CHARACTERIZATION OF THE ENVIRONMENTAL RESISTOME IN THE GALAPAGOS ISLANDS, ECUADOR: A ONE HEALTH PERSPECTIVE

Alyssa Grube

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering in the Gillings School of Global Public Health.

Chapel Hill
2021

Approved by:

Jill Stewart

Greg Lewbart

Valeria Ochoa-Herrera

Rachel Noble

Kun Lu

ABSTRACT

Alyssa Grube: Characterization of the Environmental Resistome in the Galapagos Islands, Ecuador: A One Health Perspective
(Under the direction of Jill Stewart)

Antibiotic resistance represents one of our generation's most pressing public health challenges, with some exports warning of an approaching post-antibiotic era. Mitigating this threat requires an understanding of the evolutionary ecology of resistance, including the unique ability of microorganisms to move between humans, animals, and the environment. However, significant questions remain regarding the role of the environment as a source and reservoir for antibiotic resistance. Moreover, there few environments left on earth where we can study background antibiotic resistance in the absence of significant anthropogenic influence.

The Galapagos Islands of Ecuador, where the human population is restricted to 3% of the landmass, represent a unique model system to study how human activity influences antibiotic resistance patterns in wildlife and the environment in a largely protected ecosystem. With samples from humans, animals, and the environment, we designed a *One Health* study aimed at answering what, where, and who: *what* antibiotic resistance genes are present, *where* are they located in regards to mobile genetic elements, and *who* may be the presumptive bacterial host? We employed shotgun metagenomic sequencing to achieve a broad characterization of 90 environmental, wildlife, and human resistomes and mobilomes, and paired this data with targeted detection of the class I integron-integrase gene using a novel ddPCR assay in > 250 Galapagos samples. Additionally, we used a combination of 16S rRNA amplicon sequencing and taxonomic inference from metagenomes to profile the microbial communities associated with these samples.

Our results suggest that human, environmental, and wildlife reservoirs are characterized by distinct resistomes and mobilomes, with overall abundance and diversity of antibiotic resistance genes (ARGs) increasing along a gradient of anthropogenic influence. Overall, we found wildlife to harbor fewer ARGs than wastewater and humans, though some exceptions were noted among land iguanas. Differential abundance analysis revealed ARGs unique to each wildlife species with possible bacterial hosts identified in taxonomic assignments in some cases. We recorded overall agreement between resistome and mobilome data sets, and correlation between taxa, ARGs, and MGEs pointed to a key relationship with *Enterobacteriaceae*.

“The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music.”

--Lewis Thomas, *The Lives of a Cell*

ACKNOWLEDGEMENTS

I am forever grateful to the many people who have helped me grow as a scientist and as a person. As a first generation college student, I have been humbled by the mentorship and kindness of so many that allowed me not only to pursue a PhD, but also combine Spanish with research and travel to places my younger self could not have imagined. Your support and encouragement mean more than you know.

To my advisors, committee, and colleagues at UNC, USFQ, and the GSC:

To my advisor, Dr. Jill Stewart: Going all the way back to the beginning, thank you for responding to my cold email asking if you would support my NSF-GRFP application – we'd only met briefly in person once and yet you wrote a letter to support my dream project studying antibiotic resistance in the Galapagos. Thank you for fiercely advocating for me at every step and opening so many doors that allowed me to conduct research in the Galapagos, present our work in Hong Kong, and learn new techniques in Finland. I could not have imagined a more unique PhD experience from the actual research question to the many places it took me. Thank you for your trust, encouragement, and enthusiasm in supporting both my research project and me as a person and scientist. Your passion for environmental health and genuine empathy for your students and those impacted by your research have been a constant source of inspiration over the last five years. Your example is a continual reminder that being a scientist is both an honor and a privilege. I will forever cherish our many trips to the Galapagos – especially the penguin kayak excursion!

To Dr. Greg Lewbart: The inclusion of so many wildlife samples in this work was only possible with your generosity, time, and efforts. The samples from giant tortoises, marine iguanas, red-footed boobies, sea turtles, and especially land iguanas elevated this project so far beyond what I was able to do myself. Thank you for so enthusiastically supporting my research and for including me in sampling efforts – these are experiences I will never forget. Your passion for veterinary and conservation medicine and energy for research are infectious. Also, thank you for your efforts to rally folks for early morning runs – I count the run up to the highlands and Cristo's infamous rescue among my favorite Galapagos memories.

To Dr. Valeria Ochoa-Herrera: I can't help but smile thinking of the many times and ways our paths have crossed: from you studying abroad at Juniata to me studying abroad at USFQ and tinkering with the bioreactor in your lab; from finding out while applying to graduate school that you and Jill were close collaborators to working on several projects and publishing together. Clearly, it was meant to be! You are a force for environmental quality and health in Ecuador, and I've been so lucky to learn from your example of how to be a fierce and determined scientist. You are also an example of the power of interdisciplinary, multi-lingual collaborations, something I hope I can achieve in my career. Presenting our water quality work in Spanish at the symposium remains the accomplishment I am most proud of from my PhD – thank you for allowing me to be part of that research and for encouraging me to give the talk in Spanish. I hope our paths continue to cross. *Un abrazo apretado*.

To Dr. Rachel Noble: I have learned so much from you and your lab over the last year with the SARS-CoV-2 project. Thank you for being an example of excellence: your attention to detail and rigorous methodology transform good science into exceptional science. You have

challenged me to think more carefully about experimental design, especially with the selection of controls.

To Dr. Kun Lu: Thank you for your support and helpful feedback on my written exams and this dissertation. You challenged me to think more deeply about the 16S rRNA gene and community profiling.

To my lab mates, Corinne, Connor, Knox, Nikhil, David, Megan, Elizabeth, Sarah, and Jennifer: Thank you for being a constant source of knowledge, support, commiseration, and fun over the last five years. I will always remember post-Hurricane sampling trips with stops at Melvin's for lunch.

To the Center for Galapagos Studies, including Dr. Steve Walsh, Phil Page, Kelly Weaver, and Katya McKerr: Thank you for your support, both logistical and emotional, in conducting research in the Galapagos. You are a wonderful group of people and I will miss the Galapagos gatherings and happy hours.

To the Galapagos Science Center and USFQ staff, including Dr. Carlos Mena, Sylvia Sotamba, Soledad Sarzosa, Sofia Tacle, Selene Escobar, and Cristina Vintamilla: This work would not be possible without your support. I am eternally grateful for your help in navigating the sample permitting and export process and for all you to do facilitate research in the Galapagos. It is inspiring to see the impactful work being done by such a powerful team of (mostly) women at the GSC.

To Dr. Amanda Thompson: Thank you for generously contributing the children's fecal samples to this work. They provided important context to the human side of AMR in the Galapagos.

To Dr. Diego Páez-Rosas: Thank you for your generous contribution of sea lion and fur seal fecal samples. These constituted our largest intra-species comparison in the study and led to the first 16S rRNA microbial community profiling for Galapagos sea lions. Thank you for your collaboration, time, and efforts in collecting these important samples.

To Juan Pablo Muñoz and Daniela Alarcon: Thank you for your impressive swimming skills in sampling sea turtles. I am indebted to your kindness and time in helping me obtain these samples for the study. Galapagos is lucky to have such bright, passionate scientists.

To my fellow Galapa-Grads, Hannah and Kristopher: I can't imagine my time in the Galapagos without your company and friendship. Thank you for the laughs, support, and keeping me sane while navigating research in the Galapagos.

To Dr. Sally Bornbusch: I am so grateful that TriCEM brought us together and for our numerous conversations about wildlife poop, AMR, and all things methods. I am forever indebted to your generosity in including 50 of my wildlife samples in your 16S rRNA run. This data set constitutes the first gut community profiling for Galapagos sea lions and it would not exist without your kindness. Thank you for answering my QIIME 2 questions and for performing the initial pre-processing of the data. I look forward to getting that paper out into the world, and wish you the best in your post-doctoral fellowship at the Smithsonian and beyond.

To my family, friends, and community:

To my parents, Stan and Carol Grube: First and foremost, thank you for teaching me the value of hard work and humility. Your examples of hard work and sacrifice not only provided a happy, fulfilling childhood, but also taught me how to live for those you love. Thank you for supporting my plans and encouraging (or at least tolerating) my travels – I never could have

imagined that a kid from Brunnerville would get to see so many places in the name of science. Thank you for trusting me and believing in me.

To my sister, Dr. Hillary Lobar: Thank you for paving the path to advanced academic study and for inspiring me to push myself in school and science. Your grace, patience, work ethic, intelligence, empathy, and humor as you balance being a full-time large animal veterinarian and mom to three (soon to be four!) kiddos never cease to amaze me. You are the best big sister and role model I could ask for.

To my dear friend, Dr. Caroline Solomon: Over the last eleven years of friendship you have shown me what it means to live a life examined. You actively make the world a better place every day with your bravery, radical empathy, and dare I say, stubbornness. You have played an irreplaceable role in shaping my passion for biology and research, and challenge me to be a better scientist and human. Your patients are so fortunate to have you as their advocate.

To Dr. Regina Lamendella: I cannot understate your role in shaping my path in science and research. Thank you for welcoming me into your lab at Juniata, teaching me bioinformatics, and exemplifying how to be a dangerous scientist and passionate teacher – all while still having fun. You are a force to be reckoned with and an inspiration to many.

To my Galapa-family, including Marcelo, Jessie, Romina, and Sasha: How lucky I am to have spent my first summer in Galapagos with you. From fishing excursions to discussing politics at the kitchen table, you made subsequent summers in the Galapagos memorable and always entertaining. Thank you for your unending hospitality, kindness, and concern for my long hours spent in lab. I will never forget your generosity and standing offer for a place at the dinner table.

To my partner, Sean Coots: You have been with me through almost the entirety of this PhD journey. Thank you for supporting my many trips and travels, even when it meant we were apart for months – I deeply appreciate the sacrifices you made so I could follow where science took me. Thank you for being my example of how to manage pressure and stress with grace and grit. Thank you for believing I could do this when I doubted and questioned myself. Thank you especially for your patience, reassurance, hugs, and beers in the sprint marathon that was the last three months of completing this dissertation. I cannot wait to marry you next month and take on the rest of our lives together.

Finally, thank you to the National Science Foundation Graduate Research Fellowship Program, National Science Foundation Graduate Research Opportunities Worldwide Fellowship, Triangle Center for Evolutionary Medicine, Center for Galapagos Studies, and Tinker Association for funding this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABG	Galapagos Biosecurity Agency
AMR	antimicrobial resistance
ARDB	Antibiotic Resistance Genes Database
ARG	antibiotic resistance gene
ARG-OAP	Antibiotic Resistance Genes Online Analysis Platform
ASV	Amplicon Sequence Variant
BD	below the limit of detection
CARD	Comprehensive Antibiotic Resistance Database
ddPCR	droplet digital polymerase chain reaction
DPNG	Direction of the Galapagos National Park
E. coli	*Escherichia coli*
epicPCR	emulsion, paired-isolation, concatenation polymerase chain reaction
ESBL	extended spectrum beta-lactamase
FDR	False Discovery Rate
GLM	Generalized linear model
GSC	Galapagos Science Center
HTSF	High-Throughput Sequencing Facility
MAE	Ecuadorian Ministry of Environment
MGE	mobile genetic element
MLS	Macrolide, Lincosamide, Strepotogramin
NCBI	National Center for Biotechnology Information
PCoA	Principal Coordinate Analysis

PCR	polymerase chain reaction
SNP	Single-nucleotide polymorphism
SSU rRNA	small subunit ribosomal ribonucleic acid
UNC	University of North Carolina

CHAPTER 1: INTRODUCTION AND RESEARCH OBJECTIVES

Antimicrobial resistance (AMR) represents one of today's most pressing global public health challenges, with some warning of an approaching post-antibiotic era. Currently, antibiotic resistant infections claim 700,000 lives annually at the global scale, and recent reports project mortality from AMR infections to exceed cancer deaths by the year 2050 if current prescribing and disposal practices continue (O'Neill, 2016). While antibiotic resistance is well studied in the clinic, the role of the environment as a source for and in the dissemination of AMR organisms and their genes remains poorly understood. Studies in environmental microbiology have begun to estimate the scale of anthropogenic impacts on the environmental resistome through pathways such as wastewater and agricultural run-off, but distinguishing AMR from human versus environmental origin remains a significant challenge. The extent to which AMR bacteria and their genes re-enter the human sphere from the environment similarly remains unknown.

In recognition of the intrinsic connectivity between human, animal, and environmental dimensions, this project uses a One Health approach in the Galápagos Islands to characterize the resistome, the collection of all antibiotic resistance genes and their precursors in both pathogenic and non-pathogenic bacteria. The Galápagos Islands, where the human population is restricted to 3% of the land mass, represent a strategic model system that allows for comparison of environmental and wildlife microbiomes across a gradient of human activities. By applying combining phylogenetic, resistome and mobilome analyses, this project will contribute to basic science questions about environmental antibiotics resistance and reveal how anthropogenic activities shape AMR in a largely protected ecosystem. Moreover, this work offers to provide

insight into AMR carriage by vulnerable and endangered Galápagos wildlife, which may indicate disruption of their endogenous microbiomes as well as the potential for microbial transmission between humans and animals. Collectively, results from this work will advance our understanding of antibiotic resistance in natural systems, a key step in building exposure and risk assessment models for AMR bacteria in the environment. This dissertation includes the following three objectives:

Objective 1: Characterize the antibiotic resistance genes found in wildlife, environmental, and human microbiomes across a gradient of human influence using shot-gun metagenomic sequencing and ARG annotation with publicly available databases.

Hypothesis: We will observe species-differences in ARG carriage; samples from environmental reservoirs and animals proximal to human settlements and wastewater streams will harbor more acquired ARGs.

Approach: This aim will serve to answer *what* is there in terms of genotypic antibiotic resistance across humans, animals, and environmental samples. Shotgun metagenomic sequencing reads will be annotated using three approaches to allow for methods comparison: first, the Antibiotic Resistance Gene Online Analysis Platform (ARG-OAP), a publicly available Blast-based annotation tool that provides a broad view of resistance determinants (Yin et al., 2018); second, and a mapping-based approach as described by Pärnänen et al. 2018 with BowTie2 (Langmead and Salzberg, 2012) against the Resfinder database (Zankari et al., 2012), which includes only acquired antibiotic resistance genes; and third, a mapping-based approach with BowTie2 against the MegaRes database (Doster et al., 2020), which represents a more comprehensive resistance database that also includes genes for metal and biocide tolerance.

Objective 2: Explore the mobility of detected antibiotic resistance genes using annotation of mobile genetic elements (MGE) and ddPCR to quantify a general and clinical variant of the class I integron (*intI1*).

Hypothesis: Human impacted samples, including wastewater and the receiving beach waters at Playa Carola, will have greater diversity and total abundance of MGEs compared to wildlife and environmental samples. Additionally, samples will exhibit differences in the detection of the general versus clinical *intI1* variant, with higher concentrations of the clinical variant in human-associated samples.

Approach: This aim will contribute to answering *where* ARGs are located in terms of their relation to MGEs. Metagenomes will be mapped to a custom database of mobile genetic elements as described by Pärnänen et al. 2018. DNA from a larger set of approximately 260 samples will be interrogated for two variants of the class I integron, a proposed environmental marker of anthropogenic pollution (Gillings et al., 2015) using digital droplet PCR (ddPCR) and standardized to 16S rRNA gene counts as determined by ddPCR.

Objective 3: Identify potential bacterial hosts of detected antibiotic resistance genes using 16S rRNA community profiling and taxonomic classification of metagenomic sequences.

Hypothesis: Microbial communities will be different between species, with mammalian microbiomes more similar to human microbiomes than reptiles.

Approach: This aim will answer *who* in the context of bacterial host. Fulfillment of Objective 2 will involve further characterization of the 90 metagenomes produced in Objective 1 through identification and taxonomic classification of subunit ribosomal ribonucleic acid (SSU rRNA) sequences from the metagenomic libraries. A subset of samples will also be subjected to

microbial community profiling via sequencing of the 16S rRNA gene. Results from 16S rRNA amplicon sequencing will be compared to the taxonomic classifications from paired metagenomes to assess the validity of the metagenomics-based taxonomic assignments.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

Antimicrobial resistance (AMR) represents one of our generation's most pressing global public health challenges, with the number of deaths from antibiotic resistant infections projected to exceed cancer deaths by 2050 (O'Neill, 2016). While AMR is an ancient molecular phenomenon resulting from antibiotics produced naturally by bacteria (D'Costa et al., 2011), we have amplified the scale of resistance through antibiotic overuse in the clinic, poor disposal practices, and large-scale application in livestock agriculture (Davies & Davies, 2010). Growing evidence beginning with a 1969 report of 'resistance factors' in the soil and fecal samples from an isolated population in the Solomon Islands (Gardner et al., 1969) points to the environment as a potential source and reservoir for AMR bacteria and their genes. Subsequent studies in the decades since have revealed remarkable homology in the resistance profiles of endogenous soil bacteria and clinically isolated microorganisms (Benveniste & Davies, 1973; D'Costa et al., 2006), further supporting this notion. Additionally, recent reports have identified hot-spots for the introduction of AMR bacteria and their genes into the environment, including wastewater (Ng et al., 2017), landfill leachate (Zhao et al., 2018), hospital waste (Wang et al., 2018), and livestock waste (Hong et al., 2013).

Taken together, these observations have catalyzed intense efforts to understand the resistome, the collection of all antibiotic resistance genes and their precursors in both pathogenic and non-pathogenic bacteria (Wright, 2007). While we now understand antibiotic resistance genes (ARGs) to exist ubiquitously in the environment, quantifying the effects of anthropogenic inputs to the resistome – and translating them into human health risk – remains a significant

challenge (Ashbolt et al., 2013). In characterizing the resistome, key questions include: Who is carrying resistance genes, in terms of bacterial taxonomy? *Which* genes are present? What is the *genetic context* of these genes? Are they located on *mobile genetic elements* that can be shared through horizontal gene transfer? Currently, no single method exists to simultaneously answer all of these questions, which has precluded translation of environmental AMR studies into exposure and risk assessment models. However, new innovative techniques such as targeted metagenomic sequencing (Quan et al., 2019) and epicPCR (Spencer et al., 2016) offer to revolutionize our understanding of the resistome.

Defining the Resistome

The introduction of the first antibiotics in the late 1930s was met with great confidence in their ability to revolutionize modern medicine: in the war against microorganisms, humans had irreversibly taken the lead. Indeed, in the last 80 years, antibiotics have proven essential not only to treating active bacterial infections but also preventing them, permitting advanced surgical procedures and administration of chemotherapy to immunocompromised individuals (Robinson et al., 2016). However, the apparent victory over pathogenic microorganisms was short lived. Introduction of the first sulfonamide antibiotic (Protonsil) in 1937 was followed by reports of resistant organisms before the end of the decade (Wright, 2007; Davies and Davies, 2010). Discovery of bacterial penicillinase capable of inactivating beta-lactamase drugs predated introduction of penicillin in the early 1940s (Abram and Chain, 1940). Similarly, organisms resistant to streptomycin and methicillin were isolated within only a few years of their clinical introduction (Davies and Davies, 2010; Wright, 2007).

Initially, the emergence of antibiotic resistance was thought to be in direct response to clinical application of antibiotics: either intrinsically resistant members of the bacterial community survived the therapeutic attack, or bacteria acquired *de novo* mutations that permitted

their survival (Heinemann, 1999). In either case, early reports recognized the potential for resistant organisms to share their resistant elements through horizontal gene transfer (Kasuya, 1964; Davies and Davies, 2010). However, researchers soon began to note the existence of environmental antibiotic resistance in the apparent absence of therapeutic selection pressures, including the 1969 description of "resistance factors" in the soil and fecal samples from an isolated population in the Solomon Islands (Gardner et al., 1969). This work was followed by a description of soil-dwelling actinomycetes harboring aminoglycoside inactivating enzymes functionally identical to those isolated from clinically resistant organisms (Benveniste and Davies, 1973). The reality that most aminoglycoside antibiotics are produced by actinomycetes was not lost on Benveniste and Davies; they interpreted their work as a possible origin for resistance and provided keen insight into the natural function of antibiotics: "The actinomycetes may excrete antibiotics in the soil in order to compete effectively with other soil microorganisms for nutrients, and it could be that some gram-negative or gram-positive bacteria have acquired inactivating enzymes in order to protect themselves against these antibiotics" (1973). Moreover, Benveniste and Davies noted that similarity in the resistance enzymes did not imply direct transfer, but rather "…genetic transfer by conjugation or transduction may have occurred through a chain of closely related organisms, even though the initial donor and final recipient may be totally unrelated bacterial species" (1973).

These insightful interpretations supported the developing notion that antibiotic resistance is an ancient phenomenon resulting from millennia of molecular evolution. More recent work has suggested that the beta-lactamase synthetic pathways in actinomycetes date back two billion years (Baltz, 2005), preceding even the divergence of Gram-positive and Gram-negative bacteria (Wright, 2007). Taken together, the body of research since the observation of resistance to

sulfonamide antibiotics points to the reality that bacterial populations are exquisitely well-equipped to respond and adapt to antibiotics, particularly those designed to limit bacterial reproduction (Heinemann, 1999). In recognition of the ubiquity and diversity of resistance mechanisms, D'Costa and colleagues (2006) first suggested the term *resistome* in their landmark paper describing the antibiotic resistance profiles of soil-dwelling actinomycetes. Briefly, the authors constructed a library of 480 actinomycetes strains isolated from diverse soil types, and screened them against a panel of 21 antibiotics covering all major antibiotic classes from natural products to synthetics. Not only was every strain resistant to at least one antibiotic, but they demonstrated multi-drug resistance to seven or eight antibiotics on average, with two strains proving resistant to 15 of the 21 drugs. Together, these 480 strains yielded 200 distinct resistance profiles, highlighting the diversity of resistance mechanisms in soil-dwelling organisms. Even more surprisingly, 11% of the strains demonstrated resistance to the synthetic ciprofloxacin, despite no apparent exposure to fluoroquinolones. Cloning and sequencing of the quinoline resistance-determining region of these clones revealed incredible variation in amino acid sequence, suggesting mechanistic diversity. Echoing the thoughts of Benveniste and Davies, D'Costa and colleagues were conservative in their interpretation, noting that while the presence of resistance elements in environmental bacteria did not confirm transfer of resistance to pathogenic bacteria, the work emphasized the "previously underappreciated density and concentration of environmental antibiotic resistance." This interpretation is especially compelling considering the study was performed with only culturable, soil-dwelling actinomycetes and at high antibiotic concentrations, suggesting even greater diversity of resistance mechanisms in the broader, unculturable soil bacterial community.

Following the first mention of *resistome* in 2006, Wright offered a more comprehensive definition, calling the resistome, "The collection of all antibiotic resistance genes and their precursors in both pathogenic and non-pathogenic bacteria" (Wright, 2007). His definition also includes cryptic resistance genes (which may not be expressed) and covers both antibiotic-specific genes and those that function more broadly to inactivate xenobiotics. This definition, though comprehensive, is undeniably nebulous. How can we study resistance mechanisms that *might* emerge? How do we make sense of an efflux system in a non-pathogenic environmental microorganism? How can we predict the selection of a seemingly unrelated heat shock protein in the presence of antibiotics? However, a broad, inclusive definition is essential as a framework for understanding the resistome: as the last decades of research have shown, antibiotic resistance is universal, context-dependent, transferrable, and ever-changing. Focusing narrowly on resistance conferred by chromosomal mutations or only within pathogenic organisms ignores both the incredible genotypic and phenotypic plasticity of microorganisms in response to natural selection and their movement between human, animal, and environmental contexts (Robinson et al., 2016). Perhaps in recognition of the functional short-comings of his definition, Wright followed with a model for the evolution of antibiotic resistance: "Antibiotic resistance proteins evolve from proteins with alternative biochemical functions that function as precursors to resistance elements. Some of these precursor proteins might have modest or fortuitous antibiotic resistance functions or other affinities for the antibiotic that, in the face of selective pressure, evolves into a robust resistance mechanism" (Wright, 2007). While we cannot perfectly predict the exact mechanism of antibiotic resistance in a particular setting, proceeding with a deep appreciation for the diversity and resourcefulness of bacterial genomes will prove key to advancing our understanding of the resistome.

The Role of Naturally Occurring Resistance

As described above, antibiotic resistance mechanisms are as ancient as the biosynthetic pathways that produce these small bioactive molecules. Historically, naturally-occurring resistance was thought to have emerged as a self-protection strategy by antibiotic producers (Davies and Davies, 2010; Benveniste and Davies, 1973). Indeed, antibiotic-producing organisms must devise mechanisms to avoid the toxicity of their products. Descriptions of resistance in antibiotic-producing organisms, including aminoglycoside resistance in Actinomycetes (Benveniste and Davies, 1973), methylenomycin and actinorhodin resistance in *Streptomyces coelicolor* (Chater and Bruton, 1985; Tahlan et al., 2007, respectively), and oxytetracycline resistance in *Streptomyces rimosus* (Mak et al., 2014) support this notion. As a continuation of this paradigm, the intended targets of antibiotics may be intrinsically resistant or may acquire resistance through mutation, genetic rearrangement, or horizontal gene transfer. Intrinsic resistance includes the absence of the intended antibiotic target or the presence of physiological characteristics that confer resistance. For example, Gram-negative bacteria are intrinsically resistant to vancomycin due to their protective cell membrane; in contrast, Gram-positive bacteria are more vulnerable to glycopeptide antibiotics such as vancomycin because the absence of a cell membrane exposes the peptidoglycan layer (Wright, 2007). In this example, Gram-negative bacteria are naturally resistant due to an inherent physiological property. Acquired resistance mechanisms may fall under one of several categories, including mutation of the antibiotic target, chemical transformation/inactivation of the antibiotic, and expulsion of the antibiotic from the cell. Examples of resistance conferred by mutation of the antibiotic target include modification of the 50S ribosome, the target of macrolide antibiotics (Garza-Ramos et al., 2001), and mutations in the in the DNA gyrase which confer resistance to quinolines (Seminati et al., 2005). The diversity and ubiquity of beta-lactamases in bacteria that do not

produce antibiotics, including the chromosomally located *ampC* gene in *E. coli*, serve as good examples of naturally occurring modification enzymes (Fisher et al., 2005). Finally, bacteria may exhibit natural antibiotic resistance because of efflux pumps that expel diverse chemical stressors out the cell, including xenobiotics, heavy metals, and other toxins (Poole, 2005a).

Within these broad classes of resistance mechanisms, bacteria have acquired a myriad of species-specific and antibiotic-specific strategies contributing to the whole of the resistome. The contribution of multiple molecular mechanisms to a resistance phenotype is not uncommon, such as the general efflux pump MexXY combined with the collection of aminoglycoside-modifying enzymes in *Pseudomonas aeruginosa* (Poole, 2005b). Moreover, recent reports acknowledge the consequence of silent resistance genes, which do not confer resistance in their endogenous host but manifest in resistance when transferred to a heterologous expression system. The dissemination of *ampC* genes from chromosomes, where they are minimally expressed, onto plasmids with robust promoters has played a significant role in the spread of beta-lactamase resistance (Dantas and Sommer, 2012). Examples such as these underscore the need for a broad understanding of the resistome as resistance genes and their precursors in both pathogenic and non-pathogenic bacteria (Wright, 2007).

In addition to naturally occurring resistance that arises from synthesis of antibiotics by the producer and adaptation by the target, select works have suggested alternative roles for antibiotic resistance. Research in the 1990s suggested a role for beta-lactamases in cell wall recycling, demonstrating that strains lacking either of the beta-lactamase regulatory genes *ampD* and *ampG* lost 40% of their peptidoglycan per generation (Jacobs et al., 1994). More recently, an illuminating assessment by Dantas and colleagues demonstrated that some soil microorganisms can use antibiotics as a sole carbon source (Dantas et al., 2008). In a sense, resistance permitted

these organisms to not only survive but also thrive (in specific, laboratory-controlled conditions). Nonetheless, this work underscores the need to consider the broader ecological role of antibiotics in the environment.

The Anthropogenic Impact on the Resistome

Our interest in the anthropogenic impact on the resistome is largely self-serving, as we need effective antibiotics for modern medicine. Except for some inquiry into the effect of antibiotics on the primary productivity of environmental microorganisms involved in nutrient cycling (Song et al., 2016; Robinson et al., 2016), our interest in the anthropogenic impact on the environmental resistome relates to mitigating the dire projections of the post-antibiotic era (O'Neill, 2016). We have produced and discharged millions of pounds of antibiotics into the environment since their clinical introduction in the 1940s, in applications ranging from human medicine and personal care products to livestock agriculture and aquaculture (Davies and Davies, 2010). While it is not difficult to imagine that pollution with antibiotics paired with metals and other contaminants that co-select for antibiotic resistance has altered the resistome (Gillings et al., 2015), quantifying these effects – and translating them into human health risk – remains a significant challenge (Ashbolt et al., 2013). Research over the last several decades has identified reservoirs of antibiotic resistant organisms, genes, and antibiotics, including the human gut microbiome (Bengtsoon-Palme et al., 2015; Feng et al., 2018), wastewater effluent (Munck et al., 2015; Zhang et al., 2011; Ng et al., 2017), landfill leachate (Zhao et al., 2018), waste from livestock animals (Hong et al., 2013), and hospital waste (Wang et al., 2018). Efforts have been made to correlate the presence of particular antibiotic resistance genes (ARGs) with anthropogenic impact. For example, Zhao et al. (2018) identified four genes (*sul1, sul2, aadA, and bacA*) that could quantitatively predict the total ARG abundance in landfill leachate, while Li et al. (2015) demonstrated that ARG abundances conferring resistance to common human

antibiotics, including aminoglycosides, beta-lactams, sulfonamides, tetracyclines, and quinolines correlated with anthropogenic impact in a range of environmental samples. Recently, application of the new human-specific bacteriophage crAssphage revealed that fecal pollution, rather than in-site selection, influences ARG abundances in human-impacted environments (Karkmen et al., 2019). This newly discovered indicator, along with the class I integron-integrase gene which has been proposed as a general marker or anthropogenic influence and correlates well with ARG abundance (Gillings et al., 2015), represent two promising tools for enhancing our understanding of humans impacts on the resistome.

Our concerns about how the last several decades of antibiotic use have shaped the resistome are warranted. Extensive research demonstrates that bacteria can share resistance and virulence traits through horizontal gene transfer mechanisms including conjugation, transformation, and transduction (Kasuya, 1964; Barlow, 2009; Wright, 2007). Additionally, antibiotic resistance has been shown to correlate with virulence (Robinson et al., 2005), making resistant organisms even more threatening to public health. Moreover, we know that antibiotic resistance genes are often located on mobile genetic elements such as transposons and plasmids which allow for rapid dissemination of resistance traits within a bacterial community (Gillings et al., 2015; Wright, 2017). Ample evidence points to the co-selection of antibiotic resistance determinants with other stressors such as heavy metals and xenobiotics (Gillings et al., 2015). Finally, there is notable homology in the antibiotic resistance gene sequences of particular environmental bacterial and clinically significant pathogens, suggesting a possible origin for clinical ARG as suggested by Benveniste and Davies (1973) and D'Costa and colleagues (2006). Appearance of the chromosomal CTX-M genes from the environmental, rarely pathogenic *Kluyvera spp*. on plasmids of clinically resistant isolates suggests recent transfer; confirmation of

mobility of these genes into E. coli plasmids provides further support for this notion (Lartigue et al., 2006). If clinical resistance has environmental origins in at least some cases, then it is plausible that transmission can take place in the opposite direction, from humans to the environment, and back again. Acknowledgement of the scale of antibiotic pollution paired with an appreciation for the mobility of ARGs further supports this idea, as does the understanding that bacteria and their genes have little regard for organizational divisions between the human, animal, and environmental spheres.

Clinical Significance of Environmental Resistance

There is compelling evidence for a link between environmental and clinical resistance, including similarity in the resistance profiles of soil bacteria and human pathogens (Forsberg et al., 2012), the apparent origin of CTX-M genes in environmental *Kluyvera spp.* (Lartigue et al., 2006), and detection of ARGs conferring resistance to modern antibiotics in 30,000-year-old permafrost sediments (D'Costa et al., 2011). Examples of carbapenem resistance also appear to have environmental origins (Potron et al., 2011; Poirel et al., 2008). In Europe, detection of wildlife carrying extended-spectrum beta-lactamase (ESBL) producing *E. coli*, particularly in birds (Simões et al., 2010; Guenther et al., 2011) has mirrored the types of ESBLs observed in the clinic. In a paired study of human clinical isolates and black-headed gulls in Sweden, two ESBL producing bacteria were isolated from the birds, belonging to types CTX-M-14 and CTX-M 15 (Bonnedahl et al., 2010). Incidentally, these were the most common types of ESBL observing in clinical isolates. While this observation does not confirm direct transfer, it certainly suggests some linkage between the environment and clinic. Elsewhere in the world, the emerging story of the New Dehli metallo-beta-lactamase (bla_{NDM-1}) in India appears to support a clinical environmental linkage. Yong et al. (2009) first described bla_{NDM-1} carried by in *Klebsiella pneumoniae*, the causative agent of a urinary tract infection of a woman who had recently

traveled to India. Subsequent work detected NDM-1 positive bacteria New Delhi drinking water and seepage samples but not in sewage effluent from Wales (Walsh et al., 2011), indicating the resistance element likely originated in India. The authors suggested poor sanitation (i.e. contaminated drinking water) as a potential source of antibiotic resistant bacteria.

Despite this growing body of evidence, to date no published studies exist that definitively confirm transmission of resistant organisms or elements from the environment to the clinic. Before the term resistome was coined, Smith and Morris (2005) argued that tracking an antibiotic resistant organism between human, environmental, and animal spheres was virtually impossible: "The complexity of bacterial population biology and genetics makes it practically impossible to trace bacteria (or resistance factors) from the farm to the hospital, or to directly attribute some fraction of new infections to agricultural antibiotic use." Several years later, Dantas and Sommer (2012) attributed the apparent 'missing genetic link' to under sampling of the environmental resistome, and insisted that our best chance of detecting crossover to the clinic would be through corroborative approaches. Recently, Larsson and colleagues (2018) echoed this sentiment, explaining the need for baseline knowledge on the abundance of resistant organisms and ARGs in a particular environmental compartment or human/animal population. While the challenges described by Smith and Morris (2005) are legitimate, methodological advances in the last several years could help identify the missing genetic link. These include publicly available databases and pipelines to annotate antibiotic resistance genes from metagenomic datasets (Yin et al., 2018) and epicPCR, a culture-independent molecular method that links phylogeny and function in an entire bacterial community (Spencer et al., 2016). Strategically coupling methods that answer *who*, *what*, and *where* in terms of host, ARGs, and their genetic context and designing spatiotemporal studies in the overlap between humans,

animals, and the environment is a promising way to improve our understanding of the transmission and persistence of AMR organisms and their genes across environments.

Antimicrobial Resistance: The Quintessential One Health Challenge

The US Centers for Disease Control and Prevention (CDC) defines One Health as "a collaborative, multisectoral, and transdisciplinary approach – working at the local, regional, national, and global levels – with the goal of achieving optimal health outcomes recognizing the interconnection between people, animals, plants, and their shared environment" (CDC, 2018). While this idea may seem straightforward, historical divisions in the practice and philosophy of these fields – along with our tendency to prioritize problems of human concern – has led practitioners to work in isolation on intrinsically multidisciplinary problems. Since its conceptualization, One Health principles have been applied to areas such as zoonotic disease transmission, food safety, and algal blooms. Coordinated reporting of rabies in domestic animals and humans in the United States (Ma et al., 2018) and surveillance of emerging zoonotic viruses in African bushmeat (Mwangi et al., 2016) are good examples of One Health in practice. The One Health paradigm is also featured in global agreements and efforts including the Sustainable Development Goals, the International Health Regulations, the UN Political Declaration on Antimicrobial Resistance, and the UN Paris Agreement on Climate Change (Essack 2018). The One Health approach seeks to coordinate efforts across disciplines and borders because the challenges it seeks to address are inherently multidisciplinary and often transnational in an increasingly interconnected world.

The spread of resistance among microorganisms is one such multidisciplinary problem that fits exceptionally well into the One Health framework. Recently deemed the "quintessential" One Health issue (Robinson et al., 2016; Tiedje et al., 2019). AMR exemplifies at the molecular level the connection between the health of humans, animals, and the environment. Reports of the

similarity between clinical extra-intestinal pathogenic *E. coli* isolates and those cultured directly from meat products (Liu et al., 2018) as well as the mirrored resistance profiles of clinical isolates and bacteria cultured from black-headed gulls in Sweden (Bonnedahl et al., 2010) are among examples that suggest movement of AMR bacteria between human, animal, and environmental reservoirs. **Figure 2.1** places the challenge of antibiotic resistance within a One Health framework, illustrating key relationships and proposed directionality of transmission of AMR bacteria and their genes. The entirety of antibiotic resistance genes and their precursors in both pathogenic and not pathogenic bacteria, called the resistome (Wright, 2007), can be envisioned at the intersection of the three spheres.

If AMR is the quintessential One Health issue, then efforts to curb the projections of the post-antibiotic era will require an understanding of the resistome across human, animal, and environmental reservoirs. Studies designed to sample where these spheres overlap can provide insight into how AMR bacteria and their genes cross boundaries, revealing critical control points to prevent further transmission. Understanding the resistome will also involve inquiry into naturally occurring resistance determinants that exist in the absence of anthropogenic pollution. For the reasons developed below, the Galapagos Islands represent an unmatched model system to help fill some of these knowledge gaps using the One Health approach.

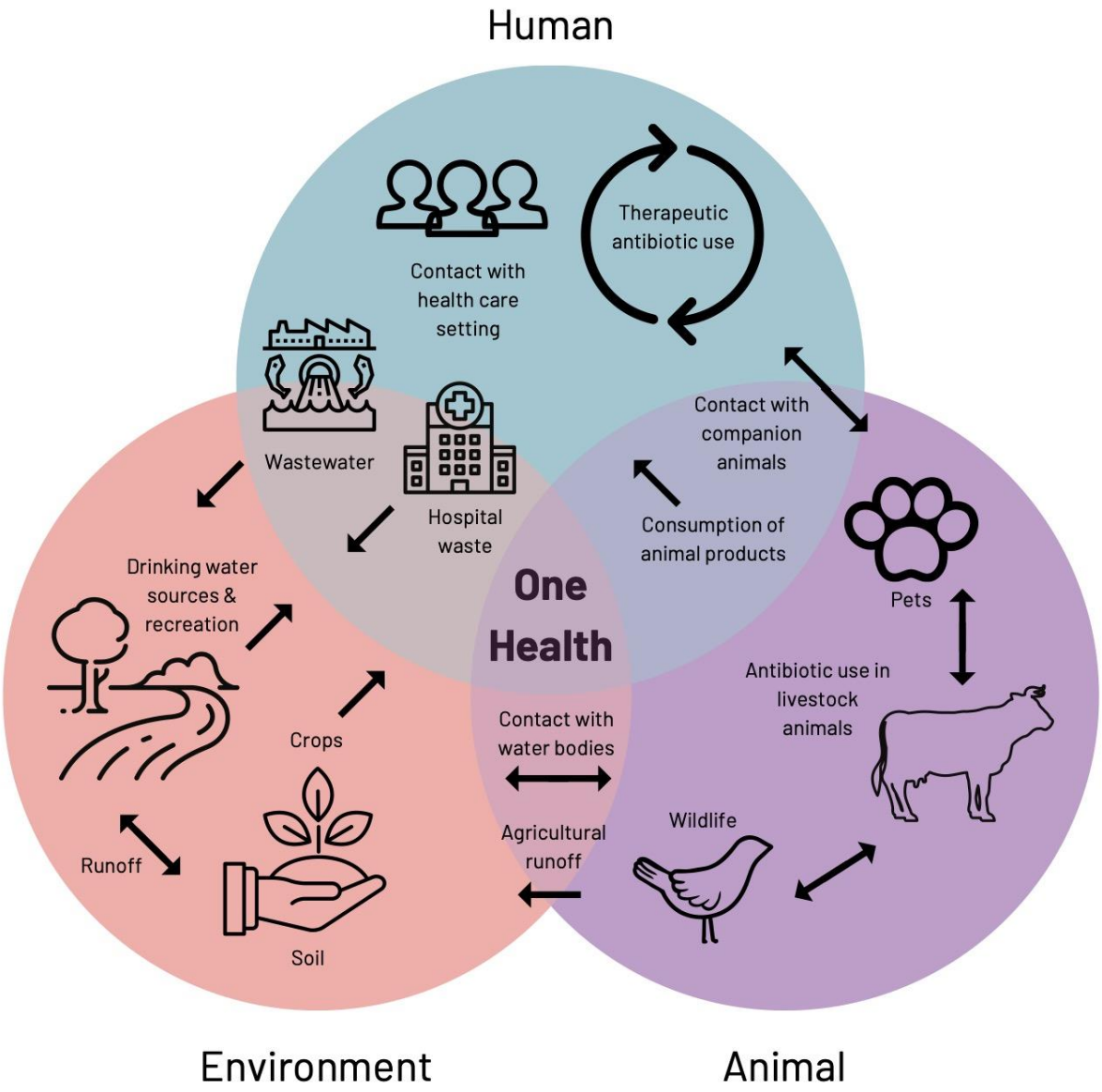

Figure 2.1: *Antimicrobial resistance in a One Health world.* The intrinsic connectivity between human, animal, and environmental spheres makes AMR a One Health challenge. Key relationships and proposed directionality of transmission are noted with arrows. Within the human sphere, therapeutic antibiotic use and contact with the health care setting selects for AMR. Discharge of wastewater from municipal WWTPs and hospitals introduces AMR organisms, genes, and other pollutants into the environment, which can be transported in aquatic and soil matrices and back to humans through drinking water and recreation. Endogenous soil bacterial communities harbor diverse AMR genes which could be transferred via the food chain. Antibiotic use in livestock animals introduces AMR bacteria into the environment through land application of waste and to humans via the food chain. The ***resistome*** encompasses all resistance genes and their precursors in pathogenic and non-pathogenic bacteria circulating between the total of human, animal, and environmental spheres (Wright, 2007). This schematic represents simplification of the network activities and additional transmission routes are likely.

The Galapagos Islands: A Model System for One Health Studies

The Galapagos Islands provide a unique setting for One Health studies due to the physical overlap of humans, domestic animals, wildlife, and endemic plant species within a shared, fragile ecosystem. Researchers have long recognized the islands as a prime example of the conflict between development and conservation (Gonzalez et al., 2008, Walsh & Mena, 2013;), describing the 'Galapagos Paradox' as the tension between preserving the islands and marketing them for tourism (Villacis and Carrillo 2013). The 'Galapagos Paradox' implicitly recognizes the connection between human, animal, and environmental health by describing how actions in one sphere (i.e. human development) affect the others (i.e. environmental quality). In turn, declines in environmental quality can negatively affect both human and animal health (i.e. contaminated water) and economic stability (i.e. tourism and fishing). Recognition of these pathways and feedback loops parallels the One Health principle of interconnectedness.

Accordingly, much of the research conducted in the Galápagos has been influenced by the One Health paradigm, if not explicitly by name, at least in practice. Studies have investigated disease dynamics between domestic animals and wildlife, including possible viral transmission between chickens and the Floreana mockingbird (Deem et al., 2012) and zoonotic diseases from dogs and cats (Levy et al., 2008; Gingrich et al., 2010). In recognition of One Health connectivity, Deem and colleagues (2008) advocated for connecting veterinary medicine and conservation biology to improve the health of wildlife and pointed to several exemplary studies in the Galápagos. While this report related to wildlife health, the recommendation was made in the broader context of anthropogenic change and decreased biodiversity. In an example of the dynamics at the human-wildlife-environment nexus, Alava and coworkers (2009) investigated plastic derivatives in the tissue of sea lions on San Cristobal, connecting a health outcome in wildlife to environmental contamination. Taken together, these examples demonstrate that

Galápagos researchers have long been cognizant of One Health principals, and support the rationale application of such an approach.

Several characteristics make the Galapagos a strategic system for One Health studies aimed at expanding our understanding of the resistome. First, geographic restriction of human settlements to 3% of the landmass allows for comparison between areas under intense anthropogenic influence to protected, uninhabited areas over a short geographic range. It would be difficult to find a comparable system in highly-developed settlings, where the boundary between anthropogenic activities and the environment is unclear if not nonexistent. For example, the presence of a wastewater treatment facility on the inhabited side of San Cristobal in relation to the uninhabited side of Puerto Chino allows for the comparison of heavily impacted receiving waters to a protected beach over the span of a few dozen kilometers. Studies designed to characterize AMR along this gradient could disentangle human-mediated versus naturally occurring sources of antibiotic resistance in the environment, which is currently a challenge in the field (Ashbolt et al., 2013; Larsson et al., 2018). Smalla and colleagues (2018) warn that we may have few to zero "pristine" environments left on earth to serve as baselines for understanding natural AMR. However, the protected areas of the Galapagos, even if not 100% pristine per the definition of Smalla and et al. (2018) can serve as useful representations of very minimally impacted areas.

Secondly, the human colonized areas of Galapagos may represent intense hot spots for AMR introduction. While still restricted to 3% of the land mass, human populations have grown nearly exponentially over the last several decades in concert with the expanding tourism industry (Epler, 2007; Watkins and Cruz, 2007; Walsh et al., 2010). The 2015 census estimated local population at 25,244 (Instituto Nacional de Estadísticas y Censos, 2015), a significant leap over

the 1990 estimate of 9,000 inhabitants (Watkins and Cruz, 2007). Similarly, tourism has seen explosive growth with 275,800 visitors in 2018 (Parque Nacional Galapagos, 2019) compared to 17,000 in 1980, corresponding to an average annual growth rate of 8% (Parque Nacional Galapagos, 2017). Moreover, the nature of tourism in the Galápagos has changed over time, with a marked transition from boat to land-based tourism. In 2007, 51% of tourists stayed on boats and 49% on the islands, versus 68% and 32%, respectively, in 2015 (Parque Nacional Galapagos, 2017). The transition from boat to land-based tourism paired with the significant increase in total tourist numbers bodes significant consequences for the islands. We hypothesize that both the resident and tourist populations represent sources of AMR organisms and genes into the Galápagos ecosystem. Notably, San Cristobal is the only inhabited island with a wastewater treatment plant, and work by Overbey et al. (2015) demonstrated a higher prevalence of antibiotic resistant *E. coli* in coastal waters receiving wastewater effluent than background beaches. Other have proposed the human gut to be a transporter of ARGs across international borders (Bengtsson-Palmee et al., 2015), and metagenomic surveys of human gut samples from across the globe note significant differences in ARG profiles from different geographical regions (Feng et al., 2018). Paired with the observation that travelers often carry or take antibiotics prophylactically or therapeutically, this suggests that tourists could be introducing antibiotics or AMR organisms particularly through wastewater. Notably, the CDC guidance page for travel to Ecuador and the Galápagos recommends antibiotics for traveler's diarrhea on a standard packing list (CDC, 2019).

Finally, conducting One Health studies of AMR in the Galapagos may provide insight into the significance of finding resistant organisms and genes in wildlife. There is a growing consensus that detection of AMR in wildlife signals anthropogenic pollution (Allen et al., 2010;

Vittecoq et al., 2016). On the surface, antibiotic resistance would seem to be of little consequence to wildlife, since they are not administered antibiotics. However, the presence of AMR bacteria in wildlife may have several implications. First, reports of increased virulence in antibiotic resistant bacteria (Robinson et al., 2005; Thomasen et al., 2017) could suggest that AMR carriage in wildlife could exacerbate morbidity and mortality. Secondly, the presence of bacteria carrying anthropogenically-associated ARGs could indicate the possibility for zoonotic disease transmission between humans and wildlife (Vittecoq et al., 2016). Still others interpret AMR bacteria in wildlife as evidence for anthropogenic pollution in general (Al-Bahrey et al 2009, Al-Bahrey et al 2011). Finally, exposure to antibiotics could disrupt the normal commensal flora of wildlife, which has significant complications for the conservation of endangered species. In connection to human health, many speculate that wildlife may be reservoirs for antibiotic resistance which could subsequently be transmitted to domestic animals or humans (Vittecoq et al., 2016; Allen et al., 2010). Finally, examining AMR in wildlife could help to elucidate naturally occurring, rather than anthropogenically selected, forms of resistance.

The Significance of Antimicrobial Resistance in Galapagos Wildlife

On the surface, antibiotic resistance would seem to be of little consequence to wildlife, since they are not administered antibiotics (except possibly in cases of rehabilitation). However, detection of antibiotic resistant organisms and antibiotic resistance genes (ARGs) in wildlife bodes several implications. First, there are observations in human pathogens that antibiotic resistance increases virulence (Robinson et al., 2005; Thomasen et al., 2017). This can be explained by co-selection or co-transmission of virulence factors on mobile genetic elements such as transposons and plasmids (Gillings et al., 2015). In this way, acquisition of an antibiotic resistant pathogen by wildlife may result in increased morbidity and mortality. Secondly, the presence of resistant organisms and genes, particularly those not intrinsic to the bacterial host,

may indicate the potential for disease transmission between human and wildlife, as others have suggested (Wheeler et al., 2012; Thaller et al., 2010b; and Vittecoq et al., 2016). Along this line, antibiotic resistant bacteria in wildlife may signal anthropogenic pollution in general, which was the interpretation of antibiotic resistant bacteria in sea turtle eggs and within the female turtle reproductive tract (Al-Bahrey et al., 2009; Al-Bahrey et al., 2011). Indeed, studies of the environmental resistome point to the scale of pollution with antibiotics and/or antibiotic resistance organisms through pathways such as wastewater effluent (Munck et al., 2015; Zhang et al., 2011; Ng et al., 2017), landfill leachate (Zhao et al., 2018), waste from livestock animals (Hong et al., 2013), and hospital waste (Wang et al., 2018). Antibiotic resistance genes and the genetic elements that facilitate their horizontal transfer are now considered pollutants (Gillings et al., 2015). Exposure to antibiotics or antibiotic residues in the environment could disrupt the commensal microbial flora of wildlife, which as described above has key functions in nutrition, immunity, and overall health status (Hanning and Diaz-Sanchez, 2015). Changes to the commensal microbial community following antibiotic administration are well-documented in humans (Palleja et al., 2018) and livestock animals (Looft and Allen, 2012). Therefore, it is reasonable to predict that antibiotic exposure, even at subtherapeutic levels, may alter the functional microbiome of wildlife. This may be especially problematic for vulnerable and endangered wildlife already threatened by habitat degradation, climate change, and invasive species, among other anthropogenic pressures. Taken together these consequences may act synergistically: exposure to antibiotics could disrupt the normal microbial floral, resulting in decreased nutrition and immune function and rendering wildlife more susceptible to colonization by pathogenic organisms.

Lastly, wildlife animals are thought to serve as potential reservoirs for antibiotic resistant organisms or their genes, which could then be transmitted to domestic animals or humans (Vittecoq et al., 2016; Allen et al., 2010). Bovine tuberculosis is just one example of a bacterial pathogen capable of transmission between wildlife, domestic animals, and humans (Fitzgerald and Kaneene, 2013); using a "One Health" framework, it is not difficult to imagine additional scenarios in which closely related bacterial species share resistance traits and move freely between spheres (Robinson et al., 2016). To this point, there is growing evidence for homology between the antibiotic resistance profiles of wildlife and clinical isolates. For example, *E. coli* producing extended-spectrum beta-lactamases (ESBLs) have been isolated from seagulls in Portugal (Simões et al., 2010) and black headed gulls in Sweden (Bonnedahl et al., 2010), and found to mirror the types of ESBLs observed clinically in those areas. While these observations do not confirm direct transfer between wildlife and humans, the existence of multiple transmission pathways – including soil and water contaminated with feces – make this notion more plausible than speculative. In any event, robust studies are needed to definitely establish the link between environmental (including wildlife) and clinical resistance.

State of Knowledge Regarding Antimicrobial Resistance in the Galapagos

Inquiry into AMR in the Galapagos has focused predominately on wildlife, with one study assessing the resistance profiles of *E. coli* isolated from recreational beaches, as mentioned above (Overby et al., 2015). Thaller and colleagues (2010) performed the first survey of AMR in Galapagos wildlife, culturing bacteria from land iguanas on Santa Fe, an uninhabited island, and testing isolates for antibiotic resistance. With the exception of two *Escherichia coli* isolates resistant to nalidixic acid and gentamicin, resistance profiles of the isolates were consistent with the intrinsic resistance profile for the dominant bacterial species identified in the sample. The authors interpreted these two *E. coli* isolates as examples of acquired antibiotic resistance, and

pointed to the land iguanas' proximity to a site accessed by fisherman and film crews as a possible exposure source. Several years later, Wheeler et al. (2012) expanded upon this work by culturing *E. coli* and *Salmonella enterica* from the feces of marine iguanas, land iguanas, and giant tortoises in addition to marine water at sites under various levels of human influence across the archipelago. This group found that antibiotic resistance was more common among *E. coli* (18/59 isolates) than *S. enterica*, with only 5 of 46 isolates showing an intermediate resistance profile and no clinical resistance detected. The majority of the 18 *E. coli* isolates that exhibited antibiotic resistance came from Punta Carola marine water samples, with several also collected from giant tortoises at La Galapaguera on San Cristobal and from land iguanas and marine iguanas on Plaza Sur. Notably, isolates collected from water and iguana samples on Fernandina, as well as those collected from the feces of iguanas on Santa Fe and La Loberia on San Cristobal, were susceptible to all antibiotics tested. Collectively, these results point to increasing antibiotic resistance among *E. coli* isolates with increasing proximity to human settlements, with the highest multidrug resistance observed at Punta Carola, in agreement with later work by Overbey et al. (2015).

More recently, Nieto-Claudin and colleagues (2019) used a culture-independent approach to survey the gut microbiomes of 30 giant tortoises on Santa Cruz for antibiotic resistance genes (ARGs). Specifically, this group tested the DNA extracted from giant tortoise fecal samples against a panel of 21 antibiotic resistance genes. Thirteen of these 21 genes could be detected in at least one sample, and genes for tetracycline resistance (*tetQ* and/or *tetW*) were present in 100% of samples. Genes thought to confer resistance to aminoglycosides, beta-lactams, and quinolines were also common, with genes corresponding to these antibiotic classes detected in 42.9, 32.1, and 28.6% of samples, respectively. Detection of the *mecA* gene, which confers

resistance to beta-lactam antibiotics such as methicillin, maybe be cause for concern even at the low detection level of one sample. While it is important to note that detecting ARGs in wildlife gut microbiomes does not confirm the direction of transmission, the migratory routes of these free-ranging tortoises through agricultural and human-associated areas does suggest the possibility for ARG acquisition from human activities. On the other hand, these ARGs may reflect 'natural' sources of resistance in the environment, pointing to wildlife as reservoirs for antibiotic resistance. In either case, much remains to be discovered about the ecological role of ARGs and the bacteria carrying them in wild animal populations.

Knowledge Gaps: General Objectives and Aims of Proposed Research

The four surveys discussed above provided an important first look into the prevalence of AMR bacteria and genes resident in the gut microbiomes of Galapagos wildlife. The first two surveys (Thaller et al., 2010; Wheeler et al., 2012) employed culture-based methods to isolate enteric bacteria, confirming their resistance profiles with functional culture-based techniques (direct method plating and disk diffusion, respectively). Wheeler and colleagues (2012) further subjected tetracycline resistant isolates to genotypic profiling to confirm the presence of particular *tet* efflux pump components, while Thaller and colleagues (2010) interrogated the genotype of the study's two resistance *E. coli* isolates. The approaches used in these two works identified the bacterial host (enteric bacteria) and confirmed functional resistance through plating methods. Some inquiry into the genotypic basis for the observed resistance phenotype was performed. In contrast, Nieto-Claudin and colleagues (2019) used a purely culture-independent approach to confirm the presence of an array of antibiotic resistance genes (ARGs). This work represents an expanded survey of ARGs over the first two studies, but does not confirm the functional expression of the ARGS or identify the associated host. Further, the animals under study belonged to the same population with presumably similar exposure to human populations.

Inclusion of environmental and human samples would help contextualize the ARGs detected in this giant tortoise population.

Collectively, these works highlight gaps in our understanding of AMR in the Galapagos. First, much remains to be explored regarding ARGs carried by the entire microbiome, not just one or two species of enteric bacteria. Secondly, contemporaneous sampling of wildlife, humans, and the environmental is necessary to elucidate the possible source of particular ARGs. Finally, studying the mobilome in addition to ARGs could provide insight into the direction of their transmission between reservoirs. Using a One Health approach with representation from humans, wildlife animals, and environment, we aimed to fill some of these gaps by characterizing the resistomes and mobilomes of 90 metagenomes. In addition, we developed a novel assay to explore the distribution of class I integron-integrase variants in more than 250 samples. Finally, we contextualized the resistome and mobilome findings through microbial community profiling, comparing 16S rRNA amplicon sequencing and metagenomic taxonomic assignments.

CHAPTER 3: WHAT ANTIBIOTIC RESISTANCE GENES ARE PRESENT? CHARACTERIZATION OF THE RESISTOME

Introduction

The resistome, first conceptualized by D'Costa and colleagues (2006) following the discovery of diverse antibiotic resistance traits in soil microbial communities, encompasses "the collection of all antibiotic resistance genes and their precursors in both pathogenic and non-pathogenic bacteria" (Wright, 2007). This necessarily broad definition parallels the scope of AMR as a global public health crisis: neither antibiotic resistance genes nor their bacterial hosts are confined to geographic or disciplinary borders (**Figure 2.1**). Accordingly, understanding the diversity, abundance, and distribution of antibiotic resistance genes across hosts and habitats has become a critical research priority. While initial efforts to characterize the resistome focused on culture-based methods (Benveniste and Davies, 1973; D'Costa et al., 2006) the expansion of next-generation sequencing technologies paired with the development of ARG databases has provided new tools for understanding the resistome. Annotation of antibiotic resistance genes from shotgun metagenomic sequences has become an attractive strategy to broadly characterize the resistome, capturing potentially many thousands of ARGs at once. These methods have been used to further our understanding of the human gut (Bengtsoon-Palme et al., 2015; Feng et al., 2018; Pärnänen et al., 2018), wastewater (Munck et al., 2015; Zhang et al., 2011), surface water (Ng et al., 2017), and landfill leachate (Zhao et al., 2018) resistomes, among other reservoirs.

Despite these advances in our ability to characterize the resistome, the role of the environment as a source and reservoir for antibiotic resistance genes remains poorly understood.

The Galapagos Islands, where the human population is limited to 3% of the landmass, represents a strategic model system to study how anthropogenic inputs shape the environmental resistome. Specifically, we can compare over the scale of kilometers the resistomes in samples with intense anthropogenic impacts (i.e. marine waters receiving wastewater discharge) with highly protected beaches. Several studies have investigated antibiotic resistance in the Galapagos, particularly among wildlife, using culture-based and targeted PCR approaches (Thaller et al., 2010; Wheeler et al., 2012; Nieto-Claudin et al., 2019; Nieto-Claudin et al., 2021). In the present work, we expanded upon these studies by presenting a One Health, metagenomic survey of antibiotic resistance genes in wastewater, marine water, freshwater, humans, and wildlife animals. As a first aim, we compared three publicly available antibiotic resistance gene annotation approaches to assess their agreement in characterizing the resistomes of 90 metagenomes. Then, we performed an in-depth characterization of human, wildlife, and environmental resistomes, including investigation of interspecies and intraspecies differences in ARG carriage. Finally, we performed a targeted analysis of beta-lactam ARGs detected in the metagenomes and explored possible bacterial hosts using taxonomic assignment of small subunit ribosomal ribonucleic acid sequences in the metagenomes.

Materials and Methods

Site description and study area

The Galapagos Islands are an archipelago located approximately 900 km west of mainland Ecuador. In our study, the majority of samples were collected on and around San Cristobal Island, the eastern-most island of the archipelago and one of only four human-inhabited islands, with a population of around 7,100 residents. **Table S3.1** provides information about the sampling sites included in the study. Specific details regarding sample collection are described below.

Water samples

Samples collected in 2017: 1 L freshwater and marine water samples were collected in autoclave-sterilized polypropylene bottles during June-July of 2017 at the sites described in **Table S3.1**. Samples were transported on ice to the Galapagos Science Center (GSC) and immediately vacuum filtered through 0.45 μM, 47 mm mixed cellulose ester filters (MilliporeSigma, Burlington, MA). The filtration volume for marine water samples was 1 L while for freshwater the volume varied from 500 mL to 1 L depending on the turbidity of the sample. Additionally, influent and effluent samples were collected from the San Cristobal municipal wastewater treatment plant (WWTP), and 1 mL influent and 50 mL effluent was filtered as above. Filters were aseptically transferred to sterile microcentrifuge tubes without buffer and stored at -20°C. Filters were transported on ice to the University of North Carolina, Chapel Hill (UNC), on July 23, 2017 under permit #062-ABG-2017 from the Galapagos Biosecurity Agency (ABG) and permit #PC-03-17/075-2017 from the Direction of the Galapagos National Park (DPNG). DNA was extracted from filters using the Qiagen DNeasy PowerSoil kit according to the manufacturer's instructions and quantified using the Qubit. Samples were subsequently stored at -80°C until sequencing and ddPCR analysis.

Samples collected in 2018: 1 L freshwater and marine water samples were collected in autoclave-sterilized polypropylene bottles during June-July of 2018 at the sites described in **Table S3.1.** Samples were transported on ice to the GSC and immediately vacuum filtered through 0.45 μM, 47 mm mixed cellulose ester filters (MilliporeSigma, Burlington, MA). The filtration volume for marine water samples was 1 L while for freshwater the volume varied from 500 mL to 1 L depending on the turbidity of the sample. During the 2018 sampling period, the San Cristobal WWTP was inoperative and could not be sampled. Instead, wastewater samples

were collected from the central tank where wastewater is collected before being pumped to the WWTP. Samples were collected on two occasions and 50 mL volumes were filtered as above. Filters were aseptically transferred to sterile microcentrifuge tubes without buffer and stored at -20°C until DNA extraction with the Qiagen DNeasy PowerSoil kit within ten days of sampling. Extracts were stored at -20°C at the GSC until exportation on ice to UNC on July 22, 2018 under permit #64-ABG-2018 from the ABG, permit #074-2018 from the DPNG, and permit #050-2018-EXP-CM-FLO-DNB/MA from the Ecuadorian Ministry of Environment (MAE). Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Giant Tortoise (Chelonoidis chathamensis) fecal samples

Samples collected in 2018: Fecal samples were collected from 23 juvenile *Chelonoidis chathamensis* individuals housed at the Galapaguera breeding facility on San Cristobal during March 2018 as part of routine physiological monitoring associated with permit #PC-21-18 issued to G.A. Lewbart. Animals were handled by a licensed veterinarian under the supervision of Galapagos National Park rangers, and all procedures were approved by the Galapagos National Park. Fresh feces were collected immediately following defecation using a sterile spatula and transferred to a sterile microcentrifuge without buffer. Samples were transported on ice to the GSC and frozen at -20°C until DNA extraction in June 2018. DNA was extracted from approximately 0.25 g fecal material using the Qiagen DNeasy PowerSoil kit according to the manufacturer's instructions. DNA extracts were stored at -20°C at the GSC until transport on ice to UNC on July 22, 2018 under permit #65-ABG-2018 from the ABG, permit #073-2018 from the DPNG, and permit #050-2018-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Samples collected in 2019: Additional samples were collected in the same manner from the 12 individuals at the Galapaguera and 6 individuals at Otoy Ranch in March 2019 as part of routine physiological monitoring associated with permit #PC-57-19 issued to G.A. Lewbart. Animals were handled by a licensed veterinarian under the supervision of Galapagos National Park rangers, and all procedures were approved by the Galapagos National Park. Samples were transported on ice to the Galapagos Science Center and frozen at -20°C until DNA extraction in May 2019. DNA was extracted from approximately 0.25 g fecal material using the Qiagen DNeasy PowerSoil kit according to the manufacturer's instructions. DNA extracts were stored at -20°C at the GSC until transport on ice to UNC on May 18, 2019 under permit #047-ABG-2019 from the ABG, permit #064-2018 from the DPNG, and permit #152-2019-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Marine Iguana (Amblyrhynchus cristatus) cloacal swabs

Cloacal swabs were collected from marine iguanas (*Amblyrhynchus cristatus)* at two locations on San Cristobal (La Loberia beach, n=14; and Los Lobos island, n=23) during June-July 2018 as part of routine physiological monitoring associated with permit #PC-59-18 issued to G.A. Lewbart. Animals were handled by a licensed veterinarian and all procedures were approved by the Galapagos National Park. Cloacal swabs were collected using sterile polyester tipped applicators (Puritan Medical Products Company, Guilford, Maine), cut with flame sterilized scissors, and transferred to a sterile microcentrifuge tube without buffer. Samples were stored on ice during the duration of the sampling day and transported to the Galapagos Science Center for storage at -20°C until DNA extraction in mid-July 2018. For DNA extraction, the cotton applicator tip was transferred to the PowerBead tube of the Qiagen DNeasy PowerSoil kit

using flame-sterilized tweezers. The cotton applicator tip was left in the PowerBead tube during vortexing, and was removed following the first centrifugation step. From this point the extraction proceeded in the same manner as with other sample types. DNA extracts were stored at -20°C until transport on ice to UNC on July 22, 2018 under permit #65-ABG-2018 from the ABG, permit #073-2018 from the DPNG, and permit #050-2018-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Sea Turtle (Eretmochelys imbricate and Chelonia mydas) cloacal swabs

Cloacal swabs were collected from Hawksbill (*Eretmochelys imbricate*) and green sea turtles (*Chelonia mydas*) at two beaches on San Cristobal (La Loberia beach, total n=4 including 1 Hawksbill; and Punta Carola beach, total n=5 including 1 Hawksbill) as part of routine physiological monitoring and population tracking associated with permit #PC-27-18 issued to J.P. Muñoz Pérez during July 2018. Individuals were captured by trained swimmers and brought to the shore for approximately 10 minutes for measurement and collection of biological samples associated with permit #PC-27-18. All procedures were approved by the Galapagos National Park and Universidad San Francisco de Quito. Cloacal swabs were collected using sterile polyester tipped applicators (Puritan Medical Products Company, Guilford, Maine), cut with flame sterilized scissors, and transferred to a sterile microcentrifuge tube without buffer. Samples were stored on ice during the duration of the sampling day and transported to the Galapagos Science Center for storage at -20°C until DNA extraction in mid-July 2018. For DNA extraction, the cotton applicator tip was transferred to the PowerBead tube of the Qiagen DNeasy PowerSoil kit using flame-sterilized tweezers. The cotton applicator tip was left in the PowerBead tube during vortexing, and was removed following the first centrifugation step. From this point the extraction proceeded in the same manner as with other sample types. DNA extracts were stored

at -20°C until transport on ice to UNC on July 22, 2018 under permit #65-ABG-2018 from the ABG, permit #073-2018 from the DPNG, and permit #050-2018-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Red-Footed Booby (Sula sula) fecal samples

Fecal samples were collected from red-footed boobies (*Sula sula*) at Punta Pitt on the north eastern side of San Cristobal in July 2018 as part of physiological monitoring associated with permit # TBD issued to S. Cardenas. Samples were collected using a previously reported procedure (Lewbart et al., 2017). All procedures were approved by the Galapagos National Park. Samples were stored on ice during the duration of the sampling day and transported to the Galapagos Science Center for storage at -20°C until DNA extraction in mid-July 2018. DNA was extracted from approximately 0.25 g fecal material using the Qiagen DNeasy PowerSoil kit according to the manufacturer's instructions. DNA extracts were stored at -20°C until transport on ice to UNC on July 22, 2018 under permit #65-ABG-2018 from the ABG, permit #073-2018 from the DPNG, and permit #050-2018-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Land Iguana (Conolophus subcristatus) cloacal swabs

Cloacal swabs were collected from 52 land iguanas across three uninhabited islands including North Seymour (n=20), Plaza Sur (n=11), and Santa Fe (n=21) during July 2018 in association with permit #PC-70-18 issued to G.A. Lewbart. Animals were handled by a licensed veterinarian and all procedures were approved by the Galapagos National Park and Universidad San Francisco de Quito. Cloacal swabs were collected using sterile polyester tipped applicators

(Puritan Medical Products Company, Guilford, Maine), cut with flame sterilized scissors, and transferred to a sterile microcentrifuge tube without buffer. Samples were stored on ice during the duration of the sampling trip and transported to the GSC for storage at -20°C until DNA extraction in April 2019. For DNA extraction, the cotton applicator tip was transferred to the PowerBead tube of the Qiagen DNeasy PowerSoil kit using flame-sterilized tweezers. The cotton applicator tip was left in the PowerBead tube during vortexing, and was removed following the first centrifugation step. From this point the extraction proceeded in the same manner as with other sample types. DNA extracts were stored at -20°C at the GSC until transport on ice to UNC on May 18, 2019 under permit #047-ABG-2019 from the ABG, permit #064-2018 from the DPNG, and permit #152-2019-EXP-CM-FLO-DNB/MA from the MAE. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Sea lion (Zalophus wollebaeki) and Fur Seal (Arctocephalus galapagoensis) fecal samples

Fecal samples were collected from Galapagos sea lions (*Zalophus wollebaeki*) and fur seals (*Arctocephalus galapagoensis*) by D. Páez-Rosas during October 2018 as part of routine population monitoring coordinated by the Galapagos National Park (DPNG) and Galapagos Science Center (GSC). Sample collection took place under permit # TBD issued to D. Páez-Rosas. All procedures were approved by the Galapagos National Park. Samples were stored on ice during the duration of the sampling expedition until arrival at the GSC for storage at -20°C until DNA extraction in April to May of 2019. For DNA extraction, a plastic fecal loop or cotton applicator tip was transferred to the PowerBead tube of the Qiagen DNeasy PowerSoil kit using flame-sterilized tweezers. The loop or cotton applicator tip was left in the PowerBead tube during vortexing, and was removed following the first centrifugation step. From this point the extraction proceeded in the same manner as with other sample types. DNA extracts were stored

at -20°C until transport on ice via courier service to UNC on April 19, 2021 under permit #005-2021 from the DPNG and permit #010-2021-EXP-CM-FAU-DBI/MAAE from the MAE. Upon arrival at UNC, samples were stored at -80°C until sequencing and ddPCR analysis.

