc azar, J. (2015). Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water Research*, *69*, 234–242. https://doi.org/10.1016/j.watres.2014.11.021

Roque-malo, S., & Kumar, P. (2017). Patterns of change in high frequency precipitation variability over North America. *Scientific Reports*, (August), 1–12. https://doi.org/10.1038/s41598-017-10827-8

Schwab, K. J., Gibson, K. E., Williams, D. L., Kulbicki, K. M., Lo, C. P., Mihalic, J. N., … Geyh, A. S. (2007). Microbial and chemical assessment of regions within New Orleans, LA impacted by Hurricane Katrina. *Environmental Science and Technology*, *41*(7), 2401–2406. https://doi.org/10.1021/es062916x

Sinigalliano, C. D., Gidley, M. L., Shibata, T., Whitman, D., Dixon, T. H., Laws, E., … Solo-Gabriele, H. M. (2007). Impacts of Hurricanes Katrina and Rita on the microbial landscape of the New Orleans area. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 9029–9034. https://doi.org/10.1073/pnas.0610552104

Stachler, E., Akyon, B., Carvalho, N. A. De, Ference, C., & Bibby, K. (2018). Correlation of crAssphage qPCR Markers with Culturable and Molecular Indicators of Human Fecal Pollution in an Impacted Urban Watershed. *Environmental Science & Technology*, *52*, 7505–7512. research-article. https://doi.org/10.1021/acs.est.8b00638

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., & Shanks, O. C. (2017). Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environ. Sci. Technol*. https://doi.org/10.1021/acs.est.7b02703

Storteboom, H., Arabi, M., Davis, J. G., Crimi, B., & Pruden, A. (2010). Tracking antibiotic resistance genes in the south platte river basin using molecular signatures of urban, agricultural, and pristine sources. *Environmental Science and Technology*, *44*(19), 7397–7404. https://doi.org/10.1021/es101657s

Suzuki, M. T., & Taylor, L. T. (2000). Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5 J -Nuclease Assays Downloaded from http://aem.asm.org/ on November 26 , 2014 by SERIALS CONTROL Lane Medical Library. *Applied and Environmental Microbiology*, *66*(11), 4605–4614. https://doi.org/10.1128/AEM.66.11.4605-4614.2000.Updated

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., … Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, *18*(3), 318–327. https://doi.org/10.1016/S1473-3099(17)30753-3

ten Veldhuis, J. A. E., Clemens, F. H. L. R., Sterk, G., & Berends, B. R. (2010). Microbial risks associated with exposure to pathogens in contaminated urban flood water. *Water Research*,

44(9), 2910–2918. https://doi.org/10.1016/j.watres.2010.02.009

United Nations. (2017). *Frontiers 2017 - Emerging Issues Of Environmental Concern*. https://doi.org/10.1016/S0140-6736(15)00298-6

Vazquez, G. J., Robledo, I. E., Arroyo, A., Nadal, E., Rodriguez, R., Bermudez, M., & Colon, M. (2003). A Comparison of the Antimicrobial Resistance Patterns of Gram-Negative Bacilli Isolated from Community-Private and University-Affiliated Hospitals from Puerto Rico. *Puerto Rico Health Sciences Journal*, *22*(3).

WHO. (2017). Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics, *43*(148), 348–365. https://doi.org/10.1590/S0100-15742013000100018

Wolter, D. J., Khalaf, N., Robledo, I. E., Vázquez, J. G., Santé, I. M., Aquino, E. E., … Hanson, N. D. (2009). Surveillance of carbapenem-resistant pseudomonas aeruginosa isolates from Puerto Rican Medical Center Hospitals: Dissemination of KPC and IMP-18 β-lactamases. *Antimicrobial Agents and Chemotherapy*, *53*(4), 1660–1664. https://doi.org/10.1128/AAC.01172-08

World Health Organization. (2015). *Global Action Plan on Antimicrobial Resistance*.

Xiang, Q., Chen, Q., Zhu, D., An, X., & Yang, X. (2018). Spatial and temporal distribution of antibiotic resistomes in a peri-urban area is associated significantly with anthropogenic activities Spatial and temporal distribution of antibiotic resistomes in a peri- urban area is associated signi fi cantly with a. *Environmental Pollution*, (January). https://doi.org/10.1016/j.envpol.2017.12.119

Xu, J., Xu, Y., Wang, H., Guo, C., Qiu, H., He, Y., … Meng, W. (2015). Occurrence of antibiotics and antibiotic resistance genes in a sewage treatment plant and its effluent-receiving river. *Chemosphere*, *119*, 1379–1385. https://doi.org/10.1016/j.chemosphere.2014.02.040

Xu, Y., Guo, C., Luo, Y., Lv, J., Zhang, Y., Lin, H., … Xu, J. (2016). Occurrence and distribution of antibiotics, antibiotic resistance genes in the urban rivers in Beijing, China. *Environmental Pollution*, *213*, 833–840. https://doi.org/10.1016/j.envpol.2016.03.054

Yu, P., Zaleski, A., Li, Q., He, Y., Mapili, K., Pruden, A., … Stadler, L. B. (2018). Elevated Levels of Pathogenic Indicator Bacteria and Antibiotic Resistance Genes after Hurricane Harvey's Flooding in Houston. *Environmental Science and Technology Letters*, *5*(8), 481–486. https://doi.org/10.1021/acs.estlett.8b00329

Zhang, S. Y., Tsementzi, D., Hatt, J. K., Bivins, A., Khelurkar, N., Brown, J., … Konstantinidis, K. T. (2019). Intensive allochthonous inputs along the Ganges River and their effect on microbial community composition and dynamics. *Environmental Microbiology*, *21*(1), 182–196. https://doi.org/10.1111/1462-2920.14439

Zheng, J., Zhou, Z., Wei, Y., Chen, T., Feng, W., & Chen, H. (2018). High-throughput profiling of seasonal variations of antibiotic resistance gene transport in a peri-urban river. *Environment International*, *114*(November 2017), 87–94. https://doi.org/10.1016/j.envint.2018.02.039

Zhu, Y., Zhao, Y., Li, B., Huang, C., Zhang, S., Yu, S., … Su, J. (2017). Continental-scale pollution of estuaries with antibiotic resistance genes. *Nature Microbiology*, (March). https://doi.org/10.1038/nmicrobiol.2016.270

SUPPLEMENTAL MATERIALS FOR CHAPTER 2

Section S2-1: Metagenomic Sequence Quality and Detection of ARGs

Shotgun metagenomic sequencing of 24 samples across the three differentially-impacted watersheds and one WWTP produced 952 million paired-end short reads representing 238 Gb, averaging 39.6 million reads (9.9 Gb) per metagenome. A total of 320,528 metagenomic short-reads were aligned to the CARD database, representing 816 unique ARGs belonging to 18 different antibiotic classes. This constituted 0.034% of the complete short-read dataset, whereas the MGE and plasmid-protein sequences were more abundant, representing 0.47% and 1.39%, respectively. Very strong correlations were observed between sul1 and intI1 annotated in metagenomic libraries versus measured via qPCR (sul1 Pearson R2 = 0.951, p = 9.5e-13; intI1 R2 = 0.921, p = 1.8e-10) (Figure S2-3). Along with the uniform sample collection and processing applied here, this result provided validation for subsequent semi-quantitative and quantitative comparisons applied across metagenomic data sets. An overview of percent relative and absolute abundances of ARG classes detected across watersheds is provided in Table S2-5. All qPCR targets measured below detection in trip, DNA extraction, and no template control blanks.

Section S2-2: Metagenomic Assembly Quality

Nonpareil coverage estimates[1], a measurement of sequencing redundancy in a sample and an indication of biological sequencing depth, were dependent on sample type and location. For example, more polluted samples had less inherent sequence diversity and therefore had higher sequencing coverage. This discrepancy in sequencing coverage was reflected in overall assembly qualities, with short-read assembly percentages ranging from less than 1% in pristine river water to 64% in raw sewage (Table S2-3). Still, de novo assembly of the metagenomic short-reads generated 9,162,997 contigs with N50 length of 915 bp containing 13,378,273 open-reading frames, allowing for broad ARG contextualization. Of the total library, 225,342 contigs were identified as carrying an ARG (2.46%). To account for inherent limitations of metagenomic sequence assembly, the MetaCompare algorithm[2] used to assess relative resistome risk is normalized to total contig library size, facilitating comparison across assembly qualities.

Section S2-3: Microbial Community Analysis

Microbiomes (i.e., taxonomic microbial community composition derived from metagenomic sequencing) were profiled, with microbial classification performed using Centrifuge[3] with a preconstructed index of all bacterial and archaeal genomes (4/15/2018 update) contained within the RefSeq database (http://www.ccb.jhu.edu/software/centrifuge/). The engine was run with the default parameters using both the forward and reverse, quality trimmed4 reads from each sample. Classified reads were filtered for Centrifuge Scores ≥ 300 (hit length > 31) and the engine's calculated abundances were used for all downstream analyses. Procrustes analysis of NMDS ordinations of the microbiomes annotated using Centrifuge indicated strong association with the resistome (Bray-Curtis dissimilarity index; sum of squares M12 = 0.4372, R = 0.751, $p < 0.001$, 999 permutations). This observation was further corroborated with the Mantel test (R = 0.4588, $p < 0.001$). While these statistical tests confirm the expectation that there is a relationship between the microbiome and resistome, lack of a perfect relationship (i.e., R=1.0) is consistent with the assumption that at least some portion of the resistome is mobile across bacterial species.

Section S2-4: Influence of Land-Use and WWTP Inputs Across Watersheds

In general, in moving from less impacted to more impacted sites, the proportion of Proteobacteria was reduced from as high as 91% in pristine headwaters (RLI_Wat_WP), down to 23% in a UHI sample (UHI_Wat_MID). The relative abundance of Proteobacteria was diminished and replaced by the phyla Actinobacteria, Bacteroidetes, Firmicutes, and Nitrospirae in these more impacted environments (Figure S2-7). In every instance, sediments directly downstream of WWTP discharge showed high proportions of Nitrospirae, as high as 31.8% of the relative abundance downstream of the 40 MGD WWTP (UHI_Sed_DN_2), which is likely due to their emission from nitrification processes in the WWTP. Chemical analysis showed that the concentration of nitrite was significantly enriched downstream of these plants (Wilcox < 0.05) (Table S11). The shift in microbial communities is also reflected by NMDS, where the location within each respective watershed (e.g., pristine, residential, upstream and downstream of effluent discharge) was associated with an altered microbiome (ANOSIM, $R2=0.206$, $p<0.05$) (Figure 2-3B). Microbial communities across each watershed were also distinctly separated (ANOSIM, $R2=0.254$, $p<0.001$), indicating a microbiome that is sensitive to watershed-specific, non-point source effects as well.

Section S2-5: Environmental Indicators of Pollution also Correlate with ARGs

According to vegan 'envfit' analysis, the total ARG absolute abundances were further correlated with total and soluble phosphorus concentrations, nitrite, and temperature variations. Notably, the average measured temperature was 22.0°C for pristine samples versus 27.7°C for urban samples and samples downstream of WWTP discharges (Table S2-11).

Section S2-6: Sediment Resistome Harbored Much Greater Abundance but Lower Gene Richness and Diversity Compared to Bulk Water

Absolute ARG concentrations in the sediments were high, with an average of 4.91 x 108 copies per gram compared to raw influent/treated effluent and river water with average abundances of 5.45 x 107 and 3.45 x 105 copies per mL, respectively. When comparing sediment samples directly upstream (n=3) and downstream (n=3) of the three WWTP discharge points, there was an increase in absolute abundance of all 18 detected ARG classes (Table S2-8). The Shannon diversity also increased across 12/18 antibiotic classes in downstream sediments. Comparing diversity metrics from wastewater impacted bulk water samples (Table S2-7) to downstream sediment samples (Table S2-8) though, there was a marked decrease in the number of detected ARGs for the aminoglycoside, beta-lactam, fluoroquinolone, fosfomycin, glycopeptide, MLS, multidrug, other, phenicol, tetracycline, and trimethoprim classes.

Section S2-7: Performance of Human Fecal Markers

All three watersheds had little to no agricultural activity, although it was common for the rural households in the RLI and PMI areas to maintain small to medium sized livestock including chickens, ducks, and goats. These rural households were also primarily on individual septic systems and in some cases, informal waste detention mechanisms, e.g., cesspools or direct discharge onto land or streams[5]. These factors likely contributed to the detection of both crAssphage and enterococci in every sample, including pristine samples with no direct human influence, albeit at levels below quantification. Detection of both crAssphage and enterococci was confirmed via Sanger sequencing in samples below the limit of quantification (10

copies/mL). The wastewater treatment plant sampled saw a 4-log reduction of crAssphage and a 6-log reduction of enterococcus markers from influent to final effluent. Sediment samples harbored much greater abundances of both markers per gram (4.16 – 5.2 log10) compared to the water column per milliliter (1.76 – 1.97 log10). The detection of human feces was corroborated via cross correlation of each marker where remarkably, the two targets exhibited a near perfect correlation (Pearson R = 0.9987, p < 1e-16). We further note that the original environmental crAssphage studies[6,7] employed a 0.45 um filter to quantify the human fecal bacteriophage whereas the present study used a 0.22 um filter, which would be expected to provide even greater viral capture efficiency.

References:

1. Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems 3, 1–9 (2018).

2. Oh, M. et al. MetaCompare : a computational pipeline for prioritizing environmental resistome risk. FEMS Microbiol. Ecol. 1–9 (2018). doi:10.1093/femsec/fiy079

3. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and accurate classification of metagenomic sequences. bioRxiv 26, 054965 (2016).

4. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

5. Sin, A. P. Impacto Ambiental de Pozos Sépticos en Puerto Rico y su Diseño y Control. 1, (2011).

6. Stachler, E., Akyon, B., Carvalho, N. A. De, Ference, C. & Bibby, K. Correlation of crAssphage qPCR Markers with Culturable and Molecular Indicators of Human Fecal Pollution in an Impacted Urban Watershed. Environ. Sci. Technol. 52, 7505–7512 (2018).

7. Stachler, E., Crank, K. & Bibby, K. Co-Occurrence of crAssphage with Antibiotic Resistance Genes in an Impacted Urban Watershed. Environ. Sci. Technol. Lett. 6, 216–221 (2019).

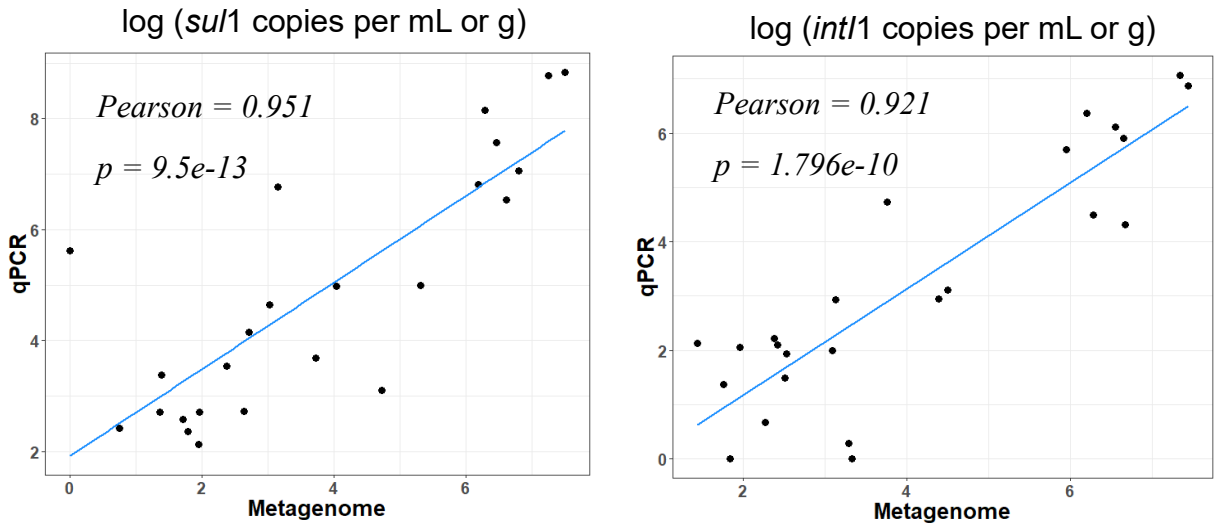


Figure S2-1: Comparison of quantitative metagenomics and qPCR for quantifying both the sul1 and intI1 genes across all samples. Absolute abundances in the metagenomes were calculated by multiplying the relative abundances of the functional genes by total abundance of 16S rRNA quantified by qPCR.

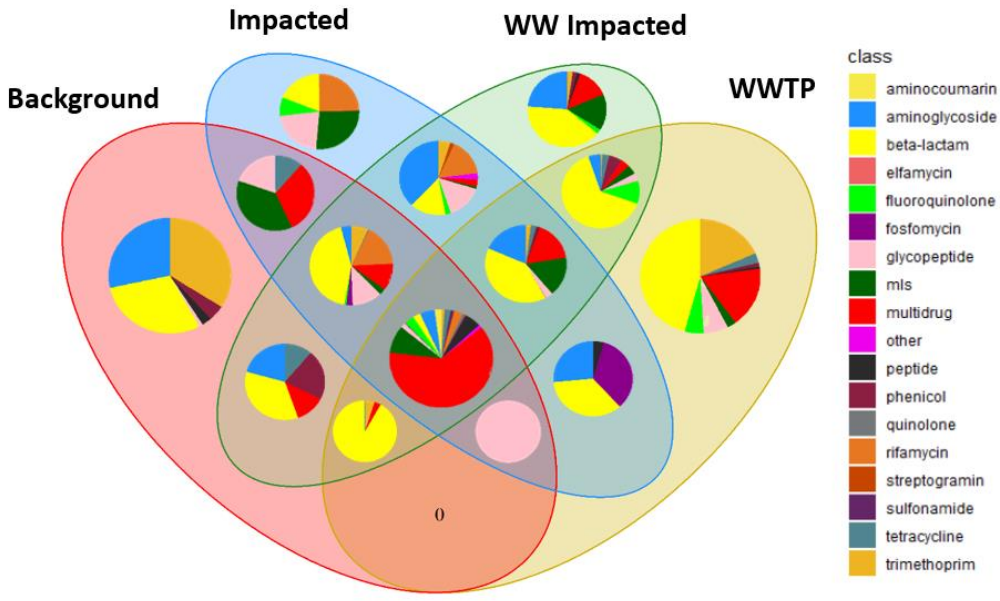


Figure S2-2: Resistome composition by Venn diagram compartment. Pie charts were generated using the absolute abundance of detected ARGs by class located in their respective compartments.

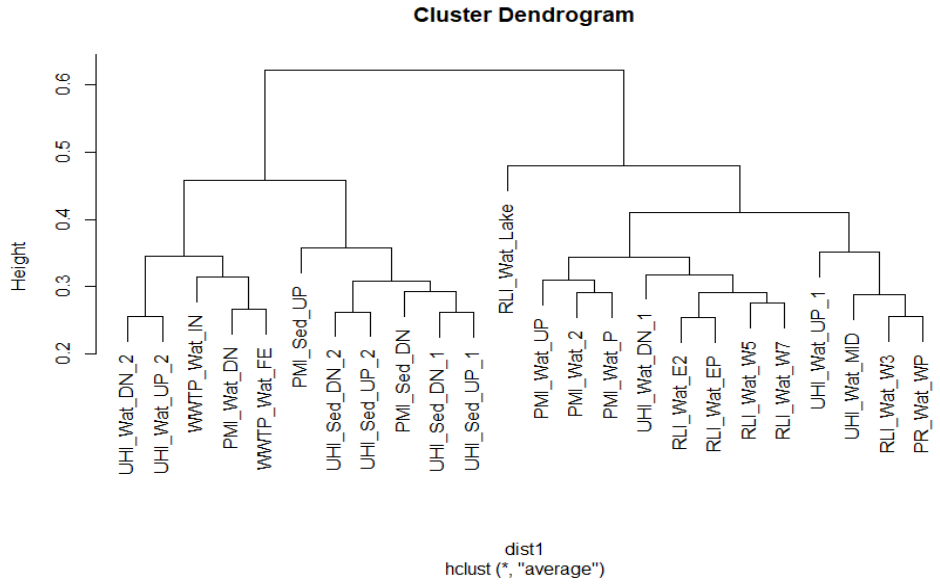


Figure S2-3: Hierarchical clustering dendrogram of absolute abundance of ARGs across all samples by Bray-Curtis dissimilarity matrix. Generated using the 'vegan' package in R.

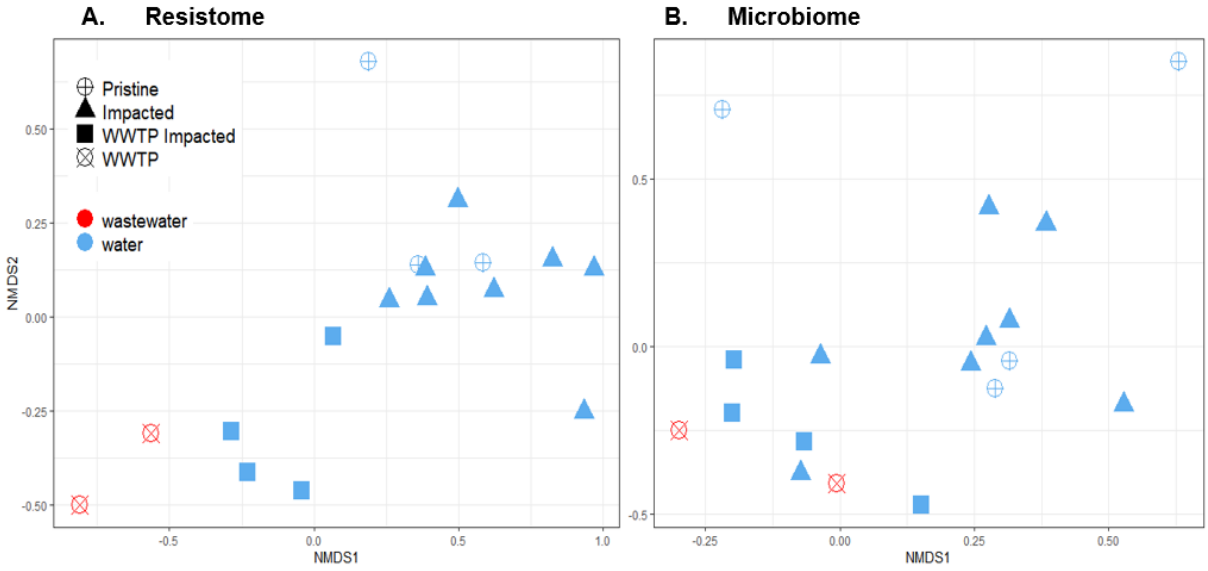


Figure S2-4: (A) Non-metric Multidimensional Scaling (NMDS) plot of calculated absolute abundances of individual ARGs detected via alignment of metagenomic reads to CARD (v2.0.1). (B) NMDS plot generated from Bray-Curtis dissimilarity matrix of Centrifuge's calculated taxonomic relative abundances at the genus level across all water and wastewater samples (n=18). "Pristine" are those samples furthest upstream in each catchment, "Impacted" ARGs are those from sites influenced by adjacent land-use, "WWTP Impacted" are those from any sites where their bulk water consists of some portion of wastewater effluent, and WWTP are those from the influent and effluent of the 1.0 MGD WWTP in the PMI watershed. (ANOSIM) Microbiome by Location (e.g., pristine, residential, upstream and downstream of effluent

discharge), R2=0.206, p<0.05; Microbiome by Watershed, R2=0.254, p<0.001; Resistome by Watershed, R2 0.1963, p < 0.05; Resistome by WWTP Impact, R2 0.2465, p < 0.001.

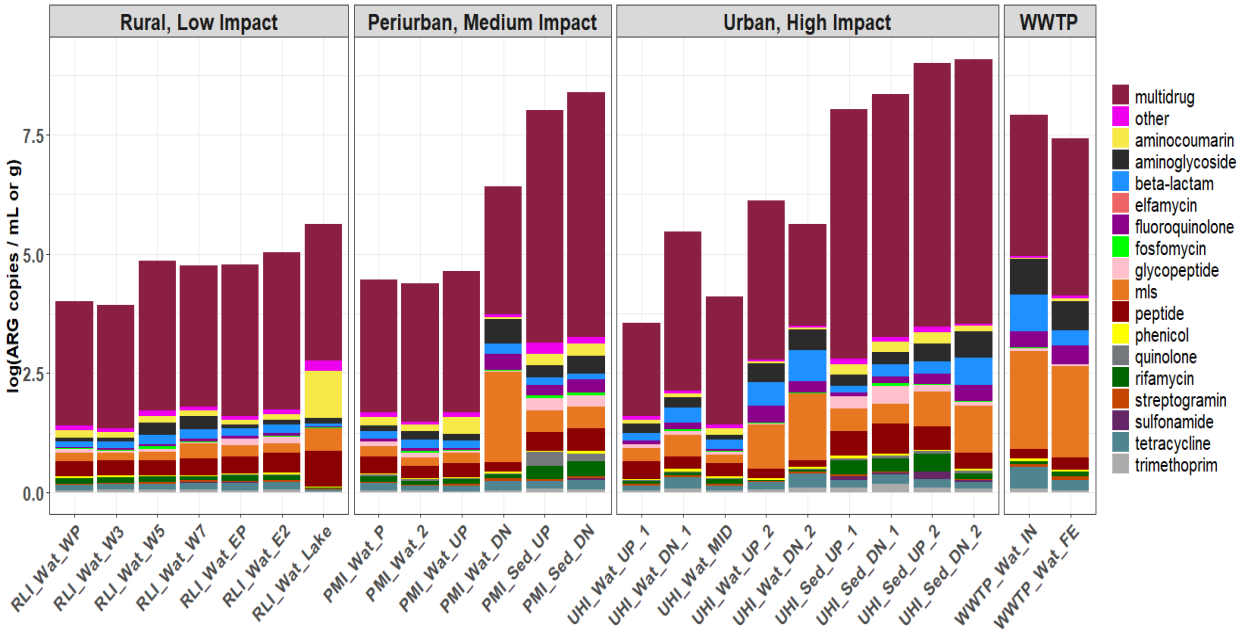


Figure S2-5: Absolute abundance of ARGs per mL or g by antibiotic class as enumerated by alignment of shotgun metagenomic sequences to the Comprehensive Antibiotic Resistance Database (v2.0.1). Samples are designated by their sample type and position within the watershed (P=Pristine, UP=Directly Upstream of WWTP effluent, DN= Directly Downstream of WWTP, IN=Influent to Patillas WWTP, FE=Final Effluent of Patillas WWTP, Wat=Bulk water, Sed=Sediment).

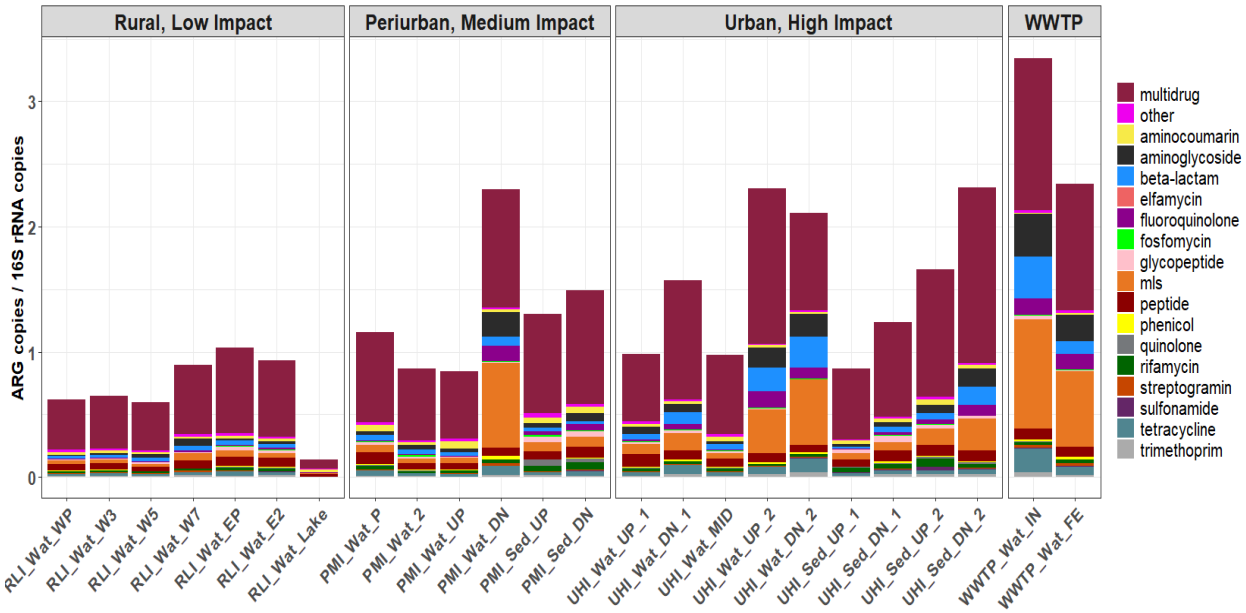


Figure S2-6: Relative abundance of ARGs per 16S rRNA copies by antibiotic class as enumerated by alignment of shotgun metagenomic sequences to the Comprehensive Antibiotic Resistance Database (v2.0.1). Samples are designated by their sample type and position within the watershed (P=Pristine, UP=Directly Upstream of WWTP effluent, DN= Directly Downstream of WWTP, IN=Influent to Patillas WWTP, FE=Final Effluent of Patillas WWTP, Wat=Bulk water, Sed=Sediment).

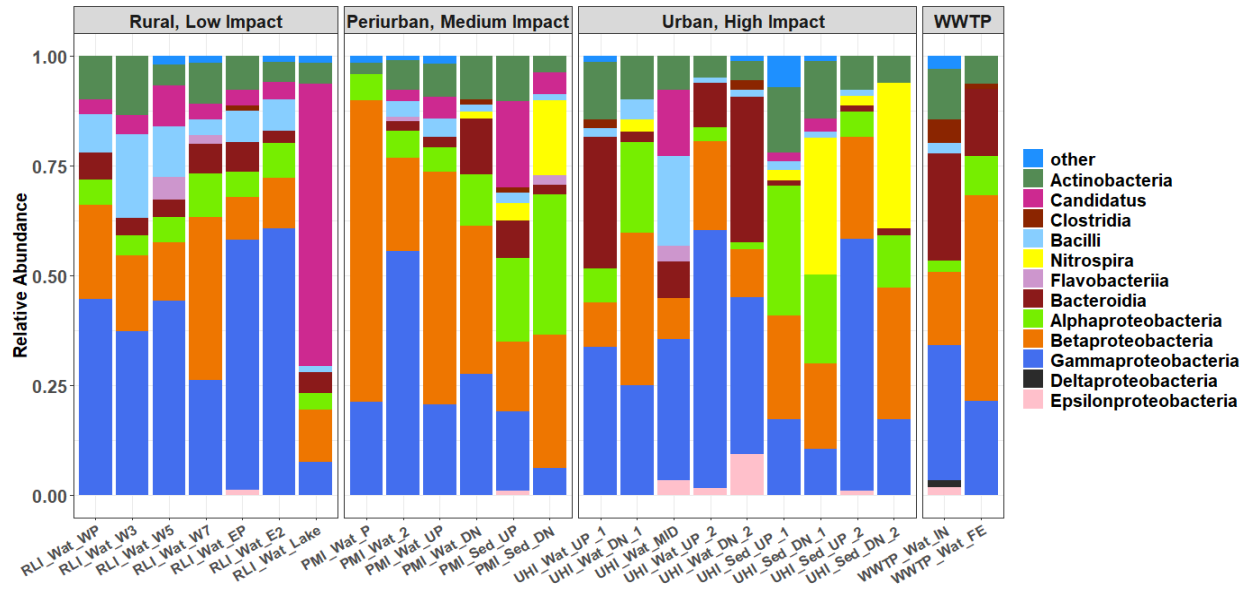


Figure S2-7: Relative abundance of taxonomic assignments generated using the Centrifuge engine against all bacterial and archaeal genomes contained in the RefSeq database (updated

4/15/2018). The Proteobacteria phyla is divided into its individual class assignments. Phyla with less than 1% abundance representation were binned as “other”.

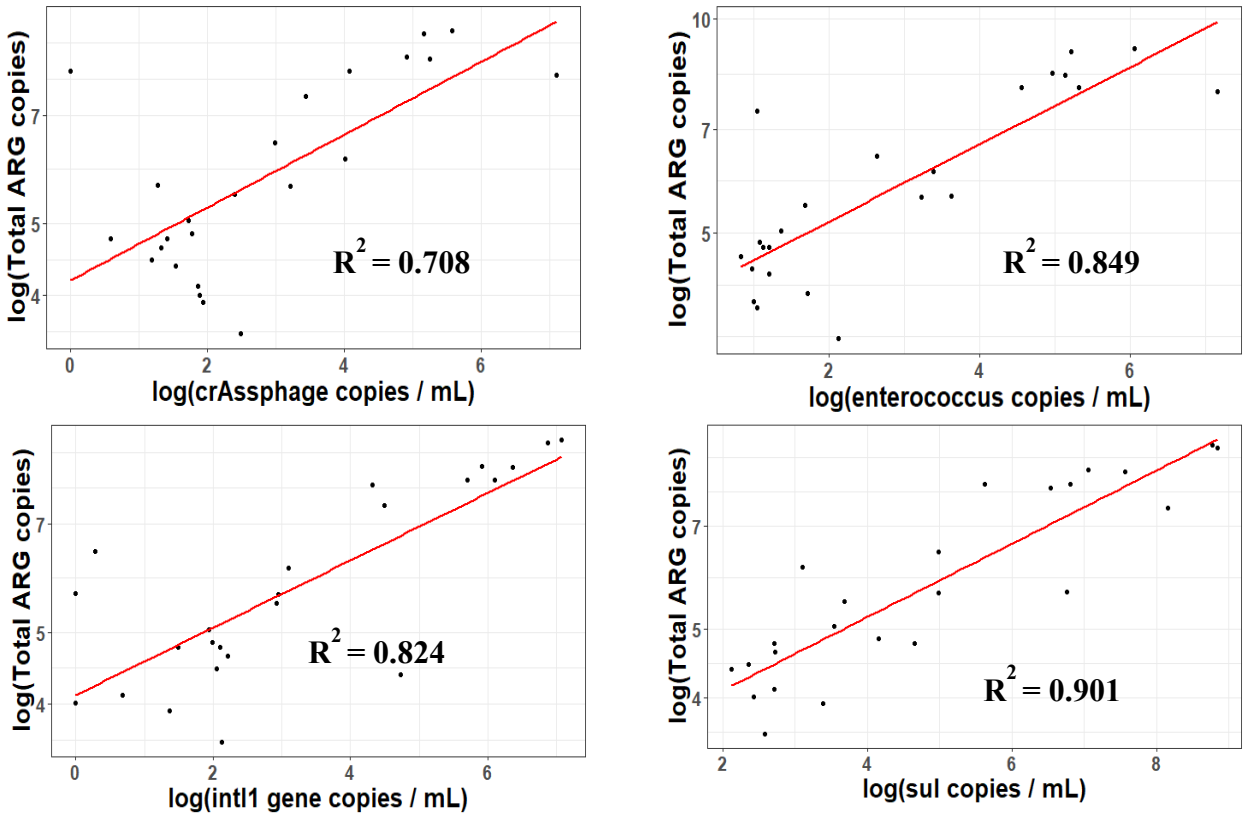


Figure S2-8: Correlations between human fecal markers (HFMs) (crAssphage and enterococcus) and anthropogenic antibiotic resistance markers (AARMs) (intI1 and sul1) with log10 Total ARG copies per sample.

Supplemental data tables can be found under my Open Science Framework projects:
https://osf.io/xty5v/

CHAPTER 3: EVALUATION OF RESISTOME RISK REDUCTION THROUGH SECONDARY WASTEWATER TREATMENT ACROSS AN INTERNATIONAL TRANSECT

Benjamin C. Davis[1], Emily D. Garner[2], Peter J. Vikesland[1], Amy Pruden[1]

1 Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia

2 Wadsworth Department of Civil and Environmental Engineering, West Virginia University, Morgantown, West Virginia

ABSTRACT

Wastewater treatment plants are at the forefront of mitigating the impact of human fecal pollution on the environmental dissemination of antibiotic resistance. Understanding the dynamics of antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARGs) through conventional activated sludge (AS) treatment plants is critical for evaluating both the human health and ecological resistome risks that treated wastewater effluents pose to receiving environments. To investigate ARB and ARG dynamics, and their relationship to the human fecal microbiome and resistome risks, 11 WWTPs representing distinct influent compositions from Europe/US and Asia were sampled and investigated using shotgun metagenomic sequencing and qPCR. From primary effluents (PE) to final treated effluents (FE), total ARG abundance was significantly reduced (~2 log reduction), regardless of PE composition or the bioinformatic normalization strategy employed. These reductions directly reflect the efficiency of individual WWTPs at removing both total bacterial cells (measured as 16S rRNA gene copies) and the abundance of human fecal indicators (crAssphage). Procrustes analysis revealed high structural symmetry of microbiome and resistome compositions in each treatment compartment, suggesting each stage of treatment represents distinct wastewater ecosystems with limited horizontal ARG transfer. Contig-based analysis indicated that a majority of ARGs were chromosomally-bound and directly tied to their host taxa. Resistome risks, measured as the prevalence of high-risk ARGs and their co-localization with mobile genetic elements and pathogens, were significantly reduced in final treated effluents. However, several high-priority and mobile ARGs persisted globally across effluents. These findings support the use of conventional biological treatment as a significant barrier to the dissemination of antibiotic resistance.

INTRODUCTION

Antibiotic resistance is an urgent threat to global human health [1,2]. Wastewater treatment plants (WWTPs) have been identified as critical barriers to, but also potential foci for, the dissemination of antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARGs) to receiving environments [3,4]. WWTPs receive diverse anthropogenic biotic and abiotic pollutants in raw sewage including human pathogens, antibiotics, biocides, and heavy metals that could possibly select or co-select for ARB and ARGs. Although WWTPs are efficient at attenuating total

microbial load, they are not designed to specifically remove ARB and may be engendering the mobilization of ARGs during biological treatment.

Conventional activated sludge (AS) WWTPs are the most widely applied in world, particularly in urban areas, and typically involve three fundamental treatment processes: biological (i.e., aeration), physical (i.e., gravimetric settling), and disinfection [5]. Several concerns have been raised about the potential for biological treatment to facilitate the mobilization and proliferation of ARGs [6]. The high cell density and prevalence of stressors (e.g., antimicrobials and heavy metals) in aeration basins may facilitate horizontal gene transfer (HGT) and potentially increase the total abundance of the collection of ARGs (i.e., the resistome). The increased stressor to planktonic cell concentration ratio present in secondary clarifiers have also been demonstrated to exhibit stronger selection potential for ARB and ARGs [7]. Disinfection processes, which undoubtedly reduce pathogenic bacteria and further reduce microbial loads, have been shown to increase the relative abundance of ARGs in final effluents [8,9]. Together, these treatment processes represent dynamic and complex microbial networks that undergo various levels of stress, selective pressures, and ultimately, potential for attenuating or proliferating resistance as unique wastewater "ecosystems" [10,11]. Due to this complexity, there have been several conflicting reports regarding the extent to which ARGs are removed through treatment [12–14]. Many factors may attribute to these disparities, including differences in influent microbial community composition, heterogenous sampling strategies, variable ARG databases and bioinformatic strategies employed (e.g., sequence-based homology-based cutoffs), and ultimately, individual WWTP performances.

Many studies have attempted to use operational conditions (e.g., temperature, pH, biochemical oxygen demand, etc.) to predict ARG dynamics, but to date, no single physicochemical factor has been shown to reliably predict the fate of ARB and ARGs through treatment [15]. The use of human fecal indicators is an attractive alternative as a holistic measurement of overall treatment performance and ARG removal. The human fecal microbiome is one of the dominant sources of high-priority ARGs and pathogenic taxa to environmental systems [16], and thus understanding the behavior and the fate of ARB, ARGs, and human fecal bacteria through conventional secondary treatment is critical for evaluating the risks that treated effluents pose to both the transmission and evolution of resistance into receiving waters around the world. The human fecal microbiome represents not only a human health risk, but also an ecological risk for the dissemination of mobile ARGs [17].

Managing human waste streams is a critical facet of efforts to combating the spread of AMR [19]. Here we seek to establish the extent to which conventional wastewater treatment; including biological, physical, and disinfection processes, act to shape the resistome of the final effluent released to the environment and the relative risk it poses to human health and the dissemination of antibiotic resistance. To accomplish this, we examined an international transect of WWTPs that receive influents that represent extremes in influent composition (i.e., Europe/US versus Asia), comparing shifts in resistome composition with each stage of treatment [18]. Primary effluent (PE), activated sludge (AS), secondary effluent (SE), and final effluent (FE) was collected across eleven WWTPs located in Switzerland, India, Hong Kong, Sweden, the United States, and the Philippines. We utilized qPCR, metagenomic sequencing, and *de novo* assembly to gain a quantitative and mechanistic understanding into how each stage of treatment can act either to amplify or to attenuate ARG abundance. We further relate ARG removal to the removal of the human fecal marker, crAssphage, as well as a collection of human gastrointestinal bacterial genomes, the Unified Human Gastrointestinal Genome (UHGG) collection, to establish the extent

to which treatment processes that are well functioning according to design also remove ARGs. The findings will help prioritize where improvements to wastewater treatment may be warranted to further reduction of mobile, pathogen-associated ARGs (i.e., "resistome risk"), both on a geographic and treatment scale.

MATERIALS AND METHODS

Sample Collection and Processing

A total of 52 wastewater samples were collected over the course of 2 years (2016-2018) from 11 WWTPs employing conventional biological treatment in Switzerland, India, Hong Kong, Sweden, the United States, and the Philippines. The Hong Kong WWTPs were sampled on two separate occasions with a year of separation. A summary of the WWTP characteristics can be found in Table S3-1. The collection and processing of samples from PE, AS, SE, and FE was uniform as previously described [20]. Briefly, grab samples were taken in autoclaved polypropylene bottles and transported to nearby laboratories on ice. Samples were homogenized by shaking and biomass was filter concentrated in triplicate on three individual 0.22-µm mixed cellulose ester filters at equal volumes, fixed in 50% ethanol, and stored at -20 °C. Ethanol-fixed filters were shipped back to Virginia Tech on ice packs for further processing. DNA extraction was performed with the FastDNA Spin Kit for Soil (MP Biomedicals, Solon, OH) and total double-stranded DNA was quantified via the dsDNA High Sensitivity Assay kit on a Qubit® Fluorometer (Invitrogen, USA) and its quality was assessed with a NanoPhotometer® Pearl (Implen, USA). One liter of deionized water was also filter concentrated and extracted to ensure no contamination occurred during DNA extraction.

Quantitative Polymerase Chain Reaction

Using qPCR, the abundance of total bacteria (16S rRNA genes) [21], *sul*1 [22], and *intI*1 [23] were quantified for each filter triplicate. Each assay was then performed in analytical triplicate with autoclaved deionized water as no template control. Double-stranded gBlock gene fragments (IDT) were used to generate standard curves. The minimum accepted efficiencies and R^2 values for standard curves were 80% and 0.980, respectively. Before sample quantification, dilution curves were performed on random samples from each treatment stage using the 16S rRNA gene assay to determine the per-compartment dilution factor appropriate to minimize PCR inhibition.

Shotgun Metagenomic Sequencing

Before library preparation, equal mass of DNA from each of the triplicate DNA extracts were pooled as a composite sample. For primary and secondary effluent composites, libraries were constructed using a NexteraXT DNA Library Prep Kit and sequenced by Diversigen, Inc (Houston, TX) via 2 × 150 bp paired-end shotgun metagenomic sequencing on an Illumina NovaSeq 6000. For final effluent samples, TrueSeq libraries (Illumina, San Diego, CA) were constructed and sequenced via 2 × 100 bp paired-end sequencing on an Illumina HiSeq 2500 by Virginia Tech Biocomplexity Institute Genomic Sequencing Center (Blacksburg, VA). Activated sludge libraries were constructed using a NEB Ultra II DNA Library Prep kit for Illumina (New England Biolabs, USA) and sequenced via 2 × 75 bp paired-end sequencing on an Illumina NextSeq500 by the Scripps Research Institute (La Jolla, CA). All sequences were deposited to the Sequence Read Archive (SRA) under accession number PRJNA527877.

ARG and Taxonomic Analysis

All reads were trimmed and filtered using Trimmomatic to remove adapters and low quality nucleotides [24]. Paired-end reads were then merged using Vsearch [25]. ARG annotation was performed using the DIAMOND blastx [26] (v.0.9.14) aligner against the Comprehensive Antibiotic Resistance Database (CARD, v.3.0.7, protein homolog model) with the representative hit approach (e-value ≤ 1e-10, bitscore ≥ 50, identity ≥ 80%, query coverage ≥ 80%). To normalize the probability of differing read lengths finding representative alignments, the minimum amino acid length was set to 80% of the length of the query (i.e., 150 bp = 40 aa, 100 bp = 27aa, 75 bp = 20 aa). The CARD database was modified for metagenomic analysis, as previously described [16], by removing ARGs known as global regulators and those that confer resistance by point mutations. The list of removed ARGs can be found in Table S3-2. Relative abundance was estimated by normalizing gene hits to 16S rRNA gene copies as enumerated by alignment to the Greengenes database [27]. The absolute abundance (gene copies/mL) was then derived as the relative abundance multiplied by the total abundance of 16S rRNA gene quantified using qPCR [28]. Fragments per kilobase million mapped reads (FPKM) was further used to account for differences in library sizes and confirm aggregate trends in resistome dynamics. Where both a R1 and a R2 read were aligned, the best hit based on bitscore was used in the FPKM calculation. Taxonomic analysis was performed in Kraken2 [29] (v.2.0.7) with default settings using the standard database of complete genomes in RefSeq for bacterial, archaeal, and viral domains. Relative abundances at each taxonomic ranking were then calculated from Kraken reports using Bracken [30] with a 100mer distribution classification.

Analysis of Human Fecal Indicators

The abundance of crAssphage was estimated by mapping the short-read library to the crAssphage genome using Bowtie2 [31] and calculating the average genome coverage depth using Samtools [32] as previously described [33]. The abundance of total human gastrointestinal bacteria was estimated in the same manner as crAssphage using all 204,928 reference genomes from the UHGG catalogue [34] (http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/). Both crAssphage and UHGG abundances were normalized to sample library size in giga base pairs.

***De Novo* Assembly and ARG Contextualization**

Quality filtered and trimmed reads were *de novo* assembled using IDBA-UD [35] with default parameters. The resulting contigs were then filtered for assemblies ≥ 1000 bp and protein-coding regions (CDS) were predicted using Prodigal [36] (v.2.6.3) with the "-p meta" option. CDSs were annotated against the CARD and the mobileOG dataset using the DIAMOND blastp aligner with stringent parameters (90% identity, aa length ≥ 100, e-value ≤ 1e-10, bitscore ≥ 50). We additionally checked for and removed overlapping ARG and MGE annotations as this has been shown to bias co-occurrence investigations [37]. To determine if ARGs were putatively positioned on plasmids, contigs were also subjected to the PlasFlow [38] pipeline, a software used to predict plasmid sequences in metagenomic data using genomic signatures. PlasFlow predictions were then supplemented by aligning the contigs to the COMPASS [39] database, a comprehensive dataset of over 12,000 complete plasmids, using blastn. Metagenomic contigs that were annotated as plasmids by PlasFlow or exhibited ≥ 90% identity over an alignment length ≥ 1000 bps to the COMPASS database were identified as plasmid-like contigs. Correspondingly, any contig that was not annotated as plasmid-like or that was predicted as "chromosome" by PlasFlow, was determined to be chromosomal. To illuminate putative ARG hosts, each contig was assigned taxonomy using Kraken2 as previously described. The normalized abundance of each CDS was calculated as the

coverage depth of the contig containing the CDS divided by the per-sample library size. This was performed by mapping the short-reads to the contig library using Bowtie2 and calculating coverage using "idxtstats" within Samtools. The relative contig abundance (i.e., library-normalized coverage) calculation is as follows:

$$Relative\ Contig\ Abundance\ (Coverage) = \sum_{1}^{N} \frac{n\ x\ \frac{R}{L}}{G}$$

where N is the number of contigs in a sample, n is the number of reads mapped to the contig, R is the length of the short-read, L is the length of the contig, and G is the size of the sequence library in total number of reads [40].

Diversity and Statistical Analysis

Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity matrices, permutational multivariate analysis of variance (adonis), and procrustes analysis was also performed in *vegan.* Boxplots, bar charts, linear models, and NMDS coordinates were plotted in ggplot2 (v3.2.1). Heat maps were generated using the *pheatmap* (v1.0.12) package. Nonpareil was also used to estimate metagenomic sequence diversity and coverage as a function of library size to determine the relative representativeness of individual metagenomes in downstream analyses [42]. Significant differences ($\alpha < 0.05$) in abundance and diversity metrics between groups were determined using pairwise Wilcox Rank Sum tests with Bonferroni corrections.

RESULTS AND DISCUSSION

Sequencing Resulted in Differential Coverage Across Treatment Compartments

Shotgun metagenomic sequencing yielded a wide range of library sizes per sample (0.8-16.3 Gb;10.6-112.2 million paired-end reads). Given that discrepancies in coverage could conflate ecologically-relevant differences and random variation [43], we estimated the relative sequencing coverage per sample using Nonpareil. The intrinsic sequence diversities varied significantly across wastewater ecosystems, with AS and PE displaying higher nucleotide diversity over the SE and FE compartments (Wilcox, $p < 0.05$). As a result, the mean metagenomic sequencing coverage (0-100 scale) was dependent on sample library size and varied by treatment process (PE=65.0; AS=26.8; SE=69.9; FE=51.5), where the deeper sequenced PE (mean=11.6 Gb) and SE samples (11.2 Gb) had significantly higher metagenomic coverages than the AS (1.3 Gb) and FE (2.6 Gb). To account for discrepancies in coverage, all subsequent analyses entailing comparison to the AS ecosystem were normalized to the abundance of total ARGs or all ARGs pertaining to the resistance class of interest.

de novo assembly resulted in a large collection of contigs (24,483,089 total: 14,455-1,256,698 per sample) with an average N50 of 1,234 bp (509-4771 bp) representing 4.9 Gb. Libraries were filtered for contigs ≥1000 bps to minimize false and fragmented assemblies [44], which reduced the collection by 82.4% to 4,305,338 contigs (889-220,150 per sample). This approach supported proportional comparisons of ARG-MGE and ARG-taxonomy associations, while also reducing potential for false assemblies. This truncated contig library resulted in a collection of 14,928,407 CDSs with an average contig N50 of 4,459 bp (1490-31,572 bp). We identified 304,132 individual contigs annotated with an ARG that allowed further determination of mobility potential and putative bacterial hosts carrying the ARGs through wastewater treatment.

Abundance of Total ARGs Decrease through Activated Sludge Treatment but Increase in Final Treated Effluents

We first calculated the abundances of total annotated ARGs across all 11 treatment plants to identify aggregate trends in resistome fluctuations through each stage of treatment. We compared trends resulting from normalizing abundances to biologically-relevant (gc/16S rRNA genes), volumetric (gc/mL), and library size corrected (FPKM) denominators. The coefficients of variation were 0.651, 4.65, 0.285 for each of the three normalization approaches, respectively. We found strong correlations between concentrations derived from our quantitative metagenomic method and qPCR for both *sul*1 (Pearson; $R^2 = 0.959$; $p < 2.2e\text{-}16$) and *intI*1 (Pearson; $R^2 = 0.892$; $p < 2.2e\text{-}16$), supporting further quantitative analysis of the metagenomic dataset (Figure S3-1).

Assessing the metagenomic dataset in aggregate, all three normalization strategies indicated a significant reduction in total ARG abundance as a result of activated sludge treatment, as apparent from comparison of PE to SE compartments (Wilcox, $p<0.001$) despite differences in influent resistome composition, climate, geographic location, and individual plant configurations (Figure 3-1ABC). Similar trends were observed when comparing PE to AS compartments or PE to FE compartments, but significant differences in resistome reduction or amplification in final treated effluents were dependent on the normalization strategy applied. Normalized to the bacterial fraction (i.e., 16S rRNA gene copies), total resistome abundance was reduced by an average of 51.7% from PE to the AS basins and remained significantly reduced through the FE compartments (Figure 3-1A). From an engineering perspective, relative to the PE compartment, the absolute abundance of total ARGs was reduced by ~2 logs after secondary clarification, and significantly reduced in 14 out of the 16 detected ARG classes (Figure 3-1BD, Table S3-3). Reductions in resistome abundance using FPKM normalizations resulted in an average reduction of 19.9% from PE to FE. Further, using *sul*1 and *intI*1 as proxies for resistome dynamics in urban WWTPs [45], we observed comparable magnitudes of attenuation in the complementary qPCR data as well (Figure S3-2) [46].