Collection of fecal samples from children on San Cristobal

Fecal samples from children under age two living on San Cristobal were generously provided by Dr. Amanda Thompson. Information regarding the study population and fecal sample collection has been reported previously (Thompson et al., 2019). This study was approved by the UNC Institutional Review Board (#15-0863) and the Universidad San Francisco de Quito. Fecal samples were stored at the GSC at -20°C between sample collection in 2016 and transport on ice to UNC on July 12, 2019 under permit #066-ABG-2010. Upon arrival at UNC, DNA was extracted from approximately 0.25 g fecal material using the Qiagen DNeasy PowerSoil kit according to the manufacturer's instructions. DNA extracts were stored at -80°C until sequencing and ddPCR analysis.

Metagenomic sequencing

Shot-gun metagenomic sequencing of 90 metagenomes was performed by the University of North Carolina High-Throughput Sequencing Facility (HTSF) across four sequencing runs. Details for each sequencing run including the samples, library preparation method, and sequencing platform are provided in **Table S3.2.** Changes in the sequencing platform across runs reflect operational changes made at the HTSF between the fall of 2017 and spring of 2019, specifically with the transition from the Illumina HiSeq 4000 platform to the NovaSeq6000SP. The first sequencing run included DNA extracted from 11 water filters collected in 2017. Libraries were prepared using the KapaHyper kit and pooled for sequencing on the Illumina HiSeq 4000 platform, generating 123 GB of 2x150 bp paired end reads with an average of 11.2

GB/sample and 17,139,882 sequence pairs/sample. The second sequencing run included DNA extracted from 23 wildlife samples collected in 2018, including 4 giant tortoises, 8 marine iguanas, 7 sea turtles, and 4 red-footed boobies. Libraries were again prepared using the KapaHyper kit and pooled for sequencing on the Illumina HiSeq 4000 platform, generating 202 GB of 2x150 bp paired end reads with an average of 8.8 GB/sample and 15,416,625sequence pairs/sample. The third run included 54 samples in total: fecal DNA extracts from 24 sea lions, 10 land iguanas, 12 children under age two, and 8 water filters from 2018. Libraries were prepared using the KapaHyper kit, pooled, and sequenced across two lanes of the NovaSeq6000SP generating 669 GB of 2x150 bp paired end reads with an average of 12.4 GB/sample and 18,221,565 sequence pairs/sample. Finally, two additional giant tortoise fecal DNA extracts collected in 2018 were sequenced as part of the "Workshop in a Box" offered at UNC by Illumina Sequencing in January 2019. Libraries were prepared using the Nextera Flex kit, pooled, and sequenced across two lanes of the NovaSeq6000SP generating 287 GB of 2x150bp paired ends reads with an average of 143 GB/sample and 196,476,760 sequence pairs/sample. Sequences from all runs were demultiplexed with bcl2fastq ver 2.20.0 with 1 mismatch allowed.

Bioinformatic analysis and antibiotic resistance gene (ARG) annotation

FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016) were used to check the quality of metagenomic reads before and after trimming. Adapter trimming, quality filtering, and length filtering were simultaneously implemented in bbduk (**https://sourceforge.net/projects/bbmap/**) as part of bbmap/38.82 using right-trimming with 23 bp kmers and discarding reads with quality scores <Q20 and/or length <75 bp. High-quality reads were annotated using three approaches: the Antibiotic Resistance Genes Online Analysis

Platform (ARGs-OAP, v2) (Yin et al., 2018), MegaRes (Doster et al., 2020), and ResFinder (Zankari et al., 2012; Bortolaia et al., 2020). Briefly, the ARGs-OAP aligns reads against sequences in a structured antibiotic resistance gene (SARG) database containing 12,307 ARG sequence variants, corresponding to 1,208 gene subtypes (i.e. *tetM*) and 24 antibiotic classes (i.e. tetracycline). Annotation of sample reads as ARGs required 75% alignment length (37 aa for 150 bp reads); e-value 1e-07, and 80% identity. Results were normalized against 16S rRNA copy number as described by the pipeline developers. Version 2 of the ARGs-OAP was implemented in two stages: Stage one, an initial classification of ARGs with lose alignment parameters was performed locally; and Stage Two, which was performed on a Galaxy webserver.

For annotation approaches two and three, paired metagenomic reads were mapped to either the MegaRes or ResFinder database using Bowtie2 version 2.4.1 (Langmead and Salzberg, 2012) with "highly sensitive" parameters -D 20 -R 3 -N 1 -L 20 -i S,1,0.50. Following the procedure described by Pärnänen et al. 2018, mapped reads were tabulated using SAMtools (Li et al., 2009). ARG counts resulting from MegaRes and ResFinder were normalized to small subunit ribosomal ribonucleic acid (SSU rRNA) counts as tabulated by Metaxa2 version 2.2 (Bengtsson-Palme et al., 2015) in paired-end mode, where SSU rRNA counts were considered as the sum of bacterial and archaeal SSU hits. SSU rRNA classified as *Eukaryota*, *Chloroplast*, *Mitochondria*, or *Uncertain* were excluded.

Statistical analysis

Statistical analyses were performed in R software version 4.0.5. Results from ARG annotation with either the ARGs-OAP, MegaRes, or ResFinder were combined with their respective taxonomy tables and sample metadata to produce three distinct phyloseq objects for analysis in phyloseq (McMurdie and Holmes, 2013) version 1.34.0. Estimation of alpha diversity

with the Simpson and Shannon indices was performed in phyloseq, and differences in alpha diversity between sample groupings were calculated using analysis of variance (ANOVA). In accordance with the procedure described by Pärnänen et al. 2018, p-values were adjusted using Tukey's post hoc test. Group mean MGE sum abundances/16S rRNA were compared using negative binomial generalized linear models (GLMs) in the R software package MASS (Venables and Ripley, 2002). The Bray-Curtis dissimilarity index was calculated for inter-species and intra-species comparisons of ARG composition. The Multivariate Analysis of Variance Using Distance Matrices (ADONIS) in vegan was implemented with 9,999 permutations to assess to extent to which categorical variables (i.e. species or location) explained variation in the distance matrix. Differential abundance of ARGs by sample type was performed using the R package DESeq2 version 1.30.1 (Love et al., 2014). MGE/16S rRNA abundances were transformed to integers by multiplying each observation by 10^5 and rounding the result (Pärnänen et al. 2018). A pseudo-count of 1 was added to all observations to allow for inclusion and log transformation of zero observations.

Table 3.1: *Antibiotic resistance gene (ARG) annotation approaches and associated modifications included in the comparison.* The number of genes retained in each modification and corresponding number of genes detected in the 90 metagenomes in the study is provided.

Database or pipeline	Underlying databases	Description of annotation strategy	Pipeline Modification	Description of modification	# genes	# genes in dataset
ARGs-OAP version 2	ARDB (Liu and Pop, 2009) and CARD (McArthur et al., 2013), NCBI-NR	BLAST and UBLAST-based similarity search with Hidden Markov Models	ARG-OAP.0	None	12,307	7,167
			ARG-OAP.1	"Multidrug" and "unclassified" types excluded	7,170	3,446
MegaRes version 2.0.0	CARD, ResFinder, NCBI's Bacterial Antimicrobial Resistance Reference Gene Database, ARG-ANNOT	Alignment against MegaRes database using bowtie2; 16S rRNA sequences tabulated using Metaxa2	MegaRes.0	None	7,868	3,271
			MegaRes.1	"Biocides", "Metals", and "Multi-compound" types excluded	6,720	2,597
			MegaRes.2	MegaRes.1, plus genes requiring single-nucleotide polymorphism (SNP) confirmation excluded	6,246	2,270
ResFinder version 4.0.0	ResFinder	Alignment against ResFinder database using bowtie2; 16S rRNA sequences tabulated using Metaxa2		None	3,085	1,301

Results

Comparison of ARG annotation approaches

We compared three approaches for annotating antibiotic resistance genes from metagenomic libraries: The Antibiotic Resistance Genes Online Analysis Platform version 2.0 (ARGs-OAP, v2); MegaRes, and ResFinder (**Table 3.1).** ARG-OAP represents a BLAST and UBLAST based approach with Hidden Markov Models in which alignment is performed against the structured antibiotic database (SARG) containing 12,307 ARG sequence variants, corresponding to 1,208 gene subtypes (i.e. *tetM)* and 24 antibiotic classes (i.e. tetracycline). The underlying databases used to construct the SARG implemented in ARGs-OAP include the Antibiotic Resistance Genes Database (ARDB; Lui and Pop, 2009), the Comprehensive Antibiotic Resistance Gene Database (CARD; McArthur et al., 2013), and the National Center for Biotechnology Information protein database (NCBI-NR).

The second and third approaches involved alignment of metagenomic libraries against the MegaRes 2.0 (Doster et al., 2020) and ResFinder version 4.0.0 (Zankari et al., 2012; Bortolai et al., 2020) databases, respectively, using bowtie2 version 2.4.1 (Langmead and Salzberg, 2012) with standardization of ARG counts against 16S rRNA counts as tabulated by Metaxa2 (Bengtsson-Palme et al., 2015). MegaRes combines CARD, ResFinder, NCBI's Bacterial Antimicrobial Resistance Reference Gene Database, and ARG-ANNOT. MegaRes includes 7,868 gene sequence variants, organized into four types (drugs, biocides, multi-compound, and metals) which are further characterized into class (i.e. tetracycline), mechanism (i.e. ribosomal protection protein), and group (i.e. *tetM*). ResFinder, which more exclusively includes acquired resistance genes, contains 3,085 ARG sequence variants organized into 19 types (i.e. tetracycline).

Due to the variation in inclusion criteria and resistance gene definition between these three approaches, with ARG-OAP and MegaRes encompassing significantly more ARG sequence variants than ResFinder, we also considered modified versions of ARG-OAP and MegaRes. Specifically, the ARG.OAP includes several thousand genes corresponding to the "multidrug" resistance class, such as *mdtA* and *TolC*, which are subunits of efflux systems Gram-negative bacteria used to export a range of chemical stressors (Kim et al, 2010; Zgurskaya et al., 2011). While these efflux systems can potentially confer resistance to multiple classes of antibiotics, detection of subunits in metagenomic libraries does not confirm a specific role in or selection for a specific resistance phenotype. Additionally, ARG.OAP includes several hundred genes belonging to "unclassified" resistance classes such as *sdiA*, a quorum sensing regulator (Kanamaru et al., 2000) and the cAMP regulatory protein, a global regulatory molecule involved in numerous cellular processes (Soberón-Chavez et al., 2017). While regulatory molecules such as *sdiA* and cAMP may have some downstream effects on antibiotic resistance in specific cases (Tavío et al., 2010), their primary functions do not relate strictly to antibiotic resistance. Therefore, we considered two versions of ARG-OAP in our analyses: the base version with no modifications, henceforth designated as ARG-OAP.0, and ARG-OAP.1, which excluded all genes corresponding to "multidrug" and "unclassified" resistance classes resulting in the inclusion of 7,170 ARG sequence variants.

Similarly, MegaRes contains broad categories of genes which may have dual roles in antibiotic resistance and tolerating other chemical stressors such as biocides in metals. Therefore, we considered MegaRes.0 with no modifications and MegaRes.1 which included only the 6,720 gene sequence variants corresponding to the "drugs" type. Among the genes retained in MegaRes.1, 474 required confirmation of a single nucleotide polymorphism (SNP) associated

with a resistance phenotype. For example, MEG_1 in the MegaRes database confers resistance to aminoglycoside antibiotics through mutation in the 16S ribosomal subunit **(Figure S3.1).** While the sequence reported in MegaRes does confer resistance to aminoglycoside antibiotics, receiving a hit for a metagenomic sequence mapped against that gene leaves two unknowns: First, whether the mapped region covered the SNP in question, and second, whether coverage was sufficient to accurately call the SNP. Indeed, recommended coverage for accurately calling SNPs from next-generation sequencing data ranges from 5-20x (Nielsen et al., 2011; Song et al., 2016). Shotgun metagenomic sequences originating from mixed microbial communities generally suffer from low coverage (Andreu-Sánchez et al., 2021). While several tools exist to call SNPs from metagenomes with low coverage, this was not the focus of the analysis. Due to this uncertainty surrounding reads mapped to genes requiring SNP confirmation in the MegaRes database, we considered a second modification of the MegaRes database designated MegaRes.2, comprised of 6,246 ARG sequence variants. **Table 3.1** summarizes the number of genes included in each modification as well as the number of genes detected across all 90 metagenomes for each approach.

Total gene observations and sum abundance by ARG annotation approach

The pipelines and their associated modifications differed in both total gene observations and sum abundance of ARGs. The base condition of ARG-OAP generally resulted in the highest total gene observations across all metagenomes (**Figure 3.1a, Figure 3.2**), consistent with its greater database size relative to MegaRes and ResFinder. When genes corresponding to multidrug and unclassified types were excluded for modification ARG-OAP.1, total gene observations more closely mirrored MegaRes.0, the unmodified condition of the MegaRes database. Total gene observations for MegaRes.1 and MegaRes.2 were less than the unmodified

database, but generally greater than ResFinder across all sample types. In terms of ARG sum abundance/16S, MegaRes.0 and MegaRes.1 yielded inflated values relative to the other approaches, particularly for human and land iguana samples (**Figure 3.1b-c, Figure 3.3**).

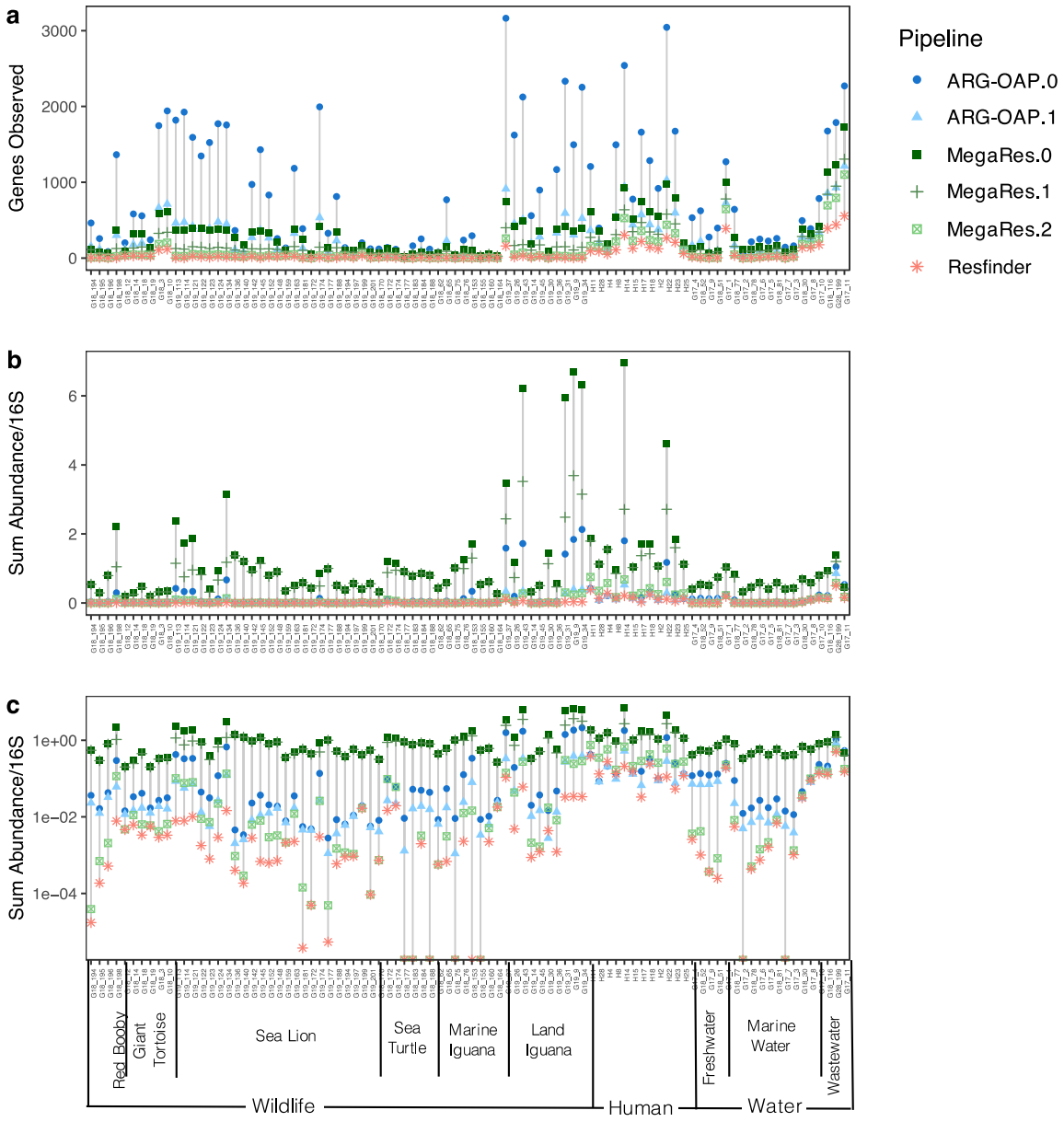

Figure 3.1: *Total ARG observations and ARG sum abundance for three annotation approaches and association modifications.* **a)** Total ARG observations**. b)** ARG sum abundance/16S rRNA, untransformed**. c)** Log-transformed ARG sum abundance/16S rRNA for improved visual separation. Sample numbers are provided and labeled according to wildlife host or sample type.

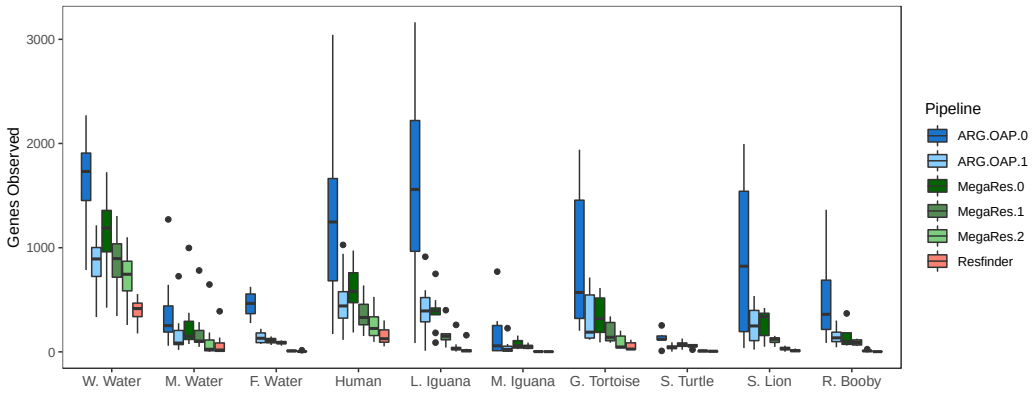

Figure 3.2: *Mean total ARGs observed by sample type for three ARG annotation approaches and associated modifications.*

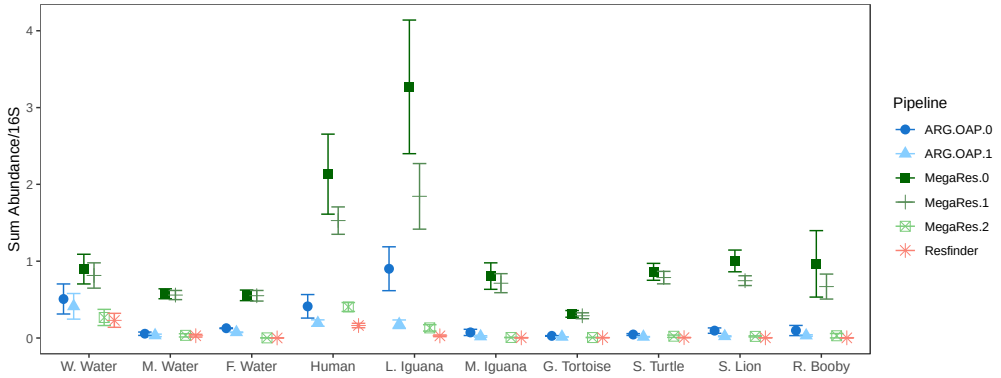

Figure 3.3: *Mean ARG sum abundance/16S rRNA ± SE by sample type for three ARG annotation approaches and associated modifications.*

Comparison of ARG sum ranks by pipeline

We next investigated whether the pipelines, though different in their total observations and ARG sum abundances, would arrive at the same conclusions regarding the relationships between samples and aggregate sample types. **Figure 3.4** shows the ranking of each sample by sum abundance/16S for each pipeline, categorized as wastewater, human, water or wildlife. Across all approaches, wastewater and human samples were generally ranked highly, most notably when using ResFinder for ARG annotation in which wastewater and human samples accounted for 13 of the top 14 ranks (**Figure 3.4**). In contrast, select wildlife samples ranked

higher than wastewater and human samples in ARG-OAP.0, MegaRes.0, and MegaRes.1. The modified ARG-OAP.1 and MegaRes.2 more closely agreed with ResFinder in placing wastewater and human samples among the highest ranks. **Figure S3.2** reports the same rankings with samples colored by subtype. The significance of these relationships was examined using the Kendall rank correlation between sum abundances/16S rRNA, such that pipelines in perfect agreement in ranking samples would yield a tau (τ) coefficient of 1. Conversely, pipelines in perfect disagreement in ranking samples would yield a tau coefficient of -1. The highest correlations were observed between MegaRes.0 and MegaRes.1 (τ=0.87, p<0.0033) followed by ResFinder and MegaRes.2 (τ=0.81, p<0.0033) and ARG-OAP.0 and ARG-OAP.1 (τ=0.78, p<0.0033). Agreement was the lowest between Megares.1 and ResFinder, though the correlation was still positive and significant (τ=0.26, p<0.0033).

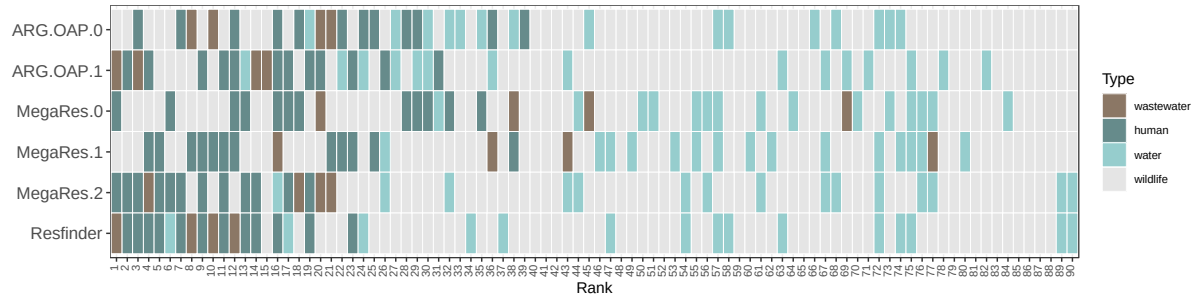

Figure 3.4: *Relative ranking of 90 samples by ARG sum abundance/16S rRNA, with samples categorized as wastewater, human, water, or wildlife.*

Table 3.2: *Kendall rank correlation (τ) of ARG sum abundance/16S rRNA on paired samples between ARG annotation approaches. ARG sums were considered individually for 90 samples.* All correlations are significant after Bonferroni correction for 15 comparisons (p<0.0033).

	ARG-OAP.0	ARG-OAP.1	MegaRes.0	MegaRes.1	MegaRes.2
ARG-OAP.1	0.78	-	-	-	-
MegaRes.0	0.51	0.39	-	-	-
MegaRes.1	0.43	0.37	0.87	-	-
MegaRes.2	0.61	0.66	0.43	0.39	-
Resfinder	0.50	0.60	0.28	0.26	0.81

We then interrogated the agreement between pipelines when metagenomes were aggregated into specific sample types. Mean sum abundance/16S rRNA was calculated for each of ten sample types and ranked (**Figure 3.5).** ARG-OAP.0, MegaRes.0, and MegaRes.1 ranked mean sum abundance/16S rRNA highest among land iguanas, while ARG-OAP.1 and ResFinder ranked wastewater samples highest. Using the Kendall rank correlation (**Table 3.3**), the highest agreement was again recorded between MegaRes.0 and MegaRes.1 (τ=0.78, p<0.0033) followed by the ARG-OAP.0 and ARG-OAP.1 (τ=0.73, p<0.0033). All remaining pairwise comparisons yielded insignificant p-values following Bonferroni correction for multiple comparisons.

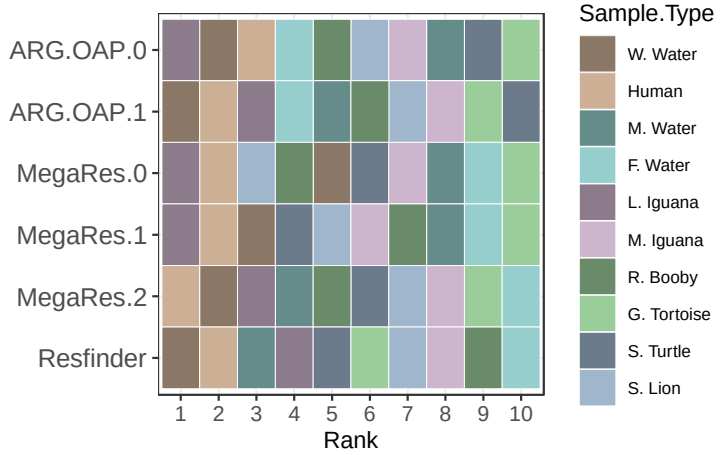

Figure 3.5: *Relative ranking of sample subtypes by mean ARG sum abundance/16S rRNA.*

Table 3.3: *Kendall rank correlation (τ) of mean ARG sum abundance/16S rRNA on paired sample subtypes.* Only tau coefficients noted with an asterisk (*) were associated with significant p-values after Bonferroni correction for 15 comparisons (p<0.0033).

	ARG-OAP.0	ARG-OAP.1	MegaRes.0	MegaRes.1	MegaRes.2
ARG-OAP.1	0.73*	-	-	-	-
MegaRes.0	0.51	0.24	-	-	-
MegaRes.1	0.47	0.20	0.78*	-	-
MegaRes.2	0.38	0.56	0.51	0.56	-
Resfinder	0.11	0.29	0.24	0.47	0.64

Comparison of alpha diversity by ARG annotation approach

The six annotation approaches were further compared on the basis of alpha diversity of ARGs using the Simpson and Shannon Diversity indices (**Figure S3.3a-b**). The Kendall rank correlation was again implemented to compare how pipelines ranked samples according to each alpha diversity metric. Consistent with the pairwise comparisons of ARG sum abundances, the highest agreement in Shannon Diversity (**Table 3.4**) was observed between MegaRes.0 and MegaRes.1 (τ=0.78, p<0.003). The lowest agreement was observed between MegaRes.0 and ResFinder (τ=0.14, p>0.003) though the tau coefficient was not associated with a significant p-value. All other pairwise comparisons with the exception of ARG-OAP.0 and ResFinder as well as MegaRes.1 and ResFinder were significant at alpha = 0.997. When considering Simpson diversity (**Table 3.5**), tau coefficients decreased slightly across all pairwise comparisons, with only 7/15 yielding significant p-values at alpha = 0.997. Again, agreement in ranks of Simpson diversity were highest between MegaRes.0 and MegaRes.1 (τ=0.74, p-value < 0.003). The lowest agreement was observed between ARG-OAP.0 and ResFinder (τ=0.02, p-value > 0.003), implying a near random relationship between the two pipelines.

Table 3.4: *Kendall rank correlation (τ) of ARG diversity based on the Shannon diversity index on paired samples between ARG annotation approaches.* The Shannon diversity index was considered individually for 90 samples. Only tau coefficients noted with an asterisk (*) were associated with significant p-values after Bonferroni correction for 15 comparisons ($p<0.0033$).

	ARG-OAP.0	ARG-OAP.1	MegaRes.0	MegaRes.1	MegaRes.2
ARG-OAP.1	0.76*	-	-	-	-
MegaRes.0	0.47*	0.48*	-	-	-
MegaRes.1	0.30*	0.36*	0.78*	-	-
MegaRes.2	0.36*	0.49*	0.24*	0.23*	-
Resfinder	0.17	0.33*	0.14	0.19	0.63*

Table 3.5: *Kendall rank correlation (τ) of ARG diversity based on the Simpson diversity index on paired samples between ARG annotation approaches.* The Simpson diversity index was considered individually for 90 samples. Only tau coefficients noted with an asterisk (*) were associated with significant p-values after Bonferroni correction for 15 comparisons ($p<0.0033$).

	ARG-OAP.0	ARG-OAP.1	MegaRes.0	MegaRes.1	MegaRes.2
ARG-OAP.1	**0.73***	-	-	-	-
MegaRes.0	**0.38***	**0.40***	-	-	-
MegaRes.1	0.15	**0.21***	**0.74***	-	-
MegaRes.2	0.22	**0.30***	0.070	0.051	-
Resfinder	0.022	0.17	0.030	0.17	**0.42***

Alpha diversity by sample type according to ARG annotation approach

Based on the results of the Kendall rank correlation of ARG sum abundance and alpha diversity between different annotation approaches, ARG-OAP.1, MegaRes2, and ResFinder were selected for additional analyses due to their moderate to high agreement. This analysis aimed to resolve if different pipelines would reveal similar relationships between the diversity of ARGs in different sample types. Simpson and Shannon diversity indices were calculated when broadly classifying samples as human, wastewater, water, or wildlife (**Figure 3.6a-f**). Using data from ARG-OAP.1, no significant differences were observed by sample type for either the Shannon (**Figure 3.6a**) or Simpson (**Figure 3.6b**) diversity index (data not shown). Sample type

differences in ARG alpha diversity in metagenomes annotated with MegaRes.2 were observed for the Shannon (**Figure 3.6c**) but not the Simpson (**Figure 3.6c**) diversity index. Specifically, ARG alpha diversity as measured by the Shannon index was significantly higher in wastewater compared to humans, water, and wildlife samples (**Table S3.3**, adjusted $p<0.05$). Finally, metagenomes annotated with ResFinder exhibited the greatest overall differences in ARG alpha diversity, with wastewater presenting higher Shannon alpha diversity over humans, water, and wildlife (**Table S3.4**, adjusted $p<0.05$.) Using this diversity index, human and water samples also showed significantly higher diversity than wildlife samples (**Table S3.4**, adjusted $p <0.05$). In contrast to ARG-OAP.1 and MegaRes.2 which exhibited no significant differences in Simpson alpha diversity between these four sample types, ResFinder highlighted three significant differences with wildlife, humans, and water each greater than wildlife (**Table S3.5**, adjusted $p<0.05$).

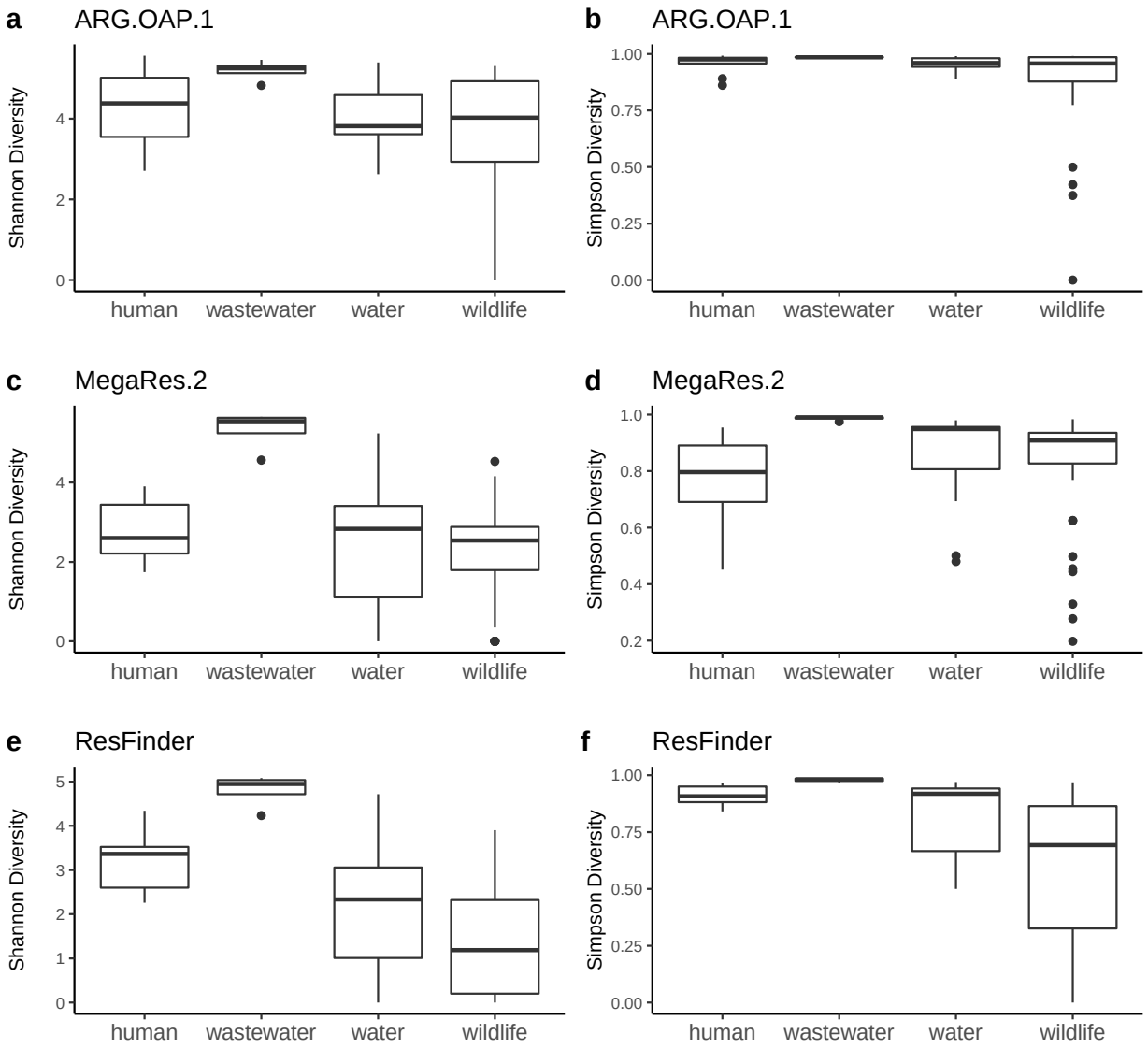

Figure 3.6: *Mean ARG diversity by sample type for three ARG annotation approaches.* **a)** ARG-OAP.1, Shannon index. **b)** ARG-OAP.1, Simpson index. **c)** MegaRes.2, Shannon index. **d)** MegaRes.2, Simpson index. **e)** ResFinder, Shannon index. **f)** ResFinder, Simpson index.

ARG composition by sample type according to annotation approach

Pipeline differences were further interrogated through calculation of the Bray-Curtis dissimilarity index, which can be used to capture compositional differences between groups (Bray and Curtis, 1957). Principal Coordinate Analysis (PCoA) was performed on the Bray-Curtis dissimilarity index matrix (**Figure 3.7a-c**) and the resulting PCoA plots were colored

according to sample type (human, wastewater, water, wildlife). Using the ADONIS test with 9,999 permutations, sample type significantly explained differences in the Bray-Curtis dissimilarity index for all three approaches, though the effect size is modest with R-square values of 0.16, 0.17, and 0.17 for ARG-OAP.1, MegaRes.2, and ResFinder, respectively ($p < 0.001$ in all cases). Ellipses drawn at a 90% confidence level revealed the closest grouping (i.e. least variation) for wastewater samples across all three annotation approaches. Using MegaRes.2 data, humans grouped distinctly from wastewater and wildlife (**Figure 3.7b),** whereas these groups overlapped when using ARG-OAP.1 or ResFinder for annotation. All three approaches pointed to subsets of wildlife samples quite distinct from humans, wastewater, and water in terms of ARG composition.

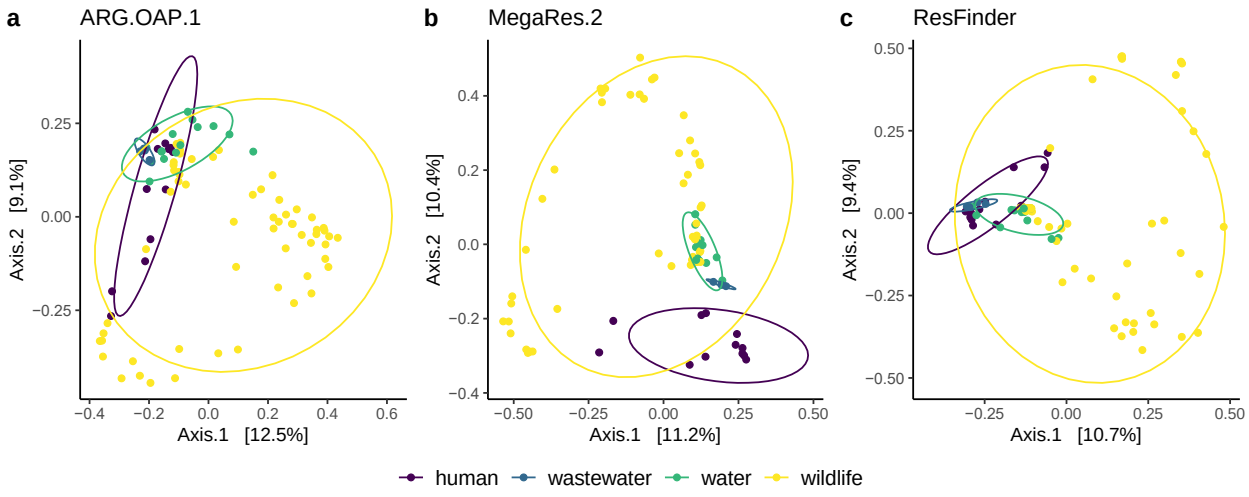

Figure 3.7: *ARG composition by sample type according to annotation approach.* Between sample distances were calculated based on the Bray-Curtis dissimilarity index. **a)** ARG-OAP.1 **b)** MegaRes.2 **c)** ResFinder

Further characterization of resistomes annotated by ARG-OAP.1 and ResFinder

We chose to consider annotations from ARG-OAP.1 and ResFinder for more in-depth characterization of environmental, human, and wildlife resistomes in the Galapagos. This decision was based on differences in their underlying methodologies and scopes: even after

excluding genes conferring resistance to multidrug and unclassified classes, ARG-OAP.1 still contains many more ARGs than ResFinder, a database whose inclusion criteria is limited to acquired resistance genes. We reasoned that annotations from ARG-OAP.1 may better capture background antibiotic resistance which is particularly relevant in the more protected regions of the Galapagos. In contrast, annotations from ResFinder may more specifically point to ARGs of anthropogenic origin. Moreover, we were interested in comparing results from annotation strategies that differed in their underling methodologies. Whereas ARG-OAP.1 uses a BLAST based approach with Hidden Markov Models, the annotations from ResFinder result from mapping with bowtie2. Overall, we aimed to assess the agreement of these approaches in downstream characterization of the resistome and examine their applicability in different sample types and environments.

Characterization of resistomes using ARG-OAP.1

Initial analyses on ARG-OAP.1 annotations were performed by broadly characterizing samples as human, wastewater, water, or wildlife (**Figure 3.8**). Mean sum abundance of ARGs was highest among wastewater samples (4.17E-01 ± 2.50E-01 copies ARG/copies 16S rRNA, negative binomial GLM predicted mean ± SE), followed by humans (1.96E-01 ± 6.78E-02 copies ARG/copies 16S rRNA), wildlife (4.84E-02 ± 7.55E-03 copies ARG/copies 16S rRNA), and water (4.53E-02 ± 1.40E-02 copies ARG/copies 16S rRNA). Group mean differences were significant when comparing humans or wastewater to water and wildlife (**Table S3.6**, $p<0.05$, Tukey's post hoc test), but no additional pairwise comparisons were significant.

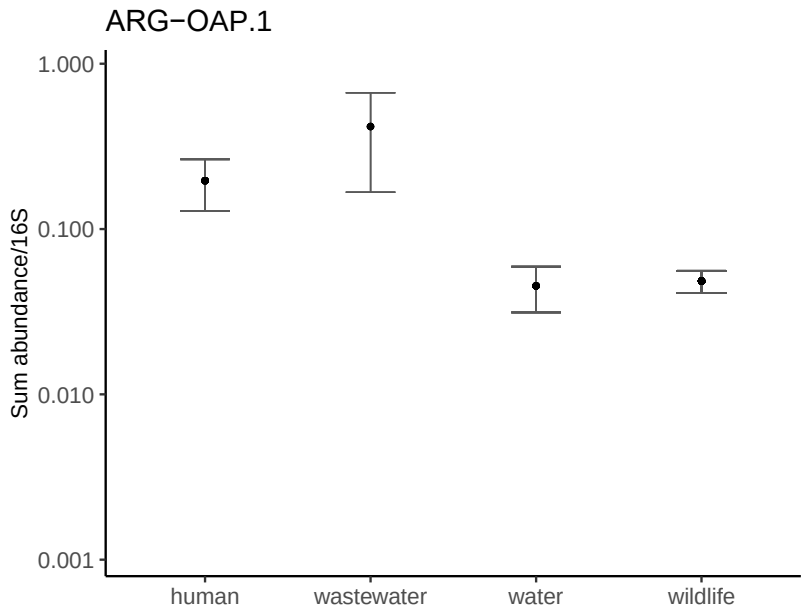

Figure 3.8: *ARG sum abundance/16S rRNA by sample type from annotation with ARG-OAP.1.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

Further characterization of samples into specific subtypes revealed additional differences within these categories **(Figure 3.9, Table 3.6).** Wastewater mean sum abundance of ARGs was significantly higher than all wildlife samples with the exception of land iguanas (**Table S3.7**, $p<0.05$, Tukey's post hoc test). Wastewater samples also had greater mean sum abundance of ARGs compared marine water samples, but not freshwater samples ($p<0.05$, Tukey's post hoc test). Mirroring wastewater samples, ARG sum abundance was significantly greater in humans compared to marine water and all wildlife samples with the exception of land iguanas ($p<0.05$, Tukey's post hoc test). Among wildlife samples, land iguanas exhibited the highest mean sum abundance of ARGs, with significant differences over giant tortoises, marine iguanas, sea turtles, sea lions, and marine water ($p<0.05$, Tukey's post hoc test).

Based on the observation that ARG sum abundances varied considerably among water samples (**Figure 3.10**), we further discretized marine water samples by confirmed or suspected wastewater impact. Both the northern most sampling point at Playa Punta Carola (designated

Carola 1) and Playa Marinero on San Cristobal have documented impacts of wastewater discharge (Overby et al., 2015; Grube et al., 2020), with Carola 1 being the discharge point of treated wastewater effluent and Playa Marinero located immediately in front of the central collection pump for municipal wastewater. The two metagenomes originating from Carola 1 and the two from Playa Marinero were considered impacted while the remaining seven marine metagenomes were designated unimpacted. As shown in **Table 3.6**, Mean ARG sum abundance was an order of magnitude higher in impacted than unimpacted marine waters (1.50E-02 ± 4.24E-03 versus 1.34E-03 ± 2.91E-04 copies ARG/copies 16S rRNA, negative binomial GLM predicted mean ± SE, $p<0.05$, Tukey's post hoc test).

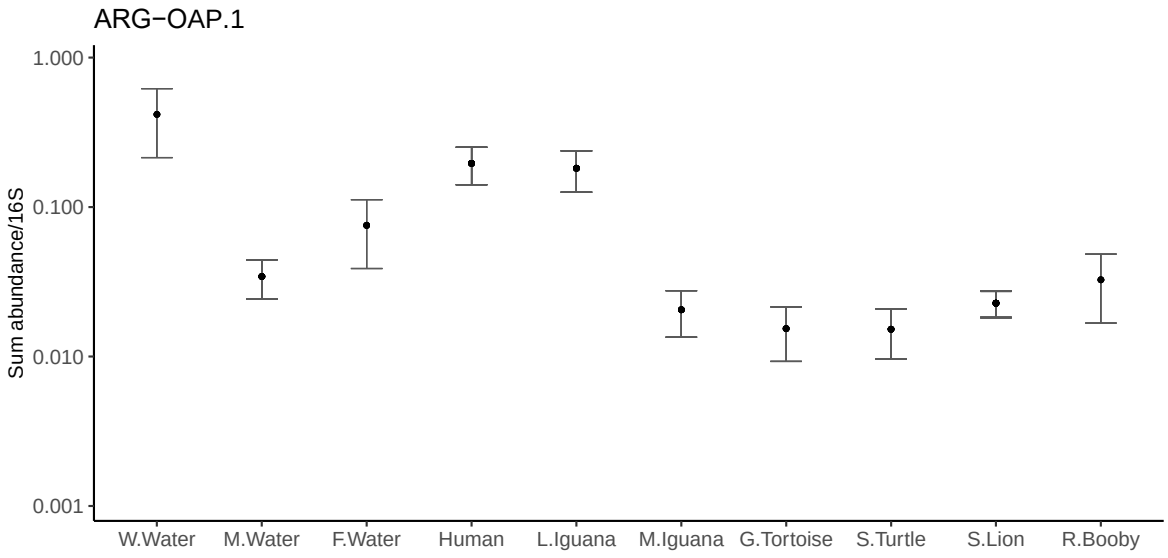

Figure 3.9: *ARG sum abundance/16S rRNA by sample subtype from annotation with ARG-OAP.1.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

Table 3.6: *Mean sum ARG abundance/16S rRNA and most abundant ARGs (type and subtype) by sample type using annotations from ARG-OAP.1.*

Sample Type	n	Mean sum ARG abundance/16S ± SE	Top 3 classes	Mean sum/16S	Top 3 subtypes	Mean sum/16S
Wastewater	4	4.13E-01 ± 1.67E-01	Aminoglycoside	9.80E-02	*sul1*	6.83E-02
			Sulfonamide	9.11E-02	*aadA*	3.81E-02
			Beta-lactam	6.36E-02	*tetC*	3.01E-02
Marine Water	11	3.40E-02 ± 1.70E-02				
Marine impacted	4	8.08E-02 ±3.86E-02	Aminoglycoside	1.75E-02	*sul1*	6.90E-03
			Tetracycline	1.46E-02	*aadA*	5.30E-03
			Beta-lactam	1.40E-02	*bacA*	4.75E-03
Marine background	7	7.20E-03 ± 1.04E-03	Tetracycline	2.18E-03	*bacA*	1.42E-03
			Bacitracin	1.51E-03	*tet34*	1.06E-03
			Vancomycin	7.59E-04	*tet35*	8.49E-04
Freshwater	4	7.47E-02 ± 3.48E-03	Bacitracin	5.52E-02	*bacA*	5.49E-02
			Vancomycin	4.82E-03	*vanR*	3.89E-03
			Fosmidomycin	4.03E-03	*macB*	3.18E-03
Human	12	1.94E-01 ± 4.19E-02	Tetracycline	5.88E-02	*ermX*	4.08E-02
			MLS	5.74E-02	*tetW*	3.78E-02
			Aminoglycoside	2.17E-02	*bacA*	1.01E-02
Sea Lion	24	2.26E-02 ± 6.28E-03	Bacitracin	4.59E-03	*bacA*	4.48E-03
			MLS	4.50E-03	*vanR*	2.32E-03
			Vancomycin	3.24E-03	*ksgA*	2.25E-03
Land Iguana	10	1.80E-01 ± 5.51E-02	MLS	4.95E-02	*ksgA*	2.67E-02
			Fosmidomycin	3.14E-02	*bacA*	2.66E-02
			Kasugamycin	2.67E-02	*macA*	2.46E-02

Giant Tortoise	6	1.52E-02 ± 1.12E-03	MLS	4.83E-03	*vatB*	3.25E-03
			Vancomycin	3.95E-03	*bacA*	1.79E-03
			Tetracycline	2.56E-03	*vanR*	1.77E-03
Sea Turtle	7	1.51E-02 ± 3.50E-03	MLS	4.88E-03	*macB*	4.86E-03
			Tetracycline	2.91E-03	*bacA*	2.13E-03
			Bacitracin	2.20E-03	*tetR*	1.74E-03
Marine Iguana	8	2.03E-02 ± 9.12E-03	Bacitracin	5.20E-03	*bacA*	5.11E-03
			Beta-lactam	3.94E-03	*arnA*	2.41E-03
			MLS	2.67E-03	*class A beta-lactamase*	2.29E-03
Red Footed Booby	4	3.23E-02 ± 1.04E-02	Vancomycin	1.15E-02	*vanS*	6.95E-03
			Bacitracin	5.13E-03	*bacA*	4.95E-03
			MLS	3.38E-03	*vanR*	4.25E-03

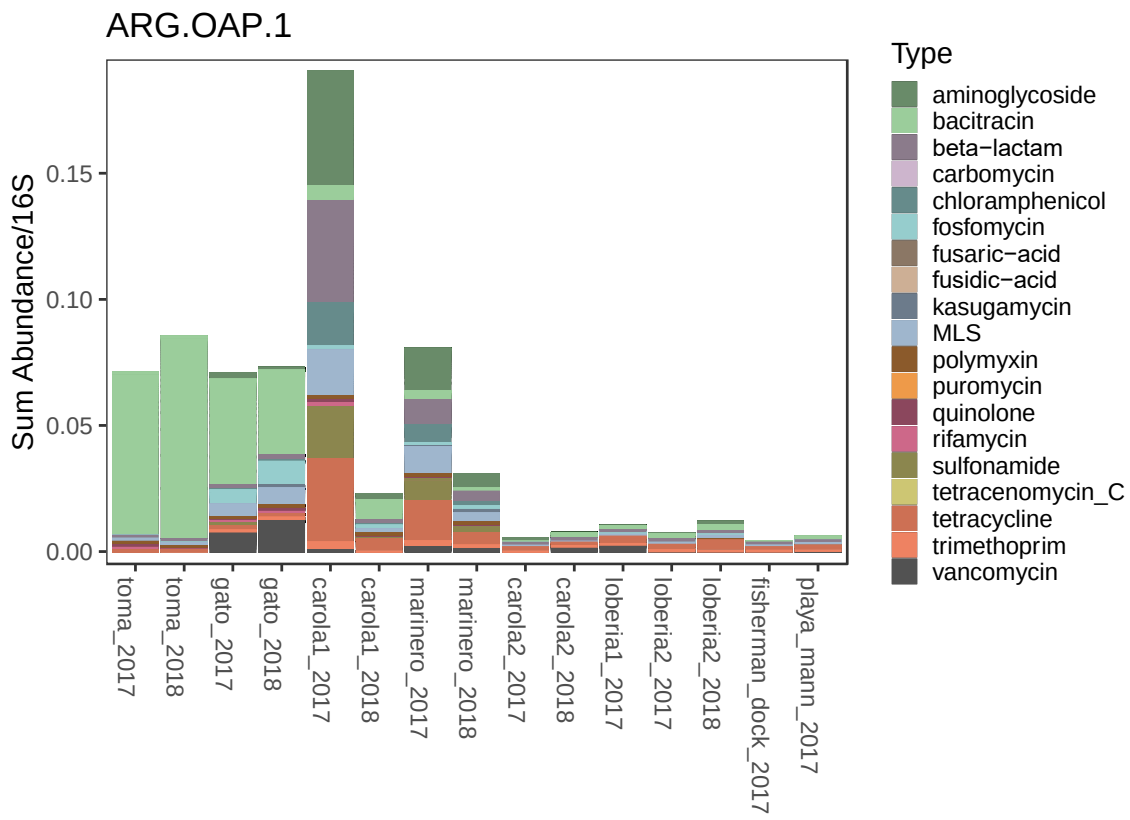

Figure 3.10: *ARG sum abundance/16S rRNA by ARG type for water samples.* Impacts of wastewater discharge are clear at the sites Carola 1, and point of sewage effluent discharge, and Marinero, a beach located immediately in front of the central collection point for municipal sewage.

Metagenomes originating from distinct sample types also showed differences in the most prevalent ARG classes and gene subtypes (**Figure 3.10**, **Table 3.6**). Wastewater samples were dominated by ARGs conferring resistance to aminoglycoside, sulfonamide, and beta-lactam antibiotics, a pattern mirrored in wastewater impacted marine waters where aminoglycoside and beta-lactam ARGs also comprised the top three classes. The sulfonamide resistance gene *sul1* and the aminoglycoside resistance gene *aadA* accounted for the first and second-most abundant subtypes, respectively, in both wastewater and wastewater-impacted marine waters. Marine waters without apparent wastewater discharge were instead dominated by tetracycline, bacitracin, and vancomycin ARGs. Similarly, bacitracin and vancomycin constituted the two

most abundant ARG classes in freshwater samples. The bacitracin resistance gene *bacA* represented the most abundant subtype in both unimpacted marine water and freshwater metagenomes. Like wastewater samples, human fecal metagenomes had aminoglycoside ARGs among the top three most abundant classes, with tetracycline and macrolide-lincosamide-streptogramin (MLS) as the top two most abundant classes. The MLS gene *ermX* accounted for the most abundant gene subtype in human samples followed by the tetracycline resistance gene *tetW*. Bacitracin ARGs represented the most abundant class for both sea lion and marine iguana samples, and was among the top three most abundant classes for all wildlife species except land iguanas. Land iguanas, giant tortoises, and sea turtles all shared MLS as the most abundant ARG class, with land iguanas and sea turtles showing overlap with *macB* in their top three most abundant gene subtypes. Giant tortoises were instead dominated by a different MLS gene, *vatB*, which originates in *Staphylococcus aureus* (Allignet and Solh, 1995). Metagenomes from red-footed boobies were characterized by vancomycin resistance genes, with *vanR* and *VanS* among the top three most abundant subtypes.

We next investigated intra-species differences in ARG sum abundances among sea lions and land iguanas based on location. The twenty-four sea lion fecal metagenomes in this study originate from six distinct sampling locations: Cabo Douglas and Punta Mangle on Fernandina island; Puerto Egas on Santiago island; Champion on Floreana island; and El Malecon and Punta Pitt on San Cristobal island. Among these, San Cristobal and Floreana are the only islands with permanent resident human populations estimated at 7,000 and 150 residents, respectively. While both Fernandina and Santiago islands are uninhabited by humans, Douglas and Punta Mangle on Fernandina represent more remote sites with less tourist traffic compared to Puerto Egas on Santiago.

We originally hypothesized that sea lions from human-inhabited islands, particularly those from El Malecon on San Cristobal where sea lions have close contact with both humans and marine waters receiving wastewater discharge, would harbor more ARGs than individuals from more remote sampling locations. However, no significant differences in group means were observed between sampling locations (**Figure 3.11**) with the exception of individuals from Punta Mangle on Fernandina compared to Punta Pitt on San Cristobal (1.48E-01 ± 9.45E-02 and 1.68E-02 ± 6.81E-03 copies ARG/copies 16S rRNA, respectively, negative binomial GLM predicted means, $p<0.05$, Tukey post hoc test). While differences in ARG sum abundances/16S were generally insignificant between the populations of sea lions, sampling location did significantly explain differences in the Bray-Curtis dissimilarity index (**Figure 3.12**) with an R-square of 0.39 ($p<0.001$, ADONIS with 9,999 permutations).

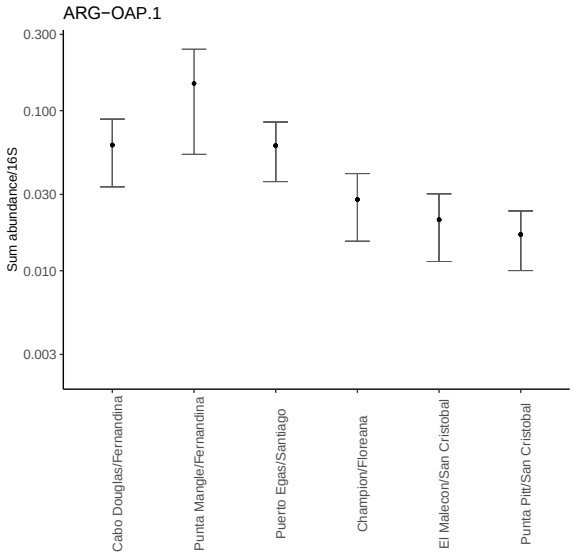

Figure 3.11: *ARG sum abundance/16S rRNA for sea lions by sampling location.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

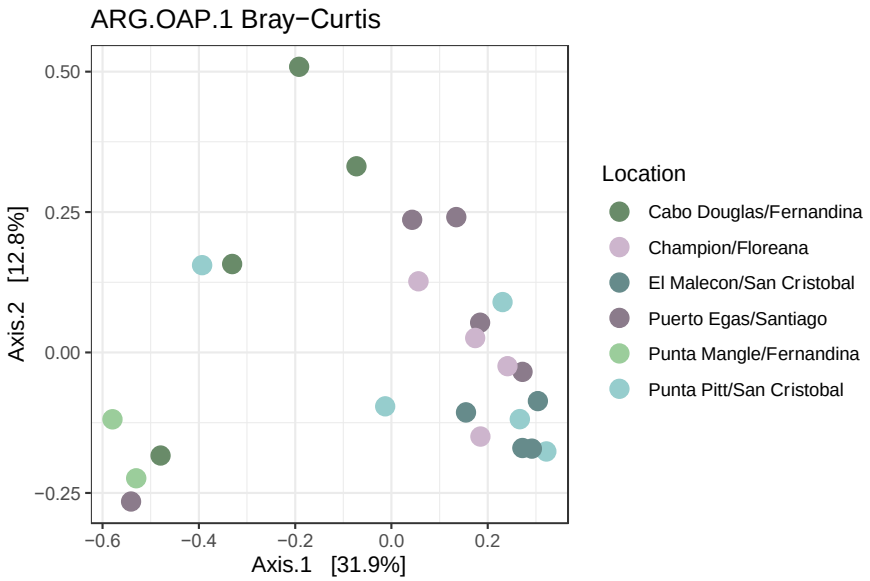

Figure 3.12: *Composition of ARGs in sea lion gut microbiomes by sampling location based on distances calculated using the Bray-Curtis dissimilarity index.*

The ten land iguana fecal metagenomes in this study originate from three sampling locations: North Seymour, Plaza Sur, and Santa Fe islands. All three islands are uninhabited by humans but receive varying levels of tourist traffic, with North Seymour and Plaza Sur situated close to the most populated island of Santa Cruz. When comparing ARG mean sum abundances between sampling locations (**Figure 3.13**), both North Seymour (1.22E-01 ± 3.93E-02 copies ARG/copies 16S rRNA) and Santa Fe (1.78E-01 ± 5.75E-02) were significantly higher than Plaza Sur (5.30E-03 ±1.49E-03, negative binomial GLM predicted means, $p<0.05$, Tukey post hoc test). Likewise, location significantly explained differences in the Bray-Curtis dissimilarity index (**Figure 3.14**) with an R-square of 0.47 ($p<0.01$, ADONIS with 9,999 permutations).

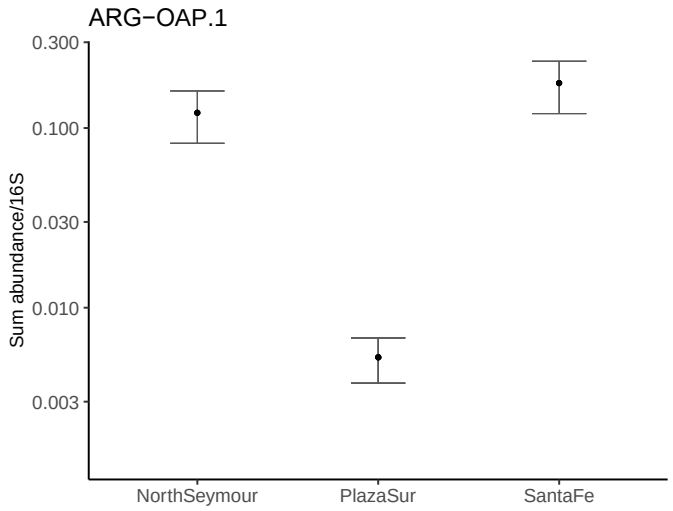

Figure 3.13: *ARG sum abundance/16S rRNA for land iguanas by sampling location.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

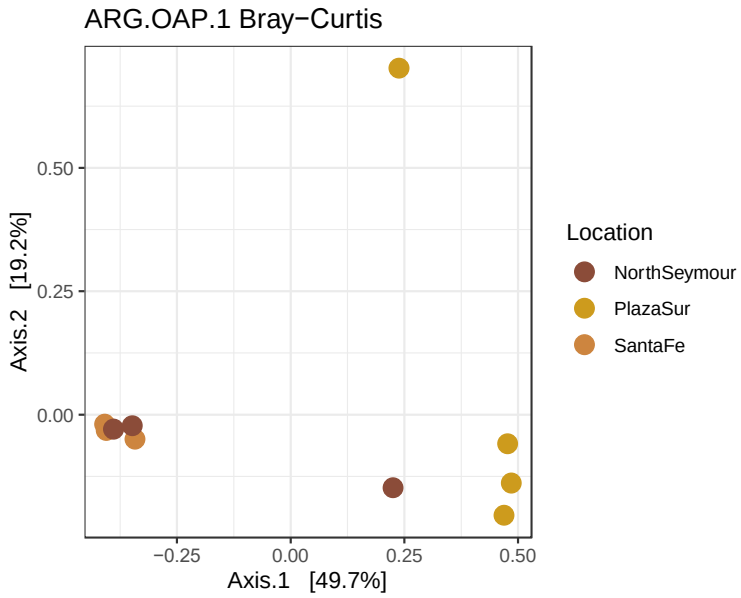

Figure 3.14: *Composition of ARGs in land iguana gut microbiomes by sampling location based on distances calculated using the Bray-Curtis dissimilarity index.*

Finally, we examined if ARG sum abundances differed by mode of delivery and nutrition among the twelve fecal samples from children under age two on San Cristobal island. While mean ARG sum abundance was slightly higher among children born via Caesarean section than

those born vaginally (**Figure S3.4**), the difference was insignificant (p=0.40, Tukey post hoc test). Likewise, mode of delivery did not significantly explain differences in the Bray-Curtis dissimilarity index (R-squared 0.15, p=0.08, ADONIS with 9,999 permutations). Similarly, ARG mean sum abundance was not significantly different by nutrition (**Figure S3.5**, p>0.5 for all pairwise comparisons, Tukey post hoc test), and nutrition did not significantly explain differences in the Bray-Curtis dissimilarity index (R-square 0.21, p=0.27, ADONIS with 9,999 permutations).

While no significant differences were observed for ARG sum abundance by mode of delivery, select ARGs were differentially abundant in children born via Caesarean section compared to those born vaginally **(Figure 3.15).** Specifically, of the 318 genes differentially abundant by birth mode, 297 were significantly more abundant in children born via Caesarean section versus 21 genes more abundant in babies born vaginally (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted p<0.05). The top five ARGs more abundant in children born via Caesarean section include a tetracycline resistance protein; *aadA* and *aph(3)-I*, which both confer resistance to aminoglycosides; a class C beta-lactamase; and *sul2*, a sulphonamide resistance gene (**Table S3.8**). The top five genes differentially abundant in children born vaginally include *aac(6)-I*, *tetW*, and three gene variants for the subtype *msrC*, which confers resistance to macrolide, lincosamide, and streptogramin (MLS) antibiotics.

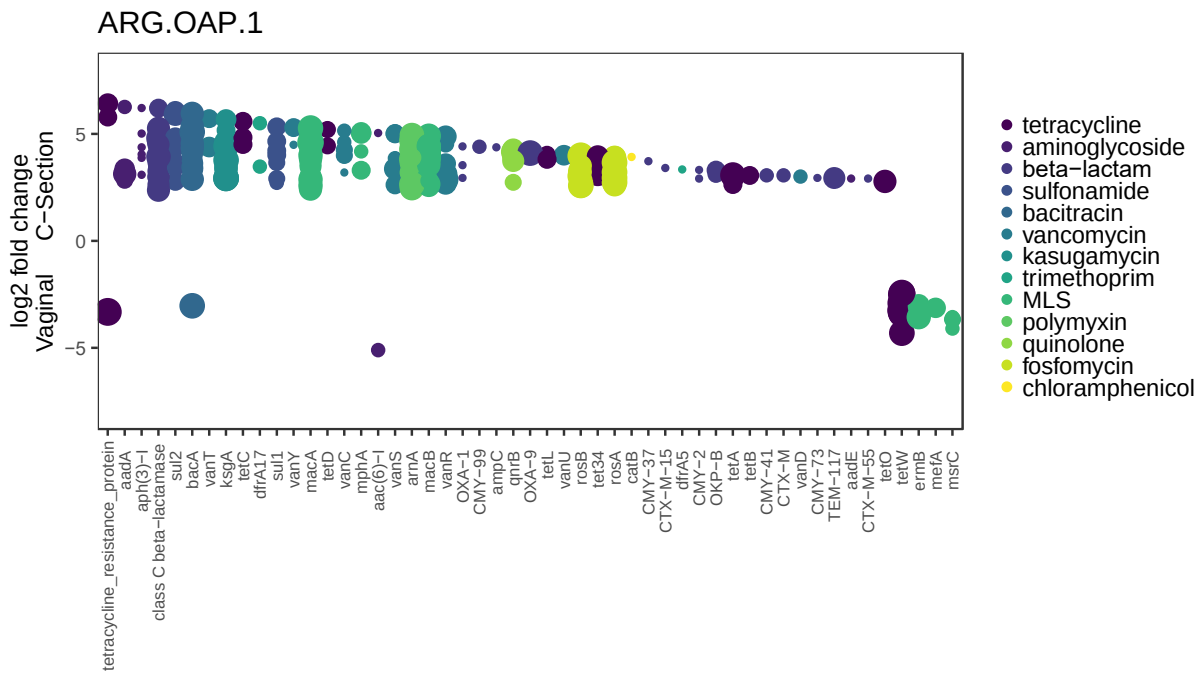

Figure 3.15: *Antibiotic resistance genes differentially abundant between babies born via Caesarean section versus vaginally based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in the Caesarean section group, while ARGs with negative fold changes were differentially abundant in the vaginal birth group. Sizes of data points correspond to the number of individuals (1-12) in which the ARG was detected.

ARGs differentially abundant by wildlife species

We next investigated if specific ARGs were differentially abundant in certain wildlife species. The greatest differences were recorded when comparing land iguanas to all other wildlife species (**Figure 3.16**), where 484 genes were differentially abundant between metagenomes originating from land iguanas compared to other wildlife (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). Of these, 474 genes were significantly higher in land iguanas, while only 12 were more abundant in non-land iguana wildlife samples, principally bacitracin and tetracycline resistance genes. ARGs with the biggest fold change in land iguanas over other wildlife included *macB* and *macA*, conferring resistance

to MLS antibiotics, class A beta-lactamases, and the kasugamycin resistance protein *ksgA* (**Table S3.9**).

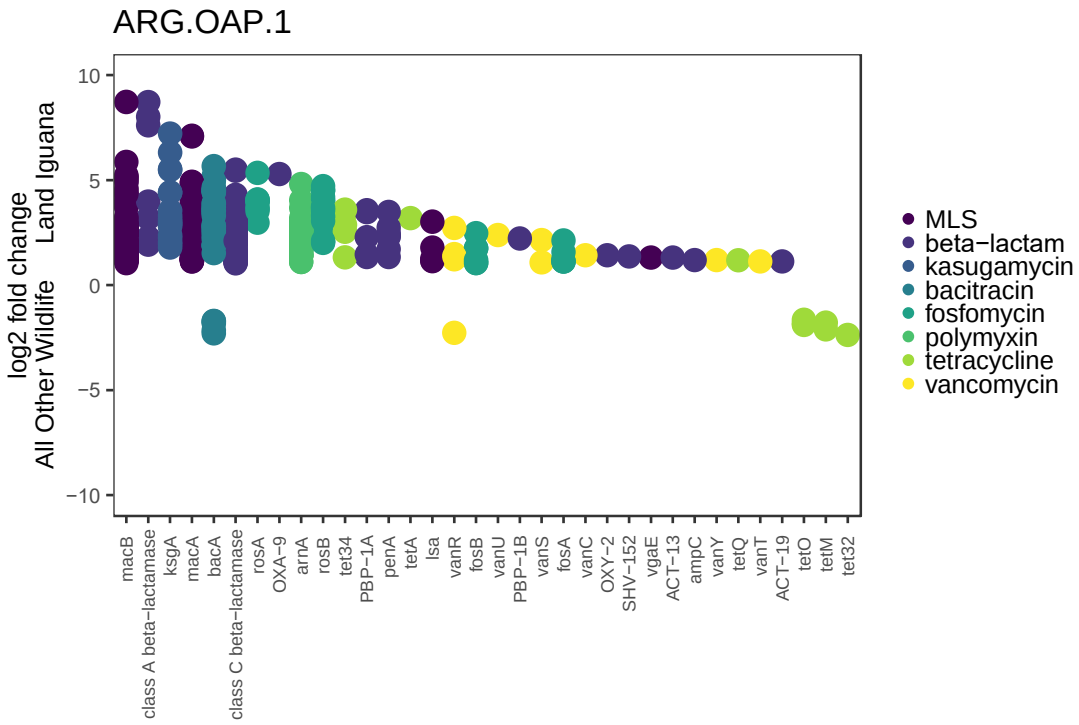

Figure 3.16: *Antibiotic resistance genes differentially abundant between land iguanas and all other wildlife based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in land iguanas, while ARGs with negative fold changes were differentially abundant in other wildlife. Sizes of data points correspond to the number of individuals in which the ARG was detected.

Sea lions were also characterized by select ARGs differentially abundant in their fecal metagenomes compared to other wildlife species (**Figure 3.17**). Of the 252 ARGs differentially abundant between sea lion and non-sea lion wildlife samples, only 57 were greater in sea lions while the remaining 195 ARGs were differentially more abundant in non-sea lion wildlife samples (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). ARGs differentially abundant in sea lion fecal metagenomes included vancomycin resistance

genes *vanR* and *vanS* as well as specific sequence variants of *bacA*, a bacitracin resistance gene (**Table S3.10**).

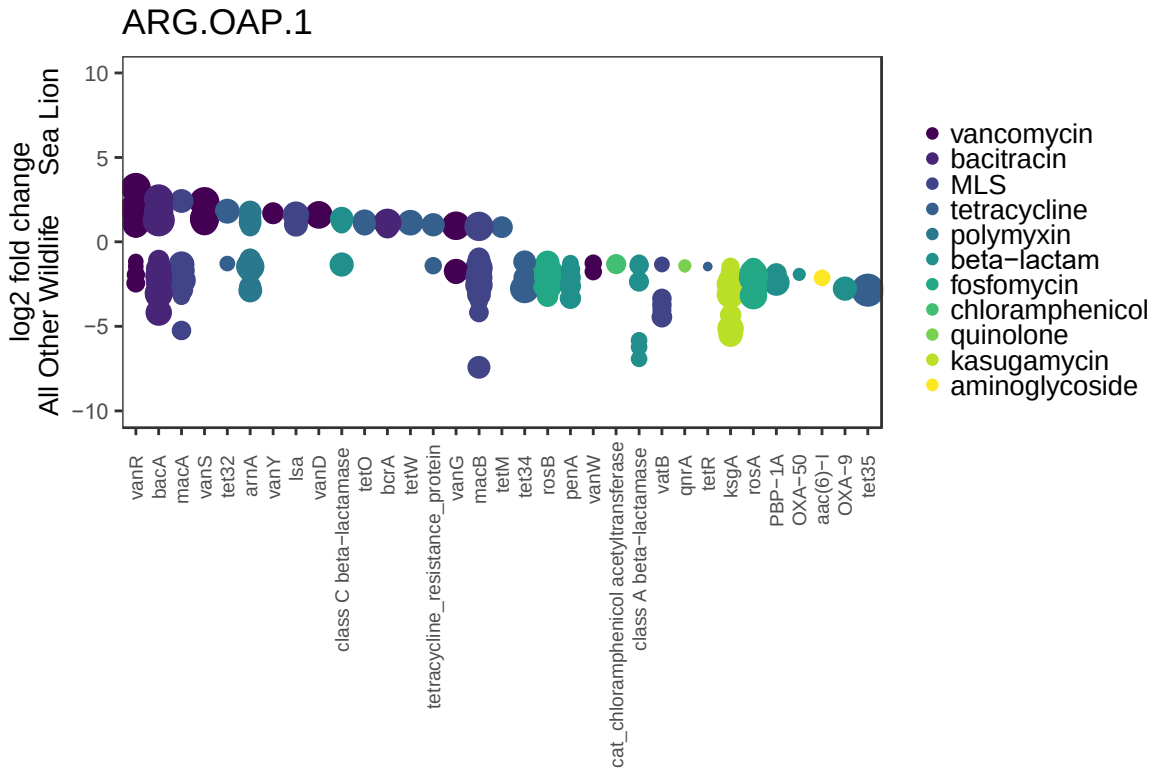

Figure 3.17: *Antibiotic resistance genes differentially abundant between sea lions and all other wildlife based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in sea lions, while ARGs with negative fold changes were differentially abundant in other wildlife. Sizes of data points correspond to the number of individuals in which the ARG was detected.

This analysis was repeated for giant tortoises, sea turtles, marine iguanas, and red-footed boobies against all other wildlife types. Among 142 ARGs differentially abundant in giant tortoise vs non-giant tortoise wildlife samples, 74 were greater in giant tortoises, including variants of MLS resistance gene *vatB* and vancomycin resistance gene *vanG* (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$; **Figure S3.6, Table S3.11).** The 28 ARGs differentially abundant in sea turtle fecal metagenomes compared to non-sea turtle wildlife samples (**Figure S3.7, Table S3.12**) included variants of *macB*, *tetR*, and *bla*$_{\text{OXA-50}}$, a

beta-lactamase found in *Pseudomonas aeruginosa* (Girlich et al., 2004). Only 15 differentially abundant ARGs were observed in marine iguanas compared to non-marine iguana samples, principally sequence variants of the polymyxin resistance gene *arnA* (**Figure S3.8, Table S3.13).** Finally, as the only avian species represented in the dataset, red footed boobies exhibited 70 differentially abundant ARG sequence variants, predominately corresponding to vancomycin resistance genes *vanR* and *vanS* (**Figure S3.9, Table S3.14**).

ARGs differentially abundant by water type

Building on the observation regarding increased ARG sum abundances in wastewater impacted marine environments compared to background marine and freshwater sites, we subsequently examined if specific ARGs were differentially abundant by water type. Among the 148 genes differentially abundant between freshwater and unimpacted marine waters, 130 were more abundant in freshwater (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$), principally variants of the bacitracin resistance gene *bacA* (**Figure 3.18, Table S3.15**). Unimpacted marine waters were distinct from freshwater primarily through higher abundances of tetracycline and vancomycin ARG variants, including *tet34*, *tet35*, *vanR*, and *vanS*. As *tet34* and *tet35* have been described in Vibrio species (Nonaka and Suzuki, 2002; Teo et al., 2002), their detection in marine waters is not surprising.

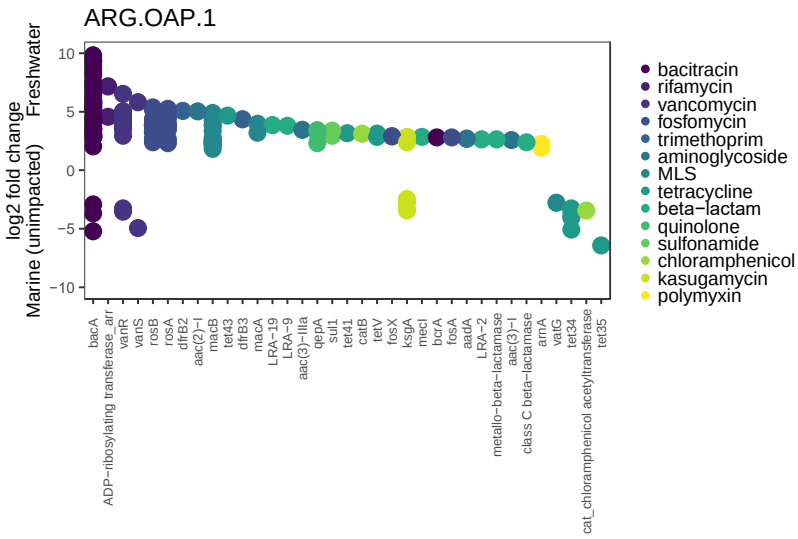

Figure 3.18: *Antibiotic resistance genes differentially abundant between freshwater and unimpacted marine waters based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant freshwater. Sizes of data points correspond to the number of individuals in which the ARG was detected.

When comparing wastewater-impacted marine water with freshwater (**Figure 3.19**), 228 differentially abundant genes were recorded, with 216 significantly more abundant in impacted marine waters (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). The top five genes with the largest fold change over freshwater included two variants of the aminoglycoside resistance gene *aph(3)-I*, the often-mobile chloramphenicol resistance gene *cmlA* (Bischoff et al., 2005), and tetracycline resistance subtypes *tet39* and *tetC* (**Table S3.16**). As with the comparison to unimpacted marine water, freshwater metagenomes were distinct in their abundant of *bacA* variants.

Finally, comparison of wastewater-impacted marine waters to sites without apparent wastewater discharge corroborated the findings above regarding differences in most abundant ARG classes (**Figure 3.20**). Among the 305 ARGs differentially abundant, 300 were significantly higher in wastewater-impacted marine water metagenomes (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). ARGs with the greatest fold

change over unimpacted marine sites included the chloramphenicol transporter *cmlA*, *tetC*, *aac(6)-I*, and variants of *sul1* (**Table S3.17**).

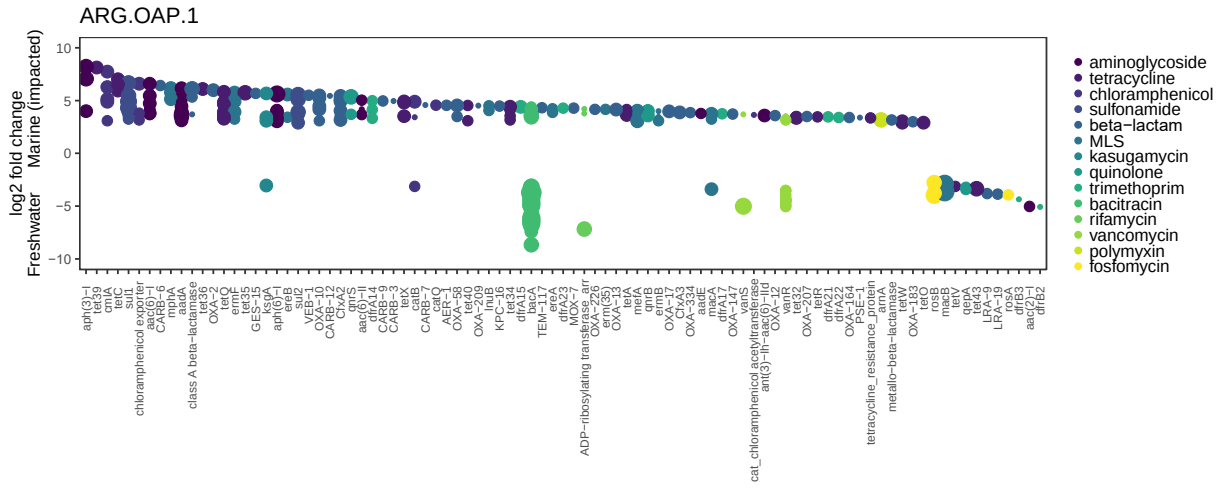

Figure 3.19: *Antibiotic resistance genes differentially abundant between wastewater-impacted marine sites and freshwater based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in impacted marine water. Sizes of data points correspond to the number of individuals in which the ARG was detected.

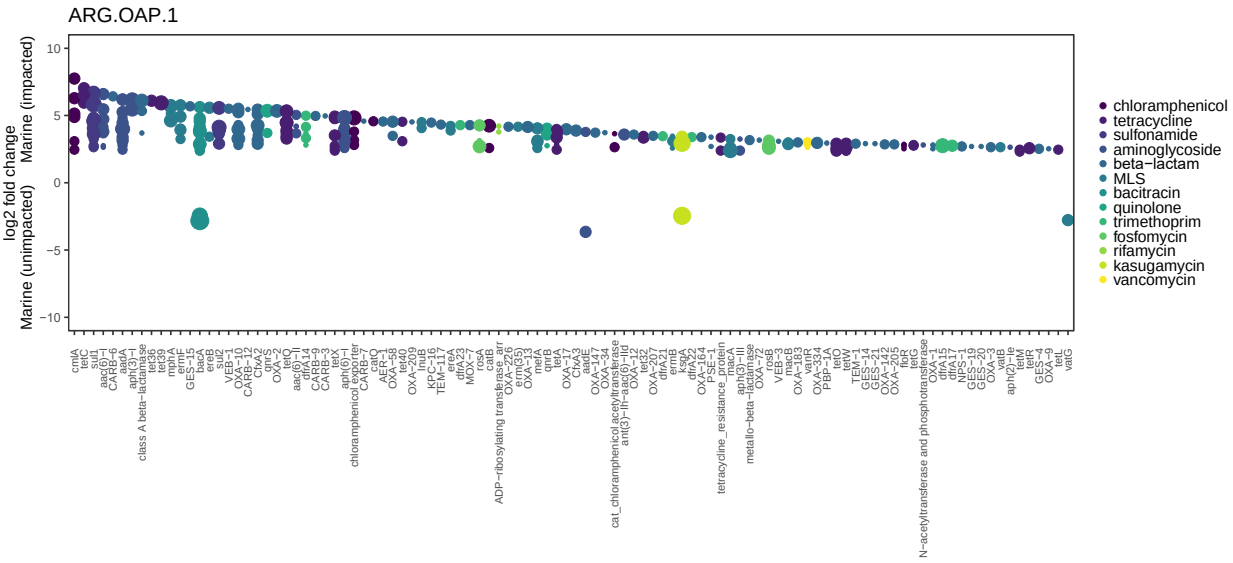

Figure 3.20: *Antibiotic resistance genes differentially abundant between impacted versus unimpacted marine sites based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in impacted marine water. Sizes of data points correspond to the number of individuals in which the ARG was detected.

Characterization of resistomes using ResFinder

As with the ARG-OAP.1 data, initial analyses on the ResFinder data were performed by broadly characterizing samples as human, wastewater, water, or wildlife (**Figure 3.21**). Mean sum abundance of ARGs was again highest among wastewater samples (2.32E-01 ± 2.01E-01 copies ARG/copies 16S rRNA, negative binomial GLM predicted mean ± SE) followed by humans (1.64E-01 ± 8.24E-02), water (2.23E-02 ± 1.00E-02) and wildlife (7.77E-03 ± 1.76E-03). In agreement with data annotated with ARG-OAP.1, group means were significantly different when comparing humans or wastewater to wildlife and water ($p<0.05$, Tukey's post hoc test), but differences between wastewater and humans as well as wildlife and water were not significant (**Table S3.18**).

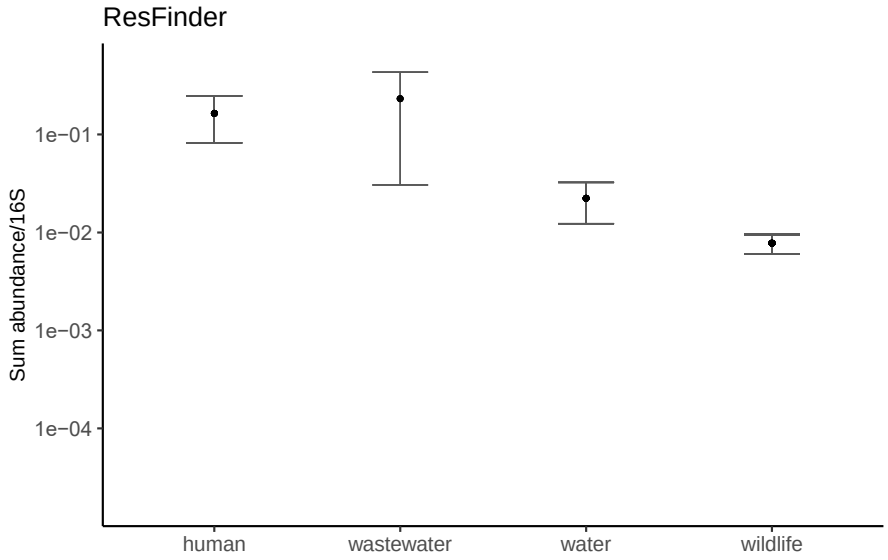

Figure 3.21: *ARG sum abundance/16S rRNA by sample type from annotation with ResFinder.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

When further categorizing samples into subtypes (**Figure 3.22**), additional significant pairwise differences were recorded beyond those associated with ARG-OAP.1 (**Table S3.19**). Wastewater mean sum abundance of ARGs was significantly higher than all wildlife samples

except for land iguanas ($p<0.05$, Tukey's post hoc test). In a pattern opposite ARG annotations from ARG-OAP.1, wastewater samples yielded significantly higher mean sum abundances than freshwater but not marine water samples ($p<0.05$, Tukey's post hoc test). Human samples yielded higher means than freshwater samples and all wildlife species including land iguanas ($p<0.05$, Tukey's post hoc test). Marine waters contained higher mean sum ARGs than giant tortoises, red footed boobies, sea lions and freshwater ($p<0.05$, Tukey's post hoc test). Moreover, the significant differences between land iguanas and other wildlife species were fewer compared to annotations with ARG-OAP.1, including only sea lions and red footed boobies as well as freshwater ($p<0.05$, Tukey's post hoc test).

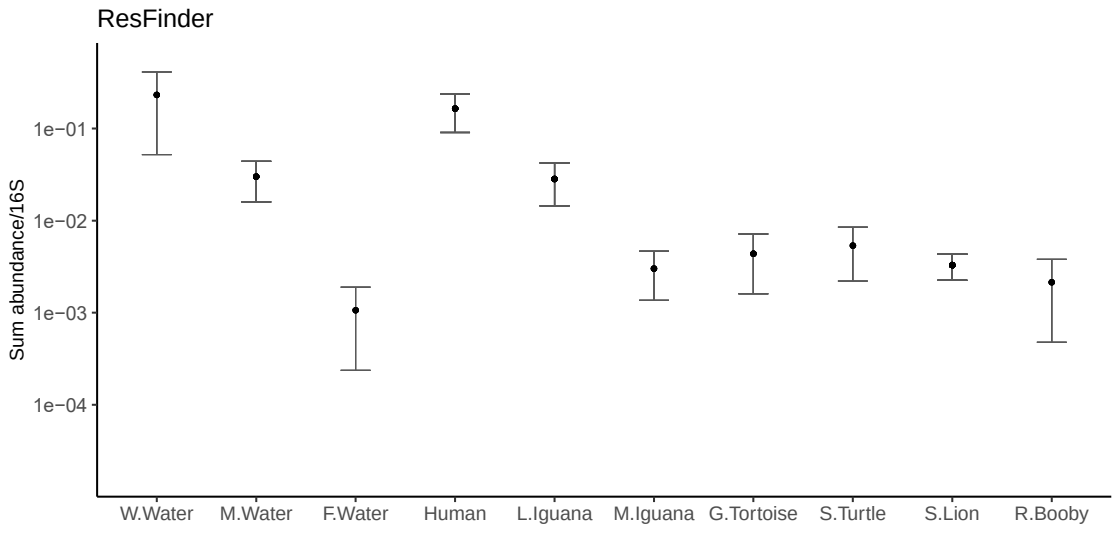

Figure 3.22: *ARG sum abundance/16S rRNA by sample subtype from annotation with ResFinder.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

Annotation with ResFinder also revealed metagenomes from distinct sample types to be characterized by different ARG classes and subtypes (**Table 3.7).** Consistent with data from ARG-OAP.1, wastewater samples featured aminoglycoside ARGs among the top three most abundant classes, in addition to macrolides and tetracyclines. Mirroring wastewater, impacted

marine waters also contained macrolide ARGs as their most abundant ARG class (**Figure 3.23**), followed by tetracyclines and aminoglycosides. Analogous to ARG-OAP.1 annotation, the sulfonamide resistance gene *sul1* was placed among the top three most abundant gene subtypes in both wastewater and marine impacted environments, thought *tetC* occupied the highest rank in wastewater while *msrE* and *mphE* accounted for the top two most abundant gene subtypes in wastewater-impacted marine environments. Interestingly, these two genes also represented the two most abundant gene subtypes in marine sites unimpacted by wastewater, though overall abundance was an order of magnitude lower. Among freshwater samples, quinoline, sulfonamide, and beta-lactam ARGs accounted for the most dominant classes, with *sul2* as the most abundant gene subtype. In a pattern similar to wastewater, human samples also featured tetracycline, macrolide, and aminoglycoside as the top three most abundant classes, with *tetW* and *ermX* representing the two most abundant gene subtypes. Notably, annotation with ARG-OAP.1 also identified *tetW* and *ermX* as the top two gene subtypes in human fecal metagenomes. Land iguanas, marine iguanas, and sea turtles all shared beta-lactam ARGs as the most abundant class, while no other wildlife had beta-lactams in the top three most abundant classes. Phenicols constituted the most abundant ARG class for sea lions and red-footed boobies and was second among giant tortoises and land iguanas. Examining specific gene subtypes, the *E. coli* multidrug efflux pump *mdfA* (Bohn and Bouloc, 1998) accounted for the top ARG among sea lions, land iguanas, and red-footed boobies. Giant tortoise fecal metagenomes were characterized by tetracycline resistance genes *tetO* and *tetW*, while marine iguanas were uniquely dominated by variants of the *cepA* beta-lactamase.

Table 3.7: *Mean sum ARG abundance/16S rRNA and most abundant ARGs (class and subtype) by sample type using annotations from ResFinder.*

Sample Type	n	Mean sum ARG abundance/16S ± SE	Top 3 classes	Mean sum/16S	Top 3 subtypes	Mean sum/16S
Wastewater	4	2.30E-01 ± 9.09E-02	Macrolide	5.32E-02	*tetC*	2.71E-02
			Aminoglycoside	4.68E-02	*sul1*	2.63E-02
			Tetracycline	4.63E-02	*sul2*	1.46E-02
Marine Water	11	2.97E-02 ± 1.79E-02				
Marine impacted	4	7.90E-02 ± 4.08E-02	Macrolide	2.66E-02	*msrE*	1.11E-02
			Tetracycline	1.61E-02	*mphE*	5.87E-03
			Aminoglycoside	1.24E-02	*sul1*	3.79E-03
Marine background	7	1.53E-03 ± 9.22E-04	Macrolide	6.49E-04	*msrE*	3.35E-04
			Tetracycline	4.12E-04	*mphE*	2.18E-04
			Beta-lactam	2.00E-04	*tet39*	2.10E-04
Freshwater	4	1.05E-03 ± 5.38E-04	Quinolone	2.49E-04	*sul2*	1.52E-04
			Sulphonamide	2.37E-04	*oqxA*	1.26E-04
			Beta lactam	1.76E-04	*oqxB*	1.22E-04
Human	12	1.63E-01 ± 2.87E-02	Tetracycline	9.44E-02	*tetW*	4.85E-02
			Macrolide	3.48E-02	*ermX*	3.10E-02
			Aminoglycoside	1.26E-02	*tetOW*	1.72E-02
Sea Lion	24	3.25E-03 ± 9.57E-04	Phenicol	1.85E-03	*mdfA*	1.85E-03
			Tetracycline	1.31E-03	*tetO/32/O*	5.24E-04
			Miscellaneous	3.68E-05	*tetO*	4.43E-04
Land Iguana	10	2.81E-02 ± 1.09E-02	Beta lactam	9.73E-03	*mdfA*	8.82E-03
			Phenicol	8.82E-03	*blaSED-1*	7.13E-03
			Quinolone	7.69E-03	*oqxB*	5.97E-03
Giant Tortoise	6	4.33E-03 ± 5.44E-04	Tetracycline	3.76E-03	*tetO/32/O*	1.06E-03
			Phenicol	2.56E-04	*tetO*	1.00E-03
			Oxazolidinone and Phenicol	1.49E-04	*tetW*	3.36E-04

Sea Turtle	7	5.29E-03 ± 3.12E-03	Beta lactam Tetracycline Aminoglycoside	3.24E-03 7.99E-04 4.85E-04	*blaPAO* *blaOXA-SHE* *tet35*	1.65E-03 7.70E-04 7.14E-04
Marine Iguana	8	2.98E-03 ± 2.18E-03	Beta lactam Aminoglycoside Lincosamide	2.42E-03 3.41E-04 1.82E-04	*cepA* *cepA-49* *cepA-29*	1.13E-03 7.53E-04 3.77E-04
Red Footed Booby	4	2.12E-03 ± 1.88E-03	Phenicol Aminoglycoside Macrolide	1.94E-03 1.77E-04 3.27E-06	*mdfA* *aadA7* *aadA6*	1.94E-03 1.29E-04 1.17E-05

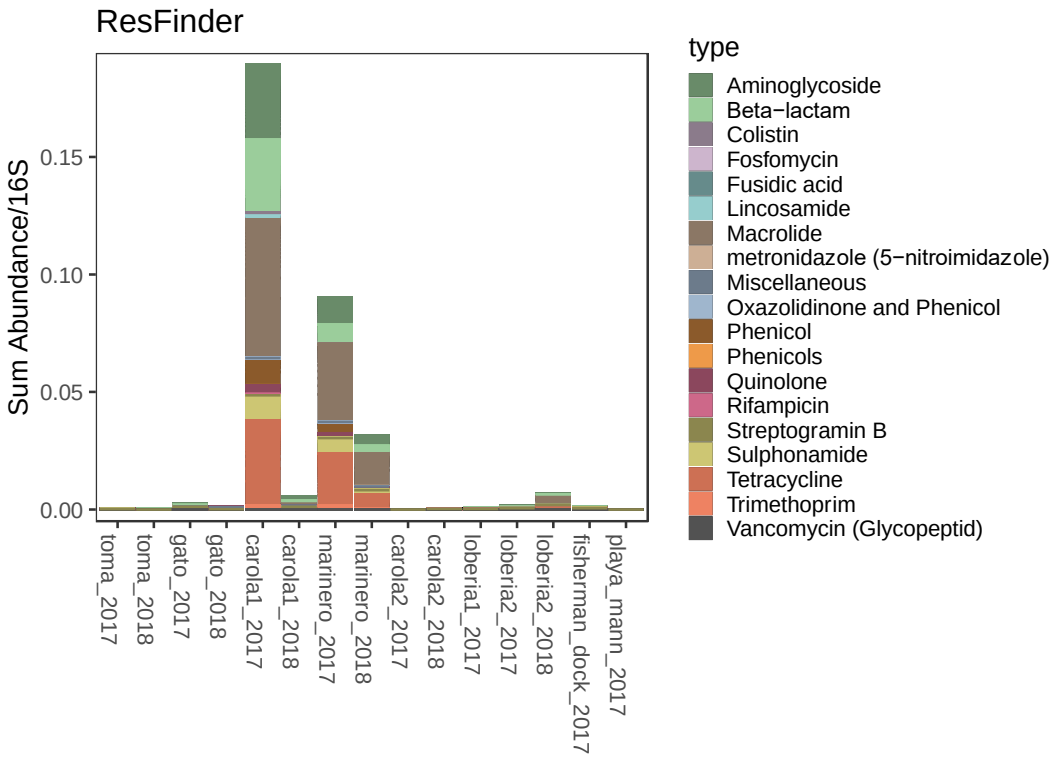

Figure 3.23: *ARG sum abundance/16S rRNA by ARG class for water samples based on ResFinder annotations.* Compared to ARG-OAP.1, only wastewater-impacted marine sites have appreciable sums of ARGs.

We subsequently explored intra-species differences in ARG sum abundance in sea lion and land iguana fecal metagenomes annotated with ResFinder. While no significant differences were observed in ARG sum abundance between sea lions from different sampling locations (**Figure S3.10**, negative binomial GLM predicted means, $p<0.05$, Tukey post hoc test), location was an important explanatory variable in the Bray-Curtis dissimilarity index, with an R-squared of 0.41 and $p<0.001$ (**Figure 3.24**, ADONIS with 9,999 permutations).

Annotation with ResFinder corroborated the results from ARG-OAP.1 regarding ARG mean sum abundance in land iguanas from distinct islands (**Figure 25**), with populations from North Seymour and Santa Fe demonstrating significantly greater means compared to Plaza Sur (2.93E-02 ± 1.19E-02, 1.71E-02 ± 6.94E-03, and 9.83E-04 ± 3.50E-04 copies ARG/copies 16S

rRNA, respectively, negative binomial GLM predicted mean ± SE, $p<0.05$, Tukey's post hoc test). Likewise, location was a significant explanatory variable in the Bray-Curtis dissimilarity index, with an R-squared of 0.42 and p=0.0061 (**Figure 3.26**, ADONIS with 9,999 permutations).

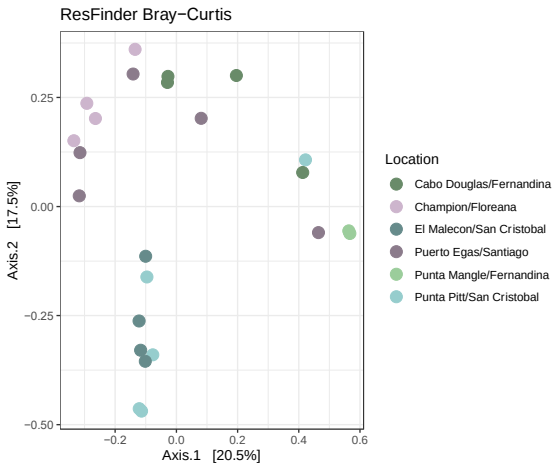

Figure 3.24: *Composition of ResFinder ARGs in sea lion gut microbiomes by sampling location based on distances calculated using the Bray-Curtis dissimilarity index.*

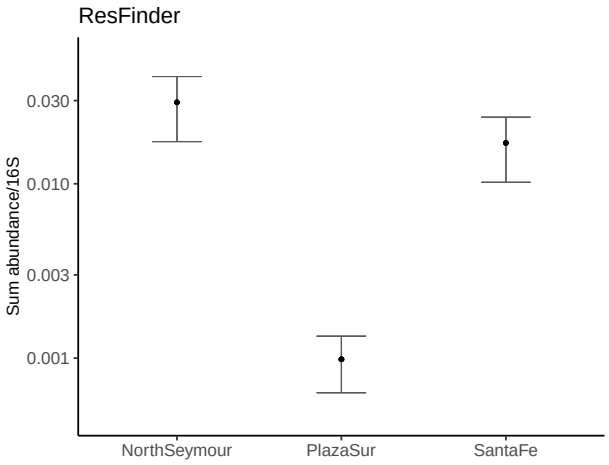

Figure 3.25: *ResFinder ARG sum abundance/16S rRNA for land iguanas by sampling location.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

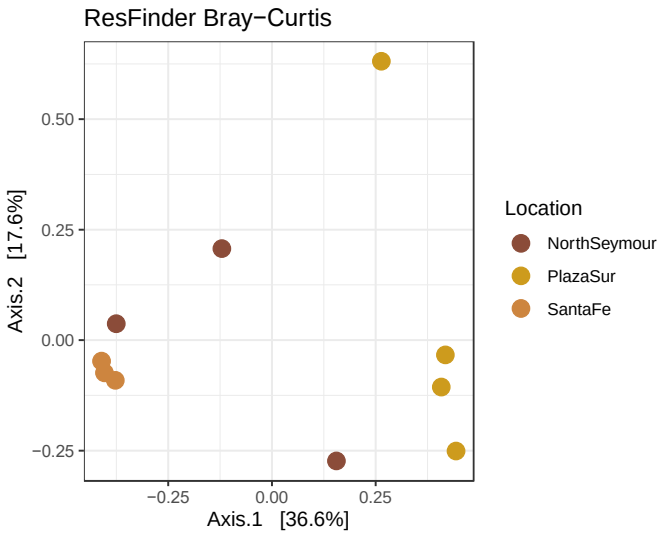

Figure 3.26: *Composition of ResFinder ARGs in land iguana gut microbiomes by sampling location based on distances calculated using the Bray-Curtis dissimilarity index.*

Finally, we asked if metagenomes from children under age two annotated with ResFinder differed in ARG sum abundance my mode of delivery and nutrition status. In a pattern opposite the results of ARG-OAP.1, babies born vaginally had slightly higher ARG sum abundance, though the difference was not statistically significant (**Figure S3.11**, negative binomial GLM predicted mean, p=0.322, Tukey's post hoc test). Similarly, no significant differences were observed in ARG sum abundance by nutrition (all pairwise p>0.05, Tukey's post hoc test). In terms of ARG composition, neither mode of delivery nor nutrition status significantly explained differences in the Bray-Curtis dissimilarity index, with respective effect sizes of 0.14 (p=0.08) and 0.23 (p=0.17, ADONIS with 9,999 permutations for both tests). Though ARG sum abundances were not significantly different by mode of delivery, select ARGs were differentially abundant in babies born via Caesarean section compared to those delivered vaginally (**Figure 3.27**). Specifically, of the 63 genes differentially abundant by mode of delivery, 54 were greater in the Caesarean section group (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted p-value <0.05). The top five genes with the greatest fold change over babies born

vaginally included *vanC*, *tetB*, *aadA*, *aph(3')-Ia*, and *mphA* (**Table S3.20**). Notably, the same analysis performed using data annotated by ARG-OAP.1 also identified *aadA* and *aph(3)-I* among the top five genes differentially abundant in children born via Caesarean section. Conversely, ARGs differentially abundant in the vaginal birth group included *msrC*, *aac(6')-Ii*, *ermB*, and *tetW*. ARG-OAP.1 corroborated this observation with gene variants for *aac(6)-I*, *tetW*, and *msrC* differentially abundant in babies born vaginally compared to those born via Caesarean section.

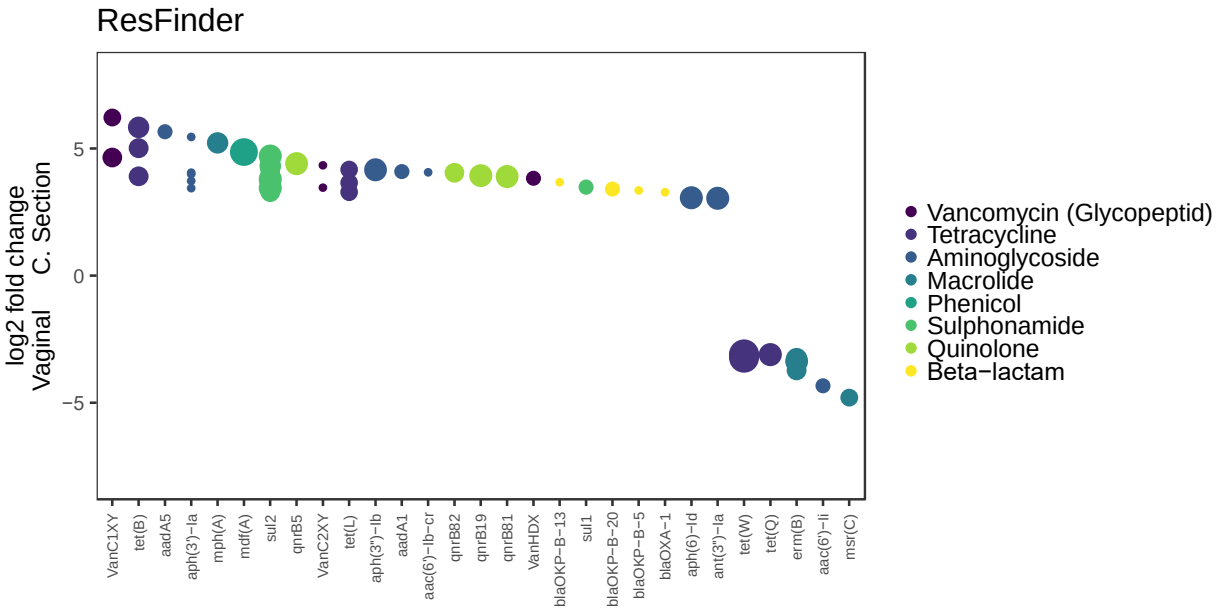

Figure 3.27: *Antibiotic resistance genes differentially abundant between babies born via Caesarean section versus vaginally based on annotations from ResFinder.* ARGs with positive fold changes were differentially abundant in the Caesarean section group, while ARGs with negative fold changes were differentially abundant in the vaginal birth group. Sizes of data points correspond to the number of individuals (1-12) in which the ARG was detected.

ARGs differentially abundant by wildlife species

Wildlife fecal metagenomes annotated with ResFinder had fewer overall differentially abundant ARGs. The greatest number of differences were observed when comparing giant tortoises to all other wildlife species (**Figure 3.28**), with a total of 31 differentially abundant

gene sequence variants (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). Tetracycline variants accounted for the many of the 29 genes differentially abundant in giant tortoises, including *tetO, tetW, tet32,* and *tet39* (**Table S3.21**). This agrees with the observation regarding tetracycline as the most abundant ARG class in giant tortoises when annotating with ResFinder (**Table 3.7).** Annotation with ARG-OAP.1 instead pointed to *vatB* and *vanG* variants as the ARGs most differentially abundant in giant tortoises compared to other wildlife, though *tet32* was also among the top ten ARGs by magnitude of fold change (**Table S3.11**). Comparison of land iguanas with non-land iguana wildlife species revealed 25 differentially abundant genes (**Figure 3.29**), with 20 of these higher in land iguanas (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted $p<0.05$). This included variants of the beta-lactam ARGs $bla_{\text{SED-1}}$ and $bla_{\text{MAL-1}}$ as well as quinolone resistance variants *oxqA* and *oxqB* among the top four differentially abundant ARGs (**Table S3.22**). Differentially abundant genes were fewer among sea turtles and marine iguanas versus other wildlife species, with only three ARG sequence variants differentially abundant in sea turtles and two differentially abundant in marine iguanas compared to other wildlife species (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted p-value <0.05). Sea turtles were uniquely characterized by sequence variants of bla_{PAO} and $bla_{\text{OXA-SHE}}$, ARGs conferring resistance to beta-lactam antibiotics, as well as quinolone ARG *qnrA5* (**Table S3.23**). A sequence variant of the lincosamide ARG *lnuP* and aminoglycoside *aac(6')-Iaa* accounted for the two ARGs differentially abundant in marine iguanas compared to other wildlife species (**Table S3.24**). No ARGs were differentially abundant in sea lions compared to non-sea lion wildlife species, and only one ARG was higher in red-footed boobies, though it was only detected in one individual (data not shown.)

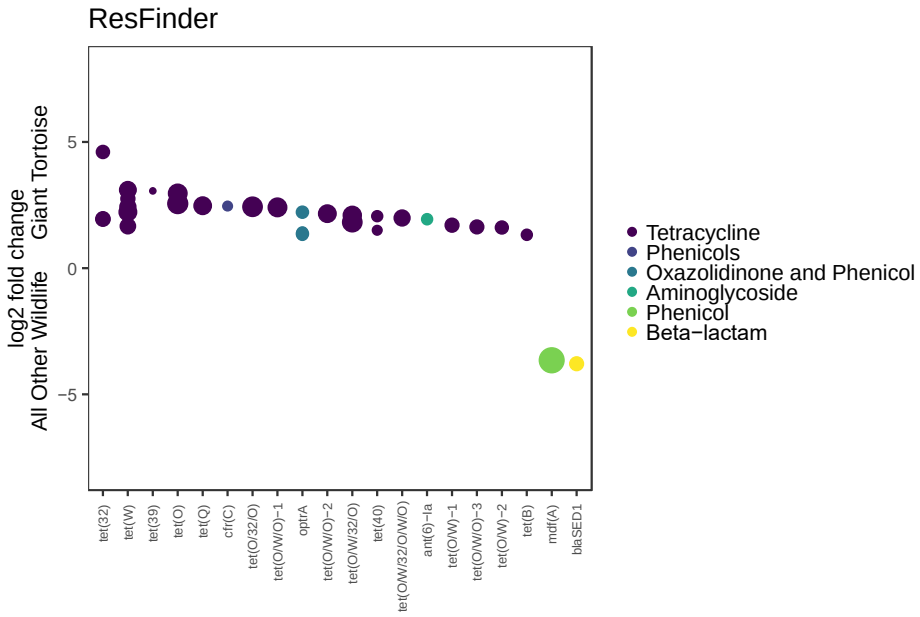

Figure 3.28: *Antibiotic resistance genes differentially abundant between giant tortoises and all other wildlife based on annotations from ResFinder.* ARGs with positive fold changes were differentially abundant in giant tortoises compared to other wildlife. Sizes of data points correspond to the number of individuals in which the ARG was detected.

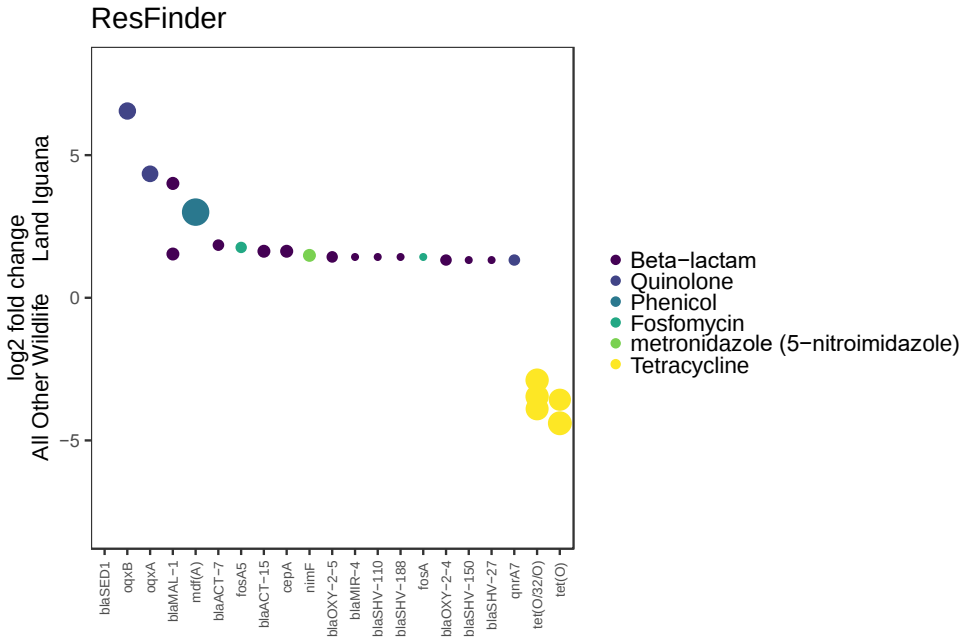

Figure 3.29: *Antibiotic resistance genes differentially abundant between land iguanas and all other wildlife based on annotations from ResFinder.* ARGs with positive fold changes were differentially abundant in land iguanas compared to other wildlife. Sizes of data points correspond to the number of individuals in which the ARG was detected.

ARGs differentially abundant by water type

Metagenomes from wastewater-impacted marine sites exhibited differentially abundant ARGs compared to freshwater and marine background sites when annotating with ResFinder. Specifically, all 161 gene sequence variants differentially abundant in marine impacted and freshwater metagenomes were higher in the former (**Figure 3.30**), with ARGs *mphE*, *tet39*, *cmlA*, *tetQ*, and *tetC* accounting for the largest fold changes (**Table S3.25**). The same analysis using ARG-OAP.1 data also highlighted *cmlA*, *tetC*, and *tet39* among the top five ARGs differentially abundant in impacted marine environments compared to freshwater. Even more differentially abundant ARG sequence variants were observed between marine impacted versus unimpacted sites, with all 269 differentially abundant genes greater in the wastewater impacted marine waters (**Figure 3.31**). Phenicol ARG *cmlA* constituted the largest fold change, while variants of tetracycline, macrolide, and aminoglycoside ARGs were also among the top ten ARGs (**Table S3.26**). This finding agrees with the results of the same analysis using data annotated with ARG-OAP.1, where the most differentially abundant genes in wastewater impacted marine waters over background sites included *cmlA*, *tetC*, *aac(6)-I.* In contrast to results from ARG-OAP.1, no ARGs were found to be differentially abundant between freshwater and unimpacted marine water metagenomes when annotating with ResFinder (data not shown).

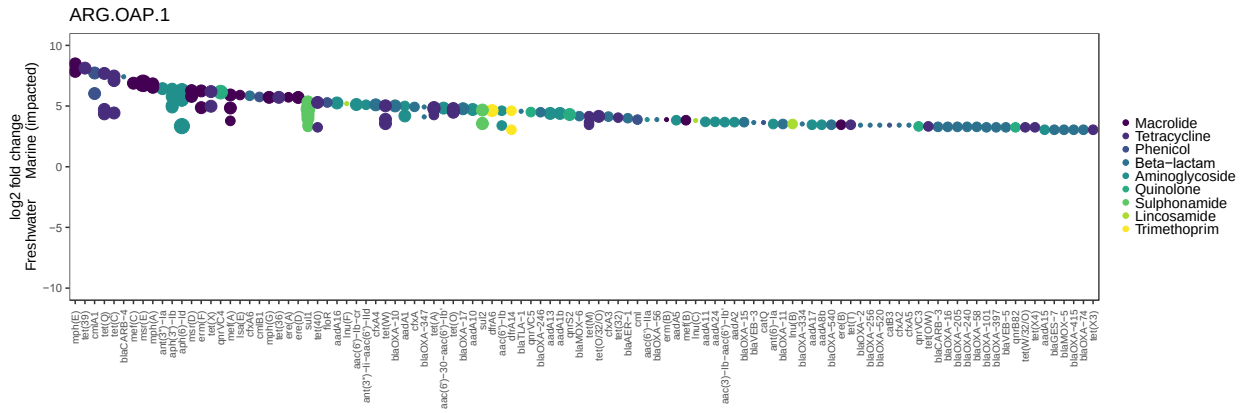

Figure 3.30: *Antibiotic resistance genes differentially abundant between wastewater-impacted marine sites and freshwater based on annotations from ResFinder.* ARGs with positive fold changes were differentially abundant in impacted marine waters. Sizes of data points correspond to the number of individuals in which the ARG was detected.

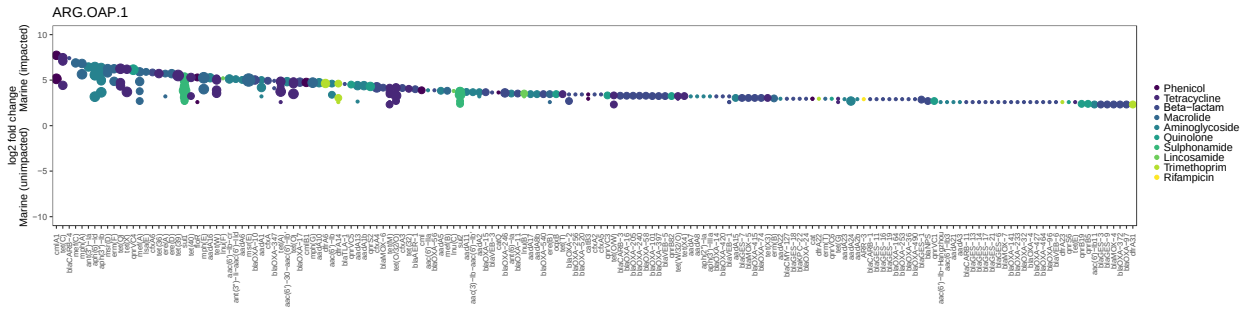

Figure 3.31: *Antibiotic resistance genes differentially abundant between wastewater-impacted marine sites and unimpacted marine sites based on annotations from ResFinder.* ARGs with positive fold changes were differentially abundant in impacted marine waters. Sizes of data points correspond to the number of individuals in which the ARG was detected.

Focus on Tetracycline and Beta-lactam ARGs

Based on the observation that tetracycline resistance gene subtypes differed between wastewater, humans, and wildlife (**Table 3.6-7, Table S3.21**) and that unique beta-lactam ARGs were noted in wildlife (**Figure 3.29**, **Table S3.23**), we further examined specific ARGs conferring resistance to tetracycline and beta-lactam antibiotics. Both ARG-OAP.1 (**Figure 3.32**) and ResFinder (**Figure 3.33**) annotations suggest *tetW* to be intensely human-associated. This coincides with the observation of *tetW* among the most abundant subtypes overall in humans

when annotating with ResFinder. However, *tetW* can also be found across wildlife species though at comparatively lower concentrations. Using ARG-OAP.1 data, *tetW* was detected in 16/24 sea lions, all six giant tortoises, two land iguanas, and one sea turtle. ResFinder mapped *tetW* in the same two land iguanas, sea turtle, all six giant tortoises, but only six sea lions. Moreover, *tetC* seems to be specifically enriched in wastewater and wastewater-impacted environments, as this subtype was entirely absent in wildlife with the exception of one giant tortoise and one sea lion (ARG-OAP.1 data) or one sea turtle (ResFinder data). Variants of *tetM* and *tetO* were common to sea lions, giant tortoises, humans, and wastewater, though the particular ResFinder *tetO/W* variant found to be highly abundant in humans (**Figure 3.33**) was detected less frequently and at lower concentrations in wildlife. Finally, *tet32*, *tet34*, and *tet35* were frequently detected across wildlife species, with *tet34* in particular detected in 100% of land iguanas. Interestingly, these tetracycline subtypes were comparatively less abundant in wastewater.

In regards to ARGs conferring resistance to beta-lactam antibiotics, we observed less overlap in the annotations from ARG-OAP.1 and ResFinder (**Table 3.8**). In total, beta-lactam subtypes were detected in 50/59 wildlife samples when annotating with ARG-OAP.1 compared to 20 with ResFinder. With one exception, the 20 wildlife samples with beta-lactam ARG subtypes detected by ResFinder were also in the ARG-OAP.1 detection group. In both cases, the greatest number of beta-lactam subtypes was recorded in a land iguana from North Seymour (G19_37). Consistent with the differential abundance analysis performed with DESeq2 which identified bla_{SED-1} as unique to land iguanas using ResFinder annotations, all land iguanas from North Seymour and Santa Fe harbored this particular ARG, whereas it was undetected among all other sample types including wastewater. ResFinder annotations also mapped bla_{ACT-15} and

bla_{MAL-1} among land iguanas. Using ARG-OAP.1, the land iguana with the most beta-lactam subtypes (G19_37 from North Seymour) was found to also carry extended-spectrum beta-lactamases bla_{SHV-53} and bla_{SHV-59}, while bla_{OXA-2} and bla_{OXA-9} were detected in several land iguanas.

Sea lion gut microbiomes annotated with ARG-OAP.1 were found to carry mostly class C beta-lactamases, bla_{OXA-9}, and penicillin binding proteins, with a pattern of increasing beta-lactam ARG diversity from western to eastern (i.e. human-inhabited) sampling sites. Specifically, extended spectrum beta-lactamases bla_{TEM-91} and $bla_{TEM-178}$ were detected in an individual from Punta Pitt, San Cristobal, while bla_{OXA-63}, $bla_{OXA-136}$, and $bla_{OXA-192}$ were detected among several sea lions from El Malecon, San Cristobal, the sampling site with the most proximity to humans and human-associated waste. ResFinder annotations followed a similar pattern with increasing beta-lactam ARG detection and diversity from west to east: $bla_{TEM-146}$ and $bla_{TEM-212}$ were found in the same individual from Punta Pitt as above, while sea lions from El Malecon were found to carry $bla_{OXA-136}$, $bla_{OXA-137}$, $bla_{OXA-192}$, $bla_{OXA-470}$, and $bla_{OXA-471}$.

A single giant tortoise was the only wildlife species found to carry CTX-M ESBLs, though it should be noted that this individual's metagenome was sequenced as part of an Illumina workshop and the library was orders of magnitude deeper than other samples (**Table S3.2**, 230,460,728 raw sequence pairs versus 11,298,249.5 average raw sequence pairs for four giant tortoises sequenced prior to workshop). Detection of $bla_{CTX-M-8}$, $bla_{CTM-M-40}$, and $bla_{CTX-M-63}$ at this sequencing depth suggests their prevalence is likely quite low. The second giant tortoise sequenced as part of the Illumina workshop (with 191,066,264 raw sequence pairs) was found

between to harbor $bla_{TEM-117}$, $bla_{TEM-118}$, $bla_{TEM-166}$, and $bla_{TEM-214}$ between ARG-OAP.1 and ResFinder annotations.

Sea turtles also carried unique beta-lactam ARGs. Consistent with the observation that $bla_{OXA-SHE}$ and $bla_{OXA-PAO}$ were found to be differentially abundant in this species compared to other wildlife with ResFinder data (**Table S3.23**), these subtypes were detected in four and two individuals, respectively. ARG-OAP.1 annotations also showed sea turtles to harbor bla_{OXA-50} and bla_{OXA-55}.

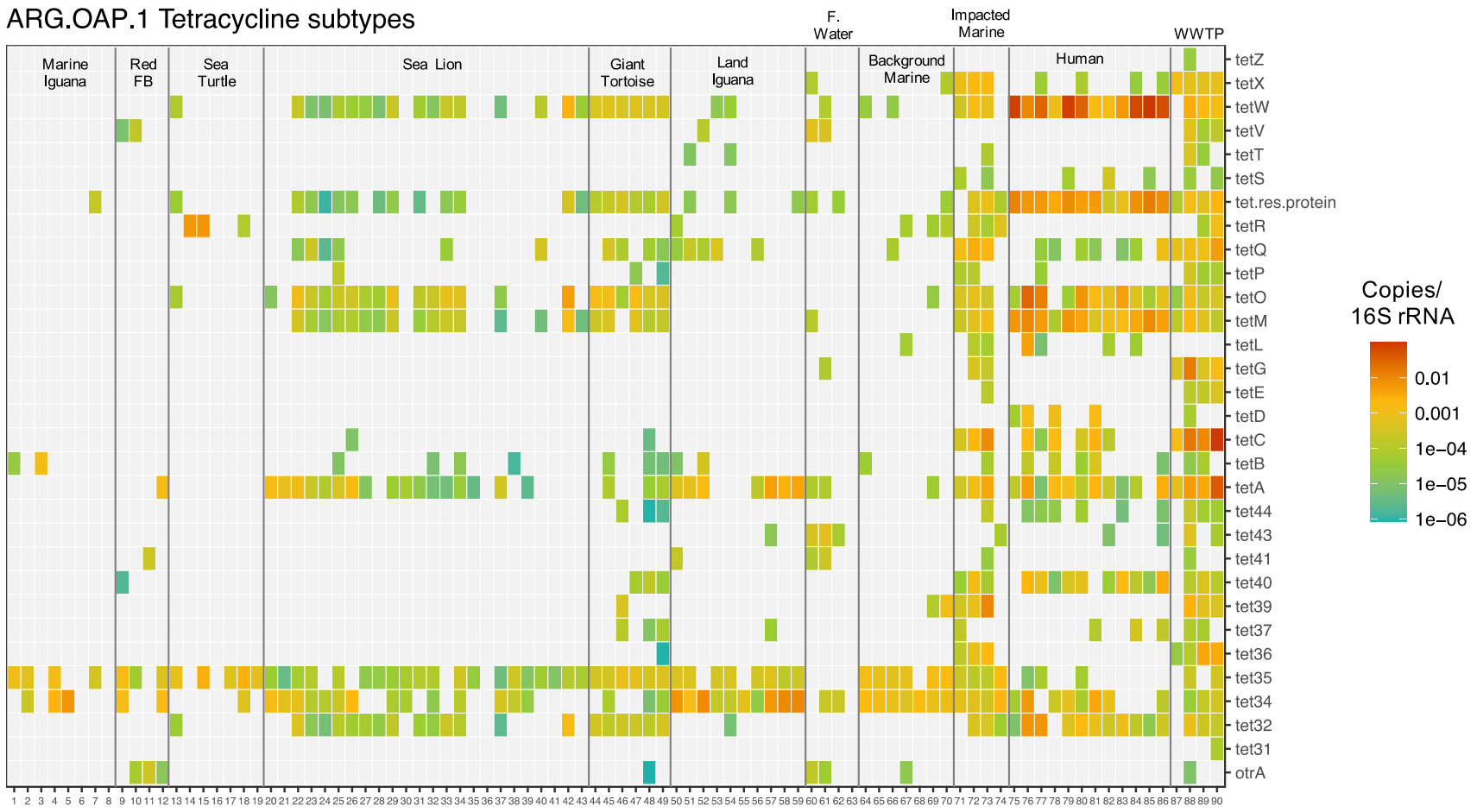

Figure 3.32: *Tetracycline subtypes detected using ARG-OAP.1.*

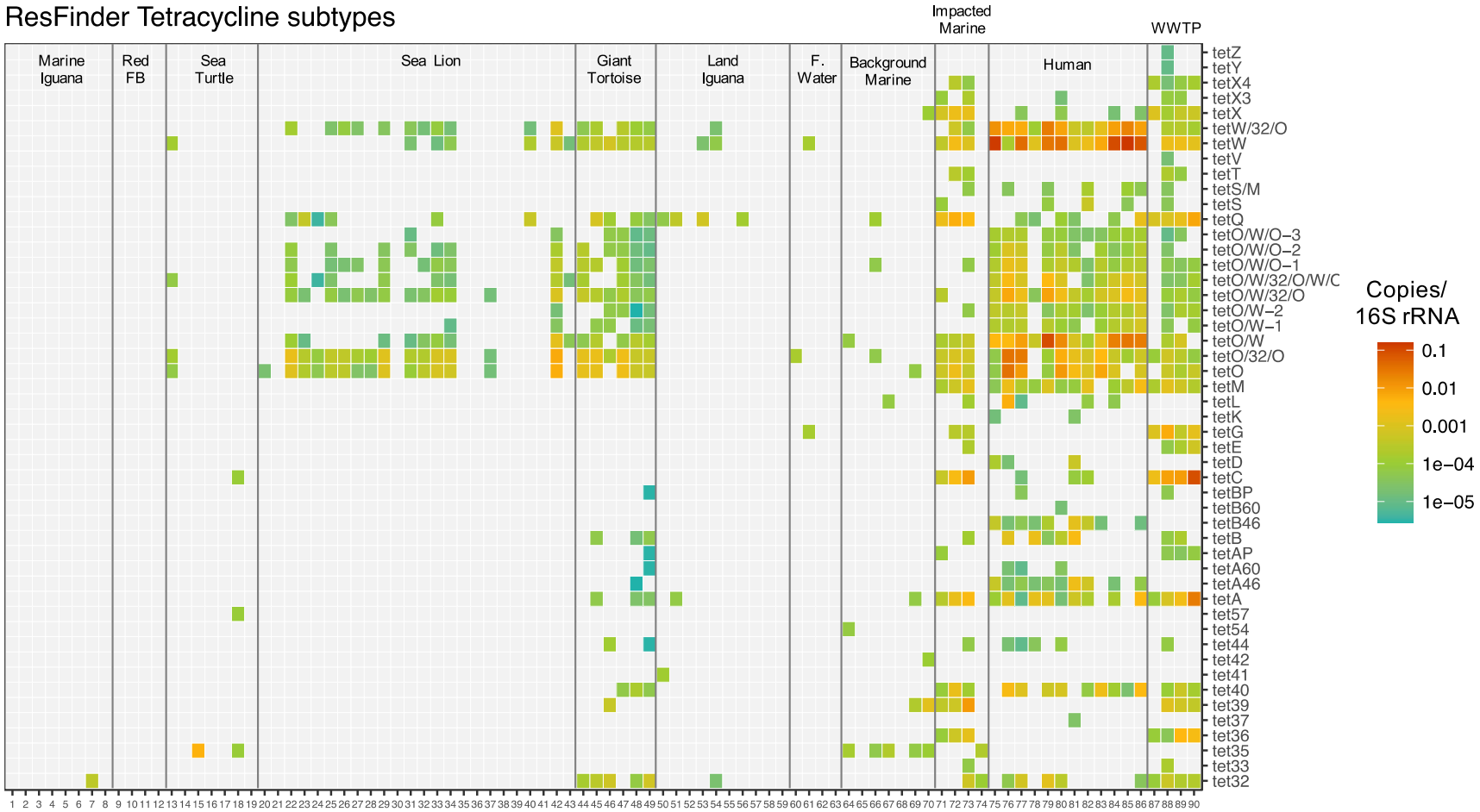

Figure 3.33: *Tetracycline subtypes detected using the ResFinder database.*

Table 3.8: *ARGs conferring resistance to beta-lactam antibiotics detected in wildlife samples using ARG-OAP.1 or ResFinder for annotation. N denotes the total number of beta-lactam subtypes detected in each sample.*

ID	Wildlife Species	Location	ARG-OAP.1						ResFinder					
				Top Five Beta-Lactamase Subtypes						Top Five Beta-Lactamase Subtypes				
			N	1	2	3	4	5	N	1	2	3	4	5
G19_31	L. Iguana	Santa Fe	5	class C	class A	OXA-9	PBP-1A	PBP-1B	1	SED1				
G19_9	L. Iguana	Santa Fe	6	class C	class A	OXA-9	PBP-1A	penA	1	SED1				
G19_34	L. Iguana	Santa Fe	5	class C	class A	OXA-9	PBP-1A	PBP-1B	1	SED1				
G19_37	L. Iguana	N. Seymour	37	class A	penA	class C	SHV-53	SHV-39	125	SED1	ACT-15	MAL-1	ACT-7	CKO -1
G19_26	L. Iguana	N. Seymour	11	class C	class A	OXA-9	OXY-2	ampC	4	SED1	ACT-15	ACT-16	OXY-2-1	
G19_43	L. Iguana	N. Seymour	9	class A	class C	OXY-2	penA	OXA-9	15	SED1	OXY-2-1	OXY-2-6	OXY-2-9	MAL -1
G19_14	L. Iguana	Plaza Sur	2	class A	penA				--	--	--	--	--	--
G19_45	L. Iguana	Plaza Sur	8	class A	penA	OXY-2	class C	ampC	6	MAL-1	OXY-2-1	OXY-2-5	ACT-15	OXY -2-7
G19_30	L. Iguana	Plaza Sur	1	penA					--	--	--	--	--	--
G19_36	L. Iguana	Plaza Sur	6	class C	class A	OXY-1	PBP-1A	OXA-9	4	OXY-1-1	OXY-1-3	OXY-1-7	OXY-6-2	
G19_113	Sea Lion	Punta Mangle	6	class C	OXA-9	PBP-1A	penA	PBP-1B	--	--	--	--	--	--
G19_114	Sea Lion	Punta Mangle	4	class C	OXA-9	PBP-1B	PBP-1A		--	--	--	--	--	--
G19_121	Sea Lion	Cabo Douglas	3	class C	OXA-9	PBP-1B	OXA-55	PBP-1A	1	OXA-55				
G19_122	Sea Lion	Cabo Douglas	4	class C	OXA-9	PBP-1A	PBP-1B		--	--	--	--	--	--
G19_123	Sea Lion	Cabo Douglas	4	class C	OXA-9	PBP-1B	PBP-1A		--	--	--	--	--	--
G19_124	Sea Lion	Cabo Douglas	4	class C	OXA-9	PBP-1A	PBP-1B		--	--	--	--	--	--
G19_134	Sea Lion	Puerto Egas	4	class C	OXA-9	PBP-1B	PBP-1A		--	--	--	--	--	--
G19_136	Sea Lion	Puerto Egas	1	class C					--	--	--	--	--	--
G19_142	Sea Lion	Puerto Egas	2	class C	PBP-1B				--	--	--	--	--	--
G19_145	Sea Lion	Puerto Egas	4	class C	OXA-9	PBP-1A	PBP-1B		--	--	--	--	--	--
G19_152	Sea Lion	Champion	4	class C	penA	PBP-1B	PBP-1A		--	--	--	--	--	--
G19_148	Sea Lion	Champion	2	OXA-192	penA				1	OXA-192				
G19_159	Sea Lion	Champion	3	class A	OXA-192	OXA-63			--	--	--	--	--	--
G19_163	Sea Lion	Champion	3	class C	class A	PBP-1B			--	--	--	--	--	--
G19_181	Sea Lion	Punta Pitt	3	PBP-1A	class C	penA			--	--	--	--	--	--
G19_174	Sea Lion	Punta Pitt	6	class C	penA	OXA-9	PBP-1A	PBP-1B	--	--	--	--	--	--
G19_177	Sea Lion	Punta Pitt	4	PBP-1A	TEM-178	TEM-91	PSE-1		2	TEM-146	TEM-212			
G19_188	Sea Lion	Punta Pitt	3	class C	penA	OXA-9			--	--	--	--	--	--
G19_194	Sea Lion	El Malecon	4	OXA-192	CfxA2	class A	OXA-63		2	OXA-470	OXA-136			
G19_197	Sea Lion	El Malecon	4	OXA-192	OXA-63	class C	OXA-136		5	OXA-192	OXA-136	OXA-471	OXA-470	OXA -137
G19_199	Sea Lion	El Malecon	1	penA					--	--	--	--	--	--
G19_201	Sea Lion	El Malecon	2	OXA-192	OXA-136				1	OXA-470				

G18_65	M. Iguana	Playa Loberia	4	penA	class A	OXA-9	PBP-1B		--	--	--	--	--	--
G18_76	M. Iguana	Playa Loberia	1	class A					--	--	--	--	--	--
G18_153	M. Iguana	Los Lobos	1	penA					--	--	--	--	--	--
G18_160	M. Iguana	Los Lobos	1	ccrA										
G18_164	M. Iguana	Los Lobos	1	class A					--	--	--	--	--	--
G18_12	G. Tortoise	Galapaguera	1	class A					--	--	--	--	--	--
G18_14	G. Tortoise	Galapaguera	1	class C					--	--	--	--	--	--
G18_18	G. Tortoise	Galapaguera	2	class C	class A				--	--	--	--	--	--
G18_19	G. Tortoise	Galapaguera	1	class C					--	--	--	--	--	--
G18_3	G. Tortoise	Galapaguera	10	class C	penA	KLUG-1	CTX-M-40	PBP-1A	4	CTX-M-40	CTX-M-63	SST-1	CTX-M-8	
G18_10	G. Tortoise	Galapaguera	9	class C	class A	penA	TEM-117	TEM-118	2	TEM-166	TEM-214			
G18_194	Red F.B.	Punta Pitt	8	PBP-1A	PSE-1	OXA-50	THIN-B	class A	--	--	--	--	--	--
G18_195	Red F.B.	Punta Pitt	2	class A	THIN-B				--	--	--	--	--	--
G18_198	Red F.B.	Punta Pitt	3	class C	VIM-34	PBP-1B			--	--	--	--	--	--
G18_170	Sea Turtle	Playa Carola	--	--	--	--	--	--	2	OXA-SHE	OXA-372			
G18_172	Sea Turtle	Playa Carola	3	OXA-50	class C	PDC-5			6	PAO	OXA-486	OXA-SHE	OXA-395	OXA-485
G18_174	Sea Turtle	Playa Carola	4	class C	OXA-55	OXA-50	PDC-7		2	PAO	OXA-SHE			
G18_183	Sea Turtle	Playa Loberia	1	PBP-1A					--	--	--	--	--	--
G18_184	Sea Turtle	Playa Loberia	5	OXA-55	TUS-1	PBP-1A	MUS-1	ampC	1	OXA-SHE				

Discussion

The purpose of this study encompassed two aims: first, compare three publicly available antibiotic resistance gene annotation approaches in their ranking of wildlife, water, wastewater, and human resistomes from the Galapagos islands; and second, perform a deep characterization of these metagenomes to elucidate how resistomes vary over space and between distinct reservoirs. Overall, we found varying levels of agreement between ARG annotation strategies, indicating that pipeline selection should perhaps be based on the research question and study system. We selected two approaches of distinct scope and underlying methodology, the Antibiotic Resistance Gene Online Analysis Platform (ARG-OAP.1) and ResFinder, for further characterization of Galapagos resistomes. Our results point to distinct patterns in both ARG abundance and diversity across a gradient of anthropogenic influence, with several key findings in wildlife.

Comparison of ARG annotation approaches

We compared three approaches for annotating ARGs in metagenomes: Antibiotic Resistance Gene Online Analysis Platform (ARG-OAP), MegaRes, and ResFinder. Upon observing that ARG-OAP and MegaRes featured databases with broader inclusion criteria, we additionally considered modifications of these pipelines. This included ARG-OAP.1, in which genes conferring resistance to “multidrug” and “unclassified” types were omitted; MegaRes.1, which considered only genes assigned to the “drugs” class; and Megares.2, which further omitted genes requiring SNP confirmation. Inclusion of these modifications brought the total comparison to six approaches.

The base condition of ARG-OAP with no modifications was associated with the most overall gene observations, which is consistent with its status as the largest database in the comparison (12,307 genes). Exclusion of multidrug and unclassified genes reduced this total to

7,170 genes, of which 3,446 where detected in our dataset. On the other hand, MegaRes in the base condition with no modifications generally yielded the highest ARG sum abundance/16S rRNA, with sums especially inflated among humans and wildlife. Inclusion of only genes conferring resistance to “drugs” and exclusion of genes requiring SNP confirmation yielded ARG sums more comparable to the other approaches. When considering how pipelines ranked different sample types based on ARG sum abundance, we observed agreement in ResFinder, ARG-OAP.1, and MegaRes.2 in ranking wastewater and human samples highly, whereas ARG-OAP.0, MegaRes.0, and MegaRes.1 ranked wildlife, and specifically land iguanas, as the highest. Kendall rank correlation on ARG sums revealed the highest correlation between MegaRes.0 and MegaRes.1 (τ=0.87), but a high degree of agreement was also recorded between Mega.Res.2 and ResFinder (τ=0.81). In terms of ARG diversity, we noted good agreement between MegaRes.2 and ResFinder when using the Shannon index. Overall, ResFinder showed the most significant differences in alpha diversity between sample types. This was driven by the high diversity noted for wastewater compared to other sample types, and coincides with the description of ResFinder as a database of acquired ARGs. Finally, categorization of samples as human, wastewater, wildlife, or water had a roughly equal effect in explaining between sample distances according to the Bray-Curtis dissimilarity index, with R-square values ranging from 0.16 to 0.17 for ARG-OAP.1, MegaRes.2, and ResFinder.

Based on the results of our pipeline comparison, we suggest that all approaches are defensible, but rationale should be provided for the selection of a given pipeline or database. For example, ResFinder has more stringent inclusion criteria in representing acquired antibiotic resistance genes, while ARG-OAP and MegaRes are clearly more comprehensive. If researchers seek to characterize the resistome downstream from a known pollution source, such as for

example discharge of wastewater effluent into a surface water, ResFinder may be an appropriate choice to track the distance-decay of intensely anthropogenic ARGs. On the other hand, ARG-OAP and MegaRes may be better suited to explore the larger resistome including ARGs not of immediate clinical concern. In our case, we chose ARG-OAP.1 and ResFinder for further characterization of 90 Galapagos resistomes to capture both acquired antibiotic resistance genes and more general resistance functions, which may be considered background in our study system.