Interestingly, an increase in ARG abundance was observed from secondary clarifiers to the final treated wastewater across several plants, although this was statistically insignificant in aggregate (Figure 3-1BC). As our derivation of absolute concentrations of total ARGs are a function of 16S rRNA gene copies determined via qPCR, the recovery in abundance was due to a significant increase in total bacterial cells from the secondary clarifiers to the final treated effluents across several WWTPs ($p < 0.05$, Figure S3-3). Several plants that were sampled were not employing a disinfection step prior to discharge (Table S3-1), but we found no significant differences between total ARG abundances in the final effluents employing disinfection versus those that did not. The increase in abundance was primarily due to outlier WWTPs in India that had disproportionately elevated 16S rRNA gene copies per mL in their effluents (Figure 3-1D, S3-3). This same trend was further highlighted in the FPKM normalized data, as well as *sul*1 and *intI*1 qPCR data, where the two sampled WWTP effluents in India and one in the Philippines were clear outliers (Figure S3-2, S3-4). The uptick in overall ARG abundance was primarily a result of enrichment of glycopeptide, trimethoprim, and 'other' ARG classes from the SE to FE across the dataset ($p<0.05$, Table S3-3). Together, these observations strongly suggest that conventional activated sludge WWTPs are significant barriers to the dissemination of ARGs to receiving environments and do not facilitate the relative enrichment of ARGs through treatment. However,

these observations are WWTP-specific and in direct correspondence to an individual WWTP's ability to remove total bacteria.

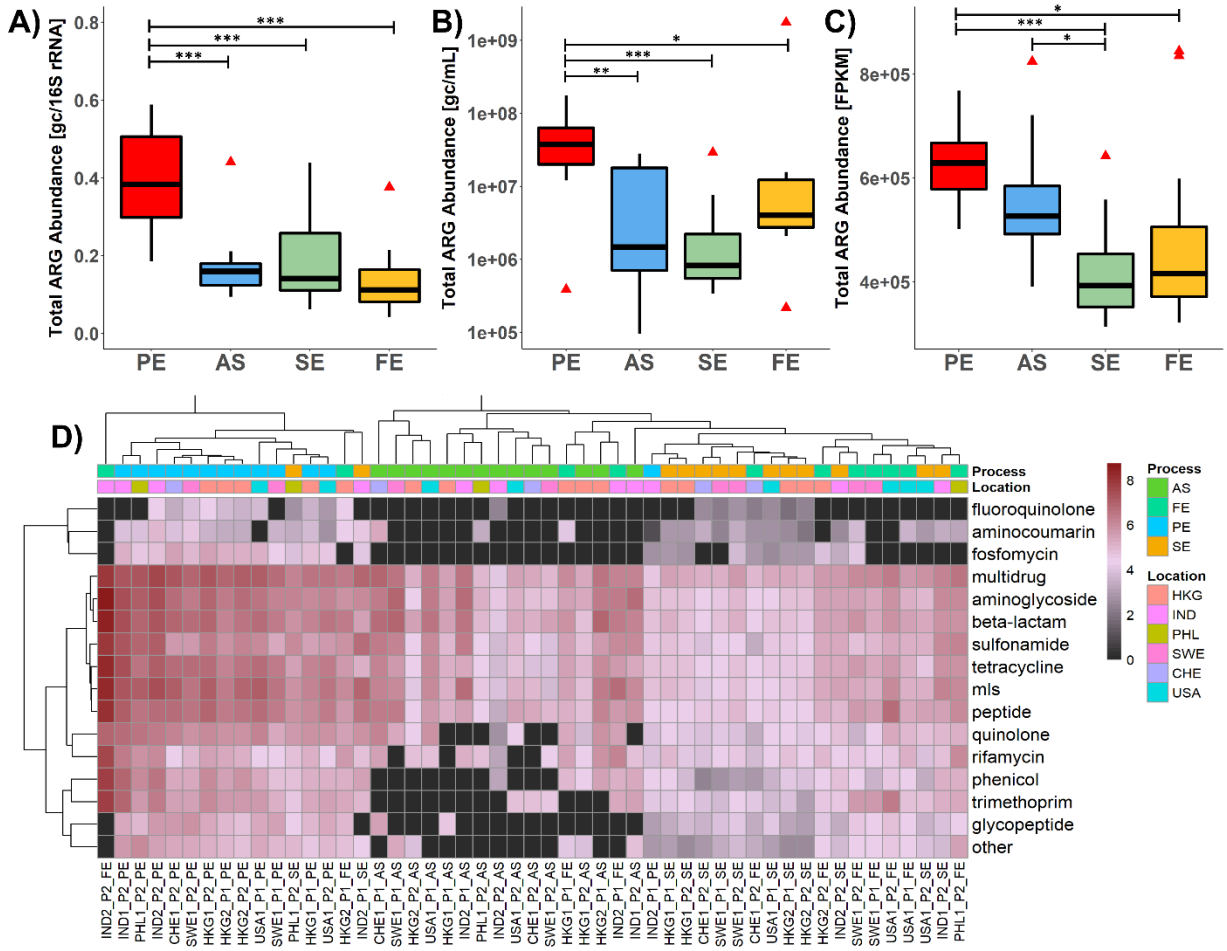


Figure 3-1: Resistome summary statistics across all 52 samples identified by metagenomic read matching to CARD (v3.0.7). (A) Relative abundance of total ARGs [gc/16S rRNA genes] , (B) volumetric (absolute) abundance [copies/mL], and (C) fragments per kilobase million [FPKM]. (D) Heat map of ARG abundance by resistance class. Color gradient represents $\log_{10}$ transformed absolute abundances in [gc/mL]. Primary effluent (PE, n = 13), activated sludge (AS, n=14), secondary effluent (SE, n=13), and final effluent (FE, n =12). Significant differences determined by pairwise Wilcox Rank Sum test with a Bonferroni correction: * 0.05, ** 0.001, *** 0.0001. Outliers are represented as red triangles.

Succession of Resistomes and Microbiomes Through the WWTP Trains

Across all samples, we detected 1094 unique ARGs derived from the CARD protein homolog database using the representative hit approach. The dominant ARG classes across all WWTPs were aminoglycoside, beta-lactam, tetracycline, macrolide-lincosamide-streptogramin (MLS), and sulfonamide ARGs (Figure 3-1D), which is consistent with larger scale international

surveys of wastewater resistomes and the geographic regions represented in this dataset [47]. Although sequencing coverages varied across treatment compartments, FPKM normalizations allowed us to reveal overarching ecological observations. For instance, based on Bray-Curtis dissimilarities, the resistome compositions were primarily shaped by the treatment compartments themselves (ADONIS, R^2=0.117, p<0.001) (Figure 3-2A), owing to the semi-uniform selectivity of engineered systems on wastewater resistomes. Further, we observed clear successional dynamics in resistome compositions whereby each treatment stage resulted in distinct resistome compositions over their previous compartments (PE vs AS, ADONIS, R^2=0.0804, p = 0.03; AS vs SE, ADONIS, R^2=0.061, p = 0.051; and SE vs FE, ADONIS, R^2 = 0.123, p = 0.006) (Figure 3-2A). Similar successional resistome dynamics have been reported in the literature, even across treatment strategies [30]. These dynamics are especially highlighted for time-sampled studies of the same WWTP [31] as well as collections of geographically-proximal WWTPs [32]. Here, the overarching dynamics spanning 11 treatment plants across climactically and demographically distinct regions indicates that the shifts in resistome composition through conventional biological treatment are relatively convergent.

Parallel shifts in composition were also observed in the microbiome data, but were far more defined (Figure 3-2B). The largest shift in composition was observed between the PE to the AS compartments (ADONIS, R2 =0.651, p < 0.001), where there was an immediate shift from obligate anaerobic phyla and classes common to human and animal gastrointestinal tracts (e.g., *Bacteroides, Firmicutes, Epsilonproteobacteria*) to facultative, motile, and filamentous genera that are members of the *Pseudomonadota* (e.g., *Thauera, Burkholderia, Pseudomonas, Dechloromonas*) **(**Figure S3-5) [33]. Similarly strong shifts were observed from AS to SE (ADONIS, R2=0.617, p < 0.001) and from SE to FE (ADONIS, R2=0.469, p<0.001), suggesting a strong turnover of the bacterial community through each progressive stage of the treatment. We further investigated the direct relationship between resistome and microbiome compositions through Procrustes analysis, wherein pairs of data matrices, in this case pairwise distance matrices, are compared using a rotational fit algorithm that minimizes the sum-of-squares residuals between them [34] (Figure 3-2C). We found strong correspondence between the structures of the total resistome and microbiome datasets (protest, R=0.771, p < 0.001). Remarkably, when we investigated each treatment stage independently, we found even stronger structural symmetry: PE (R=0.8706, p <0.001), AS (0.8589, p <0.001), SE (R=0.9011, p<0.001), and FE (0.894, p=0.003) (Figure 3-2C), illustrating distinct wastewater ecosystems. Additionally, we observed strong correspondence between the proportion of phyla putatively harboring ARGs in the contig dataset with that of the total microbiome data (Figure S3-5, S3-6), further highlighting the interdependence of the microbiome and resistome compositions and validating the relative accuracy and representativeness of the contig library for the overall structure of each original metagenome.

Several studies have shown that microbial phylogeny structures the resistome [35,36]. In other words, most ARGs are immobile, directly associated with their host taxa of origin, and are not actively being exchanged across taxa. If HGT events were occurring at appreciably high frequencies through wastewater treatment, we would expect significant deviations in compositional symmetry between the microbiomes and resistomes. Here we observe significant structural symmetry through each successional stage of treatment and that these observations are convergent even across an international transect of WWTPs. This suggests that vertical transmission of ARGs is likely the dominant mechanism of ARG transference and that any shift

in resistome abundance is likely tied to microbiome shifts driven by deterministic physicochemical factors (e.g., dissolved oxygen, temperature, pH, and COD).

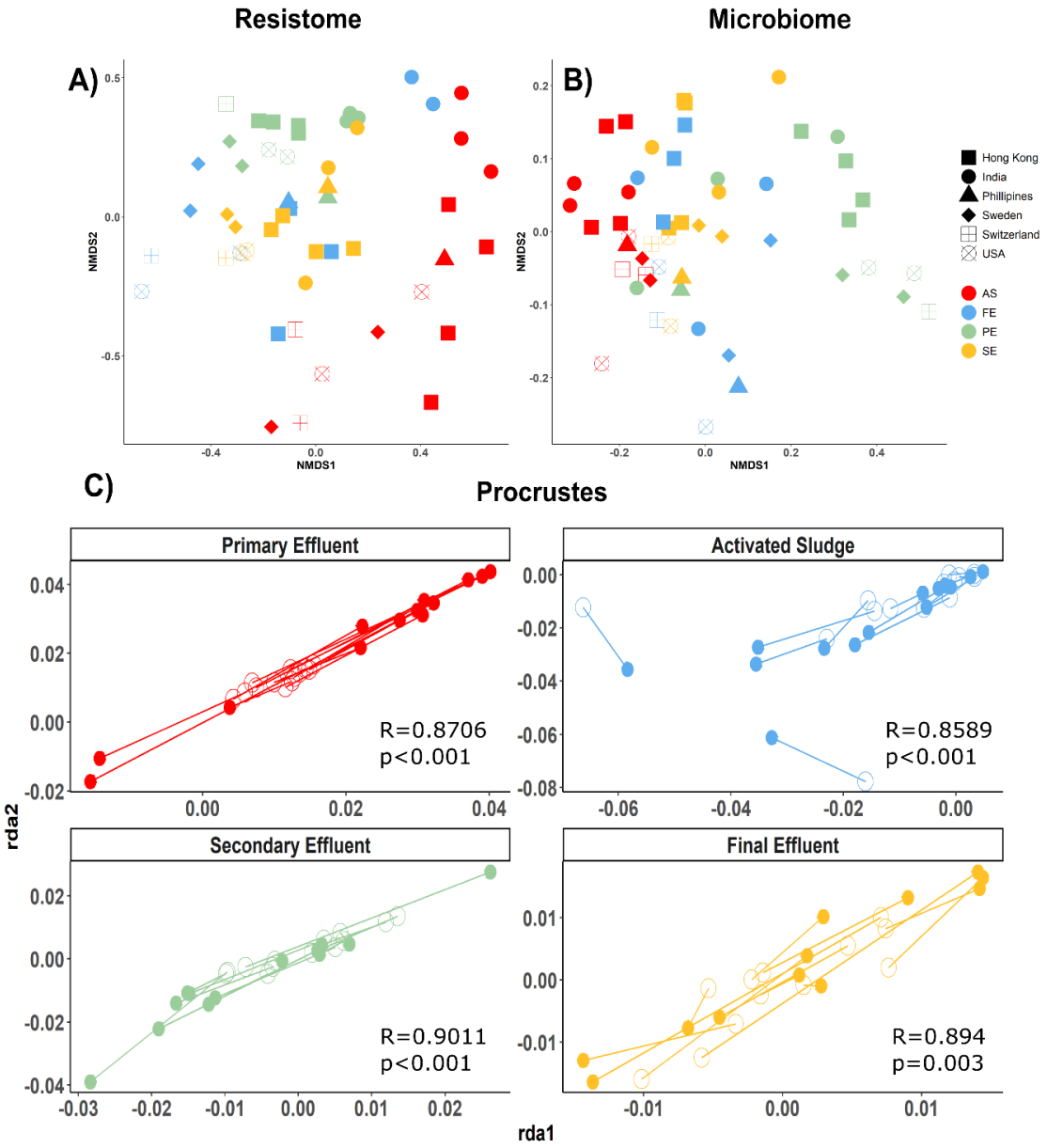


Figure 3-2: Non-metric multidimensional scaling (NMDS) plots on Bray-Curtis dissimilarity matrices of microbiome and resistome. (A) The resistome non-metric multidimensional scaling (NMDS) plot was generated from Bray-Curtis dissimilarity matrices of ARG FPKM abundance data. (B) The microbiome plot was generated from Bray-Curtis dissimilarity matrices on rarefied Bracken estimated read counts at the genus level. (C) Procrustes rotations were generated using the "protest" function in "vegan" and are paneled by treatment process with independent x- and y-axes.

Relationship Between Human Fecal Indicators and the Resistome

It has been illustrated that the human fecal indicator, crAssphage, is a strong predictor of resistome abundance in anthropogenically impacted environments, including WWTP ecosystems [33]. To test if this was the case with our dataset, we quantified both crAssphage and a dataset of 208,928 complete genomes collected from human gastrointestinal samples, the UHGG catalogue. Because crAssphage is a unique biomarker (i.e., a single viral particle), we hypothesized that a more comprehensive dataset representing the full diversity of human gastrointestinal taxa would provide greater resolution in predicting resistome fluctuation. Human feces is also one of the greatest sources of human pathogenic bacteria and high-priority ARGs and estimating its removal efficiency is advantageous for monitoring WWTPs.

The relative abundance of crAssphage (library-normalized coverage) had a 5-log range across all samples ($8.4x10^{-4}$ – 13.2) and was significantly reduced (~2-log reduction) in the AS basins, remaining significantly reduced through the SE and FE compartments (Wilcox, $p<0.0001$) (Figure 3-2A). The UHGG catalogue abundance was much lower (due to the size of the database) and had a much narrower range than crAssphage ($3.3x10^{-4}$ – $4.7x10^{-2}$), but mirrored its dynamics, showing significant reductions through secondary clarification (Figure 3-2A). The relative abundance of both indicators correlated strongly with each other (Pearson, $R^2 = 0.771$, p=2.23e-11) (Figure 3-2B), consistent with crAssphage being largely representative of the behavior of the cumulative human gastrointestinal microbiome in wastewater. Interestingly, there was a significant increase in the UHGG abundance from SE to FE compartments across the dataset (Figure 3-2A). There were weak correlations between 16S rRNA gene copies and crAssphage (Spearman, $\rho=0.332$, $p=0.016$) and UHHG abundances (Spearman, $\rho=0.241$, $p=0.084$), indicating that the apparent recovery of total bacterial cells may be in part due to human fecal-related microbiota in discharged water. The resistome abundance significantly correlated with crAssphage through wastewater treatment (Pearson, $R^2 = 0.3756$, $p=0.0061$), although not as strongly as the UHGG catalogue (Pearson, $R^2 = 0.6146$, p=1.25e-6). While crAssphage is an excellent indicator of human fecal pollution in environmental microbiomes [48], it is still a singular biomarker that may be dependent on other factors for its abundance on a per sample basis (e.g., particulate adhesion, decay). The crAssphage capsid is also ~70 nm in diameter and may have passed through our 0.22 um filters. This suggests that using UHGG normalized coverage may be a more robust metric for assessing both human gastrointestinal bacteria abundance and its relationship to the resistome in complex environmental metagenomes utilizing the commonly applied filter concentration method.

It is often difficult to discern whether an increase in resistome abundance is due to the selection of ARB and ARGs by antimicrobials, metabolic selection of taxa harboring ARGs due to shifts in ambient physicochemical parameters, or simply the result of an increased prevalence of fecal bacteria [49]. The clearest evidence for resistance selection in environmental samples, where there is a deviation from the correlation of levels of fecal pollution and resistome abundance, has come from sediments impacted by pharmaceutical wastes with antimicrobials exceeding clinical concentrations [33]. If there is an observed increase in resistome abundance relative to the human fecal bacteriome, this is therefore evidence for the environmental selection of ARB or ARGs. To investigate this hypothesis, we normalized total resistome abundance (as FPKM) to the

abundance of the UHHG catalogue through each stage of treatment (Figure 3-2C). Relative to the PE, we found a significant increase in the fecal normalized resistome abundance in the activated sludge basins (Wilcox, p=0.00093) (Figure 3-3C). This increase was sustained through secondary clarification, where it eventually returned to levels comparable to the PE in the final treated effluents. This increase in fecal normalized abundance in aeration basins may be in part due to the overall shift in taxonomic structure that was incurred via the shift to aerobic conditions. The fecal taxa were replaced by WWTP-taxa that also harbored ARGs. This is an observation of metabolic selection that allow specific genera of bacteria to proliferate in aeration basins, not necessarily the selection of ARB or the induction of HGT.

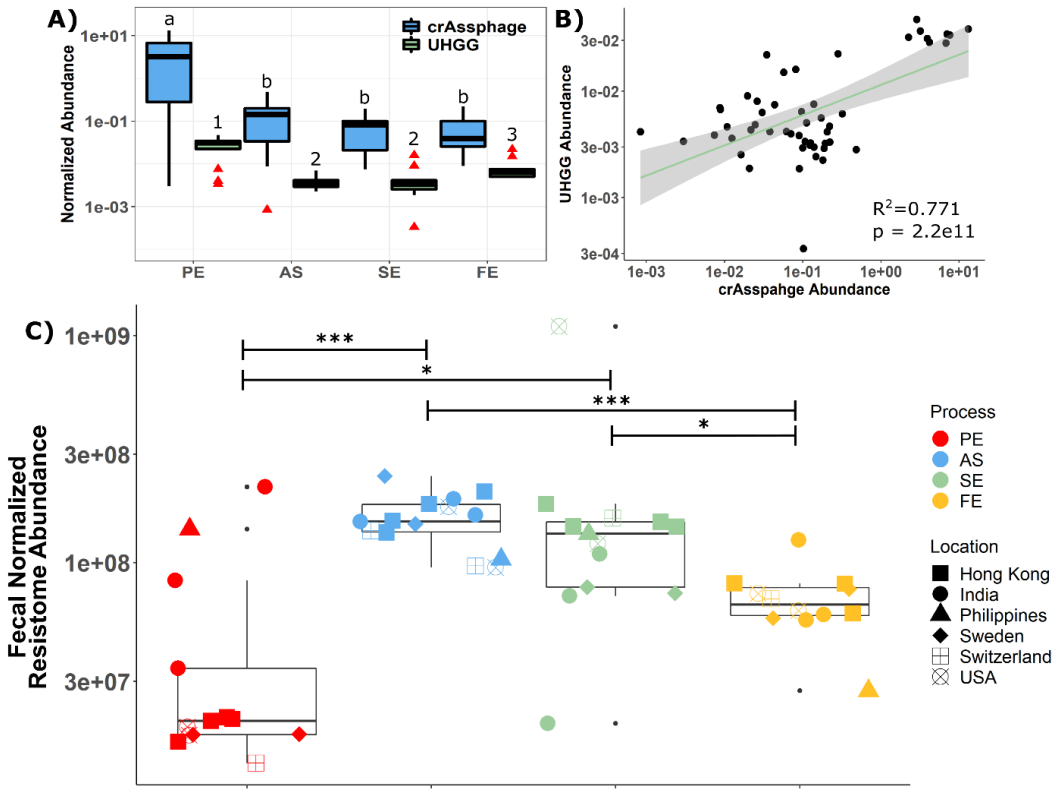


Figure 3-3: Behavior of fecal indicators through wastewater treatment and relationship to total resistome. (A) Normalized abundance of crAssphage and Unified Human Gastrointestinal Genome (UHGG) catalogue through wastewater treatment. Abundance calculated as the genome coverage normalized to total library size in reads queried. (B) Pearson correlations of $\log_{10}$ transformed data between crAssphage and UHGG abundance. (C) Fecal normalized resistome abundance across treatment compartments and individual treatment plants. Fecal normalized resistome abundance calculated as FPKM of ARGs/normalized UHGG abundance. Significant differences determined with Bonferroni corrected pairwise Wilcox Rank Sum tests: * 0.05, ** 0.001, *** 0.0001.

The Genetic Context of ARGs

We next investigated the genetic context of ARGs and their mobility potential through wastewater treatment by analyzing the library of assembled contigs. Results from PlasFlow analysis indicated that, the ratio of chromosomal to plasmid contigs was 80:20 across the contigs that could be classified. This ratio was remarkably consistent through each treatment compartment (Figure 3-4D). These proportions were ARG class-specific, however, with sulfonamide, aminoglycosides, and beta-lactam ARGs disproportionately occurring on plasmids (Wilcox, $p < 0.001$) (Figure S3-7). We then analyzed ARG co-occurrence patterns with a database of MGE hallmark genes related to insertion sequences, integration/excision proteins, conjugation machinery, and replication initiation modules, to estimate the frequency of these co-occurrences. If an increase in the relative proportion of ARGs co-occurring with MGEs through treatment is observed, this may reflect evidence for selection of ARGs or ARB or provide evidence for HGT. The average mobility association of ARGs, calculated as the number of ARG contigs co-located with an MGE (24,770 contigs) as a proportion of all ARG contigs (304,132 contigs), was 6.6%, 7.9%, 6.9%, and 7.2% for PE, AS, SE, and FE samples respectively, with no significant increase in co-occurrence with respect to PE (Figure 3-4A). Across the dataset, the classes with the highest median mobility association were multidrug (34.8%), MLS (16.0%), and tetracycline (12.7%) ARGs, with the lowest belonging to phenicol (0.58%), fosfomycin (0.50%), and elfamycin (0.23%). At the individual gene variant level, we observed highly mobilized ARGs (100% MGE co-locations) notably conferring beta-lactam resistance; specifically, several variants of high-priority ARGs identified by the CDC as posing serious health threats including blaCTX-M-15, blaKPC-2, blaOXA-48, and blaVIM [50].

To further investigate the context of these potentially mobile ARGs, we subset the ARG-MGE contigs and determined the proportion of contigs that were associated with plasmids. We found that only a small fraction of mobile ARGs were annotated as plasmid-associated. Specifically, the median percent plasmid associations were 14.8%, 6.1%, 7.4%, 8.3% for PE, AS, SE, and FE, respectively, and these proportions were significantly reduced after secondary clarification (Wilcox, $p<0.0001$) (Figure 3-4B). The ARG classes most co-located with MGEs and annotated as plasmid-associated were sulfonamides, beta-lactams, aminoglycosides, and MLS (Figure 3-4B). These data indicate that, overall, there was not a proliferation of plasmid-borne ARGs capable of intercellular transfer after treatment and that over 90% of potentially mobile ARGs were chromosomally-bound. We further estimated the relative abundance of ARG-containing contigs by mapping the short-read data back to each contig and calculating the library-normalized coverage depth. We found a significant reduction in the relative abundance of chromosomal ARG contigs, and ARG contigs co-occurring with MGE or plasmid sequences from PE to FE compartments (Wilcox, $p<0.001$) (Figure 3-4C). Although certain clinically-relevant ARGs and ARG classes were associated with mobility, these contigs represented a small fraction of the total metagenome and their abundances were being incrementally attenuated. As the majority of ARG and ARG-MGE contigs were found to be chromosomal, this indicates that their removal was primarily driven by the removal of their host bacterium.

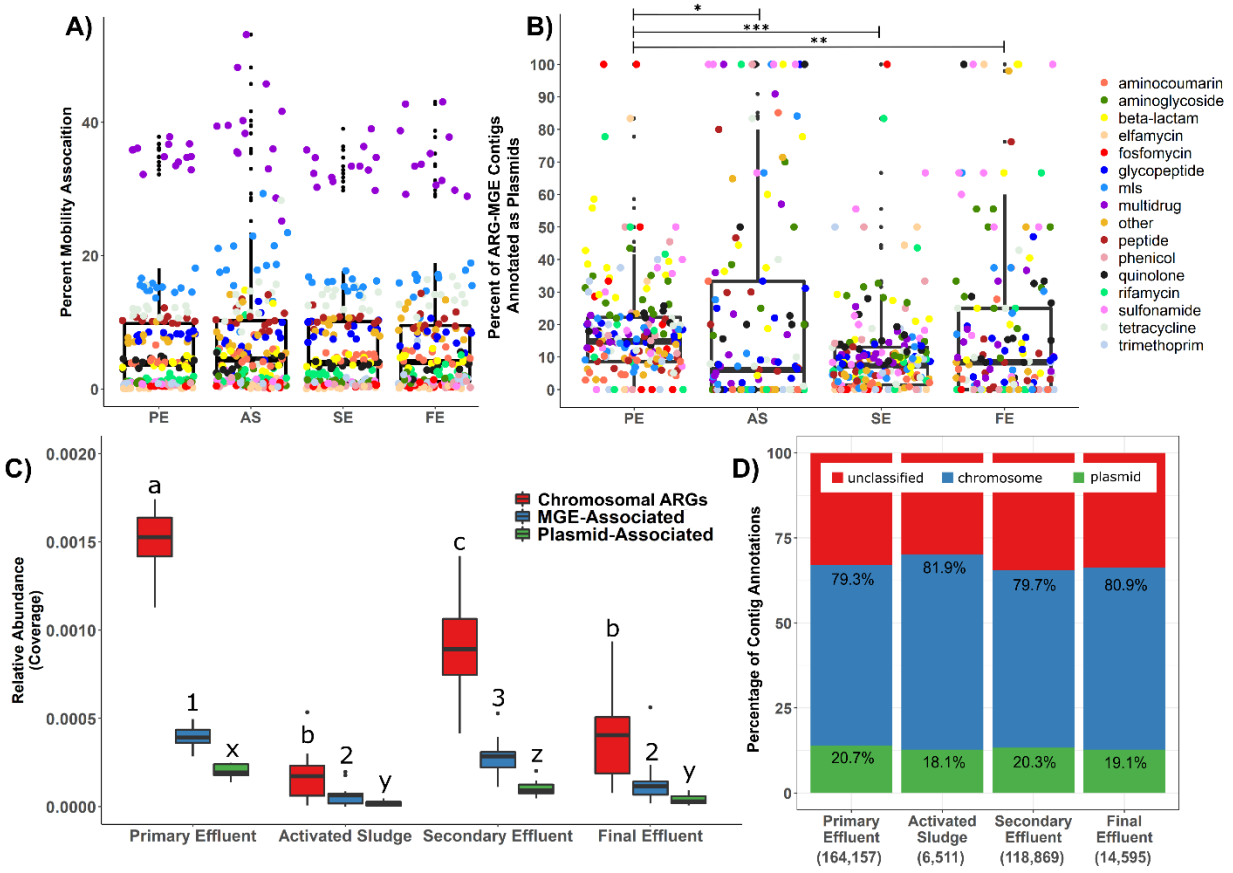


Figure 3-4: Genetic context of ARGs across treatment compartments. (A) Percent mobility association by treatment stage and ARG class. Percent mobility association determined as the number of ARG contigs co-located with an MGE (24,770 contigs) as a proportion of all ARG contigs (304,132 contigs) within each sample. (B) Percent of ARG-MGE contigs annotated as a plasmid-like sequence. (C) Relative abundance of ARG contigs and ARG contigs co-located with MGEs and plasmid-like sequences. Relative contig abundance calculated as the coverage of each contig normalized to the library size in number of reads. (D) Results of PlasFlow analysis. The number of ARG contigs representing each treatment stage are in parentheses. Significant differences determined using a pairwise Wilcox Rank Sum test with Bonferroni corrections. Significance cutoffs: * 0.05, ** 0.001, *** 0.0001

Resistome Risk Assessment Across Wastewater Resistomes

To conduct a comprehensive resistome risk evaluation through treatment, we utilized two existing bioinformatic pipelines that analyze both the contig and short-read datasets. The first is MetaCompare [53], which calculates a "relative resistome risk" score as a function of the cooccurrence patterns of ARGs, MGEs, and pathogens within contigs. The second is "arg_ranker", which utilizes a similar framework for rank ordering individual ARGs detected within the short-read dataset, considering their enrichment in human-associated environments, their history of mobility, and their host pathogenicity [54]. The "arg_ranker" pipeline further demarcates

individual ARGs into Ranks 1-4, Rank 1 being the 'current threats' that are already present in pathogens and represent a hazard to human health, which we utilized in this analysis. Similar to trends observed in the total resistome abundance, we observed significant reductions in relative resistome risk and Rank 1 ARG Risk across most treatment plants (Figure 3-5AB). WWTPs in India and the Philippines, however, displayed marked increases in resistome risk and Rank 1 ARG risk scores from PE to final treated effluents. Within the MetaCompare dataset we recognized a clear regional divide in relative resistome risk scores across all samples between WWTPs located in Asia versus those in Europe/US (Wilcox, p=0.0002) (Figure 3-5A). This divide was statistically significant between PE (p=0.045), SE (p=0.0016), and FE compartments (p=0.048). These regional divides were not present in the arg_ranker dataset either across the full dataset (p=0.299) nor within individual treatment stages.

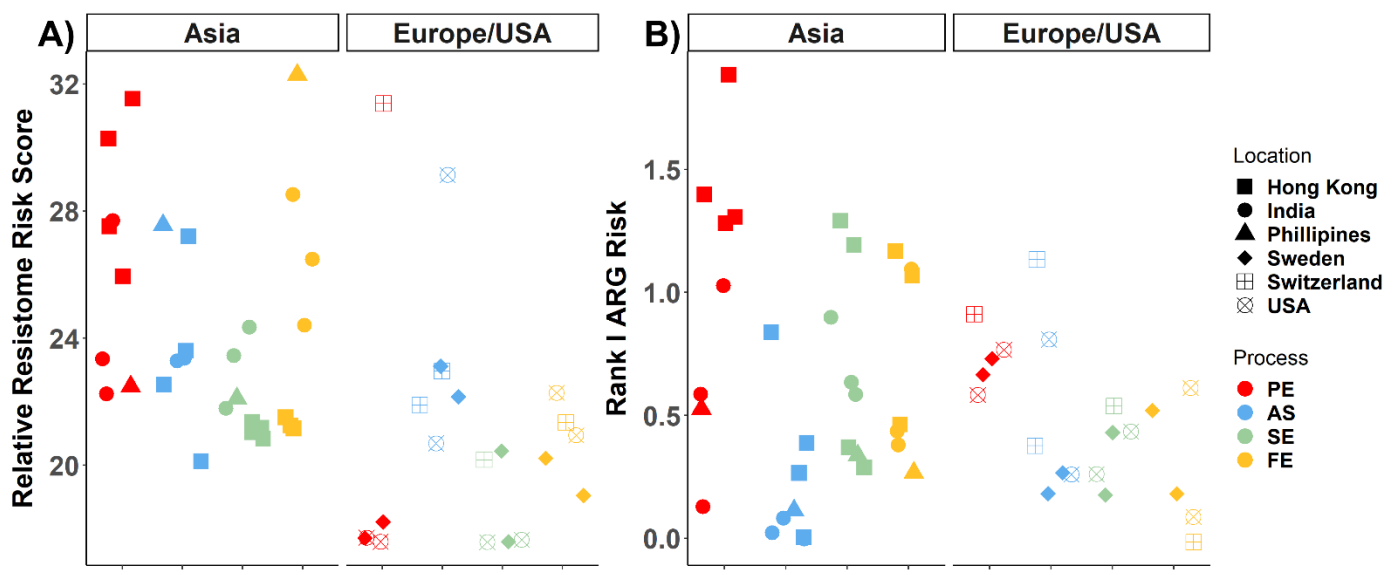


Figure 3-5: Resistome risk assessment across all samples. A) Relative resistome risk score derived from MetaCompare [53] using assembled metagenomic data. Risk scores are calculated using (i) the number of occurrences of ARGs on assembled contigs, (ii) the number of co-occurrences of ARGs and MGEs, and (iii) the number of co-occurrences of ARGs, MGEs, and human pathogen-like sequences normalized to the library size. B) Risk scores for individual high-priority ARGs derived from "arg_ranker" pipeline [54] using short-read data. Risk calculation considers relative abundance of individual ARGs for anthropogenic enrichment, mobility, and putative host pathogenicity.

CONCLUSIONS

Overall, we found strong evidence for the attenuation of the resistome through conventional biological wastewater treatment. Reduction in ARGs was primarily driven by the removal of the bacteria carrying ARGs and corresponded with an overall removal of fecal-derived microbiota due to (i) the metabolic selection of aerobic bacteria in aeration basins, and (ii) an out-competition of fecal bacteria by naturalized activated sludge communities. We found evidence for

potential resistance selection using fecal normalized resistome data and class 1 integrons, but this is likely due to the proliferation of ARGs related to taxa native to aeration basins becoming subdominant over fecal-related ARGs. This is supported by the NMDS observations where there was a clear turnover of ARGs through each stage of treatment and that these ARGs were directly correlated to their host taxa. Co-occurrence analysis revealed that there was not an overall mobilization of ARGs because of biological treatment trains and the mobility was dependent upon ARG class. Approximately 80% of ARGs were found to be chromosomally-bound and not associated with plasmid sequences. This conflicts with studies utilizing Nanopore data that suggest a much larger fraction of the resistome is plasmid-borne, however, our data display strong correspondence between short-read and assembled data. Resistome risk was significantly reduced as well as resistant ESKAPE pathogens, suggesting that secondary wastewater treatment is a significant barrier to the dissemination of high-priority ARB and ARGs, although there are still significant fluxes of resistance in effluents around the world and advanced treatment should be considered in many high transmission-risk areas.

REFERENCES

1. Kwon JH, Powderly WG. The post-antibiotic era is here. Science (80-). 2021;373:471–471.

2. World Health Organization. Global Action Plan on Antimicrobial Resistance. 2015.

3. An XL, Su JQ, Li B, Ouyang WY, Zhao Y, Chen QL, et al. Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR. Environ Int [Internet]. Elsevier; 2018;117:146–53. Available from: https://doi.org/10.1016/j.envint.2018.05.011

4. Guo J, Li J, Chen H, Bond PL, Yuan Z. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. Water Res [Internet]. Elsevier Ltd; 2017;123:468–78. Available from: http://dx.doi.org/10.1016/j.watres.2017.07.002

5. Jenkins D, Wanner J. Activated Sludge-100 years and counting. IWA Publshing; 2014.

6. Jin M, Liu L, Dong DW, Liu YW, Yin J, Wang ZYH, et al. Chlorine disinfection promotes the exchange of antibiotic resistance genes across bacterial genera by natural transformation. ISME J [Internet]. Springer US; 2020; Available from: http://dx.doi.org/10.1038/s41396-020-0656-9

7. Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, Singer HP, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes: Supplementary material. ISME J. 2019;1–39.

8. Zheng J, Su C, Zhou J, Xu L, Qian Y, Chen H. Effects and mechanisms of ultraviolet, chlorination, and ozone disinfection on antibiotic resistance genes in secondary effluents of municipal wastewater treatment plants. Chem Eng J [Internet]. Elsevier B.V.; 2017;317:309–16. Available from: http://dx.doi.org/10.1016/j.cej.2017.02.076

9. Lin W, Zhang M, Zhang S, Yu X. Can chlorination co-select antibiotic-resistance genes? Chemosphere [Internet]. Elsevier Ltd; 2016;156:412–9. Available from: http://dx.doi.org/10.1016/j.chemosphere.2016.04.139

10. Quintela-Baluja M, Abouelnaga M, Romalde J, Su JQ, Yu Y, Gomez-Lopez M, et al. Spatial

ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. Water Res [Internet]. Elsevier Ltd; 2019;162:347–57. Available from: https://doi.org/10.1016/j.watres.2019.06.075

11. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. Nature Publishing Group; 2015;9:207–16.

12. Karkman A, Johnson TA, Lyra C, Stedtfeld RD, Tamminen M, Tiedje JM, et al. High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. FEMS Microbiol Ecol. 2016;92:1–7.

13. Ng C, Tan B, Jiang X, Gu X, Chen H, Schmitz BW, et al. Metagenomic and Resistome Analysis of a Full-Scale Municipal Wastewater Treatment Plant in Singapore Containing Membrane Bioreactors. Front Microbiol. 2019;10:1–13.

14. Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, Singer HP, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J [Internet]. Springer US; 2019;13:346–60. Available from: http://dx.doi.org/10.1038/s41396-018-0277-8

15. Manaia CM, Rocha J, Scaccia N, Marano R, Radu E, Biancullo F, et al. Antibiotic resistance in wastewater treatment plants : Tackling the black box. Environ Int [Internet]. Elsevier; 2018;115:312–24. Available from: https://doi.org/10.1016/j.envint.2018.03.044

16. Lee K, Kim D, Lee D, Kim Y, Bu J, Cha J, et al. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome. Microbiome; 2020;1–14.

17. Martínez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2015;13:116–23. Available from: http://dx.doi.org/10.1038/nrmicro3399

18. Riquelme M V., Garner E, Gupta S, Metch JW, Zhu N, Blair MF, et al. Wastewater Based Epidemiology Enabled Surveillance of Antibiotic Resistance. bioarxiv. 2021;

19. Bürgmann H, Frigon D, Gaze WH, Manaia CM, Pruden A, Singer AC, et al. Water and sanitation: An essential battlefront in the war on antimicrobial resistance. FEMS Microbiol Ecol. 2018;94:1–14.

20. Li AD, Metch JW, Wang Y, Garner E, Zhang AN, Riquelme M V., et al. Effects of sample preservation and DNA extraction on enumeration of antibiotic resistance genes in wastewater. FEMS Microbiol Ecol. 2018;94:1–11.

21. Suzuki MT, Taylor LT. Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5 J -Nuclease Assays Downloaded from http://aem.asm.org/ on November 26 , 2014 by SERIALS CONTROL Lane Medical Library. Appl Environ Microbiol. 2000;66:4605–14.

22. Pei R, Kim S-C, Carlson KH, Pruden A. Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). Water Res [Internet]. 2006 [cited 2018 Jan 10];40:2427–35. Available from:

http://linkinghub.elsevier.com/retrieve/pii/S0043135406002338

23. Hardwick SA, Stokes HW, Findlay S, Taylor M, Gillings MR. Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. FEMS Microbiol Lett. 2008;278:207–12.

24. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

25. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016;2016:1–22.

26. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2014;12:59–60.

27. Larsen N, Brodie EL, Andersen GL, Huber T, DeSantis TZ, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72.

28. Garner E, Chen C, Xia K, Bowers J, Engelthaler DM, Mclain J, et al. Metagenomic Characterization of Antibiotic Resistance Genes in Full- Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. Environ Sci Technol. American Chemical Society; 2018;52:6113–25.

29. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. Genome Biology; 2019;20:1–13.

30. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. PeerJ Comput Sci. 2017;2017:1–17.

31. Langmead Ben, Steven S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2013;9:357–9.

32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

33. Karkman A, Pärnänen K, Larsson DGJ. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. Nat Commun [Internet]. Springer US; 2019;1–8. Available from: http://dx.doi.org/10.1038/s41467-018-07992-3

34. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol [Internet]. Springer US; 2020; Available from: http://dx.doi.org/10.1038/s41587-020-0603-3

35. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.

36. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11.

37. Slizovskiy IB, Mukherjee K, Dean CJ, Boucher C, Noyes NR. Mobilization of Antibiotic Resistance: Are Current Approaches for Colocalizing Resistomes and Mobilomes Useful? Front Microbiol. 2020;11:1–20.

38. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res. Oxford University Press; 2018;46:e35.

39. Douarre PE, Mallet L, Radomski N, Felten A, Mistou MY. Analysis of COMPASS, a New Comprehensive Plasmid Database Revealed Prevalence of Multireplicon and Extensive Diversity of IncF Plasmids. Front Microbiol. 2020;11.

40. Dang C, Xia Y, Zheng M, Liu T, Liu W, Chen Q, et al. Metagenomic insights into the profile of antibiotic resistomes in a large drinking water reservoir. Environ Int [Internet]. Elsevier; 2020;136:105449. Available from: https://doi.org/10.1016/j.envint.2019.105449

41. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, et al. Package "vegan" Community Ecology Package [Internet]. 2019. p. 1–297. Available from: https://cran.r-project.org/web/packages/vegan/vegan.pdf

42. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems. 2018;3:1–9.

43. Rodriguez-R LM, Konstantinidis KT. Estimating coverage in metagenomic data sets and why it matters. ISME J. Nature Publishing Group; 2014;8:2349–51.

44. Ghurye JS, Cepeda-Espinoza V, Pop M. Metagenomic assembly: Overview, challenges and applications. Yale J Biol Med. 2016;89:353–62.

45. Zheng W, Huyan J, Tian Z, Zhang Y, Wen X. Clinical class 1 integron-integrase gene – A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. Environ Int [Internet]. Elsevier; 2020;135:105372. Available from: https://doi.org/10.1016/j.envint.2019.105372

46. Mao D, Yu S, Rysz M, Luo Y, Yang F, Li F, et al. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Res [Internet]. Elsevier Ltd; 2015;85:458–66. Available from: http://dx.doi.org/10.1016/j.watres.2015.09.010

47. Hendriksen RS, Munk P, van Bunnik B, McNally L, Lukjancenko O, Röder T, et al. Global monitoring of antimicrobial resistance based on human sewage. Nat Commun. 2019;

48. Ahmed W, Lobos A, Senkbeil J, Peraud J, Gallard J, Harwood VJ. Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa , Florida. Water Res [Internet]. Elsevier Ltd; 2018;131:142–50. Available from: https://doi.org/10.1016/j.watres.2017.12.011

49. Larsson DGJ, Flach C-F. Antibiotic resistance in the environment. Nat Rev Microbiol. Springer US; 2021;0123456789.

50. CDC. Antibiotic Resistance Threats in the United States, 2019. Cent Dis Control [Internet]. 2019; Available from: http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های نوین&option=com_dbook&task=readonline&book_id=13650&page=73&chkhashk=ED9C9491B4&Itemid=218&lang=fa&tmpl=component

51. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nature [Internet]. Nature Publishing Group;

2014;509:612–6. Available from: http://dx.doi.org/10.1038/nature13377

52. Erica C. Pehrsson1,†, Pablo Tsukayama1,†, Sanket Patel1, 2, Melissa Mejía-Bautista1, 3, Giordano Sosa-Soto1, 3, Karla M. Navarrete3, Maritza Calderon4, Lilia Cabrera5, William Hoyos-Arango3, M. Teresita Bertoli3, Douglas E. Berg6, 7, Robert H. Gilman4, 5, 9. Interconnected microbiomes and resistomes in low-income human habitats. Nature. 2016;176:139–48.

53. Oh M, Pruden A, Chen C, Heath LS, Xia K, Zhang L. MetaCompare : a computational pipeline for prioritizing environmental resistome risk. FEMS Microbiol Ecol. Oxford University Press; 2018;1–9.

54. Zhang A-N, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun [Internet]. Springer US; 2021;12:1–11. Available from: http://dx.doi.org/10.1038/s41467-021-25096-3

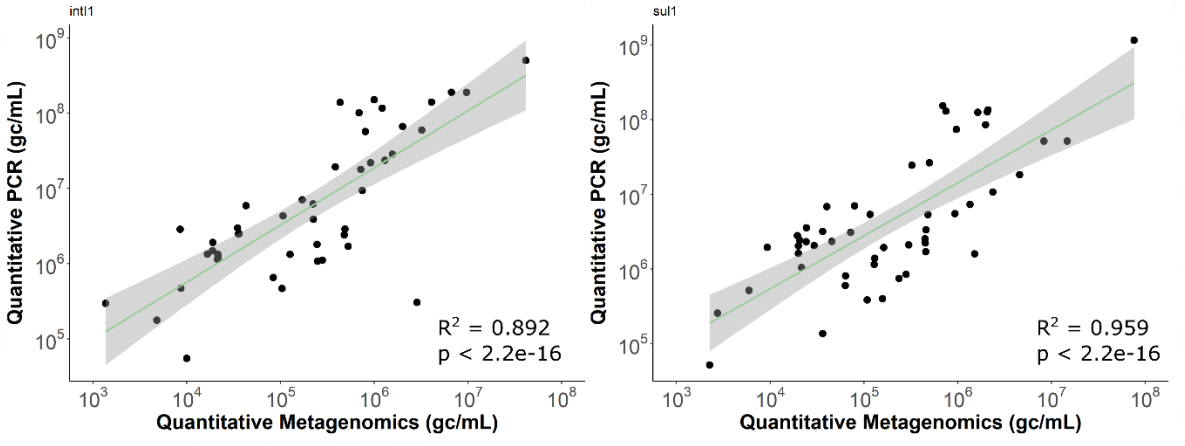


Figure S3-1: Comparison of ARG targets derived by quantitative metagenomics and qPCR.

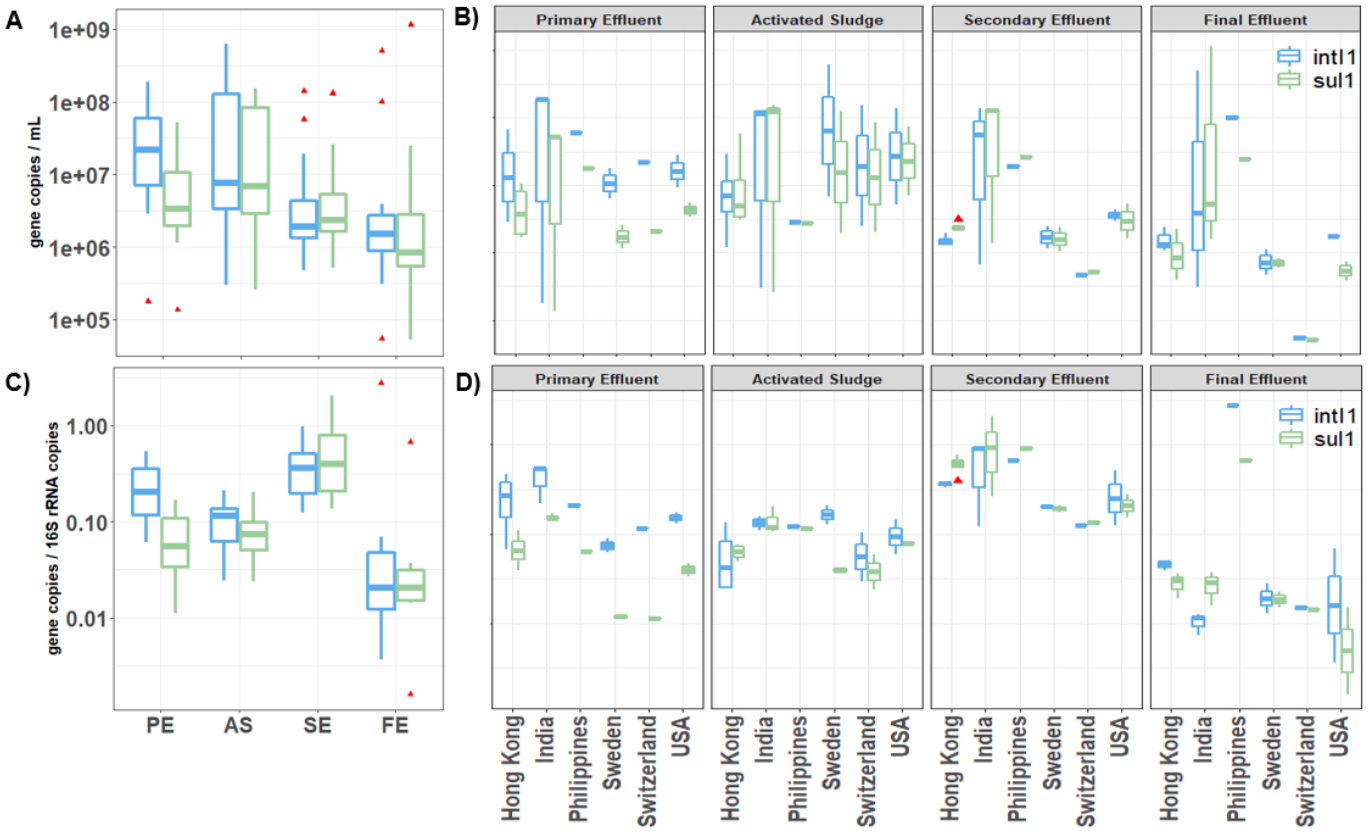


Figure S3-2: Summary boxplots of qPCR data. (A) Abundance of intI1 and sul1 determined via qPCR by treatment process (B) and location. (C) Abundance of intI1 and sul1 normalized to 16S rRNA copies determined vi qPCR by treatment process (D) and location.

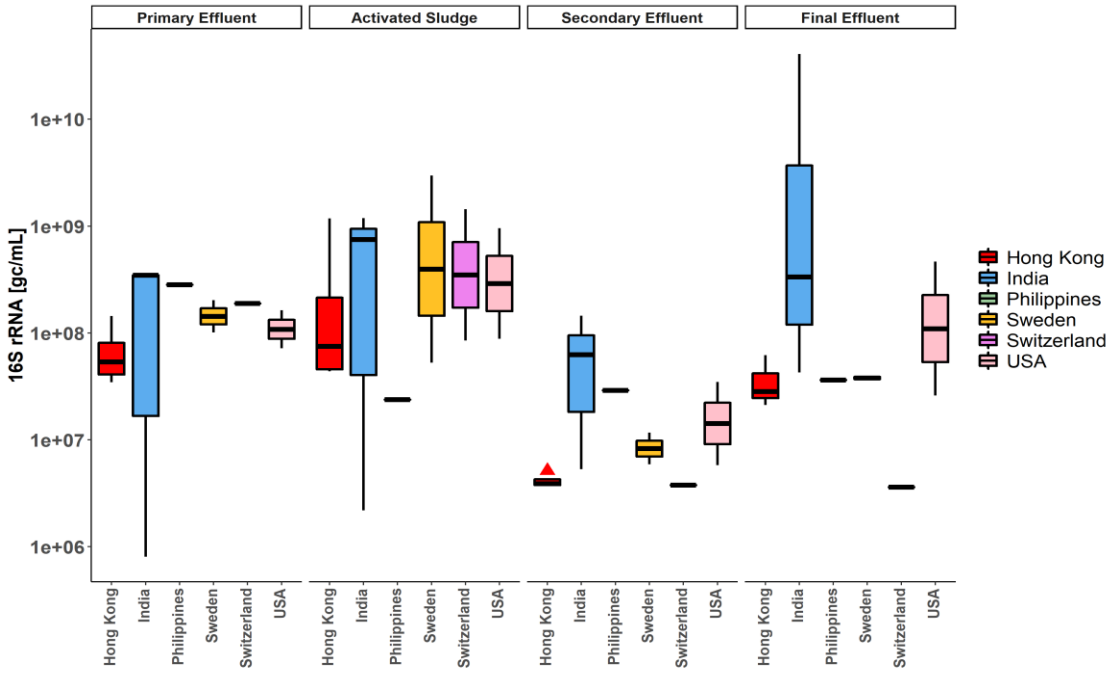


Figure S3-3: Abundance of 16S rRNA gene copies (gc) determined via qPCR deliminated by treatment process and location of WWTP.