Characterization of Galapagos resistomes: ARG-OAP.1

Annotation of antibiotic resistance genes with ARG-OAP.1 revealed sum abundance to be highest in wastewater followed by humans, wildlife, and water. Our data corroborates previous reports on the impacts of wastewater discharge at marine sites Playa Carola and Playa Marinero Both wastewater (Overby et al., 2015; Grube et al., 2020). Both wastewater and wastewater-impacted marine sites were dominated by ARGs conferring resistance to aminoglycoside and beta-lactam antibiotics and shared *sul1* as the most abundant subtype, consistent with previous reports regarding *sul1* predominance in wastewater (Che et al., 2019) and impacted surface waters (Makowska et al., 2016). In contrast, tetracycline and bacitracin ARGs accounted for the most abundant classes in marine sites without apparent wastewater discharge, with *tet34* and *tet34* among the top three most abundant subtypes. This finding is not surprising as *tet34* and *tet35* have been described in *Vibrio* species (Nonaka and Suzuki et al., 2002; Teo et al., 2002). Reflecting the abundance pattern of wastewater, impacted marine waters harbored many differentially abundant genes compared to background marine sites, including variants of chloramphenicol, tetracycline (i.e. *tetC*), aminoglycoside, and sulfonamide (i.e. *sulI*). In contrast, freshwater was differentially characterized by variants of the bacitracin ARG *bacA*,

who presence is well documented across diverse habitats (Neseme et al., 2014; Gupta et al., 2019; Scott et al., 2020) and may be considered background resistance.

The human samples in our study, which originated from children under age two, were dominated at the class level by ARGs conferring resistance to tetracycline, MLS, and aminoglycoside antibiotics. At the subtype level, the macrolide ARG *ermX* and tetracycline ARG *tetC* were the top two by abundance. Interestingly, both *ermX* and *tetC* are well-documented in the genus *Bifidobacterium* (Bottacini et al., 2018; Taft et al., 2018; Cao et al., 2020) which accounted for the most abundant genera in these metagenomes (see Chapter 5). The study by Taft and colleagues (2018) in particular focused on children under age two in Bangladesh and found a singular association between elevated *ermX* in children with a high relative abundance of *Bifidobacterium* in their gut microbiomes. Notably, when asking if ARG sum abundance differed in our cohort depending on birth mode, the difference was insignificant; however, using DESeq2 we found a significant number of genes to be differentially abundant in children born via Caesarean section compared to those born vaginally (**Figure 3.15**). Studies have reported both altered microbiota and elevated ARG load among infants born via Caesarean section (Shao et al., 2019), presumably due to intrapartum antibiotic prophylaxis (Pärnänen et al., 2018; Tapiainen et al., 2019). Our inability to detect significant differences in ARG sum abundance by birth mode could be due to the relatively small sample size. Alternatively, it could be that enrichment of specific ARGs is more consequential than changes in overall sum abundance (Pärnänen et al., 2018).

Among wildlife, ARGs conferring resistance to macrolide, lincosamide, and streptogramins (MLS) accounted for one of the top three classes in all species, with bacitracin and the subtype *bacA* in particular widespread. As discussed above, the ubiquity of *bacA* across

diverse environments, ranging from anthropogenically-impacted to pristine, has built evidence for *bacA* as a more background marker ARG. Land iguanas were found to have the highest ARG sum abundance and were dominated by ARGs conferring resistance to MLS, fosmidomycin, and kasugamycin antibiotics, with subtype *ksgA* as the most abundant subtype. The greatest number of differentially abundant genes were observed when comparing land iguanas to all other wildlife samples, with the genes accounting for the highest fold change including *macB* and *macA*, class A beta-lactamases, and ksgA. However, *macA/macB* together with *TolC* comprise an efflux pump in many Gram-negative bacteria (Xu et al., 2010), so their detection may not necessarily indicate a strong selective pressure. Notably, ARG sums were significantly different by sampling location for land iguanas, with three individuals from each North Seymour and Santa Fe exhibiting higher sums than the four land iguanas from Plaza Sur (**Figure 3.13**).

As for the other reptilian species, giant tortoises and sea turtles shared tetracycline among their top three most abundant ARG classes, though the top subtype in giant tortoises was *vatB*, an acetyltransferase originally identified in *Staphylococcus aureus* (Allignet and Solh, 1995). Accordingly, when performing differential abundance analysis, *vatB* in addition to *vanG* counted among the ARGs differentially abundant in giant tortoises compared to wildlife species. In contrast, differential abundance analysis pointed to *macB*, *tetR*, and bla_{OXA-50} as subtypes relatively more abundant among sea turtles compared to other wildlife. Finally, the top ARGs among red footed boobies were assigned to vancomycin, particularly through predominance of *vanR* and *vanS* subtypes. Vancomycin resistance has been previously reported in Australian gulls (Oravcova et al., 2017) and Canadian crows (Oravcova et al., 2014), though with genotypic resistance confirmed through detection of *vanA* and/or *vanB* which are more commonly associated with vancomycin resistance phenotypes in *Enterococci* (Dolejska, 2020).

Characterization of Galapagos resistomes: ResFinder

Annotation of antibiotic resistance genes with ResFinder revealed a pattern similar to ARG-OAP.1, with the highest sums recorded in wastewater followed by humans. However, the order of water and wildlife was reversed compared to ARG-OAP.1, with mean ARG sums in water exceeding those in wildlife. Aminoglycosides were again noted among the top ARG classes in wastewater, along with macrolide and tetracycline ARGs, and *tetC* accounted for the most abundant subtype in wastewater. Mirroring results from ARG-OAP.1, *sul1* remained among the top three subtypes in both wastewater and wastewater impacted marine sites. Collectively, annotations from ARG-OAP.1 and ResFinder corroborate the predominance of *sul1*in wastewater and wastewater-impacted waters. Importantly, *sul1* was frequently detected in human gut metagenomes (7/12, 58.3%) but was not detected in any wildlife gut microbiomes. Taken together, our findings align with previous reports of *sul1* as an intensely anthropogenic ARG, particularly in the context of association with the class I integron (Gillings et al., 2015; Koczura et al., 2016).

Using ResFinder, only wastewater-impacted marine sites have appreciable sums of ARGs compared to freshwater and background marine sites (**Figure 3.10** vs **Figure 3.23**), again reflecting the description of ResFinder as a database of acquired resistance genes. Despite differences in inclusion criteria, ARG-OAP.1 and ResFinder identified many of the same genes to be differentially abundant between water types, such as the increased abundance of *cmlA*, *tetC*, and *tet39* in impacted marine waters compared to freshwater (**Table S3.25**). Similarly, both ARG-OAP.1 and ResFinder pointed to *cmlA* as the most differentially abundant subtype between impacted and background marine sites (**Table S3.26**).

Annotations of ARGs in human samples using ResFinder aligned well with the results from ARG-OAP.1, with *tetW* and *ermX* again identified as the top two most abundant gene subtypes. While sum abundance of ARGs was not significantly different by birth mode, the majority of differentially abundant genes belonged to the Cesarean section birth group compared to babies born vaginally. Also consistent with ARG-OAP.1 data, ResFinder annotations pointed to two of the same subtypes, *aadA* and *aph(3)-I*, among the top ARGs comparatively higher in the Caesarean birth group. There was less overlap observed for wildlife however, with the top ARG class in land iguanas assigned to the beta-lactam rather than MLS class. Again, land iguanas from North Seymour and Santa Fe were found to have significantly higher ARG sum abundance over those sampled on Plaza Sur (**Figure 3.25**). In addition to land iguanas, marine iguanas and sea turtles shared beta-lactam ARGs as the most abundant class, while no other wildlife had beta-lactams in the top three most abundant classes. Instead, phenicol ARGs constituted the most abundant class in both sea lions and red footed boobies, and accounted for the second most abundant class in giant tortoises and land iguanas.

Consistent with the results from ARG-OAP.1, tetracycline was among the top three most abundant ARG classes in giant tortoises, with *tetO* and *tetW* as the two most abundant subtypes. When performing differential abundance analysis, tetracycline variants, rather than *vatB* and *vanG* as indicated by ARG-OAP.1, were among those comparatively elevated in giant tortoises compared to other wildlife species. Nieto-Claudin and colleagues have communicated two recent reports on ARG carriage among giant tortoises (*Chelonoidis porteri*) on the island of Santa Cruz. In the first, the authors analyzed fecal samples from 28 individuals for a panel of 21 genes using quantitative PCR (Nieto-Claudin et al., 2019). The most frequently detected genes were *tetQ* and *tetW*, which were detected in 100% and 96.4% of individuals, respectively. Using ARG-OAP.1

or ResFinder, *tetW* was detected in 6/6 (100%) of giant tortoises in our metagenomic survey, while both approaches found *tetQ* in the same 4/6 (75%) of individuals. That we can consistently detect *tetW* and *tetQ* using metagenomic sequencing, which is less sensitive than the qPCR methods used by Nieto-Claudin and colleagues, provides solid evidence for their presence in the gut microbiomes of Galapagos giant tortoises. As discussed above, *tetW* was among the top subtypes in human gut microbiomes in our study. Interestingly, this gene was first reported in *Butyrivibrio fibrisolvens* from rumen guts (Barbosa et al., 1999) before subsequent reports in *Fusobacterium prausnitzii* and *Bifidobacterium longum* hosts in the human gut microbiome (Scott et al., 2000). From just these two initial reports, it seems *tetW* can be found in a range of hosts, as the three species belong to distinct phyla (Firmicutes, Fusobacteria, and Actinobacteria, respectively.) In the case of giant tortoises, it would be interesting to place *tetW* within its host or hosts using a method like epicPCR (Spencer et al., 2016).

Other tetracycline genes detected in Nieto-Claduin and colleagues' first study included *tetB*, *tetA*, and *tetM*, though at prevalence ranging from 3.6-10.7%. In contrast, we detected *tetA* and *tetB* in the same 3/6 (50%) of individuals using either ARG-OAP.1 or ResFinder, while *tetM* was detected in 5/6 (83%) giant tortoises with ARG-OAP.1 but undetected with ResFinder. Outside of tetracycline, aminoglycoside subtype *aadA* constituted the third most detected ARG in the first study of giant tortoises on Santa Cruz at 39.3%. In contrast, we detected *aadA* in only one giant tortoise using either annotation approach. Finally, in contrast to detection of bla_{TEM} in 32.1% of individuals via qPCR, we detected bla_{TEM} variants in only one giant tortoise in our metagenomic survey, and only with very deep sequencing. Recently, Nieto-Claudin and team published a second study comparing fecal samples from giant tortoises on Santa Cruz with those of a remote population of *Chelonoidi vandenburghi* giant tortoises on Alcedo Volcano on Isabela

island (Nieto-Claudin et al., 2021). Again, *tetW* and *tetQ* were detected most frequently across both populations, with 93% and 71% detection in 200 *C. porter* individuals on Santa Cruz and 51.4% and 50% detection in 70 *C. vandenburghi* giant tortoises on Isabela. Interestingly, this study found a higher incidence of *tetM* among Santa Cruz giant tortoises (35.5%) compared to the first study, while this subtype went undetected among individuals from Isabela. Surprisingly, incidence of bla_{TEM} was slightly higher among the Isabela population, with detection at 37.1% versus 33% on Santa Cruz. Overall, this study found a pattern of increasing ARG abundance and diversity among giant tortoises with more anthropogenic exposure (i.e. farming and tourism), with significant differences recorded between the Santa Cruz and Isabela populations, but also within the Santa Cruz population itself when comparing individuals from remote versus human-impacted sites. It is worth noting that all six metagenomes in our study came from giant tortoises housed at La Galapaguera, the captive breeding facility on San Cristobal. Accordingly, we would expect our data to more closely match those reported by Nieto-Claudin and colleagues from individuals on Santa Cruz, particularly for those in close contact with human activities.

Connecting the resistome to the microbiome

Given the growing reports of antibiotic resistant *Enterobacteriaceae* in wildlife, including mobile, plasmid-associated resistance (Dolejska and Papagiannitsis, 2018) and resistance conferred through ESBLs (reviewed in Guenther et al., 2011), we examined all beta-lactam ARG subtypes observed in wildlife metagenomes in our study considering both ARG-OAP.1 and ResFinder annotations. This analysis was additionally motivated by the observation that particular beta-lactam ARGs were differentially abundant between wildlife species. For example, consistent with the differential abundance analysis performed with ResFinder data, land iguanas were found to uniquely harbor bla_{SED-1}, a CTX-M type ESBL (Petrella et al., 2004).

This subtype was exclusively detected in the three land iguanas from each North Seymour and Santa Fe, and was absent from all other metagenomes. Interestingly, this gene was first described in a clinical isolate of *Citrobacter sedlakii* cultured from the bile of a hospitalized patient (Petrella et al., 2001; Petrella et al., 2004). More recently, bla_{SED-1} was detected in *E. coli* cultured from ready to eat lettuce (Lio et al., 2020). Surprisingly, when examining the taxonomic assignments of SSU rRNA units with Metaxa2 (Chapter 5), we detected hits for *Citrobacter sedlakii* in the same land iguanas from North Seymour and Santa Fe and no other samples (**Table S3.27**). Additional beta-lactam ARGs apparently unique to land iguanas using ResFinder data include bla_{MAL-1}, which has been detected *Citrobacter koseri* isolated from sewage sludge (Ekwanzala et al., 2020) and bla_{ACT-15}, which has been described previously in *Enterobacter cloacae* isolated from wildlife (Literak et al., 2014). Annotation with ARG-OAP.1 pointed to different beta-lactam ARGs among land iguanas, with detection of bla_{OXA-2} and bla_{OXA-9} in several individuals, which were both originally described in *Enterobacteriaceae* hosts (reviewed in Poirel et al., 2010). Finally, the individual from North Seymour with the most beta-lactam subtypes among wildlife also harbored bla_{SHV-53} and bla_{SHV-59}. It is worth nothing that *Klebsiella* are often the hosts of bla_{SHV} genes (reviewed in Liakopoulos et al., 2016), and using Metaxa2 taxonomic assignments, this genus was detected in all land iguanas from North Seymour and Santa Fe, but only one land iguana from Plaza Sur.

Sea lions also presented an interesting case with beta-lactam ARGs. While sea lions living closer to human activity did not have statistically greater sum abundances of ARGs compared to those living in more remote colonies, we did observe a pattern of increasing beta-lactam detection and diversity along a gradient of human influence moving west to east. For example, in an individual sampled at Punta Pitt we detected $bla_{TEM-178}$, $bla_{TEM-91}$with ARG-

OAP.1 data and $bla_{TEM-146}$ and $bla_{TEM-212}$ among ResFinder annotations. Previous studies have placed bla_{TEM-91} in *E. coli* (Kurokawa et al., 2003) and $bla_{TEM-178}$ in *Serratia marcescens* (Perilli et al., 1997), while reports of $bla_{TEM-146}$ in *E. coli* and $bla_{TEM-212}$ in *Providencia stuartii* come from unpublished NCBI entries (accession numbers AAZ14084.2 and WP_063864901.1, respectively.) Whereas bla_{OXA-9} was frequently detected in sea lions from the western sampling sites on Fernandina and Santiago, sea lions from El Malecon on San Cristobal, the site closest to humans and human activity, were found to harbor additional bla_{OXA} genes, including $bla_{OXA-470}$, $bla_{OXA-471}$, $bla_{OXA-136}$, $bla_{OXA-137}$, and $bla_{OXA-192}$. Interesting, the latter three have been documented in *Brachyspira* species (Mortimer-Jones et al., 2008; Jansson and Pringle, 2011). This genus was detected in the Metaxa2 taxonomic assignments of seven sea lions, including the four from El Malecon, as well as wastewater influent, one freshwater, and one marine water sample.

Giant tortoises and sea turtles were also found to carry unique beta-lactam ARGs. Notably, a single giant tortoise was the only wildlife sample found to harbor CTX-M ESBLs, including $bla_{CTX-M-40}$, $bla_{CTX-M-8}$, and $bla_{CTX-M-63}$, with the first two previously reported in *E. coli* (Hopkins et al., 2006) and the third in *Salmonella* (Pornruangwong et al., 2011). As mentioned in the results, it should be noted that this metagenome was one of two giant tortoise fecal DNA extracts sequenced as part of an Illumina workshop, which resulted in shotgun metagenomic sequence libraries orders of magnitude deeper than those produced in our main study. The second giant tortoise sequenced at this depth was found to carry $bla_{TEM-117}$, $bla_{TEM-118}$, $bla_{TEM-166,}$ and $bla_{TEM-214}$ between ResFinder and ARG-OAP.1 annotations. Notably, both $bla_{TEM-117}$ and $bla_{TEM-118}$ have been associated with clinical *Enterobacteriaceae* isolates (Leverstein-van Hall et al., 2002; Livermore et al., 2019) while $bla_{TEM-166}$ and $bla_{TEM-212}$ are found in *E. coli* as

unpublished NCBI entries (accession numbers ACI25375.1 and AJO16044.1, respectively). Finally, sea turtles were found to uniquely harbor $bla_{OXA\text{-}SHE}$, which originated in *Shewanella algae* (Walther-Rasmussen and Høiby, 2006), and $bla_{OXA\text{-}PAO}$, which has been reported in *Pseudomonas aeruginosa* (D'Souza et al., 2019). Notably, both reports in the literature (Pace et al., 2019; Blasi et al., 2020) and our taxonomic annotations corroborate the presence of these taxa among sea turtle gut microbiomes. We detected *S. algae* SSU rRNA units in 3/7 sea turtle metagenomes and the genus *Shewanella* in 6/7 individuals. Interestingly, *Shewanella* was also detected in two land iguanas, four sea lions, one red footed booby, and most marine water samples. *Pseudomonas aeruginosa* SSU rRNA was detected in two sea turtles along with wastewater and wastewater-impacted marine sites, while at the genus level *Pseudomonas* was found in 6/7 sea turtles.

Taken together, these data provide compelling evidence for linkages between beta-lactam ARGs and specific taxa within wildlife gut microbiomes. While our methods preclude absolute confirmation of bacterial hosts, cases such as paired $bla_{SED\text{-}1}$ and *Citrobacter sedlakii* detection exclusively among six land iguanas at minimum present opportunities for further hypothesis testing. Moreover, the observation that many of the beta-lactam ARGs described above are resident in E*nterobacteriaceae* hosts supports the selection of ESBL *E. coli* for global, One Health surveillance of AMR by the WHO Tricycle Program (WHO, 2021).

Conclusion

In conclusion, we find that while different antibiotic resistance gene annotation approaches will likely yield similar overall conclusions, consideration should be given to the research question and study system. Characterization of 90 Galapagos resistomes with ARG-OAP.1 and ResFinder revealed wastewater and human samples to harbor the greatest sum abundance of ARGs, followed by wildlife and water. Our data further confirm the impacts of

wastewater discharge on two coastal sites on San Cristobal and point to a distinct marine resistome in the absence of wastewater pollution. Among wildlife samples, we found ARG sums to be highest among land iguanas, which is surprising given their degree of isolation from humans and strict site fidelity as a terrestrial species. Differential abundance analysis revealed ARGs unique to each wildlife species, and further inspection of beta-lactam classes in particular revealed unique species and geographic patterns. While our methods preclude confirmation of ARG host, we provide compelling evidence for the linkage between select beta-lactams and bacterial taxa. Taken together, our data suggest that *Enterobacteriaceae* may be common hosts of beta-lactam ARGs in these species, but additional work is needed to resolve the host of specific ARGs, such as *tetW* in giant tortoises and bla_{SHV} in land iguanas. In alignment with efforts by the WHO, we propose that ESBL *E. coli* may be a useful, One Health indicator of AMR in the Galapagos and recommend that future research efforts focus on host and functional confirmation of antibiotic resistance in these systems.

Chapter 3: Supplemental Figures

```
>>MEG_1|Drugs|Aminoglycosides|Aminoglycoside-
resistant_16S_ribosomal_subunit_protein|A16S|RequiresSNPConfirmation
AGAATTTGATCTTGGTTCAGATTGAACGCTGGCGGCGTGGATGAGGCATGCAAGTCGAACGGAATAATGACTTCGGT
TGTTATTTAGTGGCGGAAGGGTTAGTAATACATAGATAATCTGTCCTCAACTTGGGAATAACGGTTGGAAACGACCG
CTAATACCGAATGTGGTATGTTTAGGCATCTAAAACATATTAAAGAAGGGGATCTTCGGACCTTTCGGTTGAGGGAG
AGTCTATGGGATATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCTTTGACGTCTAGGCGGATTGAGAGATTGAC
CGCCAACACTGGGACTGAGACACTGCCCAGACTTCTACGGAAGGCTGCAGTCGAGAATCTTTCGCAATGGACGAAAG
TCTGACGAAGCGACGCCGCGTGTGTGATGAAGGCTCTAGGGTTGTAAAGCACTTTCGCTTGGGAATAAGAGAGATTG
GCTAATATCCAATCGATTTGAGCGTACCAGGTAAAGAAGCACCGGCTAACTCCGTGCCAGCAGCTGCGGTAATACGG
AGGGTGCTAGCGTTAATCGGATTTATTGGGCGTAAAGGGCGTGTAGGCGGAAAGGAAAGTTAGATGTTAAATCTTGG
GGCTCAACCCCAAGCCAGCATCTAATACTATCTTTCTAGAGGGTAGATGGAGAAAAGGGAATTCCACGTGTAGCGGT
GAAATGCGTAGATATGTGGAAGAACACCAGTGGCGAAGGCGCTTTTCTAATTTACACCTGACGCTAAGGCGCGAAAG
CAAGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCTTGCCGTAAACGATGCATACTTGATGTGGATAGTCTCAAC
CCTATCCGTGTCGTAGCTAACGCGTTAAGTATGCCGCCTGAGGAGTACACTCGCAAGGGTGAAACTCAAAAGAATTG
ACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGGCTTGACATGT
ATTTGACCGCGGCAGAAATGTCGTTTTCCGCAAGGACAGATACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGCCG
TGAGGTGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCGTTAGTTGCCAACACTTAGGGTGGGAACTCTAACG
AGACTGCCTGGGTTAACCAGGAGGAAGGCGAGGATGACGTCAAGTCAGCATGGCCCTTATGCCCAGGGCTACACACG
TGCTACAATGGCCAGTACAGAAGGTAGCAATATCGTGAGATGGAGCAAATCCTCAAAGCTGGCCCCAGTTCGGATTG
TAGTCTGCAACTCGACTACATGAAGTCGGAATTGCTAGTAATGGCGTGTCAGCTATAACGCCGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACATCATGGGAGTTGGTTTTGCCTTAAGTCGTTGACTCAACCTGCAAAGGAGAGA
GGCGCCCAAGGTGAGGCTGATGACTGGGATGAAGTCGTAACAAG
```

Figure S3.1: *Example of sequence from MegaRes database requiring SNP confirmation.*

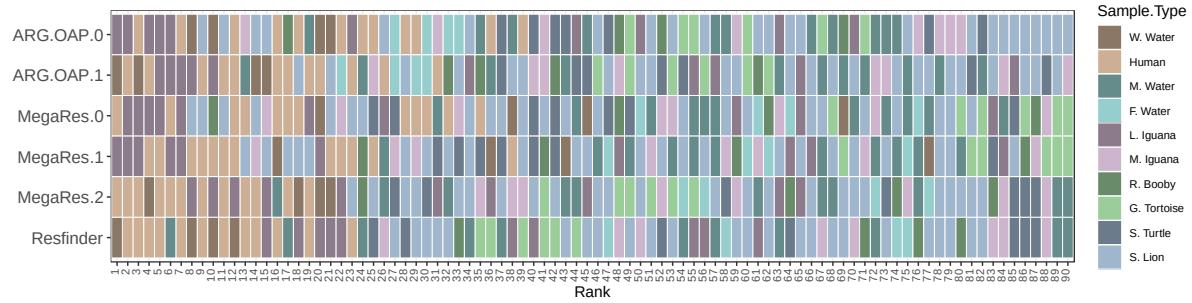

Figure S3.2: *Relative ranking of 90 samples by ARG sum abundance/16S rRNA, with samples categorized into ten subtypes.*

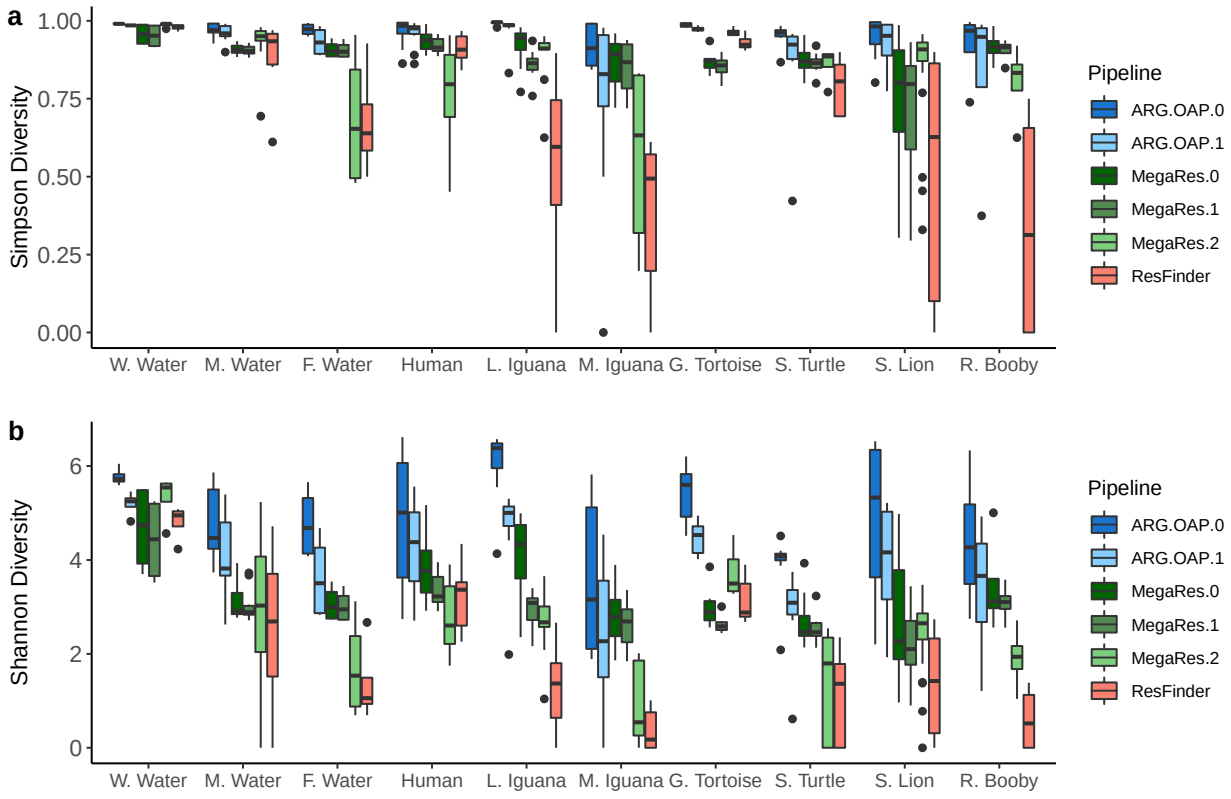

Figure S3.3: *Alpha diversity of ARGs by ten sample subtypes. a) Simpson diversity index. b) Shannon diversity index.*

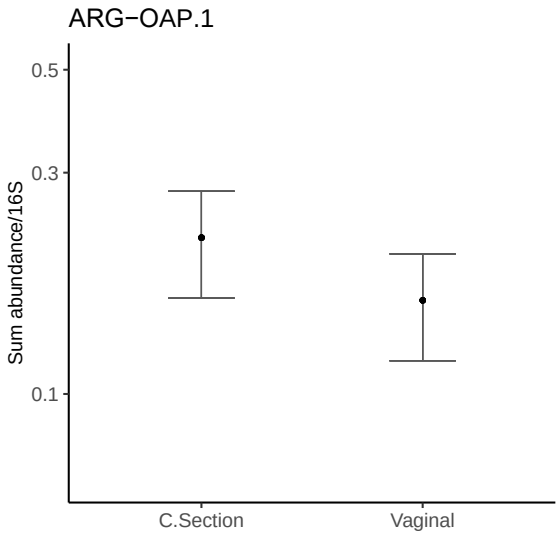

Figure S3.4: *ARG sum abundance/16S rRNA for children under two based on mode of delivery.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

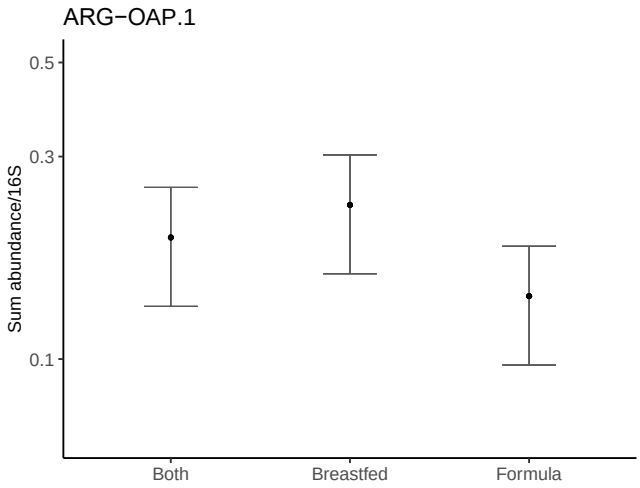

Figure S3.5: ARG sum abundance/16S rRNA for children under two based on nutrition. Error bars and black points represent negative binomial GLM-predicted means ± SE.

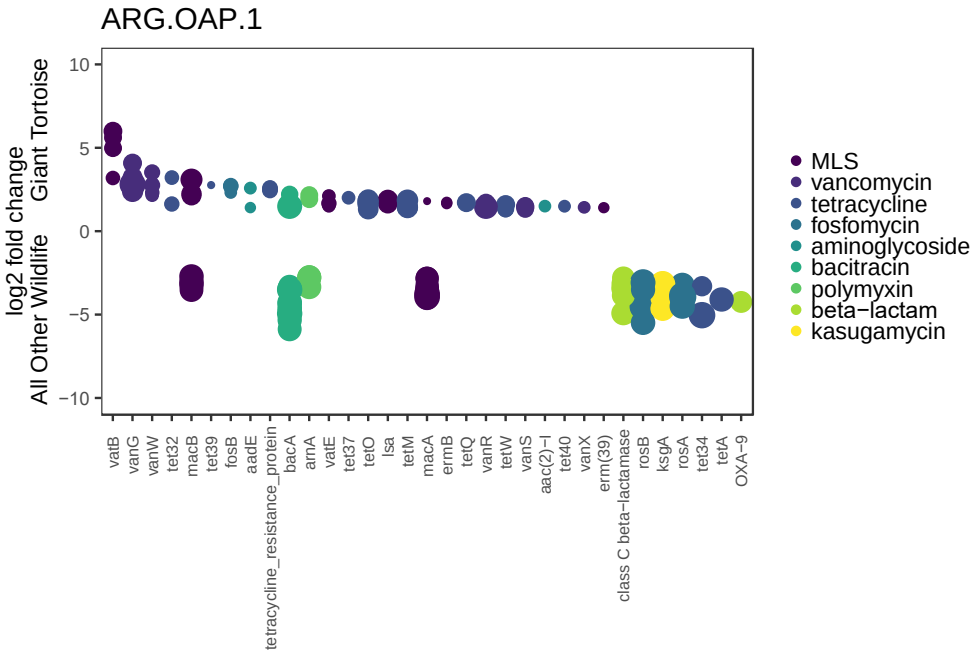

Figure S3.6: *Antibiotic resistance genes differentially abundant between giant tortoises and all other wildlife based on annotations from ARG-OAP.1.* ARGs with positive fold changes were differentially abundant in giant tortoises, while ARGs with negative fold changes were differentially in other wildlife. Sizes of data points correspond to the number of individuals in which the ARG was detected.

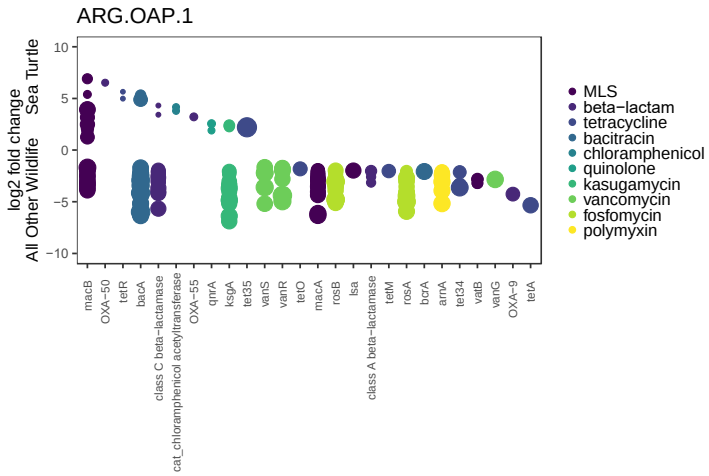

Figure S3.7: *ARGs differentially abundant between sea turtles and all other wildlife based on ARG-OAP.1.*

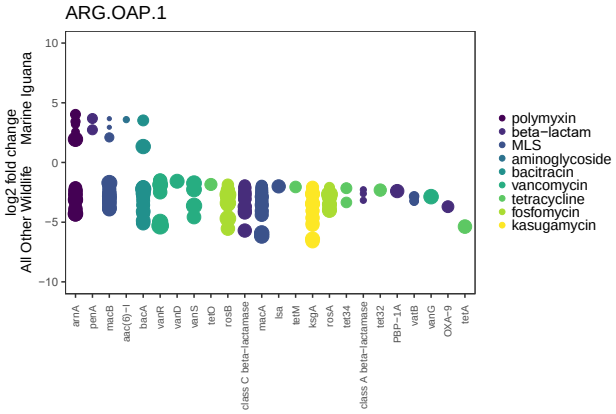

Figure S3.8: *ARGs differentially abundant between marine iguanas and all other wildlife based on ARG-OAP.1.*

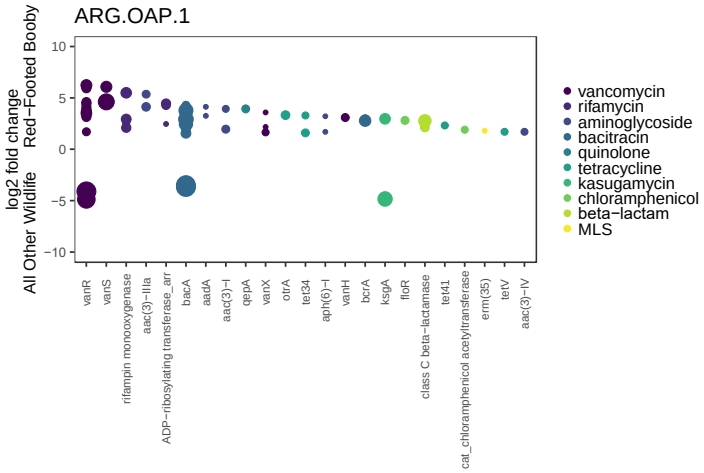

Figure S3.9: *ARGs differentially abundant between red-footed boobies and all other wildlife based on ARG-OAP.1.*

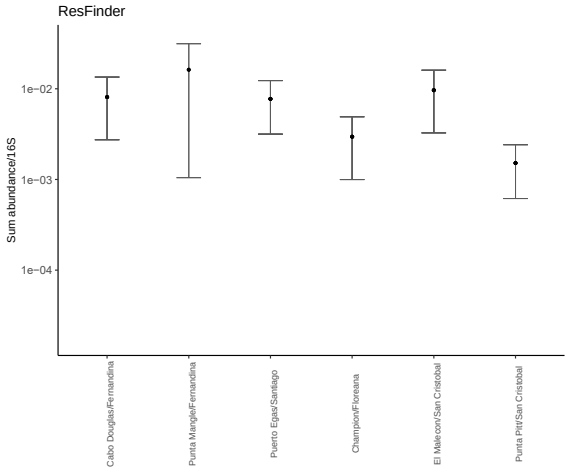

Figure S3.10: *ResFinder ARG sum abundance/16S rRNA for sea lions based on sampling location.* Error bars and black points represent negative binomial GLM-predicted means ± SE.

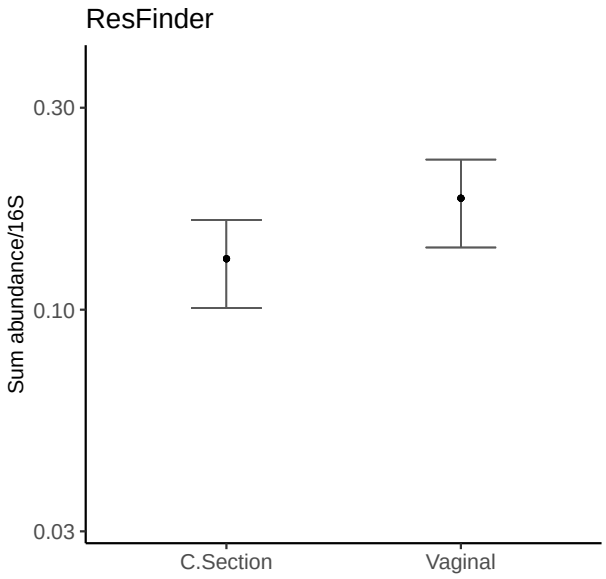

Figure S3.11: *ResFinder ARG sum abundance/16S rRNA for children under two based on mode of delivery*. Error bars and black points represent negative binomial GLM-predicted means ± SE.

Chapter 3: Supplemental Tables

Table S3.1: *Description of sampling locations and sample types collected*

Sampling Location	Island	Description	Sample Types Collected
Municipal WWTP	San Cristobal	Wastewater treatment plant	Influent, effluent
La Toma	San Cristobal	Highlands	Freshwater
Cerro Gato	San Cristobal	Highlands	Freshwater
Playa Carola	San Cristobal	Beach	Marine water, sea turtle cloacal swab
Playa Mann	San Cristobal	Beach	Marine water
Muelle de los Pescadores	San Cristobal	Beach	Marine water
Playa Marinero	San Cristobal	Beach	Marine water
La Loberia	San Cristobal	Beach	Marine water, sea turtle cloacal swab, marine iguana cloacal swab
Los Lobos	Los Lobos	Beach	Marine iguana cloacal swab
Punta Pitt	San Cristobal	Beach	Red footed boobies, sea lion fecal
La Galapaguera	San Cristobal	Captive breeding facility	Giant tortoise fecal
Otoy Ranch	San Cristobal	Reserve	Giant tortoise fecal
Puerto Baquerizo Moreno	San Cristobal	Human samples	Fecal samples from children under two
North Seymour	North Seymour	Inland	Land iguana cloacal swab
Plaza Sur	Plaza Sur	Inland	Land iguana cloacal swab
Santa Fe	Santa Fe	Inland	Land iguana cloacal swab
Punta Mangle	Fernandina	Beach	Sea lion fecal
Cabo Douglas	Fernandina	Beach	Sea lion fecal, fur seal fecal
Puerto Egas	Santiago	Beach	Sea lion fecal
Champion	Floreana	Beach	Sea lion fecal

Table S3.2: *Sequencing metadata data for 90 shotgun metagenomic libraries*

Sample ID	Sample Type	Location	Yr.	Seq. Run	Library Prep.	Sequencing Platform	Total seq. pairs (raw)
G4/GAT-5	freshwater	El Cerro Gato	2017	1	KapaHyper	Illumina HiSeq4000	18940775
G9/TOM-5	freshwater	La Toma	2017	1	KapaHyper	Illumina HiSeq4000	8757552
G1/CA1-5	marine water	Playa Carola	2017	1	KapaHyper	Illumina HiSeq4000	13027801
G2/CA2-5	marine water	Playa Carola	2017	1	KapaHyper	Illumina HiSeq4000	5836789
G3/FIS-5	marine water	Fisherman's Dock	2017	1	KapaHyper	Illumina HiSeq4000	35485713
G6/LO1-5	marine water	Playa Loberia	2017	1	KapaHyper	Illumina HiSeq4000	23850926
G5/LO2-5	marine water	Playa Loberia	2017	1	KapaHyper	Illumina HiSeq4000	32017172
G8/MAR-5	marine water	Playa Marinero	2017	1	KapaHyper	Illumina HiSeq4000	6696779
G7/MAN-5	marine water	Playa Mann	2017	1	KapaHyper	Illumina HiSeq4000	22457295
G10/WEF-5	wastewater	WWTP	2017	1	KapaHyper	Illumina HiSeq4000	16229883
G11/WIN-5	wastewater	WWTP	2017	1	KapaHyper	Illumina HiSeq4000	71217543
G18_12	giant tortoise	La Galapaguera	2018	2	KapaHyper	Illumina HiSeq4000	11986692
G18_14	giant tortoise	La Galapaguera	2018	2	KapaHyper	Illumina HiSeq4000	12075702
G18_18	giant tortoise	La Galapaguera	2018	2	KapaHyper	Illumina HiSeq4000	10883303
G18_19	giant tortoise	La Galapaguera	2018	2	KapaHyper	Illumina HiSeq4000	10247301
G18_62	marine iguana	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	10883409
G18_65	marine iguana	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	10584496
G18_75	marine iguana	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	5172288
G18_76	marine iguana	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	15872415
G18_153	marine iguana	Los Lobos	2018	2	KapaHyper	Illumina HiSeq4000	9804215
G18_155	marine iguana	Los Lobos	2018	2	KapaHyper	Illumina HiSeq4000	11118366
G18_160	marine iguana	Los Lobos	2018	2	KapaHyper	Illumina HiSeq4000	10201497
G18_164	marine iguana	Los Lobos	2018	2	KapaHyper	Illumina HiSeq4000	12601141
G18_194	red foot. booby	Punta Pitt	2018	2	KapaHyper	Illumina HiSeq4000	51353024
G18_195	red foot. booby	Punta Pitt	2018	2	KapaHyper	Illumina HiSeq4000	12555456
G18_196	red foot. booby	Punta Pitt	2018	2	KapaHyper	Illumina HiSeq4000	19978496
G18_198	red foot. booby	Punta Pitt	2018	2	KapaHyper	Illumina HiSeq4000	24259179
G18_170	sea turtle	Playa Carola	2018	2	KapaHyper	Illumina HiSeq4000	14400991
G18_172	sea turtle	Playa Carola	2018	2	KapaHyper	Illumina HiSeq4000	14916914
G18_174	sea turtle	Playa Carola	2018	2	KapaHyper	Illumina HiSeq4000	14158877
G18_177	sea turtle	Playa Carola	2018	2	KapaHyper	Illumina HiSeq4000	12577195
G18_183	sea turtle	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	13826693
G18_184	sea turtle	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	14262139
G18_188	sea turtle	Playa Loberia	2018	2	KapaHyper	Illumina HiSeq4000	14430508
G18_51	freshwater	La Toma	2018	3	KapaHyper	Illumina NextSeq6000	19745052
G18_52	freshwater	El Cerro Gato	2018	3	KapaHyper	Illumina NextSeq6000	16033201
H11	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	22650865
H14	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	18307620

H15	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	26599512
H17	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	18722792
H18	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	16407812
H2	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	15072347
H22	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	17742235
H23	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	13874089
H25	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	17070755
H28	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	17158312
H4	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	18076089
H8	human	San Cristobal	2016	3	KapaHyper	Illumina NextSeq6000	18196029
G19_37	land iguana	NorthSeymour	2018	3	KapaHyper	Illumina NextSeq6000	17532601
G19_26	land iguana	NorthSeymour	2018	3	KapaHyper	Illumina NextSeq6000	17523732
G19_43	land iguana	NorthSeymour	2018	3	KapaHyper	Illumina NextSeq6000	14406833
G19_14	land iguana	PlazaSur	2018	3	KapaHyper	Illumina NextSeq6000	17705384
G19_45	land iguana	PlazaSur	2018	3	KapaHyper	Illumina NextSeq6000	17438406
G19_30	land iguana	PlazaSur	2018	3	KapaHyper	Illumina NextSeq6000	22903967
G19_36	land iguana	PlazaSur	2018	3	KapaHyper	Illumina NextSeq6000	18115124
G19_31	land iguana	SantaFe	2018	3	KapaHyper	Illumina NextSeq6000	16243532
G19_9	land iguana	SantaFe	2018	3	KapaHyper	Illumina NextSeq6000	12752925
G19_34	land iguana	SantaFe	2018	3	KapaHyper	Illumina NextSeq6000	13371053
G18_77	marine water	Playa Carola	2018	3	KapaHyper	Illumina NextSeq6000	18740130
G18_78	marine water	Playa Carola	2018	3	KapaHyper	Illumina NextSeq6000	18395860
G18_30	marine water	Playa Marinero	2018	3	KapaHyper	Illumina NextSeq6000	18161938
G18_81	marine water	Playa Loberia	2018	3	KapaHyper	Illumina NextSeq6000	20065448
G19_113	sea lion	Punta Mangle/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	15535250
G19_114	sea lion	Punta Mangle/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	20606335
G19_121	sea lion	Cabo Douglas/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	17310507
G19_122	sea lion	Cabo Douglas/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	22361972
G19_123	sea lion	Cabo Douglas/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	17524368
G19_124	sea lion	Cabo Douglas/Fernandina	2018	3	KapaHyper	Illumina NextSeq6000	18566447

G19_134	sea lion	Puerto Egas/Santiago	2018	3	KapaHyper	Illumina NextSeq6000	20127768
G19_136	sea lion	Puerto Egas/Santiago	2018	3	KapaHyper	Illumina NextSeq6000	21033623
G19_140	sea lion	Puerto Egas/Santiago	2018	3	KapaHyper	Illumina NextSeq6000	20781649
G19_142	sea lion	Puerto Egas/Santiago	2018	3	KapaHyper	Illumina NextSeq6000	19444644
G19_145	sea lion	Puerto Egas/Santiago	2018	3	KapaHyper	Illumina NextSeq6000	18011502
G19_152	sea lion	Champion/Floreana	2018	3	KapaHyper	Illumina NextSeq6000	15913452
G19_148	sea lion	Champion/Floreana	2018	3	KapaHyper	Illumina NextSeq6000	15337994
G19_159	sea lion	Champion/Floreana	2018	3	KapaHyper	Illumina NextSeq6000	16190737
G19_163	sea lion	Champion/Floreana	2018	3	KapaHyper	Illumina NextSeq6000	18609172
G19_181	sea lion	Punta Pitt/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	18918944
G19_172	sea lion	Punta Pitt/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	20463069
G19_174	sea lion	Punta Pitt/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	18595986
G19_177	sea lion	Punta Pitt/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	18769825
G19_188	sea lion	Punta Pitt/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	18958936
G19_194	sea lion	El Malecon/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	21261779
G19_197	sea lion	El Malecon/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	16155210
G19_199	sea lion	El Malecon/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	18322717
G19_201	sea lion	El Malecon/San Cristobal	2018	3	KapaHyper	Illumina NextSeq6000	15780482
G18_199	wastewater	WWTP	2018	3	KapaHyper	Illumina NextSeq6000	15115714
G18_200	wastewater	WWTP	2018	3	KapaHyper	Illumina NextSeq6000	25252819
G18_3	giant tortoise	La Galapaguera	2018	4	NexteraFlex	Illumina NextSeq6000	230460728
G18_10	giant tortoise	La Galapaguera	2018	4	NexteraFlex	Illumina NextSeq6000	191066264

Table S3.3: *Pairwise comparison of mean Shannon diversity index of ARGs by sample type as annotated by MegaRes.2*

Comparison	**diff**	**lwr**	**upr**	**p adj**
wastewater-human	2.545265	1.0424149	4.0481151	0.000164
water-human	0.1388843	-0.9031547	1.1809234	0.9851983
wildlife-human	-0.3656453	-1.1963769	0.4650863	0.6566307
water-wastewater	-2.4063807	-3.8947101	-0.9180512	0.000341
wildlife-wastewater	-2.9109103	-4.2597595	-1.5620611	0.0000014
wildlife-water	-0.5045296	-1.3086946	0.2996353	0.3588469

Table S3.4: *Pairwise comparison of mean Shannon diversity index of ARGs by sample type as annotated by ResFinder*

Comparison	diff	lwr	upr	p adj
wastewater-human	1.6022702	0.02347665	3.1810638	0.0453569
water-human	-0.6865913	-1.7812877	0.4081051	0.3589634
wildlife-human	-1.752381	-2.6265876	-0.8781743	0.0000072
water-wastewater	-2.2888616	-3.8524008	-0.7253224	0.0013814
wildlife-wastewater	-3.3546512	-4.7725835	-1.936719	0.0000001
wildlife-water	-1.0657897	-1.9121364	-0.2194429	0.0076632

Table S3.5: *Pairwise comparison of mean Simpson diversity index of ARGs by sample type as annotated by ResFinder*

Comparison	diff	lwr	upr	p adj
wastewater-human	0.06800007	-0.3622578	0.49825796	0.975765
water-human	-0.0864024	-0.3847326	0.21192776	0.8719561
wildlife-human	-0.3439959	-0.5822375	-0.1057543	0.0016463
water-wastewater	-0.1544025	-0.5805032	0.27169823	0.7772344
wildlife-wastewater	-0.411996	-0.7984154	-0.0255765	0.0320813
wildlife-water	-0.2575935	-0.4882426	-0.0269444	0.0224654

Table S3.6: *Negative binomial GLM predicted means by sample type as annotated by ARG-OAP.1*

Comparison	Estimate	Std. Error	z value	Pr(>|z|)
wastewater - human == 0	0.75349	0.6914	1.09	0.68047
water - human == 0	-1.46609	0.46383	-3.161	0.00777
wildlife - human == 0	-1.39891	0.37924	-3.689	0.00111
water - wastewater == 0	-2.21958	0.67391	-3.294	0.00465
wildlife - wastewater == 0	-2.1524	0.61874	-3.479	0.00267
wildlife - water == 0	0.06718	0.34633	0.194	0.99721

Table S3.7: *Negative binomial GLM predicted means by sample subtype as annotated by ARG-OAP.1*

Comparison	Estimate	Std. Error	z value	Pr(>|z|)
G. Tortoise - F. Water == 0	-1.590472	0.627258	-2.536	0.23644
Human - F. Water == 0	0.955449	0.560923	1.703	0.78119
L. Iguana - F. Water == 0	0.878877	0.574774	1.529	0.87206
M. Iguana - F. Water == 0	-1.301187	0.595022	-2.187	0.44858
M. Water - F. Water == 0	-0.788123	0.567293	-1.389	0.92516
R. Booby - F. Water == 0	-0.839112	0.687054	-1.221	0.96618
S. Lion - F. Water == 0	-1.198047	0.524726	-2.283	0.38347
S. Turtle - F. Water == 0	-1.600115	0.609064	-2.627	0.19364
W. Water - F. Water == 0	1.708939	0.686967	2.488	0.26114
Human - G. Tortoise == 0	2.54592	0.48592	5.239	< 0.001
L. Iguana - G. Tortoise == 0	2.469348	0.501846	4.921	< 0.001
M. Iguana - G. Tortoise == 0	0.289285	0.524914	0.551	0.99993
M. Water - G. Tortoise == 0	0.802348	0.49326	1.627	0.82449
R. Booby - G. Tortoise == 0	0.751359	0.627316	1.198	0.97029
S. Lion - G. Tortoise == 0	0.392425	0.443644	0.885	0.99659
S. Turtle - G. Tortoise == 0	-0.009643	0.540779	-0.018	1
W. Water - G. Tortoise == 0	3.299411	0.627222	5.26	< 0.001
L. Iguana - Human == 0	-0.076572	0.41597	-0.184	1
M. Iguana - Human == 0	-2.256635	0.443528	-5.088	< 0.001
M. Water - Human == 0	-1.743572	0.40557	-4.299	< 0.001
R. Booby - Human == 0	-1.794561	0.560988	-3.199	0.04197
S. Lion - Human == 0	-2.153496	0.343515	-6.269	< 0.001
S. Turtle - Human == 0	-2.555563	0.462195	-5.529	< 0.001
W. Water - Human == 0	0.753491	0.560883	1.343	0.93865
M. Iguana - L. Iguana == 0	-2.180063	0.46092	-4.73	< 0.001
M. Water - L. Iguana == 0	-1.667	0.424521	-3.927	0.00332
R. Booby - L. Iguana == 0	-1.717989	0.574838	-2.989	0.07684
S. Lion - L. Iguana == 0	-2.076924	0.365695	-5.679	< 0.001
S. Turtle - L. Iguana == 0	-2.478991	0.47891	-5.176	< 0.001
W. Water - L. Iguana == 0	0.830062	0.574735	1.444	0.90652
M. Water - M. Iguana == 0	0.513063	0.451557	1.136	0.97913
R. Booby - M. Iguana == 0	0.462074	0.595084	0.776	0.99876
S. Lion - M. Iguana == 0	0.103139	0.396761	0.26	1
S. Turtle - M. Iguana == 0	-0.298928	0.503032	-0.594	0.99986
W. Water - M. Iguana == 0	3.010126	0.594984	5.059	< 0.001
R. Booby - M. Water == 0	-0.050989	0.567358	-0.09	1
S. Lion - M. Water == 0	-0.409924	0.353821	-1.159	0.97621
S. Turtle - M. Water == 0	-0.811991	0.469905	-1.728	0.76662
W. Water - M. Water == 0	2.497063	0.567253	4.402	< 0.001
S. Lion - R. Booby == 0	-0.358935	0.524795	-0.684	0.99955
S. Turtle - R. Booby == 0	-0.761002	0.609124	-1.249	0.96089
W. Water - R. Booby == 0	2.548051	0.68702	3.709	0.00762
S. Turtle - S. Lion == 0	-0.402068	0.417524	-0.963	0.99355
W. Water - S. Lion == 0	2.906986	0.524682	5.54	< 0.001
W. Water - S. Turtle == 0	3.309054	0.609027	5.433	< 0.001

Table S3.8: *ARGs differentially abundant by birth mode based on annotation with ARG-OAP.1*

Gene Name	log2FoldChange	padj	Subtype	Type	n
Elevated in Cesarean section birth group (positive fold change)					
gi|804348366|emb|CQW48663.1|	6.42087891	1.41E-05	tetracycline_resistance_protein	tetracycline	5
NC_010558.1.6275994.p01	6.26076401	0.00037574	aadA	aminoglycoside	2
XP_001893601	6.21383879	0.00058301	aph(3)-I	aminoglycoside	1
gi|481023147|ref|WP_001295185.1|	6.20388278	0.00034146	class C beta-lactamase	beta-lactam	4
gi|499772318|ref|WP_011453052.1|	6.05991005	0.0001049	sul2	sulfonamide	5
ZP_02900713	5.95920035	0.0001049	bacA	bacitracin	7
gi|504873101|ref|WP_015060203.1|	5.91724451	4.94E-05	sul2	sulfonamide	7
gi|599938140|gb|EYI48463.1|	5.80958839	0.00024408	tetracycline_resistance_protein	tetracycline	3
gi|254966944|gb|ACT97499.1|	5.78843959	0.00015859	tetracycline_resistance_protein	tetracycline	4
gi|657686198|ref|WP_029487032.1|	5.713945	0.00058301	vanT	vancomycin	4
Elevated in vaginal birth group (negative fold change)					
L12710.gene.p01	-5.103788	0.0024441	aac(6)-I	aminoglycoside	2
gi|168258996|gb|ACA23181.1|	-4.3093944	0.00083985	tetW	tetracycline	9
AY004350.gene.p01	-4.1009669	0.00869535	msrC	MLS	2
gi|488247627|ref|WP_002318835.1|	-3.6671787	0.01566969	msrC	MLS	3
gi|488231473|ref|WP_002302681.1|	-3.5633165	0.01969379	msrC	MLS	2
gi|695273519|ref|WP_032495453.1|	-3.5626849	0.00989607	ermB	MLS	8
gi|1028100289|ref|WP_063856424.1|	-3.4069364	0.01009444	tetW	tetracycline	10
CAD13485	-3.359491	0.01130011	tetW	tetracycline	11
gi|1028100281|ref|WP_063856416.1|	-3.3175623	0.00851978	tetracycline_resistance_protein	tetracycline	11
AAY62597	-3.2843882	0.01071307	tetW	tetracycline	12

Table S3.9: *ARGs differentially abundant between land iguanas and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in land iguanas (positive fold change)					
gi|754927849|ref|WP_042284850.1|	8.7238417	7.76E-47	macB	MLS	16
AAK63223	8.71956352	8.95E-51	class A beta-lactamase	beta-lactam	6
gi|983401724|ref|WP_060569629.1|	8.0196864	1.67E-45	class A beta-lactamase	beta-lactam	6
gi|754930806|ref|WP_042287632.1|	7.61217324	1.02E-42	class A beta-lactamase	beta-lactam	6
YP_002240485	7.21609698	3.61E-34	ksgA	kasugamycin	25
gi|749609302|ref|WP_040231863.1|	7.11036953	1.70E-38	macA	MLS	8
gi|693054423|ref|WP_032222656.1|	7.08122232	8.12E-39	macA	MLS	10
B7LVU5	6.31678469	1.55E-29	ksgA	kasugamycin	13
gi|992390012|ref|WP_061077002.1|	5.86821356	1.08E-28	macB	MLS	5
YP_002236533	5.66461317	1.35E-23	bacA	bacitracin	11
Elevated in wildlife that are not land iguanas (negative fold change)					
AJ295238.gene.p01	-2.3692907	0.00406685	tet32	tetracycline	21
YP_581244	-2.2959826	0.02138266	bacA	bacitracin	9
AAZ98835	-2.2698245	0.00940987	vanR	vancomycin	35
YP_264987	-2.1519453	0.01458911	bacA	bacitracin	19
ZP_01072284	-2.1087296	0.01208273	tetM	tetracycline	18
P23835	-1.8933395	0.01923998	tetO	tetracycline	20
ZP_03223548	-1.8123964	0.03203653	tetO	tetracycline	16

ZP_02865688	-1.8091577	0.03367449	bacA	bacitracin	16
ZP_03993170	-1.7871906	0.03536393	tetM	tetracycline	16
ZP_04528247	-1.7241397	0.04078169	bacA	bacitracin	43
AJ295238.gene.p01	-2.3692907	0.00406685	tet32	tetracycline	21

Table S3.10: *ARGs differentially abundant between sea lions and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in sea lions (positive fold change)					
AAZ98835	3.20821548	3.23E-10	vanR	vancomycin	35
ZP_02865688	2.55218699	1.68E-08	bacA	bacitracin	16
ZP_03949893	2.54413893	4.38E-09	bacA	bacitracin	33
ZP_03930663	2.45610836	3.71E-09	bacA	bacitracin	23
gi|545166026|ref|WP_021520619.1|	2.42410355	1.67E-06	macA	MLS	18
AAZ98836	2.36836113	1.67E-06	vanS	vancomycin	35
ZP_02635322	1.97679374	5.30E-06	bacA	bacitracin	18
DQ212986.1.gene4.p01	1.95342817	0.00138923	vanR	vancomycin	42
gi|507059692|ref|WP_016130597.1|	1.86790652	1.79E-06	vanR	vancomycin	26
CAB61229	1.85731449	0.00050539	vanR	vancomycin	26

Table S3.11: *ARGs differentially abundant between giant tortoises and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in giant tortoises (positive fold change)					
YP_001785579	5.99523917	1.14E-38	vatB	MLS	12
YP_002861121	5.95731593	9.13E-69	vatB	MLS	9
YP_001389621	5.61834807	1.01E-55	vatB	MLS	10
U19459.gene.p01	4.98480663	2.23E-30	vatB	MLS	10
gi|405945042|pdb|4FU0|A	4.08634005	4.15E-15	vanG	vancomycin	12
gi|765411304|ref|WP_044689514.1|	3.92071049	1.30E-12	vanG	vancomycin	7
gi|918416459|ref|WP_052467635.1|	3.52687826	6.90E-13	vanW	vancomycin	7
gi|1028100561|ref|WP_063856696.1|	3.23731681	5.01E-11	vanG	vancomycin	14
ABV82118	3.21592969	2.04E-13	tet32	tetracycline	5
YP_001779894	3.19246151	1.20E-14	vatB	MLS	5

Table S3.12: *ARGs differentially abundant between sea turtles and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in sea turtles (positive fold change)					
gi|544902100|ref|WP_021313340.1|	6.91438069	1.28E-29	macB	MLS	8
AY306130.1.gene1.p01	6.52981671	3.75E-24	OXA-50	beta-lactam	3
gi|504416866|ref|WP_014603968.1|	5.65102685	2.50E-20	tetR	tetracycline	2
gi|1016521294|ref|WP_063100554.1|	5.39457357	1.31E-19	macB	MLS	4
YP_581244	5.31564538	1.36E-17	bacA	bacitracin	9
gi|602747670|gb|EYU08172.1|	4.98117747	1.20E-17	tetR	tetracycline	2
YP_264987	4.88730716	3.22E-22	bacA	bacitracin	19
gi|550049059|ref|WP_022580963.1|	4.3187608	1.02E-14	class C beta-lactamase	beta-lactam	2

1XAT	4.1656061	1.80E-15	cat_chloramphenicol acetyltransferase	chloramphenicol	3
gi|446110311|ref|WP_000188166.1|	3.92969202	5.21E-06	macB	MLS	28

Table S3.13: *ARGs differentially abundant between marine iguanas and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in marine iguanas (positive fold change)					
gi|446571412|ref|WP_000648758.1|	4.02839941	2.64E-11	arnA	polymyxin	13
ZP_03164918	3.68495761	0.00015945	penA	beta-lactam	13
gi|983424273|ref|WP_060588471.1|	3.66480358	1.86E-11	macB	MLS	5
AF144880.1.gene1.p01	3.58937286	1.54E-08	aac(6)-I	aminoglycoside	6
YP_002479998	3.52595676	3.82E-05	bacA	bacitracin	16
gi|1002397390|ref|WP_061380650.1|	3.42857282	2.05E-10	arnA	polymyxin	11
gi|554961629|ref|WP_023206204.1|	3.21151815	1.86E-11	arnA	polymyxin	9
gi|554958544|ref|WP_023203571.1|	3.0435794	2.01E-09	arnA	polymyxin	5
gi|446048047|ref|WP_000125902.1|	2.94919289	3.70E-09	macB	MLS	5
ZP_02681785	2.73701681	0.00378072	penA	beta-lactam	13

Table S3.14: *ARGs differentially abundant between red footed boobies and all other wildlife based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
gi|747653988|emb|CEL19270.1|	6.23910712	1.94E-12	vanR	vancomycin	9
KF478993.1.gene7.p01	6.08178668	6.40E-21	vanS	vancomycin	9
gi|943675889|ref|WP_055502070.1|	5.94356268	1.80E-22	vanR	vancomycin	5
gi|544923113|ref|WP_021332509.1|	5.50060303	5.79E-21	rifampin monooxygenase	rifamycin	8
L06161.1.gene1.p01	5.36370898	8.55E-21	aac(3)-IIIa	aminoglycoside	3
AY082011.1.gene2.p1	4.61558605	0.00896513	vanS	vancomycin	27
gi|983429529|ref|WP_060593182.1|	4.58996605	5.79E-21	vanR	vancomycin	5
gi|944023697|ref|WP_055613131.1|	4.46432789	1.25E-19	vanR	vancomycin	5
AF001493.1.orf0.gene.p01	4.4529337	1.41E-16	ADP-ribosylating transferase_arr	rifamycin	5
gi|943670402|ref|WP_055497228.1|	4.39819574	1.48E-13	vanR	vancomycin	2

Table S3.15: *ARGs differentially abundant between freshwater and unimpacted marine water based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in freshwater (positive fold change)					
YP_001797193	9.83169141	1.26E-17	bacA	bacitracin	4
YP_550152	9.39766792	3.74E-26	bacA	bacitracin	4
YP_001155044	9.31456824	7.52E-12	bacA	bacitracin	2
ZP_03552050	8.85174304	1.11E-24	bacA	bacitracin	4
YP_981592	8.46658043	4.76E-22	bacA	bacitracin	4
YP_523088	8.14274018	5.02E-22	bacA	bacitracin	5
ZP_04577926	7.50778839	8.05E-15	bacA	bacitracin	4
YP_970399	7.49584975	2.20E-17	bacA	bacitracin	4
YP_001563294	7.40428542	3.70E-18	bacA	bacitracin	5
AL939114.1.orf1.gene.p01	7.17242248	5.18E-16	ADP-ribosylating transferase_arr	rifamycin	4
Elevated in unimpacted marine water (negative fold change)					
AF353562.gene.p01	-6.423857	8.31E-08	tet35	tetracycline	6
YP_264987	-5.2150054	2.03E-05	bacA	bacitracin	6
BAC58936	-5.063003	2.02E-05	tet34	tetracycline	7
AY082011.1.gene2.p1	-4.9408264	0.0013398	vanS	vancomycin	3
YP_001444376	-4.0630042	0.00082703	tet34	tetracycline	6
ZP_01813310	-3.9205584	0.00496119	tet34	tetracycline	4
YP_581244	-3.6844921	0.00605571	bacA	bacitracin	5
AAZ98835	-3.5324884	0.01531544	vanR	vancomycin	3
YP_002666586	-3.4964212	0.0036566	tet34	tetracycline	6
ZP_04417104	-3.4405668	0.01188437	cat_chloramphenicol acetyltransferase	chloramphenicol	4

Table S3.16: *ARGs differentially abundant between wastewater impacted marine water and freshwater based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in impacted marine water (positive fold change)					
DQ464881.1.gene4.p01	8.21794505	7.85E-09	aph(3)-I	aminoglycoside	4
AY743590.gene.p01	8.13440796	4.56E-06	tet39	tetracycline	3
AF078527.gene.p01	7.74817667	5.49E-06	cmlA	chloramphenicol	3
AF313472.2.gene15.p01	7.06607785	2.25E-06	aph(3)-I	aminoglycoside	4
AY171578.gene.p01	7.03065189	8.03E-05	tetC	tetracycline	3
EU675686.2.gene7.p01	6.67594634	8.59E-06	sul1	sulfonamide	4
AM296481.1.gene2.p01	6.59989926	0.00012815	chloramphenicol exporter	chloramphenicol	3
DQ303918.1.gene1.p01	6.59617784	5.04E-05	aac(6)-I	aminoglycoside	3
NP_511233	6.53914599	0.00010641	tetC	tetracycline	3
AF030945.1.gene1.p01	6.42624355	0.00355813	CARB-6	beta-lactam	2
Elevated in freshwater (negative fold change)					
YP_523088	-8.6573041	1.16E-11	bacA	bacitracin	4
YP_001155044	-7.4076489	0.0014059	bacA	bacitracin	3
AL939114.1.orf1.gene.p01	-7.1724158	2.12E-08	ADP-ribosylating transferase_arr	rifamycin	4
ZP_03541894	-6.9599889	4.77E-07	bacA	bacitracin	4
YP_981592	-6.7661352	1.79E-08	bacA	bacitracin	5
ZP_03552050	-6.6038114	6.01E-09	bacA	bacitracin	6
ABM94007	-6.519619	0.00010641	bacA	bacitracin	3
YP_550152	-6.353268	4.04E-07	bacA	bacitracin	6
YP_001797193	-6.1037629	0.00015595	bacA	bacitracin	6
YP_294981	-5.707349	5.04E-05	bacA	bacitracin	4

Table S3.17: *ARGs differentially abundant between wastewater impacted marine water and unimpacted marine water based on annotation with ARG-OAP.1*

ARG	log2FoldChange	padj	Subtype	Type	n
Elevated in impacted marine water (positive fold change)					
AF078527.gene.p01	7.74818253	6.99E-09	cmlA	chloramphenicol	3
AY171578.gene.p01	7.03065738	3.42E-07	tetC	tetracycline	3
AY162283.2.gene7.p01	6.70389691	6.99E-09	sul1	sulfonamide	4
EU675686.2.gene7.p01	6.67595035	7.05E-09	sul1	sulfonamide	4
DQ303918.1.gene1.p01	6.59618227	1.22E-07	aac(6)-I	aminoglycoside	3
NP_511233	6.53915068	4.10E-07	tetC	tetracycline	3
AF030945.1.gene1.p01	6.42625448	1.08E-05	CARB-6	beta-lactam	2
gi|446289936|ref|WP_000367791.1|	6.28539434	1.13E-06	cmlA	chloramphenicol	3
AM087405.1.gene3.p01	6.19474984	4.19E-07	aadA	aminoglycoside	3
AJ809407.1.gene1.p01	6.17491849	6.31E-07	aadA	aminoglycoside	3
Elevated in unimpacted marine water (negative fold change)					
gi|486458841|gb|EOE03251.1|	-3.6520692	0.0144674	aadE	aminoglycoside	3
YP_264987	-2.8226916	0.00866032	bacA	bacitracin	9
GQ205627.2.gene3.p01	-2.7776026	0.04874339	vatG	MLS	3
YP_001476958	-2.4556767	0.01866334	ksgA	kasugamycin	8
EAS48788	-2.4405689	0.04406754	bacA	bacitracin	6
ABD46539	2.32192574	0.04997735	tetM	tetracycline	2
ABC18245	2.39231484	0.04997735	tetW	tetracycline	2
ACI02041	2.39231484	0.04997735	tetW	tetracycline	2
gi|1004359922|gb|AMP42228.1|	2.39231484	0.04997735	tetracycline_resistance_protein	tetracycline	2
gi|226044|prf||1408188A	2.39231484	0.04997735	tetO	tetracycline	2

Table S3.18: *Negative binomial GLM predicted means by sample type as annotated by ResFinder*

Comparison	Estimate	Std. Error	z value	Pr(>|z|)
wastewater - human == 0	0.3446	0.804	0.429	0.97179
water - human == 0	-1.8542	0.5575	-3.326	0.00412
wildlife - human == 0	-2.9059	0.4468	-6.503	< 0.001
water - wastewater == 0	-2.1988	0.7962	-2.761	0.02692
wildlife - wastewater == 0	-3.2504	0.7231	-4.495	< 0.001
wildlife - water == 0	-1.0516	0.4327	-2.43	0.06596

Table S3.19: *Negative binomial GLM predicted means by sample subtype as annotated by ResFinder*

Comparison	Estimate	Std. Error	z value	Pr(>|z|)
giant tortoise - freshwater == 0	1.41487	0.76237	1.856	0.683
human - freshwater == 0	5.04094	0.68208	7.391	< 0.001
land iguana - freshwater == 0	3.28325	0.69883	4.698	< 0.001
marine iguana - freshwater == 0	1.5121	0.7921	1.909	0.64584
marine water - freshwater == 0	3.54164	0.70975	4.99	< 0.001
red footed booby - freshwater == 0	0.70037	0.83536	0.838	0.99773
sea lion - freshwater == 0	1.2142	0.64262	1.889	0.65947
sea turtle - freshwater == 0	2.17525	0.83453	2.607	0.20229
wastewater - freshwater == 0	5.38549	0.83429	6.455	< 0.001
human - giant tortoise == 0	3.62608	0.58889	6.157	< 0.001
land iguana - giant tortoise == 0	1.86838	0.60821	3.072	0.06098
marine iguana - giant tortoise == 0	0.09723	0.71343	0.136	1
marine water - giant tortoise == 0	2.12677	0.62073	3.426	0.0204
red footed booby - giant tortoise == 0	-0.7145	0.76118	-0.939	0.99463
sea lion - giant tortoise == 0	-0.20067	0.54271	-0.37	1
sea turtle - giant tortoise == 0	0.76038	0.76026	1	0.99144
wastewater - giant tortoise == 0	3.97062	0.76	5.224	< 0.001
land iguana - human == 0	-1.75769	0.50393	-3.488	0.01627
marine iguana - human == 0	-3.52885	0.6269	-5.629	< 0.001
marine water - human == 0	-1.49931	0.51896	-2.889	0.1013
red footed booby - human == 0	-4.34058	0.68074	-6.376	< 0.001
sea lion - human == 0	-3.82675	0.42255	-9.056	< 0.001
sea turtle - human == 0	-2.8657	0.67972	-4.216	< 0.001
wastewater - human == 0	0.34455	0.67943	0.507	0.99996
marine iguana - land iguana == 0	-1.77115	0.64509	-2.746	0.14549
marine water - land iguana == 0	0.25838	0.54079	0.478	0.99998
red footed booby - land iguana == 0	-2.58288	0.69752	-3.703	0.00752
sea lion - land iguana == 0	-2.06905	0.44908	-4.607	< 0.001
sea turtle - land iguana == 0	-1.10801	0.69653	-1.591	0.84228
wastewater - land iguana == 0	2.10224	0.69624	3.019	0.07089
marine water - marine iguana == 0	2.02954	0.6569	3.09	0.05759
red footed booby - marine iguana == 0	-0.81173	0.79095	-1.026	0.98967
sea lion - marine iguana == 0	-0.2979	0.58373	-0.51	0.99996
sea turtle - marine iguana == 0	0.66315	0.79007	0.839	0.9977
wastewater - marine iguana == 0	3.8734	0.78982	4.904	< 0.001
red footed booby - marine water == 0	-2.84127	0.70846	-4.01	0.00222
sea lion - marine water == 0	-2.32744	0.46589	-4.996	< 0.001
sea turtle - marine water == 0	-1.36639	0.70748	-1.931	0.63011
wastewater - marine water == 0	1.84386	0.7072	2.607	0.20218
sea lion - red footed booby == 0	0.51383	0.6412	0.801	0.99841
sea turtle - red footed booby == 0	1.47488	0.83343	1.77	0.73988
wastewater - red footed booby == 0	4.68512	0.8332	5.623	< 0.001
sea turtle - sea lion == 0	0.96105	0.64012	1.501	0.88359
wastewater - sea lion == 0	4.17129	0.63981	6.52	< 0.001
wastewater - sea turtle == 0	3.21025	0.83236	3.857	0.00409

Table S3.20: *ARGs differentially abundant by birth mode based on annotation with ResFinder*

ARG	Log2FoldChange	P adjusted	Class	Gene	n
Elevated in Caesarean section birth group (positive fold change)					
VanC1XY_1_AF162694	6.21942607	0.00091225	VanC1XY	Vancomycin (Glycopeptid)	3
tet(B)_2_AF326777	5.83333763	3.96E-05	tet(B)	Tetracycline	5
aadA5_1_AF137361	5.66198623	0.00113051	aadA5	Aminoglycoside	2
aph(3)-Ia_1_V00359	5.45239465	0.0027647	aph(3')-Ia	Aminoglycoside	1
mph(A)_2_U36578	5.22728245	0.00091225	mph(A)	Macrolide	4
mph(A)_1_D16251	5.22196935	0.00112568	mph(A)	Macrolide	5
tet(B)_1_AP000342	5.01541604	0.00071162	tet(B)	Tetracycline	4
mdf(A)_1_Y08743	4.86161387	0.00197744	mdf(A)	Phenicol	10
sul2_9_FJ197818	4.70302563	0.00091225	sul2	Sulphonamide	6
sul2_8_AJ877041	4.65397421	0.00197744	sul2	Sulphonamide	4
Elevated in vaginal birth group (negative fold change)					
msr(C)_1_AY004350	-4.7983117	0.00942462	msr(C)	Macrolide	3
aac(6)-Ii_1_L12710	-4.3341134	0.01088402	aac(6')-Ii	Aminoglycoside	2
erm(B)_21_U35228	-3.7265255	0.012844	erm(B)	Macrolide	4
erm(B)_18_X66468	-3.3728566	0.02520042	erm(B)	Macrolide	6
erm(B)_10_U86375	-3.2669872	0.03549519	erm(B)	Macrolide	5
tet(W)_3_AJ427421	-3.2300753	0.01711745	tet(W)	Tetracycline	12
tet(W)_5_AJ427422	-3.130482	0.02520042	tet(W)	Tetracycline	12
tet(Q)_1_L33696	-3.1067037	0.03644875	tet(Q)	Tetracycline	6
tet(W)_1_DQ060146	-3.0960117	0.02488438	tet(W)	Tetracycline	12
ant(3)-Ia_1_X02340	3.04048397	0.02188336	ant(3")-Ia	Aminoglycoside	6

Table S3.21: *ARGs differentially abundant between giant tortoises and all other wildlife based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
Elevated in giant tortoises (positive fold change)					
tet(32)_1_EU722333	4.60486125	7.66E-26	tet(32)	Tetracycline	5
tet(W)_5_AJ427422	3.10231555	7.64E-13	tet(W)	Tetracycline	10
tet(39)_1_KT346360	3.05889288	9.46E-08	tet(39)	Tetracycline	1
tet(O)_3_Y07780	2.96151014	6.34E-06	tet(O)	Tetracycline	14
tet(W)_2_AY049983	2.75216758	1.70E-10	tet(W)	Tetracycline	6
tet(O)_1_M18896	2.55596585	0.03416918	tet(O)	Tetracycline	17
tet(Q)_1_L33696	2.47190516	0.01033071	tet(Q)	Tetracycline	12
cfr(C)_2_CANB01000378	2.45943113	1.12E-06	cfr(C)	Phenicols	2
tet(W)_3_AJ427421	2.43246409	1.93E-06	tet(W)	Tetracycline	9
tet(O/32/O)_1_JQ740052	2.43108284	4.22E-06	tet(O/32/O)	Tetracycline	16

Table S3.22: *ARGs differentially abundant between land iguanas and all other wildlife based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
Elevated in land iguanas (positive fold change)					
blaSED1_1_AF321608	9.47937288	3.30E-57	blaSED1	Beta-lactam	6
oqxB_1_EU370913	6.54898792	6.71E-23	oqxB	Quinolone	8
oqxA_1_EU370913	4.34936408	1.24E-14	oqxA	Quinolone	7
blaMAL-1_1_AJ277209	4.00898775	2.21E-13	blaMAL-1	Beta-lactam	3
mdf(A)_1_Y08743	3.00254514	0.03310372	mdf(A)	Phenicol	31
blaACT-7_1_FJ237368	1.84799652	0.00143852	blaACT-7	Beta-lactam	2
fosA5_1_EU195449	1.76553443	0.00117379	fosA5	Fosfomycin	2
blaACT-15_1_JX440356	1.63226797	0.0036768	blaACT-15	Beta-lactam	3
cepA_6_FR688022	1.63226795	0.00143852	cepA	Beta-lactam	3
blaMAL-1_2_AJ609506	1.53605263	0.00600042	blaMAL-1	Beta-lactam	3

Table S3.23: *ARGs differentially abundant between sea turtles and all other wildlife based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
blaPAO_4_AY083592	6.44530814	3.29E-24	blaPAO	Beta-lactam	2
blaOXA-SHE_1_AY066004	5.19142343	4.56E-11	blaOXA-SHE	Beta-lactam	5
qnrA5_1_DQ058663	1.60131824	0.03397546	qnrA5	Quinolone	3

Table S3.24: *ARGs differentially abundant between marine iguanas and all other wildlife based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
lnu(P)_1_FJ589781	4.25738683	1.46E-16	lnu(P)	Lincosamide	3
aac(6)-Iaa_1_NC_003197	2.56550959	7.34E-07	aac(6')-Iaa	Aminoglycoside	7

Table S3.25: *ARGs differentially abundant between wastewater-impacted marine waters and freshwater based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
mph(E)_1_DQ839391	8.47571607	1.16E-06	mph(E)	Macrolide	3
tet(39)_1_KT346360	8.12539444	5.44E-06	tet(39)	Tetracycline	3
mph(E)_2_JF769133	7.86107122	3.01E-06	mph(E)	Macrolide	3
cmlA1_1_M64556	7.72449754	5.44E-06	cmlA1	Phenicol	3
tet(Q)_1_L33696	7.68823525	3.41E-06	tet(Q)	Tetracycline	3
tet(C)_2_AY046276	7.46962589	1.35E-05	tet(C)	Tetracycline	3
blaCARB-4_1_FJ785525	7.4135902	0.00187161	blaCARB-4	Beta-lactam	1
tet(C)_3_AF055345	7.0980167	4.34E-05	tet(C)	Tetracycline	3
mef(C)_1_AB571865	6.88568232	4.34E-05	mef(C)	Macrolide	3
msr(E)_1_FR751518	6.87133005	3.01E-06	msr(E)	Macrolide	6

Table S3.26: *ARGs differentially abundant between wastewater-impacted marine waters and unimpacted marine waters based on annotation with ResFinder*

ARG	log2FoldChange	padj	subtype	type	n
cmlA1_1_M64556	7.72450346	4.22E-09	cmlA1	Phenicol	3
tet(C)_2_AY046276	7.46963166	1.72E-08	tet(C)	Tetracycline	3
blaCARB-4_1_FJ785525	7.4136148	3.26E-05	blaCARB-4	Beta-lactam	1
tet(C)_3_AF055345	7.09802225	9.36E-08	tet(C)	Tetracycline	3
mef(C)_1_AB571865	6.88568744	9.36E-08	mef(C)	Macrolide	3
mph(A)_2_U36578	6.8297143	7.29E-08	mph(A)	Macrolide	3
ant(3)-Ia_1_X02340	6.43878376	3.55E-07	ant(3")-Ia	Aminoglycoside	3
aph(6)-Id_1_M28829	6.43462225	4.22E-09	aph(6)-Id	Aminoglycoside	4
aph(3)-Ib_5_AF321551	6.38800984	1.71E-07	aph(3")-Ib	Aminoglycoside	3

Table S3.27: *Citrobacter taxonomic assignments from Metaxa2*

Sample ID	Sample Type	Location	C. freundii	C. rodentium ICC168	C. sedlakii	Citrobacter sp. I91-5	Citrobacter sp. SL-DB7	Citrobacter sp. T1	Unclassified Citrobacter
G17_1	Marine Water*	Playa Carola 1	0	0	0	0	0	0	**1**
G18_77	Marine Water*	Playa Carola 1	0	0	0	0	0	0	**2**
H15	Human	San Cristobal	**1**	0	0	0	0	0	0
H22	Human	San Cristobal	**4**	0	0	0	0	0	**2**
G19_37	Land Iguana	North Seymour	0	**427**	**34**	0	**1**	**3**	**2499**
G19_26	Land Iguana	North Seymour	0	**5**	**1**	0	0	0	**26**
G19_43	Land Iguana	North Seymour	0	**150**	**17**	0	0	**2**	**710**
G19_14	Land Iguana	Plaza Sur	0	0	0	0	0	0	**1**
G19_45	Land Iguana	Plaza Sur	0	**1**	0	0	0	0	0
G19_30	Land Iguana	Plaza Sur	0	0	0	0	0	0	0
G19_36	Land Iguana	Plaza Sur	0	0	0	0	0	0	**2**
G19_177	Sea Lion	Punta Pitt	**2**	0	0	**1**	0	0	0
G19_121	Sea Lion	Cabo Douglas	0	0	0	0	0	**1**	0
G19_124	Sea Lion	Cabo Douglas	0	0	0	0	0	0	**63**
G19_134	Sea Lion	Puerto Egas	0	0	0	0	0	**1**	0
G19_31	Land Iguana	Santa Fe	0	**217**	**22**	0	0	**1**	**1209**
G19_9	Land Iguana	Santa Fe	0	**55**	**5**	0	0	0	**495**
G19_34	Land Iguana	Santa Fe	0	**292**	**49**	0	0	**3**	**3283**
G18_3	Giant Tortoise	Galapaguera	0	0	0	0	0	**1**	**1**
G18_116	WWTP influent	WWTP	**1**	0	0	0	0	0	**2**

CHAPTER 4: WHERE ARE THE ANTIBIOTIC REISTANCE GENES? CHARACTERIZATION OF THE MOBILIOME

Introduction

The dissemination of antibiotic resistance through horizontal gene transfer (HGT) represents a challenging dimension to mitigating AMR (Ellabaan et al., 2021). Antibiotic resistant bacteria not only have the potential to survive therapeutic attacks and replicate, thereby increasing the proportion of resistance in a given population, but also present with the capacity to disseminate AMR determinants in certain conditions. Mobile genetic elements (MGEs), including transposases, insertion sequences, integrons and their associated cassettes, as well as the plasmids and integrative conjugative elements in which these elements are often embedded, are integral to the dissemination of ARGs through HGT (reviewed in Partridge et al., 2018). Horizontal gene transfer of ARGs via MGEs has been documented in the human gut microbiome (McInnes et al., 2020; Kent et al., 2020) and wastewater treatment plants (Che et al., 2019). Moreover, the correlation between ARGs and MGEs is well established in a range of environments, including wastewater treatment plants (Makowska et al., 2016; Yin et al., 2019, Zheng et al., 2020), the infant gut microbiome (Pärnänen et al., 2018), and pharmaceutical waste bioreactors (Tao et al., 2016), with one recent study suggesting that MGEs, rather than fecal pollution, drive the continued proliferation of AMR bacteria in a river ecosystem (Lee et al., 2020). In the context of environmental AMR, the extent to which human-associated, clinically relevant pathogens exchange resistance determinants with endogenous bacteria – and, the extent to which environmental bacteria may serve as novel sources of ARGs – remains an ongoing area

of study. Regardless of the scale or direction of gene flow, MGEs are thought to play a critical role in these processes.

One mobile genetic element of particular relevance to the horizontal dissemination of ARGs is the class I integron (reviewed in Gillings et al, 2008; Zhang et al., 2018). While integrons themselves are not ARGs, they facilitate the collection of gene cassettes conferring resistance to antibiotics and other chemical stressors, such as disinfectants and heavy metals (Gillings et al., 2015). Integrons may be located chromosomally or embedded in transposons and/or plasmids, such as the Tn*402* transposon (Post et al., 2007). Integrons are comprised of three essential components: the integron-integrase gene *intI*, which allows for the incorporation of exogenous DNA into the integron; a recombination site (*attI*) where new gene cassettes are integrated, and a promotor (P_C) which drives the expression of genes within the integron (Gillings et al., 2014). The class I integron in particular is characterized by a variable array of gene cassettes bookended by a 5' conserved segment (CS) which contains *intI1*, P_C, and *attI*, and the 3' CS which typically contains the genes *qacEΔ1* and *sul1* (Yang et al., 2021). Over the course of the last century, class I integrons have proliferated among human-associated microorganisms to the extent that detection of *intI1* in the environment is now considered a marker of anthropogenic influence (Gillings et al., 2105). As class I integrons have been found to correlate with ARGs (Makowska et al., 2016, Zheng et al., 2020), they are often targeted alongside ARGs in molecular assays in resistome characterization studies. Importantly, class I integrons recovered from clinical isolates have nearly perfect sequence identify in the *intI* gene, allowing for discrimination between anthropogenic and more diverse class I integrons of environmental origin.

While detection of *intI1* represents a powerful and strategic tool in characterizing anthropogenic impacts on the environmental resistome, it represents just one of the many mobile genetic elements involved in ARG transmission. Recently, Pärnänen and colleagues (2018) developed a database of mobile genetic elements suitable for identifying MGEs from metagenomic sequencing reads in a high throughput manner with the mapping tool Bowtie2 (Langmead and Salzberg, 2012). In the present study, we used a hybrid approach to characterize the mobilomes of the 90 metagenomic libraries described in Chapter 3, pairing broad characterization of MGEs using the Pärnänen et al. (2018) database with a novel ddPCR assay discriminating between clinical and environmental variants of the class I integron-integrase gene, *intI1*. Our ddPCR survey included more than 250 samples originating from water, wildlife, and humans in the Galapagos islands. Taken together, this data set constitutes the first exploration of the mobilome in Galapagos environmental and animal reservoirs.

Materials and Methods

ddPCR assay design and optimization

We aimed to design a ddPCR assay to distinguish between clinical and environmental variants of the class I integron-integrase gene. In proposing *intI1* as an environmental marker of anthropogenic pollution, Gillings et al. (2015) pointed to primer pair intI1F165/intIR476, which generates a 311 bp product, to specifically target the clinical variant (blue region, **Figure 4.1**). Waldron and Gillings (2015) demonstrated the utility of this assay in identifying clinical *intI1* variants in foodstuffs following initial amplification with the primer pair HS463a/HS464 (Stokes et al., 2006) which generates a 473 bp product. Waldron and Gillings' work (2015) proposed a paradigm in which primer pair HS463a/HS464 will amplify both clinical and environment variants of *int1,* while intI1F165/intIR476 can subsequently distinguish those more likely to be of clinical origin. Both assays, however, produce amplicon products above the 60-200 bp size

recommended by BioRad for use in ddPCR (https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf), and neither has a published probe sequence for use in probe-based PCR assays. Barraud et al. (2010) published a probe-based, 196 bp amplicon *intI1* assay for RT-qPCR which has recently been adapted to ddPCR platforms (Wang et al., 2018; Dungan and Bjorneberg, 2020). The forward primer of this assay, which is given relative to the antisense strand, partially overlaps with intIR476, the reverse primer of the proposed clinical *intI1* assay (Gillings et al. 2015). As such, the Barraud et al. (2010) amplicon product corresponds to the 3' end of the clinical *intI1* variant. This product is indicated between the green regions highlighted in **Figure 4.1**, while the probe for this assay is highlighted in orange. Both primer pairs (green and blue) can be detected in a range *intI1* sequences from clinical and environmental origin, though environmental sequences have only the forward or reverse primer in some cases (**Table 4.1**). The probe, however, is highly specific to the clinical *intI1* variant. Due to this specificity and the 196-bp product size, we selected the Barraud et al. (2010) assay for use in discriminating *intI1* variants of anthropogenic origin. To amplify class I integron-integrases of both clinical and environmental origin (henceforth called the general variant), we designed a novel assay targeting the 5' region upstream of the clinical sequence. MEGA X software (Kumar et al., 2018) was used to identify homologous regions in clinical integron-integrase sequences from clinical and environmental origins (**Figure S4.1,** accession numbers provided in **Table 4.1**). Candidate primers and probes were checked for self-complementarity and estimated T_m in OligoCalc (Kibbe, 2007). Our forward primer matched HS464 (Stokes et al., 2006), but three nucleotides were removed from the 3' end to reduce the

T_m relative to the probe sequence. The final assay yielded a 77 bp amplicon corresponding to the product between the purple highlighted region in **Figure 4.1. Table 4.2** reports all primer and probe sequences used in the study.

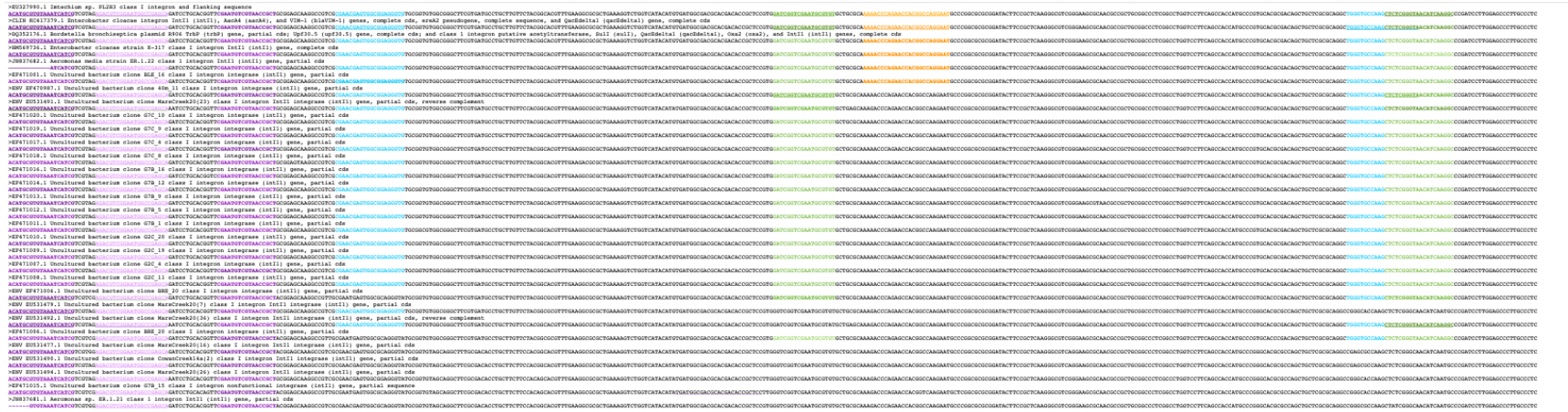

Figure 4.1: *Visual representation of regions amplified by primers targeting general and clinical class 1 integron-integrase variants.* The first five sequences (top) correspond to the clinical class I integron sequence, while the remaining sequences are sourced from primarily uncultured environmental bacteria. Accession codes for these sequences are reported below in **Table 4.1**. Regions highlighted in blue correspond to the clinical *intI1* primer pair proposed by Gillings et al. (2015). Regions highlighted in green and orange correspond to the primers and probe, respectively, for the Barraud et al. (2010) assay. Regions highlighted in dark and light purple correspond to the primers and probe, respectively, for the general *intI1* assay developed in the present study. Based on this analysis, the Barraud et al. (2010) primers should strictly amplify the clinical class I integron-integrase sequence, while the general *intI1* assay should yield amplification products across a range of clinical and environmental *intI1* sequences.

Table 4.1: *Presence and absence of intI1 primers and probe sequences among a selection of clinical and environmental class I integron-integrase sequences.*

		General *intI1*			Clinical *intI1*				
		Stokes et al. 2006	This study		Barraud et al. 2010			Gillings et al. 2015	
Accession No.	**Description**	FP HS464*	Probe	RP	FP	Probe	RP	FP IntI1F165	RP IntIR476
EU327990.1	Imtechium sp. PL2H3 class I integron and flanking sequence	✓	✓	✓	✓	✓	✓	✓	✓
KC417379.1	Enterobacter cloacae integron IntI1 (intI1), AacA4 (aacA4), and VIM-1 (blaVIM-1) genes, complete cds, ereA2 pseudogene, complete sequence, and QacEdelta1 (qacEdelta1) gene, complete cds	✓	✓	✓	✓	✓	✓	✓	✓
DQ352176.1	Bordetella bronchiseptica plasmid R906 TrbP (trbP) gene, partial cds; Upf30.5 (upf30.5) gene, complete cds; and class 1 integron putative acetyltransferase, SulI (sulI), QacEdelta1 (qacEdelta1), Oxa2 (oxa2), and IntI1 (intI1) genes, complete cds	✓	✓	✓	✓	✓	✓	✓	✓
HM569736.1	Enterobacter cloacae strain K-317 class I integron IntI1 (intI1) gene, complete cds	✓	✓	✓	✓	✓	✓	✓	✓
JN837682.1	Aeromonas media strain ER.1.22 class 1 integron IntI1 (intI1) gene, partial cds	✓	✓	✓	✓	✓	✓	✓	✓
EF471001.1	Uncultured bacterium clone BLE_16 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓	✓	✓	✓	✓
EF470987.1	Uncultured bacterium clone 40m_11 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EU531491.1	Uncultured bacterium clone MarsCreek20(23) class I integron IntI1 integrase (intI1) gene, partial cds, reverse complement	✓	✓	✓	✓		✓	✓	✓
EF471020.1	Uncultured bacterium clone G7C_10 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471019.1	Uncultured bacterium clone G7C_9 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471017.1	Uncultured bacterium clone G7C_4 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471018.1	Uncultured bacterium clone G7C_8 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓

EF471016.1	Uncultured bacterium clone G7B_16 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471014.1	Uncultured bacterium clone G7B_12 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471013.1	Uncultured bacterium clone G7B_9 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471012.1	Uncultured bacterium clone G7B_5 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471011.1	Uncultured bacterium clone G7B_1 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471010.1	Uncultured bacterium clone G2C_20 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471009.1	Uncultured bacterium clone G2C_19 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471007.1	Uncultured bacterium clone G2C_4 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471008.1	Uncultured bacterium clone G2C_11 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓	✓	✓
EF471006.1	Uncultured bacterium clone BRE_20 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓		✓
EU531479.1	Uncultured bacterium clone MarsCreek20(7) class I integron IntI1 integrase (intI1) gene, partial cds	✓	✓	✓	✓			✓	✓
EU531492.1	Uncultured bacterium clone MarsCreek20(36) class I integron IntI1 integrase (intI1) gene, partial cds, reverse complement	✓	✓	✓	✓			✓	✓
EF471006.1	Uncultured bacterium clone BRE_20 class I integron integrase (intI1) gene, partial cds	✓	✓	✓	✓		✓		✓
EU531477.1	Uncultured bacterium clone MarsCreek20(16) class I integron IntI1 integrase (intI1) gene, partial cds	✓	✓	✓					
EU531490.1	Uncultured bacterium clone CowanCreek14a(2) class I integron IntI1 integrase (intI1) gene, partial cds	✓	✓	✓					
EU531494.1	Uncultured bacterium clone MarsCreek20(26) class I integron IntI1 integrase (intI1) gene, partial cds	✓	✓	✓					
EF471015.1	Uncultured bacterium clone G7B_15 class I integron nonfunctional integrase (intI1) gene, partial sequence	✓	✓	✓					
JN837681.1	Aeromonas sp. ER.1.21 class 1 integron IntI1 (intI1) gene, partial cds	✓	✓	✓					

Table 4.2: *Primer and probe sequences for intI1 and 16S rRNA assays used in this study.*

Target	**Description**	**Sequence** (5'→ 3')	**Reference**
General class I integron (77 bp)	Forward Primer (modification of HS464)	ACATGCGTGTAAATCATCG	Stokes et al., 2006
	Reverse Primer	AGCGGTTACGACATTCG	This study
	Probe	FAM-AGACGTCGGAATGGCCGAGCA-BHQ-1	
Clinical class I integron (196 bp)	Forward Primer	GCCTTGATGTTACCCGAGAG	Barraud et al., 2010
	Reverse Primer	GATCGGTCGAATGCGTGT	
	Probe	FAM-ATTCCTGGCCGTGGTTCTGGGTTTT-BHQ-1	
16S rRNA (466 bp)	Forward Primer	TCCTACGGGAGGCAGCAGT	Nadkarni et al., 2002
	Reverse Primer	GGACTACCAGGGTATCTAATCCTGTT	
	Probe	HEX-CGTATTACCGCGGCTGCTGGCAC-BHQ-1	

Annealing temperature optimization was performed for each target using the respective positive control. Both class I integron targets demonstrated increased separation with annealing temperatures under 60°C, with 58°C selected as the optimal condition (**Figure S4.2, S4.3).** With a larger amplicon size, the 16S rRNA assay showed improved separation with increasing temperature, and 61.8°C was chosen as the annealing temperature (**Figure S4.4**). To further improve separation of the 16S rRNA target, the annealing/extension time was increased to 2 minutes and the number of cycles was increased to 45. Final cycling conditions for each target are presented in **Table 4.3**.

Table 4.3: *ddPCR cycle conditions for intI1 and 16S rRNA targets.*

Target	Enzyme activation	Denaturation	Annealing /Extension	Cycles	Enzyme deactivation	Signal stabilization
General *intI1*	95°C, 5 min	94°C, 30 sec	58°C, 60 sec	40	95°C, 10 min	4°C, 30 min
Clinical *intI1*						
16S rRNA		94°C, 30 sec	61.8°C, 2 min	45		

ddPCR quantification of intI1 and 16S rRNA in Galapagos samples

The general class I integron-integrase, clinical class I integron-integrase (Barraud et al., 2010), and 16S rRNA gene (Nadkarni et al., 2002) were quantified using the Bio-Rad QX200 Droplet Digital PCR System in >250 samples from the Galapagos, including wildlife, wastewater, freshwater, marine water, and fecal DNA extracts from children under two years of age. ddPCR reaction mixtures contained 11 uL Bio-Rad ddPCR Supermix for Probes (No dUTP, #1863024), 3 uL primer/probe mixture (final concentration 900 nM for each primer and 250 nM for probe), 5 uL nuclease free water, and 3 uL DNA template or control for a total prepared reaction volume of 22 uL. For the integron assays, samples were run undiluted with the exception of wastewater samples which were diluted 1:100 or 1:1,000. For the 16S rRNA assay, samples were generally diluted 1:1,000 with the exception of some low-DNA samples run at 1:100 and wastewater samples which were run at 1:10,000 to 1:100,000 for this target. Extraction blanks were analyzed undiluted for all targets. A 514-bp double-stranded gBlocks® gene fragment spanning nucleotide position 46 through 560 (5'→ 3') of an *Enterobacter cloacae* class I integron sequence (NCBI accession# KC417379.1) was purchased from Integrated DNA Technologies (IDT, Coralville, Iowa) for use as the positive control for both the general and clinical class I integron. For the 16S rRNA positive control, *E. coli* strain ATCC 25922 was

grown up overnight on tryptic soy agar (TSA) at 37°C. A small mass of cells was transferred into 100 uL nuclease free water using a flame-sterilized loop and boiled at 100°C for 20 minutes. The mixture was then centrifuged at 10,000 g for 1 minute and the DNA concentration of the supernatant was quantified using a Qubit4 fluorometer. The supernatant was diluted 1:1,1000, aliquoted into single-use volumes, and stored at -20°C. When running the 16S rRNA assay, a single aliquot of positive control was thawed and further diluted to 1:100,000. For negative controls, 3 uL nuclease free water was used in the place of DNA template. All samples were analyzed in duplicate with the exception of negative controls which were run in quadruplicate for each 96-well plate. PCR reaction mixtures were briefly vortexed and spun-down in a Mini-Plate centrifuge spinner (Fisher, Hampton, New Hampshire) for 30 seconds.

Using a multichannel pipette, 20 uL of the prepared 22 uL of each reaction mixture was loaded into an 8-well cartridge for droplet generation on the BioRad QX200 droplet generator along with 70 μL of droplet generation oil. The cartridge was covered with a gasket and placed in the droplet generator per manufacturer instructions. Generated droplets were transferred to a skirted 96-well plate, which was sealed in a Bio-Rad PX1 PCR Plate Sealer. PCR was performed in a deep well BioRad C1000 Touch Thermal Cycler using the conditions described in **Table 4.3** for each target.

ddPCR and statistical analysis

The BioRad QuantaSoft Analysis Pro version 1.0.596 software was used to calculate droplet count and gene-copy concentrations. The fluorescence range was calculated as the difference between the mean fluorescence amplitude of the positive control wells and the mean fluorescence amplitude of the negative control wells. A unique positivity threshold corresponding to the midpoint of the fluorescence range was established for each run. Duplicate

reactions were merged and a sample was considered positive if three or more droplets measured above the positivity threshold. Across the data set, merged droplet counts ranged from 17,452 to 41,271; 22,195 to 39,742; and 24,988 to 41,810 for the general integron, clinical integron, and 16S rRNA targets, respectively, indicating successful droplet generation. Resulting concentrations as copies/uL in the ddPCR reaction were corrected for dilution of the template as appropriate. Since equal volumes of template were used for each target and the diversity of sample types (i.e. fecal swab versus water filter) precluded absolute quantification normalized to a volume or mass of sample, concentrations were instead reported directly as general or clinical class I integron-integrase copies/16S rRNA copies.

The Shapiro-Wilkes test was used to assess the normality of untransformed and log transformed integron-integrase/16S rRNA concentrations. Since neither the untransformed nor transformed data were normally distributed, the non-parametric Kruskal-Wallis test was used to compare differences in integron-integrase/16S rRNA concentrations between different samples. Samples with non-detect values were excluded when performing the Kruskall-Wallis test and for reporting group means. When performing pairwise comparisons of general *intI1*, marine iguana and sea turtle samples were excluded as sample categories due to small sample size resulting from low detection events (n=2). Mean integron-integrase/16S rRNA concentrations ± standard error (SE) are reported but were not used in testing mean differences between groups.

Mobile genetic element (mge) mapping

Additional characterization of the 90 metagenomes described in Chapter 3 was performed by mapping reads to a custom database of mobile genetic elements (MGE) compiled by Pärnänen et al. 2018. Analogous to the procedure for ARG mapping, metagenomic reads were mapped to the MGE database using Bowtie2 version 2.4.1 (Langmead and Salzberg, 2012) with

"highly sensitive" parameters -D 20 -R 3 -N 1 -L 20 -i S,1,0.50. Mapped reads were tabulated using SAMtools (Li et al., 2009) and normalized to small subunit ribosomal ribonucleic acid (SSU rRNA) counts as tabulated by Metaxa2 version 2.2 (Bengtsson-Palme et al., 2015) in paired-end mode, where SSU rRNA counts were considered as the sum of bacterial and archaeal SSU hits. SSU rRNA classified as *Eukaryota, Chloroplast, Mitochondria*, or *Uncertain* were excluded. The resulting observation matrix was imported into the R software package phyloseq version 1.34.0 (McMurdie and Holmes, 2013). Group mean MGE sum abundances/16S rRNA were compared using negative binomial generalized linear models (GLMs) in the R software package MASS (Venables and Ripley, 2002). The Bray-Curtis dissimilarity index was calculated for intra-species comparisons of MGE composition. The Multivariate Analysis of Variance Using Distance Matrices (ADONIS) in vegan was implemented with 9,999 permutations to assess to extent to which categorical variables (i.e. location) explained variation in the distance matrix. Differential abundance of MGEs by sample type was performed using the R package DESeq2 version 1.30.1 (Love et al., 2014). MGE/16S rRNA abundances were transformed to integers by multiplying each observation by 10^5 and rounding the result (Pärnänen et al. 2018). A pseudo-count of 1 was added to all observations to allow for inclusion and log transformation of zero observations.

Table 4.4: *Proportion of samples with detectable general intI1 and clinical intI1, mean general intI1/16S rRNA concentration, and mean clinical intI1/16S rRNA concentration by sample type.*

Sample Type	General *intI1*		Clinical *intI1*	
	Detection Events	Mean ± SE copies/16S rRNA copies	Detection Events	Mean ± SE copies/16S rRNA copies
Freshwater	12/12 (100%)	4.23E-04 ± 2.44E-04	11/12 (92%)	1.40E-04 ± 7.45E-05
Marine impacted	8/8 (100%)	2.62E-02 ± 6.72E-03	8/8 (100%)	2.10E-02 ± 5.27E-03
Marine background	22/22 (100%)	2.48E-03 ± 1.81E-03	21/22 (95%)	1.72E-03 ± 1.39E-03
WWTP influent	6/6 (100%)	7.89E-02 ± 2.19E-02	6/6 (100%)	5.70E-02 ± 1.76E-02
WWTP effluent	4/4 (100%)	1.96E-01 ± 1.03E-01	4/4 (100%)	5.51E-02 ± 1.66E-02
Human	27/28 (96%)	1.88E-02 ± 1.43E-02	27/28 (96%)	1.13E-02 ± 7.59E-03
Sea Lion	58/61 (95%)	2.16E-06 ± 5.92E-07	14/61 (23%)	8.80E-07 ± 4.50E-07
Fur Seal	9/12 (75%)	3.72E-06 ± 2.05E-06	0/12 (0%)	NA
Land Iguana	47/51 (92%)	3.29E-04 ± 2.79E-04	10/45 (22%)	4.49E-04 ± 3.37E-04
Marine Iguana	2/14 (14%)	1.82E-06 ± 1.75E-06	0/14 (0%)	NA
Giant Tortoise	30/32 (94%)	1.84E-05 ± 9.53E-06	19/32 (59%)	2.29E-05 ± 1.11E-06
Sea Turtle	2/4 (50%)	8.49E-07 ± 8.21E-07	1/4 (25%)	1.81E-08
Red-Footed Booby	0/5 (0%)	NA	0/5 (0%)	NA

Results

We aimed to differentiate clinical and environmental variants of the class I integron-integrase in wildlife, human, water, and wastewater samples from the Galapagos islands by developing a novel ddPCR assay targeting a conserved region of *intI1*. The forward primer for the 77 bp target matched the published primer HS464 (Stokes et al., 2006) with a modification to remove three nucleotides on the 3' end in order to reduce the primer T_m relative to the probe. We adapted a previously published RT-qPCR assay for *intI1* (Barraud et al., 2010) to identify variants of anthropogenic or clinical origin. In total, we analyzed 259 samples for the general *intI1* variant. Due to insufficient material in six samples, a total of 253 samples from this set were analyzed for the clinical *intI1* variant. Results are reported as general *intI1* copies/16S rRNA copies or clinical *intI1* copies/16S rRNA copies.

Detection of intI1 variants by sample type

Across the data set, the general *intI1* variant was detected in 87.6% of samples (227/259) compared to the clinical variant with detection in 47.8% of samples (121/253). Detection of *intI1* variants differed by sample type in regards to both prevalence and concentration (**Table 4.4, Figure 4.2**). The proportion of samples testing positive for general *intI1* was highest in water samples, with 100% of wastewater, marine water, and freshwater samples positive. All but one of 28 human samples tested positive for general *intI1* (96%). Among wildlife, sea lions exhibited the highest proportion of general *intI1* detection (58/61, 95%), followed by giant tortoises (30/32, 94%) and land iguanas (47/51, 92%). Fur seals showed moderate detection with 9/12 samples testing positive (75%), followed by less frequent detection in sea turtles (2/4, 50%) and marine iguanas (2/14, 14%). The general *intI1* variant was not detected in the five red footed booby samples analyzed. While similar proportions of different sample types tested positive for general *intI1*, significant differences were recorded in terms of concentration. For example, while general

intI1 was detected in 100% of both freshwater and wastewater-impacted marine samples, mean concentrations between the two sample types differed in magnitude, with concentrations of 2.62E-02 ± 6.72E-03 copies/16S rRNA copies observed for impacted marine water compared to 4.23E-04 ± 2.44E-04 copies/16S rRNA copies among freshwater samples (Kruskall-Wallis test with FDR-corrected $p<0.05$). Likewise, while detection of general *intI1* was ≥ 92% in land iguana, sea lion, and giant tortoise samples, the concentration was significantly higher in land iguanas (3.29E-04 ± 2.79E-04 copies/16S rRNA copies) compared to giant tortoises (1.84E-05 ± 9.53E-06 copies/16S rRNA copies, FDR-corrected $p<0.05$) and sea lions (2.16E-06 ± 5.92E-07 copies/16S rRNA copies, FDR-corrected $p<0.05$). Mean concentration of general *intI1* among human samples (1.88E-02 ± 1.43E-02 copies/16S rRNA copies) was of the same order of magnitude as wastewater influent samples (7.89E-02 ± 2.19E-02 copies/16S rRNA copies). Mirroring the lower detection prevalence among fur seals, marine iguanas, and sea turtles, general *intI1* concentration was lowest among these sample types, ranging from 1.82E-06 ± 1.75E-06 copies/16S rRNA copies in marine iguanas to 8.49E-07 ± 8.21E-07 copies/16S rRNA copies in sea turtles. When performing pairwise Kruskall-Wallis tests, general *intI1* concentrations ranked significantly higher in wastewater effluent, influent, and wastewater-impacted marine samples compared to other sample types, and human samples ranked significantly higher than all wildlife samples (FDR-corrected $p<0.05$).

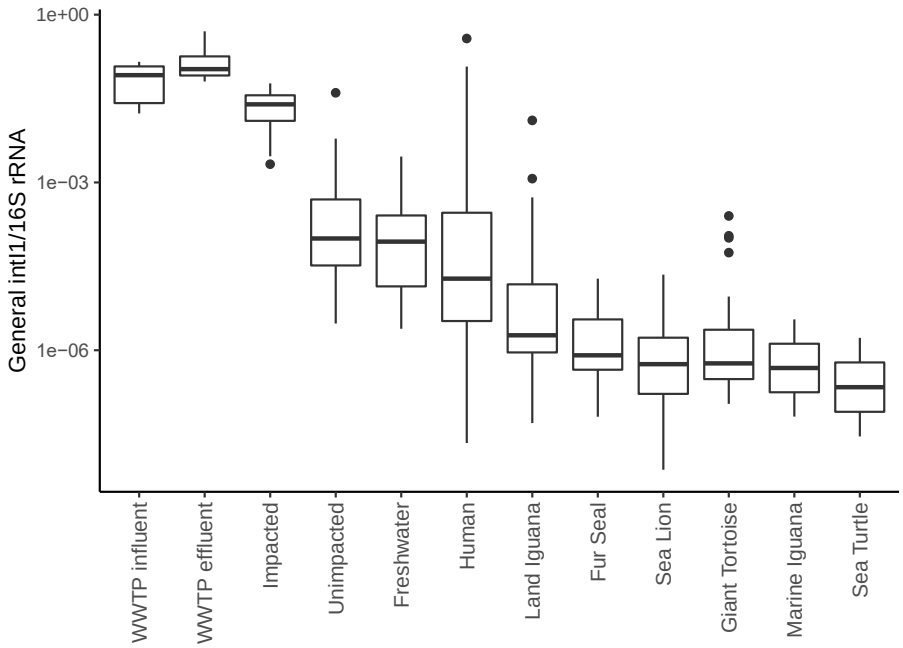

Figure 4.2: *Mean general intI1/16S rRNA concentration by sample type.* General *intI1* was detected in all sample types except red footed boobies.

Overall, the clinical *intI1* variant was detected less frequently across the dataset, with differences again observed by sample type in regards to the proportion of samples testing positive and the concentration of the target (**Figure 4.3, Table 4.4**). Clinical *intI1* was detected in 100% of wastewater influent, wastewater effluent, and wastewater-impacted marine sites, with the highest mean concentration observed in wastewater effluent (5.51E-02 ± 1.66E-02 copies/16S rRNA copies.) The clinical target was detected in nearly all freshwater (11/12, 92%) and unimpacted marine water (21/22, 95%) samples, with respective mean concentrations of 1.40E-04 ± 7.45E-05 copies/16S rRNA copies and 1.72E-03 ± 1.39E-03 copies/16S rRNA copies. Mirroring the pattern with general *intI1*, all but one of 28 human samples were positive for clinical *intI1*, and mean concentrations (1.13E-02 ± 7.59E-03 copies/16S rRNA copies) were of the same order of magnitude as wastewater and wastewater-impacted marine sites. Clinical

intI1 detection was markedly lower among wildlife samples, with the highest prevalence observed in giant tortoises (19/32, 59%) followed by sea lions (14/61, 23%), and land iguanas (10/45, 22%). Additionally, the clinical target was detected in one of four sea turtles (25%) but at the lowest concentration recorded for all samples (1.81E-08 copies/16S rRNA copies.) Whereas general *intI* could be detected in marine iguana and fur seal samples, the clinical target was undetected in samples from these two species. Consistent with the general *intI1* data, the clinical variant was again undetected in red-footed booby samples. As with general *intI1* in wildlife samples, land iguanas were associated with the highest clinical *intI1* concentrations (4.49E-04 ± 3.37E-04 copies/16S rRNA copies) compared to giant tortoises (2.29E-05 ± 1.11E-06 copies/16S rRNA copies) and sea lions (8.80E-07 ± 4.50E-07 copies/16S rRNA copies). When performing pairwise Kruskall-Wallis tests, clinical *intI1* concentrations were ranked highest among wastewater influent, wastewater effluent, and wastewater-impacted marine sites over all other samples (FDR corrected $p<0.05$) though the three sample types not significantly different from one another. Concentrations from human samples ranked higher than all wildlife samples with the exception of land iguanas (FDR-corrected $p<0.05$). Among wildlife samples, the difference in clinical *intI1* between land iguanas and giant tortoises over sea lions was significant (FDR-corrected $p<0.05$) but not between land iguanas and giant tortoises.

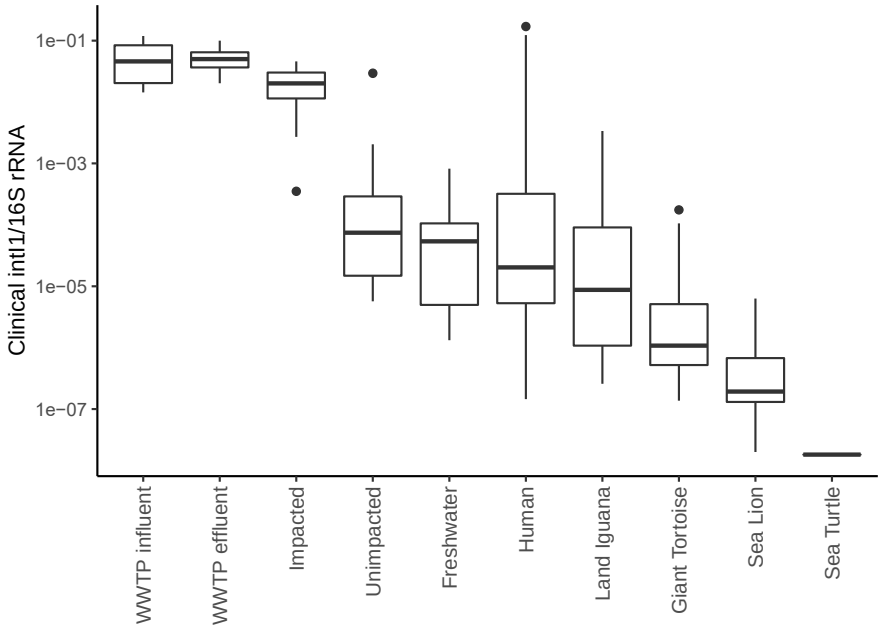

Figure 4.3: *Mean clinical intI1/16S rRNA concentration by sample type.* Among wildlife, clinical *intI1* was undetected in marine iguanas, fur seals, and red footed boobies.

Ratio of clinical to general intI1

Based on the work of Waldron and Gillings (2015) and Gillings et al. (2015) regarding the conservation of the region of *intI1* amplified by primer pair HS464/HS463a, we hypothesized that the general *intI1* variant would be detected more frequently and at a higher concentration than clinical *intI*. Moreover, we hypothesized that samples with intense anthropogenic impacts (i.e. human fecal samples, wastewater, or wastewater-impacted environments) would exhibit ratios of clinical *intI1* to general *intI1* close to 1, indicating that the majority of *intI1* variants in the sample were of anthropogenic or clinical origin. Our data support the first part of this hypothesis in that general *intI1* was detected more frequently than clinical *intI1*, with detection in 87.6% versus 47.8% of samples, respectively. Moreover, the clinical *intI1* variant was found exclusively in samples where the general *intI1* variant was also detected, meaning that there were no cases in which clinical *intI1* was present but general *intI1* was absent. However, counter to the

premise that the ratio of clinical to general *intI1* would be less than 1, indicating that general *intI1* was more abundant, this ratio was greater than 1 in 18% (22/121) of samples (**Figure 4.4).** Human samples accounted for twelve inverted ratios, along with seven marine water samples and three giant tortoises. The seven marine water samples originated from costal sites without clear wastewater impacts, with 5/7 from La Loberia. Two of the three giant tortoises were housed at La Galapaguera while the remaining individual came from Otoy Ranch. Among these 22 samples, the ratio of clinical to general *intI1* was less than 1.25 in 10 cases and less than 2 in 14 cases. The remaining samples with ratios greater than 2 included six human samples and two marine water samples.

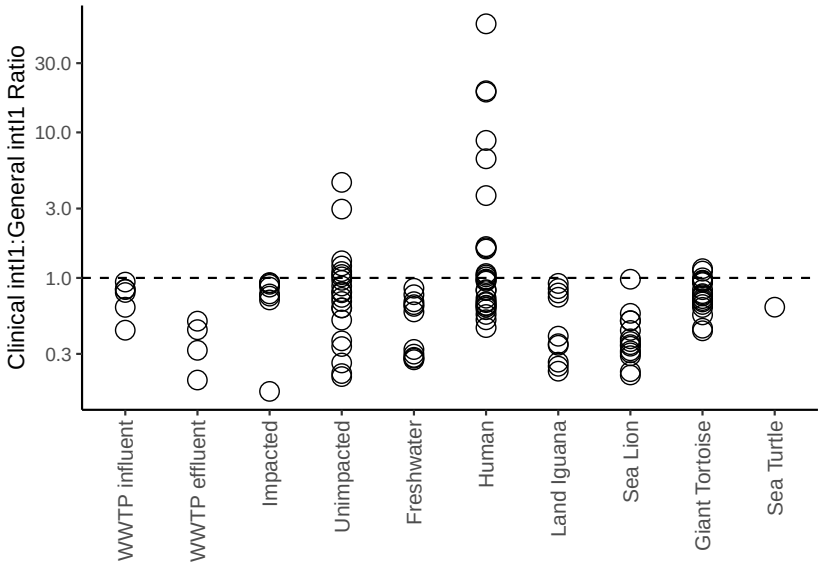

Figure 4.4: *Ratio of clinical to general intI1by sample type.* Ratios > 1, indicating a higher concentration of clinical *intI1* relative to general *intI1*, were observed in a selection of human, giant tortoise, and marine water samples.

Intra-species comparisons of general and clinical intI1

We next examined intra-species differences in general *intI1* and clinical *intI1* detection by location. Among sea lion samples, no significant differences in general *intI1* were observed

by location according to pairwise Kruskall-Wallis tests (all FDR-corrected $p>0.05$). In contrast, detection of clinical *intI1* varied considerably by location, with detection in 75% (8/12) of individuals from Punta Pitt, San Cristobal; 20% (2/10) from El Malecon, San Cristobal; 20% (1/5) from Cabo Douglas, Fernandina; 17% (2/12) from Champion, Floreana; and 8% (1/12) from Puerto Egas, Santiago. None of the 10 individuals from Punta Mangle, Fernandina tested positive for clinical *intI1*. The ratio of clinical to general *intI1* ranged from 0.21 to 0.97, with both the minimum and maximum values observed in individuals from Punta Pitt on San Cristobal.

Similarly, no significant differences were observed in general *intI1* concentrations among land iguanas from three distinct islands (Kruskall-Wallis pairwise tests, all FDR- corrected $p>0.05$). Clinical *intI* was detected in 28% (5/18) and 22% (4/18) of individuals from North Seymour and Santa Fe, respectively, compared to 11% (1/9) individuals on Plaza Sur. Moreover, while neither concentration of general *intI1* nor clinical *intI* was significantly different between giant tortoises from La Galapaguera and Otoy Ranch (Wilcoxon rank sum test, $p>0.05$), the detection proportion varied by location with 100% (6/6) of individuals from Otoy Ranch testing positive for both targets compared to 92% (24/26) and 50% (13/26) of individuals at La Galapaguera testing positive for the general and clinical class I integron-integrase, respectively.

Finally, we asked if integron-integrase detection varied among human children under age two based on birth mode. Among the 27 human samples positive for general and clinical *intI1*, no significant difference was observed between individuals born vaginally (n=9) versus via C-section (n=18) for either target according to the Wilcoxon rank sum test ($p>0.5$ in both cases).

Table 4.5: *Mean sum MGE abundance/16S rRNA and most abundant MGEs by sample type.*

Sample Type	n	Mean sum MGE abundance/16S ± SE	Top 3 MGE classes	Mean sum/16S	Top 3 MGEs	Mean sum/16S
Wastewater	4	1.31E+00 ± 4.67E-01	transposase	9.97E-01	tnpA	8.48E-01
			tniA	6.77E-02	tniA	6.77E-02
			IS91	6.26E-02	IS91	6.26E-02
Marine impacted	4	3.82E-01 ± 1.43E-01	transposase	3.22E-01	tnpA	2.78E-01
			IS91	1.39E-02	IS91	1.39E-02
			istA	6.32E-03	tnpAIS50.A	1.03E-02
Marine background	7	1.80E-02 ± 8.06E-03	transposase	1.33E-02	tnpA	1.26E-02
			IS91	2.41E-03	IS91	2.41E-03
			istA	6.72E-04	istA	6.72E-04
Freshwater	4	2.18E-02 ± 5.71E-03	transposase	1.19E-02	tnpA	1.11E-02
			istB1	1.92E-03	istB1	1.92E-03
			IS91	1.87E-03	IS91	1.87E-03
Human	12	4.43E-01 ± 1.79E-01	transposase	3.75E-01	tnpA	3.41E-01
			IS91	2.06E-02	IS91	2.06E-02
			plasmid	1.97E-02	tnpA.IS683	1.33E-02
Sea Lion	24	5.08E-02 ± 1.78E-02	transposase	3.76E-02	tnpA	2.59E-02
			plasmid	3.35E-03	tnpA4	6.53E-03
			integrase	2.92E-03	tnpA.IS683	5.21E-03
Land Iguana	10	2.29E+00 ± 8.56E-01	transposase	1.30E+00	tnpA	1.03E+00
			IS91	4.80E-01	IS91	4.80E-01
			istA2	1.72E-01	istA2	1.72E-01
Giant Tortoise	6	3.32E-03 ± 1.08E-03	transposase	2.60E-03	tnpA	2.43E-03
			IS91	2.56E-04	IS91	2.56E-04
			IS621	1.94E-04	IS621	1.94E-04
Sea Turtle	7	8.82E-03 ± 3.15E-03	transposase	7.12E-03	tnpA	6.25E-03
			ISCR	8.76E-04	ISCrsp1	8.76E-04
			IS91	7.84E-04	IS91	7.84E-04
Marine Iguana	8	7.22E-02 ± 4.98E-02	transposase	4.62E-02	tnpA	4.34E-02
			IS91	1.23E-02	IS91	1.23E-02
			ISBf10	8.28E-03	ISBf10	8.28E-03
Red Footed Booby	4	1.42E-02 ± 1.00E-02	IS91	5.10E-03	IS91	5.10E-03
			transposase	5.05E-03	tnpA	5.05E-03
			IS621	2.59E-03	IS621	2.59E-03

Mapping to mobile genetic elements (MGEs)

The mobilomes of the 90 metagenomes were explored by mapping sequences to a custom mobile genetic element (MGE) database compiled by Pärnänen et al. 2018. Counts were normalized to SSU rRNA sequences as tabulated by Metaxa2 following the procedure described in Chapter 3 and reported as MGE copies/16S rRNA copies. As with the ARG annotation in Chapter 3, initial analyses were performing by broadly classifying samples as human, wastewater, water, or wildlife. Calculation of the Bray-Curtis dissimilarity index revealed sample type to be a significant explanatory variable in MGE composition, though the effect size was modest with R-square 0.14 (**Figure 4.7,** p=1e-04, ADONIS test with 9,999 permutations). As depicted in **Figure 4.5**, mean sum abundance of MGEs was highest among wastewater samples (1.32E+00 ± 1.26E+00 copies MGE/copies 16S rRNA, negative binomial GLM predicted mean ± SE), followed by humans (4.39E-01 ± 2.41E-01), wildlife (4.26E-01 ± 1.05E-01), and water (1.16E-01 ± 5.70E-02). However, no group mean differences were significant according to Tukey's post hoc test (data not shown). Further categorization of samples into eleven subtypes revealed additional differences within these categories (**Figure 4.6, Table S4.1**). Wastewater mean sum abundance of MGEs was significantly higher than freshwater, unimpacted marine water, and all wildlife samples with the exception of land iguanas ($p<0.05$, Tukey's post hoc test). Human samples shared a similar pattern with increased MGE sums over freshwater, unimpacted marine water, and all wildlife samples except land iguanas as well as marine iguanas ($p<0.05$, Tukey's post hoc test). Analogous to ARG results, land iguanas presented higher mean sum MGEs compared to all other wildlife, freshwater, and unimpacted marine water samples ($p<0.05$, Tukey's post hoc test). In terms of which MGEs were observed between sample types, all shared transposases as the most abundant MGE class, with the exception of red footed

boobies for which insertion sequence 91 (IS91) accounted for the most abundant MGE class. IS91 represented one of the top three most abundant MGE classes in all sample types except sea lions. Humans and sea lions were the only two sample types for which plasmids were among the top three MGE classes.

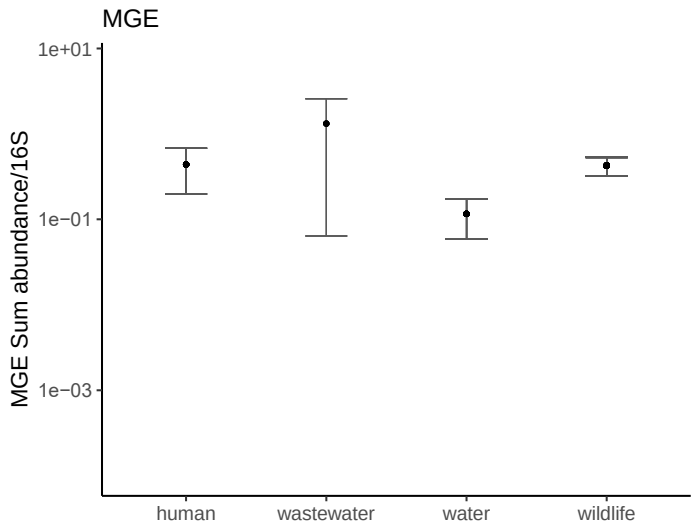

Figure 4.5: *MGE sum abundance/16S rRNA for human, wastewater, water, and wildlife samples.* Error bars and the black data point represent negative binomial GLM-predicted means ± standard error.

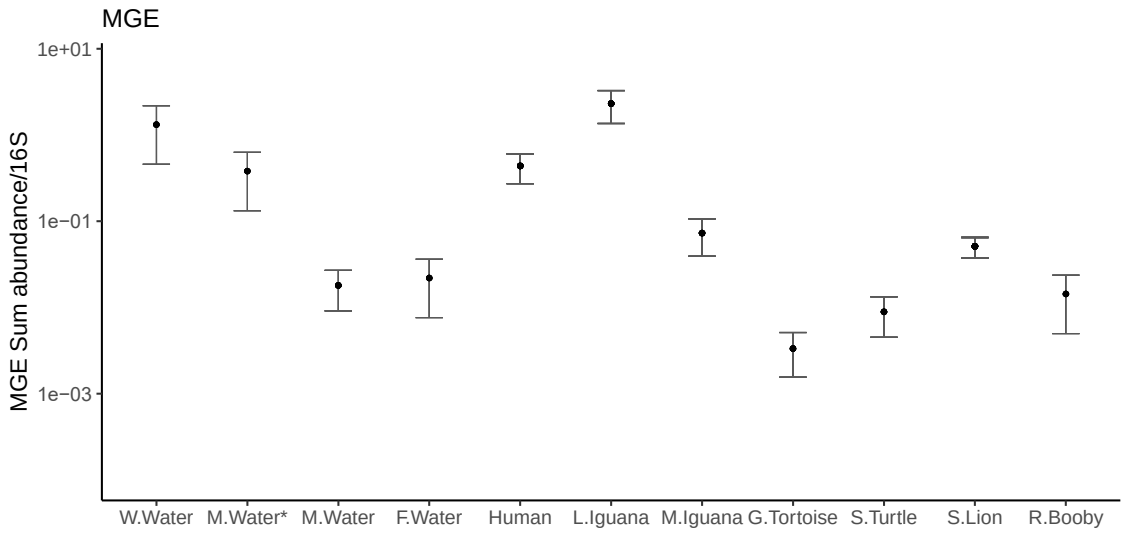

Figure 4.6: *MGE sum abundance/16S rRNA for sample subtypes.* M. Water* indicates marine sites with documented wastewater impacts. Error bars and the black data point represent negative binomial GLM-predicted means ± standard error.

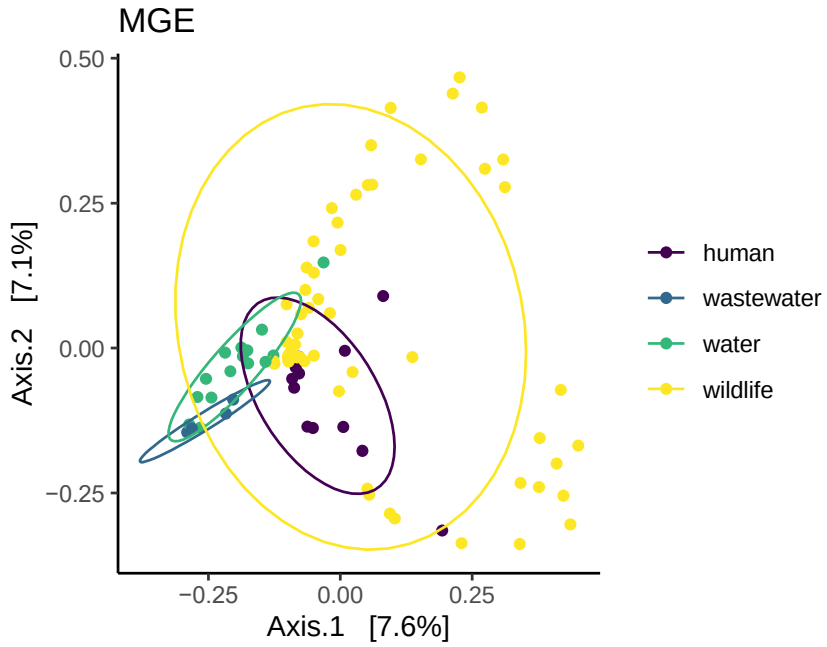

Figure 4.7: *Composition of MGEs by sample type based on sample distances calculated using the Bray-Curtis dissimilarity index.*

Intra-species MGE differences by location

We next explored if MGE sums and MGE composition varied by location between individuals of the same species. Among sea lions, group mean sum abundance of MGEs was significantly higher in individuals from Punta Mangle, Fernandina compared to those from El Malecon, San Cristobal and Champion, Floreana (**Figure 4.8, Table S4.2,** negative binomial GLM predicted means, Tukey's post hoc test, $p<0.05$). MGE sums were also found to be significantly higher in sea lions from Cabo Douglas, Fernandina, compared to those from El Malecon (negative binomial GLM predicted means, Tukey's post hoc test, $p<0.05$). No other pairwise comparisons were significant using Tukey's post hot test. Calculation of the Bray-Curtis dissimilarity index revealed location to significantly explain MGE composition, with R-square equal to 0.40 (**Figure 4.9,** ADONIS test with 9,999 permutations, p=1e-04).

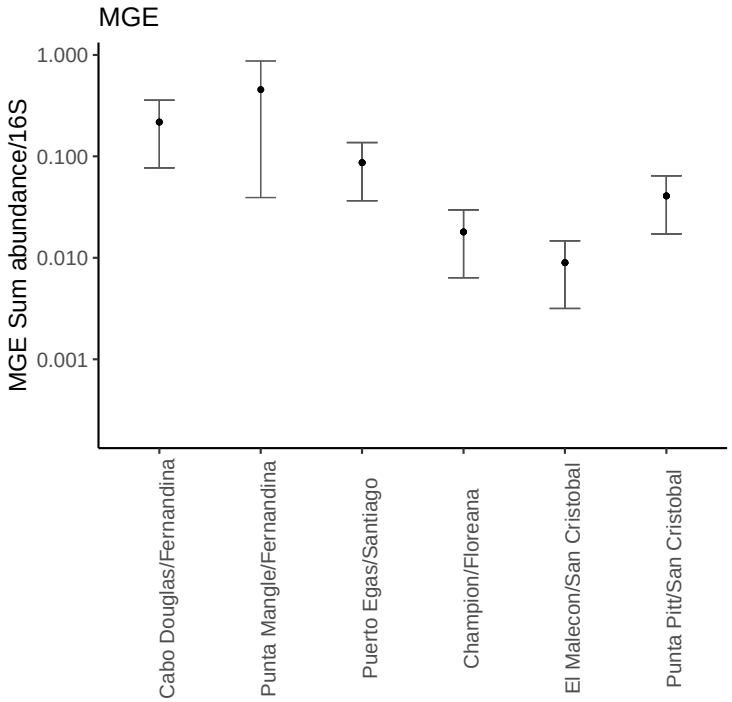

Figure 4.8: *MGE sum abundance/16S rRNA for sea lion samples from six sampling locations.* Error bars and the black data point represent negative binomial GLM-predicted means ± standard error.

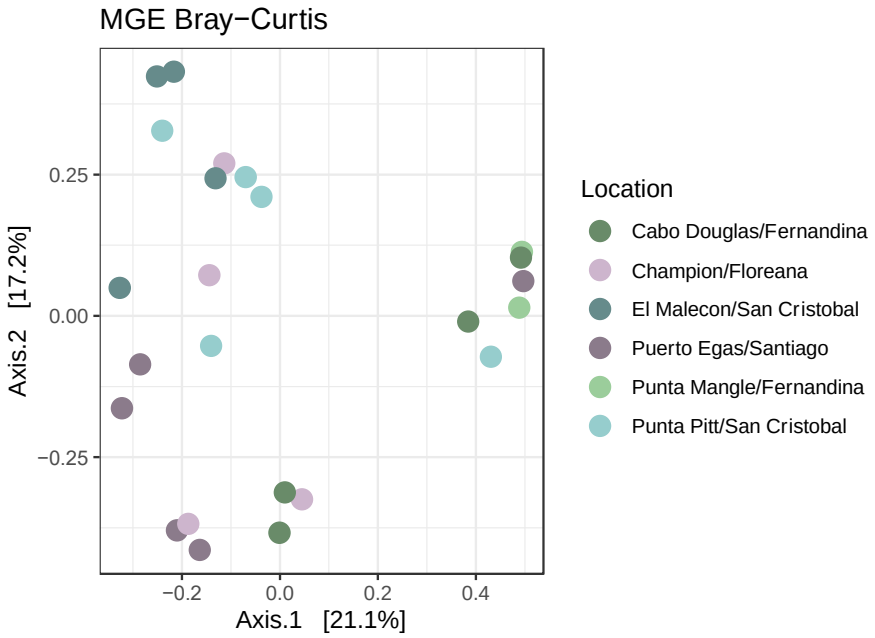

Figure 4.9: *Composition of MGEs among sea lions by sampling location based on sample distances calculated using the Bray-Curtis dissimilarity index.*

Significant differences in MGE sums and composition were also observed between land iguanas from different islands. Total MGE sum abundance/16S rRNA was significantly higher in individuals from both North Seymour and Santa Fe compared to Plaza Sur (**Figure 4.10**, negative binomial GLM predicted means, Tukey's post hoc test, $p < 0.05$). Location proved to be a significant explanatory variable in the Bray-Curtis dissimilarity index, with R-square equal to 0.48 (**Figure 4.11**, p=0.0022, ADONIS test with 9,999 permutations).

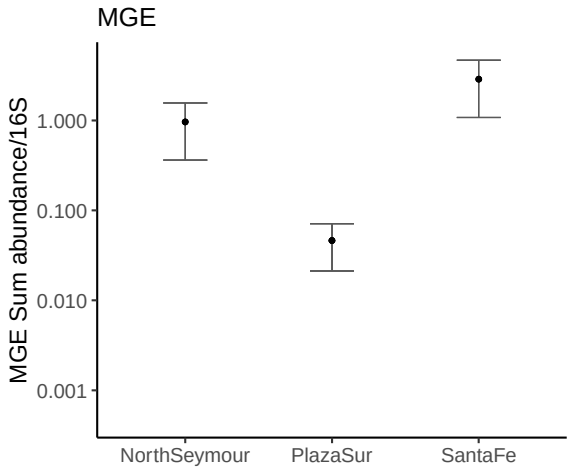

Figure 4.10: *MGE sum abundance/16S rRNA for sea lion samples from three islands.* Error bars and the black data point represent negative binomial GLM-predicted means ± standard error.