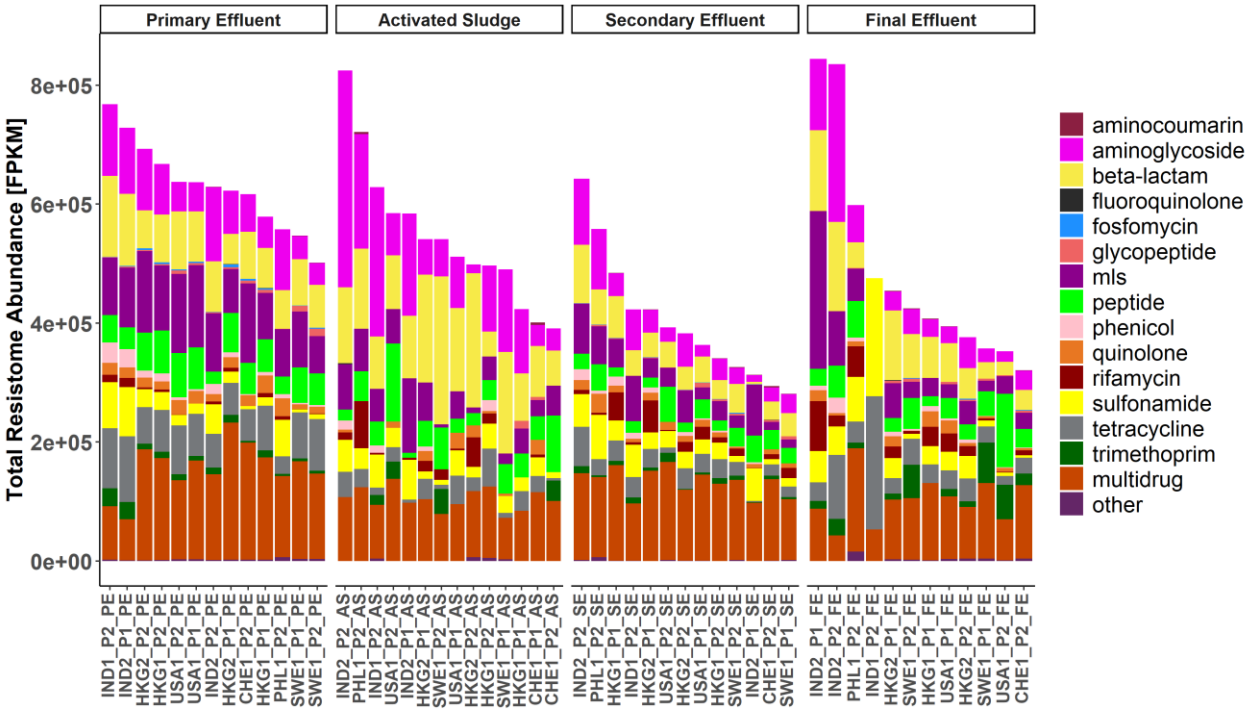


Figure S3-4: Stacked bar chart of FPKM abundance of resistomes across all WWTPs demarcated by antibiotic class. Samples have been ranked ordered by total FPKM abundance to highlight the stratification in abundances across geographies and individual treatment plants.

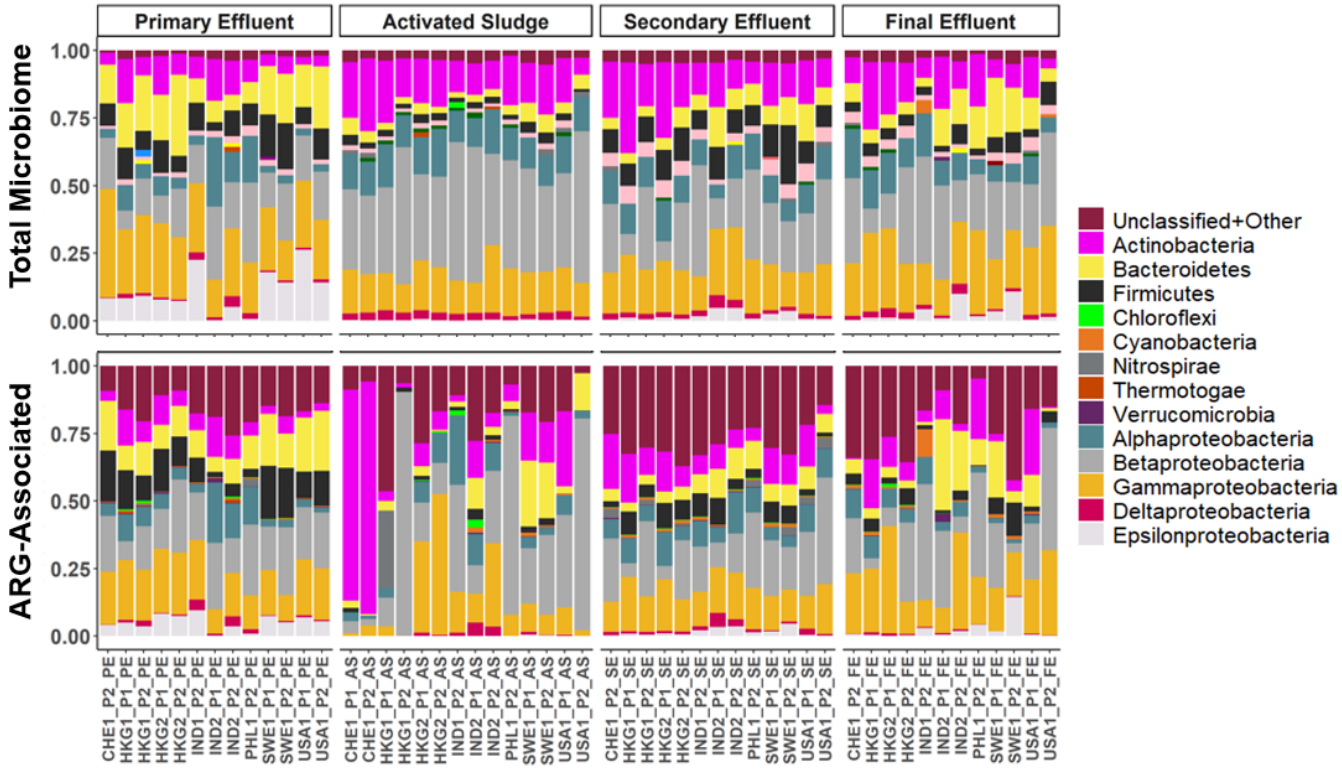


Figure S3-5: (A) Distribution of phyla determined by classification of all metagenomic short reads using Kraken2 and Bracken. (B) Distribution of phyla of all ARG-containing contigs (304,132 contigs). Proteobacteria are split into their individual classes. Any phyla with less than 1% relative abundance were categorized as "Other". Contigs that could not be taxonomically classified in panel B are labeled as "Unclassified".

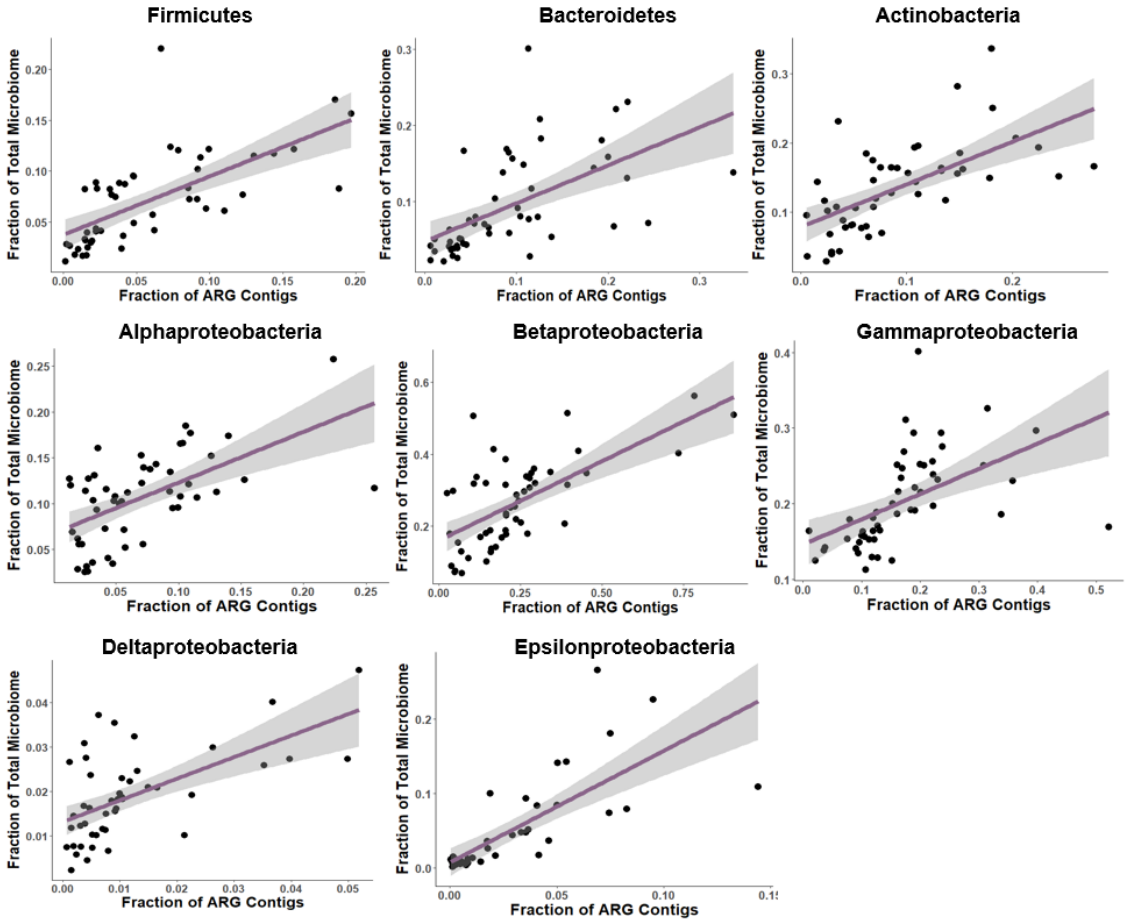


Figure S3-6: Correlations between the proportion of phyla in the total microbiome (Centrifuge classified short-reads) and those found in the ARG-containing contigs (i.e., putative ARB) in each sample. The proteobacteria are split into their individual classes. All correlations are significant at the $p<0.0001$ level (Pearson). These correlations represent a strong correspondence between the short-read and assembled data in recovering the general structure of the microbiome.

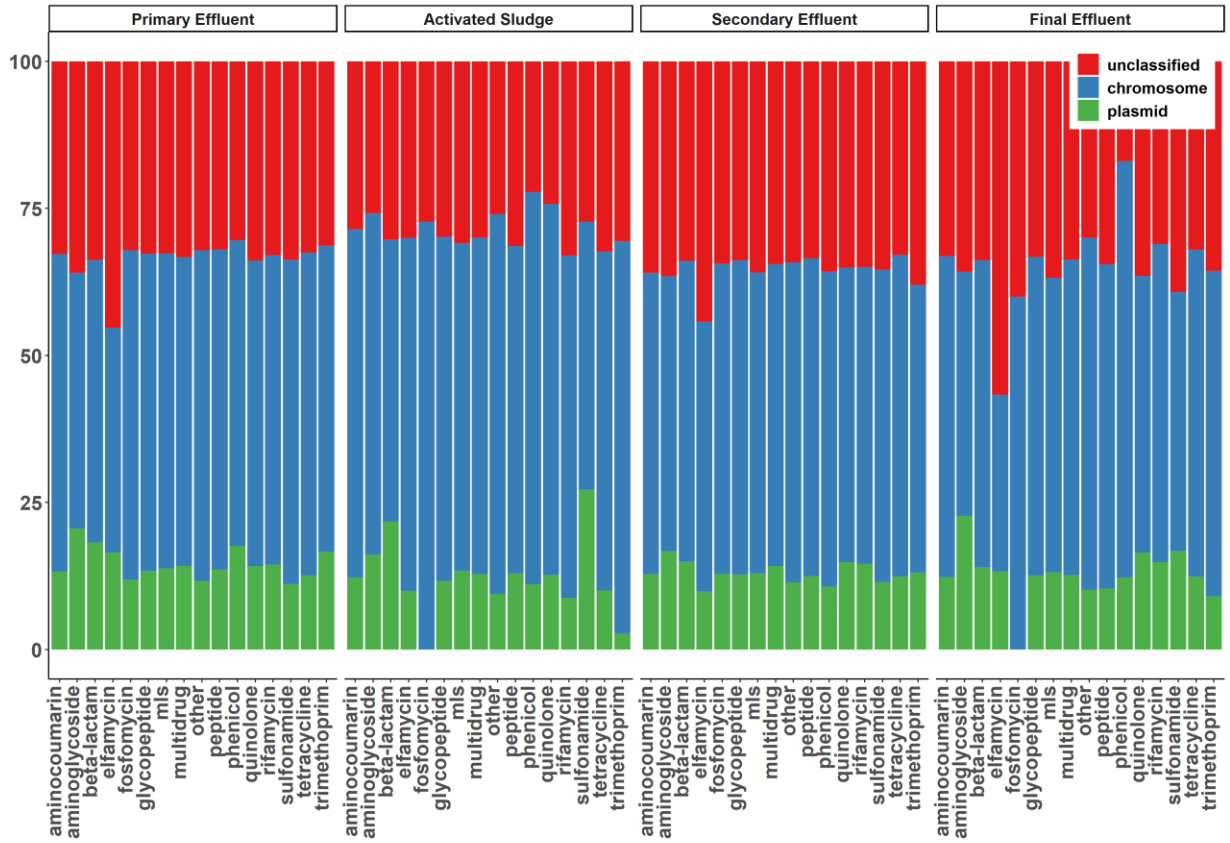


Figure S3-7: Results of PlasFlow annotations of assembled contigs demarcated by treatment compartment and ARG class.

Table S3-1: Overview of WWTP characteristics

Country	Location	WWTP Name	Sample Name	WWTP Configuration	Disinfection
India	Chennai	Nesapakkam	IND_P1	Primary treatment: grit chamber; Secondary treatment: activated sludge with diffused oxygenators	Chlorination
India	Chennai	Perungudi	IND_P2	Secondary treatment: foaming	None
Switzerland	Lucerne	Emmen	CHE_P1	Primary treatment: grit chamber; Secondary treatment: activated sludge (conventional)	None
Switzerland	Zurich	Dubendorf	CHE_P2	Primary treatment: fine rake; Secondary treatment: activated sludge (conventional)	Ozone
Sweden	Boras	Gasslosa	SWE_P1	Secondary treatment: conventional activated sludge (70%), trickling filter (30%)	None

Sweden	Skovde	Skovde	SWE_P2	Secondary treatment: conventional activated sludge	None
Hong Kong	Hong Kong	Sha Tin	HKG_P1	Secondary treatment: conventional activated sludge	Chlorination
Hong Kong	Hong Kong	Shek Wu Hui	HKG_P2	Conventional activated sludge	Chlorination
USA	Christiansburg	Christiansburg	USA_P1	Primary treatment: bar screen, grit chamber, EQ used sometimes (not on day of sampling), primary; Secondary treatment: nitrification/denitrification, conventional AS	UV
USA	Virginia Beach	Hampton Roads	USA_P2	Primary: screening, grit removal, primary clarifier; Secondary: nitrification, denitrification	Hypochlorite
Philippines	Binan	Laguna Water	PHL_P2	Primary: bar rake; Secondary: MMBR (attached growth bioreactor) with aeration basin as polishing step	Chlorine

Table S3-2: List of removed ARGs from the CARD database to reduce potential for detecting wild-types

ARO Accession	gene	function	Resistance Mechanism
ARO:3000124	*mecI*	Negative regulator	antibiotic target replacement protein
ARO:3004185	*mecD*	Variant of PBP2a	antibiotic target replacement protein
ARO:3000815	*mgrA*	Ambigous	efflux pump complex or subunit conferring antibiotic resistance
ARO:3000656	*AcrS*	Negative regulator	efflux pump complex or subunit conferring antibiotic resistance
ARO:3000559	*adeN*	Negative regulator	efflux pump complex or subunit conferring antibiotic
ARO:3000526	*cmeR*	Negative regulator	efflux pump complex or subunit conferring antibiotic

ARO:3000518	CRP	Negative regulator	efflux pump complex or subunit conferring antibiotic
ARO:3000516	*emrR*	Negative regulator	efflux pump complex or subunit conferring antibiotic
ARO:3000746	*mepR*	Negative regulator	efflux pump complex or subunit conferring antibiotic
ARO:3003710	*MexL*	Negative regulator	efflux pump complex or subunit conferring antibiotic resistance
ARO:3000160	*blaI*	Negative regulator	antibiotic inactivation enzyme
ARO:3000817	*mtrR*	Negative regulator	protein(s) and two-component regulatory system modulating antibiotic efflux
ARO:3003318	*Streptomyces rishiriensis parY* mutant conferring resistance to aminocoumarin	Mutant	antibioitc target alteration
ARO:3003784	*Mycobacterium tuberculosis intrinsic murA* conferring resistance to fosfomycin	Mutant	antibioitc target alteration
ARO:3003785	*Chlamydia trachomatis* intrinsic *murA* conferring resistance to fosfomycin	Mutant	antibioitc target alteration
ARO:3003730	*Bifidobacterium ileS* conferring resistance to mupirocin	Mutant	antibioitc target alteration
ARO:3003359	*Streptomyces cinnamoneus EF-Tu* mutants conferring resistance to elfamycin	Mutant	antibioitc target alteration
ARO:3004480	Bifidobacterium_adolescentis_rpoB_mutants_conferring_resistance_to_rifampicin	Mutant	antibioitc target alteration, target replacement
ARO:3000521	Staphylococcus_mupA_conferring_resistance_to_mupirocin	Mutant	antibioitc target alteration
ARO:3000501	*rpoB2*	Mutant	antibioitc target alteration, target replacement

Table S3-3: Pairwise Wilcox Rank Sum tests of absolute abundances (copies/mL) by treatment process across all WWTPs. Wilcox test run with alternative hypothesis set to "greater" and p-values are Bonferroni corrected. P-values < 0.05 are highlighted in green and represent significant reductions. FE-SE highlights in red indicate a significant enrichment from SE to FE.

	PE-AS	PE-SE	PE-FE	FE-SE
aminocoumarin	1.0000	0.1300	1.0000	0.4300
aminoglycoside	0.1160	0.0003	0.0038	0.6904
beta-lactam	0.1746	0.0001	0.0038	0.1282
fluoroquinolone	1.0000	0.0010	1.0000	1.0000
fosfomycin		0.0008	0.0750	1.0000
glycopeptide	1.0000	0.0006	0.0111	0.0246
MLS	0.0007	0.0008	0.0052	0.7833
multidrug	0.0012	0.0014	0.0009	0.6337
other	0.7815	0.0004	0.0424	0.0153
peptide	0.0025	0.0018	0.0152	0.3447
phenicol	1.0000	0.0006	0.0773	0.7140
quinolone	0.0596	0.0006	0.0020	1.0000
rifamycin	1.0000	0.4300	1.0000	0.1800
sulfonamide	0.1990	0.0210	0.4320	0.7150
tetracycline	0.0001	0.0000	0.0009	0.0773
trimethoprim	0.1361	0.0001	1.0000	0.0038

CHAPTER 4: TOWARDS STANDARDIZATION OF METAGENOMICS FOR MONITORING ANTIBIOTIC RESISTANCE IN WATER AND WASTEWATER

Authors: Benjamin C. Davis[1], Jeannette Calarco[2], Krista Liguori[1], Erin Milligan[1], Valerie J. Harwood[2], Amy Pruden,[1] Ishi Keenum[1*]

1 Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia

2 Department of Integrative Biology, University of South Florida, Tampa, Florida

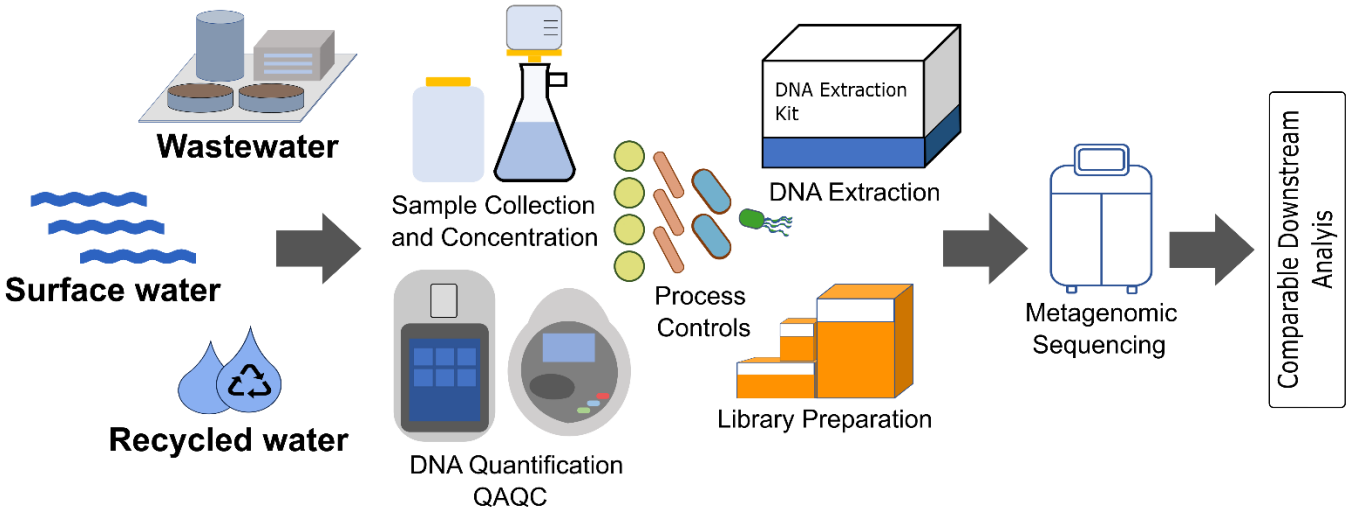
*Corresponding Author

ABSTRACT

Antibiotic resistance (AR) is a rapidly evolving global health threat. The relevance of environmental dimensions to the human clinical problem of AR is increasingly being recognized. Shotgun metagenomic sequencing of the collective genomic information carried across microbial communities is emerging as a powerful approach for monitoring AR in environmental matrices. A tremendous advantage of metagenomics is that all known and putative antibiotic resistance genes (ARGs) (i.e., the resistome), mobile genetic elements (i.e., the mobilome), and other genes of interest can be detected simultaneously, without *a priori* selection of gene targets. However, standardization of metagenomic data collection and processing is needed to support comparability across space and time. To support reproducible downstream analysis, guidance is needed with respect to sampling design, sample preservation and storage, DNA extraction, library preparation, sequencing depths, and experimental controls. Recommendations with respect to databases, sequence-based homology cutoffs, and data normalization strategies are also needed. Here we conducted a critical review to assess current practices for the application of metagenomics for AR profiling of aquatic environments and to offer recommendations to support comparability in the collection, production, and analysis of resulting data.

KEYWORDS: antibiotic resistance, metagenomics, next-generation sequencing, standardization, wastewater, surface water

GRAPHICAL ABSTRACT



INTRODUCTION

Antibiotic resistance (AR) is a growing global health threat (World Health Organization 2015; O'Neill 2016) and, increasingly, the importance of environmental dimensions to its transmission and evolution are being recognized (European Commission 2017; United Nations 2017). Correspondingly, the need for unified approaches to assessing AR in the environment is becoming apparent (Huijbers et al. 2019; JPIAMR 2019). Environmental monitoring can help to assess baselines of AR in pristine and anthropogenically-influenced environments, as well as local human and animal populations, and further aid in identifying high-risk areas for the evolution, selection, and transmission of antibiotic resistant bacteria (ARB) (Berendonk et al. 2015; Larsson et al. 2018; Pruden et al. 2021). Such information promise to be especially valuable towards informing specific policy/mitigation measures (Aarestrup and Woolhouse 2020). Monitoring of influent sewage to wastewater treatment plants (WWTPs) has especially garnered attention as a means to capture collective antibiotic resistance genes (ARGs) circulating amongst the corresponding human population and has been shown to reflect local clinical prevalence of ARB (Pärnänen et al. 2019). The WWTP itself also represents a significant barrier to the dissemination of ARB and ARGs to receiving surface waters and therefore removal efficiencies are also of interest.

Next-generation sequencing (NGS) is emerging as a highly powerful and promising tool for water and wastewater monitoring (Hendriksen et al. 2019; Garner et al. 2021a). Shotgun metagenomics applies NGS for the sequencing of DNA extracted across microbial populations inhabiting an environmental sample (e.g., water or wastewater). The resulting metagenomic library can be analyzed to characterize the resistome (i.e., the collective ARGs carried across a microbial community). The most common approach is to compare the metagenomic library against publicly-available databases to identify functionally verified ARGs, which currently number in the thousands (Alcock et al. 2020). The number of ARGs and different types can then be compared across samples of interest. Detected ARGs can be classified and ranked by various means; this includes the antibiotics they encode resistance to, the mechanism of resistance, and their degree of clinical relevance (i.e., extent to which they are found to interfere with treatment of human infections). The genetic context of various ARGs can then be explored using *de novo*

short-read, long-read, or hybrid assembly algorithms to predict whether they are intrinsic or mobile, what kinds of mobile genetic elements (MGEs, e.g., plasmids, integrons, transposons) they are carried on, or whether they occur in known human pathogens or the commensal environmental flora. Metagenomics is also being evaluated as a means of identifying recently evolved or mobilized ARGs before they become problematic in the clinic (Arango-Argoty et al. 2018; Berglund et al. 2019).

Because NGS approaches to water quality monitoring have only arisen relatively recently, the data generation techniques, software, and databases that are used are still in a period of rapid innovation and expansion. This has led to inconsistent data generation and reporting, which hinders the ability to compare measurements across studies. The overall purpose of this critical review was to assess the potential for a shotgun metagenomic approach to AR monitoring of wastewater and other impacted aquatic environments and to provide a framework towards standardizing approaches and ensuring comparability of the data and downstream analyses. Similar initiatives for consensus data reporting guidelines and improved reproducibility have recently been proposed in other fields of NGS (Mirzayi et al. 2021). The specific objectives of this review were to (1) evaluate strengths and weaknesses of existing workflows for metagenomic analysis of water and wastewater resistomes, (2) identify sources of variability introduced by different data processing techniques, including sequencing depth and coverage, (3) identify opportunities for standard reporting metrics, and (4) propose standardized workflows for generating meaningful and reproducible metagenomic data. We further provide a projection of the field as it moves toward standardized data reporting, NGS process controls, and datatypes for integration into future risk assessment models. The overall recommendations provide a framework to support the representativeness and comparability of metagenomic data and analysis for the purpose of AR monitoring of aquatic environments.

LITERATURE REVIEW PROTOCOL

To generate a systematic critical review, a three-tiered approach was used to establish search terms (**Table S1**). Tier 1 established topic level keywords that identified studies that were relevant to water reuse, wastewater, and surface water environments. Tier 2 ensured that the studies were relevant to AR, while Tier 3 established keywords to identify studies focusing on metagenomics and NGS applications. Literature returned via this search strategy were manually screened by two independent researchers to ensure that all included papers fully met the search criteria. Articles focused on aquaculture, biosolids and biosolid treatment (anaerobic digestion, composting, etc.), or laboratory-scale experiments were excluded. Studies involving only the use of secondary data (i.e., a meta-library of metagenomic sequences from other studies) were also excluded. Any disagreements between the two screeners on relevance were presented to a broader team of 5 researchers to reach a consensus on the applicability of the study towards informing a workflow that includes sampling, DNA sequencing, and data analysis. This approach produced 95 articles for inclusion in this review. Studies that met eligibility criteria were subjected to data extraction for the parameters outlined in **Table S2**.

DATA EXTRACTION AND ANALYSIS

All publicly-available metagenomes from the 95 retrieved articles were downloaded from the Sequence Read Archive (SRA), European Nucleotide Archive (ENA), and MG-RAST servers (n=1775) for depth and coverage analysis. We then filtered for paired-end Illumina

datasets pertaining to wastewater, surface water, and recycled water matrices using available metadata (n=1440). Briefly, we cleaned paired-end data using Trimmomatic (Bolger et al. 2014) (leading:3, tailing:3, slidingwindow:4:15, minlen:36) and ran forward reads through Nonpareil (Rodriguez-R et al. 2018) with option "-T kmer" to determine the relative metagenomic coverage as function of library size. Data were analyzed and visualized in R (v 4.1.2) using ggplot2 (Wickam 2009).

ESTABLISHING WATER AND WASTEWATER RESISTOME MONITORING OBJECTIVES

To ensure that metagenomic sequencing is optimally applied, it is especially critical to be clear on the objectives of the monitoring program and that the sampling design can achieve those objectives. This point was emphasized in a recent study that drew from an expert survey and workshop to develop a framework for culture-, qPCR-, and metagenomic-based monitoring of AR in water environments (Liguori et al. 2022). Here we elaborate upon monitoring objectives for which metagenomics is particularly well-suited, drawing from other recent reviews that have generally addressed the need to harmonize environmental AR monitoring efforts (Berendonk et al. 2015; Huijbers et al. 2019; Aarestrup and Woolhouse 2020; Larsson and Flach 2021).

1. Monitor ARB and ARGs circulating in human populations through municipal and hospital wastewater systems (i.e., wastewater-based surveillance)
2. Comprehensively quantify attenuation or amplification of ARGs of clinical concern through wastewater or recycled water treatment processes and determining removal efficiencies
3. Identify transmission pathways of ARGs that escape engineering controls and their potential to be assimilated by pathogens and move across human, animal, and environmental matrices
4. Assess the evolution and mobilization of new ARGs and pathogen hosts via biological wastewater treatment and various sources of anthropogenic pollution

As described in subsequent sections, the objectives will consequently dictate decisions in metagenomic workflows, particularly regarding short- versus long-read sequencing strategies and the analytical advantages provided by each. Figure 4-1 illustrates elements to consider in a standard metagenomic workflow and subsequent sections provide further detail regarding the inherent biases and considerations that may affect downstream data comparability and hypothesis testing.

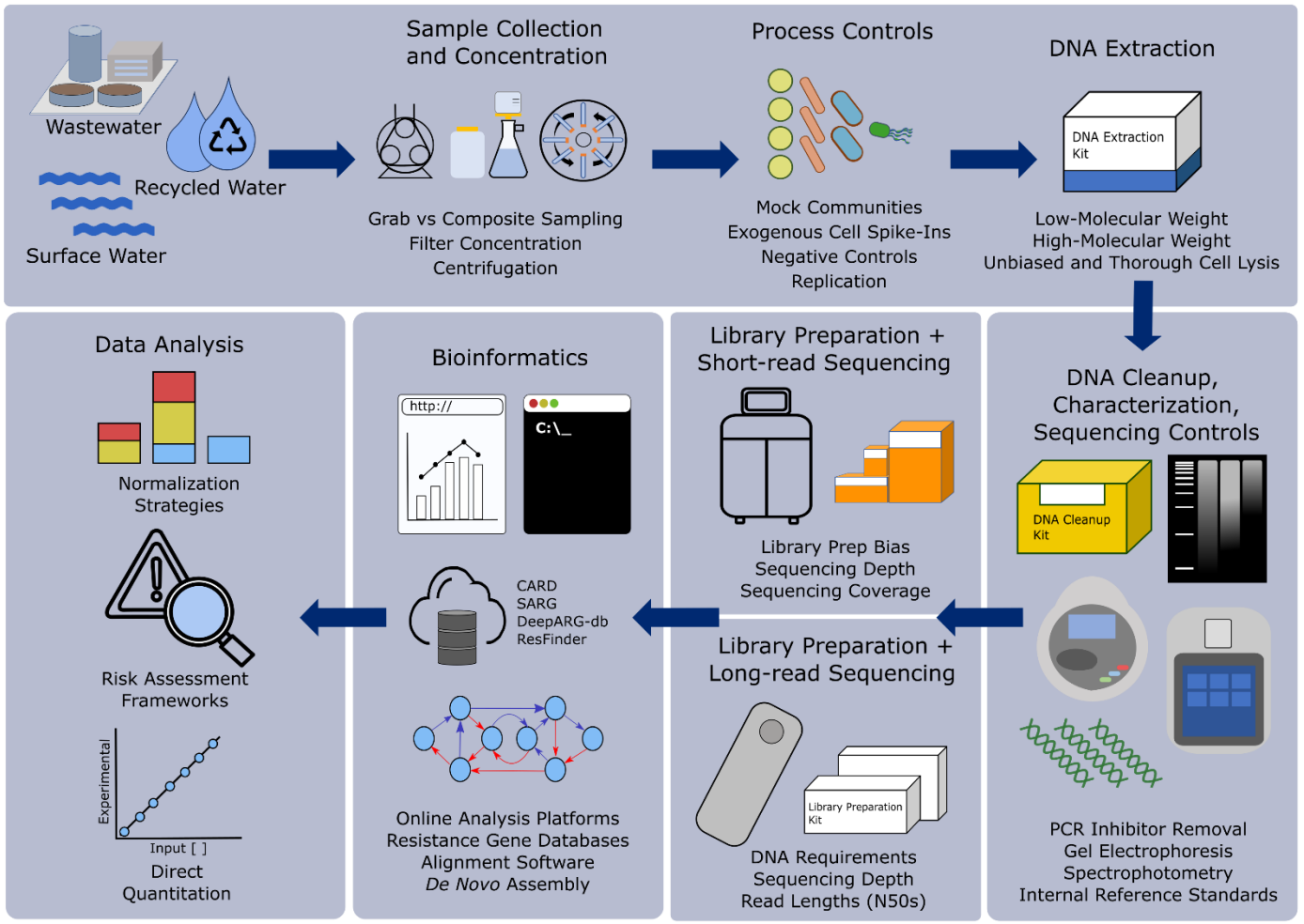


Figure 4-1: Overview of key decision points to consider in developing a workflow for metagenomic-based monitoring of ARGs in aquatic environments that are highlighted in this review.

SAMPLING AND PROCESSING

Sampling Frequency, Replication, And Controls

A growing body of research is providing insight into baseline variability of WWTP and other aquatic system resistomes (Yin et al. 2019; Majeed et al. 2021). In a landmark study of a Hong Kong WWTP, monthly sampling of the activated sludge basin was performed over a nine-year span and the resistome composition was found to turnover every 2-to-3-years (Yin et al. 2019). However, it is unclear the extent to which the observed patterns are generalizable across all WWTPs or how such dynamics vary with each stage and type of treatment. Coordinated surveillance is needed to help inform the sampling frequency and number of replicates needed to achieve metagenomic monitoring objectives. For example, if influent sewage resistomes are relatively stable across WWTPs with time, as observed in one particular conventional WWTP in the US (Majeed et al. 2021), then less frequent sampling may be necessary when the purpose is broader resistance monitoring of human populations (Objective 1). However, if the purpose is to

determine if anomalous ARGs of clinical concern are present in the influent and escaping into surface waters, then much more frequent sampling with replication and deep sequencing may be necessary (Objective 2) (Majeed et al. 2021).

The need for biological replication, in this case sequencing of multiple samples representing a given condition, is a critical consideration for any methodology and depends on the statistical power needed to capture the signal of interest and test a given hypothesis. Among the articles identified in this study, replicate sampling and sequencing was universally lacking (Figure 4-1, 2). The high per-sample cost of generating metagenomes is likely a major factor in this deficiency. Also, ecological surveys tend to favor breadth over depth in sampling design. Limitations of this nature are also inherent to field ecology in general (Filazzola and Cahill 2021). Sequencing of at least one condition in triplicate would help in revealing systematic biases in sample processing and support the statistical power needed to distinguish signal from stochastic variation and sequencing bias. The inclusion of negative controls (e.g., field, trip, and extraction blanks) were also absent. Negative controls serve as a check for any contamination events that occur during sampling, processing, and DNA extraction that may contribute to background detection of microbes and ARGs **(Figure 4-2**). This is especially useful in differentiating low abundance taxa or ARGs from technical noise or laboratory contamination (Borchardt et al. 2021).

In the studies examined in this review, samples were almost exclusively taken as grab samples. Composite samples, which may be flow- or time-weighted, may be more appropriate where replicate grab samples are infeasible (Centers for Disease Control and Prevention 2020). Studies evaluating time-sensitive wastewater-based surveillance of illicit drugs (Rodayan et al. 2014), total phosphorus and nutrients (Johannessen et al. 2012), and SARS-CoV-2 (Kopperi et al. 2021) demonstrated that time-weighted composite and grab samples yielded highly comparable results. This suggests the relative stability of wastewater compositions and treatment efficacy over diurnal timescales such that a reasonable degree of replication should be able to capture signals of interest.

Sample Preservation and Storage Affect Sample Representativeness

Appropriate preservation and storage ensure that subsequent analysis is representative of the sample at the time it was collected. This is particularly critical for time series data and comparisons across systems. A recent comprehensive analysis of storage conditions of raw pig feces and domestic wastewater samples revealed systematic biases that impacted downstream metagenomic analysis (Poulsen et al. 2021). The authors found that both storage time (immediate processing, 16 hrs, 64 hrs; and long-term storage at 4, 8, and 12 months) and temperature (deep freezer, -80°C; freezer, -20°C; refrigerator, 5°C; room temperature, 22°C) resulted in significant fluctuations in taxonomic and resistome composition; although if immediately frozen (at either -20°C or -80°C), batch effects were minimized. If freezer storage is not possible, the authors stressed that samples should be processed immediately. The need to immediately freeze or analyze the sample poses a challenge when seeking to include low income countries in global-scale studies (Hendriksen et al. 2019). Where the shipping of samples is necessary, fixing samples in 50-100% ethanol, freezing at -20°C, and shipping on ice has shown to both prevent significant fluctuation of resistomes and preserve the integrity of DNA (Li et al. 2018a). Other sample preservation reagents (e.g., Zymo DNA/RNA Shield) have been shown to preserve the integrity of soil microbiomes (Pavlovska et al. 2021), human microbiomes (Bartolomaeus et al. 2021), and fecal SARS-CoV-2 RNA (Natarajan et al. 2021), even at room temperature; although

these techniques have not been systematically assessed for analysis of aquatic resistomes. The addition of preservation reagents may also preclude sub-sampling the same sample for multiple analyses (e.g., transcriptomics, metabolomics, cultivation, pharmaceuticals analysis) (Poulsen et al. 2021).

Sample Concentration Techniques

Sample concentration serves to recover an adequate mass of microbial cells for analysis and ideally should be applied in a way that maintains representativeness of the corresponding microbial populations. Low mass samples yield low concentrations of DNA, which may preclude library preparation or necessitate amplification, which may introduce bias (see Library Preparation). Among the articles identified in this review, the two most common sample concentration methods were membrane filtration (62 studies) and centrifugation (23 studies). Membrane filtration was more common for less turbid waters (e.g., drinking water, river water, final treated wastewater effluent) and centrifugation was more common for more turbid waters (e.g., raw wastewater, activated sludge).

The most applied membrane pore sizes amongst the identified studies were 0.2 µm (12 studies), 0.22 µm (37 studies), and 0.45 µm (12 studies). Because the smallest prokaryotic cell diameter is approximately 0.2 µm, a pore size approaching that threshold will allow for the adequate representation of the bacterial and archaeal composition of a given water sample. The tradeoff of smaller pore sizes is that less water will be able to pass through due to clogging, decreasing the representative volume and increasing the detection limit. Pre-filtration of environmental samples using larger pore size membranes (1.2 µm – 1 mm) was common to reduce particulates before passing through subsequent filters, increasing the representative sample volumes. However, this effectively eliminates the particle-bound fraction of the microbiome and may significantly alter representativeness of the sample. For samples with extremely low cell densities and/or turbidity (e.g., advanced water treatment products), ultrafiltration is a means to concentrate volumes up to 100 liters, although this may still be insufficient for recovering enough nucleic acid for sequencing extremely clean samples (Stamps et al. 2018). Centrifugation workflows typically involved pelleting biomass from raw wastewater at 4,000 to 15,000xg. The supernatant is discarded, and the pellet is resuspended in buffered solution that is either then passed through an additional 0.22-µm membrane or directly subject to DNA extraction. Sample concentration techniques are unlikely to influence the representativeness and comparability of generated metagenomic libraries investigating bacteriomes, although researchers should strive for uniformity within individual experiments.

DNA EXTRACTION DICTATE REPRESENTATIVENESS AND COMPARABILITY OF METAGENOMES

Because no DNA extraction approach is 100% efficient or unbiased, DNA extraction methodologies should be consistent across sample sets intended to be compared by metagenomics. This can be challenging when seeking to compare metagenomic data across published studies, especially as DNA extraction kits and procedures continue to evolve. At a minimum, DNA extraction methods need to be reported in associated metadata so that they can be accounted for in any future meta-analyses. Ideally, a positive control sample should be included to identify potential biases in the extraction. However, mock community controls, either as standalone samples or as exogenous cell spike-in process controls (Figure 4-2), were almost entirely absent from workflows reported in the identified literature (Uyaguari-Díaz et al. 2018). Generally, these controls are comprised of known mixtures of organisms with varying

recalcitrance to cell lysis (e.g., Gram-positive versus Gram-negative bacteria) and thus serve to assess the efficiency of the DNA extraction method and give insights into the representativeness and reproducibility of NGS workflows. These process controls are standard practice in many fields of molecular biology, the most recent example being the inclusion of Bovine Coronavirus as a surrogate RNA extraction control in the wastewater monitoring of SARS-CoV-2 (Natarajan et al. 2021).

Recently, mock communities have been used to highlight DNA extraction and bioinformatic workflow bias in interlaboratory studies (Han et al. 2020; O'Sullivan et al. 2021). Irreproducibility has even been shown in replicated metagenomics work within individual experiments due to batch effects across sequencing runs (Yeh et al. 2018). Mock communities and process controls help to assess reproducibility across space, time, and laboratory groups and should be included during submission to public data repositories to evaluate the representativeness and comparability of metagenomes used in meta-analyses. However, these communities are typically much simpler (less diverse with less inhibition) than the target environment of interest and therefore cannot fully reproduce the sampling environment.

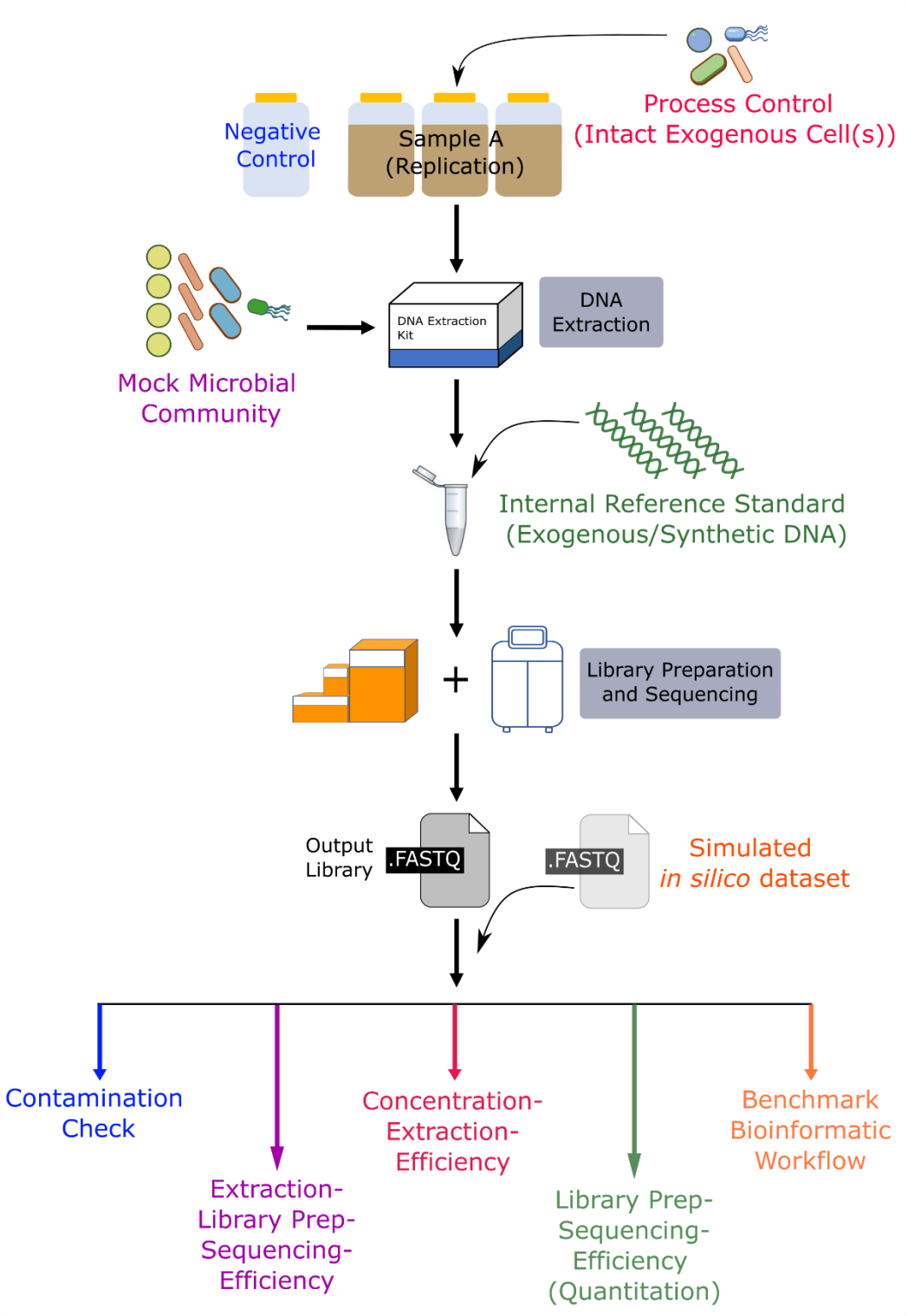


Figure 4-2: Framework of process controls for metagenomic investigations of environmental AR.

Bead Beating Kits Are Ideal for Short-Read Metagenomics

Across the studies identified, almost all DNA extractions were performed using commercial kits that employ both chemical lysis and bead beating, along with purification through a spin column (93%). The most popular were the FastDNA Spin Kit for Soil (36 studies), PowerWater or PowerSoil Kits (30), and the QIAamp DNA Stool Mini Kit (6). Previous studies comparing the efficacies of commercial DNA extraction kits for metagenomic sequencing found that the FastDNA Spin Kit for Soil (MP Biomedicals) generated the highest yield and purity of DNA from three commonly sampled WWTP compartments (influent, activated sludge, final effluent), resulting in the detection of the greatest diversity of ARGs when compared using an Illumina sequencing platform (Guo and Zhang 2013; Li et al. 2018a). Two main distinctions of the FastDNA Spin Kit for soil are that it employs a range of bead diameters, and that the DNA is suspended with the binding matrix during isolation as opposed to the binding matrix being confined to the spin column. A modified standard protocol using the QIAamp DNA stool Mini Kit (Qiagen) has also proven to be a popular and unbiased approach for aquatic resistome sampling that uses both mechanical and enzymatic lysis (Knudsen et al. 2016). These approaches aim to evenly lyse both Gram-negative and Gram-positive cells using a combination of high shear forces, enzymatic lysis of cellular membranes, and chemical precipitation of protein debris and are near ideal for large-scale environmental monitoring projects due to their ease of implementation and reproducible results.

High Molecular Weight DNA Extraction Optimizes Long-Read Sequencing

The above-cited studies were conducted for optimization of short-read sequencing platforms and therefore DNA damage during extraction is less of a concern. While bead-beating can reduce bias in DNA recovery, it also shears and fragments DNA (Quick and Loman 2019). Commercial spin column kits with bead-beating generally produce fragment lengths $\leq$ ~60 kbp (Quick and Loman 2019). Short and damaged DNA fragments can be detrimental to optimized long-read sequencing which preferentially sequence shorter sequences at higher molarity and thus high-molecular weight (HMW) DNA extraction methods should be prioritized. For instance, the traditional phenol-chloroform method can recover DNA with average fragment lengths approaching 150 kbp and maximum fragment lengths > 1 Mbp, although this method is inefficient for large numbers of samples and utilizes carcinogenic reagents. Several commercial HMW kits have been developed but have not been fully benchmarked for complex environmental matrices or resistome analysis. It should further be noted that minimum per-sample DNA inputs of 1-2 µg are required for long-read Nanopore sequencing, and this may be difficult to obtain from some aquatic sample types.

DNA Quality Control Necessary for Successful Library Preparations

Additional purification steps for the removal of PCR inhibitors from DNA extracts was uncommon among the reviewed workflows. Common PCR inhibitors; such as humic/fulvic acids, tannins, melanin, and lingering reagents from DNA extraction, have been shown to interfere with NGS library preparation and inhibit loci typing (Sidstedt et al. 2019, 2020). Inhibitors such as EDTA and other salts can also cause library preparation failure. DNA sequencing cores commonly determine the quantity and purity of submitted DNA extracts as a prerequisite for sequencing. A minimum of 1 ng of DNA per sample is generally acceptable for PCR-based library preparation (see Library Preparation). Among the reviewed articles, DNA was quantified using three different platforms: Qubit™ dsDNA HS Assay Kit (Life Technologies, 27 articles), Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, 5 articles),

and the NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific, 24 articles). Qubit and PicoGreen assays use fluorescing dyes that are highly specific to double-stranded DNA and accurately quantify 10 pg/µL to 100 ng/µL. NanoDrop uses spectrophotometry to assess the absorbance profiles of nucleic acids, proteins, and other contaminants. A 260 nm/280 nm (DNA/protein) optical density ratio (OD) of 1.8 to 2.0 is considered high quality DNA suitable for library preparation. Gel electrophoresis is also commonly used to assess the distribution of molecular weights of extracted DNA and provides a quick quality check to ensure HMW extracts. Sequencing cores will typically run gels on samples before and after shearing and after adaptor ligation using a TapeStation (Agilent) to ensure the correct insert sizes of the final library. Twenty-six of the ninety-five identified articles did not report quantification or QA/QC of their DNA extracts.

LIBRARY PREPARATION AND SEQUENCING

Different Sequencing Platforms Achieve Different Monitoring Objectives

Roche 454 Pyrosequencing was the first highly parallelized platform (released in 2005) applied for shotgun metagenomics in environmental research (Barba et al. 2013), but has since been discontinued, with the Ion Torrent (Thermo Fisher) (released in 2010) and Illumina sequencing platforms (MiSeq released in 2011) still in use today. These technologies all yield relatively short reads (75-300 bp for Ion Torrent and Illumina and 800 bp for 454) (Metzker 2005). Long-read sequencing, including PacBio (Pacific Biosciences) and Nanopore (Oxford Nanopore Technologies) platforms, entered the market more recently and are advantageous when the objective is to examine the genetic context of ARGs with greater accuracy (i.e., their association with MGEs and host organisms) (Objectives 3 & 4). Short-reads are limited in this regard because they must be assembled into longer contigs in order to examine neighboring genes, which introduces substantial uncertainty and bias (Bengtsson-Palme et al. 2017) (See below section on Metagenomic Assembly for ARG Contextualization). The tradeoff is that long-read sequencing tends to be relatively shallow (5.4 Gbp maximum identified in this review), while deep Illumina sequencing was reported to reach 77.5 Gbp (Liu et al. 2019) for wastewater samples and thus can more comprehensively profile ARGs (Figure 4-3). The base error rates for Nanopore platforms are also higher (1-20%) (Sahlin et al. 2021) compared to Illumina (~0.1%) (Stoler and Nekrutenko 2021). Studies surveyed indicated recovery of 1-500 million reads per sample for Illumina sequencing, while reports of Nanopore sequencing of aquatic matrices to date were in the tens to hundreds of thousands, limiting the absolute number of genomic inquiries per sample. These issues of sequencing depth and read lengths ultimately factor into the degree of sample coverage achievable by each platform (i.e., the fraction of the total genomic information from the microbial community that was sequenced) (Figure 4-3). However, with the advent of newer Nanopore (GridION and PromethION) and PacBio platforms (Sequel II with HiFi reads), this gap in depth, error rate, and ultimately sample coverage will continue to shrink between long-read and short-read platforms.

Among the studies identified by the search criteria, 89% utilized Illumina sequencing, 4% Oxford Nanopore sequencing, 3% Ion Torrent sequencing, and 3% Roche 454 Pyrosequencing. Thus, current understanding of optimal conditions for metagenomic monitoring of AR in water and wastewater systems is largely based on what has been learned from Illumina sequencing. However, it is important to also look to the future as long-read DNA sequencing is rapidly gaining ground and presents many advantages for certain monitoring objectives, specifically assessing the mobilization and host-context of ARGs (Che et al. 2019; Dai et al. 2022).

Hereafter, we primarily focus on what is known based on Illumina sequencing but point out distinctions and opportunities related to long-read sequencing where relevant.

Library Preparation Techniques Exhibit Inherent Biases

There are several commercial kits available for library preparation. Library preparation generally comprises three steps: DNA fragmentation to a uniform insert size (enzymatic or mechanical), repairing and end polishing of fragmented DNA, and ligation of platform-specific adaptors (Sato et al. 2019). Illumina library preparation was almost exclusively performed by the core facility performing the sequencing analysis. Consequently, available options are often restricted to the research facility providing the service.

There are two main categories of library preparation, PCR-free and PCR-based, with the latter introducing biases associated with PCR amplification. The choice between the two is typically a function of available sample DNA, where a threshold mass is required (~25 ng) for PCR-free preparations. PCR-based library preps, like the Nextera XT DNA Library Preparation Kit, use a transposome complex to simultaneously shear and ligate adaptor sequences to fragments (tagmentation) (Sato et al. 2019). Research by (Bowers et al. 2015) documented the effect of input DNA quantities and library preparation methods on the ability to reconstruct a mock community consisting of pre-extracted genomic DNA. Input DNA quantities reaching as low as 1 picogram could successfully pass library preparation using PCR-based kits, but bias towards GC rich sequences was apparent as DNA inputs fell below 1 ng, as compared to a control generated with the PCR-free TruSeq kit and 200 ng of DNA. (Sato et al. 2019) carried out a similar study and also found that PCR-based kits were unable to accurately reflect extremes in genomic GC content. The most variable reconstructions of mock communities were derived from the Nextera XT and TruSeq nano kits, presumptively due to non-random DNA fragmentation during sonication and PCR amplification. Other kits, including the newer Nextera DNA Flex (now simply Illumina DNA Prep) and the TruSeq DNA and KAPA HyperPlus PCR-free workflows, reconstructed statistically identical mock communities, even at a shallow sequencing depth (~1 Gb) (Sato et al. 2019). These studies indicate that PCR-free library prep is the best option, but that newer PCR-based methods can help to reduce bias observed in previous generation kits. Regardless, metagenomes will be most comparable when generated from the same library prep method.

Library preparations for long-read sequencing, specifically on Nanopore platforms, were done in-house and are less flexible. All four articles identified in this review used the SQK-LSK108 1D ligation genomic DNA kit (Oxford Nanopore Technologies) in-house, with 1 – 2 µg of input DNA per sample for sequencing on the MinION platform (Che et al. 2019; Hamner et al. 2019; Białasek and Miłobędzka 2020; Yadav et al. 2020). This library preparation generally involves four steps: end-repair of extracted DNA, Nanopore-specific adaptor ligation, barcoding, and purification. As noted above, the DNA extraction strategy employed will determine the suitability of DNA fragment size distributions for long-read sequencing. A study conducted by (Che et al. 2019) used a bead-beating and spin column DNA extraction approach for wastewater samples and then selected DNA fragment sizes > 8 kb by manually excising them from an agarose gel for library preparation. They then compared long-reads with sequenced and assembled Illumina data and found that the average N50 from Nanopore was 8.1 kbp (average depth 3.4 Gb), compared to 1.7 kbp from Illumina (14.5 Gb). All four articles used a bead-beating and spin column DNA extraction approach for long-read sequencing, but as HMW extraction techniques continue to emerge (Maghini et al. 2021), reconstruction of complex

microbial communities and optimization of long-read sequencing from environmental samples will continue to improve.

Sequencing Parameters Dictate Depth And Coverage Of Metagenomes

When selecting a sequencing technology and associated parameters, the platform, target read length, and depth per sample must be considered. The primary unit of "currency" for sequencing platforms is the flow cell. Each flow cell contains one to multiple lanes, or physically partitioned regions of the solid surface that enable multiple experiments to be run in parallel, but independent of each other (i.e., without cross contamination). For example, the NovaSeq 6000 platform (Illumina) has SP (2 lanes, ~ 800M reads per lane), S1 (2 lanes, ~1.5B reads per lane), S2 (2 lanes, ~1.8B reads per lane), and S4 flow cells (4 lanes, ~2.5B reads per lane). These flow cells can then be run with varying paired-end sequencing chemistries (i.e., 50-250 bp reads), which dictates the number of base pairs generated for each experiment. The number of reads generated per flow cell (and therefore per lane) are fixed, meaning the number of reads generated per sample will be a function of the number of samples multiplexed on that flow cell.

There is typically a need to strike a balance between the depth of sequencing and level of replication needed to achieve monitoring objectives, while also bearing in mind cost. The level of microbial diversity anticipated in the sample and the need to detect rare sequences and taxa will both drive the need for deeper sequencing. Careful consideration is needed when choosing sequencing depths. Comparing two environmental samples with significant differences in coverage severely inhibits accurate and ecologically-relevant insights into microbiome and resistome dynamics (Rodriguez-R and Konstantinidis 2014; Zaheer et al. 2018; Gweon et al. 2019). Metagenomic sequencing preferentially sequences the most abundant features, and shallow sequenced datasets are severely disadvantaged in their ability to detect differentially abundant features at low abundances (Rodriguez-R and Konstantinidis 2014). To provide guidance with respect to sequencing depths, we empirically estimated metagenomic coverage as a function of library size using Nonpareil (Rodriguez-R et al. 2018) across all publicly-available paired-end Illumina data from the studied articles (n=1440) (Figure 4-3). We then used the generated models to predict coverage at a depth of 10 Gb and found an average coverage of 0.779 across all water types (Influent = 0.794, Activated Sludge = 0.774, Effluent = 0.849, Recycled Water = 0.810, Surface Water = 0.757). The authors of Nonpareil observed that metagenomes with coverages ≥ 0.60 performed better in terms of assembly and detection of differentially abundant genes and can be regarded as a universal minimum. Comparing samples with greater than two-fold differences in coverage should be avoided (Rodriguez-R and Konstantinidis 2014). Effort is needed to determine whether these general guidelines are also suitable for resistome analysis, especially considering that ARG diversity does not correspond 1:1 with phylogenetic diversity.

Further, there is a trade-off with depth and length of sequences, where longer sequences improve read alignments, overall annotation, and are more amenable to assembly. Most studies assessed (58%) utilized 150 bp paired-end reads, followed by 100 bp paired end reads (32%). Across these studies, the average library size of Illumina datasets was 7.0 Gb (Figure 4-3).

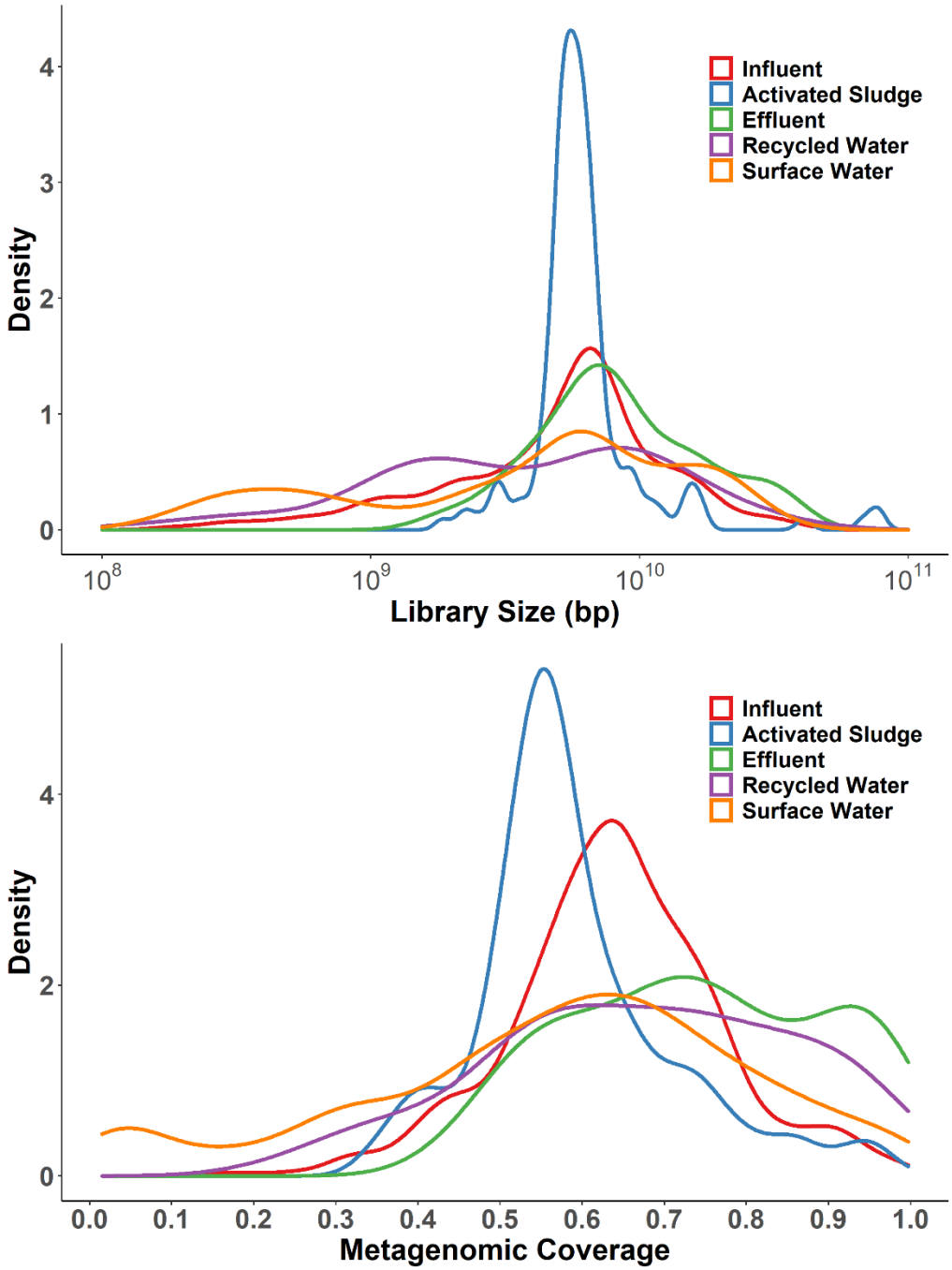


Figure 4-3: Sequencing depth and coverage by water matrix. Library sizes were determined from all publicly-available paired-end Illumina metagenomes from the 95 studied articles downloaded from the Sequence Read Archive, MG-RAST, and European Nucleotide Archive (1440 metagenomes). Metagenomic coverage was estimated using Nonpareil (Rodriguez-R et al. 2018) with option "-T kmer" on all cleaned and trimmed forward reads. Y-axis represents the frequency of individual metagenomes occurring at that library size or coverage factor.

SHARING OF COMPREHENSIVE METADATA IS NEEDED TO REAP THE VALUE OF METAGENOMIC DATA

Collection and sharing of appropriately rigorous metadata is a critical feature of any standardized framework for AR monitoring. Contextualizing resistome analysis is needed to accurately evaluate and compare samples across studies and space and time. Such metadata includes not only physicochemical parameters (e.g., temperature, pH, turbidity), but also water volumes collected, sample preservation (if any), DNA extraction methods/kits, and library preparation methods/kits (Liang et al. 2021). These metadata should be offered by researchers in all available instances, especially when uploading raw data to public repositories. Sparsely collected or vague reporting of metadata and effects on interpretation of results were common problems across the articles examined in this study. Notably, inspection of metadata reported across 1440 publicly-available metagenomes housed by SRA, ENA, and MG-RAST revealed several instances of sample types labeled "wastewater metagenome" without specifying the stage of biological wastewater treatment. Given that each stage of wastewater treatment is a distinct microbial ecosystem, lack of reporting of this nature renders the data unusable for meta-analysis, which is the intended purpose of data sharing.

BIOINFORMATICS AND DATA ANALYSIS

Online Platforms For Resistome Analysis

Depending on level of expertise, online data processing tools may be the most feasible option for resistome analysis **(**Table 4-S2). While such tools can aid in rapid analysis, it is important to be aware of any "black box" type assumptions that might be inconsistent with the nature of the samples or the monitoring objectives. Publicly-available data analysis pipelines, such as those hosted by Galaxy web portals (Giardine et al. 2005), can be beneficial where computational resources are minimal (no access to a computational server), for labs early along the adoption curve, or eventually, for when metagenomic methods for resistome monitoring become more standardized in common practice. The most commonly used online platform for environmental resistome analysis was ARGs-OAP executed in Galaxy (Yang et al. 2016; Yin et al. 2018) with its latest version utilizing DIAMOND and minimap2 against a custom, dereplicated database of ARGs, the Structured ARG Reference Database (SARG). MEGARes and its pipeline, AMR++, needs to be mentioned here. MetaStorm is another online platform with dedicated computational servers that enable the user to upload custom databases (Arango-Argoty et al. 2016).

Familiarity with command line data handling and processing for large datasets is advantageous for more advanced metagenomic analysis. This allows exploration and optimization of new analytical tools as they become available. As metagenomic profiling of ARGs is still largely implemented in the research domain, it is critical to be aware that there are numerous analytical parameters to choose from and each have implications for the research/monitoring objectives. As progress is made toward standardizing metagenomics for monitoring of resistomes in water and wastewater, agreement will be needed on default parameters (e.g., % identity, query coverage, amino acid length), depending on specific monitoring objectives, databases, and ideally, individual reference sequences.

Read QA/QC And Merging Improve Resistome Analysis

Following the generation and backup of sequencing reads, a critical first step is QA/QC assessment of the generated sequences to distinguish between natural genetic variations and sequencing errors or technical artefacts (i.e., adaptors and primer fragments). Because each

sequencing run is unique in the quality of generated data, exploratory analysis of library quality is useful in determining the degree of cleanup needed. FastQC (Andrews 2010) with MultiQC (Ewels et al. 2016) was found to be the most commonly employed software for this purpose, providing visualizations of key summary statistics of raw data, including read length, GC content, quality score distributions, number of duplicated reads, adaptor contamination, and number of Ns (unknown bases). These summary statistics can then inform appropriate read preprocessing, which involves trimming adaptors and low-quality ends, removing low quality and truncated reads, and choosing an acceptable number of Ns that define a valid sequence.

Across the studies identified in this critical review, the most frequently implemented trimming and filtering tools were Trimmomatic (31 articles) (Bolger et al. 2014), Sickle (7) (Joshi and Fass 2011), Fastx-toolkit (5) (Hannon Lab 2009), BBduk (5) (Bushnell 2017), Trim Galore! (4) (Babraham Bioinformatics 2012), and Cutadapt (3) (Marcel 2011), although many others exist and perform similar functions. The parameters used with each software were study specific, as the degree of quality filtering is dependent on the outcome of each sequencing run and the researcher's discretion. Reporting of trimming and filtering parameters, though, is essential for the reproducibility of metagenomic studies, as improperly cleaned data can result in artefacts that distort interpretation of the data (Del Fabbro et al. 2013; Bharti and Grimm 2021). The removal of reads originating from host organisms (i.e., host filtering) as a preprocessing tool was uncommon, although some chose to filter out reads aligning to *Homo sapiens* when analyzing municipal wastewater.

After reads have been filtered and trimmed, merging of the paired-end sequences via their overlapping regions was performed by a minority of studies (15 articles) using FLASH (Magoč and Salzberg 2011), Vsearch (Rognes et al. 2016), SeqPrep (St. John 2011), or PEAR (Zhang et al. 2014). When insert sizes in paired-end Illumina libraries are shorter than twice the read length, read pairs can be merged via overlapping regions to generate much longer reads (Magoč and Salzberg 2011), improving genome assembly, binning, and read mapping algorithms and therefore should be included in workflows where applicable (Bushnell et al. 2017).

Database Selection and Curation for ARG Annotation

Metagenomic sequence data must be aligned to a database to identify genes of interest. Across the included studies, the most frequently used databases for ARG annotation were the Comprehensive Antibiotic Resistance Database (CARD; 42%) (Jia et al. 2017; Alcock et al. 2020), the Antibiotic Resistance Genes Database (ARDB; 20%) (Liu and Pop 2009), Structured Antibiotic Resistance Genes (SARG; 11%) (Yin et al. 2018), ResFinder (10%) (Bortolaia et al. 2020), ARG-ANNOT (4%) (Gupta et al. 2014), and MEGARes (2%) (Doster et al. 2020) (Table 4-S3**).** ARDB and ARG-ANNOT, it should be noted, are no longer maintained and all sequences have been incorporated into several other databases. The National Database of Antibiotic Resistant Organisms (NDARO). ResFinder, SARG, MEGARes, and CARD remain actively curated. In many cases, a collection of these databases are manually combined and dereplicated on a per-study basis to increase the breadth of ARG detection (Subirats et al. 2016; Ju et al. 2019; Liu et al. 2019). ARGminer (Arango-Argoty et al. 2020) is a database that seeks to maintain active curation through crowd-sourcing, and is useful for exploratory research, bearing in mind that not all ARGs have been functionally validated in the laboratory.

When choosing an ARG database, it is important to consider that each one is curated for specific purposes and has strengths and weaknesses. In terms of ARG monitoring, a common

objective may be to conservatively identify all known and functionally-validated ARGs (Objective 1), such as those in the CARD database. ResFinder focuses specifically on acquired resistance genes. On the other hand, if the objective is to identify potentially new ARG variants that could be of concern in a community, then the deep-learning enabled DeepARG or the probabilistic gene model based fARGene (Berglund et al. 2019) pipelines might be advantageous (Objective 4).

Another concern is that many ARGs are conferred via single nucleotide polymorphisms (SNPs) of housekeeping genes, for example, the *rpoB2* gene variant in *Nocardia* spp. found in the protein homolog database of CARD. Although a match may be found in the metagenomic dataset, even at 80-100% sequence homology, it cannot be guaranteed that the variant conferring resistance was detected. This is an intrinsic limitation of short-read shotgun metagenomics, where the length of the query is only a fraction of the reference sequence, and the sequencing error rate precludes confidence in detecting a SNP without sufficient query depth (**Figure 4**). ARGs that are known to be caused by SNPs such as *par*E, *rpo*B, *pho*P, *pho*Q, *evg*S, *evg*A, *crp*, *evg*A, *env*R, *mar*A, *cpx*A, *cpx*R, *omp*F, and *bla*R should be checked for 100% peptide homology over a significant portion of the reference to prevent the overrepresentation of wild types (Doster et al. 2018). ARGs that are known as global regulators of efflux pump complexes are also commonly manually excised from databases before annotation (Lee et al. 2020a). Past efforts were made to manually remove such ARGs from the CARD database (e.g., through the development of SARG), but recent updates to CARD have continually improved this issue. Clustering ARG databases to generate consensus sequences of model reference genes may also help to eliminate these biases but would forgo gene variant level resolution. A recent review of available databases for metagenomic resistome analysis found that conclusions vary widely by which database is used and consensus references will be needed for universal data comparisons in the future (Lal Gupta et al. 2020).

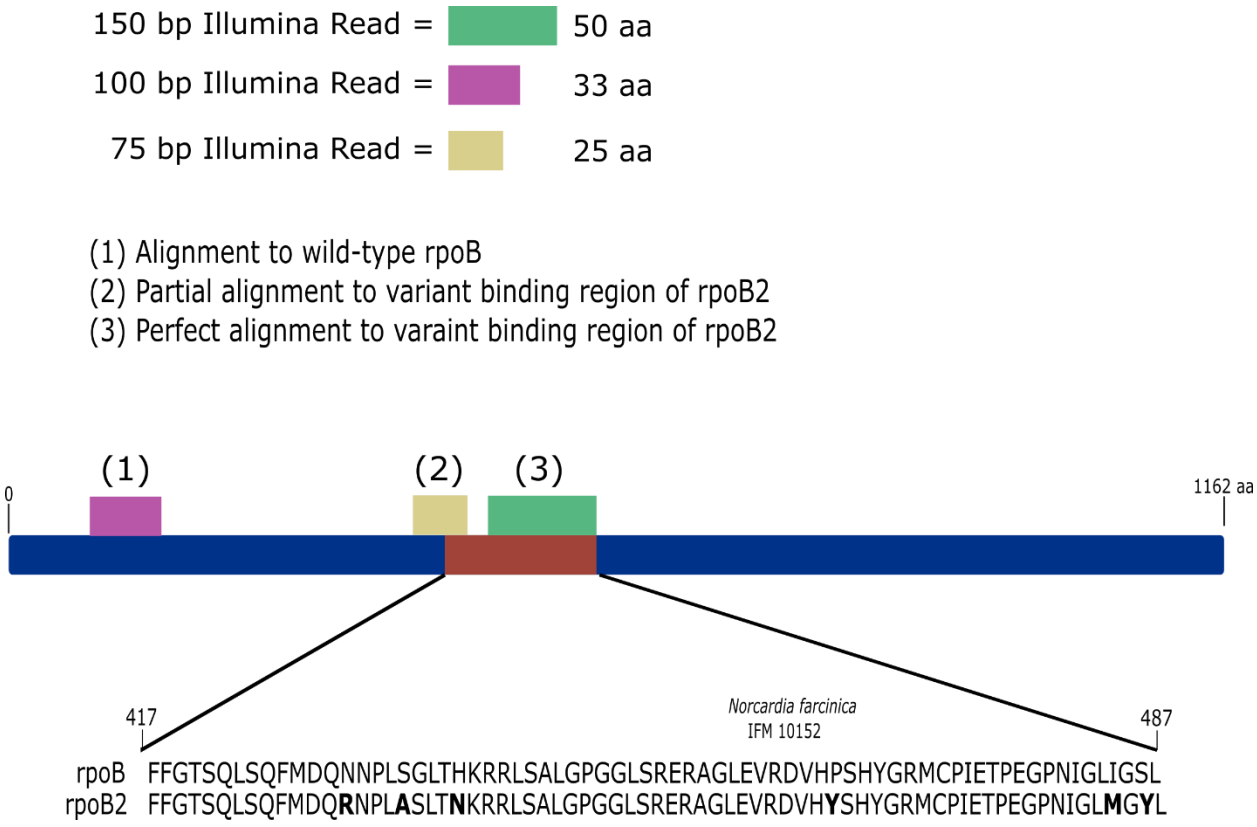


Figure 4-4: Example of the limitations of sequenced-based homology strategies for ARG detection using Illumina short reads. The size of the reference gene for the "Nocardia rifampin beta-subunit of RNA polymerase (rpoB2)" within the CARD protein-homolog database is 1162 aa long with the variable region conferring resistance (brown snippet) being 70 aa. Differences in amino acid sequences in the rpoB2 sequence are in bold. The probability that Illumina short reads align perfectly to the variable region identifying the rpoB2 variant (scenario 3) is unlikely and many databases are flooded with such variants of common orthologs, exacerbating ARG abundance calculations.

Read Alignment Tools and Parameters Dictate Detection Stringency

When performing read alignment to identify ARGs and other relevant genes, it is critical to assess what level of stringency is needed for the monitoring objective. Among the studies examined here, BLAST and its variants (Johnson et al. 2008) such as DIAMOND (Buchfink et al. 2014) and UBLAST/USEARCH (Edgar 2010) have emerged as the dominant family of read annotation tools. BLAST is known for its alignment accuracy (Buchfink et al. 2021), but DIAMOND and UBLAST provide much more reasonable turnaround time for metagenomic alignment and typically the accuracy trade-off is miniscule.

It is critical to report any cut-off parameters applied, such as the e-value, amino acid identity, query coverage, and bit score, as these will dictate the stringency of database hits. Some articles identified in this review did not report these cutoffs, particularly when using online

platforms. Across studies utilizing short-read sequences, the e-values used ranged from 1E-10 to 1E-4, while amino acid identity ranged from 50-95 percent, depending on the research question (Table 4-S4). When objectives are to conservatively identify known and functionally verified ARGs of clinical concern, parameters are stricter (e.g., query coverages of ≥ 80%, amino acid identities ≥ 80%, e-values ≤ 1e-10). The most applied alignment tool was BLASTx implemented in DIAMOND with an amino acid length of 25 at ≥ 80% identity. These parameters were first introduced by (Kristiansson et al. 2011) and have since been propagated throughout the field. Databases such as CARD also provide recommended bit score cutoffs for specific protein models, which can help to reduce guesswork in homology-based cutoffs.

However, traditional sequence-based homology frameworks are not ideal for new gene discovery (Objective 4), where expanded databases and deep learning models (e.g., DeepARG and HMD-ARG (Li et al. 2021b)), Hidden Markov Model-based approaches (e.g., ARGsOAP v2 (Yin et al. 2018) and ResFams (Gibson et al. 2015)), and probabilistic gene models (e.g., fARGenes (Berglund et al. 2019)) have been developed; Although the need for further validation has been duly noted (Bengtsson-Palme 2018). Permissive parameters are sometimes applied to more broadly capture putative ARGs. In any case, an agreed upon classification of allowable stringent to permissive alignment parameters would greatly enhance the comparability of resistome monitoring studies.

Normalization and Comparison of ARGs Across Environmental Samples

Because NGS does not directly yield absolute quantitative information, it is common practice to normalize to an internal or external parameter or to rarefy to a uniform read depth to facilitate comparisons across samples. Unfortunately, consistency in normalization is notably lacking and detracts from comparability across studies. Normalizing to the 16S rRNA gene (Li et al. 2015a) as a housekeeping gene present in all bacteria has been the most common approach and provides a biologically-relevant denominator, e.g., a proxy for ARGs/total bacteria. However, 16S rRNA gene copy numbers vary across species and therefore detracts from the biological relevance of this denominator. ARG/cell equivalent estimates are gaining ground as an alternative and can be obtained by using flow cytometry (Liang et al. 2020), by dividing by single copy genes (e.g., the β subunit of bacterial RNA polymerase, *RpoB)* (Zhang et al. 2019; Thornton et al. 2020), or by dividing the number of ARGs by the average coverage of a set of housekeeping genes (Yin et al. 2018; Dang et al. 2020; Lee et al. 2020b).

Because gene lengths vary and are generally longer than a typical 100-300 bp read (Figure 4-4), normalization should also consider the reference sequence length. Reads per kilobase million (RPKM) and fragments per kilobase million (FPKM) are metrics derived from RNA-Seq and are common normalization approaches when the aim is to compare across different projects, where there may be significant variation in sequencing depths (Hendriksen et al. 2019). It is important to be aware that RPKM is derived for single-end data whereas FPKM is designed for paired-end reads, by restricting the double-counting of pairs of sequences aligning to the same reference. Thus, these two normalizations should not be used interchangeably. A parts-per-million (ppm) normalization was also common in the literature, which simply divides the number of ARGs found by the number of million-reads queried. Rarefaction, i.e., randomly subsampling to a consistent number of reads per sample, can also be useful to support statistical analysis when sequencing depth is highly variable (Karlsson et al. 2014), but results in a substantial loss of very costly data.

Among the studied articles, 21 studies normalized ARGs to 16S rRNA genes, 11 to an estimated cell count (RpoB or collection of single copy genes), 14 to ppm, 12 to the relative percentages of reads (i.e., a 0 to 1 scale), and 11 to either RPKM or FPKM. Normalizations that are used in differential abundance analysis in RNA-Seq or traditional microbiome studies (e.g., Cumulative-Sum Scaling or Trimmed Mean of M-Values) (Lin and Peddada 2020) were only encountered twice (Rovira et al. 2019; Zaheer et al. 2019). An evaluation of 14 of these methods for determining differentially abundant genes in comparative metagenomics revealed large discrepancies in performance, concluding that sample size, effect size, and gene abundance were key factors skewing biological interpretations (Jonsson et al. 2016). Systematic evaluation of normalization parameters for the purpose of ARG monitoring would be of value.

QUANTITATIVE METAGENOMICS

Relative abundance metrics are not always ideal for downstream analysis, especially for microbial risk assessment (Garner et al. 2021b; Li et al. 2021a). A few studies to date have sought to derive absolute ARG abundances (i.e., ARGs per volume or mass of sample) from environmental metagenomic data, i.e., quantitative metagenomics (qMeta). Hybrid spike-independent approaches convert relative ARG abundances into absolute abundances by relying on supplementary quantitative analyses. For instance, (Garner et al. 2016), (Garner et al. 2018), and (Davis et al. 2020) determined the relative abundance of ARGs per 16S rRNA copies within the metagenomic dataset and correspondingly quantified the 16S rRNA copies per sample using qPCR. Applying the assumption that the target gene/16S rRNA quotient is equivalent between metagenomics and qPCR, a gene copy per unit volume metric is derived. Correlations between absolute ARG abundances derived from qPCR and hybrid spike-independent methods have shown strong correlations across several gene targets (Davis et al. 2020; Majeed et al. 2021). However, the reliability was shown to diminish for low abundance ARGs where the limit of detection (LOD) for metagenomics exceeded that of the qPCR assay targets (Davis et al. 2020; Majeed et al. 2021), or where primers fail to capture the full diversity of target ARGs (Crossette et al. 2021).

Spike-dependent methods use internal nucleic acid reference standards that are incorporated directly into samples after DNA extraction (Figure 4-2). The reference standards are selected to be highly unlikely to be present in the sample, allowing them to be distinguished from the native microbial community. Recently, (Crossette et al. 2021) spiked genomic DNA from an exogenous marine organism (*Marninobacter hydrocarbonoclasticus*) into DNA extracted from digested and undigested cow manure to quantify tetracycline ARGs. Reads were mapped to all 4,272 genes comprising the genome and the average ratio of known spiked-in gene copies to reads mapped were used to calculate absolute abundances on a per-mass basis. The authors found that qPCR and qMeta were in strong agreement, but qPCR displayed a lower LOD than qMeta (2 to 8 copies/mg versus $3x10^4$ copies/mg). The LOD for qMeta is directly proportional to the sequencing depth. Synthetic DNA reference standards (Li et al. 2021a) and quantitative ladders (Hardwick et al. 2018) have recently been developed and are worthy of exploration to support quantitative environmental monitoring of ARGs.

The addition of internal DNA reference standards has the potential to normalize datasets across space and time, regardless of sequencing depth, and provide universal library comparisons.

METAGENOMIC ASSEMBLY FOR ARG CONTEXTUALIZATION

The genetic context of ARGs is a crucial facet in defining their human and ecological health concerns (Objectives 3 & 4). In particular, mobile ARGs (i.e., carried on MGEs) are of greater concern because they can be acquired by a new bacterial host and result in a new resistant strain. Likewise, carriage in human pathogens presents greater concern than carriage in a non-pathogens (Martinez et al. 2015; Bengtsson-Palme et al. 2017). Assembly of short-read sequence data into contigs and scaffolds has been widely applied to identify genes neighboring ARGs to estimate whether they are mobile or present in pathogens. However, assembly algorithms diminish the quantitative value of the data. One workaround is to derive the relative abundance of protein coding regions (CDSs) within assembled libraries by mapping short-reads back to the assembled library (Ng et al. 2017; Zhou et al. 2019; Zhao et al. 2020). Still, the larger challenge is uncertainty in the accuracy of short-read assembly and lack of means to assess the accuracy.

Environmental metagenomes are especially difficult to assemble due to the intermingled genomes of thousands of species at unknown abundance distributions, many of which are closely related or are not represented in databases. There are numerous options to assembling short-read data, each with their own assumptions, computational requirements, and overall limitations (Ayling et al. 2020). Due to the relatively low representation of environmental bacteria in taxonomic databases, *de novo* assembly is often implemented. Across the studies identified in the literature review, MEGAHIT (Li et al. 2015b), IDBA-UD (Peng et al. 2012), and SOAPdenovo2 (Luo et al. 2015) were the most commonly used assemblers.

Long-read sequencing is a promising way to circumvent the uncertainties associated with assembly of short-reads and has recently been demonstrated for ARG monitoring (Che et al. 2019; Dai et al. 2022), but comes with the tradeoff of shallow sequencing depth and low coverage (Figure 4-3). A recent systematic evaluation of various assembly approaches for contextualizing ARGs found that a hybrid assembly approach resulted in the least number of erroneous contigs, suggesting a 10× minimum depth to minimize chimeric contigs that may skew resistome analysis (Brown et al. 2021). Following assembly, CDSs can be found using Prodigal (Hyatt et al. 2010), FragGeneScan (Rho et al. 2010), or MetaGeneMark (Zhu et al. 2010) and annotated for ARGs (Tables 4-S3, S4), MGEs, MRGs, or other functional genes.

RESISTOME RISK ASSESSMENT MODELS

Looming large over efforts to monitor aquatic resistomes is the need to take steps towards translating the measurements to human and ecological health risks. The original framework proposed by (Martínez et al. 2015) ranks the "risk" posed by individual ARGs as a function of their documented ability to cause treatment failure, their association with MGEs, their carriage by human and animal pathogens, and their propensity for being transferred into pathogens. This framework was translated into an empirical model, MetaCompare (Oh et al. 2018), where the metagenomic reads are *de novo* assembled and annotated to identify ARGs, MGEs, and pathogen markers and their co-occurrence patterns. Samples are scored and ranked in accordance with these co-occurrences to identify potential "hot spots" for AMR evolution and transmission. A key limitation to this approach is the algorithm's inability to rank the relative importance of individual ARGs and taxonomic sub-groups of bacteria. For instance, differentiating the relative importance of MGE-borne carbapenamases in *Enterobacterales* over ubiquitous efflux pumps in environmental strains of human pathogenic taxa is a critical distinction. More recently, a similar omics-based framework and software package "arg_ranker" was developed to categorize

individual gene targets by their enrichment in anthropogenically-impacted environments, their history of mobility, and their presence in ESKAPE (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* spp.) pathogens (Zhang et al. 2021). The framework made marked progress in sorting high- from low-risk ARGs from an environmental perspective and determined human-associated, functionally verified, and mobilized ARGs to constitute only 3.6% of the known resistome. ARGs found by sequenced-based homology are ubiquitous in the environment, but only a small fraction pose a direct threat to human health (Fitzpatrick and Walsh 2016), while others may serve better as indicators of conditions that are conducive to the evolution and selection of resistant strains.

Advancements are also being made in the field of quantitative microbial risk assessment (QMRA), where the goal is to quantify the risk of human exposures to environmental AMR (Ashbolt et al. 2013). This approach, though, requires knowledge of the absolute concentrations (i.e., on a target per volume bases) of ARB and ARGs at exposure sites to derive dose estimates following ingestion, skin contact, or inhalation (Huijbers et al. 2015). A key hurdle to this framework as it is applied to AMR is factoring in HGT rates between commensal bacteria and pathogens in both the environment and in the human host following exposure (Li et al. 2018b). A recent review of research needs for risk assessment of recycled water matrices duly pointed out the need for reporting absolute concentrations of ARB and ARGs on a per volume basis for integration and further development of current QMRA frameworks as opposed to the myriad of relative abundance metrics derived by traditional shotgun metagenomics (Garner et al. 2021b).

CONCLUSION

Metagenomics has emerged as a powerful tool for the monitoring of environmental resistomes. The near-random sequencing of all genomic fragments in a sample without *a priori* identification of gene targets allows for comprehensive assessments of microbial dynamics and risk factors for the development and proliferation of AR. However, several aspects of the workflow, from sample collection to NGS data generation and analysis, require careful consideration to ensure comparability of resulting data across space and time. Experimental controls were conspicuously absent from identified studies applying NGS for AR monitoring of aquatic environments and should be included in future studies. Sequencing depths should be appropriately targeted based on the research question and internal and external standards should be included to verify the accuracy and improve the quantitative capacity of resulting metagenomic data. The recommendations here can aid in the generation of universally comparable sequence libraries needed to support broader ecological studies and environmental surveys. Sharing of metadata can also support larger-scale computational modeling. Given that a major advantage of NGS is the ability to store and analyze data retrospectively, the sooner the field can move towards improved quality and consistency in application of NGS for environmental AR monitoring, the better off we will be in our ability to accurately harvest the information needed to effectively combat the spread of AR.

DISCLOSURE STATEMENT

ACKNOWLEDGEMENTS

REFERENCES

Aarestrup FM, Woolhouse MEJ (2020) Using sewage for surveillance of antimicrobial resistance. Science (80-) 367:630–632. https://doi.org/10.1126/science.aba3432

Alcock BP, Raphenya AR, Lau TTY, et al (2020) CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935

Alosaimi S, Bandiang A, Van Biljon N, et al (2018) A broad survey of DNA sequence data simulation tools. Brief Funct Genomics 19:49–59. https://doi.org/10.1093/bfgp/elz033

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arango-Argoty G, Garner E, Pruden A, et al (2018) DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6:. https://doi.org/10.1186/s40168-018-0401-z

Arango-Argoty G, Singh G, Heath LS, et al (2016) MetaStorm: A public resource for customizable metagenomics annotation. PLoS One 11:1–13. https://doi.org/10.1371/journal.pone.0162442

Arango-Argoty GA, Guron GKP, Guron GKP, et al (2020) ARGminer: A web platform for the crowdsourcing-based curation of antibiotic resistance genes. Bioinformatics 36:2966–2973. https://doi.org/10.1093/bioinformatics/btaa095

Ashbolt NJ, Amézquita A, Backhaus T, et al (2013) Human health risk assessment (HHRA) for environmental development and transfer of antibiotic resistance. Environ Health Perspect 121:993–1001. https://doi.org/10.1289/ehp.1206316

Ayling M, Clark MD, Leggett RM (2020) New approaches for metagenome assembly with short reads. Brief Bioinform 21:584–594. https://doi.org/10.1093/bib/bbz020

Babraham Bioinformatics (2012) TrimGalore!

Barba M, Czosnek H, Hadidi A (2013) Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses 6:106–136. https://doi.org/10.3390/v6010106

Bartolomaeus TUP, Birkner T, Bartolomaeus H, et al (2021) Quantifying technical confounders in microbiome studies. Cardiovasc Res 117:863–875. https://doi.org/10.1093/cvr/cvaa128

Bengtsson-Palme J (2018) The diversity of uncharacterized antibiotic resistance genes can be predicted from known gene variants-but not always. Microbiome 6:1–12. https://doi.org/10.1186/s40168-018-0508-2

Bengtsson-Palme J, Larsson DGJ, Kristiansson E (2017) Using metagenomics to investigate human and environmental resistomes. J Antimicrob Chemother 2690–2703. https://doi.org/10.1093/jac/dkx199

Berendonk TU, Manaia CM, Merlin C, et al (2015) Tackling antibiotic resistance: The environmental framework. Nat Rev Microbiol 13:310–317. https://doi.org/10.1038/nrmicro3439

Berglund F, Österlund T, Boulund F, et al (2019) Identification and reconstruction of novel antibiotic resistance genes from metagenomes. Microbiome 7:1–14. https://doi.org/10.1186/s40168-019-0670-1

Bharti R, Grimm DG (2021) Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform 22:178–193. https://doi.org/10.1093/bib/bbz155

Białasek M, Miłobędzka A (2020) Revealing antimicrobial resistance in stormwater with MinION. Chemosphere 258:. https://doi.org/10.1016/j.chemosphere.2020.127392

Bokulich NA, Ziemski M, Robeson MS, Kaehler BD (2020) Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Comput Struct Biotechnol J 18:4048–4062. https://doi.org/10.1016/j.csbj.2020.11.049
Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
Borchardt MA, Boehm AB, Salit M, et al (2021) The Environmental Microbiology Minimum Information (EMMI) Guidelines: qPCR and dPCR Quality and Reporting for Environmental Microbiology. Environ Sci Technol. https://doi.org/10.1021/acs.est.1c01767
Bortolaia V, Kaas RS, Ruppe E, et al (2020) ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500. https://doi.org/10.1093/jac/dkaa345
Bowers RM, Clum A, Tice H, et al (2015) Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. BMC Genomics 16:1–12. https://doi.org/10.1186/s12864-015-2063-6
Brown CL, Keenum IM, Dai D, et al (2021) Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. Sci Rep 11:1–12. https://doi.org/10.1038/s41598-021-83081-8
Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366–368. https://doi.org/10.1038/s41592-021-01101-x
Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176
Bushnell B (2017) BBmap
Bushnell B, Rood J, Singer E (2017) BBMerge – Accurate paired shotgun read merging via overlap. PLoS One 12:1–15. https://doi.org/10.1371/journal.pone.0185056
Centers for Disease Control and Prevention (2020) Developing a Wastewater Surveillance Sampling Strategy. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/wastewater-surveillance/developing-a-wastewater-surveillance-sampling-strategy.html#print
Che Y, Xia Y, Liu L, et al (2019) Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7:1–13. https://doi.org/10.1186/s40168-019-0663-0
Crossette E, Gumm J, Langenfeld K, et al (2021) Metagenomic quantification of genes with internal standards. MBio 12:1–12. https://doi.org/10.1128/mBio.03173-20
Dai D, Brown C, Bürgmann H, et al (2022) Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. Microbiome 10:1–16. https://doi.org/10.1186/s40168-021-01216-5
Dang C, Xia Y, Zheng M, et al (2020) Metagenomic insights into the profile of antibiotic resistomes in a large drinking water reservoir. Environ Int 136:105449. https://doi.org/10.1016/j.envint.2019.105449
Davis BC, Riquelme MV, Ramirez-toro G, et al (2020) Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. Environ Sci Technol. https://doi.org/10.1021/acs.est.0c05567
Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM (2013) An extensive evaluation of read trimming effects on illumina NGS data analysis. PLoS One 8:1–13. https://doi.org/10.1371/journal.pone.0085024
Doster E, Lakin SM, Dean CJ, et al (2020) MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res 48:D561–D569. https://doi.org/10.1093/nar/gkz1010

Doster E, Rovira P, Noyes NR, et al (2018) Investigating Effects of Tulathromycin Metaphylaxis on the Fecal Resistome and Microbiome of Commercial Feedlot Cattle Early in the Feeding Period. Front Microbiol 9:1–14. https://doi.org/10.3389/fmicb.2018.01715

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461

European Commission (2017) A European One Health Action Plan against Antimicrobial Resistance (AMR)

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Filazzola A, Cahill JF (2021) Replication in field ecology: Identifying challenges and proposing solutions. Methods Ecol Evol 12:1780–1792. https://doi.org/10.1111/2041-210X.13657

Fitzpatrick D, Walsh F (2016) Antibiotic resistance genes across a wide variety of metagenomes. FEMS Microbiol Ecol 92:1–8. https://doi.org/10.1093/femsec/fiv168

Garner E, Chen C, Xia K, et al (2018) Metagenomic Characterization of Antibiotic Resistance Genes in Full- Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. Environ Sci Technol 52:6113–6125. https://doi.org/10.1021/acs.est.7b05419

Garner E, Davis BC, Milligan E, et al (2021a) Next generation sequencing approaches to evaluate water and wastewater quality. Water Res 194:116907. https://doi.org/10.1016/j.watres.2021.116907

Garner E, Organiscak M, Dieter L, et al (2021b) Towards risk assessment for antibiotic resistant pathogens in recycled water: A systematic review and summary of research needs. Environ Microbiol 00: https://doi.org/10.1111/1462-2920.15804

Garner E, Wallace JS, Argoty GA, et al (2016) Metagenomic profiling of historic Colorado Front Range flood impact on distribution of riverine antibiotic resistance genes. Sci Rep 1–10. https://doi.org/10.1038/srep38432

Giardine B, Riemer C, Hardison RC, et al (2005) Galaxy: A platform for interactive large-scale genome analysis. Genome Res 15:1451–1455. https://doi.org/10.1101/gr.4086505

Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9:207–216. https://doi.org/10.1038/ismej.2014.106

Guo F, Zhang T (2013) Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. Appl Microbiol Biotechnol 97:4607–4616. https://doi.org/10.1007/s00253-012-4244-4

Gupta SK, Padmanabhan BR, Diene SM, et al (2014) ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13

Gweon HS, Shaw LP, Swann J, et al (2019) The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ Microbiomes 14:1–15. https://doi.org/10.1186/s40793-019-0347-1

Hamner S, Brown BL, Hasan NA, et al (2019) Metagenomic profiling of microbial pathogens in the little bighorn river, Montana. Int J Environ Res Public Health 16:. https://doi.org/10.3390/ijerph16071097

Han D, Gao P, Li R, et al (2020) Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing. J Adv Res 26:111–121. https://doi.org/10.1016/j.jare.2020.07.010

Hannon Lab (2009) FastX-Toolkit
Hardwick SA, Chen WY, Wong T, et al (2018) Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nat Commun 9:1–10. https://doi.org/10.1038/s41467-018-05555-0
Hardwick SA, Deveson IW, Mercer TR (2017) Reference standards for next-generation sequencing. Nat Rev Genet 18:473–484. https://doi.org/10.1038/nrg.2017.44
Hendriksen RS, Munk P, van Bunnik B, et al (2019) Global monitoring of antimicrobial resistance based on human sewage. Nat Commun. https://doi.org/10.1038/s41467-019-08853-3
Huijbers PMC, Blaak H, De Jong MCM, et al (2015) Role of the Environment in the Transmission of Antimicrobial Resistance to Humans: A Review. Environ Sci Technol 49:11993–12004. https://doi.org/10.1021/acs.est.5b02566
Huijbers PMC, Flach C-F, Larsson DGJ (2019) A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. Environ Int 130:104880. https://doi.org/10.1016/j.envint.2019.05.074
Hyatt D, Chen GL, LoCascio PF, et al (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:. https://doi.org/10.1186/1471-2105-11-119
Jia B, Raphenya AR, Alcock B, et al (2017) CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004
Johannessen E, Eikum AS, Krogstad T (2012) Evaluation of sampling methods for monitoring effluent phosphorus for on-site wastewater treatment systems. Water Sci Technol 65:2049–2054. https://doi.org/10.2166/wst.2012.107
Johnson M, Zaretskaya I, Raytselis Y, et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36:5–9. https://doi.org/10.1093/nar/gkn201
Jonsson V, Österlund T, Nerman O, Kristiansson E (2016) Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. BMC Genomics 1–14. https://doi.org/10.1186/s12864-016-2386-y
Joshi NA, Fass JN (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files
JPIAMR (2019) Strategic Research and Innovation Agenda on Antimicrobial Resistance
Ju F, Beck K, Yin X, et al (2019) Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J 13:346–360. https://doi.org/10.1038/s41396-018-0277-8
Karlsson FH, Nookaew I, Nielsen J (2014) Metagenomic Data Utilization and Analysis (MEDUSA) and Construction of a Global Gut Microbial Gene Catalogue. PLoS Comput Biol 10:. https://doi.org/10.1371/journal.pcbi.1003706
Keenum I, Liguori K, Calarco J, et al (2022) A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. Crit Rev Environ Sci Technol 0:1–25. https://doi.org/10.1080/10643389.2021.2024739
Knudsen BE, Bergmark L, Munk P, et al (2016) Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. mSystems 1:1–15. https://doi.org/10.1128/msystems.00095-16
Kopperi H, Tharak A, Hemalatha M, et al (2021) Defining the methodological approach for

wastewater-based epidemiological studies — Surveillance of SARS-CoV-2. Environ Technol Innov 23:. https://doi.org/10.1016/j.eti.2021.101696
Kristiansson E, Fick J, Janzon A, et al (2011) Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One 6:. https://doi.org/10.1371/journal.pone.0017038
Lal Gupta C, Kumar Tiwari R, Cytryn E (2020) Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. Environ Int 138:105667. https://doi.org/10.1016/j.envint.2020.105667
Larsson DGJ, Andremont A, Bengtsson-palme J, et al (2018) Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. Environ Int 117:132–138. https://doi.org/10.1016/j.envint.2018.04.041
Larsson DGJ, Flach C-F (2021) Antibiotic resistance in the environment. Nat Rev Microbiol 0123456789: https://doi.org/10.1038/s41579-021-00649-x
Lee K, Kim D, Lee D, et al (2020a) Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome 1–14
Lee K, Kim D, Lee D, et al (2020b) Supplemental Info 1: Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome
Li AD, Metch JW, Wang Y, et al (2018a) Effects of sample preservation and DNA extraction on enumeration of antibiotic resistance genes in wastewater. FEMS Microbiol Ecol 94:1–11. https://doi.org/10.1093/femsec/fix189
Li B, Li X, Yan T (2021a) A Quantitative Metagenomic Sequencing Approach for High-Throughput Gene Quantification and Demonstration with Antibiotic Resistance Genes. Appl Environ Microbiol 87:1–11. https://doi.org/10.1128/AEM.00871-21
Li B, Yang Y, Ma L, et al (2015a) Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J 9:2490–2502. https://doi.org/10.1038/ismej.2015.59
Li D, Liu CM, Luo R, et al (2015b) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
Li L, Dechesne A, He Z, et al (2018b) Estimating the Transfer Range of Plasmids Encoding Antimicrobial Resistance in a Wastewater Treatment Plant Microbial Community. Environ Sci Technol Lett acs.estlett.8b00105. https://doi.org/10.1021/acs.estlett.8b00105
Li Y, Xu Z, Han W, et al (2021b) HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes. Microbiome 9:1–12. https://doi.org/10.1186/s40168-021-01002-3
Liang J, Mao G, Yin X, et al (2020) Identification and quantification of bacterial genomes carrying antibiotic resistance genes and virulence factor genes for aquatic microbiological risk assessment. Water Res 168:115160. https://doi.org/10.1016/j.watres.2019.115160
Liang X, Akers K, Keenum I, et al (2021) AgroSeek: a system for computational analysis of environmental metagenomic data and associated metadata. BMC Bioinformatics 22:1–14. https://doi.org/10.1186/s12859-021-04035-5
Liguori K, Keenum I, Calarco J, et al (2022) Antimicrobial Resistance Monitoring of Water Environments: A Call for Standardized Methods and Quality Control. Environ Sci Technol
Lin H, Peddada S Das (2020) Analysis of microbial compositions: a review of normalization and differential abundance analysis. npj Biofilms Microbiomes 6:. https://doi.org/10.1038/s41522-020-00160-w

Liu B, Pop M (2009) ARDB - Antibiotic resistance genes database. Nucleic Acids Res 37:443–447. https://doi.org/10.1093/nar/gkn656
Liu Z, Klümper U, Liu Y, et al (2019) Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. Environ Int 129:208–220. https://doi.org/10.1016/j.envint.2019.05.036
Luo R, Liu B, Xie Y, et al (2015) Erratum to "SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler" [GigaScience, (2012), 1, 18]. Gigascience 4:1. https://doi.org/10.1186/s13742-015-0069-2
Maghini DG, Moss EL, Vance SE, Bhatt AS (2021) Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. Nat Protoc 16:458–471. https://doi.org/10.1038/s41596-020-00424-x
Magoč T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507
Majeed HJ, Davis BC, Riquelme M V., et al (2021) Evaluation of metagenomic-enabled antibiotic resistance surveillance at a conventional wastewater treatment plant. Front Microbiol. https://doi.org/10.3389/fmicb.2021.657954
Marcel M (2011) Cutadapt removes adaptor sequences from high-throughput reads. EMBnet J 7:2803–2809
Martinez J, Coque T, Baquero F (2015) What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol
Martínez JL, Coque TM, Baquero F (2015) What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol 13:116–123. https://doi.org/10.1038/nrmicro3399
Metzker ML (2005) Emerging technologies in DNA sequencing. Genome Res 15:1767–1776. https://doi.org/10.1101/gr.3770505
Mirzayi C, Renson A, Furlanello C, et al (2021) Reporting guidelines for human microbiome research: the STORMS checklist. Nat Med 27:1885–1892. https://doi.org/10.1038/s41591-021-01552-x
Natarajan A, Han A, Zlitni S, et al (2021) Standardized preservation, extraction and quantification techniques for detection of fecal SARS-CoV-2 RNA. Nat Commun 12:1–12. https://doi.org/10.1038/s41467-021-25576-6
Ng C, Tay M, Tan B, et al (2017) Characterization of metagenomes in urban aquatic compartments reveals high prevalence of clinically relevant antibiotic resistance genes in wastewaters. Front Microbiol 8:1–12. https://doi.org/10.3389/fmicb.2017.02200
O'Neill J (2016) Tackling Drug-Resistant Infections Globally: Final Report and Recommendations
O'Sullivan DM, Doyle RM, Temisak S, et al (2021) An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities. Sci Rep 11:1–14. https://doi.org/10.1038/s41598-021-89881-2
Oh M, Pruden A, Chen C, et al (2018) MetaCompare : a computational pipeline for prioritizing environmental resistome risk. FEMS Microbiol Ecol 1–9. https://doi.org/10.1093/femsec/fiy079
Pärnänen KMM, Narciso-Da-Rocha C, Kneis D, et al (2019) Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. Sci Adv 5:. https://doi.org/10.1126/sciadv.aau9124
Pavlovska M, Prekrasna I, Parnikoza I, Dykyi E (2021) Soil sample preservation strategy affects the microbial community structure. Microbes Environ 36:1–7.