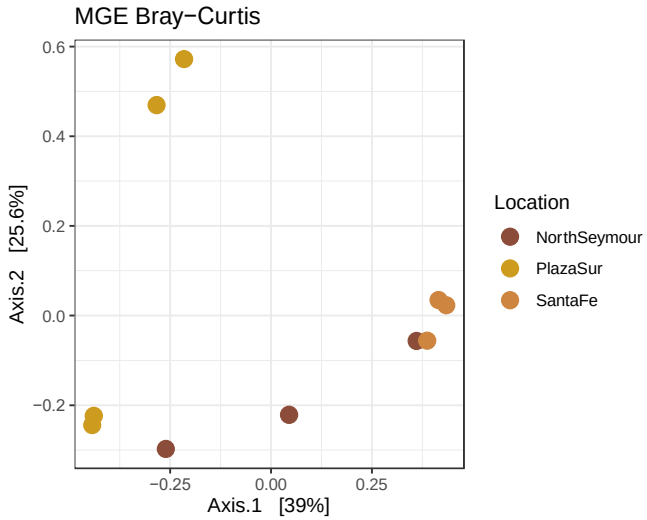

Figure 4.11: *Composition of MGEs among land iguanas by sampling location based on sample distances calculated using the Bray-Curtis dissimilarity index.*

Finally, we explored differences in MGE sum and composition by birth mode among children under age two. While mean sum abundance of MGEs/16s RNA was slightly higher in babies born via Caesarean section compared to those born vaginally (**Figure 4.12,** negative binomial GLM predicted means, Tukey's post hoc test, p-value = 0.0429), birth mode was not a significant explanatory variable in the Bray-Cutis dissimilarity index (R-square = 0.07, p-value =

0.51, ADONIS test with 9,999 permutations). Moreover, select MGEs were differentially abundant in children born via Caesarean section compared to those born vaginally (**Figure 4.12**). Of the 143 MGEs differentially abundant by birth mode, 132 were significantly more abundant in babies born via Caesarean section (negative binomial GLMs, Wald's test implemented in DESeq2, adjusted p-value <0.05). MGEs with the biggest fold change over babies born vaginally included those belonging to transposases and plasmids, though variants of the transposase tnpA dominated both groups (**Table S4.3).**

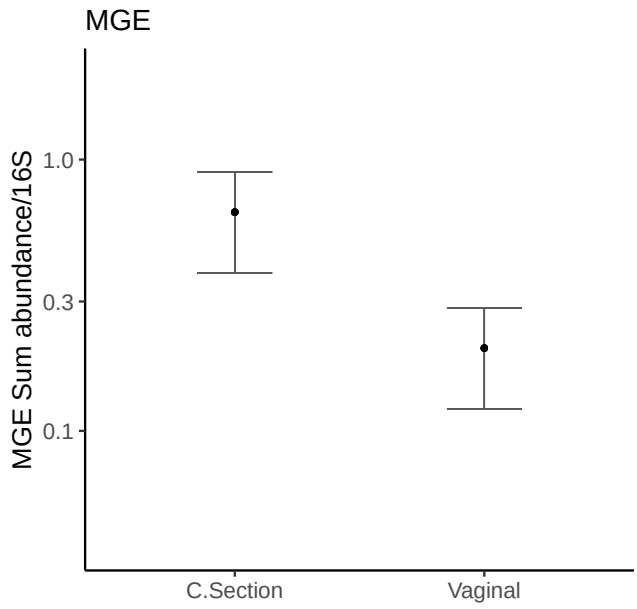

Figure 4.12: *MGE sum abundance/16S rRNA for babies born via Cesarean section (n=6) or vaginally (n=6).* Error bars and the black data point represent negative binomial GLM-predicted means ± standard error.

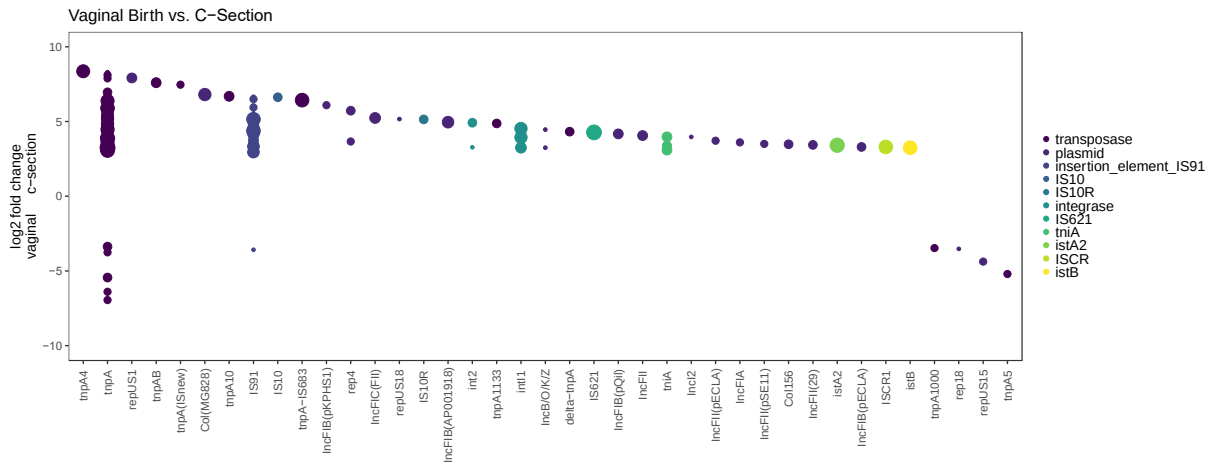

Figure 4.13: *MGEs differentially abundant between babies born via Caesarean section versus vaginally.* MGEs with positive fold changes were differentially abundant in the Caesarean section group. MGEs with negative fold changes were differentially abundant in the vaginal birth group. Sizes of data points correspond to the number of individuals (1-12) in which the MGE was detected.

Comparison of ddPCR and MGE mapping

We subsequently investigated the extent of agreement between MGE mapping and ddPCR detection of the class I integron-integrase. Using MGE mapping, the class I integron was detected in 20 of 90 metagenomes, including one freshwater, one giant tortoise, five marine water, nine human, and all four wastewater samples (**Figure 4.15**). Four of the five marine water samples originated from wastewater-impacted costal sites (Carola 1 and Marinero from both 2017 and 2018). In comparison, using ddPCR clinical *intI* was detected in a total of 40 samples, including all wastewater (n=4), marine water (n=11), freshwater (n=4), and human (n=12) samples, as well as one land iguana, two giant tortoises, five sea lions, and one sea turtle. This number increased to 60 when considering samples positive for clinical *intI1* but below the threshold of three or more positive droplets. All 20 samples with MGE mapping results to *intI1* were also ddPCR positive for clinical *intI1*, and there were no cases in which *intI1* was detected by MGE mapping and not ddPCR. As illustrated in **Figure 4.14,** correlation of log-transformed

intI1/16S rRNA from MGE mapping to log-transformed clinical *intI1*/16S rRNA from ddPCR using linear regression revealed a significant positive relationship with an effect size of 0.68 (adjusted R-square of 0.8629, p-value 2.08e-09).

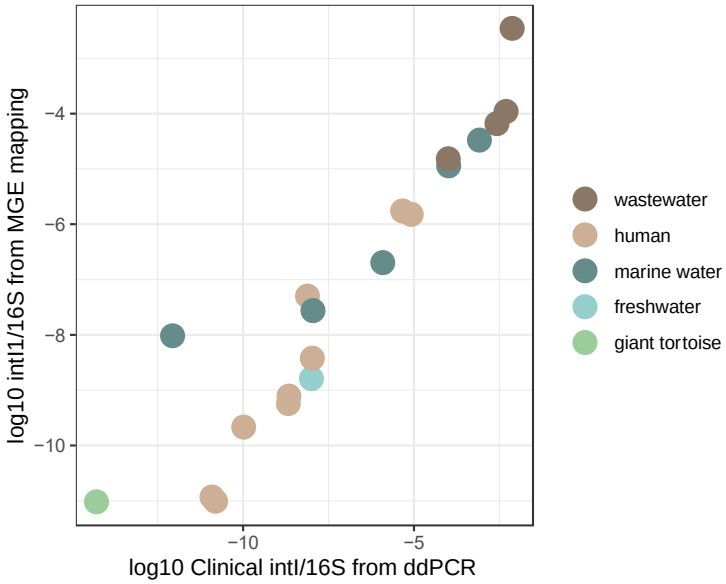

Figure 4.14: *Liner correlation between log-transformed clinical intI1/16S rRNA as detected by ddPCR and log-transformed intI1/16S rRNA as detected by MGE mapping.*

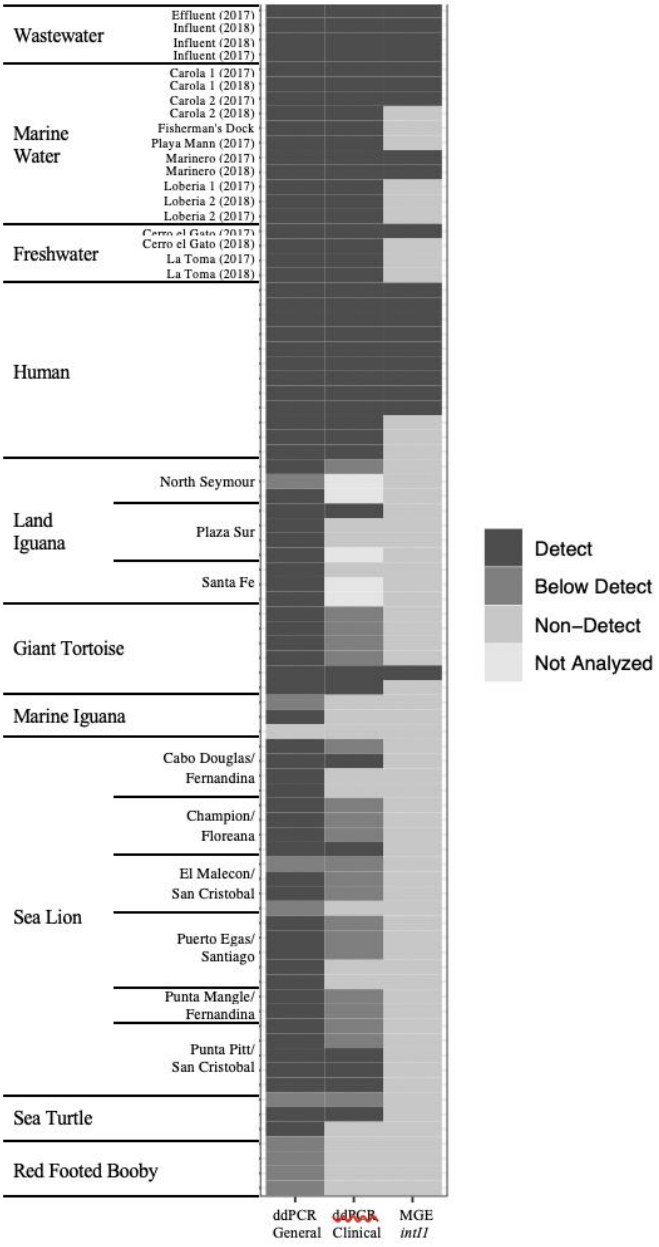

Figure 4.15: *Detection of general intI1 by ddPCR, clinical intI1 by ddPCR, and intI1 by MGE mapping for paired samples.*

Discussion

We investigated the mobilomes of Galapagos wildlife, water, wastewater, and human reservoirs by pairing a novel ddPCR assay for the class I integron-integrase with a mapping-based metagenomics approach. Collectively these approaches pointed to a general trend of

increasing mobile genetic elements (MGEs) and clinical *intI1* detection along a gradient of anthropogenic influence. While the two approaches showed considerable agreement in identifying samples with significant mobilomes, ddPCR proved a more sensitive method in detecting the class I integron-integrase.

Detection of general versus clinical class I integron-integrase variants

Among over 250 samples representing environmental, human, and wildlife reservoirs, the general *intI* variant was detected more frequently than the clinical variant, with detection in 87.6% and 47.8% of samples, respectively. Clinical *intI1* was found exclusively in the presence of general *intI1*, such that there were no instances in which the clinical variant was detected in the absence of the general variant. This finding aligns with the framework presented by Waldron and Gillings (2015) in which 1) general and clinical *intI1* variants can be distinguished based on primer selection, and 2) the clinical variant should be found exclusively in the presence of the general variant, but not vice versa. Moreover, the concentration of clinical *intI1* exceeded that of the general variant in 82% of samples, as indicated by a clinical to general ratio < 1. This too supports the notion that class I integron-integrases bearing the clinical sequence will be PCR positive for both targets, whereas integron-integrases with of environmental or non-clinical origin will amplify only the general target, contributing to an overall greater concentration of the general target in a given sample.

Detection and abundance of class I integron-integrase by sample type

For both general and clinical targets, we observed an overall pattern of increasing concentration along a gradient of human influence, with the highest values recorded in wastewater and wastewater-impacted marine environments followed by humans, freshwater, marine water, and wildlife. Both variants were ubiquitous in all wastewater and wastewater-

impacted marine environments, and nearly ubiquitous in all freshwater and background marine sites, with only one sample of each type negative for clinical *intI1* but all positive for general *intI1*. Detection of the class I integron-integrase is now common in aquatic environments, as Wang et al. (2018) used the same clinical *intI1* primers (Barraud et al., 2010) adapted for ddPCR to survey surface water samples from the Weihe River in China across a range of rural and urban areas. Clinical *intI1* was detected in 100% of samples, though concentrations were higher in urban sampling sites. Similarly, Dungan and Bjorneberg (2020) used the Barraud et al. (2010) primers adapted for ddPCR to quantify class I integrons in irrigation return flows in Idaho. The target was recovered at all eight sampling sites with detection ranging from 67-100% of 81 sampling events. Detection of the clinical variant in nearly all freshwater samples is notable, however, as these two sampling sites on San Cristobal are located in the highlands well above the large human settlement in Puerto Baquerizo Moreno. Previous work by our group (Grube et al., 2020) has documented mean total coliform and *E. coli* concentrations at these sites in the range of 100-1000 MPN/100 mL and 1-10 MPN/100 mL, respectively, representing a potential source of integron-bearing Gammaproteobacteria. In constructing a class I integron database, Zhang and colleagues (2018) estimated that 96% of class I integrons are found in Gammaproteobacteria, and in particular the family *Enterobacteriaceae*. In the case of the highlands on San Cristobal, more work is needed to discern if the *E. coli* detected in freshwater sources are of environmental origin and/or are the result of limited agricultural activity in this region. Nonetheless, the concentration of clinical *intI1* in the two freshwater sites was an order of magnitude lower compared to unimpacted marine water sites, signifying that while prevalent, the overall abundance of clinical *intI1* was the lowest among aquatic reservoirs studied.

Among 28 samples from children under age two, only one was negative for both general and clinical *intI1*. High or ubiquitous prevalence of clinical *intI1* in human samples could be expected, even among children under age two, as several studies have indicated a higher antibiotic resistance load among infants compared to adults (Gibson et al., 2016; Li et al., 2021) and *intI1* often correlates with ARG load (Gillings et al., 2015, Zheng et al., 2020). In a PCR survey of fecal samples from healthy adults (Labbate et al., 2008), the class I integron-integrase was detected in nine of fifteen total individuals using primer pair HS464/HS463a (Stokes et al., 2006), which detects both environmental and clinical *intI1* per the definition of Waldron and Gillings (2015). Culture-based approaches have yielded even lower detection levels, including only 20 of 181 (11%) *E. coli* isolates from healthy adult humans testing positive for *intI1* using a primer pair with specificity similar to that of Barraud et al. (2010) in distinguishing clinical *intI1* (Skurnik et al., 2005). The comparatively higher prevalence of clinical *intI1* among human samples in the present study could be attributed to the increased sensitivity of ddPCR on fecal metagenomic DNA extracts over initial culture-based screenings exclusively in *E. coli*. For example, in a study of integron carriage among *E. coli* cultured from wastewater (Kotlarska et al., 2015), only 29-38% of antibiotic resistant isolates from raw wastewater and 27-37% of isolates from treated effluent were PCR positive for *intI*. In contrast, 100% of wastewater samples were positive for both general and clinical *intI1* in the present study, indicating that PCR based methods on total genomic DNA likely offer increased sensitivity over initial culture-based screenings.

General *intI1* detection was similarly high among certain wildlife species, with 92-95% detection in sea lions, giant tortoises, and land iguanas. Despite similar prevalence among these species, concentrations varied considerably by more than two orders of magnitude between land

iguanas (3.29E-04 ± 2.79E-04 copies/16S rRNA copies) and sea lions (2.16E-06 ± 5.92E-07 copies/16S rRNA copies). Detection of general *intI1* was markedly lower among the other wildlife species, with 75% detection in fur seals, 50% in sea turtles (only 2/4 individuals), 14% of marine iguanas, and 0% of red footed boobies. However, we posit that lower detection in these species may be in part explained by lower DNA extraction efficiency and yields for these sample types. Notably, mean DNA concentration among general *intI1* positive wildlife samples (n=148) was 22.4 ± 21.1 ng/uL, compared to 10.9 ± 13.2 ng/uL in the non-detect group (n=31). This difference is significant according to the Wilcoxon rank sum test (W = 3178.5, p=0.0007523), and differences in DNA concentration according to detection status can be observed by sample type (**Figure S4.5).** Accordingly, we cannot conclude that these samples are truly negative for general *intI1*, but instead suspect that our methods were insufficient to capture the target.

Clinical *intI1* was detected less frequently in wildlife, ranging from 59% of giant tortoise samples to 22% of land iguanas. However, as with the general *intI1* sequence, concentrations were markedly higher among land iguanas compared to giant tortoises and sea lions, though the difference between giant tortoises and land iguanas was not statistically significant. Notably, detection of clinical *intI1* among wildlife samples positive for general *intI1* did not seem to depend on DNA concentration (**Figure S4.6),** with respective means of 24.9 ± 19.7 ng/uL and 21.9 ± 22.1 ng/uL among positive and non-detect samples (W = 2513.5, p=0.1428). Our data suggest that geography may instead play a role in intra-species differences in clinical *intI1* detection. Notably, 8/14 (57%) clinical *intI1* detection events in sea lions were among individuals from a single location, Punta Pitt, on the northeastern side of San Cristobal. While San Cristobal is inhabited by humans, this particular beach is located on the uninhabited side of

the island and can only be accessed by boat. In total, 75% of sea lions from Punta Pitt were positive for clinical *intI1*, compared to 0-20% in all other locations. Due to their proximity to both humans and wastewater-impacted marine waters, we had hypothesized that sea lions from El Malecon, San Cristobal, would be associated with the highest levels of ARGs and MGEs. In a study of wild versus captive Australian sea lions (*Neophoca cinerea*), Delport and colleagues (2015) reported detection of *intI1* using primer pair HS464/HS463a (Stokes et al., 2006) in *E. coli* isolated from 8 captive individuals but were unable to detect the target in *E. coli* isolated from 21 wild animals. Notably, the authors observed an overall lower prevalence of *E. coli* in wild sea lions compared to captive animals, corroborating the connection between class I integrons and *Enterobacteriaceae* proposed by Zhang et al. (2018) in constructing the integron database. Along the same line, our data suggest that clinical *intI1* detection among Galapagos sea lions may not depend solely on proximity to human settlements, but may instead relate to underlying differences in gut microbial community composition (discussed further in Chapter 5).

Land iguanas also presented geographic and sub-population differences in *intI1* detection. Considering the differences observed in ARG sum abundance between individuals from North Seymour, Santa Fe, and Plaza Sur (discussed in Chapter 3), it is notable that no significant differences were observed in general *intI1* concentrations among the three islands. However, detection of clinical *intI1* was higher in both North Seymour (28%) and Santa Fe (22%) compared to Plaza Sur (11%). This finding aligns with the ARG data from Chapter 3 and supports the connection between clinical *intI1* and antibiotic resistance. Chapter 5 further contextualizes these data in relation to the gut microbial community composition of land iguanas, in which *Enterobacteriaceae* appear to play a dominant role compared to other wildlife species in the present study.

Placing these results within the literature on *intI1* detection in wildlife, our data would appear to suggest a higher prevalence compared to other surveys. However, the majority of studies to date have used initial culture-based screenings overwhelmingly targeting *E. coli*. As discussed above, this approach resulted in *intI1* detection in the range of 29-38% for raw wastewater samples (Kotlarska et al., 2015), whereas our culture-independent, ddPCR approach led to 100% detection of both general and clinical targets in wastewater. Accordingly, studies of *intI1* carriage in wildlife that rely on initial culturing must be interpreted through this lens. For example, in a recent study of AMR in 386 wild animals in Italy, Gambino and colleagues (2021) detected *intI1* in only three of 61 strains including one *E. coli* from a golden eagle, one *E. coli* from an owl, and one *Enterobacter cloacae* from a rabbit, using a primer set with specificity similar to that of Barraud et al. (2010). In another study of European wildlife, Literak and colleagues (2010) detected *intI1* in a total of three antibiotic resistant isolates among a total 590 *E. coli* isolates, including one from a striped field mouse and two from wild boars. The class I integron integrase has been found at similarly low prevalence levels among *E. coli* isolated from Mexican wildlife (Cristóbal-Azkarate et al., 2014), with three *E. coli* isolates from 138 howler monkey, spider monkey, and tapir fecal samples PCR positive using a primer set similar to HS464/HS463a. Using initial culture-based methods, Dolejská and colleagues (2009) reported comparatively higher *intI1* prevalence in black-headed gulls, with detection in 9/60 (15%) of antibiotic resistant *E. coli* isolates. Others have documented *intI1* detection in animals along an anthropogenic gradient, including a survey of 341 fecal E. coli isolates from 150 wild birds and mammals, 128 farm animals, and 42 companion dogs (Skurnik et al., 2006). Notably, *intI1* was detected in only 7% of farm animals and 16% of pets, with no detection among any wildlife samples. Finally, using methods most similar to those of the present work, McDougall et al.

(2019) investigated *intI1* carriage in fecal DNA extracts of gray-headed flying foxes (*Pteropus poliocephalus*) using primers HS464/HS463a (Stokes et al., 2006) and reported a clear difference in detection between wild (5.3%) and captive individuals (41.2%). The comparatively higher prevalence among wildlife species in the present study could be attributed to increased sensitivity of ddPCR over other methods.

Inverted ratio of clinical to general intI1

In the majority of clinical *intI1* detection events, the general target was found at a higher concentration as indicated by a clinical to general ratio < 1. However, in 22/121 (18%) of samples, the concentration of clinical *intI1* exceeded that of the general variant. This finding is counter to the Waldron and Gillings (2015) paradigm wherein a DNA fragment containing the clinical *intI1* sequence should also be PCR positive for the general target, resulting in a clinical:general ratio of 1:1. Among the 22 samples with an inverted ratio, twelve originated from humans, seven from unimpacted marine water, and three from giant tortoises. Approximately one third (8/22) of samples have a clinical:general ratio > 2, representing a significant departure from the one to one paradigm. While ratios close to 1 may simply be the result of technical variation in the assays, the more likely explanation is that the underlying sequence targeted by the general primers is more diverse than previously reported. Our understanding of class I integron-integrase variants, along with their ecology and distribution, is limited to existing sequences in databases which are in turn the product of a limited set of primers. Recently, Yang and colleagues (2021) proposed a sequencing-based approach to recover diverse class I integron sequences using newly designed primers. The newly described primer pair spans the entirety of the class I integron from the 5' conserved segment (CS) to the 3' CS. While the authors demonstrated superior specificity over existing primer pairs, this assay is

not well suited to ddPCR given the long and variable length of complete class I integrons. Nonetheless, efforts like those of Yang and colleagues will undoubtedly improve our understanding of these genetic elements and offer opportunities for further refinement of *intI1* targets suitable for environmental application.

Detection and abundance of mobile genetic elements (MGEs) by sample type

Additional MGEs beyond integrons were explored by mapping 90 metagenomes to the database constructed by Pärnänen et al. 2018, revealing a general trend of increasing MGE sum abundance along a gradient of anthropogenic influence. Mirroring the pattern in ARG sum abundance when annotating with the ARG-OAP.1, the highest mean MGE sum abundance/16S rRNA was recorded in wastewater, followed by humans, wildlife, and animals, though no pairwise comparisons were associated with significant test statistics. In agreement with observations regarding ARG sum abundance/16S rRNA and both *intI1* targets, MGEs were higher among land iguanas compared to other wildlife species. The inter-island differences recorded for ARGs and *intI1* between land iguanas were again corroborated by MGEs, with mean sum abundance/16S rRNA higher in individuals from both North Seymour and Santa Fe compared to Plaza Sur. Finally, and in contrast to ARG sums tabulated by ARG-OAP.1 or ResFinder, mean MGE sum abundance/16S rRNA was slightly greater in the Cesarean section group. Additionally, and in agreement with ARG data, more MGEs were differentially abundant in the Cesarean section birth group compared to the vaginal birth group. Collectively, these data align with prior observations regarding the correlation between the resistome and mobilome, including a study in a wastewater treatment train where Spearman correlation of the overall abundance of ARGs and MGEs yielded a significant rho equal to 0.748 (Makowska et al., 2016). Pärnänen et al. 2018 similarly demonstrated a high level of agreement between resistome and

mobilome distance matrices when annotating metagenomic sequences from infant gut microbiomes with the MGE database.

However, there were some unusual observations in the MGE data, specifically with the MGE sum abundance/16S rRNA in four land iguanas exceeding that of the highest wastewater sample. This raises important questions regarding the specificity and quantitative nature of mapping-based approaches. Given that land iguana gut microbial communities have been minimally studied, it could be that DNA fragments mapping to the MGE database have similar sequences but distinct ecological functions. Notably, none of the ten land iguana metagenomes mapped to *intI1*, whereas all metagenomes originating from wastewater and wastewater impacted marine sites mapped to *intI1*.

Agreement of MGE mapping and ddCPR quantification of intI1

Overall, we found that MGE mapping and ddPCR of the clinical class I integron-integrase agreed in distinguishing samples with significant mobilomes. A significant linear relationship was observed among metagenomes that mapped to *intI1* and the corresponding clinical *intI1* concentration in the same sample (R=0.86, p=2.08e-09). As could be anticipated, the clinical class I integron-integrase was detected more frequently by ddPCR compared to the mapping-based approach. Metagenomic approaches are inherently biased against low abundance targets (Podar et al., 2007) and therefore are better suited to describe broad differences in the most abundant features rather than quantify a specific target.

Conclusion

This dataset constitutes the first exploration of mobilomes in human, environmental, and wildlife reservoirs in the Galapagos islands. Overall, both MGE mapping and quantification of the class I integron-integrase gene point to increasing mobility along a gradient of anthropogenic influence. In agreement with antibiotic resistance gene annotations, land iguanas appear to

harbor gut microbial communities with elevated mobility potential compared to other wildlife species. Moreover, carriage of clinical *intI1* among wildlife animals of the same species seems to vary by location and associate closely with overall ARG burden, especially in the case of land iguanas. Detection of the general *intI1* sequence more frequently than the clinical *intI1* sequence using the ddPCR assay developed herein supports the notion that class I integrons are widely circulated across environments. While overall detection of *intI1* was higher compared to other studies in the literature, we hypothesize this to be the result of the increased sensitivity of ddPCR over initial culture-based methods, which account for the majority of *intI1* surveys in wildlife. Additionally, observation of several samples with clinical:general ratios > 1 suggests that the region targeted by general primer pairs is likely not as conserved as previously described and represents an opportunity for refinement of the assay. Finally, we noted considerable agreement between the values for *intI1*/16S rRNA from MGE mapping and clinical *intI1*/16S rRNA as determined by ddPCR, though ddPCR was more sensitive in terms of overall detection than the mapping approach. Taken together, data from this study further support the use of clinical *intI1* as a marker of anthropogenic influence and show that with the exception of land iguanas, Galapagos wildlife are characterized by lower sum abundances of MGEs compared to wastewater and humans.

Chapter 4: Supplemental Figures

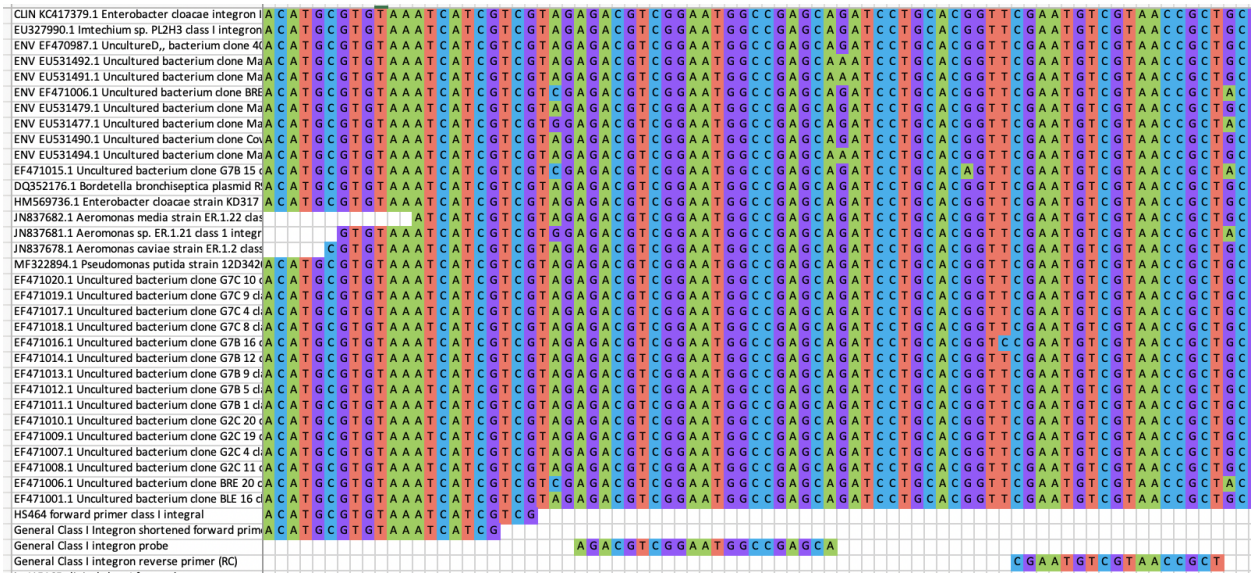

Figure S4.1: *MEGA X alignment of clinical and environmental class I integron-integrase sequences highlighting region of general intI1 assay, including the forward primer (5'- ACATGCGTGTAAATCATCG-3'), probe (5'- FAM-AGACGTCGGAATGGCCGAGCA-BHQ-1-3'), and reverse primer (5'- AGCGGTTACGACATTCG-3').*

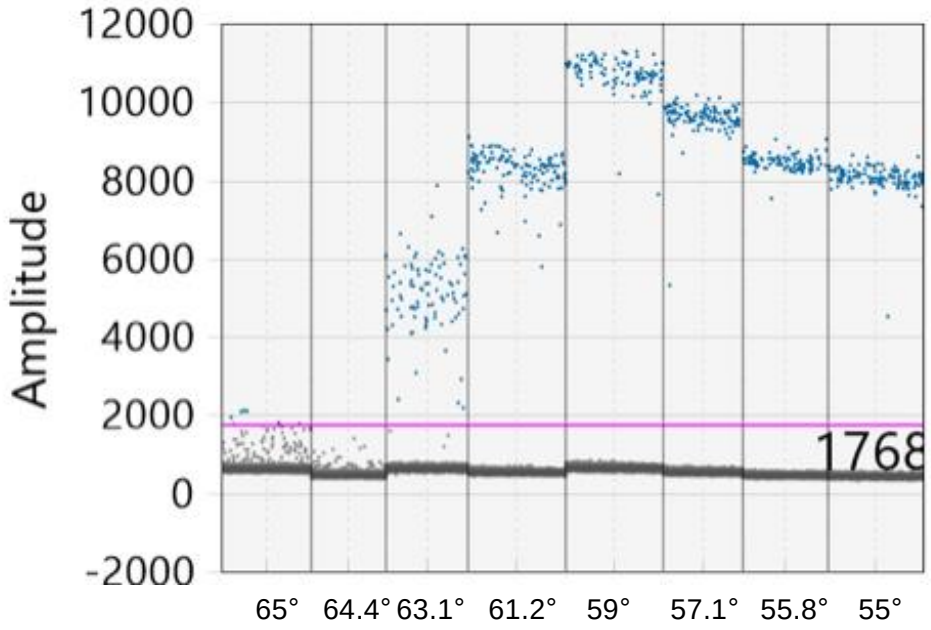

Figure S4.2: *Annealing temperature optimization for the general intI1 assay was performed across a range of 55℃ to 65℃.*

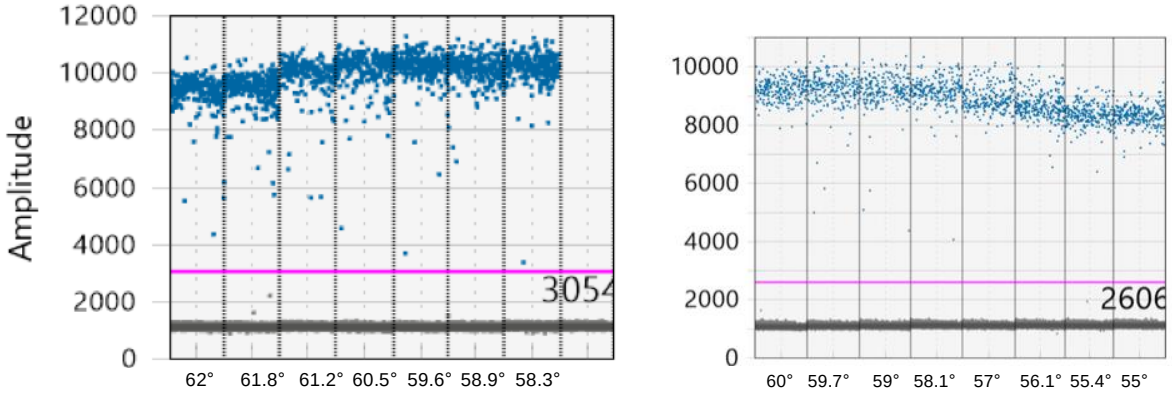

Figure S4.3: *Annealing temperature optimization for the clinical intI1 assay (Barraud et al., 2010) was performed across a range of 58.3 ℃ to 62 ℃ (left) and 55 ℃ to 60 ℃ (right).*

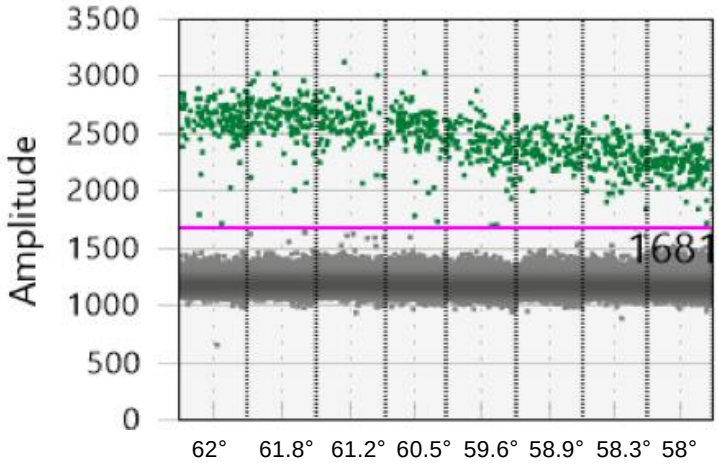

Figure S4.4: *Annealing temperature optimization for the 16S rRNA assay (Nadkarni et al., 2002) was performed across a range of 58 ℃ to 62 ℃.*

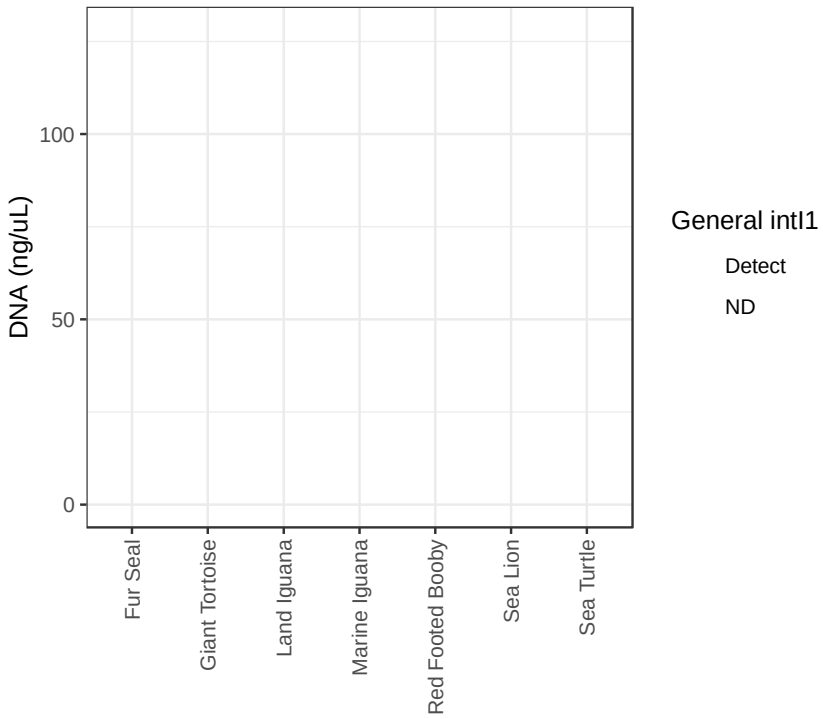

Figure S4.5: *DNA concentration among wildlife samples and detection of general class I integron-integrase.*

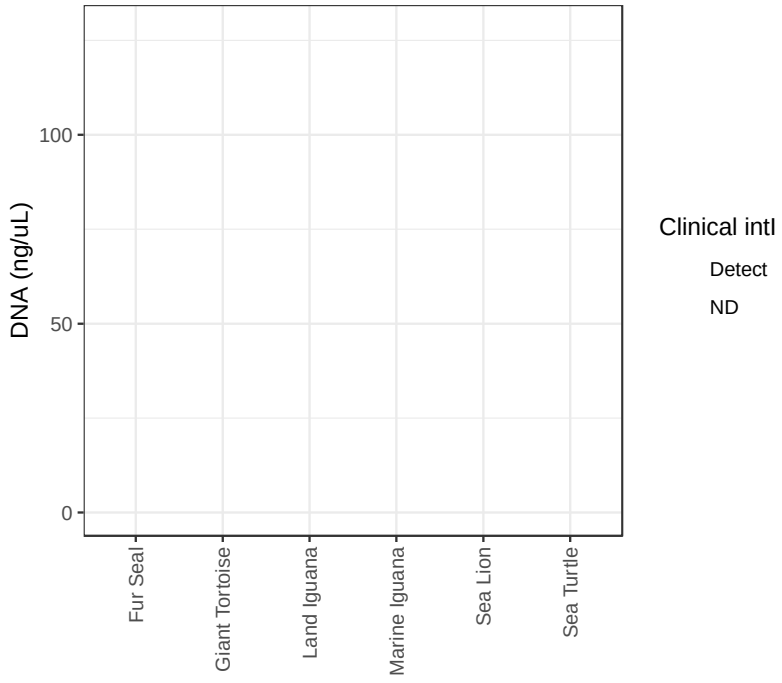

Figure S4.6: *DNA concentration among wildlife samples and detection of clinical class I integron-integrase.*

Chapter 4: Supplemental Tables

Table S4.1: *Tukey's post hoc test of pairwise comparisons of GLM-predicted MGE sum abundance/16S by sample type*

Comparison	Estimate	Std. Error	z value	Pr(>|z|)
W.Water - M.Iguana == 0	2.8958	0.7994	3.623	0.0126
R.Booby - M.Water* == 0	-3.2812	0.9232	-3.554	0.0152
S.Turtle - S.Lion == 0	-1.7504	0.561	-3.12	0.0623
S.Turtle - M.Iguana == 0	-2.1029	0.6758	-3.112	0.065
M.Water* - F.Water == 0	2.8558	0.9231	3.094	0.0672
M.Iguana - Human == 0	-1.7944	0.5958	-3.012	0.0853
L.Iguana - Human == 0	1.6635	0.5589	2.976	0.0931
S.Lion - M.Water* == 0	-2.0076	0.705	-2.848	0.1316
M.Water* - L.Iguana == 0	-1.8028	0.7722	-2.334	0.3927
M.Water - G.Tortoise == 0	1.6863	0.7269	2.32	0.4029
G.Tortoise - F.Water == 0	-1.8835	0.8432	-2.234	0.4632
M.Water* - M.Iguana == 0	1.6551	0.7994	2.071	0.5816
M.Water - M.Iguana == 0	-1.3978	0.6757	-2.069	0.5827
R.Booby - M.Iguana == 0	-1.6261	0.7995	-2.034	0.6075
S.Lion - M.Water == 0	1.0454	0.5608	1.864	0.7261
S.Lion - R.Booby == 0	1.2737	0.7052	1.806	0.7634
R.Booby - G.Tortoise == 0	1.458	0.8432	1.729	0.8085
M.Iguana - F.Water == 0	1.2006	0.7995	1.502	0.9129
W.Water - Human == 0	1.1013	0.7536	1.461	0.9263
S.Turtle - G.Tortoise == 0	0.9812	0.727	1.35	0.9561
W.Water - M.Water* == 0	1.2407	0.923	1.344	0.9572
S.Lion - F.Water == 0	0.8482	0.7051	1.203	0.9805
S.Turtle - F.Water == 0	-0.9022	0.8184	-1.102	0.9899
S.Turtle - M.Water == 0	-0.7051	0.698	-1.01	0.9949
W.Water - L.Iguana == 0	-0.5621	0.7722	-0.728	0.9997
S.Lion - M.Iguana == 0	-0.3524	0.5329	-0.661	0.9999
M.Water - F.Water == 0	-0.1972	0.8183	-0.241	1
R.Booby - F.Water == 0	-0.4255	0.9233	-0.461	1
M.Water* - Human == 0	-0.1393	0.7536	-0.185	1
R.Booby - M.Water == 0	-0.2283	0.8184	-0.279	1
S.Turtle - R.Booby == 0	-0.4768	0.8185	-0.582	1
Human - F.Water == 0	2.9951	0.7538	3.974	<0.01
L.Iguana - F.Water == 0	4.6585	0.7724	6.032	<0.01
W.Water - F.Water == 0	4.0964	0.9231	4.438	<0.01
Human - G.Tortoise == 0	4.8786	0.6533	7.468	<0.01
L.Iguana - G.Tortoise ==	6.542	0.6747	9.696	<0.01
M.Iguana - G.Tortoise ==	3.0841	0.7056	4.371	<0.01
M.Water* - G.Tortoise ==	4.7392	0.8431	5.621	<0.01
S.Lion - G.Tortoise == 0	2.7317	0.5965	4.579	<0.01
W.Water - G.Tortoise == 0	5.9799	0.8431	7.093	<0.01
M.Water - Human == 0	-3.1922	0.6209	-5.141	<0.01
R.Booby - Human == 0	-3.4205	0.7538	-4.538	<0.01
S.Lion - Human == 0	-2.1469	0.4615	-4.652	<0.01
S.Turtle - Human == 0	-3.8973	0.621	-6.276	<0.01
M.Iguana - L.Iguana == 0	-3.4579	0.6192	-5.585	<0.01
M.Water - L.Iguana == 0	-4.8557	0.6434	-7.547	<0.01
R.Booby - L.Iguana == 0	-5.084	0.7724	-6.582	<0.01
S.Lion - L.Iguana == 0	-3.8103	0.4913	-7.755	<0.01

S.Turtle - L.Iguana == 0	-5.5608	0.6435	-8.642	<0.01
M.Water* - M.Water == 0	3.0529	0.8182	3.731	<0.01
W.Water - M.Water == 0	4.2936	0.8182	5.247	<0.01
S.Turtle - M.Water* == 0	-3.758	0.8183	-4.592	<0.01
W.Water - R.Booby == 0	4.5219	0.9232	4.898	<0.01
W.Water - S.Lion == 0	3.2482	0.705	4.608	<0.01
W.Water - S.Turtle == 0	4.9986	0.8183	6.108	<0.01

Table S4.2: *Tukey's post hoc test of pairwise comparisons of GLM-predicted MGE sum abundance/16S among sea lions by location*

Comparison	Estimate S	td. Error z	value P	r(>|z|)
Punta Mangle/Fernandina - El Malecon/San Cristobal == 0	3.9279	1.1192	3.51	0.00586
El Malecon/San Cristobal - Cabo Douglas/Fernandina == 0	-3.191	0.9139	-3.492	0.00615
Punta Mangle/Fernandina - Champion/Floreana == 0	3.2287	1.1191	2.885	0.04443
Champion/Floreana - Cabo Douglas/Fernandina == 0	-2.4918	0.9138	-2.727	0.06865
Puerto Egas/Santiago - El Malecon/San Cristobal == 0	2.2708	0.867	2.619	0.09102
Punta Pitt/San Cristobal - Punta Mangle/Fernandina == 0	-2.4114	1.081	-2.231	0.22052
Punta Pitt/San Cristobal - Cabo Douglas/Fernandina == 0	-1.6744	0.8668	-1.932	0.37853
Puerto Egas/Santiago - Champion/Floreana == 0	1.5716	0.8669	1.813	0.45326
Punta Pitt/San Cristobal - El Malecon/San Cristobal ==	1.5166	0.867	1.749	0.49502
Punta Mangle/Fernandina - Puerto Egas/Santiago == 0	1.6571	1.081	1.533	0.639
Puerto Egas/Santiago - Cabo Douglas/Fernandina == 0	-0.9202	0.8668	-1.062	0.89493
Punta Pitt/San Cristobal - Champion/Floreana == 0	0.8173	0.8669	0.943	0.93429
Punta Pitt/San Cristobal - Puerto Egas/Santiago == 0	-0.7542	0.8172	-0.923	0.93975
El Malecon/San Cristobal - Champion/Floreana == 0	-0.6992	0.914	-0.765	0.97283
Punta Mangle/Fernandina - Cabo Douglas/Fernandina == 0	0.7369	1.119	0.659	0.98606

Table S4.3: *MGEs differentially abundant according to birth mode*

MGE name	Log2FoldChange	P adjusted	Class	Gene	n
Elevated in Caesarean section birth group (positive fold change)					
505_tnpA4_AF550679.1	8.36272865	1.00E-08	transposase	tnpA4	8
533_tnpA_CBTV010000137.1	8.29842459	0.0001034	transposase	tnpA	1
1761_tnpA_HM370391.1	8.11762207	0.0001034	transposase	tnpA	2
123_repUS1__ORF(E.faeciumContig1258)_JDOE	7.91751856	0.0001034	plasmid	repUS1	4
718_tnpA_KF680002.1	7.89046131	0.0001034	transposase	tnpA	2
921_tnpAB_EU402605.1	7.59956049	0.00024864	transposase	tnpAB	4
545_tnpA(ISnew)_AJ698325.1	7.46642205	0.00016392	transposase	tnpA(ISnew)	2
1337_tnpA_AP001918.1	6.95183439	0.00051289	transposase	tnpA	3
246_Col(MG828)_1__NC_008486	6.80644254	0.00045343	plasmid	Col(MG828)	7
765_tnpA_HE613569.1	6.72790536	0.00048337	transposase	tnpA	1
Elevated in vaginal birth group (negative fold change)					
2756_tnpA_KX810026.1	-6.9483439	0.00027628	transposase	tnpA	2
521_tnpA_AJ318089.1	-6.3951563	0.00053337	transposase	tnpA	2
320_tnpA_AJ534881.1	-5.4429293	0.00112729	transposase	tnpA	3
688_tnpA5_HG916826.1	-5.2029264	0.00170554	transposase	tnpA5	2
137_repUS15__ORF(E.faecium287)_NZAAAK010000287	-4.3750268	0.01140971	plasmid	repUS15	2
1180_tnpA_AM286690.1	-3.7548789	0.01896396	transposase	tnpA	2

1544_tnpA_AF408195.1	-3.6629557	0.02784127	transposase	tnpA	1
1881_IS91_MNQS01000105.1	-3.5849535	0.03142428	insertion_element_IS91	IS91	1
80_rep18_1_repA(p200B)_AB158402	-3.5235532	0.03389981	plasmid	rep18	1
1585_tnpA1000_KX709966.1	-3.4712957	0.0424654	transposase	tnpA1000	2

CHAPTER 5: WHO ARE POSSIBLE BACTERIAL HOSTS OF ANTIBIOTIC REISSTANCE GENES? CHARACTERIZATON OF THE MICROBIAL COMMUNITY

Introduction

Antibiotic resistance as a global public health challenge is often framed in the context of pathogens, as we rely on antibiotics to treat infections caused by bacteria harmful to humans and animals. While the increase of ARGs and MGEs circulating among the entirety of bacteria, both those human associated and of environmental origin is worrying in regards to horizontal gene transfer, our concerns ultimately center on the acquisition of resistance by pathogenic organisms and consequent morbidity and mortality from these infections. In 2017, the World Health Organization (WHO) released a list of twelve priority pathogens for which research and development of new antimicrobial therapies is desperately needed in order to mitigate projections likes those from the Review on Antimicrobial Resistance (O'Neil, 2016). Therefore, the question of *who* in regards to the host of antibiotic resistance genes and their mobility is an essential component of environmental AMR surveillance.

The question of *who*, however, is difficult to answer at scale: researchers can choose to culture a group of microorganisms and test for phenotypic resistance and genotypic resistance. This method, used by Thaller and colleagues (2010) and Wheeler and colleagues (2012) in the first surveys of antibiotic resistance among Galapagos wildlife, provides the most reliable confirmations of host identify and functional resistance, though at the cost of missing resistance among organisms not amenable to culture or among the larger bacterial community. As discussed in Chapter 4, screenings of the class I integron-integrase that rely on initial culturing of

target organisms yielded considerably lower detection compared to ddPCR interrogation of total genomic DNA. On the other hand, researchers can employ culture-independent approaches involving 16S rRNA amplicon or shotgun metagenomic sequencing and recover DNA from many thousands of organisms. This approach, however, relies on taxonomic inference, and any correlations between specific ARGS and taxa are speculative. Recently, Emulsion, Paired Isolation, and Concatenation (epic) PCR has been proposed as molecular tool linking the 16S rRNA gene with target ARGs of interest in the same reaction (Spencer et al., 2016). However, this strategy is technically difficult, requires significant optimization for each gene target, and permits surveillance of only several ARGs at a time.

In this chapter, we aimed to provide phylogenetic context to characterizations of the resistome (Chapter 3) and mobilome (Chapter 4) of wastewater, water, wildlife, and human samples from the Galapagos using a combined approach of 16S rRNA amplicon sequencing for a subset of 50 samples and taxonomic inference on the larger metagenomic dataset. While we acknowledge that this approach cannot definitely link specific taxa with ARGs, we consider microbial community profiling of Galapagos wildlife species to be a worthwhile scientific contribution in itself: Only two studies have provided high-throughput characterizations of the land iguana gut microbiome (Hong et al., 2011; Hong et al., 2015) and only one dataset exists for Galapagos giant tortoises, with data originating from only four individuals on San Cristobal (Hong et al., 2011). To our knowledge, this represents the first microbial community profiling of the Galapagos sea lion. Given the importance of the gut microbiome to overall health, including key roles in digestion, nutrition, and immune system training (Hanning and Diaz-Sanchez, 2015), some researchers have recommended gut microbial profiling as a wildlife conservation strategy (Bahrndorff et al., 2016; Redford et al., 2012). Bacteria play both beneficial and harmful roles in

wildlife health, with the nature of the particular interaction sometimes changing across space and time (Redford et al., 2012). For example, commensal bacteria may start to breakdown the host's intestinal mucin in the absence of sufficient nutrients or during fasting (Hanning and Diaz-Sanchez, 2015), while increasing temperature can shift bacteria from mutualistic to pathogenic interactions in coral (Redford et al., 2012). These examples are particularly relevant to the Galapagos, where El Niño events drastically influence algal availability to marine iguanas (and disrupt the food chain in general), and climate change is already warming waters (Casey, 2018). In light of these imminent environmental pressures, better understanding of the relationships between bacteria and their wildlife hosts may reveal information useful to the conservation of Galapagos wildlife.

Materials and Methods

16S rRNA community profiling

A subset of 50 wildlife fecal DNA extracts was subjected to microbial community profiling via amplicon sequencing of the 16S rRNA gene. This included 24 sea lion, 4 giant tortoise, and 10 land iguana samples from the metagenomic data set, as well as an additional 6 giant tortoise samples from each Otoy Ranch and La Galapaguera collected in March 2019. In the case of the land iguana samples, a replacement land iguana from Santa Fe island (G19_23) was selected due to insufficient material from the sample G19_9 included in the metagenomic dataset. Sea lion, giant tortoise, and land iguana samples were chosen for 16S rRNA sequencing specifically to 1) explore differences in the fecal microbial communities of reptilian and mammalian Galapagos wildlife, and 2) investigate intra-species differences by location. Additional details about the samples sent for 16S rRNA sequencing are provided in **Table S5.1**. Sample collection and DNA extraction was performed as described in Chapter 3.

Library preparation and sequencing

Extracted DNA was shipped on ice to the Argonne National Laboratory Environmental Sequencing facility (Lemont, IL) for library preparation and 16S rRNA amplicon sequencing. Amplification of the V4 variable region of the 16S rRNA gene (515F-806R) was performed using primers adapted for the Illumina MiSeq platform (Caporaso et al., 2012). Forward primers included a 12-base barcode sequence to facilitate pooling of samples in the flow cell. Pooled libraries were cleaned using AMPure XP Beads (Beckman Coulter, Pasadena, CA) and quantified using the Qubit4 fluorometer. Amplicons were sequenced on the Illumina MiSeq platform generating 2x151 bp reads.

Bioinformatics and statistical analysis

Initial processing of 16S rRNA sequence data was performed in QIIME 2 (Bolyen et al., 2019) version 2020.11. Briefly, paired reads were joined, demultiplexed, and assessed for quality. The median quality score was ≥ 33 across the length of both the forward and reverse reads and no additional trimming was performed prior to denoising. Denoising, including filtering of low quality, unmerged, and chimeric reads, was executed with the DADA2 plug-in (dada2 denoise-paired). As a result of denoising, 88.2 ± 3.1% of sequences were retained. The resulting amplicon sequence variant (ASV) table included 4,056 features with a median of 50,390 sequences/sample (ranging from 22,362 to 72,420 sequences/sample). For taxonomy assignment, a Naïve Bayes classifier was trained using the SILVA-138 database (Quast et al., 2013) at 99% sequence identity against the 515F-806R region of the 16S rRNA gene. Prior to differential abundance analyses, the ASV table was filtered to exclude ASVs observed 10 or fewer times across the dataset and ASVs observed in only one sample. The filtered table was summarized at the level of phyla (L2), order (L4), family (L5), and genus (L6) and ANCOM was

implemented in QIIME 2 to estimate significant differences in differential abundance between wildlife species. ANCOM was utilized in the same manner to evaluate intra-species differential abundance by location. For visualization of taxonomic differences with relative abundance bar charts, additional ASV tables were produced where ASVs observed fewer than 10 times across the dataset were filtered, but those observed in only one sample were retained.

Additional analyses were performed in R software version 4.0.5 using the packages phyloseq (McMurdie and Holmes, 2013) version 1.34.0 and vegan (Oksanen et al., 2019) version 2.5.7. Briefly, the ASV table (table.qza), tree (rooted-tree.qza), taxonomy table (taxonomy-silva-trained-515-806.qza), and QIIME 2 compliant metadata file were imported as a phyloseq object using qza_to_phyloseq from the R package qiime2R (Bisanz, 2021) version 0.99.6. The resulting phyloseq object was rarified at a depth of 20,000 sequences/sample. Alpha diversity estimates were calculated in phyloseq using the Shannon and Simpson diversity indices. Differences between sample groupings (i.e. wildlife species) were calculated using analysis of variance (ANOVA) and p-values were adjusted using Tukey's post hot test. Prior to calculation of beta diversity estimates, the phylogenetic tree within the phyloseq object was corrected to a binary tree using multidi2di within the R package ape (Paradis et al., 2021) version 5.5. Beta diversity estimates were calculated in phyloseq based on weighted Unifrac distances and the Bray-Curtis dissimilatory index and visualized through Principal Coordinate Analysis (PCoA). The Multivariate Analysis of Variance Using Distance Matrices (ADONIS) in vegan was implemented with 9,999 permutations to assess to extent to which categorical variables (i.e. species) explained variation in the distance matrix. To perform intra-species comparisons by location, the original phyloseq object was filtered by wildlife species, and each wildlife species-specific phyloseq object re-rarified according to the lowest sequence count in the group (22,362

for sea lions; 44,086 for land iguanas, and 38,556 for giant tortoises.) Estimations of alpha and beta diversity as well as the associated statistical comparisons by location were carried out as described above.

To explore taxa explaining differences in ARG and MGE sums, the rarified ASV table was summarized at the level of phyla (L2) and family (L5). The resulting tables were filtered to include 1) only samples for which paired metagenomes were available, including 4 giant tortoises, 24 sea lions, and 9 land iguanas; and 2) only taxa present in 20% of samples (≥7 of 37 samples.) Spearman rank correlation coefficients and associated FDR adjusted p-values were calculated between taxa counts, ARG sums (from both ARG-OAP.1 and ResFinder), and MGE sums. Finally, to assess the agreement of bacterial community profiling from16S rRNA amplicon sequencing and assignment of SSU rRNA sequences in metagenomes, the Horn-Morisita similarity index was calculated in vegan for the L7 Metaxa table for the same 37 samples and compared to the distance matrix from the 16S rRNA (i.e. QIIME 2) ASV table using the Mantel test.

Taxonomic classification of Metagenomes

The taxonomic classifications of SSU rRNA sequences generated using Metaxa2 version 2.2 (Bengtsson-Palme et al., 2015) in Chapter 3 were further analyzed to summarize observations at different taxonomic levels. First, the Metaxa2 Taxonomic Traversal Tool (metaxa2_ttt) was used to summarize observations at the level of domain, phyla, class, order, family, genus, and species for SSU rRNA sequences assigned to bacteria and archaea with the flag -t b,a. Next, the Metaxa2 Data Collector (metaxa2_dc) was implemented to aggregate results from each sample into abundance matrices at each taxonomic level. The species-level abundance matrix (L7) was imported into R and combined with its respective taxonomy matrix and metadata to produce a

phyloseq object. Samples with SSU rRNA counts totaling less than 5,000 were excluded, and the resulting table was rarified at a depth of 5,000 SSU rRNA counts/sample. This eliminated 14 of 90 total samples, including two marine water samples, six marine iguanas, three sea turtles, two land iguanas, and one red-footed booby. Taxa relative abundance was calculated on the rarified table at the levels of phyla (L2), family (L5), and genus (L6), and mean relative abundance was calculated by sample type. Taxa accounting for ≥ 1% relative abundance in at least one sample type were retained for visualization with bar charts.

Statistical analysis

Alpha diversity estimates were calculated in phyloseq using the Shannon and Simpson diversity indices. Differences between sample groupings were calculated using analysis of variance (ANOVA) and p-values were adjusted using Tukey's post hot test. Beta diversity estimates were calculated in phyloseq based on the Bray-Curtis dissimilatory index and visualized through Principal Coordinate Analysis (PCoA). The Multivariate Analysis of Variance Using Distance Matrices (ADONIS) in vegan was implemented with 9,999 permutations to assess to extent to which categorical variables (i.e. species) explained variation in the distance matrix.

To explore possible relationships between specific taxa and ARG and MGE sums, Spearman rank correlations were performed at the levels of phyla (L2), family (L5), and genus (L6) against the sum abundance ARGs/16S (both ARG-OAP.1 and ResFinder data) and sum abundance MGEs/16s (MGE mapping). As above, taxa tables were filtered to exclude samples with fewer than 5,000 SSU rRNA counts, corresponding to 76 samples. The resulting tables were rarified at 5,000 SSSU rRNA counts/sample. Only taxa present in ≥ 20% of samples were considered in the analysis. Resulting p-values were corrected using the false discovery rate

(FDR) correction method for the respective number of pairwise comparisons. This analysis was performed on all 76 samples included in the rarified tables and on a subset that included only water and wildlife samples. The relationship between the family *Enterobacteriaceae* and ARG and MGE sums was further examined by calculating the Kendall rank correlation coefficient between counts of *Enterobacteriaceae* from the table rarified at 5,000 SSU rRNA counts/sample and ARG sum abundance/16S (as tabulated by both ARG-OAP.1 and ResFinder) and MGE sum abundance/16S. This comparison was performed on all 76 samples included in the rarified table and on subsets of sample types (i.e. human, wildlife, water.) The resulting p-values were corrected using the false discovery rate (FDR) correction method for the respective number of pairwise comparisons.

Results

16S rRNA amplicon sequencing of giant tortoise, sea lion, and land iguana gut microbiomes

Samples from a total of 50 Galapagos wildlife animals, including 16 giant tortoises, 24 sea lions, and 10 land iguanas were subjected to microbial community profiling via 16S rRNA amplicon sequencing. To our knowledge, this effort represents the first 16S rRNA data set for Galapagos sea lion gut microbial communities. Initial data processing in QIIME 2 (Bolyen et al., 2019) produced an amplicon sequence variant (ASV) table with 4,056 features, with a median of 50,390 sequences/sample (minimum 22,362).

Alpha and beta diversity: interspecies comparison

Giant tortoise gut microbial communities harbored the highest alpha diversity, followed by land iguanas and sea lions (**Figure 5.1**). All pairwise comparisons of the Shannon diversity index yielded significant test statistics, while only the difference between giant tortoises and sea lions was significant using the Simpson index (ANOVA, Tukey's post hoc test, adjusted $p<0.05$). In terms of microbial community composition, samples grouped according to wildlife

host species using both taxonomic (weighted Unifrac) and non-taxonomic (Bray-Curtis dissimilarity index) estimations of distances between samples (**Figure 5.2**). Using the Bray-Curtis dissimilarity index, wildlife host species explained 38.8% of distances (ADONIS with 9,999 permutations, p=1e-04), and all samples grouped distinctly by species with no overlap when ellipses were drawn at the 90% confidence level. When taxonomy was considered using weighted Unifrac distances, the effect size decreased slightly to 35.6% (ADONIS with 9,999 permutations, p=1e-04) with some overlap observed between the reptilian species when ellipses were drawn at the 90% confidence level.

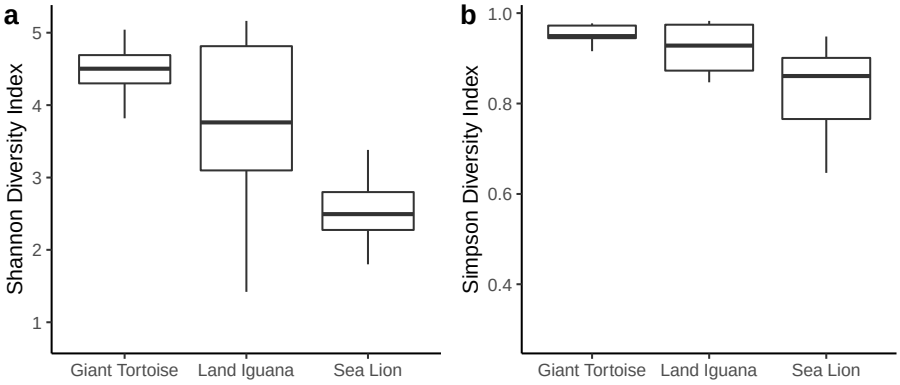

Figure 5.1: *Alpha diversity of the gut microbial communities of giant tortoise, land iguana, and sea lion samples.* **a)** Shannon diversity index. All pairwise comparisons are significant (ANOVA, Tukey's post hoc test, adjusted $p < 0.05$). **b)** Simpson diversity index. Only the difference between giant tortoises and sea lions is significant (ANOVA, Tukey's post hoc test, adjusted $p < 0.05$)

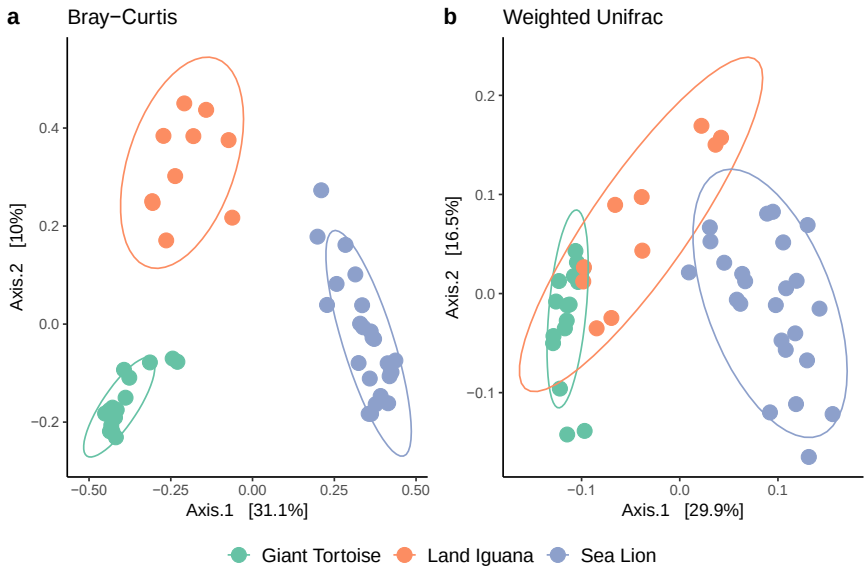

Figure 5.2: *Host species type influences bacterial community composition.* **a)** Bray-Curtis dissimilarity index (R=0.388, p=1e-04). **b)** Weighted Unifrac distances (R=0.356, p=1e-04). ADONIS test with 9,999 permutations.

Alpha and beta diversity: intraspecies comparison by location

We next examined whether diversity differed between individuals of the same wildlife species sampled at different locations. Among sea lions, the lowest alpha diversity was recorded among individuals from Puerto Egas, Santiago (**Figure 5.3, left**), with the difference significant compared to all other sampling locations except Cabo Douglas when using the Simpson diversity index (ANOVA, Tukey's post hoc test, adjusted p <0.05). Only individuals from El Malecon, San Cristobal and Punta Pitt, San Cristobal had significantly higher alpha diversity compared to the Puerto Egas population when using the Shannon index (ANOVA, Tukey's post hoc test, adjusted p <0.05). In terms of beta diversity, sampling location was associated with a significant test statistic for both the Bray-Curtis dissimilarity index (R-square=0.36, p=2e-04) and weighted Unifrac distances (R-square=0.34, p=0.0167), but unlike the interspecies comparison, clear clustering patterns were not observed (**Figure S5.1**). Considering land iguanas, the highest alpha diversity was recorded at Plaza Sur and the lowest among individuals from Santa Fe (**Figure**

S5.2), but no pairwise comparisons were significant for either the Simpson or Shannon diversity metric (ANOVA, Tukey's post hoc test, adjusted p>0.05). Sampling location explained differences in the gut microbial composition of land iguanas when considering the Bray Curtis dissimilarity index (R-square=0.33, p=0.016) but not weighted Unifrac distances (R-square=0.28, p=0.20, ADONIS with 9,999 permutations, data not shown). Finally, comparison of the two populations of giant tortoises revealed significantly higher alpha diversity at La Galapaguera when using the Simpson index (ANOVA, Tukey's post hoc test, adjusted p <0.05), but the difference when using the Shannon index was not significant (**Figure 5.3, right**). As was noted for land iguanas, the sampling location of giant tortoises significantly explained differences in the Bray-Curtis dissimilarity index (R-square=0.15, p=0.016) but not weighted Unifrac distances (R-square=0.14, p=0.068, ADONIS with 9,999 permutations, data not shown).

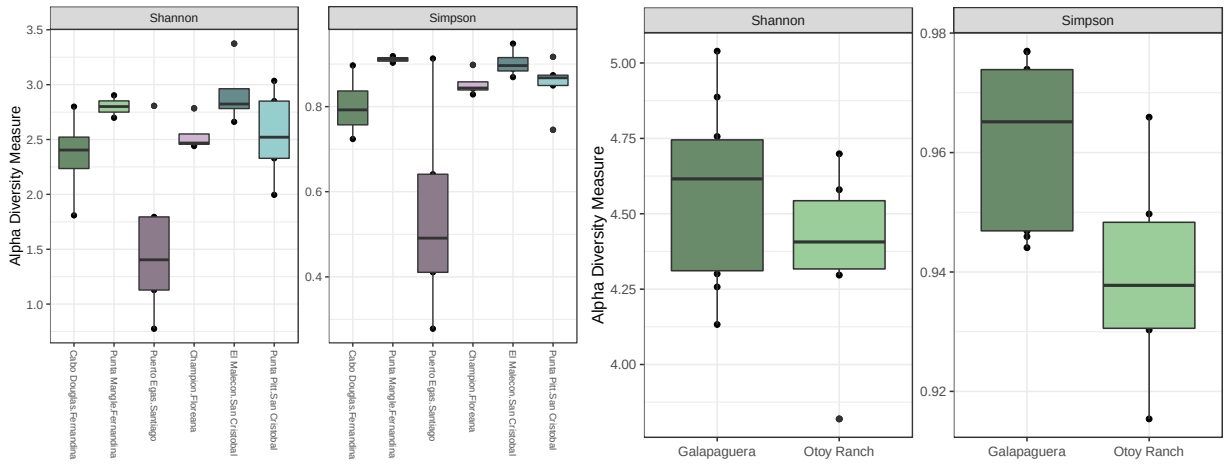

Figure 5.3*: Alpha diversity in sea lions (left) and giant tortoises (right) by location.*

Relative abundance of bacterial taxa by wildlife species

We subsequently investigated the relative abundance of bacterial taxa between wildlife species (**Figure 5.4, Table 5.1**). Across the data set, five phyla accounted for 95% of the

observed ASVs including Firmicutes (72.1%), Bacteroidetes (7.5%), Proteobacteria (6.7%), Fusobacteria (5.4%), and Actinobacteria (3.3%). Only two additional phyla accounted for overall abundances >1%, including Verrucomicrobia (1.5%) and Planctomycetes (1.5%). Firmicutes constituted the most abundant phyla in each of the three wildlife species, ranging from 62.9% in land iguanas to 78.6% in giant tortoises. Bacteroidetes was the only other phyla to occupy a position among the top five most abundant phyla in each wildlife species. Giant tortoises featured Planctomycetes, Verrucomicrobia, and Halobacterota (archaea) among its top five most abundant phyla, while land iguanas and sea lions shared Proteobacteria and Actinobacteria. Like giant tortoises, land iguanas also had Verrucomicrobiota among their top five most abundant phyla, while sea lions were uniquely characterized by ASVs belonging to Fusobacteria.