https://doi.org/10.1264/jsme2.ME20134
Peng Y, Leung HCM, Yiu SM, Chin FYL (2012) IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174
Poulsen CS, Kaas RS, Aarestrup FM, Pamp SJ (2021) Standard Sample Storage Conditions Have an Impact on Inferred Microbiome Composition and Antimicrobial Resistance Patterns. Microbiol Spectr 9:. https://doi.org/10.1128/spectrum.01387-21
Pruden A, Vikesland PJ, Davis BC, et al (2021) Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. Curr Opin Microbiol 64:91–99. https://doi.org/10.1016/j.mib.2021.09.013
Quick J, Loman NJ (2019) DNA extraction strategies for nanopore sequencing. In: Nanopore Sequencing: An Introduction. World Scientific, pp 1–17
Rho M, Tang H, Ye Y (2010) FragGeneScan: Predicting genes in short and error-prone reads. Nucleic Acids Res 38:1–12. https://doi.org/10.1093/nar/gkq747
Rodayan A, Majewsky M, Yargeau V (2014) Impact of approach used to determine removal levels of drugs of abuse during wastewater treatment. Sci Total Environ 487:731–739. https://doi.org/10.1016/j.scitotenv.2014.03.080
Rodriguez-R LM, Gunturu S, Tiedje JM, et al (2018) Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. mSystems 3:1–9. https://doi.org/10.1128/msystems.00039-18
Rodriguez-R LM, Konstantinidis KT (2014) Estimating coverage in metagenomic data sets and why it matters. ISME J 8:2349–2351. https://doi.org/10.1038/ismej.2014.76
Rognes T, Flouri T, Nichols B, et al (2016) VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016:1–22. https://doi.org/10.7717/peerj.2584
Rovira P, McAllister T, Lakin SM, et al (2019) Characterization of the Microbial Resistome in Conventional and "Raised Without Antibiotics" Beef and Dairy Production Systems. Front Microbiol 10:1–12. https://doi.org/10.3389/fmicb.2019.01980
Sahlin K, Sipos B, James PL, Medvedev P (2021) Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. Nat Commun 12:1–13. https://doi.org/10.1038/s41467-020-20340-8
Sato MP, Ogura Y, Nakamura K, et al (2019) Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. DNA Res 26:391–398. https://doi.org/10.1093/dnares/dsz017
Sidstedt M, Rådström P, Hedman J (2020) PCR inhibition in qPCR, dPCR and MPS—mechanisms and solutions. Anal Bioanal Chem 412:2009–2023. https://doi.org/10.1007/s00216-020-02490-2
Sidstedt M, Steffen CR, Kiesler KM, et al (2019) The impact of common PCR inhibitors on forensic MPS analysis. Forensic Sci Int Genet 40:182–191. https://doi.org/10.1016/j.fsigen.2019.03.001
St. John J (2011) SeqPrep
Stamps BW, Leddy MB, Plumlee MH, et al (2018) Characterization of the microbiome at the world's largest potable water reuse facility. Front Microbiol 9:1–16. https://doi.org/10.3389/fmicb.2018.02435
Stoler N, Nekrutenko A (2021) Sequencing error profiles of Illumina sequencing instruments. NAR Genomics Bioinforma 3:1–9. https://doi.org/10.1093/nargab/lqab019
Subirats J, Sànchez-Melsió A, Borrego CM, et al (2016) Metagenomic analysis reveals that

bacteriophages are reservoirs of antibiotic resistance genes. Int J Antimicrob Agents 48:163–167. https://doi.org/10.1016/j.ijantimicag.2016.04.028
Thornton CN, Tanner WD, Vanderslice JA, Brazelton WJ (2020) Localized effect of treated wastewater effluent on the resistome of an urban watershed. Gigascience 9:1–13. https://doi.org/10.1093/gigascience/giaa125
United Nations (2017) Frontiers 2017 - Emerging Issues Of Environmental Concern
Uyaguari-Díaz MI, Croxen MA, Luo Z, et al (2018) Human activity determines the presence of integron-associated and antibiotic resistance genes in Southwestern British Columbia. Front Microbiol 9:1–20. https://doi.org/10.3389/fmicb.2018.00852
Wickam H (2009) ggplot2: Elegant Graphics for Data Analysis. Springer
World Health Organization (2015) Global Action Plan on Antimicrobial Resistance
Yadav R, Rajput V, Gohil K, et al (2020) Comprehensive metagenomic insights into a unique mass gathering and bathing event reveals transient influence on a riverine ecosystem. Ecotoxicol Environ Saf 202:110938. https://doi.org/10.1016/j.ecoenv.2020.110938
Yang Y, Jiang X, Chai B, et al (2016) ARGs-OAP: Online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. Bioinformatics 32:2346–2351. https://doi.org/10.1093/bioinformatics/btw136
Yeh Y-C, Needham DM, Sieradzki ET, Fuhrman JA (2018) Taxon Disappearance from Microbiome Analysis Reinforces the Value of Mock Communities as a Standard in Every Sequencing Run. mSystems 3:. https://doi.org/10.1128/msystems.00023-18
Yin X, Deng Y, Ma L, et al (2019) Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. Environ Int 133:105270. https://doi.org/10.1016/j.envint.2019.105270
Yin X, Jiang XT, Chai B, et al (2018) 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 34:2263–2270. https://doi.org/10.1093/bioinformatics/bty053
Zaheer R, Lakin SM, Polo RO, et al (2019) Comparative diversity of microbiomes and Resistomes in beef feedlots, downstream environments and urban sewage influent. BMC Microbiol 19:1–17. https://doi.org/10.1186/s12866-019-1548-x
Zaheer R, Noyes N, Ortega Polo R, et al (2018) Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep 8:1–11. https://doi.org/10.1038/s41598-018-24280-8
Zhang A-N, Gaston JM, Dai CL, et al (2021) An omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun 12:1–11. https://doi.org/10.1038/s41467-021-25096-3
Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. https://doi.org/10.1093/bioinformatics/btt593
Zhang SY, Tsementzi D, Hatt JK, et al (2019) Intensive allochthonous inputs along the Ganges River and their effect on microbial community composition and dynamics. Environ Microbiol 21:182–196. https://doi.org/10.1111/1462-2920.14439
Zhao R, Yu K, Zhang J, et al (2020) Deciphering the mobility and bacterial hosts of antibiotic resistance genes under antibiotic selection pressure by metagenomic assembly and binning approaches. Water Res 186:. https://doi.org/10.1016/j.watres.2020.116318
Zhou S, Zhu Y, Yan Y, et al (2019) Deciphering extracellular antibiotic resistance genes

(eARGs) in activated sludge by metagenome. Water Res 161:610–620. https://doi.org/10.1016/j.watres.2019.06.048
Zhu W, Lomsadze A, Borodovsky M (2010) Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:1–15. https://doi.org/10.1093/nar/gkq275

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Towards Standardization of Metagenomics for Monitoring Antibiotic Resistance in Water and Wastewater

Authors: Benjamin C. Davis[1], Jeannette Calarco[2], Krista Liguori[1], Erin Milligan[1], Valerie J. Harwood[2], Amy Pruden,[1] Ishi Keenum[1*]

1 Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia
2 Department of Integrative Biology, University of South Florida, Tampa, Florida
*Corresponding Author

Table S1: *Search parameters*

Tier	Attribute	Keywords
1	Topic	“wastewater” OR “reclaimed water” OR “recycled water” OR “water reuse” OR “non-potable reuse” OR “greywater” OR “hospital wastewater” OR “surface water” OR “sewage” OR “wastewater treatment plant” OR “filtration” OR “direct potable reuse” OR “indirect potable reuse” OR “river” OR “watershed” OR "lake" OR "pond" OR "recreational water" OR "influent" OR "effluent" OR "aquatic" OR "water quality" OR "de facto reuse"
2	Topic	“antibiotic resistan*” OR “antimicrobial susceptibility” OR “antimicrobial resistan*” OR “drug resistan*” OR “multi-drug resistan*” OR “resistome” OR "ARG" OR "antibiotic resistan* gene"
3	Topic	“Next generation” OR “metagenom*” OR "NGS" OR “sequencing” OR “high-throughput sequencing”

Table S2*: Data extracted from each included study*

Parameter	Description
Reference	Formatted reference citation
Country	Where the study took place
Continent	
General Environment	Surface, Recycled or Wastewater
Specific Environment	What treatment stages were sampled/data extracted from?
Sample Replicates	How many samples were taken in replicate?
Sampling Volume	Minimum and Maximum of collected sample volumes
Concentration Methods	How were samples concentrated?
DNA Extraction	What kit/ method was used
DNA Clean Up	What DNA clean up kit was used (if applicable)
DNA Quantification	How was DNA quantified? (if reported)
Library Prep	What library prep kit was used?
Sequencing Platform	What sequencing platform was used?
Read Length	What length of reads were sequenced?
Giga base pairs	How many base pairs resulted from sequencing (if reported)?
Analysis Tool	What bioinformatic tools were used?
Tool function	Why was this tool applied?
Parameters	What parameters were used with the tool?
Melt Curves	Were melt curves analyzed?
Database	Which databases were used to annotate metagenomes?
Metrics	Which metrics were derived for assessing antibiotic resistance?
Normalization	What type of ARG count normalization was applied?
Statistics	What statistics were applied overall?
Metagenome Accessions	Sequence Read Archive, European Nucleotide Archive, and MG-RAST project identifications

***Table S3**: Free Web-based Metagenomic ARG Profiling Tools*

Tool	**Studies utilized (n)**	**Sequence Platform**	**Incorporated functions**	**Databases used**	**Reference**
ARG-OAPv2	10	Illumina	Annotation, Normalization	SARG	[1]
MetaStorm	3	Illumina	Annotation, Assembly, Normalization	Customizable	[2]
DeepARG	2	Illumina	Annotation, Prediction, Normalization	DeepARG-DB	[3]
NanoARG	2	Nanopore	Annotation	DeepARG-DB, NCBI-NR + I-VIP, BacMet, Centrifuge, ESKAPE+WHO	[4]
ARGPORE	1	Nanopore	Annotation	SARG	[5]
AMRPlusPlus	2	Illumina	Annotation, Normalization	MEGARes(1.0,2.0)	[6,7]

Table S4: Databases used across all reviewed articles and motivations for ARG analysis

Database	Studies utilized (n)	Approach	Genes Contained	Last Version Update	Reference
CARD	38	Highly curated ARG database	3,146	October 2021	[8]
ARDB	18	All ARGs	4,545	July 2009	[9]
SARG	14	Dereplicated ARG database	12,307	January 2021	[1]
ResFinder	9	ARGS from whole genome datasets	2,236	April 2021	[10]
DeepARG-db	5	Machine Learning ARG Predication Database	14,933	April 2020	[3]
ARG-ANNOT	4	Point Mutation and all ARGs	2,038	May 2018	[11]
ResFams	3	HMM model for ARG Predication Database	177	January 2015	[12]
MEGARes	2	ARGs, MRGs- hand-curated AMR database and annotation structure	7,868	October 2019	[7]
Resqu	2	Horizontally transferred ARGs	3,018		https://www.1928diagnostics.com/resdb/

Table S5: *Alignment parameters used in short-read and long-read queries of ARG databases.*

	Parameter	Alignment Tools Used	Studies Utilized (n)	Range Observed
Short-Read Alignments	Query Coverage	BLAT, BLASTx	12	50 - 90 %
	Amino Acid Identity	BLASTx, RAPsearch, ShortBRED, UBLAST, Vmatch	31	50 - 95 %
	Amino Acid Length	BLASTx, RAPsearch, ShortBRED, UBLAST, Vmatch	26	25 - 60 aa
	E value	BLASTx, BLASTn, RAPsearch, HMMsearch, UBLAST	31	1e-10 - 1e-3
	Bitscore		0	
	Query Coverage	BLASTn, BLASTp, LAST	20	40 - 95 %
	Nucleotide Identity	BLASTn, LAST	5	70 - 97 %
Predicted ORF Alignments (long-reads)	Nucleotide Length	BLASTn, LAST	5	50 - 100 nt
	Amino Acid Identity	BLASTp	21	30 - 85 %
	Amino Acid Length	BLASTp	4	25 - 100 aa
	E value	BLASTp	24	1e-50 - 1e-3
	Bitscore	BLASTp	3	50

CHAPTER 5: TOWARDS THE STANDARDIZATION OF ENTEROCOCCUS CULTURE METHODS FOR WATERBORNE ANTIBIOTIC RESISTANCE MONITORING: A CRITICAL REVIEW AND META-ANALYSIS OF TRENDS ACROSS STUDIES

Authors: Benjamin C. Davis[1], Ishi Keenum[1], Jeannette Calarco[2], Krista Liguori[1], Erin Milligan[1], Amy Pruden,[1] Valerie J. Harwood[2*]

1 Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia

2 Department of Integrative Biology, University of South Florida, Tampa, Florida

*Corresponding Author

ABSTRACT

Antibiotic resistance is a major 21st century One Health (humans, animals, environment) challenge whose spread limits options to treat bacterial infections. There is growing interest in monitoring water environments, including surface water and wastewater, which have been identified as key recipients, pathways, and sources of antibiotic resistant bacteria (ARB). *Enterococcus* spp. often carry clinically-important antibiotic resistance genes and are of interest as environmental monitoring targets. *Enterococcus* spp. are Gram-positive bacteria that are typically of fecal origin; however, they are also found in relevant environmental niches, with various species and strains that are opportunistic human pathogens. Although the value of environmental monitoring of antibiotic-resistant *Enterococcus* has been recognized by both national and international organizations, lack of procedural standardization has hindered generation of comparable data needed to implement integrated surveillance programs. Here we provide a comprehensive methodological review to assess the techniques used for the culturing and characterization of antibiotic-resistant *Enterococcus* across water matrices for the purpose of environmental monitoring. We analyzed 105 peer-reviewed articles from 33 countries across six continents. The goal of this review is to provide a critical analysis of (i) the various methods applied globally for isolation, confirmation, and speciation of *Enterococcus* isolates, (ii) the different methods for profiling antibiotic resistance among enterococci, and (iii) the current prevalence of resistance to clinically-relevant antibiotics among *Enterococcus* spp. isolated from various environments. Finally, we provide advice regarding a path forward for standardizing culturing of *Enterococcus* spp. for the purpose of antibiotic resistance monitoring in wastewater and wastewater-influenced waters within a global surveillance framework.

INTRODUCTION

*Enterococcu*s spp. are important members of the natural enteric microbiome of both humans and animals and have emerged as important antibiotic-resistant pathogens in clinical medicine (Arias and Murray, 2012). There are currently 60 published *Enterococcus* genomes in the National Center for Biotechnology Information database, most of which are commensal

microorganisms, although some act as opportunistic pathogens in humans. *E. faecalis* and *E. faecium* are among the most important etiological agents of nosocomial infections; including urinary tract infections (UTIs), central nervous system infections, endocarditis, bacteremia, neonatal infections, and surgical site infections (Moellering, 1992; Murray, 1990). From 2006 to 2017, *Enterococcus* spp. were responsible for approximately 14% of all healthcare-associated infections in the US, ranking second overall behind *Staphylococcus aureus* (Hidron et al., 2008; Sievert et al., 2013; Weiner-Lastinger et al., 2020; Weiner et al., 2016). *Enterococcus* spp. possess full or partial intrinsic chromosomal resistance to cephalosporins, aminoglycosides, lincosamides, trimethoprim-sulfamethoxazole, and penicillins (Hollenbeck and Rice, 2012). Their rapid development of multi-drug resistance has been attributed in part to their highly malleable genomes that lack CRISPR (clustered regularly interspaced palindromic repeats) elements, which has facilitated the ready acquisition of allochthonous mobile DNA (e.g., vancomycin resistance gene clusters) (Palmer and Gilmore, 2010). Nearly 25% of the genomes of many clinical *E. faecalis* and *E. faecium* isolates consist of acquired genetic elements (Hegstad et al., 2010; Paulsen et al., 2003). Recently, over 85% of *E. faecium* and 15% of *E. faecalis* isolates responsible for catheter-associated UTIs and central line-associated bloodstream infections diagnosed in the US have been found to be vancomycin resistant (Weiner-Lastinger et al., 2020). The US Center for Disease Control and Prevention (CDC) and the World Health Organization (WHO) have set VRE to "high" priority and a "serious" threat level (Centers for Disease Control and Prevention, 2019; Tacconelli et al., 2018).

Enterococcus spp. are members of the larger, phenotypically-defined group known as enterococci, which are Gram-positive, catalase-negative, obligately fermentative chemoorganotrophs that can survive over a wide range of temperatures, pH, and salinity (Teixeira et al., 2015). Enterococci are found in many extraenteric environmental niches, including soils and sediments, beach sands, aquatic vegetation, and terrestrial vegetation. Enterococci have been extensively isolated from wastewaters, marine waters, and freshwaters (Byappanahalli et al., 2012). Because of their abundance in human and animal feces, their extraenteric persistence, and the ease with which they are cultured, enterococci have been targeted for decades as fecal indicators for the purpose of water quality monitoring (Jang et al., 2017; Schoen et al., 2011; Sinclair et al., 2012). Enterococci (formerly classified within the larger group known as "fecal streptococci") have been widely used to assess the microbiological safety of surface waters, drinking waters, recreational beaches, and as a target for assessing process removal efficiencies during wastewater treatment. Enterococci have also been found to correlate directly to public health measures; for example, across the US, the rate of gastrointestinal illness in swimmers has been correlated with *Enterococcus* spp. levels in recreational beach waters that are impacted by wastewaters (Prüss, 1998; Wade et al., 2006, 2003). Their importance as water quality indicators and their inclusion in governmental regulatory frameworks has led to a great deal of method development for isolation and enumeration from environmental samples (Boehm and Sassoubre, 2014; Health Canada, 2020).

Global and national action plans set in place to combat the spread of antibiotic resistance have generally embraced a One Health approach (humans-animals-environment) (European Commission, 2017; Hernando-amado et al., 2019), but a better understanding of the role of environmental dimensions is needed. Aquatic environments have been identified as a key recipient and transmission pathway of antibiotic resistant bacteria into and out of human and animal populations (Amarasiri et al., 2020; Larsson and Flach, 2021). Multidrug-resistant

pathogens and mobile antibiotic resistance genes enter the environment via treated and untreated wastewater across the globe (Alexander et al., 2020; Marathe et al., 2017; Zhang et al., 2020). Recently, the WHO put forth the Tricycle protocol as a standardized method for monitoring the dissemination, transmission, and evolution of antibiotic resistance along the One Health continuum: humans (hospitals and community), the food chain (animal husbandry), and the environment (human and animal fecal contamination) (WHO, 2021). Specifically, the protocol targets Gram-negative extended-spectrum beta-lactamase (ESBL) producing *Escherichia coli*, which display phenotypic resistance to third-generation cephalosporins. While ESBL *E. coli* was selected, in part, to coordinate with global surveillance of *Enterobacterales* (Marano et al., 2020), the extent to which it is truly a representative indicator of resistome dynamics has not been established. Thus, the present is a critical moment to also consider other potential targets.

Antibiotic-resistant enterococci present many advantages as a potential target for monitoring antibiotic resistance in the water environment. The level of standardization for methods targeting *Enterococcus* spp. is arguably second only to *E. coli* in environmental waters. As such, *Enterococcus* spp. could present an attractive complimentary target to *E. coli*. Notably, as Gram-positive organisms, they provide insight into distinct genotypes and phenotypes of antibiotic resistance that would not be captured by monitoring only Gram-negative organisms. In particular, the plasticity of enterococcal genomes and their propensity for horizontal gene transfer and exchange of virulence and antibiotic resistance determinants from clinical strains to environmental reservoirs of enterococci (Ekwanzala et al., 2020b; Gouliouris et al., 2019, 2018) marks them as potentially comprehensive targets for antibiotic resistance monitoring.

Recent progress has been made in applying culture-based methods for monitoring antibiotic resistant *Enterococcus* spp. in the environment, with emphasis on human and animal wastewater, and hospital wastewater pollution (Gouliouris et al., 2019; Savin et al., 2020; Zaheer et al., 2020). The phenotypic and morphological similarity of other Gram-positive organisms to the enterococci, however, makes isolation on selective media prone to false-positives, with cross-selectivity with other cocci (e.g., *Streptococcus, Pediococcus, Weisella*) (Harwood et al., 2001). False positives are especially problematic for environmental samples (Pagel and Hardy, 1980). Non-selectivity can actually be exacerbated with addition of antibiotics to media, because several members of non-target *Bacillus* and other genera are intrinsically resistant to clinically-relevant concentrations of certain antibiotics (Woodford et al., 1995).

The selection of any monitoring target or strategy entails consideration of the overarching purpose or questions to be addressed, and these may vary depending on the focus across the spectrum from wastewater to surface water. The following are examples of key monitoring goals and considerations addressed by this review:

- Monitoring antibiotic resistance among clinically-relevant strains of *Enterococcus* spp. in sewage as a means of assessing their levels carried in the human population
- Assessing whether clinically-relevant *Enterococcus* spp. or specific resistance phenotypes are effectively removed during wastewater treatment and if they persist in impacted aquatic environments
- Evaluating evidence that clinically-relevant *Enterococcus* spp. acquire antibiotic resistance genes from the environment

- Comparing resistant *Enterococcus* spp. in various water matrices both locally and globally to assess factors that may be contributing to antibiotic resistance in Gram-positive organisms

LITERATURE REVIEW PROTOCOL

This systematic review was conducted in a four-tiered approach using search terms presented in Table 5-S1 to collect studies published between January 2000 and December 2020. Briefly, Tier 1 was designed to isolate topic relevant search terms for surface water, wastewater, recycled water, or reclaimed water (3,828,792 articles). Tier 2 was designed to select for those articles addressing antibiotic resistance (15,043). Tier 3 further narrowed the search to culturing techniques (5,439) and Tier 4 specifically to enterococci/*Enterococcus* (479). Initially, these 479 articles were independently screened by two researchers for containing a complete workflow from environmental sampling through to characterization of individual isolates. Articles were excluded that were focused on secondary data. Further exclusions were articles that exclusively investigated: biofilms, mesocosms, drinking water, sediments, or digested sludge (e.g., anaerobic digestion). Fecal source tracking articles based on antibiotic resistance analysis were also excluded (Harwood et al., 2000). Articles that used a non-selective media for initial isolation, such as R2A or TSA, were also excluded. Disagreements on article inclusion from the initial screening were presented to a larger group of five researchers to reach a consensus. The resulting 105 peer-reviewed articles were then subject to data extraction using parameters outlined in Table 5-S2.

META-ANALYSIS OF PUBLISHED DATA

A meta-analysis was performed to extract data relevant to the species and phenotypic distribution of all isolated enterococci. First, the number of isolates per species was extracted from articles in which libraries were speciated (86 articles) to reveal general population statistics across *Enterococcus* spp. Second, antibiotic susceptibility testing (AST) data were extracted from all articles that provided the percent of resistant isolates compared to total enterococci isolated in the absence of any antibiotic (66 articles). Studies that summed isolates with "intermediate" or "resistant" classifications of resistance without providing individual statistics, as well as studies that did not cite standardized methodology for classifying resistance (e.g., current CLSI breakpoints at the time of sampling), were excluded.

Methods for Culturing Environmental Enterococci

In the US and Canada, enterococci are recommended for monitoring saline (brackish or marine) and recreational freshwaters. In the EU, enterococci are regulated in both drinking water and recreational water by standardized culture methods (Agency, 2020). Several standardized culture methods have been developed, including the US Environmental Protection Agency (USEPA) Methods 1106.1 and 1600 for ambient waters and wastewaters (U.S. Environmental Protection Agency, 2009, 2006), the International Organization for Standardization (ISO) Methods 7899-1 and 7899-2 (International Organization for Standardization, 2000), and Method 9230 (A-D) as part of the American Public Health Association's (APHA) "Standard Methods for the Examination of Water and Wastewater" (APHA, 1999; Rice and Baird, 2017) (Table 5-1). These methods include three distinct techniques: membrane filtration (MF), multiple tube fermentation (MTF), and defined substrate techniques (e.g., Enterolert). The current "gold standard" for enterococci enumeration from the environment is considered the MF technique (Byappanahalli et al., 2012) and was used by over 90% of articles included in this review (Table 5-1).

The principal selective and differential solid media used in standard MF assays are Slanetz-Bartley (SB), mEnterococcus (mE), and membrane-Enterococcus Indoxyl-β-D-Glucoside (mEI). These media use various peptone and yeast extract-based nutrients with the addition of sodium azide and/or nalidixic acid. Sodium azide obstructs the growth of Gram-negative bacteria through the inhibition of cytochrome oxidase. Both SB and mE agars include 2,3,5-triphenyltetrazolium chloride (TTC), which dyes viable colonies red. Differentiated colonies grown on SB or mE are then confirmed as enterococci by their ability to hydrolyze esculin in the presence of bile using either Bile Esculin Azide or Esculin Iron agars. The hydrolyzed esculin product, esculetin, reacts with iron salt in the media to produce black to reddish colonies for enumeration. mEI is similar to mE medium, but contains the chromogen, indoxyl-β-D-glucoside. When cleaved by β-D-glucosidase positive enterococci, blue halos are formed around positive colonies. mEI is typically used a standalone media as all colonies with blue halos are considered enterococci.

Table 5-1: Published standardized methods for the detection and enumeration of enterococci in different water matrices

Organization and Method	Recommended Matrix	Media	Number of Citations[a]	Assay Turnaround (hours)	General Procedure
Membrane Filtration (MF); (CFU/mL); Number of Studies Identified: 93/105					
EPA Method 1600	drinking water; source water; wastewater; marine and freshwater	mEI[b]	18	24	mEI (41°C for 24 hrs); Count blue halos
ISO 7899-2	surface water; wastewater	Slanetz-Bartley; Bile Esculin Azide	15	48	Slanetz-Bartley (36°C for 44 hrs); Bile Esculin Azide Agar (44°C for 2 hrs)
EPA Method 1106.1	marine and freshwater (not applicable to wastewater)	mEnterococcus; Esculin Iron Agar	0	48	mEnterococcus (41°C for 48 hrs); Esculin Iron Agar (41°C for 20 min); Count pink to red colonies
APHA SM 9230C.2a	drinking water; source water; wastewater; marine and freshwater	mEnterococcus; Esculin Iron Agar	3	48	mEnterococcus (41°C for 48 hrs); Esculin Iron Agar (41°C for 20 min); Count pink to red colonies
APHA SM 9230C.2b	drinking water; source water; wastewater; marine and freshwater	mEI[b]	0	24	mEI (41°C for 24 hrs); Count blue halos
APHA SM 9230C.2c	drinking water; source water;	mEnterococcus	10	48	mEnterococcus (35°C for 48 hrs); Count

	wastewater; marine and freshwater				light and dark red colonies
Multiple Tube Fermentation (MTF); (MPN/mL); Number of Studies Identified: 1					
APHA SM 9230B	drinking water; source water; marine and freshwater (not applicable to wastewater)	Azide Dextrose Broth; Bile Esculin Azide Agar	1	48-72	Azide Dextrose Broth (35°C for 24-48hrs); Bile Esculin Azide Agar (35°C for 24 hrs); Compute MPN
Fluorogenic Substrate Test (MTF and MPN); (MPN/mL); Number of Studies Identified: 3					
APHA SM 9230D	drinking water; source water; wastewater; marine and freshwater	Enterolert®	3	24	Enterolert Media (41°C for 24 hrs); Compute MPN
ISO 7899-1	surface water; wastewater	MUD[c] Media	0	36-72	MUD Media in Microtitre Wells (44°C for 36-72 hrs); Compute MPN

[a] The 'Number of Citations' under Membrane Filtration do not correspond to the number of citations in the table as many articles did not follow or cite a standard method. Techniques not listed are direct plating after serial dilutions (8 articles). [b] membrane-Enterococcus Indoxyl-β-D-Glucoside Agar (mEI), [c] 4-methylumbelliferyl-β-D-glucoside (MUD), Environmental Protection Agency (EPA), International Organization for Standardization (ISO), American Public Health Association (APHA)

Performance of Standard Enterococci Culture Assays

Several comparative studies have been conducted over recent decades to assess each medium's selectivity for *Enterococcus* (Table 5-2). Pagel et al. 1980 used pure cultures to assess PSE (Pfizer), KF Streptococcus, mE, and SB agars against over 100 pure cultures of clinical and environmental isolates of various cocci. The highest selectivity for enterococci was observed for PSE (94%) and mE (94%) agars, with the lowest being KF Streptococcus (80%) and SB (78%) (Pagel and Hardy, 1980). Compared to mE, however, PSE was found to yield lower recovery efficiencies from wastewaters with much higher rates of background colony growth. Other comparative studies found enterococci selectivity on PSE and KF Streptococcus agars as low as 86% and 54%, respectively (Brodsky and Schiemann, 1976). The original mE agar formulation study found a false positive rate of 10% and false negative rate of 11.2% for surface water isolates (Levin et al., 1975). Subsequent studies have confirmed false positive rates for mE agar as low as 2.5% when testing pure *Enterococcus* cultures (Dionisio and Borrego, 1995) and 1.7% in marine, riverine, and treated wastewater effluent (Adcock and Saint, 2001). The inclusion of the indoxyl-β-D-glucoside chromogen to mE agars resulted in an increase in specificity of *Enterococcus* to upwards of 99.7% in ambient freshwaters (Adcock and Saint, 2001). In a recent benchmarking study, ISO method 7899-2 (SB media) was found to have false positive rates as high as 18% and false negative rates as high as 57.1%, depending on the colony count on the filter membrane of recreational marine water (Tiwari et al., 2018). Differences in *Enterococcus* selectivity have also been documented between MF and defined substrate techniques, where *E. faecalis* is differentially selected for in wastewater using Enterolert, leading to the conclusion

that these methods should not be used interchangeably for regulatory purposes (Ferguson et al., 2013, 2010; Kinzelman et al., 2003; Maheux et al., 2009).

Significant differences *Enterococcus* concentrations have also been reported. For instance, several studies were conducted in the wake of the advent of Enterolert assays in the mid-1990s to compare its efficacy against established MF techniques for water quality monitoring (Fricker and Fricker, 1996). Significant differences in concentrations were reported between Enterolert and SB agar in marine and recreational freshwaters (Valente et al., 2010), while no significant differences were found between mE agar and Enterolert concentrations across surface water, wastewater, or marine waters (Abbott et al., 1998; Budnick et al., 1996; Eckner, 1998; Fricker and Fricker, 1996). No significant differences in enterococci concentrations were identified between mE mEI agar (Adcock and Saint, 2001). Importantly, no studies were identified that directly compared the specificities and concentrations derived from SB and mE or mEI across water matrices. Such a comparison should be considered in future studies that assess their utility for regulatory frameworks for antibiotic resistance monitoring internationally. Any biases in species distributions and total enterococci concentrations originating from the selective media could skew downstream distributions in resistance frequencies and introduce bias if the data are used for risk assessment.

Table 5-2: Performance of Enterococcus selective media used in standard membrane filtration assays

Medium	Matrix Tested	Presumptive Colonies	Specificity (%)[1]	Selectivity (%)[2]	Reference
mEI	Marine	1361	-	82.4	(Ferguson et al., 2005)
mEI	Pure Cultures	101	97.3	100	(Maheux et al., 2009)
mEI	Surface	54	-	100	(Nishiyama et al., 2015)
mEI	Surface; Wastewater; Marine	1279	-	94.9	(Ferguson et al., 2013)
mEI	Surface; Wastewater; Marine	641	-	94.5	(Ferguson et al., 2010)
mEI	Surface; Wastewater; Marine	361	-	93.9	(Messer and Dufour, 1998)
mEnterococcus	Marine	80	-	97.5	(Dionisio and Borrego, 1995)
mEnterococcus	Marine	624	-	94.2	(de Oliveira and Watanabe Pinhata, 2008)
mEnterococcus	Pure Cultures	93	91.0	88.2	(Pagel and Hardy, 1980)
mEnterococcus	Surface	2231	-	88.5	(Levin et al., 1975)

mEnterococcus	Surface; Wastewater; Marine	1043	-	90.2	(Adcock and Saint, 2001)
Slanetz-Bartley	Marine	97	-	93.8	(Audicana et al., 1995)
Slanetz-Bartley	Marine	234	-	92.7	(Tiwari et al., 2018)
Slanetz-Bartley	Pure Cultures	82	78	74.4	(Pagel and Hardy, 1980)
Slanetz-Bartley	Surface	321	-	95.3	(Łuczkiewicz et al., 2010)
Slanetz-Bartley	Surface; Wastewater	385	-	93.8	(Fricker and Fricker, 1996)

1. Specificity = (True Negatives)/(True Negatives + False Positives)

2. Selectivity = True Positives/(True Positives + False Positives) or (Colonies Confirmed to Enterococcus Genus)/(Total Presumptive Enterococcus Colonies in Collection)

Workflows for Antibiotic-Resistant *Enterococcus* Monitoring

A useful method for culturing and enumerating both generic and antibiotic resistant environmental *Enterococcus* would strike a balance between sensitivity (i.e., detect all *Enterococcus* spp. that are present), specificity (i.e., avoid detecting other genera), and the high-throughput needed for large-scale environmental monitoring. Aquatic matrices display a large and dynamic range of enterococci concentrations, and a method for their enumeration would also need an appropriately low limit of detection for "cleaner" samples and a sufficiently high limit of quantification for matrices like wastewater where enterococci are concentrated. Specific logistical considerations are also warranted, such as the ability to perform the assay in low-tech laboratories using materials, techniques, and media that are economically feasible for AMR monitoring in low- and middle-income countries. In this instance, standard methods that have been developed for enumerating generic enterococci can be leveraged for their extensive vetting with respect to quality assurance/quality control and adapted to the increased throughput needs of AMR monitoring projects.

Culture-based approaches for investigating antibiotic resistance amongst environmental *Enterococcus* must be modified based on the aquatic matrix being investigated and the purpose of the assessment. Here we delineate these approaches into three general categories: population-level surveys, targeted monitoring for specific antibiotic resistant phenotypes, and recovery of low concentration or viable but non-culturable (VBNC) populations (Figure 5-1), each with their own benefits and limitations.

For population-level monitoring (66 articles) (Figure 5-1A), where the objective is to achieve an unbiased snapshot of the distribution of resistance phenotypes, a collection of isolates can be generated using an *Enterococcus* selective method (e.g., Table 5-1). After colonies have been counted, isolates with the specified morphologies can then be selected randomly off plates for phenotypic antibiotic susceptibility testing, generating an antibiotic resistance profile as a function of the total number of isolates subsampled. The disadvantage of this approach is that

most colonies screened may not be antibiotic resistant and finding colonies with the resistant phenotypes of interest, and to achieve required statistical power, can be akin to "searching for a needle in a haystack". The advantage to this approach is that it provides a denominator for total *Enterococcus* in the sample and an unbiased distribution of both enterococcal species and their genotypes and phenotypes, resulting in an ecologically-relevant analysis (Cho et al., 2019).

If the phenotype of interest is already known (e.g., high-level VRE), targeted monitoring approaches may be more efficient for in-depth characterizations of sub-populations of *Enterococcus.* Such approaches use an antibiotic at clinically-relevant breakpoints to select for specific resistance phenotypes (Figure 5-1B). The use of low/intermediate breakpoints of antibiotics may be useful for capturing a broad range of phenotypes in the environment but will frequently capture clinically-irrelevant organisms, especially in the case of glycopeptide resistance (Figure 5-1C). The sample can also be plated in tandem on the selective media without the antibiotic (Figure 5-1D), thus allowing the quantification of the resistant population as a fraction of the total enterococci measured in CFU/unit volume, a universally comparable monitoring value (e.g., see WHO Tricycle Program recommendations (WHO, 2021)). Studies utilizing targeted approaches often screen the identified resistant colonies against a panel of antibiotics, which can include the original selective antibiotic to confirm clinically-relevant levels of resistance. This approach requires the choice of initial selective antibiotic and therefore will exclude strains that are not resistant to the primary selective antibiotic. Using a selective antibiotic will also skew the distribution of *Enterococcus* spp. away from the true distribution, often selecting for closely related genera that share the same resistance phenotype, such as *Lactobacillus, Leuconostoc, Weissella,* and *Pediococcus* in the case of high levels of glycopeptide resistance (Harwood et al., 2001; Nishiyama et al., 2017, 2015).

In some scenarios, the recovery of very dilute phenotypes (rare targets) or stressed cells is desirable, for instance in advanced treated wastewater intended for reuse or other disinfected waters. Pre-enrichment of samples in concentrated selective broth (e.g., Enterococcosel or Azide Dextrose Broth) amended with the selected antibiotic at low/intermediate concentrations can greatly increase the detection limit of rare phenotypes by helping to recover VBNC colonies (Blanch et al., 2003; Vilanova and Blanch, 2006). These recovered, resistant colonies can then be streaked on high-levels of the antibiotic to recover clinically-relevant phenotypes of interest. However, any protocol employing a target enrichment step will preclude the ability to quantify the resistant *Enterococcus* population or normalize to the total population, a necessity for universally comparable datatypes (Figure 5-1E).

After the collection of isolates is generated, purification and confirmation of the genus can be performed on brain-heart infusion (BHI), bile esculin azide (BEA), or esculin iron agar (EIA) and the level of resistance (low, intermediate, full resistance) is determined. Further isolate characterization is advised, including speciation and genotyping for ARGs and virulence factors. Whole genome sequencing (WGS) is recommended as it is the most accurate and comprehensive method for speciation and genotyping and enables global isolate comparisons. A comprehensive evaluation of the suite of methods used for isolate characterization is presented in the following sections.

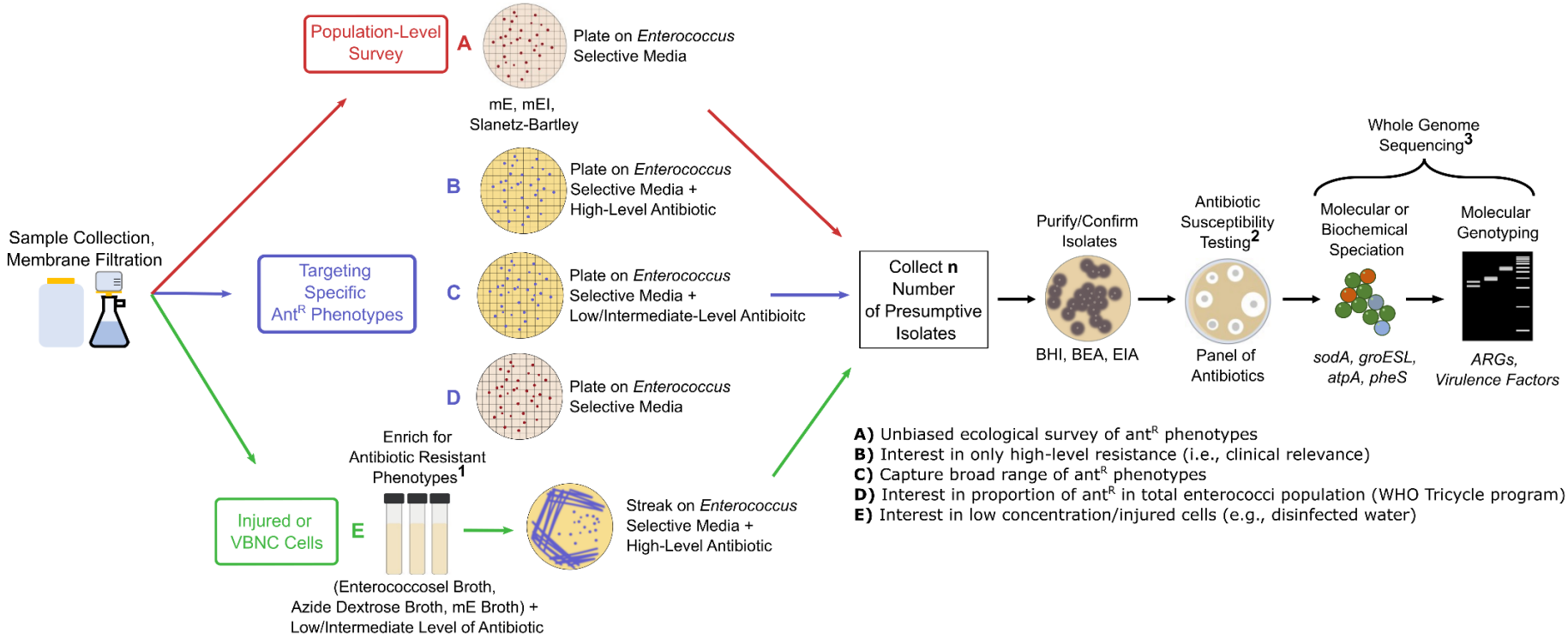


Figure 5-1: Workflows for monitoring antibiotic-resistant Enterococcus in the environment. [1]Note that pre-enrichment for resistance phenotypes (injured or VBNC cells) prevents their quantification. [2]Antibiotic susceptibility testing of subsampled colonies often includes original selective antibiotic to confirm full "resistant" classification. [3]Whole genome sequencing is recommended for the most accurate speciation and comprehensive genotyping for global isolate comparisons. Ant^R = antibiotic resistance, VBNC = viable but non-culturable, BHI=brain-heart infusion, BEA=bile esculin azide, EIA=esculin iron agar, ARGs = antibiotic resistance gene

Antibiotic Susceptibility Testing Methods

Several techniques and automated platforms exist for generalized antibiotic susceptibility testing (AST) or the determination of minimum inhibitory concentrations (MICs) of isolate libraries; including Kirby-Bauer disk diffusion assays (~ 47% of articles), various commercial automated systems (24%), manual broth or agar dilutions (20%), or strip test methods (9%) (Figure 5-2A). Over 90% of AST were performed on either Mueller-Hinton agar or in Mueller-Hinton broth. Concentrations of antibiotics chosen for AST were predominantly determined by referencing the Clinical and Research Standards Institute (CLSI; 70%) breakpoints for *Enterococcus*, according to the most currently available guidelines. Other standardized breakpoint concentrations were specific to a particular nation or governmental body, e.g., including the European Committee on Antimicrobial Susceptibility Testing (EUCAST; 7%), the National Antimicrobial Resistance Monitoring System (NARMS; USA; 2%), and the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS; 1%). Approximately 15% of studies utilizing AST that also differentiated degrees of resistance (i.e., "susceptible", "intermediate", or "resistant") did not cite a justification for antibiotic concentrations nor a standardized method for determining the level of resistance observed.

The panel of antibiotics used to screen isolates varied across studies, but overarching trends were apparent. VRE was mentioned in the title of ~35% of articles identified, and vancomycin was included in the screening panel in over 75% of the articles (Figure 5-2B). The number of observed phenotypes among clinical and environmental isolates of *Enterococcus* are wide-ranging, which warrants a diverse range of antibiotics included in the panels. The antibiotics tested were further categorized into twelve distinct classes by activity. The most prominent classes across all studies were glycopeptides (76%), macrolide-lincosamide-streptogramin (MLS; 72%), tetracyclines (72%), beta-lactams (especially penicillins) (70%), and aminoglycosides (64%). Interestingly, antibiotics that are either approved by the FDA to treat VRE infections or are commonly used to treat VRE (Arias and Murray, 2012) were less commonly included in panels. These compounds include linezolid (27%), quinupristin/dalfopristin (21%), daptomycin (8%), and the synthetic glycopeptide, teicoplanin (28%). High-level aminoglycoside resistance (e.g., gentamicin, streptomycin, and kanamycin) in enterococci isolated from the environment was the focus of small subset of studies.

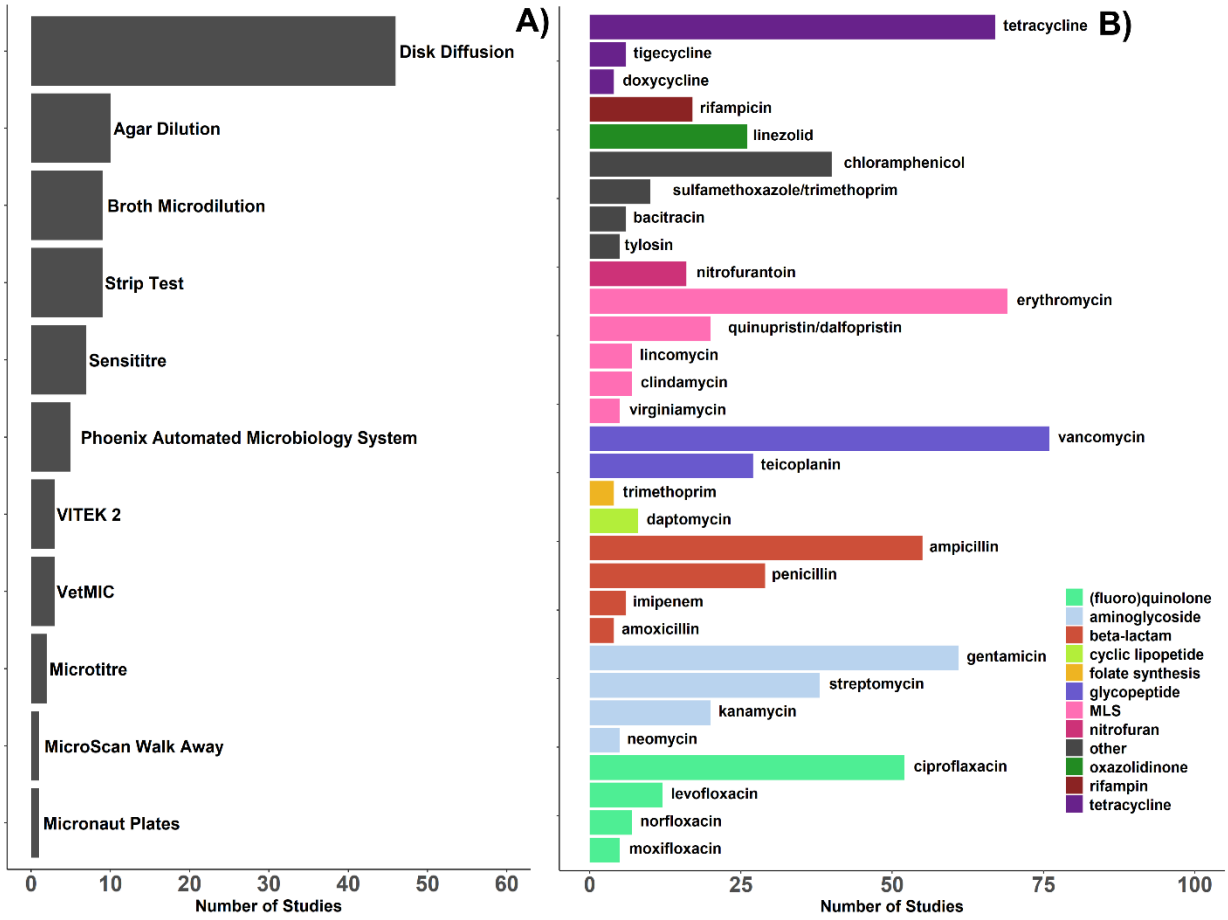


Figure 5-2: Summary of antibiotic susceptibility testing (AST) methods applied across the reviewed articles. A) Distribution of assays and commercialized platforms used for AST. B) Distribution of antibiotics used for screening enterococci isolate collections.

Multidrug Resistance Profiling

Because enterococci are intrinsically resistant to several antibiotics; including cephalosporins, penicillins, clindamycin, and aminoglycosides, resistant phenotypes are commonly found in environmental samples. Plasmid- and transposon-mediated resistance to tetracyclines, erythromycin, chloramphenicol, trimethoprim, vancomycin, and clindamycin; however, have further allowed the genus to become a leading cause of multidrug resistant nosocomial infections, particularly in the US (Murray, 1998). Modern nosocomial *E. faecium* isolates, for example, are commonly resistant to ampicillin, vancomycin, and high levels of aminoglycosides (Miller et al., 2014). Recently, the emergence of multidrug-resistant VRE to newer, last-resort antibiotics; including oxazolidinone-linezolid, daptomycin, quinupristin/dalfopristin, and tigecycline, have caused frequent treatment failures and are of global concern (Ahmed and Baptiste, 2018). Thus, screening for multidrug resistance amongst isolated environmental enterococci is essential for monitoring the evolution of the genus over time, specifically as a function of anthropogenic pollution, as well as assessing the relative hazard posed by the isolate. The choice of antibiotics to include in screening panels is crucial, as most are not useful from a risk-based monitoring framework, although some may highlight ecological relevance. For instance, screening for phenotypes that are intrinsic to the genus may

not have clinical relevance but may be useful in determining the distribution of endemic phenotypes to contextualize the relative frequency of VRE detection.

Biochemical and Molecular Speciation

Because the virulence and resistance characteristics across different *Enterococcus* spp. vary substantially, speciation of resistant *Enterococcus* is desirable. A suite of biochemical tests has been developed by clinical microbiologists to confirm *Enterococcus* to the genus level. These tests include Gram staining, catalase testing, thermal growth range/thermotolerance (growth at 10 and 45°C), halotolerance (6.5% NaCl), growth at pH 9.7, pyrrolidonylarylamidase activity, and the ability to hydrolyze esculin in the presence of bile salts (e.g. growth on bile esculin agar) (Facklam and Collins, 1989; Teixeira et al., 2015). These tests were common features of nearly three quarters of articles and served as a prerequisite for inclusion in downstream characterization, including further speciation (Figure 5-S1). Studies that did not confirm isolates to the genus level either relied on chromogenic agar (e.g., mEI, CHROMagar VRE) to select presumptive enterococci or speciated their library without screening for characteristic metabolisms or morphologies. Genus-specific primers based on the 16S rRNA gene (Deasy et al., 2000), 23S rRNA gene (EPA Method 1611) (EPA, 2012), or the elongation factor EF-Tu (*tuf*) (Ke et al., 1999) gene have also been used for rapid identification of the genus *Enterococcus*. However, 16S rRNA primer sets are known to fail to capture all *Enterococcus* spp. (Botina and Sukhodolets, 2006).

Speciation of enterococci libraries was common and performed in 86/105 articles, the most common approach of which was PCR. There are several conserved proteins and corresponding genes that are targeted in these assays. The simultaneous detection of enterococcal species and glycopeptide resistance was the first molecular approach to improve diagnostic speeds for clinical enterococci and was based on the detection of genes encoding D-alanine:D-alanine (*ddl*) ligases and other glycopeptide resistance determinants. A reduced affinity for glycopeptides in VanA- and VanB-type resistance in enterococci are due to the integration of D-alanyl:D-lacate into peptidoglycan precursors by the chromosomally-encoded *ddl* ligases (Dutka-Malen et al., 1995a). The *ddl* enzymes in *E. faecium* ($ddl_{E.\ faecium}$) and *E. faecalis* ($ddl_{E.\ faecalis}$) are conserved, and in resistant strains, these enzymes are present in addition to *vanA* or *vanB*. Similarly conserved ligases, *vanC1* and *vanC2-3* are highly specific for *E. gallinarum* (Dutka-Malen et al., 1992) and *E. casseliflavus* (Navarro and Courvalin, 1994), respectively. The primers published by Dutka-Malen et al. (1995) for $ddl_{E.\ faecium}$, $ddl_{E.\ faecalis}$, $vanC1_{E.\ gallinarum}$, and $vanC2\text{-}3_{E.\ casseliflavus}$ have been the most widely used for the speciation of environmental enterococci, as many researchers are specifically concerned with the identification of these four most common and clinically-relevant species (Table 5-S3). Kariyama et al. (2000) (Kariyama et al., 2000) and Depardieu et al. (2004) (Depardieu et al., 2004) provided additional multiplex PCR assays for more high-throughput approaches to VRE surveillance. A multiplex PCR assay based on species-specific superoxide dismutase (*sodA*) genes developed by Jackson et al. (2004) includes primers for 23 different enterococcal species (Jackson et al., 2004).

The PCR primers described above were in part developed due to a lack of consensus between commercially-available systems and kits, such as the Analytical Profile Index (API; bioMérieux), PhenePlate (PhPlate Microplate Techniques AB), Phoenix Microbiology Systems (BD Phoenix), VITEK (bioMérieux), Micronaut-Strep2 (MERLIN), MicroScan Walk Away

(Beckman Coulter), and BBL Crystal (MG Scientific) manual or automated rapid identification systems. The principles behind these higher-throughput systems are derived from conventional biochemical phenotyping of enterococci which involve differentiating carbohydrate fermentation of mannitol, sorbitol, sorbose, inulin, arabinose, melibiose, sucrose, raffinose, trehalose, lactose, glycerol, salicin, and maltose, among others (Facklam and Collins, 1989; Teixeira et al., 2015). The commercial methods employ a panel of biochemical tests in parallel to reduce the labor costs of manual phenotyping. However, if atypical species are present, these systems will struggle to identify the organism with acceptable levels of certainty (Castillo-Rojas et al., 2013). This issue is especially problematic in matrices outside of the clinical setting, as these systems were developed and validated targeting common clinical strains and reference cultures with distinguishable biochemical characteristics and not the wide phenotypic diversity of environmental samples. A comprehensive survey of these systems has been reviewed previously (Emery et al., 2016).

An emerging technology for the rapid identification of microorganisms is matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) (Singhal et al., 2015). This method uses lasers to generate singly protonated ions from analytes in the sample. In the process of identifying unknown microbes, these analytes are primarily housekeeping and ribosomal proteins, given that they constitute a large portion of the dry weight of microbial cells. The ionized proteins are then separated by their mass-to-charge ratio and depending on their time of flight through a channel, a peptide mass fingerprint is generated that can be compared to openly sourced databases. MALDI-TOF MS for the purpose of enterococci speciation emerged in the literature in 2017, as this is an emerging technique in environmental studies. MALDI-TOF MS systems are considerably more expensive than most molecular or phenotypic methods, but the throughput is generally larger (thousands of isolates) and can better accommodate the demand in clinical laboratories.

16S rRNA gene sequencing was less frequently used than other methods due its non-specificity. Other genes may be sequenced, including the *sodA*, *rpoA*, and *pheS* genes, which have shown to be more discriminatory than the 16S rRNA gene for closely-related species, such as *E. casseliflavus* and *E. flavescens* (Naser et al., 2005; Poyart et al., 2000). In a recent comparative study of various loci to differentiate closely-related *Enterococcus* spp., it was determined that the alpha subunits of ATP synthase (*atpA*), chaperonins (*groESL*), and phenylalnyl-tRNA synthase alpha subunits (*pheS*) performed equally well or better than 16S rRNA gene sequencing against 308 enterococci isolates from untreated urban wastewater (Sanderson et al., 2019). The rate of false identification of consensus reference strains based on loci sequencing was approximately 2%, much lower than the parallelized carbohydrate phenotyping systems discussed above, such as RapID STR, which had error rates of 15.9% for *E. faecalis*, 21.5% for *E. faecium*, and 56.9% for *E. casseliflavus/ gallinarum*. The gold standard for speciating enterococci, and any organism for that matter, is WGS (Sanderson et al., 2019).

Genotyping Resistant Enterococci

Numerous genetic determinants confer antibiotic resistance across the genus *Enterococcus*. Co-occurrence of resistance genes and virulence factors is of particular concern from a clinical standpoint and is common among nosocomial strains (Guzman Prieto et al., 2016; Pöntinen et al., 2021). Most genetic determinants of antibiotic resistance in *Enterococcus* spp.

are intrinsically encoded, i.e., they exist on the chromosome within the core genome of the genus. Typically, acquired resistance, i.e., a product of horizontal gene transfer, is of greater interest for monitoring, where the purpose is to examine trends in resistance patterns and if they are changing in time and space. Acquired resistance genes are of primary concern as drivers of failure of antibiotic treatment in clinical infections. Canonical mobile resistance determinants within enterococci include those that confer resistance to glycopeptides (*van* gene clusters), aminoglycosides (aac(6')-Ie-aph(2")-Ia and aph(3′)-IIIa), MLS (*ermB*), and tetracyclines (*tetM* and *tetL*) (Figure 5-S2).

Just over a third of analyzed articles (38/105) genotyped colonies for antibiotic resistance determinants after they had been isolated on antibiotic-containing media, 34 using PCR and 4 using WGS. The *van* operon was commonly targeted, with an emphasis on *vanA* and *vanB* within VRE isolates themselves (Figure 5-S2). There are nine distinct gene clusters conferring glycopeptide resistance in enterococci (VanA, B, C, D, E, G, L, M, N) (Hancock et al., 2014; Teixeira et al., 2015) and these determinants differ both genetically and phenotypically based on their physical location (encoded on mobile genetic elements or chromosomal), whether resistance is inducible or constitutive, the type of peptidoglycan precursor that is produced, and ultimately the level of resistance conferred. VanA gene clusters are the most common in clinical isolates and are typically found on Tn*1546*-like transposons, are frequently integrated into a wide range of plasmids, and produce clinical levels of resistance to vancomycin (MIC 64-1,000 µg/mL) and teicoplanin (MIC 16-512 µg/mL) (Teixeira et al., 2015). Similar to VanA, VanB gene clusters are also typically found in clinical isolates and are present on transposons (Tn*1547* or Tn*1549* to Tn*5382*), but differ from *van*A due their inability to recognize teicoplanin, allowing strains with the VanB phenotype to remain susceptible (Miller et al., 2014). These two gene clusters are the most significant genetic determinants in clinically-resistant enterococci and several PCR assays have been developed for their detection (Dutka-Malen et al., 1995b; Kariyama et al., 2000; Nam et al., 2013; Rathnayake et al., 2011). Enterococci displaying susceptible to intermediate resistance are typically attributed to chromosomally encoded *van* clusters, like *vanC1* in *E. gallinarum* and *vanC2/3* in *E. casseliflavus*, which are commonly detected in environmental samples. The much more rare *vanD-N* genotypes were not detected in any articles that screened for them (Kotzamanidis et al., 2009; Taučer-Kapteijn et al., 2016; Zdragas et al., 2008).

Virulence Factors and Pathogenesis

The pathogenesis of infections caused by enterococci is still poorly understood (Teixeira et al., 2015). However, several PCR assays have been developed for the detection of virulence factors common to *Enterococcus*, including: surface adhesion proteins (*esp*), aggregation substances (*agg*), cytolysin (*cyl*) and hemolysin (*hyl*) secretion operons, collagen adhesion (*ace*), and gelatinase secretion proteins that are predominantly found in endocarditis isolates (*gelE*) (Eaton and Gasson, 2001; Mannu et al., 2003; Vankerckhoven et al., 2004). These virulence factors were primarily screened for in the studied articles due to their implications in pathogenesis routes from environmental reservoirs back into humans and animals. Only 18 of the studies identified screened for virulence factors, the most common being *esp* (all 18 articles), *cyl*, and *gelE* (Figure 5-S2). The first vancomycin-resistant *E. faecalis* strain documented in the US was revealed to carry a large, transmissible pathogenicity island containing both *esp* and a complete *cyl* operon, and several other functions that are non-essential to commensal behavior of

the organism (Shankar et al., 2002). The *E. faecalis* pathogenicity island is an integrative conjugative element that can mobilize between plasmids and chromosomes in *E. faecalis* and *E. faecium,* transferring virulence factors and antibiotic resistance determinants (Laverde Gomez et al., 2011; Manson et al., 2010). The co-occurrence of virulence with antibiotic resistance is a key consideration when screening the environment for the emergence of potentially hyper-virulent strains. Rathnayake et al. (2012) found significant correlations between the presence of virulence factors and phenotypic antibiotic resistance among both *E. faecium* and *E. faecalis* isolates in surface waters and regional clinical isolates in Australia (Rathnayake et al., 2012). Similarly, Lata et al. (2016) documented widespread co-occurrence of *vanA* and *vanB* genotypes with *gelE, ace, efaA,* and *esp* virulence factors in both *E. faecalis* and *E. faecium* in impacted surface waters in northern India (Lata et al., 2016). Such studies demonstrate the value of monitoring both antibiotic resistance and virulence, particularly for gaining insight into ecological factors at play in observed resistance patterns.

Multilocus Sequence Typing

Multilocus sequence typing (MLST) is a technique by which multiple loci, or specific internal DNA fragments within an organism's genome, are amplified by PCR, sequenced, and then compared across multiple isolates of that species. The loci are typically housekeeping genes common to the genus, and the allelic composition of the set of targets determines the "sequence type", allowing for the determination of clonal complexes (CCs; isolates with differences of no more than 2 loci) and potential source attribution of isolates. Sequence types can then be compared to publicly available and curated databases, such as PubMLST (Jolley et al., 2018), where global isolate comparisons can be made. Pulse-field gel electrophoresis is a similar technique to MLST and was previously known as the "gold standard" for source attribution and epidemiological linkages of bacterial isolates, but low interlaboratory reproducibility and inability to perform phylogenetic or population structure studies makes it unsuitable for global, long-term epidemiological studies (Nemoy et al., 2005). MLST profiles of *Enterococcus* have only been developed for *faecium* (Homan et al., 2002) and *faecalis* (Ruiz-Garbajosa et al., 2006) and are therefore the only two species present in the PubMLST database. The two profiles are derived from a mixed set of 11 housekeeping genes: *gdh* (glucose-6-phosphate dehydrogenase), *purK* (phosphoribosylaminoimidazol carboxylase ATPase subunit), *pstS* (phosphate ATP-binding cassette transporter), *atpA* (ATP synthase, alpha subunit), *gyd* (glyceraldehyde-3-phosphate dehydrogenase), *adk* (adenylate kinase), *ddl*, *gki* (glucokinase), *aroE* (shikimate 5-dehydrogenase), *xpt* (shikimate 5-dehydrogenase), and *yqiL* (acetyl-coenzyme A acetyltransferase), which were chosen for their low ratios of nonsynonymous to synonymous mutations and their dispersed locations on the chromosomes. MLST is useful when exploring potential clonal relationships between *Enterococcus* spp. isolated from the environment and those from critical AMR monitoring points such as hospital wastewaters and pharmaceutical production waste. The clonal complex 17 (CC17) of *Enterococcus faecium,* for example, is a nosocomial strain associated with outbreaks worldwide and is generally ampicillin and quinolone resistant and contains the *esp* surface adhesion protein (Top et al., 2008). CC17 has been detected in several environmental samples and is an indication of the interconnectedness of the environment and clinical wastewater streams (Caplin et al., 2008).

Whole Genome Sequencing

Despite the development of several sophisticated molecular biological assays for the complete characterization of enterococcal isolates over the last few decades, the advent and proliferation of next-generation sequencing techniques has allowed for comprehensive and high-throughput functionality of all the previous assays in a singular method (Figure 5-1). WGS of isolates allows for the simultaneous detection of ARGs, virulence factors, plasmids, bacteriophages, insertion sequences, transposons, and the sequence type and cladal relatedness of isolates that can be comparted with enterococcal libraries globally. Only 4 articles performed WGS in this review, indicating that comprehensive epidemiological analysis of antibiotic resistant, virulent, hospital-adapted enterococcal clades is largely absent from the environmental literature. In WGS workflows, resistant enterococci are initially screened for on selective media supplemented with antibiotics (in this case ampicillin or vancomycin) and then isolated and speciated before they are subject to sequencing. Genomic surveillance of *E. faecium* isolates from retail meat, patients with bloodstream infections, and wastewater treatment plants revealed distinct clades with limited sharing of ARGs between livestock and humans in the UK (Gouliouris et al., 2018). There was, however, extensive overlap between isolates from bloodstream infections and those from the influents and effluents of 17 different wastewater treatment plants (WWTPs) in the region, which could indicate the emergence of new lineages of *E. faecium* that are both hospital-adapted and persist in the environment (Gouliouris et al., 2019). Similarly, Ekwanzala et al. (2020) investigated the prevalence of VRE in hospital wastewater, municipal wastewater, and the receiving surface water in South Africa and found that 35% of the enterococci exiting the wastewater treatment plant were vancomycin resistant, leading to the greatest VRE loadings in the downstream sediment (Ekwanzala et al., 2020a). Subsequent comparative genomics found that ST40, a human pathogenic *E. faecalis* sequence type, and CC17 of *E. faecium* were found persisting in downstream sediments, posing a risk to human health, and demonstrating the need for more advanced wastewater treatment in this scenario. Although WGS is more expensive and difficult to perform than PCR-based genotyping, its high-throughput and robust analysis is quickly becoming commonplace as sequencing costs continue to fall. Also, the storage and sharing of sequenced genomes to public databases allows for longitudinal, phylogenetic tracking of problematic clones as they are transmitted globally (van Hal et al., 2021).

Trends in Total and Antibiotic-Resistant *Enterococcus* Found in Water Environments

A comprehensive meta-analysis was performed to identify overarching trends in the species distribution of generic and antibiotic-resistant *Enterococcus* spp. to illuminate general trends and inform sampling priorities, extracting data from all articles that reported resistance as a percentage of the total number of isolates in a collection. Together, this meta-collection consisted of 39,514 isolates extracted from 80/105 articles. To reduce the amount of bias introduced by sampling and enrichment procedures, only *Enterococcus* AST data that was generated in the absence of an initial selective antibiotic were used. This reduced the collection size for AST data to 16,593 isolates extracted from 68 articles but allowed for an estimation of the "true" phenotypic diversity of environmental antibiotic-resistant *Enterococcus*. This approach also allows for the empirical prioritization of monitoring targets for *Enterococcus* resistant to critical antibiotics as well as a baseline for further studies across different water matrices without a bias towards resistant populations.

Total Enterococci

Only 52 of the 86 articles in which libraries were speciated were the data provided in a format that could be extracted unambiguously (e.g., in tabular format) or detailed population statistics reported, resulting in a collective of 27,273 speciated isolates for meta-analysis (Figure 5-3). The three most common *Enterococcus* spp. across the 52 articles were *E. faecium* (34.0%), *E. faecalis* (29.0%), and *E. hirae* (13.0%). Minor species such as *E. raffinosus, E. avium,* and *E. pseudoavium* (Table S3) each represented less than 2% of the total isolates, although this is likely influenced by underrepresentation of these species in common PCR confirmation assays. In fact, in many articles, only *E. faecalis, E. faecium*, *E. gallinarum*, and *E. casseliflavus* were screened, as these species represent the most encountered resistant clinical isolates. Despite the uneven representation of the number isolates from each water matrix, there were clear trends among enterococcal populations. In WWTP and hospital wastewater samples, the combined proportions of *E. faecium, E. faecalis,* and *E. hirae* were nearly identical, making up approximately 90% of all isolates (Figure 5-3). These proportions are consistent with the natural distributions of *Enterococcus* spp. in the gastrointestinal tract of healthy human adults and animals (Silva et al., 2012). The proportions of non-fecal related or undefined *Enterococcus* spp. were greater in surface water and coastal water samples, consistent with dilution of fecal contamination with environmental strains. A large proportion *E. hirae*, the dominant species excreted by cattle (Jackson et al., 2011), was found in cattle feedlot drains examined by (Zaheer et al., 2020). The even proportions across comparable water matrices (e.g., WWTPs and hospital wastewater) suggest that there is a lack of systematic biases in either media selectivity or speciation techniques.

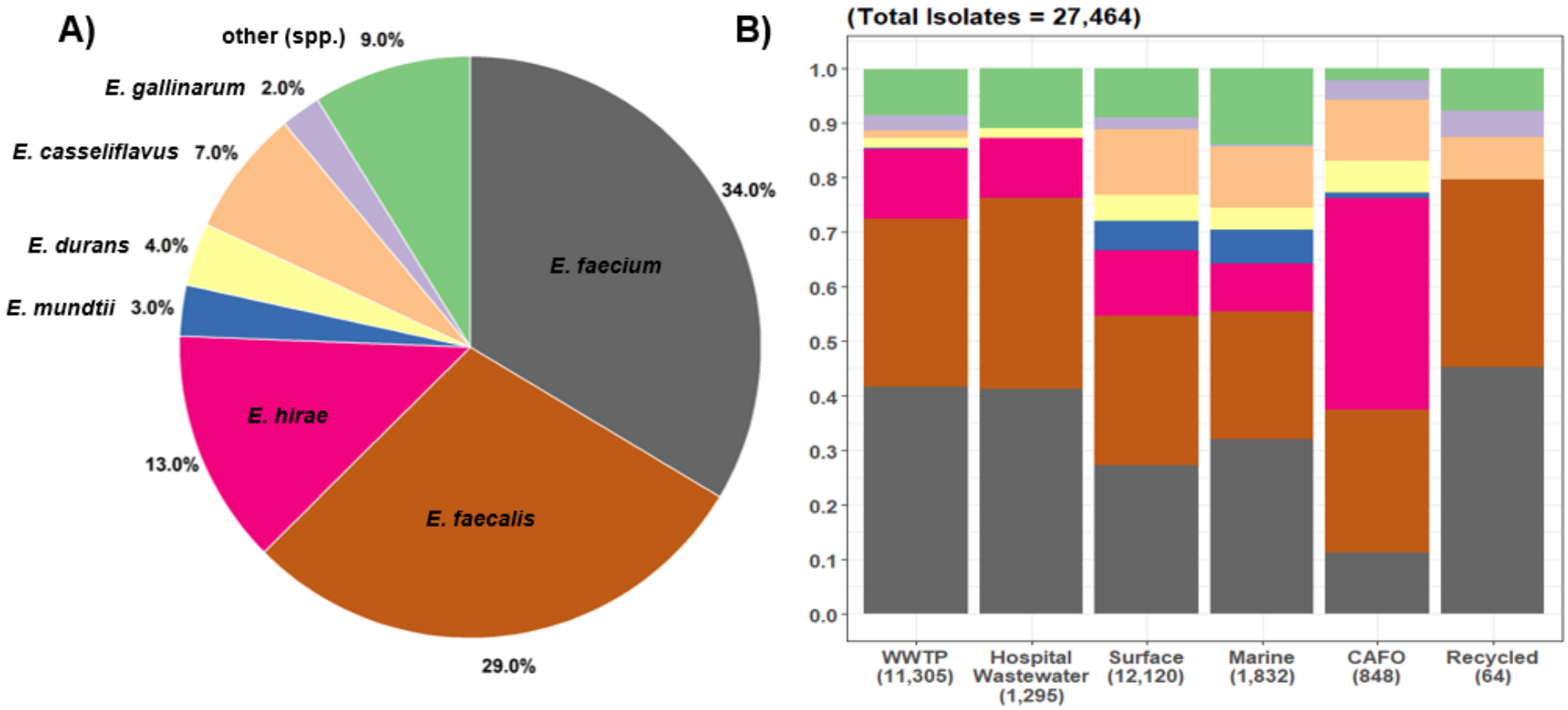


Figure 5-3: Distribution of *Enterococcus* spp. isolates across all studies. A) Species distribution of the total speciated meta-library across all water types (27,464 isolates). "other (spp.)" is any species that was not detected at >2% of the total meta-library abundance or was reported only as *Enterococcus* spp. B) The distribution of generic enterococcal isolates by water matrix. The total number of isolates representing each matrix is indicated in parentheses. Concentrated animal feeding operation (CAFO) samples encompass wastewater lagoons, feedlot drains, and any on-site treatment systems. WWTP samples encompass all treatment stages.

Surface Water

The majority of isolates obtained across studies were derived from surface water, which included both urban and rural watersheds and riverine sites directly impacted by municipal or hospital wastewater discharge (5,530 isolates) (Figure 5-3). Compared to the other environments examined, surface waters displayed some of the lowest percentages of resistant isolates to all 18 antibiotics and antibiotic classes. It should be noted that, despite dilution effects and environmental attenuation, 5.6 % of 8,538 *Enterococcus* spp. isolates were phenotypically resistant to vancomycin. The vast majority of these isolates were confined to *E. faecium* and *E. faecalis.* By contrast, vancomycin resistance was virtually non-existent in *E. casseliflavus*, *E. gallinarum* and the other less dominant species (Figure 5-3; Table 5-S3). Despite intrinsic resistance of *E. casseliflavus* and *E. gallinarum* by *vanC* genes, their phenotypes rarely exceeded CLSI breakpoints for clinical resistance. Łuczkiewicz et al. (2010) examined *Enterococcus* resistant to 13 different antibiotics in an urban river system in Poland in the absence of wastewater treatment plant discharge and found that resistance to erythromycin, ciprofloxacin, and tetracycline was common among all isolates (Łuczkiewicz et al., 2010). They also found multidrug resistance (some to all 13 antibiotics tested), including vancomycin and high-level aminoglycoside resistance, among *E. faecalis* and *E. faecium* isolates in the two main tributaries feeding the coastal waters. The authors suggested that riverine enterococci should be considered as a potential risk for downstream recreational bathers, even in the absence of point-source wastewater pollution. In contrast, studies of rural watersheds (< 1 % urban) in Ontario (Canada) and Georgia (US) found that the diversity and distribution of antibiotic resistance among *Enterococcus* were strikingly different than in more anthropogenically-impacted waterways (Cho et al., 2019; Lanthier et al., 2011). These two studies, together comprising 2,195 isolates, indicate that the enterococcal species and their phenotypes were stochastically distributed and sparse, with few multidrug (< 6 antibiotics) resistant strains and no isolates reaching the CLSI breakpoints for vancomycin, teicoplanin, or linezolid. They attributed the dispersion of the resistant fecal indicators to domesticated animal and wildlife fecal pollution and their dissemination to not be significant. The most significant rates of resistance among *Enterococcus* isolates from surface water studies came from the North West province of South Africa where 86/124 *Enterococcus* spp. were phenotypically resistant to vancomycin. These were isolated in the absence of a selective antibiotic and displayed multidrug resistance to ampicillin, amoxicillin, penicillin, ciprofloxacin, erythromycin, and tetracycline. Interestingly, a single isolate of *E. sulfureus* was found to be multidrug resistant to ampicillin, amoxicillin, penicillin, streptomycin, vancomycin, chloramphenicol, ciprofloxacin, erythromycin, and tetracycline (Molale and Bezuidenhout, 2016).

Upstream and downstream sampling of municipal wastewater discharge was carried out in several studies. The detection of enterococci with resistance to clinically-relevant antibiotics downstream of municipal wastewater discharge was regionally dependent and linked to the degree of treatment employed by the WWTP. Ben Said et al. (2014) found no ampicillin-resistant *Enterococcus* (ARE) or VRE in the influent, effluent, or receiving water in Tunisia, suggesting the regional variation in their distribution (Ben Said et al., 2015). Further, Bessa et al. (2014) found *vanA* positive *E. faecium* mostly *upstream* of WWTP discharge in Portugal (Bessa et al., 2014), a region where VRE has become endemic.

Municipal and Hospital Wastewater

Municipal and hospital wastewater were commonly sampled across the identified articles. Together they accounted for 6,450 isolates in the meta-collection and, as one might expect, they displayed the highest rates of clinically-relevant phenotypes (Figure 5-4). *Enterococcus* spp. isolated from untreated hospital wastewater displayed the highest rates resistance to fluoroquinolones, ampicillin/penicillin, macrolides, and vancomycin.

WWTPs, which encompassed raw influent, activated sludge, secondary effluents, and variably treated final effluents, yielded *Enterococcus* spp. isolates with the highest rates of resistance to cephalosporins, lincosamides, nitrofurantoin, and teicoplanin among *Enterococcus* spp. isolated from 15 different countries. Counterintuitively, within the "WWTP" category, only 2.6% of 6,519 isolates were resistant to vancomycin in the meta-collection. Interestingly, when the isolates were binned into two separate categories, "WWTP Influent" and all other compartments of the WWTP, resistance percentages were highly similar: 136 of 5,279 isolates (2.57%) for influent samples and 33 of 1140 (2.89%) for all other treatment compartments. Thus, the meta-analysis revealed striking patterns in the geographical distribution of vancomycin resistance. Because the influent was disproportionately sampled, the number of VRE was likely diluted by regions in which VRE is not endemic.

Although they represented a smaller fraction of isolates, hospital wastewater environments showed the most significant contributions of phenotypic resistance to receiving waters. For example, Novais et al. (2005) found statistically significant increases in phenotypic resistance exceeding CLSI breakpoints among *E. faecium* and *E. faecalis* for vancomycin, teicoplanin, ciprofloxacin, and ampicillin in urban sewers receiving hospital wastewater discharge in Portugal (Novais et al., 2005). Clonal analysis, Tn*1546* typing, and virulence factor assays were also consistent with local clinics being the source of downstream surface water isolates. Similar observations were made along a medical center-retirement home-wastewater treatment plant-river continuum in France in 2013, where the hospital-adapted clonal complex, CC17 *E. faecium,* was culturable along the continuum and into receiving waters (Leclercq et al., 2013). The CC17 concentration, though, was attenuated by the WWTP and the proportion of CC17 became outweighed by environmental strains. Further, epidemiological source tracking using WGS of VRE isolated from 20 WWTPs in the UK in 2019 determined that there was widespread dissemination of hospital-adapted *E. faecium* in WWTP effluents across eastern England (Gouliouris et al., 2019). They found that WWTPs receiving hospital wastewater had significantly higher VRE and ARE concentrations than non-hospital associated treatment plants and found highly similar isolates shared between the local teaching hospital and those emitted from surrounding WWTPs. Together, these studies provide strong support of *Enterococcus* spp. as both a clinically-relevant and environmentally-relevant target for waterborne monitoring. Hospital wastewaters, the municipal wastewater infrastructure responsible for treatment, and receiving waters are key monitoring points for tracking their dissemination.

Pharmaceutical Wastewater

Pharmaceutical wastewater is a critical monitoring point in the dissemination of resistant microorganisms and is currently understudied in the field. Only two such articles focused on resistant *Enterococcus* were encountered in this review. Guardabassi et al. (2002) documented invariably high numbers of VRE (20 ug/mL vancomycin; 10^6 CFU/mL) in the waste biomass

from the fermentation tanks used in the production of vancomycin (Guardabassi et al., 2002). The WWTP treating the waste was also enriched with presumptive VRE, and quantifiable CFUs were routinely emitted from the plant after secondary clarification without disinfection. PFGE analysis found identical VRE patterns between vancomycin production fermentation waste and the final effluents of the WWTP, suggesting that pharmaceutical production waste can be a direct contributor to the dissemination of VRE into the environment. Further, Guardabassi et al. (2004) investigated the relationship of Tn*1546*-like elements in *Enterococcus* spp. isolated from municipal sewage, activated sludge, vancomycin production waste, human feces, mussels, and soil using long PCR-restriction fragment length polymorphism and found indistinguishable elements shared across the ecologically distinct locations and between enterococcal species, suggesting ready transferability of the VanA genotype between clinical and environmental strains (Guardabassi and Dalsgaard, 2004).

Recycled Water

Due to the increased pressures on freshwater around the world, wastewaters are increasingly being treated and reintroduced into water and food cycles as the recycled water is used for crop irrigation, groundwater recharge, and even direct potable reuse. Only three studies in the current review examined resistant *Enterococcus* in wastewaters intended for reuse. Goldstein et al. (2014) sampled two WWTPs in the Mid-Atlantic and two from the Midwest regions of the US that reuse their treated effluents and detected VRE in 27% of wastewater samples, with higher rates in the Mid-Atlantic plants. VRE were only detected in final treated effluents when there was lack of chlorination (Goldstein et al., 2014). Subsequent studies from the same WWTPs found that VRE are detectable at low concentrations at the point of use after recovery from UV disinfection, although other phenotypes are more prevalent (Carey et al., 2016). Both WWTPs studied receive hospital wastewater and their effluents were used for spray irrigation (Goldstein et al., 2012).

Marine Waters and Recreational Beaches

Freshwater and marine water environments used for recreational bathing are at the direct interface between environmental fecal pollution and human exposure and are therefore important monitoring points. Studies of enterococcal populations of marine and freshwater beaches from Spain, Puerto Rico, Poland, Greece, Malaysia, Brazil, Italy, and Michigan confirmed that many drug resistant strains are readily culturable in recreational marine water and sand (Alm et al., 2014; Arvanitidou et al., 2001; Dada et al., 2013; de Oliveira and Watanabe Pinhata, 2008; Monticelli et al., 2019; Sadowy and Luczkiewicz, 2014; Santiago-Rodriguez et al., 2013; Tejedor Junco et al., 2001). High phenotypic and phylogenetic diversity was observed across all studies, and many non-fecal associated enterococci dominated local populations, including *E. casseliflavus* (Monticelli et al., 2019), *E. hirae* (Sadowy and Luczkiewicz, 2014), and *E. avium* and *E. raffinosus* (Arvanitidou et al., 2001). Beach sands are of particular interest as they represent a niche environment for enterococci where horizontal gene transfer can occur at higher frequencies than among planktonic bacteria, accelerating the rate at which fecal microbiota exchange genes with pathogens and engendering human-pathogen interactions (Alm et al., 2014; Oravcova et al., 2017).

CAFOs and Irrigation Water

Studies of the effects of cattle and swine concentrated animal feeding operation (CAFO) wastewater on downstream environments were also prevalent in the literature. The enterococcal isolates from CAFO studies in the meta-library showed the highest resistance prevalence to tetracyclines, which is not surprising given that tetracyclines make up nearly 40% of all antimicrobials used in animal husbandry in the US. (CDC, 2013). As many CAFOs contain and treat their wastewater on-site, their direct impact on groundwater or downstream surface water (e.g., due to runoff or unintentional discharge) was a concern. The animal products themselves also represent a potential direct line of exposure from animals to humans through the food chain.

Sapkota et al. (2007) sampled upstream and downstream of a high-density swine operation in the Mid-Atlantic region of the US and found higher MICs for clindamycin and tetracycline amongst the *Enterococcus* spp. in both downstream groundwater and surface water (Sapkota et al., 2007). Similarly, Stine et al. (2007) sampled the waste lagoons, surface waters, and well water of a swine CAFO in the US. that had been administering tetracycline-containing feed for over 20 years and found that 68% of all enterococci were resistant to tetracyclines, and a total of 60 different species displayed phenotypic resistance to tetracyclines across the sites (Stine et al., 2007). Further evidence for the direct dissemination of clinically-relevant enterococci into ambient surface waters by CAFOs was documented by Jahne et al. (2015) (Jahne et al., 2015). They documented a cattle CAFO and its on-site wastewater treatment system comprising of an infiltration basin with subsequent sequestration by a constructed wetland. Enterococci that displayed co-resistance to vancomycin, linezolid, and daptomycin were common in the wastewater and, during rain events, the increased hydraulic loading on the infiltration basin and constructed wetland resulted in the direct emission of these organisms into downstream surface waters.

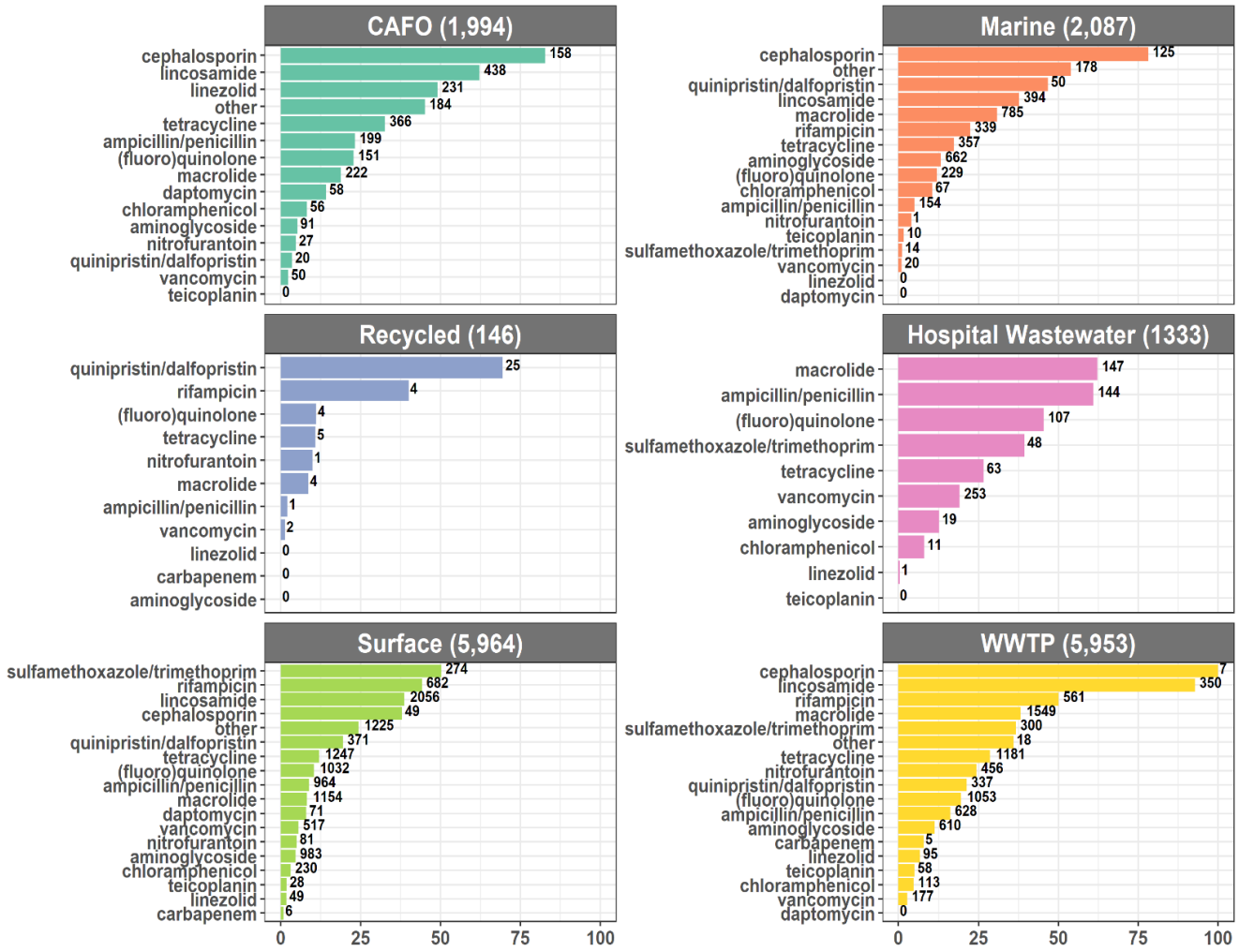


Figure 5-4: Distribution of percent of Enterococcus isolates resistant to each antibiotic by sample matrix. Isolates were only included if they were initially isolated in the absence of a selective antibiotic. The number of isolates representing each matrix is in parentheses in the headers. The number of resistant isolates per antibiotic tested is represented adjacent to each bar in each panel. CAFO = concentrated animal feeding operation, WWTP = wastewater treatment plant.

Conclusion and Recommendations

In this review we recounted the last 20 years of research assessing antibiotic-resistant enterococci in various water environments. Because *Enterococcus* has been shown to be a reliable indicator of fecal contamination of water bodies, several nationally and internationally recognized standard culture methods have been developed for their enumeration. Various *Enterococcus* spp. are both clinically-relevant and survive and persist in the environment. The studies surveyed here lay the groundwork for considering *Enterococcus* spp. as a standardized target for waterborne monitoring of antibiotic resistance.

Recently, the WHO put forward a standardized, comprehensive surveillance program for One Health-inspired monitoring, i.e., the Tricycle protocol (WHO, 2021), which targets the Gram-negative ESBL *E. coli*. Here, *Enterococcus* spp., as Gram-positive organisms, represent a compelling target to consider as a complement to such monitoring programs. *Enterococcus* spp. display resistance to critically-important antibiotics that would not be captured by an *E. coli*-targeted monitoring program alone. As observed in this critical review, *Enterococcus* spp. also display sensitive responses to anthropogenic pollution, including hospitals and CAFOs, that are apparent in their distinct geographical occurrence patterns.

The recommended path forward for standardizing environmental antibiotic-resistant *Enterococcus* monitoring should ensure the comparability of monitoring points, methods employed, and reporting metrics. Accessibility and ease of application are also important considerations. Ideally, *Enterococcus* spp. monitoring for the purpose of antibiotic resistance surveillance could be incorporated into existing monitoring programs, especially considering the high level of existing standardization and regulatory requirements. The existence of nationally and internationally recognized standard culture methods is of great value in ensuring comparability of the data gathered in space and time, however, further standardization is needed for the purpose of antibiotic resistance surveillance specifically. The conclusions and recommendations based on this critical review are as follows:

- mEI is a prime candidate for a standard selective media, given that it yields the highest selectivity over mEnterococcus and Slanetz-Bartley and is integrated into existing regulatory recommendations in the US. Still, the findings here were encouraging that studies are generally consistent, even if different media were employed, but confirmation of the genus is critical because no *Enterococcus* media is 100% selective. The addition of a selective antibiotic to the media can decrease specificity by selecting for intrinsically-resistant, non-target genera.
- In line with recommendations made in the WHO Tricycle protocol, plating the environmental sample on the selective media with and without a primary selective antibiotic produces both a percent resistance of the enterococci population and a CFU/unit volume measurement (e.g., CFU/ml). A CFU/mL measurement represents a universally comparable magnitude of antibiotic-resistant bacteria.
- Depending on the research question, vancomycin and other antibiotics used for primary selection may be added to media at the breakpoint for full, intermediate, or low resistance. The full resistance breakpoint is the most useful for clinical relevance, but in some environments, one may be interested in intermediate resistance. Where one expects bacteria to be compromised (injured), use of low-level antibiotic in enrichment cultures may be useful for lowering detection limits. The use of low/intermediate breakpoints may complicate human-health risk assessments due to the isolation of intrinsically-resistant species that are common in the environment.
- A defined set of key antibiotics aimed at treating VRE infections could also be employed for monitoring emerging phenotypes and multidrug resistance. These include ampicillin, teicoplanin, oxazolidinone-linezolid, daptomycin, quinupristin/dalfopristin, and tigecycline. The emergence of multidrug resistant VRE that are co-resistant to these 6 antibiotics are of great concern.
- AST using disk diffusion on Mueller-Hinton agar or agar dilutions in Mueller-Hinton broth should be followed according to the most up to date CLSI guidelines.

- Speciation and virulence typing of resistant enterococci is of interest for risk assessment and longitudinally tracking changing genotypes and phenotypes across the genus. PCR-based approaches are appropriate for low-tech labs, specific loci sequencing, and emerging technologies such as MALDI-TOF MS will increase throughput without sacrificing accuracy.
- Key monitoring points to consider for the dissemination of resistant enterococci are the hospital-municipal wastewater continuum and their receiving water bodies, especially where they impact recreational waters. Pharmaceutical wastewater and recycled water are critically understudied for resistant *Enterococcus.* The surrounding areas of CAFOs are of particular concern after storm events.
- Whenever possible, isolation of resistant colonies for WGS will aid in determining the sequence type, virulence genotype, plasmid type, acquired AMR genes, and chromosomal point mutations. Public sharing of WGS data will help to advance understanding of the ecology, epidemiology, and global transmission of this important pathogen.

Declaration of Competing Interest

None.

Acknowledgements

This works was supported by the Water Research Foundation (WRF) under project #5052: Standardizing Methods with QA/QC Standards for Investigating the Occurrence and Removal of Antibiotic Resistant Bacteria/Antibiotic Resistance Genes (ARB/ARGs) in Surface Water, Wastewater, and Recycled Water. Support was also provided by National Science Foundation Awards OAC 2004751, OISE 1545756, NRT 2125798, and ECCS 2025151.

References

Abbott, S., Caughley, B., Scott, G., 1998. Evaluation of Enterolert® for the enumeration of enterococci in the marine environment. New Zeal. J. Mar. Freshw. Res. 32, 505–513. https://doi.org/10.1080/00288330.1998.9516839

Adcock, P.W., Saint, C.P., 2001. Development of glucosidase agar for the confirmation of water-borne Enterococcus. Water Res. 35, 4243–4246. https://doi.org/10.1016/S0043-1354(01)00125-7

Agency, E.E., 2020. Bathing water management in Europe: Successes and challenges.

Ahmed, M.O., Baptiste, K.E., 2018. Vancomycin-Resistant Enterococci: A Review of Antimicrobial Resistance Mechanisms and Perspectives of Human and Animal Health. Microb. Drug Resist. 24, 590–606. https://doi.org/10.1089/mdr.2017.0147

Alexander, J., Hembach, N., Schwartz, T., 2020. Evaluation of antibiotic resistance dissemination by wastewater treatment plant effluents with different catchment areas in Germany. Sci. Rep. 10, 1–9. https://doi.org/10.1038/s41598-020-65635-4

Alm, E.W., Zimbler, D., Callahan, E., Plomaritis, E., 2014. Patterns and persistence of antibiotic resistance in faecal indicator bacteria from freshwater recreational beaches. J. Appl.

Microbiol. 117, 273–285. https://doi.org/10.1111/jam.12512

Amarasiri, M., Sano, D., Suzuki, S., 2020. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. Crit. Rev. Environ. Sci. Technol. 50, 2016–2059. https://doi.org/10.1080/10643389.2019.1692611

APHA, 1999. Standard Methods for the Examination of Water and Wastewater, 20th Edition.

Arias, C.A., Murray, B.E., 2012. The rise of the Enterococcus: Beyond vancomycin resistance. Nat. Rev. Microbiol. 10, 266–278. https://doi.org/10.1038/nrmicro2761

Arvanitidou, M., Katsouyannopoulos, V., Tsakris, A., 2001. Antibiotic resistance patterns of enterococci isolated from coastal bathing waters. J. Med. Microbiol. 50, 1001–1005. https://doi.org/10.1099/0022-1317-50-11-1001

Audicana, A., Perales, I., Borrego, J.J., 1995. Modification of kanamycin-esculin-azide agar to improve selectivity in the enumeration of fecal streptococci from water samples. Appl. Environ. Microbiol. 61, 4178–4183. https://doi.org/10.1128/aem.61.12.4178-4183.1995

Ben Said, L., Klibi, N., Lozano, C., Dziri, R., Ben Slama, K., Boudabous, A., Torres, C., 2015. Diversity of enterococcal species and characterization of high-level aminoglycoside resistant enterococci of samples of wastewater and surface water in Tunisia. Sci. Total Environ. 530–531, 11–17. https://doi.org/10.1016/j.scitotenv.2015.05.091

Bessa, L.J., Barbosa-Vasconcelos, A., Mendes, Â., Vaz-Pires, P., Da Costa, P.M., 2014. High prevalence of multidrug-resistant Escherichia coli and Enterococcus spp. in river water, upstream and downstream of a wastewater treatment plant. J. Water Health 12, 426–435. https://doi.org/10.2166/wh.2014.160

Blanch, A.R., Caplin, J.L., Iversen, A., Kühn, I., Manero, A., Taylor, H.D., Vilanova, X., 2003. Comparison of enterococcal populations related to urban and hospital wastewater in various climatic and geographic European regions. J. Appl. Microbiol. 94, 994–1002. https://doi.org/10.1046/j.1365-2672.2003.01919.x

Boehm, A.B., Sassoubre, L.M., 2014. Enterococci as Indicators of Environmental Fecal Contamination. Enterococci From Commensals to Lead. Causes Drug Resist. Infect. 1–21.

Botina, S.G., Sukhodolets, V. V., 2006. Speciation in bacteria: Comparison of the 16S rRNA gene for closely related Enterococcus species. Russ. J. Genet. 42, 247–251. https://doi.org/10.1134/S1022795406030033

Brodsky, M.H., Schiemann, D.A., 1976. Evaluation of Pfizer selective enterococcus and KF media for recovery of fecal streptococci from water by membrane filtration. Appl. Environ. Microbiol. 31, 695–699. https://doi.org/10.1128/aem.31.5.695-699.1976

Budnick, G.E., Howard, R.T., Mayo, D.R., 1996. Evaluation of enterolert for enumeration of enterococci in recreational waters. Appl. Environ. Microbiol. 62, 3881–3884. https://doi.org/10.1128/aem.62.10.3881-3884.1996

Byappanahalli, M.N., Nevers, M.B., Korajkic, A., Staley, Z.R., Harwood, V.J., 2012. Enterococci in the Environment. Microbiol. Mol. Biol. Rev. 76, 685–706.

https://doi.org/10.1128/mmbr.00023-12

Caplin, J.L., Hanlon, G.W., Taylor, H.D., 2008. Presence of vancomycin and ampicillin-resistant Enterococcus faecium of epidemic clonal complex-17 in wastewaters from the south coast of England. Environ. Microbiol. 10, 885–892. https://doi.org/10.1111/j.1462-2920.2007.01507.x

Carey, S.A., Goldstein, R.E.R., Gibbs, S.G., Claye, E., He, X., Sapkota, A.R., 2016. Occurrence of vancomycin-resistant and -susceptible Enterococcus spp. in reclaimed water used for spray irrigation. Environ. Res. 147, 350–355. https://doi.org/10.1016/j.envres.2016.02.030

Castillo-Rojas, G., Mazari-Hiríart, M., Ponce de León, S., Amieva-Fernández, R.I., Agis-Juárez, R.A., Huebner, J., López-Vidal, Y., 2013. Comparison of Enterococcus faecium and Enterococcus faecalis Strains Isolated from Water and Clinical Samples: Antimicrobial Susceptibility and Genetic Relationships. PLoS One 8, 1–10. https://doi.org/10.1371/journal.pone.0059491

CDC, 2013. Antibiotic Resistance Threats in the United States, 2013 37, 828–831. https://doi.org/10.1016/j.medmal.2007.05.006

Centers for Disease Control and Prevention, 2019. Vacomycin-Resistant Enterococci.

Cho, S., Hiott, L.M., McDonald, J.M., Barrett, J.B., McMillan, E.A., House, S.L., Adams, E.S., Frye, J.G., Jackson, C.R., 2019. Diversity and antimicrobial resistance of Enterococcus from the Upper Oconee Watershed, Georgia. J. Appl. Microbiol. 128, 1221–1233. https://doi.org/10.1111/jam.14550

Dada, A.C., Ahmad, A., Usup, G., Heng, L.Y., Hamid, R., 2013. High-level aminoglycoside resistance and virulence characteristics among Enterococci isolated from recreational beaches in Malaysia. Environ. Monit. Assess. 7427–7443. https://doi.org/10.1007/s10661-013-3110-x

de Oliveira, A.J.F.C., Watanabe Pinhata, J.M., 2008. Antimicrobial resistance and species composition of Enterococcus spp. isolated from waters and sands of marine recreational beaches in Southeastern Brazil. Water Res. 42, 2242–2250. https://doi.org/10.1016/j.watres.2007.12.002

Deasy, B.M., Rea, M.C., Fitzgerald, G.F., Cogan, T.M., Beresford, T.P., 2000. A rapid PCR based method to distinguish between Lactococcus and Enterococcus. Syst. Appl. Microbiol. 23, 510–522. https://doi.org/10.1016/S0723-2020(00)80025-9

Depardieu, F., Perichon, B., Courvalin, P., 2004. Detection of the van alphabet and identification of enterococci and staphylococci at the species level by multiplex PCR. J. Clin. Microbiol. 42, 5857–5860. https://doi.org/10.1128/JCM.42.12.5857-5860.2004

Dionisio, L.P.C., Borrego, J.J., 1995. Evaluation of media for the enumeration of faecal streptococci from natural water samples. J. Microbiol. Methods 23, 183–203. https://doi.org/10.1016/0167-7012(95)00014-C

Dutka-Malen, S., Evers, S., Courvalin, P., 1995a. Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. J. Clin. Microbiol. 33, 24–27.

Dutka-Malen, S., Evers, S., Courvalin, P., 1995b. Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. J. Clin. Microbiol. 33, 24–27. https://doi.org/10.1128/jcm.33.1.24-27.1995

Dutka-Malen, S., Molinas, C., Arthur, M., Courvalin, P., 1992. Sequence of the vanC gene of Enterococcus gallinarum BM4174 encoding a d-alanine:d-alanine ligase-related protein necessary for vancomycin resistance. Gene 112, 53–58. https://doi.org/10.1016/0378-1119(92)90302-6

Eaton, T.J., Gasson, M.J., 2001. Molecular Screening of Enterococcus Virulence Determinants and Potential for Genetic Exchange between Food and Medical Isolates. Appl. Environ. Microbiol. 67, 1628–1635. https://doi.org/10.1128/AEM.67.4.1628-1635.2001

Eckner, K.F., 1998. Comparison of membrane filtration and multiple-tube fermentation by the Colilert and Enterolert methods for detection of waterborne coliform bacteria, Escherichia coli, and Enterococci used in drinking and bathing water quality monitoring in southern Swed. Appl. Environ. Microbiol. 64, 3079–3083. https://doi.org/10.1128/aem.64.8.3079-3083.1998

Ekwanzala, M.D., Dewar, J.B., Kamika, I., Momba, M.N.B., 2020a. Comparative genomics of vancomycin-resistant Enterococcus spp. revealed common resistome determinants from hospital wastewater to aquatic environments. Sci. Total Environ. 719, 137275. https://doi.org/10.1016/j.scitotenv.2020.137275

Ekwanzala, M.D., Dewar, J.B., Momba, M.N.B., 2020b. Environmental resistome risks of wastewaters and aquatic environments deciphered by shotgun metagenomic assembly. Ecotoxicol. Environ. Saf. 197, 110612. https://doi.org/10.1016/j.ecoenv.2020.110612

Emery, C.L., Appleman, M.D., Siders, J.A., Davis, T.E., 2016. Rapid Devices and Instruments for the Identification of Anaerobic Bacteria. Man. Commer. Methods Clin. Microbiol. 56–86. https://doi.org/10.1002/9781119021872.ch4

EPA, 2012. Method 1611 : Enterococci in Water by TaqMan ® Quantitative Polymerase Chain Reaction (qPCR) Assay.

European Commission, 2017. A European One Health Action Plan against Antimicrobial Resistance (AMR).

Facklam, R.R., Collins, M.D., 1989. Identification of enterococcus species isolated from human infections by a conventional test scheme. J. Clin. Microbiol. 27, 731–734. https://doi.org/10.1128/jcm.27.4.731-734.1989

Ferguson, D.M., Griffith, J.F., McGee, C.D., Weisberg, S.B., 2010. Comparison of Enterococcus Species Selectivity using Enterolert and EPA Method 1600.

Ferguson, D.M., Griffith, J.F., McGee, C.D., Weisberg, S.B., Hagedorn, C., 2013. Comparison of enterococcus species diversity in marine water and wastewater using enterolert and EPA method 1600. J. Environ. Public Health 2013. https://doi.org/10.1155/2013/848049

Ferguson, D.M., Moore, D.F., Getrich, M.A., Zhowandai, M.H., 2005. Enumeration and speciation of enterococci found in marine and intertidal sediments and coastal water in southern California. J. Appl. Microbiol. 99, 598–608. https://doi.org/10.1111/j.1365-

2672.2005.02660.x

Fricker, E.J., Fricker, C.R., 1996. Use of defined substrate technology and a novel procedure for estimating the numbers of enterococci in water. J. Microbiol. Methods 27, 207–210. https://doi.org/10.1016/S0167-7012(96)00950-5

Goldstein, R.E., Micallef, S.A., Gibbs, S.G., Davis, J.A., He, X., George, A., Kleinfelter, L.M., Schreiber, N.A., Mukherjee, S., Sapkota, A., Joseph, S.W., Sapkota, A.R., 2012. Methicillin-Resistant Staphylococcus aureus (MRSA) Detected at Four U.S. Wastewater Treatment Plants. Environ. Health Perspect. 120, 1551–1558.

Goldstein, R.E., Micallef, S.A., Gibbs, S.G., George, A., Claye, E., Sapkota, A., Joseph, S.W., Sapkota, A.R., 2014. Detection of vancomycin-resistant enterococci (VRE) at four U.S. wastewater treatment plants that provide effluent for reuse. Sci. Total Environ. 466–467, 404–411. https://doi.org/10.1016/j.scitotenv.2013.07.039

Gouliouris, T., Raven, K.E., Ludden, C., Blane, B., Corander, J., Horner, C.S., Hernandez-Garcia, J., Wood, P., Hadjirin, N.F., Radakovic, M., Holmes, M.A., de Goffau, M., Brown, N.M., Parkhill, J., Peacock, S.J., 2018. Genomic surveillance of enterococcus faecium reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom. MBio 9, 1–15. https://doi.org/10.1128/mBio.01780-18

Gouliouris, T., Raven, K.E., Moradigaravand, D., Ludden, C., Coll, F., Blane, B., Naydenova, P., Horner, C., Brown, N.M., Corander, J., Limmathurotsakul, D., Parkhill, J., Peacock, S.J., 2019. Detection of vancomycin-resistant Enterococcus faecium hospital-adapted lineages in municipal wastewater treatment plants indicates widespread distribution and release into the environment. Genome Res. 29, 626–634. https://doi.org/10.1101/gr.232629.117

Guardabassi, L., Brønnum, P.T., Danø, R., Forslund, A., Dalsgaard, A., 2002. Dissemination of vancomycin-resistant enterococci harboring vanA through disposal of waste derived from industrial production of vancomycin. Microb. Drug Resist. 8, 401–406. https://doi.org/10.1089/10766290260469688

Guardabassi, L., Dalsgaard, A., 2004. Occurrence, Structure, and Mobility of Tn1546-Like Elements in Environmental Isolates of Vancomycin-Resistant Enterococci. Appl. Environ. Microbiol. 70, 984–990. https://doi.org/10.1128/AEM.70.2.984-990.2004

Guzman Prieto, A.M., van Schaik, W., Rogers, M.R.C., Coque, T.M., Baquero, F., Corander, J., Willems, R.J.L., 2016. Global emergence and dissemination of enterococci as nosocomial pathogens: Attack of the clones? Front. Microbiol. 7, 1–15. https://doi.org/10.3389/fmicb.2016.00788

Hancock, L.E., Murray, B.E., Sillanpää, J., 2014. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. Boston: Massachusetts Eye and Ear Infirmary.