The most abundant bacterial family differed for each wildlife species, with *Clostridiaceae* accounting for 33.9% of ASVs in giant tortoises, compared to *Peptostreptococcaceae* (29.7%) in sea lions and *Enterobacteriaceae* (15.8%) in land iguanas. *Clostridiaceae* and *Lachnospiraceae* were among the top five most abundant bacterial families for all wildlife species, though their relative abundances varied.

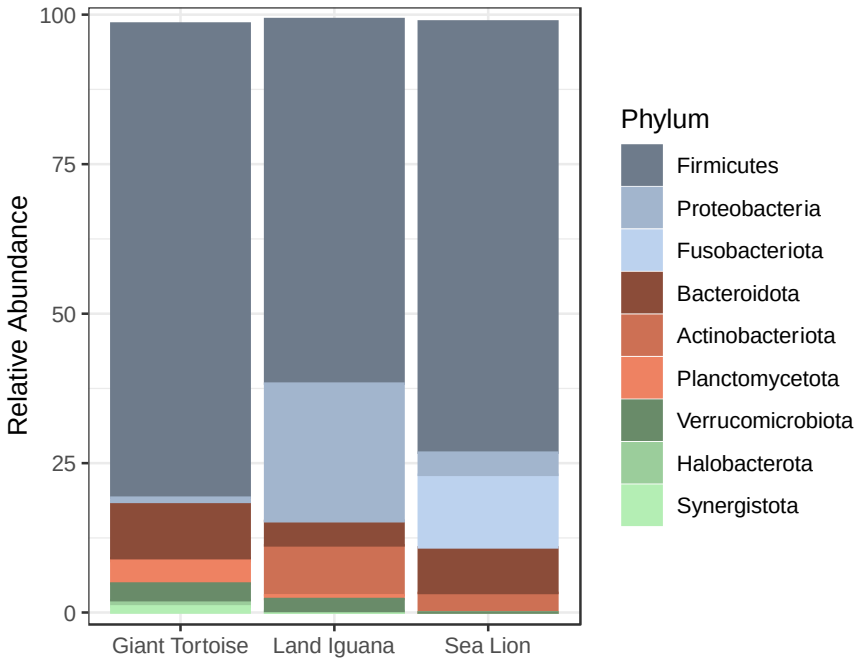

Figure 5.4: *Most abundant phyla in three wildlife species.* Phyla were filtered to retain only those accounting for mean relative abundance ≥1% in at least one wildlife species.

Table 5.1: *Mean relative abundance of phyla, families, and genera by wildlife species using 16S rRNA amplicon sequencing analyzed using QIIME 2.*

Species	Phylum	Rel. abund. (%)	*Family*	Rel. abund. (%)	*Genus*	Rel. abund. (%)
Giant Tortoise	Firmicutes	78.6	*Clostridiaceae*	33.9	*Sarcina*	23.1
	Bacteroidetes	9.7	*Lachnospiraceae*	15.8	*Clostridium sensu stricto 1*	10.6
	Planctomycetes	3.7	*Christensenellaceae*	6.7	*Lachnospiraceae, genus unknown*	9.4
	Verrucomicrobia	3.0	*Rikenellaceae*	4.6	*Christensenellaceae R-7 group*	6.1
	Halobacterota	1.1	*Pirellulaceae*	3.7	*Rikenellaceae RC9 gut group*	4.3
Sea Lion	Firmicutes	72.3	*Peptostreptococcaceae*	29.7	*Peptoclostridium*	27.5
	Fusobacteria	12.0	*Lachnospiraceae*	22.5	*Fusobacterium*	12.6
	Bacteroidetes	7.4	*Fusobacteriaceae*	12.7	*Marvinbryantia*	8.2
	Proteobacteria	3.9	*Clostridiaceae*	10.0	*Clostridium sensu stricto 1*	7.8
	Actinobacteria	3.3	*Bacteroidaceae*	6.4	*Bacteroides*	6.4
Land Iguana	Firmicutes	62.9	*Enterobacteriaceae*	15.8	*Clostridium sensu stricto 1*	9.8
	Proteobacteria	21.4	*Clostridiaceae*	14.8	*Escherichia-Shigella*	8.2
	Actinobacteria	7.7	*Lachnospiraceae*	12.6	*Enterobacteriaceae, genus unknown*	7.7
	Bacteroidetes	4.4	*Bacillaceae*	5.4	*Cellulosilyticum*	5.9
	Verrucomicrobia	2.0	*Moraxellaceae*	5.2	*Acinetobacter*	5.2

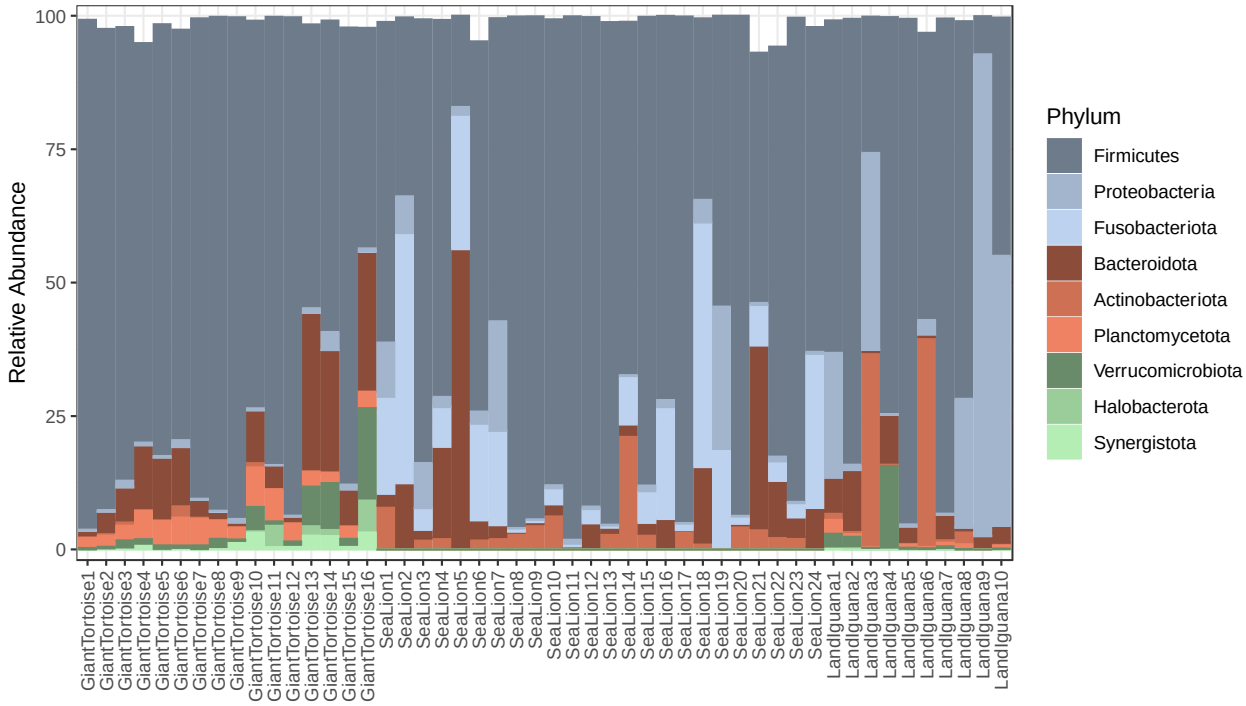

Figure 5.5: *Most abundant phyla in 50 wildlife samples.* Phyla represented are those included for visualization in Figure 5.4.

When comparing the relative abundance of phyla for each of 50 samples individually rather than by species, several differences can be appreciated according to sampling location (**Figure 5.5**). For example, giant tortoises originating from La Galapaguera (Giant Tortoise 7-16) appeared to have a higher relative abundance of Verrucomicrobia and Halobacterota compared to their counterparts at Otoy Ranch (Giant Tortoise 1-6). Similarly, land iguanas from Santa Fe (Land Iguana 8-10) had a greater proportion of Proteobacteria compared to individuals from Plaza Sur (Land Iguana 4-7).

Differential abundance of bacterial taxa between wildlife species

We subsequently explored if select phyla, families, and genera were differentially abundant between wildlife species using ANOSIM as implemented in QIIME 2. At the phylum level, 19 phyla differentially abundant between giant tortoises, sea lions, and land iguanas were

associated with significant W test statistics, with the most extreme being Planctomycetes (clr=335, W=24), a phylum relatively more abundant in giant tortoises and land iguanas compared to sea lions (**Figure 5.6**). In contrast, Fusobacteria were significantly more abundant in sea lions compared to the reptilian species (clr=280, W=25). ASVs assigned to Verrucomicrobia were significantly more abundant in land iguanas and giant tortoises compared to sea lions (clr=95, W=24), while Euryarchaeota were uniquely abundant in land iguanas (clr=93, W=25). Among the 44 differentially abundant families associated with a significant W statistic (**Figure 5.7**), the most extreme was *Christensenellaceae,* a newly described Firmicutes family which was more abundant in microbiomes from the two reptilian species compared to sea lions (clr=756, W=145). Additional families differentially abundant in reptiles include *Oscillospirales* UCG-010 (clr=474, W=144), the Bacilli family RF39 (clr=435, W=145), and *Pirellulaceae* (W=362, W=144). *Fusobacteriaceae* (clr=343, W=145) and *Peptostreptococcaceae* (clr=59, W=135) were among the families more abundant in sea lions compared to giant tortoises and land iguanas. Finally, 87 genera differentially abundant by species were associated with significant W test statistics (**Figure S5.3**).

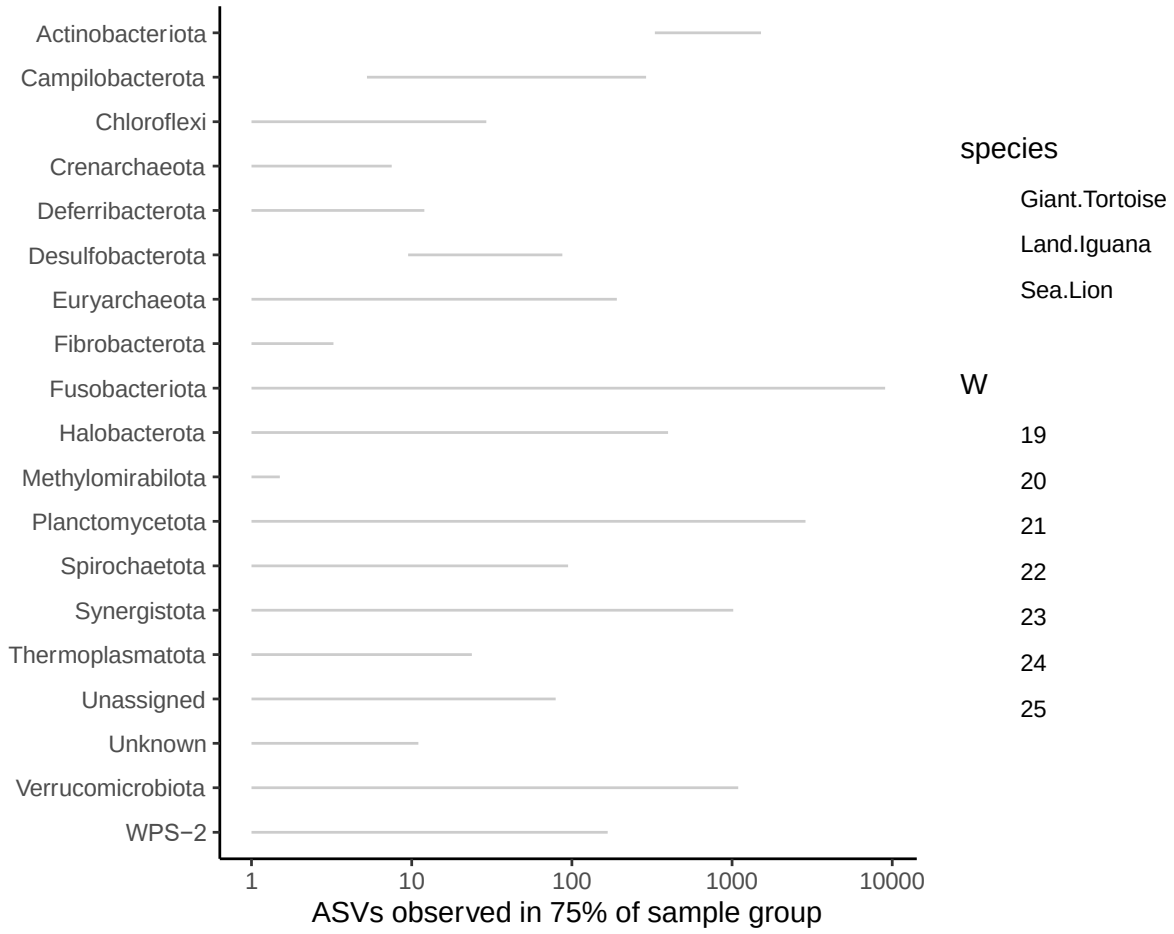

Figure 5.6: *Differentially abundant phyla between three wildlife species according to the ANOSIM test implemented in QIIME 2.* Data presented are ASVs observed in 75% of sample group (i.e. the upper quartile from the ANOSIM test.) The size of the data point corresponds to the strength of the test statistic W.

Additionally, we asked if certain taxa were differentially abundant between individuals of the same species from distinct sampling locations. The greatest number of differentially abundant taxa were recorded among sea lions, where at the level of order Actinomycetales (W=61) and Eubacteriales (W=58) were detected among sea lions from Champion, Floreana but absent at all other sampling locations. Two genera, *Clostridium sensu stricto 2* (W=296) and *Ruminococcus torques group* (W=275) and were differentially abundant among sea lions, with the former detected at Puerto Egas, Cabo Douglas, Punta Mangle, and Champion, but absent at

Punta Pitt and El Malecon. *Ruminococcus torques group* was detected at Cabo Douglas and Champion but absent from all other sites. Among giant tortoises, the Cyanobacterial order Chloroplast was differentially abundant among the Otoy Ranch population compared to La Galapaguera, though unlike sea lions no taxa were differentially abundant at the level of genus. Notably, no taxa at either the level of order or genus were found to be differentially abundant between land iguanas from the three distinct islands.

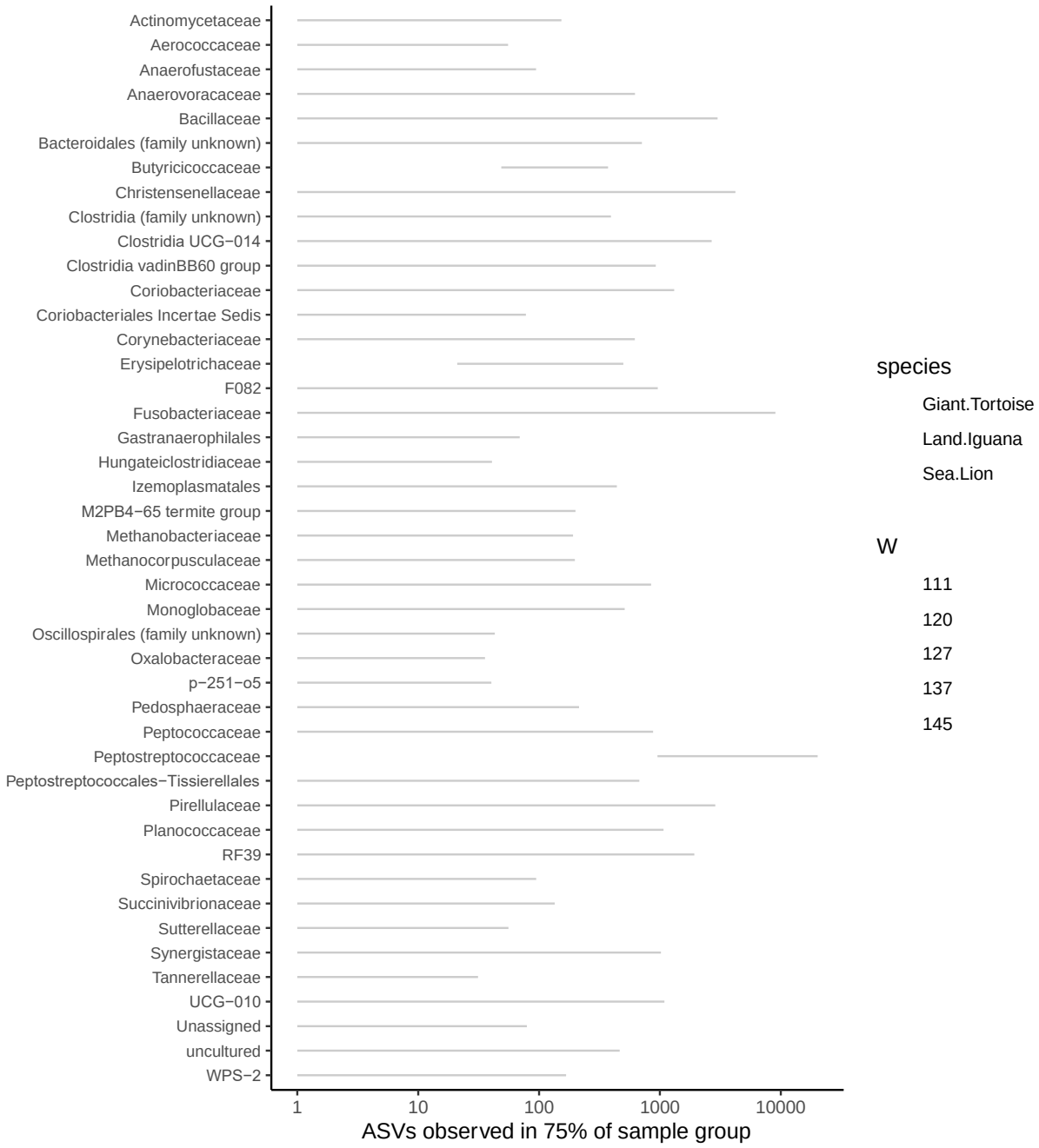

Figure 5.7: *Differentially abundant families between three wildlife species according to the ANOSIM test implemented in QIIME 2.* Data presented are ASVs observed in 75% of sample group (i.e. the upper quartile from the ANOSIM test.) The size of the data point corresponds to the strength of the test statistic W.

Correlations between bacterial taxa, ARGs, and MGEs

Given that select taxa, ARG sums (Chapter 3) and MGE sums (Chapter 4) differed between wildlife species, we next aimed to identify particular taxa associated with over ARG and MGE sum abundance/16S rRNA beginning at the level of phyla. Spearman rank correlations were calculated between the 37 samples for which paired 16S rRNA amplicon sequencing data and metagenomic annotation of ARGs and MGEs were available, including four giant tortoises, twenty four sea lions, and nine land iguanas, using an ASV table rarified at 20,000 sequences/sample and summarized at the level of phyla. Only phyla found in 20% of samples (≥7) were included in the analysis. Proteobacteria positively correlated with MGE (rho=0.66, adj. p=0.00051) and ARG sum abundance/16S rRNA for both ResFinder (rho=0.62, adj. p=0.0011) and ARG-OAP.1 (rho=0.61, adj. p=0.0012) annotation approaches (**Table 5.2**). The remaining phyla associated with significant correlations all showed negative relationships with MGE and ARG sum abundance/16S rRNA, including Synergistota with ARG-OAP.1 and MGE sums and Firmicutes with ARG sums annotated with ResFinder. Verrucomicrobia, Thermoplasmatota, and Spirochaetota all negative correlated with MGE sum abundance/16S rRNA.

Table 5.2: *Spearman rank correlation between phyla, resistome, and mobilome sum abundance/16S rRNA.*

Sum abundance/16S rRNA	Phylum	rho	Adjusted p-value
MGE	Proteobacteria	0.66	0.00051
ResFinder	Proteobacteria	0.62	0.0011
ARG-OAP.1	Proteobacteria	0.61	0.0012
ARG-OAP.1	Synergistota	-0.93	0.026
MGE	Synergistota	-0.93	0.026
ResFinder	Firmicutes	-0.46	0.031
MGE	Verrucomicrobia	-0.74	0.031
MGE	Thermoplasmatota	-0.80	0.036
MGE	Spirochaetota	-0.65	0.048

Repeating this procedure on a rarified ASV table summarized at the level of bacterial family revealed *Enterobacteriaceae* to strongly correlate with MGE (rho=0.85, adj. p=8.49E-07) and ARG sum abundances/16S rRNA for both ARG-OAP.1 (rho=0.76, adj. p=0.00013) and ResFinder (rho=0.83, adj. p=3.48E-06) annotation approaches (**Table 5.3**). The apparent negative relationship between *Peptostreptococcaceae* and ResFinder ARG sum abundance/16S rRNA (rho=-0.67, adj. p=0.00026) amounted to the only significant negative correlation recorded. The family *Bacillaceae* also correlated positively with ARG-OAP.1 sum abundance/16S rRNA (rho=0.88, adj. p=0.014), but no additional statistically significant correlations were noted.

Table 5.3: *Spearman rank correlation between families, resistome, and mobilome sum abundance/16S rRNA.*

Sum abundance/16S rRNA	Family	rho	Adjusted p-value
MGE	*Enterobacteriaceae*	0.85	8.49E-07
ResFinder	*Enterobacteriaceae*	0.83	3.48E-06
ARG-OAP.1	*Enterobacteriaceae*	0.76	0.00013
ResFinder	*Peptostreptococcaceae*	-0.67	0.00026
ARG-OAP.1	*Bacillaceae*	0.88	0.014

Given that *Enterobacteriaceae* counts, MGE sums, and ARG sums met the assumptions of the Shapiro-Wilkes test following log transformation for this subset of 37 samples (data not shown), we subsequently explored the relationships with this bacterial family using linear models. As shown in **Figure 5.8**, the positive relationship between *Enterobacteriaceae*, MGE sums, and ARG sums remained significant when using parametric statistics, with adjusted R-squared values of 0.66, 0.60, and 0.75 for ARG-OAP.1, ResFinder, and MGE sum abundance/16S rRNA, respectively (all p<4.42e-07). This relationship was most apparent for land iguanas and sea lions, while both *Enterobacteriaceae* counts and MGE/ARG sums were

generally lower for giant tortoises. Performing this analysis considering only land iguanas revealed a clear difference by sampling location (**Figure S5.4**), with *Enterobacteriaceae* counts and ARG sums high among individuals from North Seymour and Santa Fe compared to Plaza Sur. Linear models again yielded significant positive correlations, with R-square values of 0.56, 0.89, and 0.78 for ARG-OAP.1 (p=0.013), ResFinder (p=9.49e-05), and MGE (p=0.00092) sum abundance/16S rRNA, respectively.

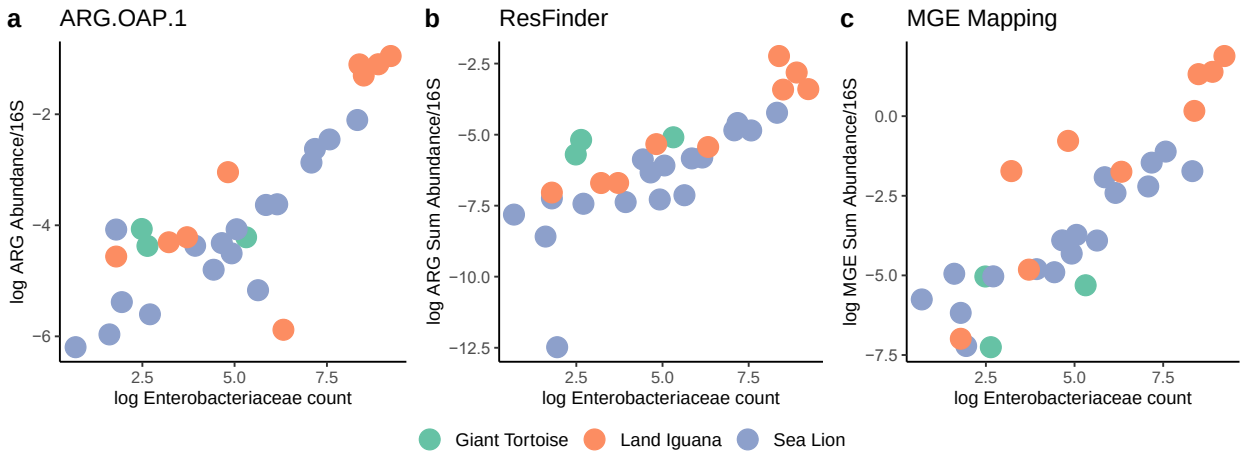

Figure 5.8: *Linear correlation between log-transformed Enterobacteriaceae counts and log-transformed MGE or ARG sum abundance/16S.* Enterobacteriaceae counts were taken from an ASV table rarified at 20,000 sequences/sample and summarized at the level of family (L5). **a)** Correlation with ARG-OAP.1 sum abundance. **b)** Correlation with ResFinder sum abundance**. c)** Correlation with MGE sum abundance.

Comparison of 16S rRNA and metagenomic taxonomic assignment

A second aim of this chapter centered on obtaining taxonomic information from the metagenomes produced in Chapter 3. To this end, we first compared the taxonomic assignments produced from 16S rRNA amplicon sequencing (i.e. the dataset generated by QIIME 2) with the taxonomic assignments produced using Metaxa2. Among the 37 paired samples between these two data sets, one land iguana from Plaza Sur was excluded due to SSU rRNA counts < 5,000 sequences. Overall, the approaches agreed well at high taxonomic levels. When mean relative

abundance was calculated by wildlife species and phyla were filtered to include only those with relative abundance ≥ 1% in at least one of the wildlife species, taxonomic assignment with Metaxa2 pointed to the same nine phyla identified using the same procedure with the 16S rRNA dataset. However, two additional phyla including Tenericutes and unclassified Bacteria accounted for ≥ 1% mean relative abundance in at least one wildlife species (**Figure 5.9**). Both taxonomic assignment methodologies pointed to Firmicutes as the most abundant phyla across the three wildlife species (**Table 5.4**), though the proportions differed considerably with the 16S rRNA data yielding overall higher relative abundances of this phyla. For example, using Metaxa2 mean relative abundance of Firmicutes was 38.7%, 54.7%, and 57.7% for giant tortoises, sea lions, and land iguanas, respectively, compared to 78.6%, 72.3% and 62.9% using QIIME 2. The two approaches generally reached consensus regarding the most abundant phyla in each wildlife species, though some departures were noted. Both approaches pointed to Firmicutes, Bacteroidetes, Fusobacteria, Proteobacteria, and Actinobacteria as the top five most abundant phyla in sea lions, though their relative abundances differed. For giant tortoises and land iguanas, the approaches agreed in identifying four of the same phyla as the top five most abundant. For sea lions, the two approaches continued to reach consensus through lower taxonomic levels, with *Peptostreptococcaceae* and *Lachnospiraceae* as the first and second-most abundant families in each data set and *Fusobacteriaceae* and *Bacteroidaceae* also among the top five families. At the level of genus, Metaxa2 data pointed to an unknown *Peptostreptococcaceae* genus as most abundant, while *Peptoclostridium* (family *Peptostreptococcaceae*) occupied the position of most abundant genus using QIIME 2. Among land iguanas, both Metaxa2 and QIIME 2 identified *Enterobacteriaceae* as the most abundant family and shared *Clostridiaceae* and *Lachnospiraceae* among the top five. Top genera shared among the top five include *Escherichia-*

Shigella and unclassified *Enterobacteriaceae.* At lower taxonomic levels, little consensus was noted for giant tortoises. The only overlap in the top five most abundant families was *Clostridiales,* and no genera were shared among the top five. This may in part be explained by poor resolution with Metaxa2 at lower taxonomic levels, as four of the five most abundant genera in giant tortoises were unclassified genera of higher taxonomic levels.

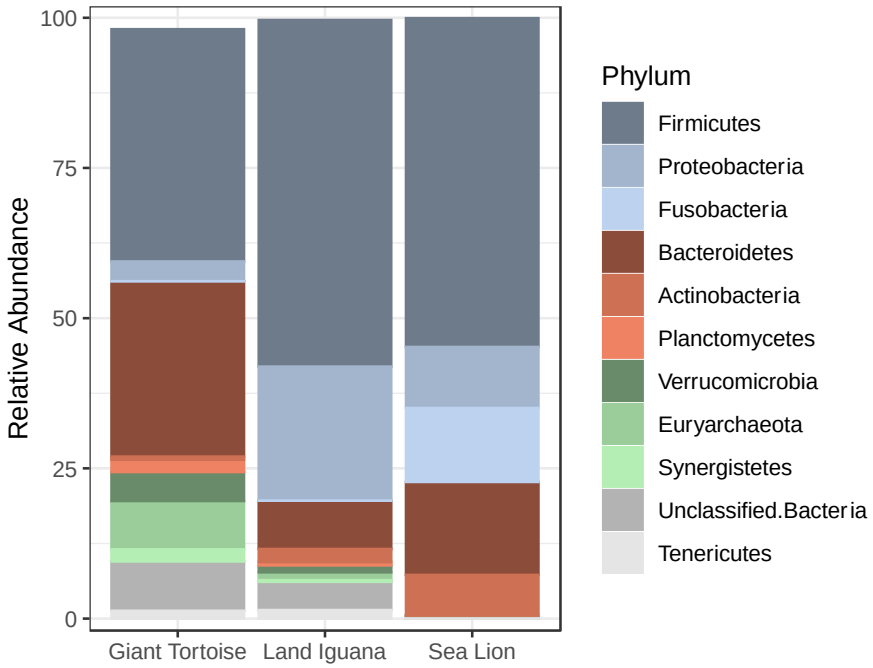

Figure 5.9: *Most abundant phyla in three wildlife species using taxonomic assignments from Metaxa2.* Phyla were filtered to retain only those accounting for mean relative abundance ≥1% in at least one wildlife species.

Finally, we assessed the extent of agreement in sample distances using each taxonomic assignment approach. Implementation of the Mantel test on distance matrices calculated from the Horn-Morisita similarity index yielded a test statistic r equal to 0.44 (sig=0.001), indicating moderate, statistically significant agreement between the two approaches.

Table 5.4: *Mean relative abundance of phyla, families, and genera by wildlife species using Metaxa2 taxonomic assignments.* Data are from samples with matched 16S rRNA amplicon sequencing data. One land iguana sample was excluded due to SSU rRNA counts <5,000 sequences/sample. Total n=36.

Species	Phylum	Rel. abund. (%)	*Family*	Rel. abund. (%)	*Genus*	Rel. abund. (%)
Giant Tortoise	Firmicutes	38.7	*Unclassified Bacteroidetes*	10.9	*Unclassified Bacteroidetes*	10.9
	Bacteroidetes	29.2	*Ruminococcaceae*	8.7	*Unclassified Bacteria*	8.2
	Unclassified	8.2	*Unclassified bacteria*	8.2	*Methanocorpusculum*	7.6
	Euryarchaeota	8.1	*Methano-corpusculaceae*	7.6	*Unclassified Clostridiales*	6.6
	Verrucomicrobia	4.8	*Unclassified Clostridiales*	6.6	*Unclassified Bacteroidales*	5.5
Sea Lion	Firmicutes	54.7	*Peptostreptococcaceae*	21.5	*Peptostreptococcaceae, Incertae Sedis*	15.9
	Bacteroidetes	15.5	*Lachnospiraceae*	15.3	*Bacteroides*	14.5
	Fusobacteria	12.3	*Bacteroidaceae*	14.6	*Fusobacterium*	10.1
	Proteobacteria	10.2	*Fusobacteriaceae*	12.2	*Unclassified Lachnospiraceae*	7.5
	Actinobacteria	6.8	*Coriobacteriaceae*	6.7	*Unclassified Clostridiales*	6.5
Land Iguana	Firmicutes	57.7	*Enterobacteriaceae*	21.2	*Unclassified Enterobacteriaceae*	12.2
	Proteobacteria	22.3	*Ruminococcaceae*	11.6	*Unclassified Ruminococcaceae*	7.1
	Bacteroidetes	8.0	*Lachnospiraceae*	9.7	*Unclassified Clostridiales*	6.5
	Unclassified	4.7	*Clostridiaceae*	8.1	*Clostridium*	5.7
	Actinobacteria	2.2	*Unclassified Clostridiales*	6.5	*Escherichia-Shigella*	5.2

Taxonomic classification of SSU rRNA in metagenomes

Taxonomic classification of SSU rRNA using Metaxa2 produced combined bacterial and archaeal counts ranging from 173 in a sea turtle to 553,329 in a sea lion (median 27,451). To reduce bias from samples with low SSU rRNA counts, samples with fewer than 5,000 counts were filtered from the observation table. This included two marine water samples, six marine iguanas, three sea turtles, two land iguanas, and one red-footed booby, with 76 of 90 samples retained in the analysis. The filtered table was rarified at an even sampling depth of 5,000 SSU rRNA counts/sample for diversity calculations.

Alpha and beta diversity

Alpha diversity was calculated using the Shannon and Simpson diversity indices when samples were categorized broadly as human, wastewater, water, or wildlife (**Figure 5.10**). For both indices, diversity was highest in water followed by wastewater, wildlife, and humans. All pairwise comparisons were significant using the Shannon diversity index (ANOVA, Tukey's post hoc test, adjusted $p < 0.05$) with the exception of the difference between water and wastewater. Considering the Simpson diversity index, wastewater, water, and wildlife all exhibited significantly higher diversity compared to human samples (ANOVA, Tukey's post hoc test, adjusted $p < 0.05$), but no additional pairwise comparisons were significant. Sample type exerted a significant effect over bacterial community composition, with samples clustering clearly according to type (**Figure 5.11**). Calculation of the Bray-Curtis dissimilarity index revealed sample type to explain 30.5% of sample distances (ADONIS with 9,999 permutations, p=1e-04). When ellipses were drawn at the 90% confidence level, human samples generally clustered distinctly from the other sample categories, while considerable overlap was observed between wastewater and water samples.

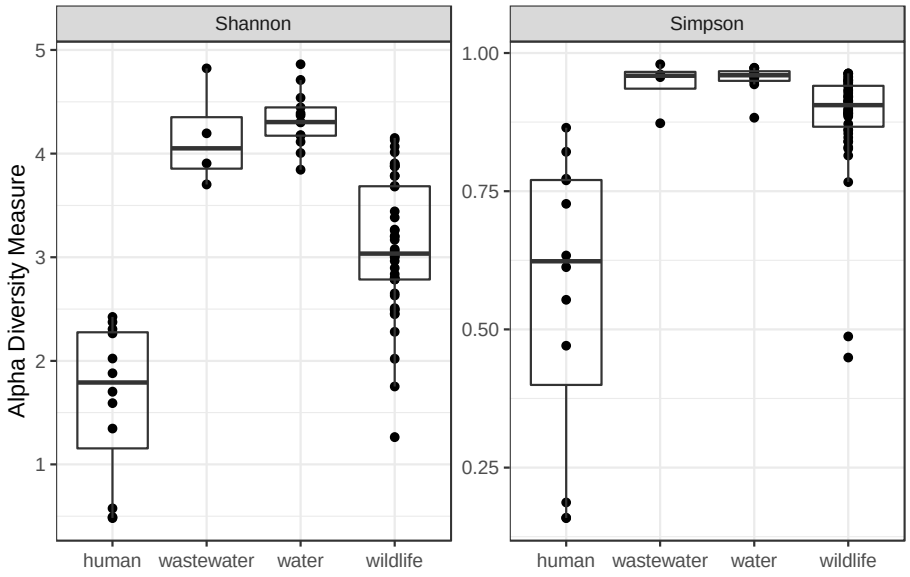

Figure 5.10: *Alpha diversity of metagenomes according to taxonomic assignments produced by Metaxa2.*

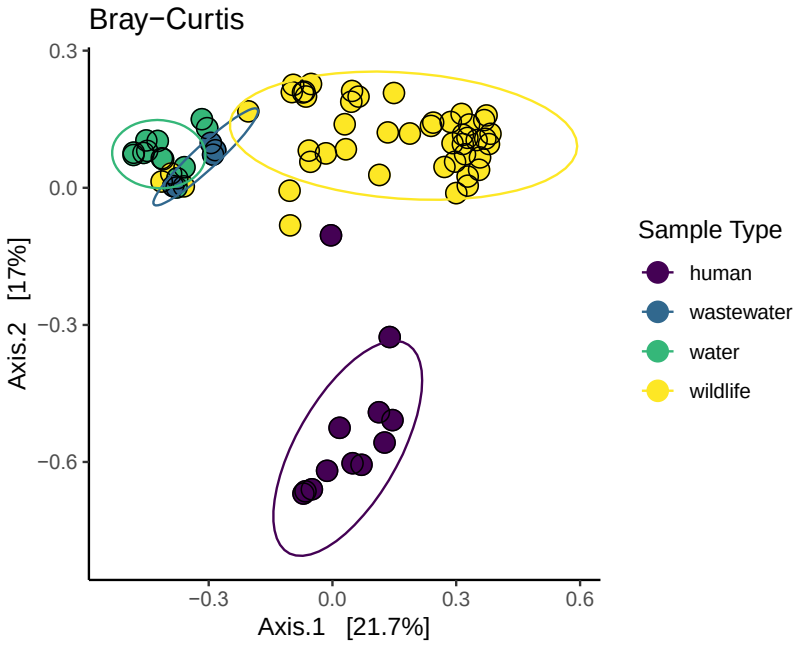

Figure 5.11: *Sample type drives bacterial community composition of metagenomes according to taxonomic assignments produced by Metaxa2.*

Relative abundance of bacterial taxa in wildlife and humans

The relative abundance of bacterial taxa differed by host species (**Figure 5.12, Table 5.5**). Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria accounted for four of the top five most abundant phyla in human and all wildlife samples except for giant tortoises (lacking Proteobacteria) and red footed boobies (lacking Bacteroidetes). All wildlife with the exception of sea turtles and red footed boobies were dominated by Firmicutes. Samples from children under age two were uniquely distinguished by predominance of SSU rRNA sequences assigned to Actinobacteria. Humans, sea lions, land iguanas, and marine iguanas shared *Lachnospiraceae* as one of the top five most abundant families. *Peptostreptococcaceae* was common to both sea lions and marine iguanas, while *Ruminococcaceae* was among the five most abundant families for giant tortoises and land iguanas. *Fusobacteriaceae* was an abundant family shared by sea lions and red footed boobies, while *Enterobacteriaceae* made the top five for land iguanas and humans.

Relative abundance of bacterial taxa in water samples

All water samples (wastewater, wastewater-impacted marine water, background marine water, and freshwater) were dominated by the phylum Proteobacteria (47.2 to 70.8%), with the highest relative abundance observed in freshwater (**Figure 5.12, Table 5.5**). Bacteroidetes accounted for the second most abundant phylum in wastewater and marine water and the fourth most abundant phyla in freshwater. Firmicutes were among the top five most abundant phyla for all water types, though in position four or five, while Cyanobacteria were uniquely abundant in marine water samples. At the family level, *Campylobacteraceae* constituted the most abundant taxa in both wastewater (9.6%) and wastewater-impacted marine environments (12.3%), but was

absent from the top five families of background marine and freshwater samples, which were instead dominated by *Flavobacteriaceae* (12.2%) and *Comamonadaceae* (16%), respectively.

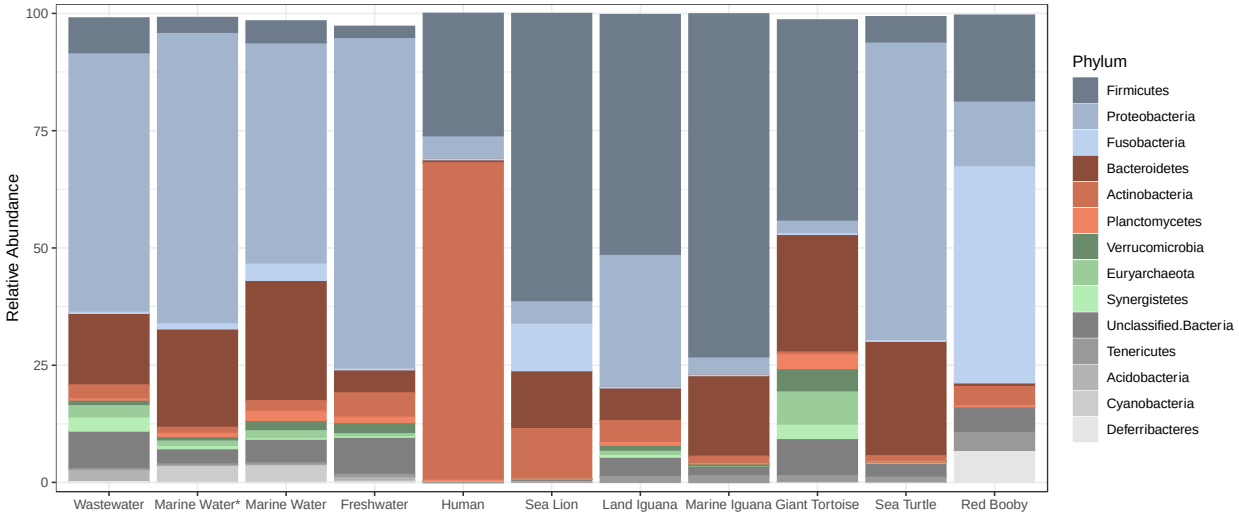

Figure 5.12: *Most abundant phyla using taxonomic assignments from Metaxa2.* Phyla were filtered to retain only those accounting for mean relative abundance ≥1% in at least one sample type.

Table 5.5: Mean relative abundance of phyla, families, and genera by wildlife species using Metaxa2 taxonomic assignments. Samples with fewer than 5,000 sequences were excluded. Total n=76.

Species	Phylum	Rel. abund (%)	*Family*	Rel. abund (%)	*Genus*	Rel. abund (%)
Wastewater	Proteobacteria	55.5	*Campylobacteraceae*	9.6	*Arcobacter*	9.4
	Bacteroidetes	15.3	*Unclassified Bacteria*	8.1	*Unclassified Comamonadaceae*	4.5
	Unclassified Bacteria	8.1	*Comamonadaceae*	6.3	*Unclassified Betaproteobacteria*	3.2
	Firmicutes	7.3	*Rhodocyclaceae*	3.9	*Bacteroides*	2.9
	Actinobacteria	3.0	*Flavobacteriaceae*	3.5	*Unclassified Bacteroidetes*	2.4
Marine Water (Impacted)	Proteobacteria	62.3	*Campylobacteraceae*	12.3	*Arcobacter*	12.2
	Bacteroidetes	21.0	*Flavobacteriaceae*	9.0	*Unclassified Flavobacteriaceae*	7.1
	Cyanobacteria	3.6	*Rhodobacteraceae*	7.9	*Unclassified Rhodobacteraceae*	6.9
	Unclassified Bacteria	3.3	*Unclassified Gammaproteobacteria*	7.2	*Unclassified Bacteroidetes*	3.4
	Firmicutes	3.1	*Alteromonadaceae*	4.5	*Unclassified Alteromonadaceae*	2.4
Marine Water (Background)	Proteobacteria	47.2	*Flavobacteriaceae*	12.2	*Unclassified Flavobacteriaceae*	9.2
	Bacteroidetes	25.7	*Rhodobacteraceae*	8.8	*Unclassified Rhodobacteraceae*	7.5
	Unclassified Bacteria	5.1	*Unclassified Alphaproteobacteria*	5.5	*Unclassified Alphaproteobacteria*	5.5
	Firmicutes	4.6	*Unclassified Gammaproteobacteria*	5.5	*Unclassified Bacteroidetes*	3.8
	Cyanobacteria	3.4	*Unclassified Bacteria*	5.1	*Fusobacterium*	2.5

Freshwater	Proteobacteria	70.8	*Comamonadaceae*	16.0	*Unclassified Comamonadaceae*	13.4
	Unclassified Bacteria	8.0	*Methylococcaceae*	10.4	*Methylomonas*	6.6
	Actinobacteria	5.2	*Unclassified Bacteria*	8.0	*Polynucleobacter*	4.9
	Bacteroidetes	5.0	*Burkholderiaceae*	5.5	*Unclassified Betaproteobacteria*	4.1
	Firmicutes	2.3	*Methylophilaceae*	4.7	*Unclassified Methylococcaceae*	2.6
Human	Actinobacteria	67.7	*Bifidobacteriaceae*	58.3	*Bifidobacterium*	57.9
	Firmicutes	26.0	*Lachnospiraceae*	10.5	*Collinsella*	5.7
	Proteobacteria	5.1	*Coriobacteriaceae*	8.3	*Unclassified Lachnospiraceae*	4.2
	Bacteroidetes	0.7	*Enterobacteriaceae*	4.9	*Enterococcus*	3.7
	Unclassified Bacteria	0.4	*Enterococcaceae*	4.0	*Streptococcus*	3.5
Sea Lion	Firmicutes	61.1	*Lachnospiraceae*	23.0	*Unknown Peptostreptococcaceae*	14.4
	Bacteroidetes	12.4	*Peptostreptococcaceae*	20.4	*Unclassified Lachnospiraceae*	11.5
	Actinobacteria	10.7	*Bacteroidaceae*	11.3	*Bacteroides*	11.2
	Fusobacteria	9.8	*Coriobacteriaceae*	10.6	*Fusobacterium*	8.4
	Proteobacteria	5.1	*Fusobacteriaceae*	9.8	*Collinsella*	8.4
Land Iguana	Firmicutes	51.0	*Enterobacteriaceae*	27.4	*Enterobacteriaceae*	27.4
	Proteobacteria	28.5	*Ruminococcaceae*	10.1	*Ruminococcaceae*	10.1
	Bacteroidetes	7.0	*Lachnospiraceae*	8.4	*Lachnospiraceae*	8.4
	Actinobacteria	4.7	*Clostridiaceae*	7.0	*Clostridiaceae*	7.0
	Unclassified Bacteria	4.2	*Unclassified Clostridiales*	5.7	*Unclassified Clostridiales*	5.7
Marine Iguana	Firmicutes	73.1	*Lachnospiraceae*	32.1	*Clostridium*	17.8

	Bacteroidetes	17.2	*Clostridiaceae*	19.5	*Unclassified Lachnospiraceae*	17.2
	Proteobacteria	4.0	*Bacteroidaceae*	14.0	*Bacteroides*	13.9
	Unclassified Bacteria	2.1	*Peptostreptococcaceae*	9.3	*Unclassified Lachnospiraceae*	9.3
	Actinobacteria	1.6	*Unclassified Clostridiales*	3.6	*Unclassified Peptostreptococcaceae*	8.6
Giant Tortoise	Firmicutes	42.6	*Ruminococcaceae*	9.6	*Unclassified Bacteroidetes*	9.5
	Bacteroidetes	25.1	*Unclassified Bacteroidetes*	9.5	*Unclassified Clostridiales*	7.1
	Unclassified Bacteria	8.0	*Unclassified Bacteria*	8.0	*Methanocorpusculum*	7.0
	Euryarchaeota	7.5	*Unclassified Clostridiales*	7.1	*Unclassified Ruminococcaceae*	6.1
	Verrucomicrobia	4.7	*Methanocorpusculaceae*	7.1	*Unclassified Bacteroidales*	4.7
Sea Turtle	Proteobacteria	63.8	*Neisseriaceae*	12.0	*Unclassified Neisseriaceae*	9.5
	Bacteroidetes	24.5	*Unclassified Gammaproteobacteria*	11.9	*Unclassified Bacteroidales.*	4.8
	Firmicutes	5.3	*Flavobacteriaceae*	7.3	*Unclassified Flavobacteriaceae*	4.5
	Unclassified Bacteria	3.2	*Unclassified Bacteroidales*	4.8	*Unclassified Bacteroidetes*	3.6
	Actinobacteria	1.2	*Pasteurellaceae*	4.5	*Unclassified Betaproteobacteria*	3.1
Red Footed Booby	Fusobacteria	46.0	*Fusobacteriaceae*	45.1	*Fusobacteriaceae*	45.1
	Firmicutes	18.2	*Deferribacteraceae*	6.8	*Deferribacteraceae*	6.8
	Proteobacteria	14.1	*Vibrionaceae*	6.6	*Vibrionaceae*	6.6
	Deferribacteres	6.8	*Unclassified Bacteria*	5.5	*Unclassified Bacteria*	5.5
	Unclassified Bacteria	5.5	*Unclassified Clostridiales*	5.3	*Unclassified Clostridiales*	5.3

Correlation between bacterial taxa, ARGs, and MGEs

Given the relationships between ARGs, MGEs, and taxonomic assignments from 16S rRNA amplicon sequencing, we asked if similar relationships could be detected when taxonomic classifications were instead inferred by Metaxa2. Using an observation table rarified at 5,000 SSU rRNA counts/sample and considering phyla found in at least 20% (≥15 of 76) samples, several significant Spearman rank correlations were recorded. Acidobacteria was found to positively correlate with ARG-OAP.1 sum abundance/16S rRNA (rho=0.61, adj. p=2.91E-02) while Proteobacteria and MGE sum abundance/16S rRNA were also positively correlated (rho=0.34, adj. p=4.26E-02). Conversely, negative correlations were noted between Chloroflexi and ResFinder ARG sum abundance/16S rRNA (rho= -0.70, adj. p=1.62E-02), Euryarchaeota and MGE sum abundance (rho= -0.50, adj. p=4.26E-02), and finally Lentisphaerae and ARG-OAP.1 sum abundance. When performing this analysis at the family level (**Table S5.2**), we noticed that many positive correlations involved taxa that were elevated in wastewater compared to other sample types, such as *Comamonadaceae* and *Rhodocyclaceae*. Accordingly, to reduce correlations between taxa and ARG sums driven solely by wastewater, we repeated this analysis on strictly environmental samples, including marine water, freshwater, and wildlife (n=60 samples). At the level of phylum, Proteobacteria positively correlated with MGE and ARG-OAP.1 sum abundance/16S rRNA, yielding respective rho values equal to 0.45 and 0.52 with p≤1.10E-02 in both cases, while Euryarchaeota negatively correlated with MGE sum abundance/16S rRNA (rho= -0.59, p=1.37E-02). Among bacterial families, *Enterobacteriaceae* positively correlated with MGE (rho=0.75, adj. p=8.20E-08) and ARG sum abundance/16S rRNA considering both ARG-OAP.1 (rho=0.66, adj. p=3.56E-05) and ResFinder ARG annotations (rho=0.65, adj. p=9.06E-05). No additional families were associated with significant

correlations after FDR p-value correction. At the level of genus, unclassified *Enterobacteriaceae* and *Escherichia-Shigella* positively correlated with MGE and ARG-OAP.1 sum abundance/16S rRNA (**Table 5.6**), while *Enterobacteriaceae* but not *Escherichia-Shigella* correlated significantly with ResFinder ARG sum abundance/16S rRNA. Four additional genera showed significant positive correlations with ARG-OAP.1 sum abundance, including *Comamonas* and unclassified genera belonging to *Comamonadaceae*, *Mycoplasmataceae*, and *Campylobacterales*, while only one additional genus, an unclassified Bacilli, significantly correlated with ResFinder ARG sum abundance/16S rRNA.

Table 5.6: *Spearman rank correlation between genera, resistome, and mobilome sum abundance/16S rRNA among water and wildlife samples using taxonomic assignments from Metaxa2.*

Sum abundance/16S rRNA	Genus	rho	Adjusted p-value
MGE	*Unclassified Enterobacteriaceae*	0.82	3.87E-10
MGE	*Escherichia-Shigella*	0.74	1.24E-05
MGE	*Unclassified Spirochaetaceae*	-0.68	4.11E-02
ARG-OAP.1	*Comamonas*	0.91	2.49E-03
ARG-OAP.1	*Unclassified Comamonadaceae*	0.78	4.57E-03
ARG-OAP.1	*Unclassified Mycoplasmataceae*	0.72	1.02E-02
ARG-OAP.1	*Unclassified Enterobacteriaceae*	0.70	1.08E-05
ARG-OAP.1	*Escherichia-Shigella*	0.62	2.49E-03
ARG-OAP.1	*Unclassified.Campylobacterales.*	0.61	4.71E-02
ARG-OAP.1	*Gracilibacter*	-0.81	1.68E-02
ResFinder	*Unclassified Enterobacteriaceae*	0.63	6.04E-04
ResFinder	*Unclassified Bacilli*	0.52	4.11E-02
ResFinder	*Peptostreptococcaceae (unknown)*	-0.50	1.02E-02

Relationship between Enterobacteriaceae, ARGs, and MGEs varies by sample type

Kendall rank correlations performed between *Enterobacteriaceae* counts, ARG, and MGE sum abundances revealed distinct relationships by sample type. While *Enterobacteriaceae*

counts correlated well with MGE sums and ARG sums for both wildlife and water samples (**Figure 5.13**, **Table 5.7**), relationships were less apparent among human samples. Specifically, *Enterobacteriaceae* counts seemed unrelated to ARG-OAP.1 and ResFinder ARG sum abundance/16S rRNA with tau equal to 0.1 (p=63) and -0.14 (p=0.58), respectively. In contrast, a significant positive correlation was noted between *Enterobacteriaceae* counts and MGE sums in human samples.

Table 5.7: *Kendall tau correlations between Enterobacteriaceae counts, MGE, and ARG sum abundance by sample type.*

Data subset	Sum abundance/16S rRNA	tau	Adjusted p-value
Full dataset	ARG-OAP.1	0.43	1.77E-07
Full dataset	MGE	0.54	6.48E-11
Full dataset	ResFinder	0.31	0.000146
wildlife	ARG-OAP.1	0.58	5.48E-08
wildlife	MGE	0.69	9.24E-11
wildlife	ResFinder	0.43	6.02E-05
water	ARG-OAP.1	0.54	0.016
water	MGE	0.54	0.016
water	ResFinder	0.27	0.24
human	ARG-OAP.1	0.1	0.63
human	MGE	0.6	0.013
human	ResFinder	-0.14	0.58

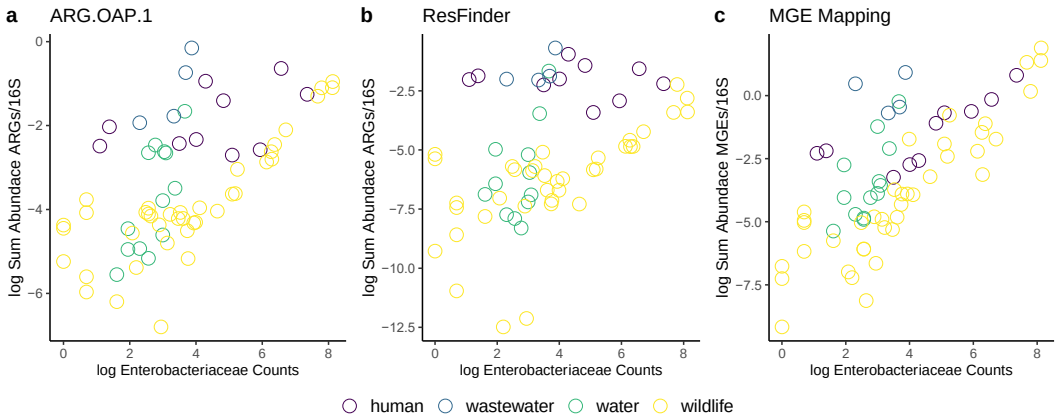

Figure 5.13: *The relationship between Enterobacteriaceae, ARGs, and MGEs varies by sample type.* For wildlife and water samples, Kendall rank correlations reveal a positive correlation between *Enterobacteriaceae* counts and **a)** ARG-OAP.1 sum abundance, **b)** ResFinder sum abundance, and **c)** MGE sum abundance. Human samples show a positive relationship between *Enterobacteriaceae* and MGE sums, but no apparent relationship with ARG sums.

Discussion

The present work aimed to add phylogenetic context to the antibiotic resistance (Chapter 3) and mobile genetic element (Chapter 4) data presented for 90 wildlife, wastewater, water, and human metagenomes. To this end, we characterized the gut microbial communities of 50 wildlife animals from the Galapagos islands using 16S rRNA amplicon sequencing and compared results to taxonomic classifications assigned to the larger metagenomic dataset. To our knowledge, this represents the first high-throughput survey of the Galapagos sea lion (*Zalophus wollebaeki*) gut microbiome; second for Galapagos giant tortoises (*Chelonoidis nigra*), though the first included data from only four individuals (Hong et al., 2011); and third for Galapagos land iguanas (*Conolophus subcristatus*). At high taxonomic levels, we found taxonomic assignments produced by targeted 16S rRNA amplicon sequencing to agree well with taxonomic assignments of shotgun metagenomic sequences, and both methods identified the same bacterial phyla and families in correlations with ARG and MGE sums.

Microbial community composition differs by wildlife host species

Among 50 Galapagos wildlife gut microbiomes from 16 giant tortoises, 24 sea lions, and 10 land iguanas, five phyla accounted for 95% of the observed ASVs. Firmicutes (62.9 to 78.6%) and Bacteroidetes (4.4 to 7.4%) were among the top five most abundant phyla in all three species, consistent with prior observations regarding the dominance of these two phyla in vertebrate gut microbiomes (Ley et al., 2008). Varying degrees of overlap were observed in the remaining top phyla between wildlife species, with giant tortoises characterized by Planctomycetes, Verrucomicrobia, and Halobacterota. Verrucomicrobia was also among the most abundant phyla in land iguanas, which otherwise shared Proteobacteria and Actinobacteria with sea lions. Sea lions, on the other hand, were uniquely characterized by the Fusobacteria phylum.

Relative abundance of bacterial taxa in giant tortoises

These observations generally align well with prior reports in the same or similar wildlife species. Microbial community profiling of bolson tortoises (Peña et al., 2019) and Galapagos giant tortoises (Hong et al., 2011) identified Firmicutes as the most dominant phyla, with relative abundances reported at 80-93% and 81%, respectively. On the other hand, studies in the Seychelles giant tortoise (Sandri et al., 2020) and gopher tortoise (Yuan et al., 2015) found more equal proportions of Firmicutes and Bacteroidetes, with respective relative abundances of 34% and 38% in the Seychelles giant tortoise and 36% and 36.5% in the gopher tortoise. At lower taxonomic levels, some overlap can be drawn between tortoise species. For example, *Rikenellaceae* was among the most abundant families in Seychelles giant tortoises (Sandri et al., 2020), a pattern shared by the giant tortoises in the present study, while bolson tortoises shared the families *Christensenellaceae* and *Lachnospiraceae* (Peña et al., 2019). Surprisingly, our data

shared only partial overlap with the study of four giant tortoises housed at La Galapaguera on San Cristobal (Hong et al., 2011), in which Firmicutes (81%), unclassified Bacteria (10.4%), Bacteroidetes (4.4%), Proteobacteria (2%), and Actinobacteria (0.8%) accounted for the most abundant phyla. *Prevotellaceae,* instead of *Clostridiaceae,* was found to be the most abundant family, though *Clostridiaceae* was more abundant in giant tortoises compared to other reptiles in the study. Despite the differences at the phylum level, both Hong et al. (2011) and our present data point to *Sarcina* and *Clostridium* as highly abundant genera in Galapagos giant tortoises. In total, the differences between the work by Hong et al. (2011) and data reported herein are likely attributable to both improvement in high-throughput sequencing technologies and expansion of 16S rRNA databases in the last ten years: indeed, in a subsequent metagenomic study in marine and land iguanas, Hong et al. (2015) reported almost complete resolution of unclassified Bacteria (0.03-0.07%) compared to the 2011 study with unclassified Bacteria accounting for 14.5-26.6% of observations.

Relative abundance of bacterial taxa in sea lions

The gut microbiomes of the twenty-four Galapagos sea lions in the present study shared similarities with those of captive Californian sea lions (Bik et al., 2016) and Australian sea lions (Delport et al., 2016). Like Galapagos sea lions, California sea lions shared Bacteroidetes (41.3%), Firmicutes (28.9%), Fusobacteria (25.4%) and Proteobacteria (4.0%) as the most dominant phyla, though Firmicutes instead of Bacteroidetes constituted the most dominant phyla in our study. Australian sea lions shared Firmicutes as the most abundant phyla in most samples (7.1 to 99.5%), with either Proteobacteria or Bacteroidetes as the most abundant phyla where Firmicutes came in second. Fusobacteria and Actinobacteria were additionally among the phyla accounting for mean relative abundance of at least 1%. In agreement with our data,

Peptostreptococcaceae, *Lachnospiraceae*, and *Clostridiaceae* were among the most predominant families, though Australian sea lions were also characterized by members of the *Ruminococcaceae* family. Gut microbiome studies in other pinnipeds such as the hooded seal (Acquarone et al., 2020) and pacific harbor seal (Pacheco-Sandoval et al., 2019) corroborated Firmicutes as the most abundant bacterial phyla in these species. Interestingly, Fusobacteria were also found to be dominant members of the pacific harbor seal gut microbiota, accounting for 26% overall abundance (Pacheco-Sandoval et al., 2019). Accordingly, *Fusobacteriaceae* and *Fusobacterium* represented the most abundant family and genera, respectively, in pacific harbor seals. These taxa were also among the top five families and genera in Galapagos sea lions.

Relative abundance of bacterial taxa in land iguanas

To date, only two studies have profiled the microbial communities of the Galapagos land iguana, including a first effort by Hong and colleagues (2011) sampling individuals from Fernandina, Plaza Sur, and Santa Fe, and a second effort on only Plaza Sur by the same group several years later (Hong et al., 2015) The first study relied on 454 pyrosequencing while the second employed shotgun metagenomics. In the first study of land iguanas from three islands, Firmicutes (63.9%), unclassified Bacteria (26.6%), Bacteroidetes (4.2%), Proteobacteria (1.4%), Actinobacteria (1.3%) accounted for the most abundant phyla, a finding in close agreement with the present study with the exception of the large proportion of unclassified Bacteria. The follow-up study using shotgun metagenomic sequencing on seven land iguanas from Plaza Sur (Hong et al., 2015) resolved many of the unclassified Bacteria, instead presenting an abundance pattern of Firmicutes (60.7-63.9%) followed by Proteobacteria (12.9-15.1%), and Bacteroidetes (7.3%). Our data agreed in the order of Firmicutes and Proteobacteria, but placed Actinobacteria above Bacteroidetes. The first study (Hong et al., 2011) detected subtle variations in community

composition by island, with individuals on Fernandina exhibiting lower microbial richness compared to those on Plaza Sur and Santa Fe. While measures of alpha diversity were not significantly different by island in the present study, sampling location did significantly influence community composition when using the Bray-Curtis dissimilarity index, but not weighted Unifrac distances.

Differential abundance of taxa between wildlife species

Taxa found to be differentially abundant between wildlife species were generally consistent with observations made regarding the top phyla, families, and genera for each species. For example, Planctomycetes and Verrucomicrobia were differentially abundant in giant tortoises and land iguanas compared to sea lions, which coincides with the placement of these phyla among the top five most abundant in the reptile species. At the family level, *Christensenellaceae* was more prominent in giant tortoises and land iguanas compared to sea lions, while genera *Oscillospirales* UCG-010, the Bacilli family RF39, and *Pirellulaceae* were also differentially abundant in the reptilian species. This observation adds to the growing list of animals with *Christensenellaceae* as a common gut microbial species (Waters and Ley, 2019). In contrast, sea lions were unique characterized by the phylum Fusobacteria and the families *Fusobacteriaceae* and *Peptostreptococcaceae,* trends that could be predicted through their placement in the top five most abundant taxa for sea lions. Overall, the observation that more taxa were differentially abundant in the reptilian species seems to agree with the lower diversity recorded among sea lions. Interestingly, a study of marine mammals also found sea lion gut microbiomes to be associated with lower diversity compared to dolphins (Bik et al., 2016).

Agreement between taxonomic assignments from QIIME 2 and Metaxa2

While 16S rRNA amplicon sequencing has become the preferred method for microbial community profiling due to the abundance of bioinformatics tools for analysis (Bolyen et al., 2019) and lower cost compared to shotgun metagenomic sequencing (Rausch et al., 2019), several reports highlight advantages to inferring taxonomy from metagenomic libraries. First, 16S rRNA sequencing efforts are inherently subject to bias from primer choice (Fouhy et al., 2016) and differential amplification of the targeted region depending on GC-content (Laursen et al., 2017). Moreover, initial amplification is thought to introduce bias against low abundance taxa. In contrast, shotgun metagenomic sequencing largely avoids these biases though at a substantially higher cost. In the present study, we compared taxonomic assignments from 16S rRNA amplicon sequencing (i.e. the QIIME 2 dataset) with those produced using Metaxa2 on paired shotgun metagenomic libraries. We observed that the two approaches generally reached consensus at high taxonomic levels, with both pointing to Firmicutes as the most abundant phyla across all sample types though with different relative abundances. Moreover, QIIME 2 and Metaxa2 identified the same five most abundant phyla for sea lions. Among giant tortoises and land iguanas, the approaches agreed in identifying four of the same phyla among as the top five most abundant. Both approaches named *Peptostreptococcaceae* and *Enterobacteriaceae* as the most abundant families in sea lions and land iguanas, respectively. Less consensus was reached for giant tortoises, especially at lower taxonomic levels where Metaxa2 lacked resolution.

For other sample types, however, Metaxa2 produced lower-level taxonomic inferences consistent with other reports in the literature. For example, Metaxa2 identified *Bifidobacterium* as the most abundant genus in fecal samples from children under age two, accounting for 57.9% of total observations. *Bifidobacterium* are known to play an important role in infant gut

development (O'Neill et al., 2017; Oki et al., 2018) and by some reports are the most abundant genera in healthy infant gut microbiomes (Favier et al., 2002; Arboleya et al., 2016). While the overwhelming predominance of *Bifidobacterium* explains the low alpha diversity of humans compared to wastewater, water, and wildlife samples (**Figure 5.10**), this result is inconsistent with previous 16S rRNA amplicon sequencing performed on this sample cohort (Thompson et al., 2019) where Actinobacteria accounted for roughly 20% of observations. However, this study used GreenGenes for taxonomic inference, which has not been updated since 2013, while both the QIIME 2 and Metaxa2 analyses herein were performed against the SILVA database (Quast et al., 2013).

Similarly, taxonomic inferences made on water samples coincided with previous studies. All water samples, including wastewater, marine water, and freshwater were dominated by Proteobacteria, an observation that has been reported for freshwater aquaculture (Fang et al., 2019), drinking water (Vaz-Moreira et al., 2017), surface waters (Ibekwe et al., 2012) and municipal wastewater (Guo et al., 2019). Predominance of the family *Campylobacteraceae* and genus *Arcobacter* in wastewater agrees with prior reports of *Arcobacter* as the most abundant genus in municipal wastewater treatment plants (Guo et al., 2019; Kristensen et al., 2020). The observation that certain marine environments were also dominated by *Campylobacteraceae* at the family level and *Arcobacter* at the genus level corroborates our previous reports of wastewater impacts at these sites (Overby et al., 2015; Grube et al., 2020). Consistent with studies in the Sargasso Sea (Venter et al., 2004) and in the Galapagos specifically (Gifford et al., 2020), Cyanobacteria constitute a dominant phylum in marine environments, as we observed. In agreement with work by Giffort and colleagues (2020) in Galapagos marine waters, we also

identified *Flavobacteriaceae* and *Rhodobacteraceae* among the most abundant families in marine waters.

Taken together, we find the taxonomic inferences from Metaxa2 to be useful and reliable, though resolution was sometimes poor at the level of genus. While numerous tools exist to infer microbial community structure from shotgun metagenomic sequences, including MetaPhlAn (Segata et al., 2012), MG-RAST (Meyer et al., 2008), and MEGAN (Huson et al., 2007) among others, we aimed to take advantage of the taxonomic assignments produced by Metaxa2 in the course of tabulating SSU rRNA for normalization of ARGs and MGEs in Chapters 3 and 4.

Correlation between taxa, resistome and mobilome datasets

The overarching aim of the present chapter was to provide phylogenetic context to the resistome and mobilome data presented in Chapters 3 and 4. To this end, we sought to identify bacterial taxa that may be associated with ARG and MGE sums. When considering taxonomic assignments from QIIME 2 at the phylum level, Proteobacteria were found to positively correlate with MGE and ARG sum abundance/16S rRNA for both ARG-OAP.1 and ResFinder annotations, with rho≥0.6 and adj. $p<0.05$ for all comparisons. This association appeared to be driven by the family *Enterobacteriaceae* in particular, which yielded rho≥0.76 with MGE, ARG-OAP.1, and ResFinder ARG sum abundances, and linear models with R-square values ranging from 0.6 to 0.75 (all $p<4.42e-07$) on log-transformed predictor (*Enterobacteriaceae* counts) and response (MGE or ARG sum abundance/16S rRNA) variables. Notably, for the QIIME 2 dataset no other families yielded significant correlations. Additional possible associations between taxa, MGEs, and ARGs were recorded when considering Metaxa2 taxonomic assignments on 76 metagenomes, where at the phylum level Spearman rank correlations revealed a significant positive relationship between Acidobacteria and ARG-OAP.1 sum abundance/16S rRNA as well

as between Protobacteria and MGE sum abundance/16S rRNA. Performing this analysis at the family level identified many taxa positively associated with MGE and ARG sum abundance (**Table S5.2**). Based on the observation that the majority of associations appeared to involve taxa with relatively high abundance in wastewater, we instead explored which taxa might explain MGE and ARG sums in specifically in environmental samples. When considering the subset of data including marine water, freshwater, and all wildlife samples, Proteobacteria significantly correlated with both MGE and ARG-OAP.1 sum abundance/16S rRNA, though the rho correlation coefficients were lower (0.45 and 0.52, respectively) compared to the QIIME 2 data set. Mirroring correlations performed with the QIIME 2 data, *Enterobacteriaceae* was the only bacterial family to significantly correlate with MGE, ARG-OAP.1, and ResFinder ARG sum abundance, with rho correlation coefficients ranging from 0.65 to 0.75. Several other bacterial families, including *Comamonas* and unclassified members of *Comomonadaceae*, *Mycoplasmataceae,* and *Campylobacterales* positively correlated with ARG sum abundance/16S rRNA from the ARG-OAP.1 dataset but not MGE sums or ResFinder ARGs.