Harwood, V.J., Brownell, M., Perusek, W., Whitlock, J.E., 2001. Vancomycin-Resistant Enterococcus spp. Isolated from Wastewater and Chicken Feces in the United States. Appl. Environ. Microbiol. 67, 4930–4933. https://doi.org/10.1128/AEM.67.10.4930-4933.2001

Harwood, V.J., Whitlock, J., Withington, V., 2000. Classification of Antibiotic Resistance Patterns of Indicator Bacteria by Discriminant Analysis: Use in Predicting the Source of

Fecal Contamination in Subtropical Waters. Appl. Environ. Microbiol. 66, 3698–3704.

Health Canada, 2020. Guidance on the use of Enterococci as an Indicator in Canadian Drinking Water Supplies.

Hegstad, K., Mikalsen, T., Coque, T.M., Werner, G., Sundsfjord, A., 2010. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clin. Microbiol. Infect. 16, 541–554. https://doi.org/10.1111/j.1469-0691.2010.03226.x

Hernando-amado, S., Coque, T.M., Baquero, F., Martínez, J.L., 2019. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat. Microbiol. 4. https://doi.org/10.1038/s41564-019-0503-9

Hidron, A.I., Edwards, J.R., Patel, J., Horan, T.C., Sievert, D.M., Pollock, D.A., Fridkin, S.K., 2008. Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Annual Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infect. Control Hosp. Epidemiol. 29, 996–1011. https://doi.org/10.1086/591861

Hollenbeck, B.L., Rice, L.B., 2012. Intrinsic and acquired resistance mechanisms in enterococcus. Virulence 3, 421–569. https://doi.org/10.4161/viru.21282

Homan, W.L., Tribe, D., Poznanski, S., Li, M., Hogg, G., Spalburg, E., Van Embden, J.D.A., Willems, R.J.L., 2002. Multilocus sequence typing scheme for Enterococcus faecium. J. Clin. Microbiol. 40, 1963–1971. https://doi.org/10.1128/JCM.40.6.1963-1971.2002

International Organization for Standardization, 2000. ISO 7899-2 2000 Water Quality – Detection and Enumeration of Intestinal Enterococci – Part 2: Membrane Filtration Method. Geneva, Switzerland.

Jackson, C.R., Fedorka-Cray, P.J., Barrett, J.B., 2004. Use of a genus- and species-specific multiplex PCR for identification of enterococci. J. Clin. Microbiol. 42, 3558–3565. https://doi.org/10.1128/JCM.42.8.3558-3565.2004

Jackson, C.R., Lombard, J.E., Dargatz, D.A., Fedorka-Cray, P.J., 2011. Prevalence, species distribution and antimicrobial resistance of enterococci isolated from US dairy cattle. Lett. Appl. Microbiol. 52, 41–48. https://doi.org/10.1111/j.1472-765X.2010.02964.x

Jahne, M.A., Rogers, S.W., Ramler, I.P., Holder, E., Hayes, G., 2015. Hierarchal clustering yields insight into multidrug-resistant bacteria isolated from a cattle feedlot wastewater treatment system. Environ. Monit. Assess. 187. https://doi.org/10.1007/s10661-014-4168-9

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M.N., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. J. Appl. Microbiol. 123, 570–581. https://doi.org/10.1111/jam.13468

Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: 2 approved]. Wellcome Open Res. 3, 1–20. https://doi.org/10.12688/wellcomeopenres.14826.1

Kariyama, R., Mitsuhata, R., Chow, J.W., Clewell, D.B., Kumon, H., 2000. Simple and reliable multiplex PCR assay for surveillance isolates of vancomycin-resistant enterococci. J. Clin. Microbiol. 38, 3092–3095. https://doi.org/10.1128/jcm.38.8.3092-3095.2000

Ke, D., Picard, F.J., Martineau, F., Ménard, C., Roy, P.H., Ouellette, M., Bergeron, M.G., 1999. Development of a PCR assay for rapid detection of enterococci. J. Clin. Microbiol. 37, 3497–3503. https://doi.org/10.1128/jcm.37.11.3497-3503.1999

Kinzelman, J., Ng, C., Jackson, E., Gradus, S., Bagley, R., 2003. Enterococci as indicators of Lake Michigan recreational water quality: Comparison of two methodologies and their impacts on public health regulatory events. Appl. Environ. Microbiol. 69, 92–96. https://doi.org/10.1128/AEM.69.1.92-96.2003

Kotzamanidis, C., Zdragas, A., Kourelis, A., Moraitou, E., Papa, A., Yiantzi, V., Pantelidou, C., Yiangou, M., 2009. Characterization of vanA-type enterococcus faecium isolates from urban and hospital wastewater and pigs. J. Appl. Microbiol. 107, 997–1005. https://doi.org/10.1111/j.1365-2672.2009.04274.x

Lanthier, M., Scott, A., Zhang, Y., Cloutier, M., Durie, D., Henderson, V.C., Wilkes, G., Lapen, D.R., Topp, E., 2011. Distribution of selected virulence genes and antibiotic resistance in Enterococcus species isolated from the South Nation River drainage basin, Ontario, Canada. J. Appl. Microbiol. 110, 407–421. https://doi.org/10.1111/j.1365-2672.2010.04893.x

Larsson, D.G.J., Flach, C.-F., 2021. Antibiotic resistance in the environment. Nat. Rev. Microbiol. 0123456789. https://doi.org/10.1038/s41579-021-00649-x

Lata, P., Ram, S., Shanker, R., 2016. Multiplex PCR based genotypic characterization of pathogenic vancomycin resistant Enterococcus faecalis recovered from an Indian river along a city landscape. Springerplus 5. https://doi.org/10.1186/s40064-016-2870-5

Laverde Gomez, J.A., Hendrickx, A.P.A., Willems, R.J., Top, J., Sava, I., Huebner, J., Witte, W., Werner, G., 2011. Intra- and interspecies genomic transfer of the Enterococcus faecalis pathogenicity Island. PLoS One 6. https://doi.org/10.1371/journal.pone.0016720

Leclercq, R., Oberlé, K., Galopin, S., Cattoir, V., Budzinski, H., Petit, F., 2013. Changes in enterococcal populations and related antibiotic resistance along a medical center-wastewater treatment plant-river Continuum. Appl. Environ. Microbiol. 79, 2428–2434. https://doi.org/10.1128/AEM.03586-12

Levin, M.A., Fischer, J.R., Cabelli, V.J., 1975. Membrane Filter Technique for Enumeration of Enterococci in Marine Waters. Appl. Microbiol. 30, 66–71. https://doi.org/10.1128/am.30.1.66-71.1975

Łuczkiewicz, A., Jankowska, K., Kurlenda, J., Olańczuk-Neyman, K., 2010. Identification and antimicrobial resistance of Enterococcus spp. isolated from surface water. Water Sci. Technol. 62, 466–473. https://doi.org/10.2166/wst.2010.909

Maheux, A.F., Picard, F.J., Boissinot, M., Huppé, V., Bissonnette, L., Bernier, J.L.T., Cantin, P., Huletsky, A., Bergeron, M.G., 2009. Analytical limits of three β-glucosidase-based commercial culture methods used in environmental microbiology, to detect enterococci. Water Sci. Technol. 60, 943–955. https://doi.org/10.2166/wst.2009.428

Mannu, L., Paba, A., Daga, E., Comunian, R., Zanetti, S., Duprè, I., Sechi, L.A., 2003. Comparison of the incidence of virulence determinants and antibiotic resistance between Enterococcus faecium strains of dairy, animal and clinical origin. Int. J. Food Microbiol. 88, 291–304. https://doi.org/10.1016/S0168-1605(03)00191-0

Manson, J.M., Hancock, L.E., Gilmore, M.S., 2010. Mechanism of chromosomal transfer of Enterococcus faecalis pathogenicity island, capsule, antimicrobial resistance, and other traits. Proc. Natl. Acad. Sci. U. S. A. 107, 12269–12274. https://doi.org/10.1073/pnas.1000139107

Marano, R.B.M., Fernandes, T., Manaia, C.M., Nunes, O., Morrison, D., Berendonk, T.U., Kreuzinger, N., Telson, T., Corno, G., Fatta-Kassinos, D., Merlin, C., Topp, E., Jurkevitch, E., Henn, L., Scott, A., Heß, S., Slipko, K., Laht, M., Kisand, V., Di Cesare, A., Karaolia, P., Michael, S.G., Petre, A.L., Rosal, R., Pruden, A., Riquelme, V., Agüera, A., Esteban, B., Luczkiewicz, A., Kalinowska, A., Leonard, A., Gaze, W.H., Adegoke, A.A., Stenstrom, T.A., Pollice, A., Salerno, C., Schwermer, C.U., Krzeminski, P., Guilloteau, H., Donner, E., Drigo, B., Libralato, G., Guida, M., Bürgmann, H., Beck, K., Garelick, H., Tacão, M., Henriques, I., Martínez-Alcalá, I., Guillén-Navarro, J.M., Popowska, M., Piotrowska, M., Quintela-Baluja, M., Bunce, J.T., Polo-López, M.I., Nahim–Granados, S., Pons, M.N., Milakovic, M., Udikovic-Kolic, N., Ory, J., Ousmane, T., Caballero, P., Oliver, A., Rodriguez-Mozaz, S., Balcazar, J.L., Jäger, T., Schwartz, T., Yang, Y., Zou, S., Lee, Y., Yoon, Y., Herzog, B., Mayrhofer, H., Prakash, O., Nimonkar, Y., Heath, E., Baraniak, A., Abreu-Silva, J., Choudhury, M., Munoz, L.P., Krizanovic, S., Brunetti, G., Maile-Moskowitz, A., Brown, C., Cytryn, E., 2020. A global multinational survey of cefotaxime-resistant coliforms in urban wastewater treatment plants. Environ. Int. 144, 106035. https://doi.org/10.1016/j.envint.2020.106035

Marathe, N.P., Pal, C., Gaikwad, S.S., Jonsson, V., Kristiansson, E., Larsson, D.G.J., 2017. Untreated urban waste contaminates Indian river sediments with resistance genes to last resort antibiotics. Water Res. 124, 388–397. https://doi.org/10.1016/j.watres.2017.07.060

Messer, J.W., Dufour, A.P., 1998. A rapid, specific membrane filtration procedure for enumeration of enterococci in recreational water. Appl. Environ. Microbiol. 64, 678–680. https://doi.org/10.1128/aem.64.2.678-680.1998

Miller, W.R., Munita, J.M., Arias, C.A., 2014. Mechanisms of antibiotic resistance in enterococci. Expert Rev. Anti. Infect. Ther. 12, 1221–1236. https://doi.org/10.1586/14787210.2014.956092

Moellering, R.C., 1992. Emergence of Enterococcus as a Significant Pathogen. Clin. Infect. Dis. 14, 1173–1176.

Molale, L.G., Bezuidenhout, C.C., 2016. Antibiotic resistance, efflux pump genes and virulence determinants in Enterococcus spp. from surface water systems. Environ. Sci. Pollut. Res. 23, 21501–21510. https://doi.org/10.1007/s11356-016-7369-7

Monticelli, L.S., Decembrini, F., Bergamasco, A., Caruso, G., 2019. Water quality assessment of transitional and coastal marine Sicilian waters (Italy): Ecological and epidemiological significance of multiple antimicrobial resistant Enterococcus spp. Estuar. Coast. Shelf Sci. 217, 173–184. https://doi.org/10.1016/j.ecss.2018.11.021

Murray, B.E., 1998. Diversity among multidrug-resistant enterococci. Emerg. Infect. Dis. 4, 37–47. https://doi.org/10.3201/eid0401.980106

Murray, B.E., 1990. The life and times of the enterococcus. Clin. Microbiol. Rev. 3, 46–65. https://doi.org/10.1128/CMR.3.1.46

Nam, S., Kim, M.J., Park, C., Park, J.G., Maeng, P.J., Lee, G.C., 2013. Detection and genotyping of vancomycin-resistant Enterococcus spp. by multiplex polymerase chain reaction in Korean aquatic environmental samples. Int. J. Hyg. Environ. Health 216, 421–427. https://doi.org/10.1016/j.ijheh.2012.12.004

Naser, S.M., Thompson, F.L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M., Swings, J., 2005. Application of multilocus sequence analysis (MLSA) for rapid identification of Enterococcus species based on rpoA and pheS genes. Microbiology 151, 2141–2150. https://doi.org/10.1099/mic.0.27840-0

Navarro, F., Courvalin, P., 1994. Analysis of genes encoding D-alanine-D-alanine ligase-related enzymes in Enterococcus casseliflavus and Enterococcus flavescens. Antimicrob. Agents Chemother. 38, 1788–1793. https://doi.org/10.1128/AAC.38.8.1788

Nemoy, L.L., Kotetishvili, M., Tigno, J., Keefer-Norris, A., Harris, A.D., Perencevich, E.N., Johnson, J.A., Torpey, D., Sulakvelidze, A., Morris, J.G., Stine, O.C., 2005. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing Escherichia coli isolates. J. Clin. Microbiol. 43, 1776–1781. https://doi.org/10.1128/JCM.43.4.1776-1781.2005

Nishiyama, M., Iguchi, A., Suzuki, Y., 2015. Identification of Enterococcus faecium and Enterococcus faecalis as vanC-type Vancomycin-Resistant Enterococci (VRE) from sewage and river water in the provincial city of Miyazaki, Japan. J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng. 50, 16–25. https://doi.org/10.1080/10934529.2015.964599

Nishiyama, M., Ogura, Y., Hayashi, T., Suzuki, Y., 2017. Antibiotic Resistance Profiling and Genotyping of Vancomycin-Resistant Enterococci Collected from an Urban River Basin in the Provincial City of Miyazaki, Japan. Water. https://doi.org/10.3390/w9020079

Novais, C., Coque, T.M., Ferreira, H., Sousa, C., Peixe, L., Cie, F., 2005. Environmental Contamination with Vancomycin-Resistant Enterococci from Hospital Sewage in Portugal. Appl. Environ. Microbiol. 71, 3364–3368. https://doi.org/10.1128/AEM.71.6.3364

Oravcova, V., Mihalcin, M., Zakova, J., Pospisilova, L., Masarikova, M., Literak, I., 2017. Vancomycin-resistant enterococci with vanA gene in treated municipal wastewater and their association with human hospital strains. Sci. Total Environ. 609, 633–643. https://doi.org/10.1016/j.scitotenv.2017.07.121

Pagel, J.E., Hardy, G.M., 1980. Comparison of selective media for the enumeration and identification of fecal streptococci from natural sources. Can. J. Microbiol. 26, 1320–1327. https://doi.org/10.1139/m80-219

Palmer, K.L., Gilmore, M.S., 2010. Multidrug-resistant enterococci lack CRISPR-cas. MBio 1, 1–10. https://doi.org/10.1128/mBio.00227-10

Paulsen, I.T., Banerjei, L., Hyers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., Fraser, C.M., 2003. Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. Science (80-.). 299, 2071–2074. https://doi.org/10.1126/science.1080613

Pöntinen, A.K., Top, J., Arredondo-Alonso, S., Tonkin-Hill, G., Freitas, A.R., Novais, C., Gladstone, R.A., Pesonen, M., Meneses, R., Pesonen, H., Lees, J.A., Jamrozy, D., Bentley, S.D., Lanza, V.F., Torres, C., Peixe, L., Coque, T.M., Parkhill, J., Schürch, A.C., Willems, R.J.L., Corander, J., 2021. Apparent nosocomial adaptation of Enterococcus faecalis predates the modern hospital era. Nat. Commun. 12. https://doi.org/10.1038/s41467-021-21749-5

Poyart, C., Quesnes, G., Trieu-Cuot, P., 2000. Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. J. Clin. Microbiol. 38, 415–418. https://doi.org/10.1128/JCM.38.1.415-418.2000

Prüss, A., 1998. Review of epidemiological studies on health effects from exposure to recreational water. Int. J. Epidemiol. 27, 1–9. https://doi.org/10.1093/ije/27.1.1

Rathnayake, I.U., Hargreaves, M., Huygens, F., 2012. Antibiotic resistance and virulence traits in clinical and environmental Enterococcus faecalis and Enterococcus faecium isolates. Syst. Appl. Microbiol. 35, 326–333. https://doi.org/10.1016/j.syapm.2012.05.004

Rathnayake, I.U., Hargreaves, M., Huygens, F., 2011. Genotyping of Enterococcus faecalis and Enterococcus faecium isolates by use of a set of eight single nucleotide polymorphisms. J. Clin. Microbiol. 49, 367–372. https://doi.org/10.1128/JCM.01120-10

Rice, E.W., Baird, A.D., 2017. Method 9230 Fecal Enterococcus/Streptococcus Groups, in: Standard Methods for the Examination of Water and Wastewater, 23rd Edition. pp. 3–6.

Ruiz-Garbajosa, P., Bonten, M.J.M., Robinson, D.A., Top, J., Nallapareddy, S.R., Torres, C., Coque, T.M., Cantón, R., Baquero, F., Murray, B.E., Del Campo, R., Willems, R.J.L., 2006. Multilocus sequence typing scheme for Enterococcus faecalis reveals hospital-adapted genetic complexes in a background of high rates of recombination. J. Clin. Microbiol. 44, 2220–2228. https://doi.org/10.1128/JCM.02596-05

Sadowy, E., Luczkiewicz, A., 2014. Drug-resistant and hospital-associated Enterococcus faecium from wastewater, riverine estuary and anthropogenically impacted marine catchment basin. BMC Microbiol. 14. https://doi.org/10.1186/1471-2180-14-66

Sanderson, H., Ortega-Polo, R., McDermott, K., Zaheer, R., Brown, R.S., Majury, A., McAllister, T., Liss, S.N., 2019. Comparison of biochemical and genotypic speciation methods for vancomycin-resistant enterococci isolated from urban wastewater treatment plants. J. Microbiol. Methods 161, 102–110. https://doi.org/10.1016/j.mimet.2019.04.019

Santiago-Rodriguez, T.M., Rivera, J.I., Coradin, M., Toranzos, G.A., 2013. Antibiotic-resistance and virulence genes in Enterococcus isolated from tropical recreational waters. J. Water Health 11, 387–396. https://doi.org/10.2166/wh.2013.005

Sapkota, A.R., Curriero, F.C., Gibson, K.E., Schwab, K.J., 2007. Antibiotic-resistant enterococci and fecal indicators in surface water and groundwater impacted by a concentrated swine feeding operation. Environ. Health Perspect. 115, 1040–1045. https://doi.org/10.1289/ehp.9770

Savin, M., Bierbaum, G., Hammerl, J.A., Heinemann, C., Parcina, M., Sib, E., Voigt, A., Kreyenschmidt, J., 2020. Antibiotic-resistant bacteria and antimicrobial residues in wastewater and process water from German pig slaughterhouses and their receiving municipal wastewater treatment plants. Sci. Total Environ. 727, 138788. https://doi.org/10.1016/j.scitotenv.2020.138788

Schoen, M.E., Soller, J.A., Ashbolt, N.J., 2011. Evaluating the importance of faecal sources in human-impacted waters. Water Res. 45, 2670–2680. https://doi.org/10.1016/j.watres.2011.02.025

Shankar, N., Baghdayan, A.S., Gilmore, M.S., 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant Enterococcus faecalis. Nature 417, 746–750. https://doi.org/10.1038/nature00802

Sievert, D.M., Ricks, P., Edwards, J.R., Schneider, A., Patel, J., Srinivasan, A., Kallen, A., Limbago, B., Fridkin, S., 2013. Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009–2010. Infect. Control Hosp. Epidemiol. 34, 1–14. https://doi.org/10.1086/668770

Silva, N., Igrejas, G., Gonçalves, A., Poeta, P., 2012. Commensal gut bacteria: Distribution of Enterococcus species and prevalence of Escherichia coli phylogenetic groups in animals and humans in Portugal. Ann. Microbiol. 62, 449–459. https://doi.org/10.1007/s13213-011-0308-4

Sinclair, R.G., Rose, J.B., Hashsham, S.A., Gerba, C.P., Haase, C.N., 2012. Criteria for selection of surrogates used to study the sate and control of pathogens in the environment. Appl. Environ. Microbiol. 78, 1969–1977. https://doi.org/10.1128/AEM.06582-11

Singhal, N., Kumar, M., Kanaujia, P.K., Virdi, J.S., 2015. MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. Front. Microbiol. 6, 1–16. https://doi.org/10.3389/fmicb.2015.00791

Stine, O.C., Johnson, J.A., Keefer-Norris, A., Perry, K.L., Tigno, J., Qaiyumi, S., Stine, M.S., Morris, J.G., 2007. Widespread distribution of tetracycline resistance genes in a confined animal feeding facility. Int. J. Antimicrob. Agents 29, 348–352. https://doi.org/10.1016/j.ijantimicag.2006.11.015

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E.M., Houchens, C.R., Grayson, M.L., Hansen, P., Singh, N., Theuretzbacher, U., Magrini, N., Aboderin, A.O., Al-Abri, S.S., Awang Jalil, N., Benzonana, N., Bhattacharya, S., Brink, A.J., Burkert, F.R., Cars, O., Cornaglia, G., Dyar, O.J., Friedrich, A.W., Gales, A.C., Gandra, S., Giske, C.G., Goff, D.A., Goossens, H., Gottlieb, T., Guzman Blanco, M., Hryniewicz, W., Kattula, D., Jinks, T., Kanj, S.S., Kerr, L., Kieny, M.P., Kim, Y.S.,

Kozlov, R.S., Labarca, J., Laxminarayan, R., Leder, K., Leibovici, L., Levy-Hara, G., Littman, J., Malhotra-Kumar, S., Manchanda, V., Moja, L., Ndoye, B., Pan, A., Paterson, D.L., Paul, M., Qiu, H., Ramon-Pardo, P., Rodríguez-Baño, J., Sanguinetti, M., Sengupta, S., Sharland, M., Si-Mehand, M., Silver, L.L., Song, W., Steinbakk, M., Thomsen, J., Thwaites, G.E., van der Meer, J.W., Van Kinh, N., Vega, S., Villegas, M.V., Wechsler-Fördös, A., Wertheim, H.F.L., Wesangula, E., Woodford, N., Yilmaz, F.O., Zorzet, A., 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. 18, 318–327. https://doi.org/10.1016/S1473-3099(17)30753-3

Taučer-Kapteijn, M., Hoogenboezem, W., Heiliegers, L., de Bolster, D., Medema, G., 2016. Screening municipal wastewater effluent and surface water used for drinking water production for the presence of ampicillin and vancomycin resistant enterococci. Int. J. Hyg. Environ. Health 219, 437–442. https://doi.org/10.1016/j.ijheh.2016.04.007

Teixeira, L.M., Siqueira, G., Shewmaker, P.L., Facklam, R.R., 2015. Enterococcus, in: Manual of Clinical Microbiology. https://doi.org/10.1128/9781555817381.ch23

Tejedor Junco, M.A.T., Martin, M.G., Pita Toledo, M.A.L., Gómez, P.L., Barrasa, J.L.M., 2001. Identification and antibiotic resistance of faecal enterococci isolated from water samples. Int. J. Hyg. Environ. Health 203, 363–368. https://doi.org/10.1078/1438-4639-00043

Tiwari, A., Hokajärvi, A.M., Santo Domingo, J.W., Kauppinen, A., Elk, M., Ryu, H., Jayaprakash, B., Pitkänen, T., Domingo, J.W.S., 2018. Categorical performance characteristics of method ISO 7899-2 and indicator value of intestinal enterococci for bathing water quality monitoring. J. Water Health 16, 711–723. https://doi.org/10.2166/wh.2018.293

Top, J., Willems, R., Van Der Velden, S., Asbroek, M., Bonten, M., 2008. Emergence of clonal complex 17 Enterococcus faecium in the Netherlands. J. Clin. Microbiol. 46, 214–219. https://doi.org/10.1128/JCM.01351-07

U.S. Environmental Protection Agency, 2009. Method 1600: Enterococci in Water by Membrane Filtration Using membrane-Enterococcus Indoxyl- B -D-Glucoside Agar (mEI).

U.S. Environmental Protection Agency, 2006. Method 1106.1 : Enterococci in Water by Membrane Filtration Using membrane- Enterococcus-Esculin Iron Agar.

Valente, M.S., Pedro, P., Alonso, M.C., Borrego, J.J., Dionísio, L., 2010. Are the defined substrate-based methods adequate to determine the microbiological quality of natural recreational waters? J. Water Health 8, 11–19. https://doi.org/10.2166/wh.2009.220

van Hal, S.J., Willems, R.J.L., Gouliouris, T., Ballard, S.A., Coque, T.M., Hammerum, A.M., Hegstad, K., Westh, H.T., Howden, B.P., Malhotra-Kumar, S., Werner, G., Yanagihara, K., Earl, A.M., Raven, K.E., Corander, J., Bowden, R., Pinholt, M., Loens, K., Xavier, B.B., Matheeussen, V., Goossens, H., 2021. The global dissemination of hospital clones of Enterococcus faecium. Genome Med. 13, 1–12. https://doi.org/10.1186/s13073-021-00868-0

Vankerckhoven, V., Van Autgaerden, T., Vael, C., Lammens, C., Chapelle, S., Rossi, R., Jabes, D., Goossens, H., 2004. Development of a multiplex PCR for the detection of asaI, gelE,

cylA, esp, and hyl genes in enterococci and survey for virulence determinants among european hospital isolates of Enterococcus faecium. J. Clin. Microbiol. 42, 4473–4479. https://doi.org/10.1128/JCM.42.10.4473-4479.2004

Vilanova, X., Blanch, A.R., 2006. Distribution and persistence of fecal bacterial populations in liquid and dewatered sludge from a biological treatment plant. J. Gen. Appl. Microbiol. 51, 361–368. https://doi.org/10.2323/jgam.51.361

Wade, T.J., Calderon, R.L., Sams, E., Beach, M., Brenner, K.P., Williams, A.H., Dufour, A.P., 2006. Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness. Environ. Health Perspect. 114, 24–28. https://doi.org/10.1289/ehp.8273

Wade, T.J., Pai, N., Eisenberg, J.N.S., Colford, J.M., 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ. Health Perspect. 111, 1102–1109. https://doi.org/10.1289/ehp.6241

Weiner-Lastinger, L.M., Abner, S., Edwards, J.R., Kallen, A.J., Karlsson, M., Magill, S.S., Pollock, D., See, I., Soe, M.M., Walters, M.S., Dudeck, M.A., 2020. Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network, 2015–2017. Infect. Control Hosp. Epidemiol. 41, 1–18. https://doi.org/10.1017/ice.2019.296

Weiner, L.M., Webb, A.K., Limbago, B., Dudeck, M.A., Patel, J., Kallen, A.J., Edwards, J.R., Sievert, D.M., 2016. Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. Infect. Control Hosp. Epidemiol. 37, 1288–1301. https://doi.org/10.1017/ice.2016.174

WHO, 2021. Global Tricycle Surveillance: ESBL E. coli, WHO integrated global surveillance on ESBL-producing E. coli using a “One Health” approach: Implementation and opportunities.

Woodford, N., Johnson, A.P., Morrison, D., Speller, D.C.E., 1995. Current perspectives on glycopeptide resistance. Clin. Microbiol. Rev. 8, 585–615. https://doi.org/10.1128/cmr.8.4.585

Zaheer, R., Cook, S.R., Barbieri, R., Goji, N., Cameron, A., Petkau, A., Polo, R.O., Tymensen, L., Stamm, C., Song, J., Hannon, S., Jones, T., Church, D., Booker, C.W., Amoako, K., Van Domselaar, G., Read, R.R., McAllister, T.A., 2020. Surveillance of Enterococcus spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. Sci. Rep. 10, 1–16. https://doi.org/10.1038/s41598-020-61002-5

Zdragas, A., Partheniou, P., Kotzamanidis, C., Psoni, L., Koutita, O., Moraitou, E., Tzanetakis, N., Yiangou, M., 2008. Molecular characterization of low-level vancomycin-resistant enterococci found in coastal water of Thermaikos Gulf, Northern Greece. Water Res. 42, 1274–1280. https://doi.org/10.1016/j.watres.2007.09.022

Zhang, S., Huang, J., Zhao, Z., Cao, Y., Li, B., 2020. Hospital Wastewater as a Reservoir for Antibiotic Resistance Genes: A Meta-Analysis. Front. Public Heal. 8, 1–12. https://doi.org/10.3389/fpubh.2020.574968

SUPPLEMENTAL MATERIALS FOR CHAPTER 5

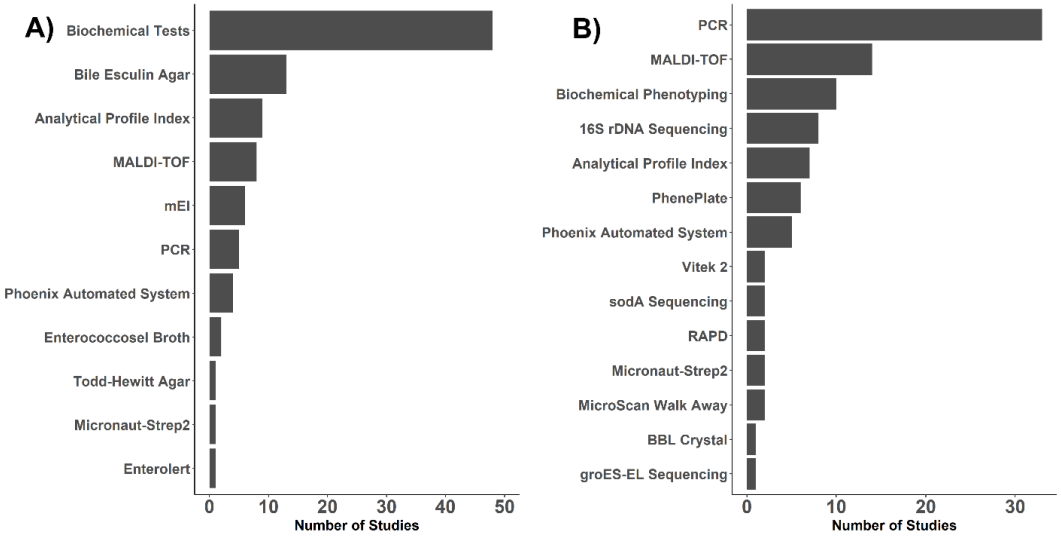


Figure 5-S1: A) Distribution of methods used for confirming enterococci to the genus (75/105 articles performed genus confirmation). B) Distribution of methods used for confirming enterococci to the species (86/105 articles performed species confirmation).

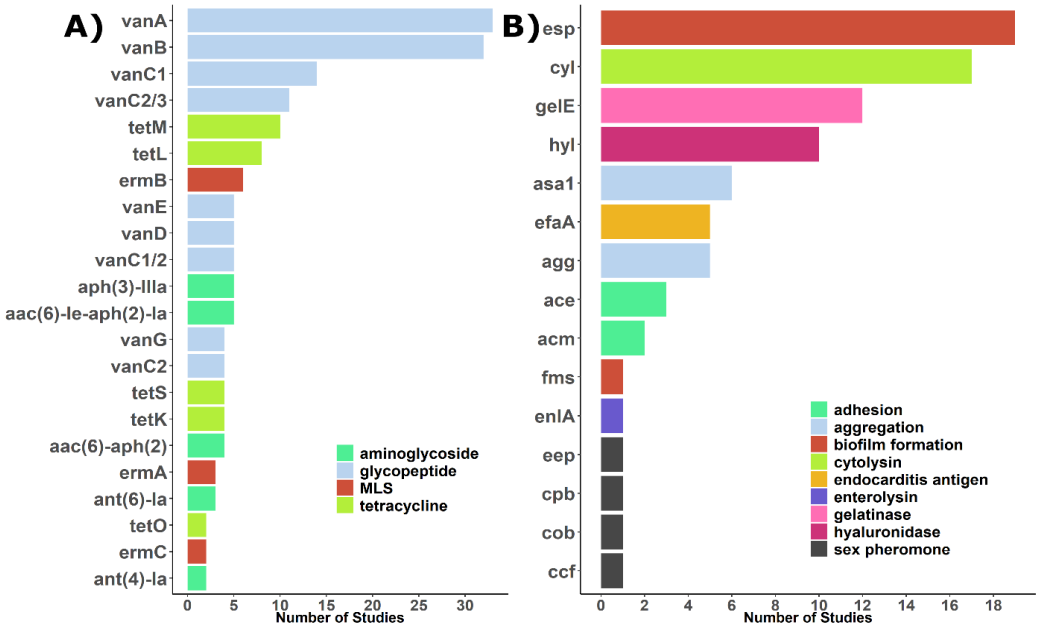


Figure 5-S2: Distribution of PCR targets used for genotyping Enterococcus spp. across all *articles.*

Table 5-S1: Tiered search term approach for systematic review. All searches were done in Web of Science.

Search Tiers	Search Terms	Number of Articles
Water Matrices	“wastewater” OR “reclaimed water” OR “recycled water” OR “water reuse” OR “non-potable reuse” OR “greywater” OR “hospital wastewater” OR “surface water” OR “sewage” OR “wastewater treatment plant” OR “filtration” OR “direct potable reuse” OR “indirect potable reuse” OR “river” OR “watershed” OR "lake" OR "pond" OR "recreational water" OR "influent" OR "effluent" OR "aquatic" OR "water quality" OR "de facto reuse" OR “water*”	3,828,792
Antibiotic Resistance	& “antibiotic resistan*” OR “antimicrobial susceptibility” OR “antimicrobial resistan*” OR “drug resistan*” OR “multi-drug resistan*” OR “resistome” OR "ARG" OR "antibiotic resistan* gene"	15,043
Culturing	& “culture” OR “isolat*” OR "membrane filtrat*" OR "spread plating" OR "IDEXX" OR "Enterolert"	5,439
Enterococci	& “enterococc*”	479

Table 5-S2: Data extraction parameters used for

Parameter	Description/Example
Reference	APA format reference
Country	Where the samples were taken
Continent	
General Environment	Surface water, wastewater, recycled water
Specific Environment	influent, effluent, river, saltwater beach
Sample Volume	What volume of water represents each sample
Concentration Technique	membrane filtration, enrichment
Standardized method referenced	ISO 7899-2, EPA Method 1600
Selective media	mEI, Slanetz-Bartley
Incubation time and temp	Recorded for any agar used in all workflows
Confirmation media	Brain Heart Infusion Agar, Iron Esculin Agar
ATCC Controls	Which reference cultures were used throughout workflows
Antibiotic used in selective media	What antibiotic was used in the initial agar and at what concentration
Antibiotic susceptibility testing technique	Disc diffusion, agar dilution, strip tests, automated platform
Antibiotic used in susceptibility testing	The antibiotic and the concentration used
Susceptibility testing referenced	CLSI, EUCAST, NARMS
DNA extraction technique	For isolates characterized molecularly
ARG screened for with PCR	In DNA extracted isolates, which ARGs were searched for
Speciation PCR target	The gene target for molecular speciation of isolates
Total enterococci isolates	Where applicable, the number of isolates in the article library
Number of resistant isolates	The number of isolates marked “resistant” to specified antibiotics
Species distribution of enterococci	Where applicable, the summary statistics of species distributions

Table 5-S3: The 10 most common Enterococcus species screened for antibiotic resistance across all published articles.

Species	Known Origins [1]	Opportunistic Human Pathogen	Confirmed Vancomycin Resistance	Reference
E. faecalis	Human, Animal, Plant, Insect	Yes	Yes	[2–4]
E. faecium	Human, Animal, Plant, Insect	Yes	Yes	[4–7]
E. gallinarum	Human, Animal, Insect	Yes	Rare*	[2,4,8]
E. casseliflavus	Human, Animal, Plant, Soil	Yes	Rare*	[3,4,8]
E. hirae	Animal, Plant	Yes	Rare	[4]
E. durans	Human, Animal, Insect	Yes	Rare	[3]
E. mundtii	Soil, Plant	Yes	Rare	[3]
E. avium	Human, Animal	Yes	No	
E. raffinosus	Human	Yes	No	
E. pseudoavium	Human	Yes	No	

*Intermediate resistance is common

References

1. Byappanahalli, M. N., Nevers, M. B., Korajkic, A., Staley, Z. R. & Harwood, V. J. Enterococci in the Environment. *Microbiol. Mol. Biol. Rev.* **76**, 685–706 (2012).

2. Goldstein, R. E. *et al.* Detection of vancomycin-resistant enterococci (VRE) at four U.S. wastewater treatment plants that provide effluent for reuse. *Sci. Total Environ.* **466**–**467**, 404–411 (2014).

3. Guardabassi, L. & Dalsgaard, A. Occurrence, Structure, and Mobility of Tn1546-Like Elements in Environmental Isolates of Vancomycin-Resistant Enterococci. *Appl. Environ. Microbiol.* **70**, 984–990 (2004).

4. Blanch, A. R. *et al.* Comparison of enterococcal populations related to urban and hospital wastewater in various climatic and geographic European regions. *J. Appl. Microbiol.* **94**, 994–1002 (2003).

5. Poole, T. L. *et al.* Vancomycin-resistant Enterococcus faecium strains isolated from community wastewater from a semiclosed agri-food system in Texas. *Antimicrob. Agents Chemother.* **49**, 4382–4385 (2005).

6. Araújo, C. *et al.* Vancomycin-resistant enterococci from Portuguese wastewater treatment plants. *J. Basic Microbiol.* **50**, 605–609 (2010).

7. Borhani, K., Ahmadi, A., Rahimi, F. & Pourshafie, M. R. Determination of Vancomycin Resistant Enterococcus faecium Diversity in Tehran Sewage Using Plasmid Profile , Biochemical Fingerprinting and Antibiotic Resistance. *Jundishapur J. Microbiol.* **7**, 1–5 (2014).

8. Shaghaghi, B. *et al.* Phenotypic diversity of multiple antibiotic resistant enterococci with emphasis on Enterococcus gallinarum carrying vanA and vanB genes. *Water. Air. Soil Pollut.* **186**, 255–261 (2007).

CHAPTER 6: BENCHMARKING QUANTITATIVE METAGENOMICS FOR WASTEWATER-BASED SURVEILLANCE OF ANTIBIOTIC RESISTANCE

Benjamin C. Davis[1], Gabriel Maldonado Rivera[1], Peter Vikesland[1], Amy Pruden[1]

1 Via Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia

ABSTRACT

Metagenomics is gaining momentum as a comprehensive, non-targeted tool for the detection of antibiotic resistance genes (ARGs) in wastewater-based surveillance (WBS) systems. However, there is a need for quantitative metagenomic (qMeta) datatypes to support risk assessment and to improve cross-study comparability. Here we spiked in internal DNA reference standards (meta sequins) into replicate wastewater samples at decreasing concentration ratios to assess the variability of gene detection at low abundances. Meta sequin ladders exhibited strong linearity at input concentrations as low as 2×10^{-5} w/w% ($R^2 > 0.98$) and displayed minimal GC bias. We calculated the limit of quantification (LoQ) and limit of detection (LoD) for the experiment to be 2.7×10^4 and 54 gene copies per µL DNA extract, respectively, and applied these cutoffs to the high-throughput quantification of ARGs. In influent, activated sludge, and secondary effluent samples, 75.8%, 86.2%, and 87.3% of detected ARGs fell below the LoQ, even at ~100 Gb per sample. Comparing qMeta to qPCR concentrations for 16S rRNA and *sul*1 genes, we found no significant differences between techniques. The result of this study will help to improve the comparability of metagenomics for WBS of antibiotic resistance and help support the statistical evaluation of ARG detection at extremely low abundances.

INTRODUCTION

Quantitative molecular approaches are needed to support momentum towards establishing and expanding wastewater-based surveillance (WBS) and risk assessment of antibiotic resistance associated with wastewater ecosystems and associated aquatic environments. Quantitative polymerase chain reaction (qPCR) has been the gold standard for highly sensitive quantification of antibiotic resistance genes (ARGs) in environmental matrices for the past two decades and has been widely used to study ARG dynamics in surface waters (Davis et al., 2020; Pruden et al., 2012), recycled water (Garner et al., 2019), and wastewater systems (Mao et al., 2015; McConnell et al., 2018). qPCR is an attractive means to reliably quantify specific ARGs of interest carried across a microbial community with high sensitivity. Advantages relative to conventional culture-based investigations have been previously documented (Keenum et al., 2022). Realistically, however, qPCR is only capable of targeting a handful of ARGs at a time. Given that there are currently thousands of known functionally-verified ARGs in public databases (Alcock et al., 2020), each with their own distinct relevance to

human and ecological health, higher-throughput techniques are needed to support comprehensive surveillance efforts.

Recent advancements in microfluidic-based high-throughput qPCR have allowed the simultaneous detection of hundreds of ARGs, but currently lack quality control and validation metrics (e.g., standard curves and melt curves). This results in uncertainty in individual target detection and quantification (Waseem et al., 2019). Additionally, *a priori* identification of gene targets requires the design and validation of suitable primers, significantly hindering the ability to detect emerging ARG variants in real time. Metagenomic sequencing represents a distinct advantage in this regard, as it provides the high-throughput capabilities necessary for environmental monitoring whilst circumventing *a priori* identification of targets (Garner et al., 2021a). Several studies have demonstrated the effectiveness of metagenomics for ARG monitoring in wastewater (Hendriksen et al., 2019; Majeed et al., 2021; Riquelme et al., 2021) and there is currently momentum in the proposition of the use of metagenomics for comprehensive surveillance of ARGs in wastewater treatment plants (WWTPs) at a global scale (Aarestrup and Woolhouse, 2020; Pruden et al., 2021). However, metagenomic data have been limited to compositional assessments of ARGs (i.e., relative abundances normalized to a denominator), where absolute concentrations (i.e., target normalized per sample volume basis) are needed for quantitative microbial risk assessments (Garner et al., 2021b; Haas, 2020). Additionally, due to the near random sequencing of all genomic DNA across complex microbiomes, metagenomic detection of individual ARGs is rife with uncertainty, and statistical thresholds are needed for defining the presence or absence of detected ARGs as they enter and traverse WWTPs.

To achieve absolute quantification of gene targets, quantitative metagenomics (qMeta) requires the addition of internal nucleic acid reference standards to serve as a "ground truth" in experimental samples (Hardwick et al., 2017). Recent advances in process controls for next-generation sequencing experiments have led to the development of synthetic DNA reference standards, termed "sequins" (sequencing spike-ins), that represent the range of features and complexity of natural microbial communities while sharing no homology to reference sequences (Hardwick et al., 2018). Sequins designed for metagenomics (meta sequins) are mixtures of 86 unique DNA oligonucleotides of varying lengths (987-9120 bp) and GC content (24-71%) that are present at 16 discrete input proportions, forming a ladder with a 3.2×10^4 fold range (Hardwick et al., 2018). Meta sequins were originally benchmarked for inter-sample normalizations as well as for measuring fold changes between microbial communities, but to date, have not been used for absolute quantification of ARGs in wastewater environments. Previous studies have utilized exogenous whole genomes for ARG quantification (Crossette et al., 2021a), but genetic homology between naturally-derived spike-ins and environmental DNA cross-detection add uncertainty to quantification. To overcome this bias, synthetic DNA oligonucleotides with embedded xenobiotic insertions (consecutive stop codons) were recently benchmarked (Li et al., 2021), but small insertion sizes (103-430 bp) significantly reduce the probability of detecting references at extremely low abundance.

The overall goal of this study was to propose and evaluate a qMeta protocol for the monitoring of WWTP associated ARGs. Specifically, the objectives were to (i) benchmark the efficacy of sequin standards for quantifying ARGs and (ii) explore the technical limitations of Illumina

sequencing for detecting low abundance ARGs by defining the limit of quantification (LoQ) and detection (LoD). Meta sequins were spiked at logarithmically decreasing proportions (w/w%; mass of meta sequins/mass of total library x 100) into replicate wastewater samples representing three typical wastewater matrices: influent (INF), activated sludge (AS), and secondary effluent (SE). These matrices represent distinct levels of DNA complexity and expected ARG concentrations. Sequins were spiked into samples after DNA extraction to isolate the effects of library preparation and sequencing on the fidelity of the ladders and individual sequins at varying concentrations. The behavior of sequins across sample matrices, input concentrations, and individual sequin compositions were then assessed. To determine the overall variability of meta sequin detection at decreasing abundances, sequencing yields were calculated and used for quantifying ARGs. The result of this study will help to improve the quality, value, and overall comparability of environmental metagenomic data produced for ARG monitoring and other purposes.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, and Cleanup

Wastewater grab samples were collected in November 2021 from a local 5 MGD conventional wastewater treatment plant. INF, AS, and SE samples were taken in 2L autoclaved polypropylene bottles and transported back to Virginia Tech on ice. Briefly, samples were vacuum filtered onto 0.45 µm mixed cellulose-ester filters in replicates of 10 (30 total filters). 50, 10, and 500 mL was filtered for INF, AS, and SE samples, respectively. 100 mL of deionized water was vacuumed through an additional filter to serve as the filter blank and negative control for the experiment. Filters were placed in 2 mL centrifuge tubes, fixed with 1 mL 100% ethanol, and stored at -20°C for DNA extraction the following day. During DNA extraction, ethanol fixed filters were torn into small (~1 cm^2) pieces with flame sterilized tweezers and placed into lysing matrix E tubes of the FastDNA Spin Kit for Soil (MPBio). Before the addition of phosphate and lysing buffer, 75 µL of the ZymoBIOMICS Spike-In Control II (Zymo Research, CA) was added to each sample. Samples were then homogenized via bead-beating (40 s at 6 m/s) with the FastPrep-24 5G (MPBio), further extracted according to manufacturer's instructions, and eluted in 100 µL of elution buffer. DNA extracts were first quantified using a dsDNA high sensitivity assay kit on a Qubit Fluorometer (Invitrogen), and 260/280 ratios were checked on a NanoPhotometer® Pearl (Implen). Each sample was then cleaned using a ZymoBIOMICS DNA Clean & Concentrator kit, eluted with 50 µL elution buffer, and re-quantified and quality checked.

Quantitative Polymerase Chain Reaction of Gene Targets

qPCR was performed on all cleaned DNA extracts to quantify the abundance of total bacteria (16S rRNA genes) (Suzuki and Taylor, 2000) and *sul*1 (Pei et al., 2006). Initially, subsampled DNA from the first replicates of each wastewater matrix were aliquoted to identify dilution factors that would minimize the effects of PCR inhibitors. 16S rRNA was quantified at 1:1, 1:50, 1:100, 1:250, and 1:500 dilutions where 1:100 dilutions were shown to maximize output concentrations for all three wastewater matrices (Figure 6-S1). The filter blank was below the limit of quantification for 16S rRNA in all analytical triplicates and was what not processed further. Using diluted aliquots, each qPCR target was then quantified in analytical triplicate with

deionized water as the no template control. Standard curves were generated from double-stranded gBlock gene fragments (IDT), resuspended according to manufacturer recommendations in molecular grade water, and quantified via Qubit. The minimum accepted qPCR standard curve efficiencies and R^2 va;ues were 80% and 0.980, respectively. The limit of quantification for each assay was set to the lowest standard that amplified in at least duplicate.

Internal Reference Standard Spiking and Sequencing

Sequins (https://www.sequinstandards.com/metagenome/) were received from the Garvan Institute of Medical Research (Sydney 2010 NSW, Australia) as lyophilized nucleic acids. Meta sequin "Mixture A" was resuspended according to manufacturer's instructions to a concentration of 2 ng/µL using molecular grade water and quantified using a Qubit Fluorometer. Mixture A contains 86 individual sequins at 16 discrete input proportions with at least 5 unique sequins at each proportion level, forming a reference ladder. The details for these meta sequins can be found in Table 6-S2. To generate a logarithmically decreasing spike-in gradient, sample DNA extracts were first normalized to 1000 ng in clean 500 µL microcentrifuge tubes. Meta sequins were serially diluted 9 times, starting at 2 ng/µL, and 10 µL of each dilution were added to corresponding replicate samples. This achieved a 2%, 0.2%, 0.02%, etc. w/w% (meta sequin mass to sample mass) of sequins per DNA extract for sequencing. Manufacturer instructions recommend adding sequins at a 2% w/w ratio and was defined here as a dilution factor of 1. Samples shipped overnight to the Duke Center for Genomic and Computational Biology on dry ice for library preparation and sequencing. All 30 samples were prepped using a KAPA HyperPrep PCR-free workflow targeting 500 bp insert sizes. Libraries were then sequenced on a single NovaSeq 6000 S4 flowcell with 150 bp chemistry targeting ~100 Gb per sample.

Bioinformatics and Statistics

Initially, R1 and R2 files from each of the 4 flow cell lanes were concatenated, respectively. Adapters were then removed and reads were quality filtered using fastp with default parameters (Chen et al., 2018). To align reads to the sequin reference standards, the Anaquin software, a dedicated package designed to analyze sequin reference standards, was run using the "meta" option (Wong et al., 2017). Calibration settings for each Anaquin command were kept at 0.01 as calibration files were not used in downstream analysis. For ARG annotation, paired reads were merged via their overlapping regions using Vsearch for instances where paired read lengths were greater than the insert size (Rognes et al., 2016). Merged clean reads were queried against the Comprehensive Antibiotic Resistance Database (CARD, v. 3.0.3, protein homolog model) using DIAMOND blastx (max-target-seqs = 1, aa length ≥ 25, bitscore ≥ 75, identity ≥ 80%). Differences in sample group means were determined with paired t-tests ($\alpha < 0.05$)

Modeling qMeta Approach

The mathematical relationships between spiked internal reference standard concentrations and experimentally-detected reads (equation 1) were derived by (Li et al., 2021) and are synthesized and repurposed below for use in the study:

$$\frac{C_{sequin-i}}{C_{Total}} \; x \; Y_{seq-i} = \frac{n_{sequin-i}}{n_{Total}} \qquad (1)$$

Where ($\frac{C_{sequin-i}}{C_{Total}}$) is the mass concentration ratio of sequin-*i* ($C_{sequin\text{-}i}$, ng/µL) to the total concentration of DNA in the sample (C_{Total}, ng/µL) and ($\frac{n_{sequin-i}}{n_{Total}}$) is the ratio of experimentally-

detected reads per sequin-*i* ($n_{sequin\text{-}i,}$) to the size of the total metagenomic library (n_{Total}, reads). The empirical parameter, $Y_{seq\text{-}i}$, represents the sequencing yield for sequin-*i*. An overall sequencing yield, Y_{seq}, can then be calculated as the sum of $Y_{seq\text{-}i}$, with a theoretical value of 1 and used to calculate the concentrations of target gene sequences (C_{Target}, ng/µL) in individual samples (equation 2). Concentrations can then be converted to gene copies (gc) per volume of DNA extract ($C_{Target\ gc/uL}$) (equation 3):

$$\frac{C_{Target}}{C_{Total}}\ x\ Y_{seq} = \frac{n_{Target}}{n_{Total}} \quad (2)$$

$$C_{Target\ gc/uL} = \frac{C_{Target}\ x\ N_A}{L_{Target}\ x\ 10^9\ x\ 650} \quad (3)$$

Where N_A is Avogadro's Number ($6.02x10^{23}$ copies/mol), L_{Target} is the length of the target gene (bp), 10^9 is the conversion from nanograms to grams, and 650 is the average molecular weight of a bp (g/mol). For final conversion to gc per volume of extract, $C_{Target\ gc/uL}$ values are multiplied by the total volume of DNA extracted ($V_{Extract}$) and divided by the volume of sample filtered (V_{sample}) (equation 4):

$$C_{Target\ gc/mL} = C_{Target\ gc/uL}\ x\ \frac{V_{Extract}}{V_{Sample}} \quad (4)$$

RESULTS

Behavior of Meta Sequins Across Spike-In Concentrations

After read quality filtering and trimming, an average of 6.3×10^8 reads (94 Gb) were generated per sample, resulting in a total library size of 2.8 Tb (Table 6-S1). Total per sample reads derived from meta sequins ranged from a single read to 1.1×10^7 (Table 6-S1). Strong linearity (Pearson, $R^2 \geq 0.980$, p<1e-16) between expected input proportions and experimental read counts was observed across dilution factors, regardless of the original sample matrix. This linearity was maintained with inputs as low as 2×10^{-5} w/w% (Figure 6-1A), suggesting that sequins are highly stable and reliably detected at low input abundance. All 86 sequins were detected at the first two dilutions across the three sample matrices, and at least a single sequin was detected in 27/30 samples (Figure 6-S2, S3). At the lowest dilution factors, we generally found that longer sequins with average GC contents were detected the most reliably. Additionally, we found strong overall linearity ($R^2 = 0.983$) between spiked concentration ratios ($\frac{C_{sequin-i}}{C_{Total}}$) and sequenced read ratios ($\frac{n_{sequin-i}}{n_{Total}}$) across a large dynamic range (1×10^{-9} – 1×10^{-2} ng/ng) (Figure 6-S4), validating the mathematical relationships presented by (Li et al., 2021).

At inputs greater than 2×10^{-4} w/w%, we found individual sequin read counts at each input proportion and dilution factor between INF, AS, and SE samples to be statistically indistinguishable (paired t-test, p < 1e-6), indicating that the inherent nucleic acid complexity of the DNA extracts did not influence detection. This observation allowed us to treat INF, AS, and SE samples at each dilution factor as technical replicates in downstream analyses. Total sequencing yield (Y_{seq}) approached saturation at the highest input w/w% and was stable across replicates (mean ± standard deviation; 0.71 ± 0.27) (Figure 6-1B). This stability was not maintained at subsequent dilution factors, and Y_{seq} values steadily fell to a minimum at 2×10^{-5}

w/w% (0.132 ± 0.06) as the detection of individual sequins became sporadic (Figure 6-S2, S3). A recovery in Y_{seq} was then observed as read counts began to surpass stochiometric inputs.

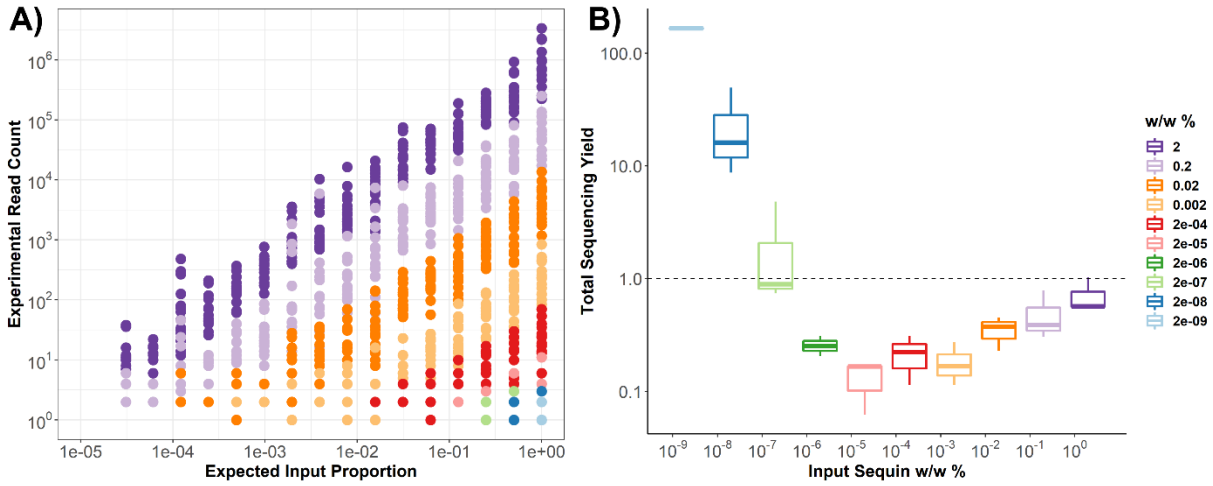


Figure 6-1: Performance of sequin ladders across dilution factors. (A) Sequin "Mixture A" ladder linearity at decreasing input w/w% and constant sequencing depth. (B) Total average sequencing yield of sequin mixtures per sample and input w/w%. Dashed line marks the theoretical Y_{seq} value of 1, representing equal proportions of input sequin mass and experimentally-detected sequin reads.

Defining the LoQ and LoD

To quantitatively explore the LoQ and LoD, the coefficient of variation (CV) of detected read counts for individual sequins were analyzed as a function of calculated input copies per µL of DNA extract (Figure 6-2). Here, INF, AS, and SE samples at each input w/w% were treated as technical replicates. Using general recommendations for qPCR experiments, we first stringently defined the LoQ as the lowest input sequin concentration that was detected across all three technical replicates with a read count CV ≤ 0.35, and the LoD as the lowest individual sequin concentration detected across all three technical replicates (Forootan et al., 2017). The LoQ of the study was determined as 4.7×10^4 copies/µL ($1.1\text{x}10^{-8}$ ng/ng) with a LoD of 2.9×10^4 copies/µL ($1.8\text{x}10^{-8}$ ng/ng). However, there are no general guidelines for qMeta approaches for gene quantification and we explored more relaxed cutoffs for defining the LoQ by increasing the acceptable CV threshold to 1 with detection of an individual sequin in at least duplicate. These relaxed cutoffs resulted in a new calculated LoQ of 2.7×10^4 copies/µL ($3.0\text{x}10^{-8}$ ng/ng). We also redefined the LoD as simply the lowest concentration of a sequin detected with at least a single read, however, this occurred in several instances. Sequins that were detected with a single read were found at an input concentration range of 54 – 5500 copies/µL (1.9×10^{-11} – 3.7×10^{-9} ng/ng), which when converted to a per volume of water filtered basis, equates to approximately 1.0 – 243.6 input copies/mL. There was not an instance of a sequin being detected at a calculated input copy number < 1, suggesting strong agreement between theoretical stochiometric calculations and the physical presence of individual sequins.

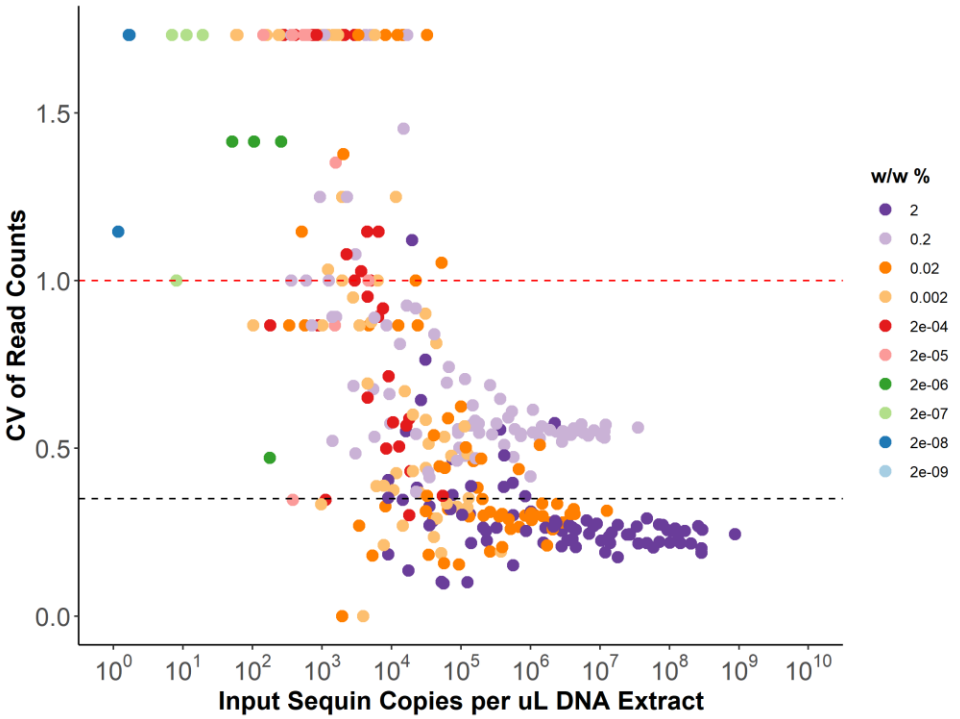


Figure 6-2: Determining the LoQ and LoD of qMeta. INF, AS, and SE at each w/w% were treated as technical replicates. The black dashed line marks the recommended threshold CV = 0.35 for qPCR experiments. The red dashed line marks an adjusted threshold CV = 1 for this study.

High Throughput Quantification of ARGs

The absolute quantification of ARGs across all 30 samples was calculated using the average Y_{seq} of the first dilution factor across the three sample matrices (0.721). We detected an average of 802 ± 62, 401 ± 28, and 542 ± 28 unique ARGs, resulting in total measured concentrations of $10^{8.6\pm0.2}$, $10^{7.9\pm0.001}$, and $10^{6.3\pm0.02}$ gc/mL in the INF, AS, and SE samples, respectively. We then applied the LoQ to each sample matrix to determine the proportion of individual ARGs and total ARG concentrations that were not statistically quantifiable. For the detection of individual ARG occurrences, we found that on average only 24.2%, 13.8%, and 12.7% of ARGs were above the LoQ, resulting in total ARG concentrations of $10^{8.3\pm0.4}$, $10^{7.8\pm0.2}$, and $10^{6.2\pm0.2}$ for INF, AS, and SE, respectively. After the application of the LoQ, the calculated concentration of total ARGs was not significantly different than before the application of the cutoff (paired t-test, p=0.86)

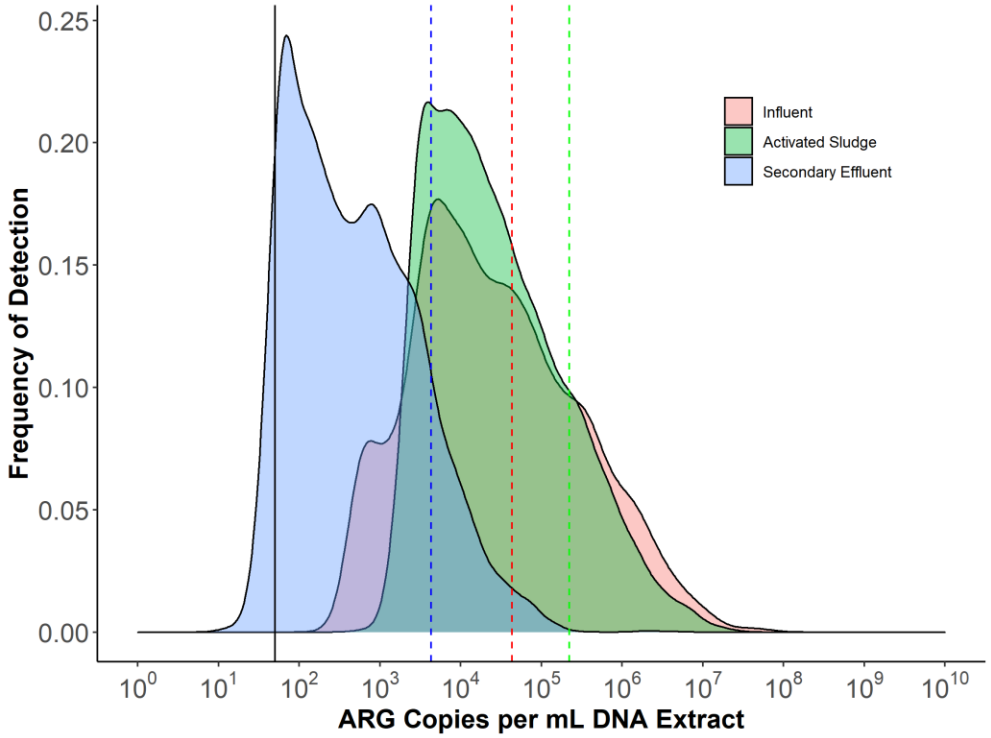


Figure 6-3: Density plot of detected ARGs across concentration range. The dashed lines represent the LoQ of the entire experiment ($2.7x10^4$ gc/uL) converted to gc/mL for each sample matrix. Red = Influent, Green = Activated Sludge, Blue = Secondary Effluent. The black solid line represents the LoD.

Comparison of qMeta and qPCR

The qMeta approach was verified by comparing absolute gene quantities (gc per volume of water filtered) to replicate qPCR measurements of two target genes, 16S rRNA and *sul*1 (Figure 6-4). For qMeta gene calculations, all ten biological replicates were used with the average Y_{seq} values as previously described. We found no significant difference between gene quantities derived from qMeta and qPCR for 16S rRNA (paired t-test, $p = 0.64$) or *sul*1 (paired t-test, $p = 0.45$), although these targets were well above the LoQ. Comparing individual gene quantities in each sample matrix, however, we found significant differences in absolute gene concentrations between *sul*1 measurements derived from qPCR and qMeta in SE samples (paired t-test, $p < 0.01$). Despite these differences in derived gene target concentrations, the variance between calculated gene concentrations by each methodology were insignificant. For the 16S rRNA gene concentrations, the CV of qPCR-derived concentrations (0.365 ± 0.03) was not significantly different than qMeta (0.445 ± 0.147) (paired t-test, $p = 0.41$). The same was observed for *sul*1, where the CV of qPCR concentrations (0.42 ± 0.209) and qMeta (0.472 ± 0.132) showed no significant difference (paired t-test, $p = 0.95$).

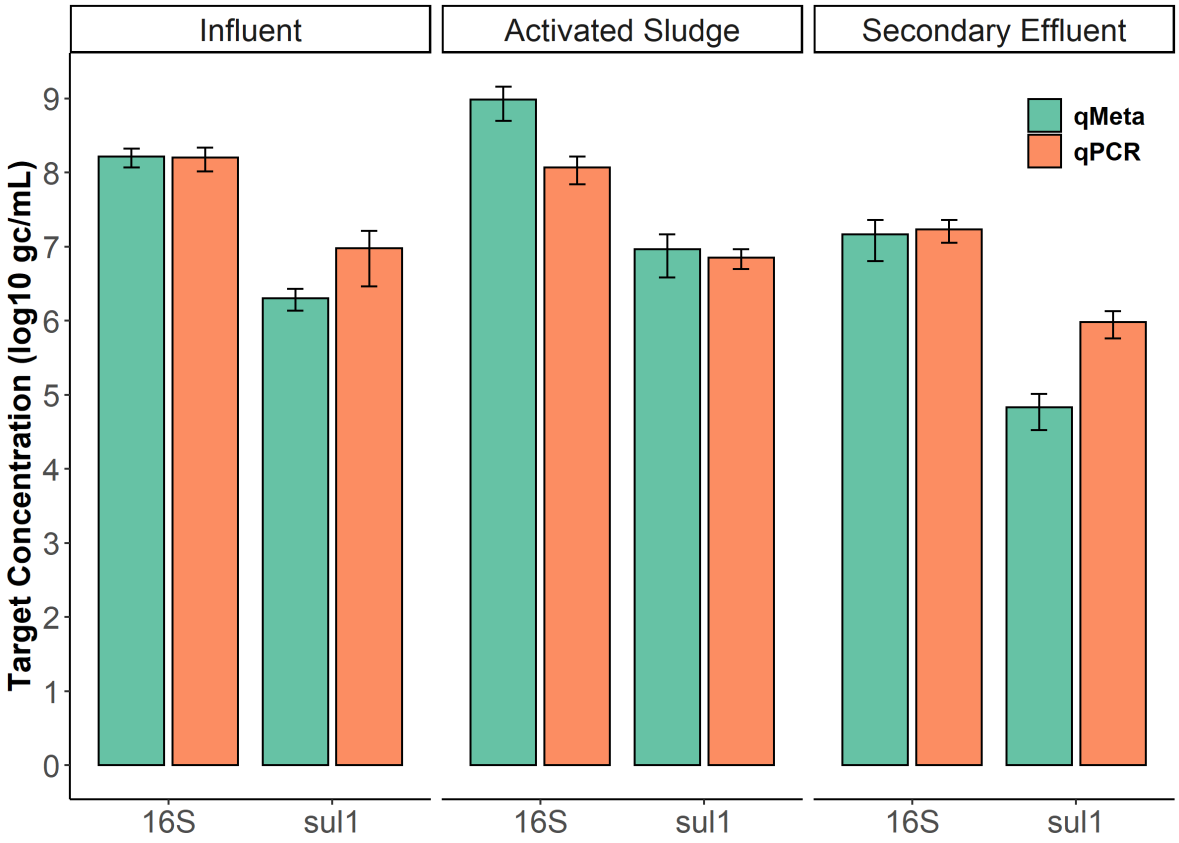


Figure 6-4: Comparison of qMeta and qPCR for absolute gene quantification. qPCR data for each sample matrix represents all 10 biological replicates in technical triplicate. qMeta represents the first 3 dilutions. Error bars represent ± the standard deviation between replicates.

DISCUSSION

This study demonstrated that qMeta is a feasible approach to broad, non-targeted monitoring of ARGs in wastewater samples. Benchmarking to qPCR, the most sensitive and quantitative method currently available for individual ARG monitoring, demonstrated no measurable difference in quantification of target genes as gc/mL. This finding could be a significant game-changer, as the lack of quantitative capacity has been cited as a key limitation of the applicability of metagenomic data to inform human health risk assessment. For qMeta to be widely and successfully adopted, it will be necessary to promote common protocols to sample preparation, including the spiking of internal reference standards, such as meta sequin ladders, and sequencing to sufficient depth to capture targets of interest. However, it is important to recognize that there are still inherent limitations of metagenomics for ARG monitoring.

In particular, rare targets, i.e., below ~10^3 gc/uL of DNA extract, will be extremely difficult to capture with acceptable consistency, even at sequence depths of 100 Gb with biological replication. Based on this study, we can estimate a theoretical LoQ and LoD of

2.7×10^4 gc/uL and 54 gc/uL, respectively. Although the LoD presented here was further calculated to reach as low as 1 gc/mL, ARGs occurring at these extremely low concentrations will likely go undetected during conventional monitoring efforts that use much shallower sequencing depths. This has implications for the design of future WBS efforts because the emergence of novel resistance determinants will appear in wastewater samples well below the technical limitations of Illumina sequencing. Conversely, the ARGs that appear above the LoD, and especially above the LoQ, have likely already undergone the evolutionary jumps to become fixed in human populations (i.e., have been mobilized and are present in human commensals or pathogens). Approaches, however, can be derived from this work to define sequencing depths appropriate for the detection of ARGs on a sliding-scale, allowing a statistically defined threshold of ARGs to go "under the radar". qPCR can be used to supplement and verify the stochastic detection of presumably novel ARGs and new ARG variants with higher specificity and verified using functional metagenomic techniques. Although sequencing costs will undoubtedly continue to fall, the sample preparation and sequencing in this proof-of-concept experimental design resulted in a per-sample cost approaching $1,000. This per-sample cost is well outside the realm of feasibility for large monitoring projects and trade-offs will inevitably need to be made for non-targeted approaches to ARG surveillance.

These limitations in gene detection at low abundances may be an artefact of the Illumina sequencing platform itself. We used a PCR-based KAPA HyperPrep library preparation kit with consistent input of ~1000 ng of environmental DNA and observed minimal bias in the reconstruction of meta sequin ladders across sequin lengths and GC content that was maintained at low input concentrations. We did observe, however, a fall-off in Y_{seq} values as meta sequin inputs fell below 2×10^{-5} w/w% which then recovered to exceed the theoretical value of 1. Based on these observations, Illumina sequencers may simply preferentially sequence the most abundant features in a metagenome, sequencing at random the highest concentration genomic fragments. Once a target sequence falls below the LoQ, however, its detection becomes a matter of chance and the number of total queries generated. In many cases, at the lowest spike-in concentrations, the number of detected meta sequins greatly surpassed the stochiometric concentration that were spiked-in ($Y_{seq} > 1$), supporting this hypothesis.

Overall, this study supports the use of qMeta approaches for continued efforts in WBS of antibiotic resistance and made headway in the statistical analyses necessary for reproducible ARG detection. This approach will be especially useful in the future application of molecular data types that inform quantitative molecular assessments of both human and ecological health risks of ARGs.

REFERENCES

Aarestrup, F.M., Woolhouse, M.E.J., 2020. Using sewage for surveillance of antimicrobial resistance. Science. 367, 630–632. https://doi.org/10.1126/science.aba3432

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., Min, S.Y., Miroshnichenko, A., Tran, H.K.,

Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G. V., McArthur, A.G., 2020. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48, D517–D525. https://doi.org/10.1093/nar/gkz935

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Crossette, E., Gumm, J., Langenfeld, K., Raskin, L., Duhaime, M., Wigginton, K., 2021a. Metagenomic quantification of genes with internal standards. MBio 12, 1–12. https://doi.org/10.1128/mBio.03173-20

Crossette, E., Gumm, J., Langenfeld, K., Raskin, L., Duhaime, M., Wigginton, K., 2021b. Metagenomic Quantification of Genes with Internal Standards: Supplemental Info. MBio.

Davis, B.C., Riquelme, M.V., Ramirez-toro, G., Bandaragoda, C., Garner, E., Rhoads, W.J., Vikesland, P., Pruden, A., 2020. Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c05567

Dutka-Malen, S., Evers, S., Courvalin, P., 1995. Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. J. Clin. Microbiol. 33, 24–27.

Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L., Kubista, M., 2017. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). Biomol. Detect. Quantif. 12, 1–6. https://doi.org/10.1016/j.bdq.2017.04.001

Garner, E., Davis, B.C., Milligan, E., Blair, M.F., Keenum, I., Maile-Moskowitz, A., Pan, J., Gnegy, M., Liguori, K., Gupta, S., Prussin, A.J., Marr, L.C., Heath, L.S., Vikesland, P.J., Zhang, L., Pruden, A., 2021a. Next generation sequencing approaches to evaluate water and wastewater quality. Water Res. 194, 116907. https://doi.org/10.1016/j.watres.2021.116907

Garner, E., Inyang, M., Garvey, E., Parks, J., Glover, C., Grimaldi, A., Dickenson, E., Sutherland, J., Salveson, A., Edwards, M.A., Pruden, A., 2019. Impact of blending for direct potable reuse on premise plumbing microbial ecology and regrowth of opportunistic pathogens and antibiotic resistant bacteria. Water Res. 151, 75–86. https://doi.org/10.1016/j.watres.2018.12.003

Garner, E., Organiscak, M., Dieter, L., Shingleton, C., Haddix, M., Joshi, S., Pruden, A., Ashbolt, N., Medema, G., Hamilton, K.A., 2021b. Towards risk assessment for antibiotic resistant pathogens in recycled water: A systematic review and summary of research needs. Environ. Microbiol. 00. https://doi.org/10.1111/1462-2920.15804

Haas, C.N., 2020. Quantitative Microbial Risk Assessment and Molecular Biology: Paths to Integration. Environ. Sci. Technol. 54, 8539–8546. https://doi.org/10.1021/acs.est.0c00664

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., Lovelock, C.E., Neilan, B.A., Mercer,

T.R., 2018. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nat. Commun. 9, 1–10. https://doi.org/10.1038/s41467-018-05555-0

Hardwick, S.A., Deveson, I.W., Mercer, T.R., 2017. Reference standards for next-generation sequencing. Nat. Rev. Genet. 18, 473–484. https://doi.org/10.1038/nrg.2017.44

Hardwick, S.A., Stokes, H.W., Findlay, S., Taylor, M., Gillings, M.R., 2008. Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. FEMS Microbiol. Lett. 278, 207–212. https://doi.org/10.1111/j.1574-6968.2007.00992.x

Hendriksen, R.S., Munk, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., Kaas, R.S., Clausen, P.T.L.C., Vogt, J.K., Leekitcharoenphon, P., Schans, M.G.M. van de, Zuidema, T., Husman, A.M. de R., Rasmussen, S., Petersen, B., Consortium, the G.S.S. project, Amid, C., Cochrane, G., Sicheritz-Ponten, T., Schmitt, H., Alvarez, J.R.M., Aidara-Kane, A., Pamp, S.J., Lund, O., Hald, T., Woolhouse, M., Koopmans, M., Vigre, H., Petersen, T.N., Aarestrup, F.M., 2019. Global monitoring of antimicrobial resistance based on human sewage. Nat. Commun. https://doi.org/10.1038/s41467-019-08853-3

Keenum, I., Liguori, K., Calarco, J., Davis, B.C., Milligan, E., Harwood, V.J., Pruden, A., 2022. A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. Crit. Rev. Environ. Sci. Technol. 0, 1–25. https://doi.org/10.1080/10643389.2021.2024739

Kim, J., Lim, Y., Jeong, Y., Seol, S., 2005. Occurrence of CTX-M-3, CTX-M-15, CTX-M-14, and CTX-M-9 Extended-Spectrum B-Lactamases in Enterobacteriaceae Clinical Isolates in Korea. Antimicrob. Agents Chemother. 49, 1572–1575. https://doi.org/10.1128/AAC.49.4.1572

Li, B., Li, X., Yan, T., 2021. A Quantitative Metagenomic Sequencing Approach for High-Throughput Gene Quantification and Demonstration with Antibiotic Resistance Genes. Appl. Environ. Microbiol. 87, 1–11. https://doi.org/10.1128/AEM.00871-21

Majeed, H.J., Davis, B.C., Riquelme, M. V., Gupta, S., Angeles, L., Aga, D.S., Garner, E., Vikesland, P.J., Pruden, A., 2021. Evaluation of metagenomic-enabled antibiotic resistance surveillance at a conventional wastewater treatment plant. Front. Microbiol. https://doi.org/10.3389/fmicb.2021.657954

Mao, D., Yu, S., Rysz, M., Luo, Y., Yang, F., Li, F., Hou, J., Mu, Q., Alvarez, P.J.J., 2015. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Res. 85, 458–466. https://doi.org/10.1016/j.watres.2015.09.010

McConnell, M.M., Truelstrup Hansen, L., Jamieson, R.C., Neudorf, K.D., Yost, C.K., Tong, A., 2018. Removal of antibiotic resistance genes in two tertiary level municipal wastewater treatment plants. Sci. Total Environ. 643, 292–300. https://doi.org/10.1016/j.scitotenv.2018.06.212

Ng, L.K., Martin, I., Alfa, M., Mulvey, M., 2001. Multiplex PCR for the detection of tetracycline resistant genes. Mol. Cell. Probes 15, 209–215. https://doi.org/10.1006/mcpr.2001.0363

Pei, R., Kim, S.C., Carlson, K.H., Pruden, A., 2006. Effect of River Landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). Water Res.

40, 2427–2435. https://doi.org/10.1016/j.watres.2006.04.017

Pruden, A., Arabi, M., Storteboom, H.N., 2012. Correlation between upstream human activities and riverine antibiotic resistance genes. Environ. Sci. Technol. 46, 11541–11549. https://doi.org/10.1021/es302657r

Pruden, A., Vikesland, P.J., Davis, B.C., Maria, A., Husman, D.R., 2021. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. Curr. Opin. Microbiol. 64, 91–99. https://doi.org/10.1016/j.mib.2021.09.013

Riquelme, M. V., Garner, E., Gupta, S., Metch, J.W., Zhu, N., Blair, M.F., Arango-Argoty, G.A., Maile-Moskowitz, A., Li, A., Flach, C.-F., Aga, D.S., Nambi, I., Larsson, D.G.J., Bürgmann, H., Zhang, T., Pruden, A., Vikesland, P.J., 2021. Wastewater Based Epidemiology Enabled Surveillance of Antibiotic Resistance. bioarxiv.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016, 1–22. https://doi.org/10.7717/peerj.2584

Suzuki, M.T., Taylor, L.T., 2000. Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5 J -Nuclease Assays Downloaded from http://aem.asm.org/ on November 26 , 2014 by SERIALS CONTROL Lane Medical Library. Appl. Environ. Microbiol. 66, 4605–4614. https://doi.org/10.1128/AEM.66.11.4605-4614.2000.Updated

Waseem, H., Jameel, S., Ali, J., Ur Rehman, H.S., Tauseef, I., Farooq, U., Jamal, A., Ali, M.I., 2019. Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: A critical review. Molecules 24. https://doi.org/10.3390/molecules24010163

Wong, T., Deveson, I.W., Hardwick, S.A., Mercer, T.R., 2017. ANAQUIN: A software toolkit for the analysis of spike-in controls for next generation sequencing. Bioinformatics 33, 1723–1724. https://doi.org/10.1093/bioinformatics/btx038

SUPPLEMENTARY MATERIALS FOR CHAPTER 6

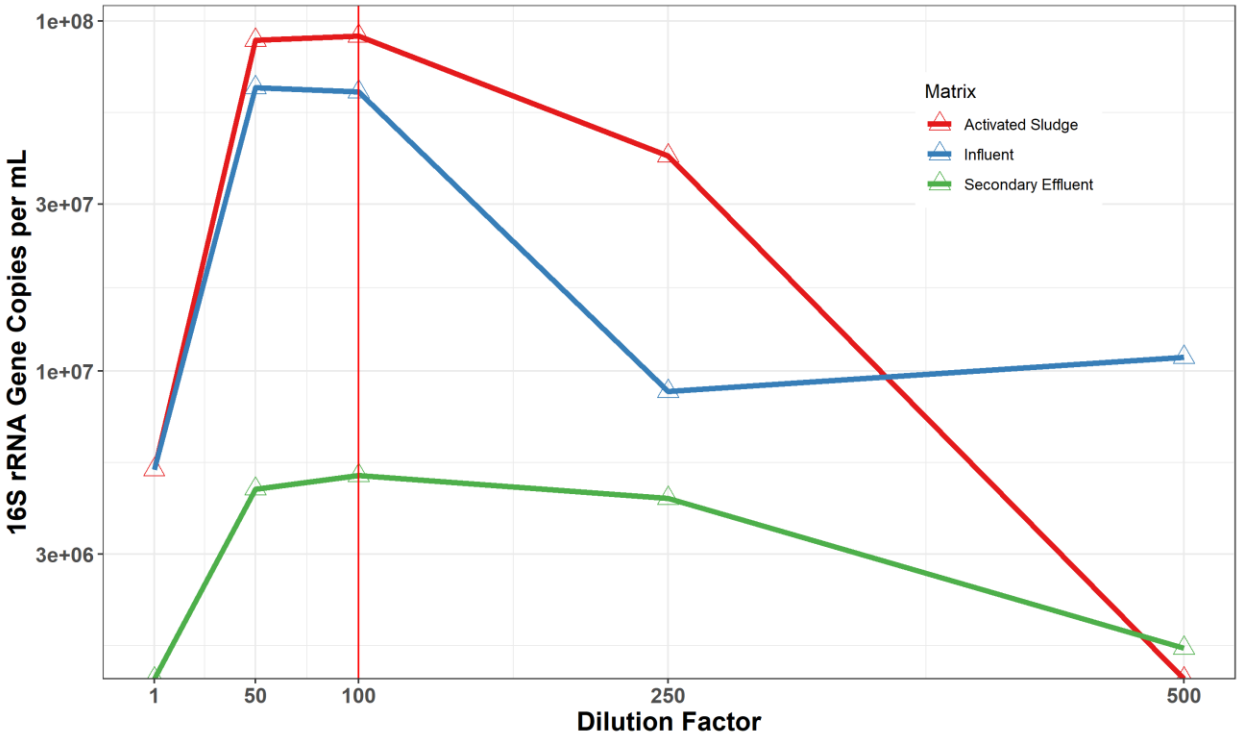


Figure 6-S1: Dilution curve optimizing qPCR gene quantification across wastewater matrices using the 16S rRNA gene as the representative assay. A dilution factor of 1:100 was identified as optimal for all sample matrices.

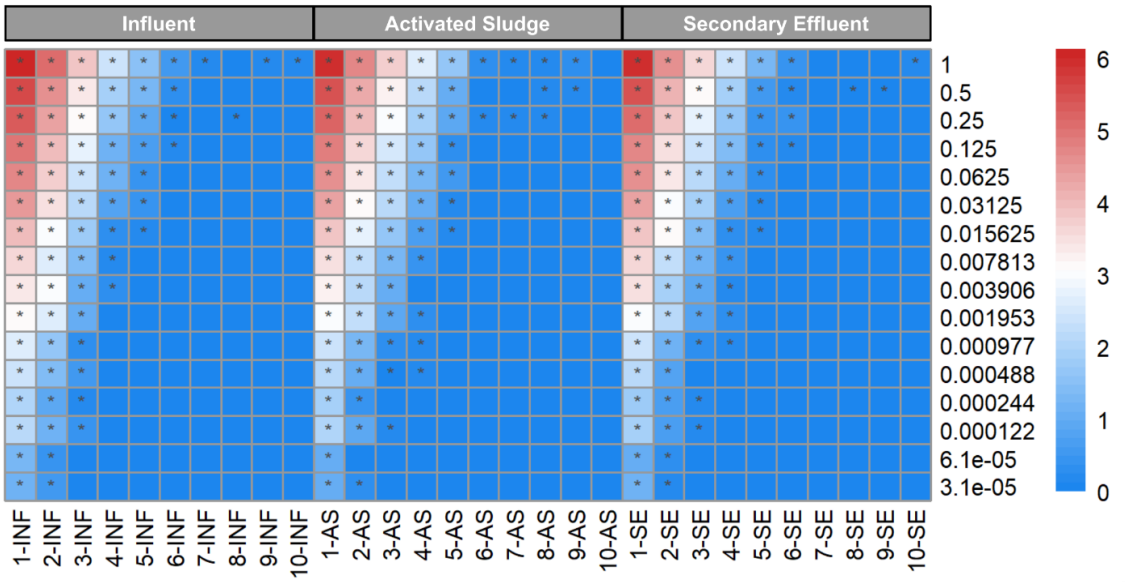


Figure 6-S2: Heat map of detected meta sequins across all 30 samples. Rows represent expected input proportions for meta sequin ladders. The color gradient is the $\log_{10}$ mean detected read count. * = at least a single sequin was detected.

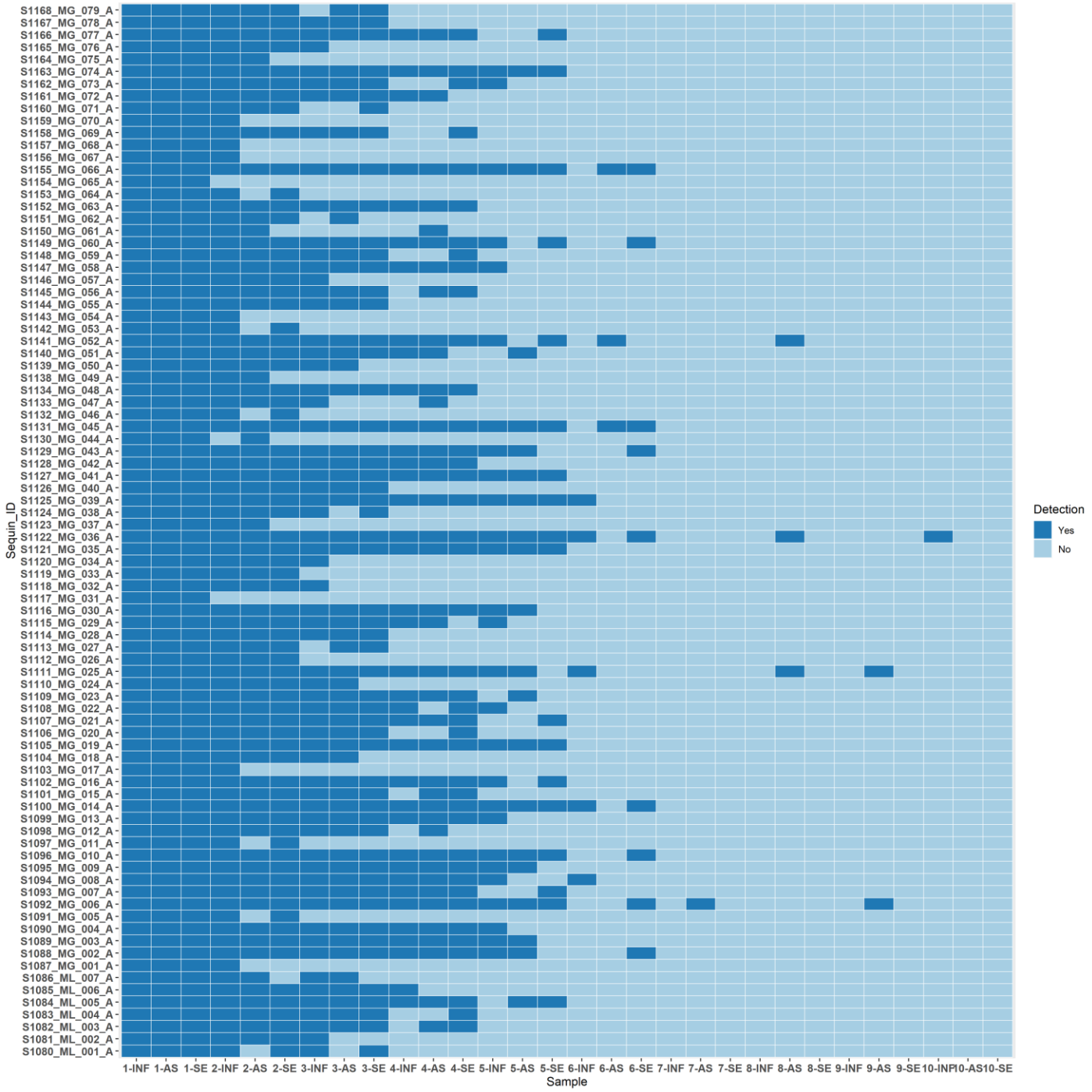


Figure 6-S3: Heat map of individual sequin detections across sample dilutions.

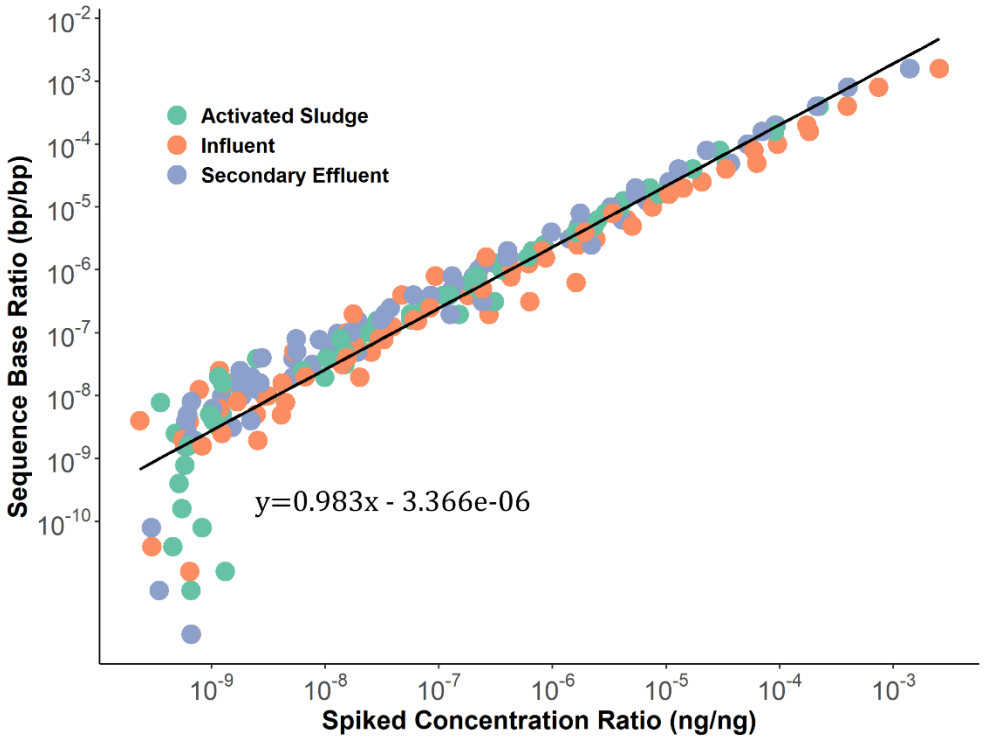


Figure 6-S4: Correlation between spiked meta sequin concentration ratios ($\frac{C_{sequin-i}}{C_{Total}}$) and experimentally-derived sequence base ratios ($\frac{n_{sequin-i}}{n_{Total}}$) across all 30 samples.

Table S3: Summary of sample characteristics and sequin spike-in masses.

Sample	Filter Volume (mL)	Final Qubit (C_{Total})	Sample Mass In (ng)	Sequin Mass In (ng)	Reads Passing QA/QC (150 bp)	Base Pairs (Gb)	Total Sequin Ladder Reads
1-INF	50	28.6	1430	20	535,013,050	80.25	11,040,198
2-INF	50	14.9	745	2	682,264,862	102.34	1,073,673
3-INF	50	19	950	0.2	626,058,174	93.91	56,592
4-INF	50	21.8	1090	0.02	852,102,356	127.82	1,942
5-INF	50	18	900	0.002	537,603,870	80.64	332
6-INF	50	25	1250	0.0002	714,076,610	107.11	24
7-INF	50	20.4	1020	0.00002	484,396,382	72.66	2
8-INF	50	26.2	1310	0.000002	559,475,636	83.92	1
9-INF	50	22.4	1120	0.0000002	621,449,370	93.22	2
10-INF	50	21.8	1090	0.00000002	599,232,550	89.88	2
1-AS	10	19	950	20	663,293,200	99.49	7,551,758
2-AS	10	17.4	870	2	644,302,050	96.65	503,012
3-AS	10	18.9	945	0.2	561,812,936	84.27	42,191
4-AS	10	21	1050	0.02	689,547,724	103.43	3,755

5-AS	10	22.4	1120	0.002	692,374,412	103.86	310
6-AS	10	16.5	825	0.0002	644,008,238	96.60	8
7-AS	10	15.6	780	0.00002	644,583,050	96.69	4
8-AS	10	16.7	835	0.000002	727,957,052	109.19	7
9-AS	10	20.8	1040	0.0000002	607,594,870	91.14	6
10-AS	10	21.6	1080	0.00000002	587,974,204	88.20	0
1-SE	500	17.8	890	20	668,746,476	100.31	7,528,242
2-SE	500	21.6	1080	2	651,113,730	97.67	398,144
3-SE	500	19.9	995	0.2	674,983,984	101.25	30,929
4-SE	500	22.2	1110	0.02	654,267,052	98.14	2,185
5-SE	500	18.1	905	0.002	538,684,262	80.80	123
6-SE	500	17.5	875	0.0002	600,034,330	90.01	20
7-SE	500	7.82	954	0.00002	516,485,834	77.47	0
8-SE	500	19	950	0.000002	675,154,342	101.27	1
9-SE	500	19	950	0.0000002	575,233,546	86.29	1
10-SE	500	19.7	985	0.00000002	604,263,924	90.64	2

Table S2: Sequin “Mixture A” ladders (https://www.sequinstandards.com/metagenome/)

Sequin_ID	LENGTH (bp)	GC	Expected Input Proportion	Molecular Weight (g/mol)
S1092_MG_006_A	4215	0.427284	1	2,566,935
S1100_MG_014_A	2878	0.341557	1	1,752,702
S1155_MG_066_A	981	0.61264	1	597,429
S1122_MG_036_A	2979	0.503525	1	1,814,211
S1125_MG_039_A	2979	0.377308	1	1,814,211
S1096_MG_010_A	3217	0.7106	0.5	1,959,153
S1105_MG_019_A	2559	0.5932	0.5	1,558,431
S1111_MG_025_A	4126	0.293989	0.5	2,512,734
S1127_MG_041_A	2979	0.53575	0.5	1,814,211
S1163_MG_074_A	1644	0.34854	0.5	1,001,196
S1095_MG_009_A	3670	0.601907	0.25	2,235,030
S1102_MG_016_A	2694	0.335932	0.25	1,640,646
S1141_MG_052_A	1780	0.385393	0.25	1,084,020
S1129_MG_043_A	2979	0.507217	0.25	1,814,211
S1131_MG_045_A	2979	0.689157	0.25	1,814,211
S1088_MG_002_A	2112	0.555398	0.25	1,286,208
S1094_MG_008_A	3972	0.496979	0.125	2,418,948
S1107_MG_021_A	2271	0.470718	0.125	1,383,039
S1149_MG_060_A	1515	0.291089	0.125	922,635
S1152_MG_063_A	1137	0.525066	0.125	692,433
S1116_MG_030_A	2979	0.644512	0.125	1,814,211
S1109_MG_023_A	2174	0.301748	0.0625	1,323,966
S1147_MG_058_A	1590	0.51195	0.0625	968,310

S1121_MG_035_A	2979	0.568983	0.0625	1,814,211
S1134_MG_048_A	2979	0.47902	0.0625	1,814,211
S1089_MG_003_A	3810	0.613386	0.0625	2,320,290
S1166_MG_077_A	1137	0.538259	0.0625	692,433
S1093_MG_007_A	4132	0.372459	0.03125	2,516,388
S1099_MG_013_A	2969	0.339508	0.03125	1,808,121
S1108_MG_022_A	2256	0.62367	0.03125	1,373,904
S1140_MG_051_A	1823	0.462425	0.03125	1,110,207
S1161_MG_072_A	996	0.711847	0.03125	606,564
S1090_MG_004_A	2571	0.443796	0.03125	1,565,739
S1084_ML_005_A	4719	0.581055	0.015625	2,873,871
S1098_MG_012_A	2991	0.396189	0.015625	1,821,519
S1114_MG_028_A	1929	0.399171	0.015625	1,174,761
S1145_MG_056_A	1660	0.328313	0.015625	1,010,940
S1115_MG_029_A	2979	0.413562	0.015625	1,814,211
S1162_MG_073_A	999	0.717718	0.015625	608,391
S1101_MG_015_A	2847	0.679663	0.007813	1,733,823
S1106_MG_020_A	2373	0.438685	0.007813	1,445,157
S1139_MG_050_A	1837	0.337507	0.007813	1,118,733
S1144_MG_055_A	1667	0.397121	0.007813	1,015,203
S1128_MG_042_A	2979	0.499161	0.007813	1,814,211
S1167_MG_078_A	1694	0.603896	0.007813	1,031,646
S1110_MG_024_A	2174	0.320147	0.003906	1,323,966
S1085_ML_006_A	9120	0.516667	0.003906	5,554,080
S1113_MG_027_A	2037	0.371625	0.003906	1,240,533
S1148_MG_059_A	1557	0.428388	0.003906	948,213
S1151_MG_062_A	1486	0.369448	0.003906	904,974
S1158_MG_069_A	987	0.678825	0.003906	601,083
S1082_ML_003_A	7059	0.420173	0.001953	4,298,931
S1146_MG_057_A	1644	0.330292	0.001953	1,001,196
S1120_MG_034_A	2979	0.418597	0.001953	1,814,211
S1126_MG_040_A	2979	0.635784	0.001953	1,814,211
S1168_MG_079_A	1929	0.372214	0.001953	1,174,761
S1083_ML_004_A	6227	0.524651	0.000977	3,792,243
S1104_MG_018_A	2571	0.457799	0.000977	1,565,739
S1112_MG_026_A	2038	0.443081	0.000977	1,241,142
S1150_MG_061_A	1515	0.333333	0.000977	922,635
S1160_MG_071_A	999	0.678679	0.000977	608,391
S1081_ML_002_A	5565	0.426415	0.000488	3,389,085
S1118_MG_032_A	2979	0.658946	0.000488	1,814,211
S1133_MG_047_A	2979	0.489426	0.000488	1,814,211
S1156_MG_067_A	999	0.244244	0.000488	608,391
S1165_MG_076_A	1823	0.456939	0.000488	1,110,207

S1080_ML_001_A	5313	0.386222	0.000244	3,235,617
S1091_MG_005_A	2783	0.655767	0.000244	1,694,847
S1143_MG_054_A	1680	0.464286	0.000244	1,023,120
S1153_MG_064_A	999	0.268268	0.000244	608,391
S1119_MG_033_A	2979	0.360524	0.000244	1,814,211
S1086_ML_007_A	4509	0.481703	0.000122	2,745,981
S1138_MG_049_A	1849	0.690644	0.000122	1,126,041
S1123_MG_037_A	2979	0.36623	0.000122	1,814,211
S1124_MG_038_A	2979	0.409198	0.000122	1,814,211
S1157_MG_068_A	999	0.258258	0.000122	608,391
S1103_MG_017_A	2648	0.635952	0.000061	1,612,632
S1142_MG_053_A	1694	0.554309	0.000061	1,031,646
S1154_MG_065_A	999	0.27027	0.000061	608,391
S1132_MG_046_A	2979	0.505539	0.000061	1,814,211
S1087_MG_001_A	3217	0.380168	0.000061	1,959,153
S1097_MG_011_A	2973	0.503195	0.000031	1,810,557
S1117_MG_031_A	2979	0.366902	0.000031	1,814,211
S1130_MG_044_A	2979	0.527694	0.000031	1,814,211
S1159_MG_070_A	999	0.703704	0.000031	608,391
S1164_MG_075_A	1817	0.336269	0.000031	1,106,553

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

In order to thwart the spread and proliferation of antibiotic resistance in the environment, integrated surveillance systems are needed to screen for and characterize ARB and ARGs coming into and out of human populations. An attractive and efficient place to start would be across the world's WWTPs and the receiving water bodies that are impacted by human fecal pollution. Not one single method will suffice, and targeted and non-targeted culture-dependent and independent methods will need to be utilized in tandem with human and animal clinical data to capture the breadth of AR that threatens both human and ecological health in real time. While metagenomics is a promising tool for the implementation of environmental monitoring schemes, significant drawbacks with respect to costs, detection limits, and the need for high-throughput computing resources and technical expertise in data analysis are obvious hindrances to such an effort, especially in low- and middle-income countries. Regardless, much research is needed for the design and implementation of environmental monitoring programs with insightful and reproducible methodologies at their forefront. This dissertation contributed to this need by demonstrating usefulness of metagenomics for describing the occurrence of ARB and ARGs in surface water and wastewater systems and through the conductance of critical literature reviews, provide guidance for the future application of standard methods and analyses for NGS and culture data.

Chapter two of this dissertation demonstrates the applicability of metagenomics for comprehensive characterizations of ARGs in surface waters impacted by anthropogenic pollution including treated wastewater effluents. We explored correlations to human fecal indicators and the genetic context of these ARGs, and bioinformatically assessed the relative risks posed by specific co-occurrences with MGEs and human pathogen gene markers. In addition, specific genetic signatures (e.g., transposon-mediated *bla*KPC ARGs) were putatively linked to local clinical presence of antibiotic-resistant *Klebsiella pneumonia* isolates, establishing clear connections between the environment and the local clinical setting.

The third chapter sought to identify the efficacy of conventional biological treatment on the removal of ARGs across an international cross-section of WWTPs representing distinct compositions of influent ARGs. Metagenomic sequencing revealed distinct successional dynamics of both the microbiome and resistome compositions through influent, activated sludge, secondary effluent, and final treated effluent compartments with strong structural symmetry. Contig analysis revealed the genetic context of ARGs to be predominantly chromosomally bound and directly associated with host taxa, suggesting a limited role of horizontal gene transfer on structuring the resistome through treatment stages. Ultimately, however, high-risk and mobile ARGs traverse WWTPs across the globe and are discharged to surface waters in large quantities, posing distinct human health and ecological health risks.

In the fourth chapter, we sought to comprehensively review next-generation sequencing methodologies for the investigation of aquatic resistomes from sampling design through to

sequencing and data analysis. We identified key shortcomings in current practice of metagenomics including the lack of experimental process controls, a dearth of properly reported metadata in public repositories, a lack of clear reporting standards for ARG detection from public databases, as well as unclear optimized normalization strategies. We further provided a critical discussion of the limitations of the use of short-read Illumina data for contextualizing environmental ARGs and suggest the use of longer Nanopore reads in future monitoring practice for efforts in contextualization. Additionally, we review emerging efforts in qMeta and risk assessment frameworks for environmental resistome analysis. We conclude with the call for the generation of representative and comparable metagenomic data for sharing in public repositories to reap the full benefits of high throughput sequencing for environmental resistome monitoring efforts globally.

Chapter five continues the discussion for standard monitoring targets and methodologies, providing a discussion for the culturing of environmental antibiotic-resistant *Enterococcus* as an attractive Gram-positive organism to complement the popularized ESBL *E. coli*. Efficient culturing and isolate characterization workflows are proposed and individual analytical techniques for phenotyping and genotyping resistant isolates are critically evaluated. We suggest the repurposing of existing standard culture methods to take advantage of extensive QA/QC and existing regulatory recognition. Additionally, we provided aggregated occurrences of phenotypic antibiotic resistance profiles to aid in referencing typical detection frequencies across environmental matrices.

Chapter six puts into practice the recommendations for use of internal reference standards during NGS experiments made by chapter four in a comprehensive benchmarking experiment of qMeta. The statistical thresholds defining the LoQ and LoD of qMeta experiments are then explored and applied to the high-throughput quantification of ARGs. The technique was further verified by proving qMeta's direct comparability to qPCR measurements.

Together, these chapters advance the state of knowledge of the environmental dimension of AR and take strides towards harmonizing monitoring techniques and frameworks for future surveillance efforts. However, these studies raise several research gaps and persistent questions. There is a pressing need for the integration of quantitative microbial risk assessment (QMRA) into AR monitoring that warrants increased research efforts. Given the sheer diversity of ARGs and ARB in natural systems, molecular methods will undoubtedly be at the forefront and much work is needed for bridging the gap between molecular detection of individual ARGs and assessment of both ecological and human health risk. To this end, the continued development, benchmarking, and implementation of internal reference standards for NGS investigations of environmental AR will help generate high-throughput, universally comparable, and quantitative monitoring data that will aid in developing such emerging QMRA models. Further, the genetic context of ARGs, including mobility potential and host viability and pathogenicity need to be better illuminated through the generation of closed bacterial genomes and plasmids at ecological scales. This can ostensibly be achieved through the application of emerging long-read sequencing technologies and hybrid *de novo* assembly techniques.