The possible connection between Proteobacteria – specifically, *Enterobacteriaceae* – ARGs, and MGEs is important but perhaps not surprising. Among the WHO priority pathogens list for research and development of new antibiotics (WHO, 2017), eight of the twelve identified antibiotic-resistant pathogens are Proteobacteria and three fall into the category of *Enterobacteriaceae,* with ESBL-producing and carbapenem-resistance *Enterobacteriaceae* noted among the three most critical pathogens. Flouroquinone-resistant *Salmonella* falls into the second, high priority category, while fluoroquinone-resistant Shigella species are in the third group of medium priority. *Enterobacteriaceae* are increasingly associated with mobile ARGs (reviewed in Patridge, 2015 and Iredell et al., 2016) and are speculated to be among the most

common hosts of class I integrons (Zhang et al., 2018). Recently, Li and colleagues (2021) found *E. coli*, and by extension *Enterobacteriaceae*, Gammaproteobacteria, and Proteobacteria to explain patterns in ARG profiles more than any other taxa in the gut microbiomes of 662 Danish children under one year of age. While the correlations we report cannot confirm *Enterobacteriaceae* as the host of specific ARGs in Galapagos water and wildlife samples, these observations can inform hypotheses for additional work in the future.

When considering only environmental samples (i.e. wastewater and human samples excluded), we also observed a positive correlation between ResFinder ARG sum abundance/16S rRNA and unclassified Bacilli (rho=0.52, p=4.11E-02). This is worth noting as Bacilli encompass several pathogens of clinical concern, including among the WHO priority pathogen list vancomycin-resistant Enterococcus faecium, methicillin-resistant, vancomycin-intermediate and resistant Staphylococcus aureus, and penicillin-non-susceptible Streptococcus pneumoniae. Antibiotic resistance among gram-positive Bacilli represents a major healthcare burden and source of hospital acquired infections (reviewed in Rice, 2006) and horizontal gene transfer of AMR determinants is thought to be common among members of this group (Lanza et al., 2015). This unclassified Bacilli genus was detected to some extent in all wildlife species, with the highest counts observed among land iguanas (data not shown). As discussed with *Enterobacteriaceae* correlations our work does not confirm Bacilli as the host of ARGs, but may serve as the basis for testing specific hypotheses in the future.

Conclusion

In conclusion, we found water, wildlife, and human samples from the Galapagos to be defined by distinct microbial communities. At the phyla level, Proteobacteria dominated all water samples while Firmicutes dominated all but two wildlife species and Actinobacteria dominated samples from children under two years of age. Amplicon sequencing of the 16S

rRNA gene of 50 giant tortoise, sea lion, and land iguana gut microbiomes revealed overlap in Firmicutes, Bacteroidetes, Actinobacteria. At the phyla level, giant tortoises were uniquely characterized by Planctomycetes while sea lions uniquely harbored Fusobacterial among the most dominant phyla. Mirroring the observations of ARGs and MGEs in Chapters 3 and 4, the impacts of wastewater discharge on marine water was reflected in the microbial community composition. In general, we noted good agreement between taxonomic inference from QIIME 2 and Metaxa 2 at high taxonomic levels, though Metaxa 2 provided less resolution at lower taxonomic levels in some cases. Both taxonomic annotation approaches suggested that Proteobacteria and Enterobacteriaceae specifically may be associated with overall ARG and MGE burden in environmental samples, though our methods preclude direct linkage. Instead, we hope that these data may serve as the foundation for additional hypothesis testing, specifically confirming if land iguana gut microbiomes harbor the functional resistance suggested by the genotypic ARG and MGE annotations. In line with the efforts of the WHO Global Tricycle Surveillance effort, ESBL E. coli could prove a useful indicator for AMR among Galapagos wildlife and the broader environment.

Chapter 5: Supplemental Figures

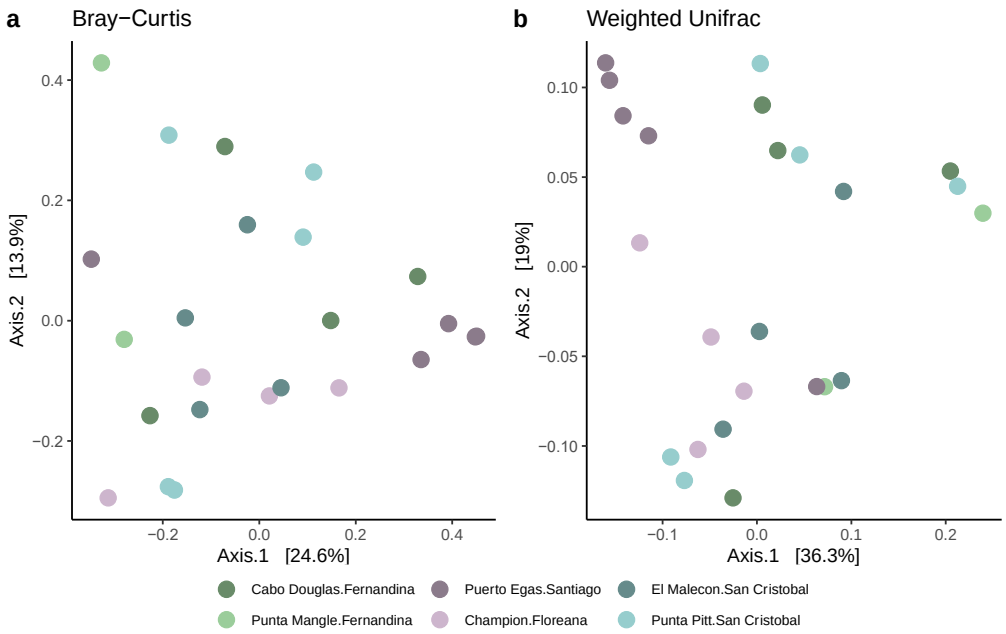

Figure S5.1: *Sampling location influences bacterial community composition among sea lions.* **a)** Bray-Curtis dissimilarity index (R-square=0.36, p=2e-04). **b)** Weighted Unifrac distances (R-square=0.34, p=0.0167). ADONIS test with 9,999 permutations.

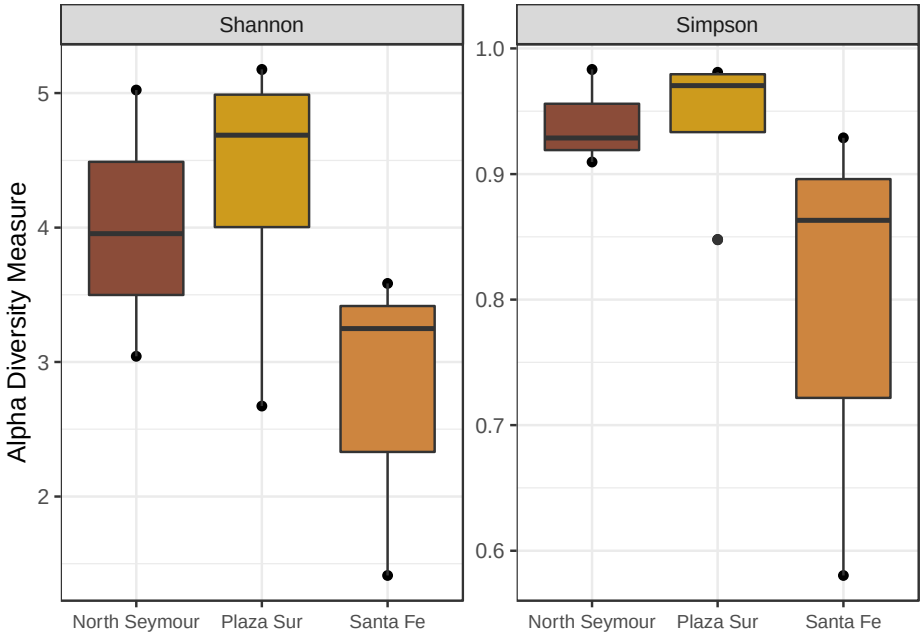

Figure S5.2: *Shannon and Simpson alpha diversity metrics for land iguanas by sampling location.*

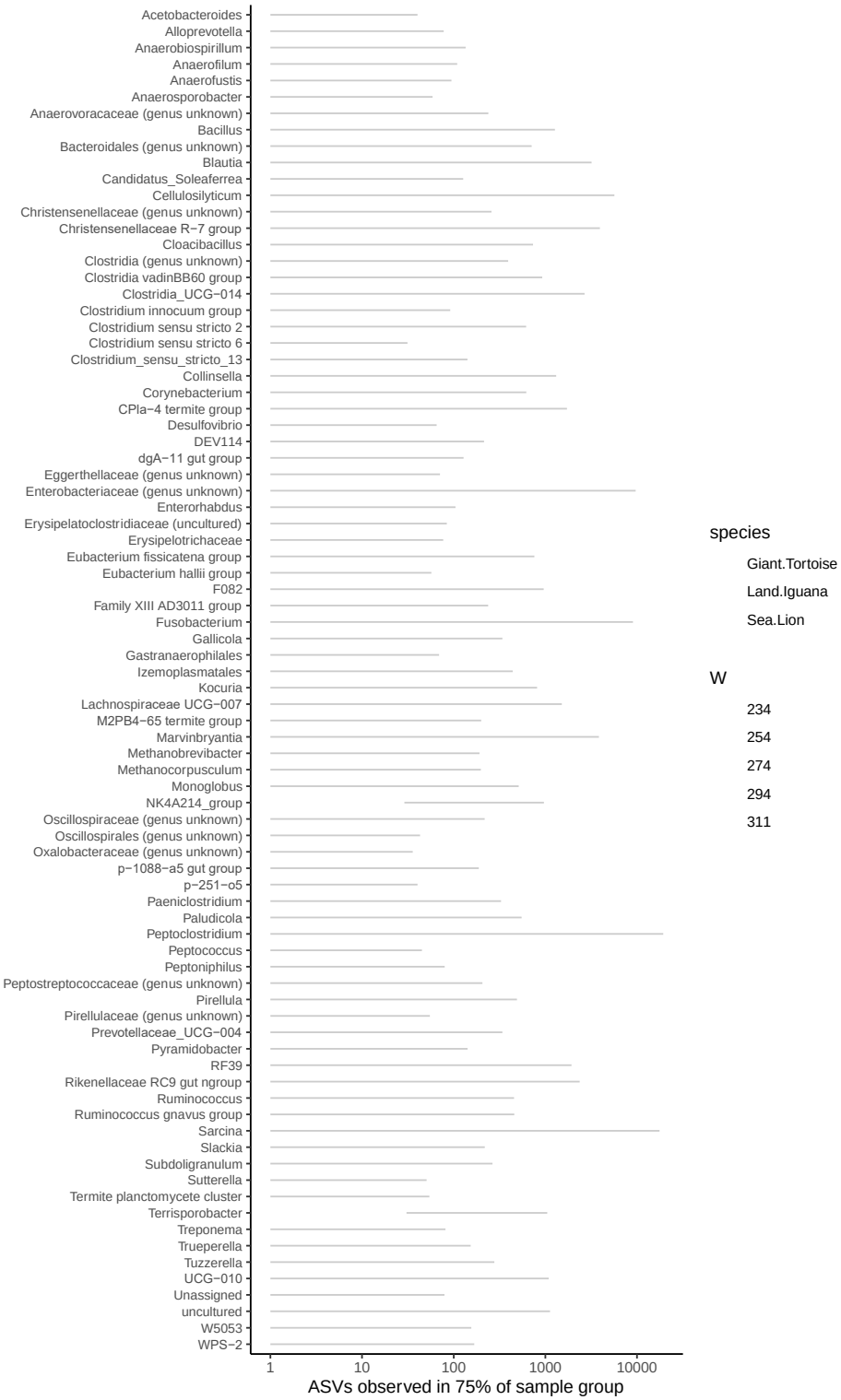

Figure S5.3: *Differentially abundant genera between three wildlife species according to the ANOSIM test implemented in QIIME 2.* Data presented are ASVs observed in 75% of sample group (i.e. the upper quartile from the ANOSIM test.) The size of the data point corresponds to the strength of the test statistic W.

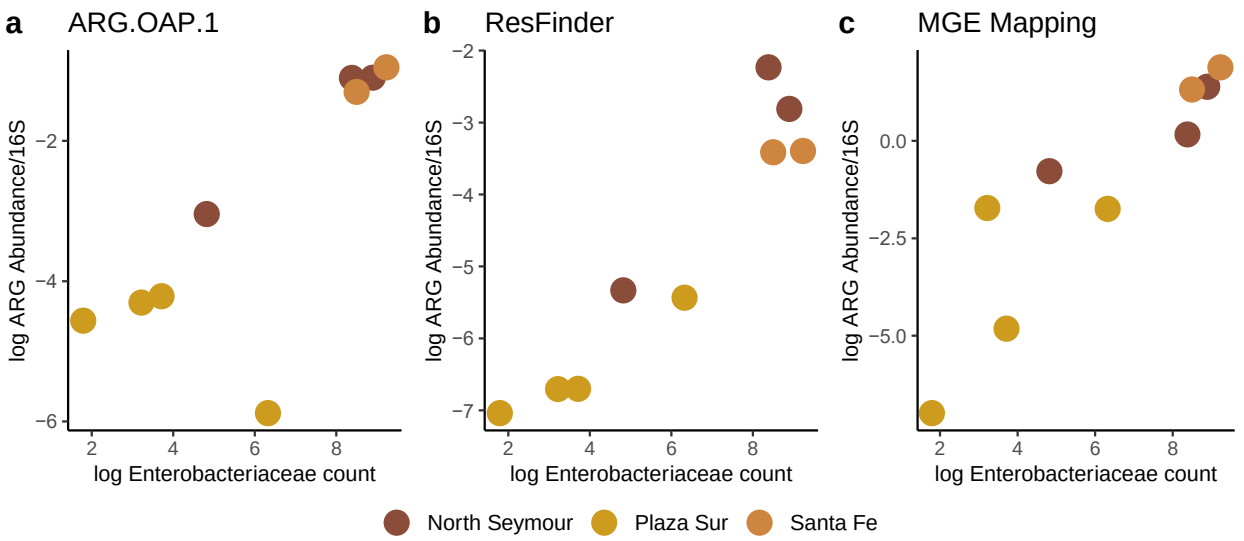

Figure S5.4: *Linear correlation between log-transformed Enterobacteriaceae counts and log-transformed MGE or ARG sum abundance/16S among land iguanas from three distinct sampling locations.* Enterobacteriaceae counts were taken from an ASV table rarified at 20,000 sequences/sample and summarized at the level of family (L5). **a)** Correlation with ARG-OAP.1 sum abundance. **b)** Correlation with ResFinder sum abundance**. c)** Correlation with MGE sum abundance.

Chapter 5: Supplemental Tables

Table S5.1: *Samples included in 16S rRNA amplicon sequencing*

Sample ID	DNA (ng/uL)	Sample Type	Location	Paired Metagenome
G19_55	23.2	Giant tortoise	Otoy Ranch	No
G19_57	11.2	Giant tortoise	Otoy Ranch	No
G19_59	40.6	Giant tortoise	Otoy Ranch	No
G19_61	24.9	Giant tortoise	Otoy Ranch	No
G19_63	36.1	Giant tortoise	Otoy Ranch	No
G19_65	22.4	Giant tortoise	Otoy Ranch	No
G19_67	33.8	Giant tortoise	Galapaguera	No
G19_69	28.6	Giant tortoise	Galapaguera	No
G19_71	44.2	Giant tortoise	Galapaguera	No
G19_74	10.9	Giant tortoise	Galapaguera	No
G19_77	22.6	Giant tortoise	Galapaguera	No
G19_78	20.8	Giant tortoise	Galapaguera	No
G18_12	34	Giant tortoise	Galapaguera	Yes
G18_14	28.6	Giant tortoise	Galapaguera	Yes
G18_18	18.4	Giant tortoise	Galapaguera	Yes
G18_19	32.5	Giant tortoise	Galapaguera	Yes
G19_113	1.5	Sea Lion	Punta Mangle/Fernandina	Yes
G19_114	11.6	Sea Lion	Punta Mangle/Fernandina	Yes
G19_121	2.4	Sea Lion	Cabo Douglas/Fernandina	Yes
G19_122	7.5	Sea Lion	Cabo Douglas/Fernandina	Yes
G19_123	10	Sea Lion	Cabo Douglas/Fernandina	Yes
G19_124	1.2	Sea Lion	Cabo Douglas/Fernandina	Yes
G19_134	17.2	Sea Lion	Puerto Egas/Santiago	Yes
G19_136	5.72	Sea Lion	Puerto Egas/Santiago	Yes
G19_140	4.82	Sea Lion	Puerto Egas/Santiago	Yes
G19_142	7.79	Sea Lion	Puerto Egas/Santiago	Yes
G19_145	5.61	Sea Lion	Puerto Egas/Santiago	Yes
G19_152	2.2	Sea Lion	Champion/Floreana	Yes
G19_148	1.1	Sea Lion	Champion/Floreana	Yes
G19_159	18	Sea Lion	Champion/Floreana	Yes
G19_163	3.6	Sea Lion	Champion/Floreana	Yes
G19_181	15.8	Sea Lion	Punta Pitt/San Cristobal	Yes
G19_172	57	Sea Lion	Punta Pitt/San Cristobal	Yes
G19_174	11.8	Sea Lion	Punta Pitt/San Cristobal	Yes
G19_177	15.8	Sea Lion	Punta Pitt/San Cristobal	Yes
G19_188	30.9	Sea Lion	Punta Pitt/San Cristobal	Yes
G19_194	22.6	Sea Lion	El Malecon/San Cristobal	Yes
G19_197	12.2	Sea Lion	El Malecon/San Cristobal	Yes
G19_199	32	Sea Lion	El Malecon/San Cristobal	Yes
G19_201	36.4	Sea Lion	El Malecon/San Cristobal	Yes
G19_37	8.25	Land Iguana	North Seymour	Yes
G19_26	1.8	Land Iguana	North Seymour	Yes

G19_43	2	Land Iguana	North Seymour	Yes
G19_14	11.2	Land Iguana	Plaza Sur	Yes
G19_45	1.1	Land Iguana	Plaza Sur	Yes
G19_30	17.8	Land Iguana	Plaza Sur	Yes
G19_36	11.8	Land Iguana	Plaza Sur	Yes
G19_31	1.6	Land Iguana	Santa Fe	Yes
G19_23	10	Land Iguana	Santa Fe	No
G19_34	14.4	Land Iguana	Santa Fe	Yes

Table S5.2: *Spearman rank correlation between bacterial families, MGE sum abundance, and ARG sum abundance using taxonomic assignments from Metaxa2.* Total n=76 samples included in analysis.

Sum abundance/16S rRNA	Family	rho	Adjusted p-value
MGE	*Aeromonadaceae*	0.75	8.09E-03
MGE	*Enterobacteriaceae*	0.69	9.78E-08
MGE	*Desulfobulbaceae*	0.67	2.40E-02
MGE	*Geobacteraceae*	0.58	4.22E-02
MGE	*Xanthomonadaceae*	0.56	4.22E-02
MGE	*Unclassified Clostridiales*	-0.38	2.41E-02
MGE	*Spirochaetaceae*	-0.53	4.22E-02
MGE	*Leuconostocaceae*	-0.71	4.64E-02
ARG-OAP.1	*Rhodocyclaceae*	0.86	4.58E-03
ARG-OAP.1	*Aeromonadaceae*	0.69	2.37E-02
ARG-OAP.1	*Geobacteraceae*	0.68	9.88E-03
ARG-OAP.1	*Burkholderiaceae*	0.62	2.40E-02
ARG-OAP.1	*Comamonadaceae*	0.59	3.60E-02
ARG-OAP.1	*Enterobacteriaceae*	0.58	8.35E-05
ARG-OAP.1	*Unclassified Burkholderiales*	0.53	2.23E-02
ARG-OAP.1	*Unclassified Micrococcales*	0.53	8.09E-03
ARG-OAP.1	*Unclassified Actinobacteria.*	0.44	1.43E-02
ARG-OAP.1	*Unclassified Clostridiales*	-0.46	4.79E-03
ARG-OAP.1	*Peptostreptococcaceae*	-0.46	8.09E-03
ARG-OAP.1	*Unclassified Cytophagales*	-0.51	4.99E-02
ARG-OAP.1	*Unclassified Flavobacteriales*	-0.56	2.40E-02
ARG-OAP.1	*Cyanobacteria SubsectionIII FamilyI*	-0.59	3.70E-02
ARG-OAP.1	*Unclassified.Cyanobacteria.*	-0.68	4.22E-02
ARG-OAP.1	*Lentisphaeraceae*	-0.75	1.09E-02
ResFinder	*Aeromonadaceae*	0.77	9.88E-03
ResFinder	*Actinomycetaceae*	0.55	2.40E-02
ResFinder	*Enterobacteriaceae*	0.51	3.26E-03

ResFinder	*Streptococcaceae*	0.47	2.40E-02
ResFinder	*Unclassified Micrococcales*	0.46	3.29E-02
ResFinder	*Unclassified Clostridiales*	-0.40	2.12E-02
ResFinder	*Peptostreptococcaceae*	-0.45	9.88E-03
ResFinder	*Cystobacteraceae*	-0.70	4.22E-02
ResFinder	*Coxiellaceae*	-0.78	1.44E-02

CHAPTER 6: LINKING WHAT, WHERE, AND WHO: CORRELATION OF RESISTIOME, MOBILOME, AND PHYLOGENETIC DATASETS

Introduction

Collectively, the data presented in Chapters 3, 4, and 5 suggest linkages between ARGs, MGEs, and the underlying phylogeny of bacterial communities. Previous studies in soil (Forsberg et al., 2014) and the infant gut (Pärnänen et al., 2018) have documented clear relationships between bacterial community structure, the resistome, and the mobilome. We aimed to determine if similar relationships exist in our data sets using Kendall rank correlation between ARG sums, MGE sums, and ddPCR data, as well as matrix correlation between the ARG, MGE, and taxonomic data sets.

Materials & Methods

ARG, MGE, and ddPCR data sets were compared on the basis of sum abundance/16S rRNA using linear regression when log-transformed data met the assumptions of the Shapiro-Wilkes test. Non-parametric statistics (i.e. Kendall rank correlation) were empolyed both when log transformation failed to produce a normal distribution and for normally distributed data as a comparison to the respective parametric test. In the case of ddPCR data, samples below the limit of detection (i.e. 2 or fewer positive droplets) were assigned a value of 0 and were excluded when performing log transformations. The similarity of ARG, MGE, and taxonomic matrices was assessed using the Mantel test (method = kendall) to compare distances in the Horn-Morisita similarity index as calculated in the R package vegan (Oksanen et al., 2019) version 2.5.7.

Results

Correlation between resistome and mobilome data sets

Calculation of the Kendall rank correlation revealed varied levels of agreement between resistome and mobilome datasets (**Table 6.1**). The greatest correlation was recorded between ARG-OAP.1 and ResFinder sum abundance/16S rRNA (tau = 0.60, p< 2.2e-16) followed by ARG-OAP.1 and MGE sum abundance/16S rRNA (tau = 0.56, p<4.086e-15). Correlation between MGE and ResFinder sum abundance/16S rRNA was similarly high (tau = 0.51, p=1.109e-12). When comparing resistome and mobilome mapping based approaches to quantification of *intI1* variants with ddPCR, MGE sum abundance/16S rRNA and clinical *intI1*/16S rRNA accounted for the strongest correlation (tau = 0.37, p=1.244e-05), closely followed by MGE and general *intI1* (tau = 0.36, p=2.012e-06). Correlation of both *intI1* targets with ResFinder and ARG-OAP.1 sum abundance/16S rRNA yielded coefficients ranging from 0.28 to 0.33, and all were associated with significant p-values.

Table 6.1: *Correlation of ARG, MGE, and ddPCR datasets using the non-parametric Kendall rank correlation*

	ARG-OAP.1		**MGE**		**ResFinder**		**General *intI1* (ddPCR)**	
	tau	p-value	tau	p-value	tau	p-value	tau	p-value
MGE	0.56	4.086e-15	--	--	--	--	--	--
ResFinder	0.60	< 2.2e-16	0.51	1.109e-12	--	--	--	--
General *intI1* (ddPCR)	0.28	0.000184	0.36	2.012e-06	0.33	1.27e-05	--	--
Clinical *intI1* (ddPCR)	0.30	0.000228	0.37	1.244e-05	0.32	0.000113	0.76	< 2.2e-16

Additionally, linear regression was performed when both predictor and response variables fit a normal distribution following log transformation (**Table 6.2**). Log sum abundance/16S rRNA correlated well between ARG-OAP.1 and MGE **(Figure 6.1),** with each increase in 1 log ARG sum abundance/16S rRNA translating to an increase of 0.47 log MGE sum abundance/16S rRNA (adjusted R-squared = 0.57, p<2e-16). Both ARGs and MGEs showed significant positive correlations with ddPCR detected clinical *intI1* (**Figure 6.1, Table 6.2**), though the adjusted R-squared was higher for MGEs.

Table 6.2: *Correaltion of ARG, MGE, and ddPCR datasets using linear regression where both predictor and response variables were normally distributed following log transformation*

	log ARG-OAP.1 sum/16S			**log MGE sum/16S**		
	Estimate	Adjusted R-squared	P-value	Estimate	Adjusted R-squared	P-value
log MGE	0.47	0.57	< 2e-16	--	--	--
log Clinical *intI1* (ddPCR)	1.52	0.34	3.77e-05	1.35	0.53	5.16e-08

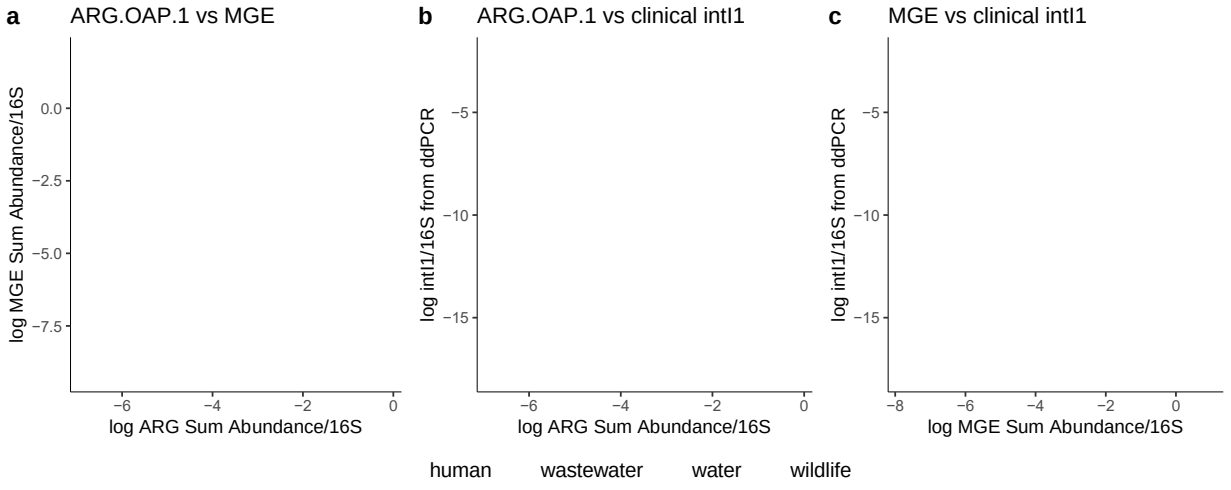

Figure 6.1: *Correlation of ARG, MGE, and ddPCR data sets using linear models.* **a)** Correlation of log-transformed ARG with MGE sum abundace/16S rRNA. **b)** Correlation of log-transformed ARG sum abundace/16S rRNA with clinical *intI1* from ddPCR. **c)** Correlation of log-transformed MGE sum abundace/16S rRNA with clinical *intI1* from ddPCR.

Finally, we examined the extent to which bacterial community composition explained ARG and MGE composition. For each data set, distances between samples were calculated based on the Horn-Morisita index and compared using the Mantel test. We first performed this analysis on the 37 samples for which paired metagenome and taxonomic assignments as produced by 16S rRNA amplicon sequencing (i.e. QIIME2) were available. Comparison of the taxonomic and ARG-OAP.1 distance matrices yielded a mantel test statistic equal to 0.2733 (sig. = 0.001), compared to test statistics of 0.1437 for ResFinder (sig. = 0.005) and 0.1492 for MGE (sig. = 0.008). When performing distance correlations on the full dataset of metagenomes, correlation of the taxonomic assignments produced using Metaxa2 with the ARG-OAP.1 distance matrix yielded a comparatively higher test statistic of 0.4842 (sig. = 0.001), followed by MGE (r = 0.2729, sig. = 0.001) and ResFinder (r = 0.1625, sig. = 0.001). ARG and MGE composition were also highly correlated, with the highest agreement recorded between ResFinder and MGE (r = 0.488, sig. = 0.001). Slightly lower agreement was observed when comparing ARG-OAP.1 with ResFinder (r = 0.339) and MGE (r = 0.3819), though both test statistics were significant.

Table 6.3: *Mantel test on sample distances based on Horn-Morisita similarity index. r = Mantel test statistic.*

	ARG-OAP.1		ResFinder		MGE	
	r	Significance	r	Significance	r	Significance
ResFinder†	0.339	0.001	--	--	--	--
MGE	0.3819	0.001	0.448	0.001	--	--
Metaxa2‡	0.4842	0.001	0.1625	0.001	0.2729	0.001

† For comparisons involving ResFinder, distance matrices were filtered to exclude samples with all-zero observations for ResFinder (n=82 retained).
‡ For comparisons involving Metaxa2, distance matrices were filtered to exclude samples with fewer than 5,000 SSU rRNA counts for Metaxa2 (n=76 retained). The comparison of ResFinder and Metaxa2 included 73 samples retained when both filters were applied.

Discussion and Conclusion

This brief chapter explored the interrelatedness of ARG, MGE, and taxonomic data sets. We found that all pairwise Kendall rank correlations of sum abundance from ARG-OAP.1, ResFinder, and MGE yielded significant p-values with tau ≥ 0.50. Correlations between clinical *intI* as detected by ddPCR and sum abundance from ARG-OAP.1, ResFinder, and MGE yielded generally lower tau coefficients, ranging from 0.28 to 0.37, though all p-values were still significant. Using qPCR arrays, Zheng and colleagues (2020) reported a non-parametric Spearman correlation of rho=0.748 ($p<0.001$) between the total abundance of ARGs and MGEs in wastewater, while Pearson correlations between clinical *intI* (using the Gillings et al., 2015 primers) and the absolute abundance of ARGs yielded a correlation of R=0.794 ($p<0.001$) in activated sludge and R=0.641 ($p<0.001$) in permeate. The lower correlations reported in the represent study may be explained by differences in methodology: whereas all targets were measured by qPCR arrays in the work by Zheng and colleagues (2020), our data was produced by two different ARG annotation pipelines plus ddPCR for the clinical *intI1* target. Moreover, the matrix correlation performed between Metaxa2 bacterial taxonomy and ARG-OAP.1 (r=0.48) was on par with values reported by Pärnänen et al. (2018) using the same methods, though correlations were lower between taxonomy and MGEs (r=0.27) or ResFinder ARGs (r=0.16). However, the work by Pärnänen et al. (2018) included only samples from infants, whereas our dataset spanned wastewater, human, water, and wildlife samples which exhibited distinct patterns in the correaltion of taxonomy, ARGs, MGEs, and clinical *intI1* (see Figure 6.1b as an example). Taken together, we view these inter-data set correlations as supporting the growing evidence that 1) bacterial community structure and resistome structure are closely related, and 2), resistome and mobiliome structures are closely related.

CHAPTER 7: DISSERTATION CONCLUSIONS AND IMPLICATIONS

This research represents the first One Health Survey of Galapagos resistomes and mobilomes across a range of human, environmental, and wildlife reservoirs. Through generation and annotation of 90 metagenomes, we sought to answer *what* antibiotic resistance genes are present, *where* they may be located in regards to mobile genetic elements, and *who* may be their presumptive bacterial hosts. In the course of resistome characterization, we compared three publicly available pipelines for annotating ARGs from metagenomic reads and found that while different approaches generally reach similar conclusions, pipeline selection should be based on the research question and study system. Annotation of ARGs in metagenomes pointed to distinct resistomes between environmental, wildlife, and human reservoirs, with ARG sum abundance and diversity generally increasing along a gradient of human influence.

In the course of mobilome characterization, we developed a novel ddPCR assay capable of distinguishing between general and clinical sequence variants of the class I integron-integrase, a proposed marker of anthropogenic influence in the environment. Accordingly, we detected the clinical variant at lower frequencies among wildlife, though differences were noted by sampling location. We found that ARG and MGE sums correlated well in regards to both sum abundance and between-sample distances.

In the course of phylogenetic characterization, we provided new and expanded data on the community composition of Galapagos wildlife gut microbiomes, producing the first ever 16S rRNA data set for Galapagos sea lions and the largest data set for Galapagos giant tortoises to date. Using taxonomic inferences made on SSU rRNA sequences in metagenomes, we were able

to identify the presumptive hosts of select beta-lactam ARGs, such as the case of bla_{SED-1} and *Citrobacter sedlakii* in land iguanas. Finally, we provide evidence for a possible relationship between the family *Enterobacteriaceae*, the resistome, and the mobilome. Taken together, our results suggest that Galapagos wildlife are at the present minimally impacted by human-associated ARGs, but increasing detection and diversity of beta-lactams in sea lions living in close proximity to humans, for example, is cause for concern. Moreover, the situation in land iguanas demands further research involving host and phenotypic confirmation of antibiotic resistance. Given the relationship between *Enterobacteriaceae* and ARGs sums and our ability to detect the presumptive hosts of specific beta-lactam genes among taxonomic annotations of the metagenomes, we recommend that future AMR surveillance efforts in the Galapagos consider ESBL *E. coli*. Importantly, ESBL *E. coli* has been selected by the WHO Tricycle program as the global, One Health indicator for AMR surveillance. Accordingly, efforts to surveil ESBL *E. coli* in the Galapagos would complement larger global efforts and provide risk information pertinent to both humans and animals.

Dissertation Strengths

This research benefited from the inclusion of sample types encompassing a One Health perspective, with representation from environmental, human, and wildlife reservoirs. Importantly, inclusion of so many sample types was only possible through the generosity, time, and efforts of our collaborators. This includes the provision of giant tortoise, land iguana, marine iguana, and red footed booby samples from Dr. Greg Lewbart; the provision of sea turtle samples by Dr. Greg Lewbart, Juan Pablo Muñoz, and Daniela Alarcon; the provision of sea lion and fur seal samples by Dr. Diego Páez Rosas; and the provision of human fecal samples by Dr. Amanda Thompson. Such a sampling effort would be beyond the financial and logistical capabilities of a single researcher. Additionally, this research was strengthened by the combination of

bioinformatic analyses which allowed for simultaneous characterization of the resistome, mobilome, and taxonomy of bacterial communities. The phylogenetic assignments produced in Chapter 5 provided useful context to the ARG and MGE data at a surprisingly specific scale (i.e. genus and species, in some cases.) Finally, the paired data between metagenomes and class I integron ddPCR detection served as a useful check between the two methods. Using the ddPCR assay developed as part of this work, we were able to assay > 250 samples for a class I integron, producing additional data on anthropogenic impacts in the Galapagos beyond what was produced from the 90 metagenomes.

Dissertation Limitations

There are several important limitations associated with this work, with the first and important limitations relating to sample collection and storage. For example, sample collection proved to be difficult for certain animal species, especially sea turtles and marine iguanas. Due to the slow digestion time of reptiles, swabbing the cloaca did not always result in sufficient, solid, fecal material. We processed what we were able to collect, but we suspect the low sequencing output among some individuals of these species (paired with high counts of eukaryotic SSU rRNA in the metagenomes) to be a direct consequence of the sample quality. As a result, we cannot confirm that ARGs were truly absent among certain individuals, but rather, were beyond our ability to detect using the sampling methodology. Related to this point are the differences in sampling methodology between wildlife species, such as the use of fecal loops in sea lions versus cotton swabs in other animals. While these differences were unavoidable due to resource and logistical constraints, we acknowledge their ability to influence DNA extraction yields and downstream results. Finally, differences in sample and DNA extract storage times were not ideal, but unavoidable due to permitting and travel logistics.

Another limitation of this research relates to our inability to confirm the host or phenotype of ARGs detected through metagenomic annotation. Our study would be improved by paired culturing of organisms such as *Enterobacteriaceae* and/or confirmation of phenotypic resistance through construction of clone libraries. We were also limited by representation of environmental and human samples from only San Cristobal Island. Indeed, characterization of freshwater and marine samples from across the archipelago would greatly enhance this work. Moreover, this research was limited by using fecal samples from children as the proxy for human resistomes. While we used data from wastewater and children's gut microbiomes as a proxy for anthropogenic resistomes, the addition of fecal samples from adults would be greatly beneficial.

Research Implications and Future Directions

Taken together, we hope that data from this study can be used to inform hypothesis testing in the Galapagos and beyond. While we found that wildlife generally harbored fewer ARGs in terms of abundance and diversity, we view the situation in land iguanas as both curious and concerning. Much remains to be learned regarding the land iguana resistome, specifically in confirming the presence of specific *Enterobacteriaceae* that appear to be associated with ARGs and phenotypic resistance. At present, it is unclear whether *Enterobacteriaceae* represent natural commensals of land iguanas, or if they are the consequence of human contact. Notably, earlier work on antibiotic resistance in land iguanas by Thaller et al. (2010) and Wheeler et al. (2012) detected low levels of antibiotic resistance when culturing *E. coli* and *Salmonella*. It could be that the land iguana resistome has changed markedly in the last ten years, or that the ARGs detected in the present study are hosted by other bacteria.

Moreover, our results support the applicability of ESBL *E.coli* as an AMR surveillance indicator across humans, animals, and the environment. While we did not culture *E. coli* in this work, the connection we observed between beta-lactam ARGs and presumptive

Enterobacteriaceae hosts is notable. Our previous work on water quality on San Cristobal demonstrated high levels of total coliforms and *E. coli* in freshwater, raising questions around the possibility of environmental *E. coli* in the Galapagos and in tropical regions more broadly.

REFERENCES

Abraham, E.P., and Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. Nature *146*, 837.

Acquarone, M., Salgado-Flores, A., and Sundset, M.A. (2020). The Bacterial Microbiome in the Small Intestine of Hooded Seals (Cystophora cristata). Microorganisms *8*, 1664.

Al-Bahry, S., Mahmoud, I., Elshafie, A., Al-Harthy, A., Al-Ghafri, S., Al-Amri, I., and Alkindi, A. (2009). Bacterial flora and antibiotic resistance from eggs of green turtles Chelonia mydas: An indication of polluted effluents. Marine Pollution Bulletin *58*, 720–725.

Al-Bahry, S.N., Mahmoud, I.Y., Al-Zadjali, M., Elshafie, A., Al-Harthy, A., and Al-Alawi, W. (2011). Antibiotic resistant bacteria as bio-indicator of polluted effluent in the green turtles, Chelonia mydas in Oman. Marine Environmental Research *71*, 139–144.

Alava, J.J., Ikonomou, M.G., Ross, P.S., Costa, D., Salazar, S., Aurioles-Gamboa, D., and Gobas, F.A.P.C. (2009). Polychlorinated Biphenyls and Polybrominated Diphenyl Ethers in Galapagos Sea Lions (Zalophus wollebaeki). Environmental Toxicology and Chemistry *28*, 2271–2282.

Alawi, M., Schneider, B., and Kallmeyer, J. (2014). A procedure for separate recovery of extra- and intracellular DNA from a single marine sediment sample. J. Microbiol. Methods *104*, 36–42.

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J., and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. Nat Rev Micro *8*, 251–259.

Allignet, J., and el Solh, N. (1995). Diversity among the gram-positive acetyltransferases inactivating streptogramin A and structurally related compounds and characterization of a new staphylococcal determinant, vatB. Antimicrob Agents Chemother *39*, 2027–2036.

Andreu-Sánchez, S., Chen, L., Wang, D., Augustijn, H.E., Zhernakova, A., and Fu, J. (2021). A Benchmark of Genetic Variant Calling Pipelines Using Metagenomic Short-Read Sequencing. Front. Genet. *12*.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P., and Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome *6*, 23.

Arboleya, S., Watkins, C., Stanton, C., and Ross, R.P. (2016). Gut Bifidobacteria Populations in Human Health and Aging. Front Microbiol *7*, 1204.

Ashbolt, N.J., Amézquita, A., Backhaus, T., Borriello, P., Brandt, K.K., Collignon, P., Coors, A., Finley, R., Gaze, W.H., Heberer, T., et al. (2013). Human Health Risk Assessment (HHRA) for environmental development and transfer of antibiotic resistance. Environ. Health Perspect. *121*, 993–1001.

Baltz, R. H. (2005). Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News* 55, 186–196.

Barbosa, T.M., Scott, K.P., and Flint, H.J. (1999). Evidence for recent intergeneric transfer of a new tetracycline resistance gene, tet(W), isolated from Butyrivibrio fibrisolvens, and the occurrence of tet(O) in ruminal bacteria. Environ Microbiol *1*, 53–64.

Barlow, M. (2009). What antimicrobial resistance has taught us about horizontal gene transfer. Methods Mol. Biol. *532*, 397–411.

Barraud, O., Baclet, M.C., Denis, F., and Ploy, M.C. (2010). Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. J Antimicrob Chemother *65*, 1642–1645.

Bengtsson-Palme, J., Angelin, M., Huss, M., Kjellqvist, S., Kristiansson, E., Palmgren, H., Larsson, D.G.J., and Johansson, A. (2015). The Human Gut Microbiome as a Transporter of Antibiotic Resistance Genes between Continents. Antimicrob. Agents Chemother. *59*, 6551–6560.

Bengtsson-Palme, J., Hartmann, M., Eriksson, K.M., Pal, C., Thorell, K., Larsson, D.G.J., and Nilsson, R.H. (2015). METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. Mol Ecol Resour *15*, 1403–1414.

Benveniste, R., and Davies, J. (1973). Aminoglycoside Antibiotic-Inactivating Enzymes in Actinomycetes Similar to Those Present in Clinical Isolates of Antibiotic-Resistant Bacteria. Proc Natl Acad Sci U S A *70*, 2276–2280.

Berglund, F., Österlund, T., Boulund, F., Marathe, N.P., Larsson, D.G.J., and Kristiansson, E. (2019). Identification and reconstruction of novel antibiotic resistance genes from metagenomes. Microbiome *7*, 52.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson, S., and Relman, D.A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nat Commun *7*, 10516.

Bisanz, J. qiime2R: Importing QIIME2 artificates and associated data into R Sessions. R package version 0.99.6. 2021.

Bischoff, K.M., White, D.G., Hume, M.E., Poole, T.L., and Nisbet, D.J. (2005). The chloramphenicol resistance gene cmlA is disseminated on transferable plasmids that confer multiple-drug resistance in swine Escherichia coli. FEMS Microbiol Lett *243*, 285–291.

Blasi, M.F., Migliore, L., Mattei, D., Rotini, A., Thaller, M.C., and Alduina, R. (2020). Antibiotic Resistance of Gram-Negative Bacteria from Wild Captured Loggerhead Sea Turtles. Antibiotics (Basel) *9*, 162.

Bohn, C., and Bouloc, P. (1998). The Escherichia coli cmlA gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl-beta-D-thiogalactopyranoside exclusion and spectinomycin sensitivity. J Bacteriol *180*, 6072–6075.

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

Bonnedahl, J., Drobni, P., Johansson, A., Hernandez, J., Melhus, A., Stedt, J., Olsen, B., and Drobni, M. (2010). Characterization, and comparison, of human clinical and black-headed gull (Larus ridibundus) extended-spectrum beta-lactamase-producing bacterial isolates from Kalmar, on the southeast coast of Sweden. J. Antimicrob. Chemother. *65*, 1939–1944.

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., et al. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy *75*, 3491–3500.

Bottacini, F., Morrissey, R., Esteban-Torres, M., James, K., van Breen, J., Dikareva, E., Egan, M., Lambert, J., van Limpt, K., Knol, J., et al. (2018). Comparative genomics and genotype-phenotype associations in Bifidobacterium breve. Sci Rep *8*, 10633.

Bray, J.R., and Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs *27*, 325–349.

Cao, L., Chen, H., Wang, Q., Li, B., Hu, Y., Zhao, C., Hu, Y., and Yin, Y. (2020). Literature-Based Phenotype Survey and In Silico Genotype Investigation of Antibiotic Resistance in the Genus Bifidobacterium. Curr Microbiol *77*, 4104–4113.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J *6*, 1621–1624.

Centers for Disease Control and Prevention (2018). One Health Basics. https://www.cdc.gov/onehealth/basics/index.html

Centers for Disease Control and Prevention (2019). Health Information for Travelers to Ecuador, including the Galápagos Islands. **https://wwwnc.cdc.gov/travel/destinations/ecuador/traveler/packing-list**

Cesare, A.D., Petrin, S., Fontaneto, D., Losasso, C., Eckert, E.M., Tassistro, G., Borello, A., Ricci, A., Wilson, W.H., Pruzzo, C., et al. (2018). ddPCR applied on archived Continuous Plankton Recorder samples reveals long-term occurrence of class 1 integrons and a sulphonamide resistance gene in marine plankton communities. Environmental Microbiology Reports *10*, 458–464.

Chater, K.F., and Bruton, C.J. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. EMBO J. *4*, 1893–1897.

Che, Y., Xia, Y., Liu, L., Li, A.-D., Yang, Y., and Zhang, T. (2019). Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome *7*, 44.

Cristóbal-Azkarate, J., Dunn, J.C., Day, J.M.W., and Amábile-Cuevas, C.F. (2014). Resistance to Antibiotics of Clinical Relevance in the Fecal Microbiota of Mexican Wildlife. PLOS ONE *9*, e107719.

D'Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W.L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., et al. (2011). Antibiotic resistance is ancient. Nature *477*, 457–461.

D'Costa, V.M., McGrann, K.M., Hughes, D.W., and Wright, G.D. (2006). Sampling the Antibiotic Resistome. Science *311*, 374–377.

D'Souza, A.W., Potter, R.F., Wallace, M., Shupe, A., Patel, S., Sun, X., Gul, D., Kwon, J.H., Andleeb, S., Burnham, C.-A.D., et al. (2019). Spatiotemporal dynamics of multidrug resistant bacteria on intensive care unit surfaces. Nat Commun *10*, 4569.

Dantas, G., and Sommer, M.O. (2012). Context matters — the complex interplay between resistome genotypes and resistance phenotypes. Current Opinion in Microbiology *15*, 577–582.

Dantas, G., Sommer, M.O.A., Oluwasegun, R.D., and Church, G.M. (2008). Bacteria Subsisting on Antibiotics. Science *320*, 100–103.

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

Deem, S.L., Cruz, M.B., Higashiguchi, J.M., and Parker, P.G. (2012). Diseases of poultry and endemic birds in Galapagos: implications for the reintroduction of native species. Animal Conservation *15*, 73–82.

Deem, S.L., Parker, P.G., and Miller, R.E. (2008). Building Bridges: Connecting the Health and Conservation Professions. Biotropica *40*, 662–665.

Delport, T.C., Harcourt, R.G., Beaumont, L.J., Webster, K.N., and Power, M.L. (2015). Molecular detection of antibiotic-resistance determinants in Escherichia coli isolated from the endangered Australian sea lion (*Neophoca cinerea*). Journal of Wildlife Diseases *51*, 555–563.

Delport, T.C., Power, M.L., Harcourt, R.G., Webster, K.N., and Tetu, S.G. (2016). Colony Location and Captivity Influence the Gut Microbial Community Composition of the Australian Sea Lion (Neophoca cinerea). Appl Environ Microbiol *82*, 3440–3449.

Dolejska, M. (2020). Antibiotic-Resistant Bacteria in Wildlife. In Antibiotic Resistance in the Environment : A Worldwide Overview, C.M. Manaia, E. Donner, I. Vaz-Moreira, and P. Hong, eds. (Cham: Springer International Publishing), pp. 19–70.

Dolejska, M., and Papagiannitsis, C.C. (2018). Plasmid-mediated resistance is going wild. Plasmid *99*, 99–111.

Dolejská, M., Bierošová, B., Kohoutová, L., Literák, I., and Čížek, A. (2009). Antibiotic-resistant Salmonella and Escherichia coli isolates with integrons and extended-spectrum beta-lactamases in surface water and sympatric black-headed gulls. Journal of Applied Microbiology *106*, 1941–1950.

Doster, E., Lakin, S.M., Dean, C.J., Wolfe, C., Young, J.G., Boucher, C., Belk, K.E., Noyes, N.R., and Morley, P.S. (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.

Dungan, R.S., and Bjorneberg, D.L. (2020). Antibiotic resistance genes, class 1 integrons, and IncP-1/IncQ-1 plasmids in irrigation return flows. Environmental Pollution *257*, 113568.

Ekwanzala, M.D., Dewar, J.B., Kamika, I., and Momba, M.N.B. (2020). Genome sequence of carbapenem-resistant Citrobacter koseri carrying blaOXA-181 isolated from sewage sludge. J Glob Antimicrob Resist *20*, 94–97.

Ellabaan, M.M.H., Munck, C., Porse, A., Imamovic, L., and Sommer, M.O.A. (2021). Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. Nat Commun *12*, 2435.

Epler, B. (2007). Tourism, the Economy, Population Growth, and Conservation in Galapagos. Puerto Ayora, Santa Cruz: Charles Darwin Research Foundation.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ *3*, e1319.

Essack, S.Y. (2018). Environment: the neglected component of the One Health triad. The Lancet Planetary Health *2*, e238–e239.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

Fang, H., Huang, K., Yu, J., Ding, C., Wang, Z., Zhao, C., Yuan, H., Wang, Z., Wang, S., Hu, J., et al. (2019). Metagenomic analysis of bacterial communities and antibiotic resistance genes in the Eriocheir sinensis freshwater aquaculture environment. Chemosphere *224*, 202–211.

Favier, C.F., Vaughan, E.E., De Vos, W.M., and Akkermans, A.D.L. (2002). Molecular monitoring of succession of bacterial communities in human neonates. Appl Environ Microbiol *68*, 219–226.

Feng, J., Li, B., Jiang, X., Yang, Y., Wells, G.F., Zhang, T., and Li, X. (2018). Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. Environ. Microbiol. *20*, 355–368.

Fisher, J.F., Meroueh, S.O., and Mobashery, S. (2005). Bacterial Resistance to β-Lactam Antibiotics: Compelling Opportunism, Compelling Opportunity. Chem. Rev. *105*, 395–424.

Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O.A., and Dantas, G. (2012). The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. Science *337*, 1107–1111.

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

Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J., and Cotter, P.D. (2016). 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. BMC Microbiology *16*, 123.

Galapagos National Park Service (2017). Informe Anual de Visitantes que Ingresaron a Las Áreas Protegidas de Galápagos 2016. Puerto Ayora, Santa Cruz, CA: Galapagos National Park.

Galapagos National Park Service (2019). Informe Anual de Visitantes que Ingresaron a Las Áreas Protegidas de Galápagos 2018. Puerto Ayora, Santa Cruz, CA: Galapagos National Park.

Gambino, D., Vicari, D., Vitale, M., Schirò, G., Mira, F., Giglia, M.L., Riccardi, A., Gentile, A., Giardina, S., Carrozzo, A., et al. (2021). Study on Bacteria Isolates and Antimicrobial Resistance in Wildlife in Sicily, Southern Italy. Microorganisms *9*.

Gardner, P., Smith, D., Beer, H., and Moellering, R. (1969). Recovery of resistance (r) factors from a drug-free community. The Lancet *294*, 774–776.

Garza-Ramos, G., Xiong, L., Zhong, P., and Mankin, A. (2001). Binding Site of Macrolide Antibiotics on the Ribosome: New Resistance Mutation Identifies a Specific Interaction of Ketolides with rRNA. J Bacteriol *183*, 6898–6907.

Gibson, M.K., Wang, B., Ahmadi, S., Burnham, C.-A.D., Tarr, P.I., Warner, B.B., and Dantas, G. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. Nat Microbiol *1*, 1–10.

Gifford, S.M., Zhao, L., Stemple, B., DeLong, K., Medeiros, P.M., Seim, H., and Marchetti, A. (2020). Microbial Niche Diversification in the Galápagos Archipelago and Its Response to El Niño. Front. Microbiol. *11*.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., and Stokes, H.W. (2008). The Evolution of Class 1 Integrons and the Rise of Antibiotic Resistance. J Bacteriol *190*, 5095–5100.

Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M., and Zhu, Y.-G. (2015). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. ISME J *9*, 1269–1279.

Gingrich, E.N., Scorza, A.V., Clifford, E.L., Olea-Popelka, F.J., and Lappin, M.R. (2010). Intestinal parasites of dogs on the Galapagos Islands. Veterinary Parasitology *169*, 404–407.

Girlich, D., Naas, T., and Nordmann, P. (2004). Biochemical Characterization of the Naturally Occurring Oxacillinase OXA-50 of Pseudomonas aeruginosa. Antimicrob Agents Chemother *48*, 2043–2048.

González, J., Montes, C., Rodríguez, J., and Tapia, W. (2008). Ecology and Society: Rethinking the Galapagos Islands as a Complex Social-Ecological System: Implications for Conservation and Management. Ecology and Society *13*.

Grube, A.M., Stewart, J.R., and Ochoa-Herrera, V. (2020). The challenge of achieving safely managed drinking water supply on San Cristobal island, Galápagos. International Journal of Hygiene and Environmental Health *228*, 113547.

Guenther, S., Ewers, C., and Wieler, L.H. (2011). Extended-Spectrum Beta-Lactamases Producing E. coli in Wildlife, yet Another Form of Environmental Pollution? Front Microbiol *2*.

Guo, B., Liu, C., Gibson, C., and Frigon, D. (2019). Wastewater microbial community structure and functional traits change over short timescales. Science of The Total Environment *662*, 779–785.

Gupta, S., Arango-Argoty, G., Zhang, L., Pruden, A., and Vikesland, P. (2019). Identification of discriminatory antibiotic resistance genes among environmental resistomes using extremely randomized tree algorithm. Microbiome *7*, 123.

Heinemann, J.A. (1999). How antibiotics cause antibiotic resistance. Drug Discovery Today *4*, 72–79.

Hong, P.-Y., Mao, Y., Ortiz-Kofoed, S., Shah, R., Cann, I., and Mackie, R.I. (2015). Metagenomic-Based Study of the Phylogenetic and Functional Gene Diversity in Galápagos Land and Marine Iguanas. Microb Ecol *69*, 444–456.

Hong, P.-Y., Wheeler, E., Cann, I.K.O., and Mackie, R.I. (2011). Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. The ISME Journal *5*, 1461–1470.

Hong, P.-Y., Yannarell, A.C., Dai, Q., Ekizoglu, M., and Mackie, R.I. (2013). Monitoring the Perturbation of Soil and Groundwater Microbial Communities Due to Pig Production Activities. Appl. Environ. Microbiol. *79*, 2620–2629.

Hopkins, K.L., Deheer-Graham, A., Threlfall, E.J., Batchelor, M.J., and Liebana, E. (2006). Novel plasmid-mediated CTX-M-8 subgroup extended-spectrum beta-lactamase (CTX-M-40) isolated in the UK. Int J Antimicrob Agents *27*, 572–575.

Hultman, J., Tamminen, M., Pärnänen, K., Cairns, J., Karkman, A., and Virta, M. (2018). Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. FEMS Microbiol Ecol *94*.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res *17*, 377–386.

Ibekwe, A.M., Leddy, M.B., Bold, R.M., and Graves, A.K. (2012). Bacterial community composition in low-flowing river water with different sources of pollutants. FEMS Microbiol Ecol *79*, 155–166.

Instituto Nacional de Estadísticas y Censos (2015). Principales Resultados Censo de Población y Viviendo Galápagos 2015

Iredell, J., Brown, J., and Tagg, K. (2016). Antibiotic resistance in Enterobacteriaceae: mechanisms and clinical implications. BMJ: British Medical Journal *352*.

Jacobs, C., Huang, L.J., Bartowsky, E., Normark, S., and Park, J.T. (1994). Bacterial cell wall recycling provides cytosolic muropeptides as effectors for beta-lactamase induction. EMBO J. *13*, 4684–4694.

Jansson, D.S., and Pringle, M. (2011). Antimicrobial susceptibility of Brachyspira spp. isolated from commercial laying hens and free-living wild mallards (Anas platyrhynchos). Avian Pathol *40*, 387–393.

Kanamaru, K., Kanamaru, K., Tatsuno, I., Tobe, T., and Sasakawa, C. (2000). SdiA, an Escherichia coli homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohaemorrhagic Escherichia coli O157:H7. Mol Microbiol *38*, 805–816.

Karkman, A., Johnson, T.A., Lyra, C., Stedtfeld, R.D., Tamminen, M., Tiedje, J.M., and Virta, M. (2016). High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. FEMS Microbiol. Ecol. *92*.

Karkman, A., Pärnänen, K., and Larsson, D.G.J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. Nature Communications *10*, 80.

Kasuya, M. (1964). Transfer of drug resistance between enteric bacteria induced in the mouse intestine. J Bacteriol *88*, 322–328.

Kent, A.G., Vill, A.C., Shi, Q., Satlin, M.J., and Brito, I.L. (2020). Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. Nat Commun *11*, 4379.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res *35*, W43-46.

Kim, H.-S., Nagore, D., and Nikaido, H. (2010). Multidrug Efflux Pump MdtBC of Escherichia coli Is Active Only as a B2C Heterotrimer. J Bacteriol *192*, 1377–1386.

Koczura, R., Mokracka, J., Taraszewska, A., and 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. Microb Ecol *72*, 909–916.

Kotlarska, E., Łuczkiewicz, A., Pisowacka, M., and Burzyński, A. (2015). Antibiotic resistance and prevalence of class 1 and 2 integrons in Escherichia coli isolated from two wastewater treatment plants, and their receiving waters (Gulf of Gdansk, Baltic Sea, Poland). Environ Sci Pollut Res Int *22*, 2018–2030.

Kristensen, J.M., Nierychlo, M., Albertsen, M., and Nielsen, P.H. (2020). Bacteria from the Genus Arcobacter Are Abundant in Effluent from Wastewater Treatment Plants. Appl Environ Microbiol *86*, e03044-19.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol *35*, 1547–1549.

Kurokawa, H., Shibata, N., Doi, Y., Shibayama, K., Kamachi, K., Yagi, T., and Arakawa, Y. (2003). A new TEM-derived extended-spectrum beta-lactamase (TEM-91) with an R164C substitution at the omega-loop confers ceftazidime resistance. Antimicrob Agents Chemother *47*, 2981–2983.

Labbate M., Chowdhury P. Roy, and Stokes H. W. (2008). A Class 1 Integron Present in a Human Commensal Has a Hybrid Transposition Module Compared to Tn402: Evidence of Interaction with Mobile DNA from Natural Environments. Journal of Bacteriology *190*, 5318–5327.

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

Lanza, V.F., Tedim, A.P., Martínez, J.L., Baquero, F., and Coque, T.M. (2015). The Plasmidome of Firmicutes: Impact on the Emergence and the Spread of Resistance to Antimicrobials. Microbiol Spectr *3*, PLAS-0039-2014.

Larsson, D.G.J., Andremont, A., Bengtsson-Palme, J., Brandt, K.K., de Roda Husman, A.M., Fagerstedt, P., Fick, J., Flach, C.-F., Gaze, W.H., Kuroda, M., et al. (2018). Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. Environment International *117*, 132–138.

Lartigue, M.-F., Poirel, L., Aubert, D., and Nordmann, P. (2006). In Vitro Analysis of ISEcp1B-Mediated Mobilization of Naturally Occurring β-Lactamase Gene blaCTX-M of Kluyvera ascorbata. Antimicrobial Agents and Chemotherapy *50*, 1282–1286.

Laursen, M.F., Dalgaard, M.D., and Bahl, M.I. (2017). Genomic GC-Content Affects the Accuracy of 16S rRNA Gene Sequencing Based Microbial Profiling due to PCR Bias. Front Microbiol *8*, 1934.

Lee, K., Kim, D.-W., Lee, D.-H., Kim, Y.-S., Bu, J.-H., Cha, J.-H., Thawng, C.N., Hwang, E.-M., Seong, H.J., Sul, W.J., et al. (2020). Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. Microbiome *8*, 2.

Leverstein-van Hall, M.A., Fluit, A.C., Paauw, A., Box, A.T.A., Brisse, S., and Verhoef, J. (2002). Evaluation of the Etest ESBL and the BD Phoenix, VITEK 1, and VITEK 2 automated instruments for detection of extended-spectrum beta-lactamases in multiresistant Escherichia coli and Klebsiella spp. J Clin Microbiol *40*, 3703–3711.

Levy, J.K., Crawford, P.C., Lappin, M.R., Dubovi, E.J., Levy, M.G., Alleman, R., Tucker, S.J., and Clifford, E.L. (2008). Infectious Diseases of Dogs and Cats on Isabela Island, Galapagos. Journal of Veterinary Internal Medicine *22*, 60–65.

Lewbart, G.A., Ulloa, C., Deresienski, D., Regalado, C., Muñoz-Pérez, J.-P., Garcia, J., Hardesty, B.D., and Valle, C.A. (2017). Health status of red-footed boobies (sula sula) determined by hematology, biochemistry, blood gases, and physical examination. J Zoo Wildl Med *48*, 1230–1233.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., and Gordon, J.I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol *6*, 776–788.

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., and Zhang, T. (2015). Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J *9*, 2490–2502.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, X., Stokholm, J., Brejnrod, A., Vestergaard, G.A., Russel, J., Trivedi, U., Thorsen, J., Gupta, S., Hjelmsø, M.H., Shah, S.A., et al. (2021). The infant gut resistome associates with E. coli, environmental exposures, gut microbiome maturity, and asthma-associated bacterial composition. Cell Host & Microbe *29*, 975-987.e4.

Li, X., Stokholm, J., Brejnrod, A., Vestergaard, G.A., Russel, J., Trivedi, U., Thorsen, J., Gupta, S., Hjelmsø, M.H., Shah, S.A., et al. (2021). The infant gut resistome associates with E. coli, environmental exposures, gut microbiome maturity, and asthma-associated bacterial composition. Cell Host & Microbe *29*, 975-987.e4.

Liakopoulos, A., Mevius, D., and Ceccarelli, D. (2016). A Review of SHV Extended-Spectrum β-Lactamases: Neglected Yet Ubiquitous. Front. Microbiol. *7*.

Liao, N., Borges, C.A., Rubin, J., Hu, Y., Ramirez, H.A., Chen, J., Zhou, B., Zhang, Y., Zhang, R., Jiang, J., et al. (2020). Prevalence of β-Lactam Drug-Resistance Genes in Escherichia coli Contaminating Ready-to-Eat Lettuce. Foodborne Pathog Dis *17*, 739–742.

Literak, I., Dolejska, M., Radimersky, T., Klimes, J., Friedman, M., Aarestrup, F.M., Hasman, H., and Cizek, A. (2010). Antimicrobial-resistant faecal Escherichia coli in wild mammals in central Europe: multiresistant Escherichia coli producing extended-spectrum beta-lactamases in wild boars. J Appl Microbiol *108*, 1702–1711.

Literak, I., Manga, I., Wojczulanis-Jakubas, K., Chroma, M., Jamborova, I., Dobiasova, H., Sedlakova, M.H., and Cizek, A. (2014). Enterobacter cloacae with a novel variant of ACT AmpC beta-lactamase originating from glaucous gull (Larus hyperboreus) in Svalbard. Veterinary Microbiology *171*, 432–435.

Liu, B., and Pop, M. (2009). ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res *37*, D443-447.

Liu, C.M., Stegger, M., Aziz, M., Johnson, T.J., Waits, K., Nordstrom, L., Gauld, L., Weaver, B., Rolland, D., Statham, S., et al. (2018). Escherichia coli ST131-H22 as a Foodborne Uropathogen. MBio *9*.

Livermore, D.M., Day, M., Cleary, P., Hopkins, K.L., Toleman, M.A., Wareham, D.W., Wiuff, C., Doumith, M., and Woodford, N. (2019). OXA-1 β-lactamase and non-susceptibility to penicillin/β-lactamase inhibitor combinations among ESBL-producing Escherichia coli. J Antimicrob Chemother *74*, 326–333.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Ma, X., Monroe, B.P., Cleaton, J.M., Orciari, L.A., Li, Y., Kirby, J.D., Chipman, R.B., Petersen, B.W., Wallace, R.M., and Blanton, J.D. (2018). Rabies surveillance in the United States during 2017. Journal of the American Veterinary Medical Association *253*, 1555–1568.

Mak, S., Xu, Y., and Nodwell, J.R. (2014). The expression of antibiotic resistance genes in antibiotic-producing bacteria. Molecular Microbiology *93*, 391–402.

Makowska, N., Koczura, R., and Mokracka, J. (2016). Class 1 integrase, sulfonamide and tetracycline resistance genes in wastewater treatment plant and surface water. Chemosphere *144*, 1665–1673.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., et al. (2013). The comprehensive antibiotic resistance database. Antimicrob Agents Chemother *57*, 3348–3357.

McDougall, F., Boardman, W., Gillings, M., and Power, M. (2019). Bats as reservoirs of antibiotic resistance determinants: A survey of class 1 integrons in Grey-headed Flying Foxes (Pteropus poliocephalus). Infection, Genetics and Evolution *70*, 107–113.

McInnes, R.S., McCallum, G.E., Lamberte, L.E., and van Schaik, W. (2020). Horizontal transfer of antibiotic resistance genes in the human gut microbiome. Curr Opin Microbiol *53*, 35–43.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE *8*, e61217.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics *9*, 386.

Mortimer-Jones, S.M., Phillips, N.D., La, T., Naresh, R., and Hampson, D.J. (2008). Penicillin resistance in the intestinal spirochaete Brachyspira pilosicoli associated with OXA-136 and OXA-137, two new variants of the class D beta-lactamase OXA-63. J Med Microbiol *57*, 1122–1128.

Munck, C., Albertsen, M., Telke, A., Ellabaan, M., Nielsen, P.H., and Sommer, M.O.A. (2015). Limited dissemination of the wastewater treatment plant core resistome. Nature Communications *6*, 8452.

Mwangi, W., Figueiredo, P. de, and Criscitiello, M.F. (2016). One Health: Addressing Global Challenges at the Nexus of Human, Animal, and Environmental Health. PLOS Pathogens *12*, e1005731.

Nadkarni, M.A., Martin, F.E., Jacques, N.A., and Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology (Reading) *148*, 257–266.

Nesme, J., Cécillon, S., Delmont, T.O., Monier, J.-M., Vogel, T.M., and Simonet, P. (2014). Large-Scale Metagenomic-Based Study of Antibiotic Resistance in the Environment. Current Biology *24*, 1096–1100.

Ng, C., Tay, M., Tan, B., Le, T.-H., Haller, L., Chen, H., Koh, T.H., Barkham, T.M.S., Thompson, J.R., and Gin, K.Y.-H. (2017). Characterization of Metagenomes in Urban Aquatic Compartments Reveals High Prevalence of Clinically Relevant Antibiotic Resistance Genes in Wastewaters. Front Microbiol *8*.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet *12*, 443–451.

Nieto-Claudin, A., Deem, S.L., Rodríguez, C., Cano, S., Moity, N., Cabrera, F., and Esperón, F. (2021). Antimicrobial resistance in Galapagos tortoises as an indicator of the growing human footprint. Environmental Pollution *284*, 117453.

Nieto-Claudin, A., Esperón, F., Blake, S., & Deem, S. L. (2019). Antimicrobial resistance genes present in the faecal microbiota of free-living Galapagos tortoises (Chelonoidis porteri). *Zoonoses and Public Health*, *66*(8), 900–908.

Nonaka, L., and Suzuki, S. (2002). New Mg2+-dependent oxytetracycline resistance determinant tet 34 in Vibrio isolates from marine fish intestinal contents. Antimicrob Agents Chemother *46*, 1550–1552.

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

O'Neill, I., Schofield, Z., and Hall, L.J. (2017). Exploring the role of the microbiota member Bifidobacterium in modulating immune-linked diseases. Emerg Top Life Sci *1*, 333–349.

O'Neill, J. (2016). Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. *The Review on Antimicrobial Resistance.*

Oh, M., Pruden, A., Chen, C., Heath, L.S., Xia, K., and Zhang, L. (2018). MetaCompare: a computational pipeline for prioritizing environmental resistome risk. FEMS Microbiol. Ecol. *94*.

Oki, K., Akiyama, T., Matsuda, K., Gawad, A., Makino, H., Ishikawa, E., Oishi, K., Kushiro, A., and Fujimoto, J. (2018). Long-term colonization exceeding six years from early infancy of Bifidobacterium longum subsp. longum in human gut. BMC Microbiol *18*, 209.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., GL Simpson, G.L, Solymos, P., Stevens, H.H., Szoecs, E., Wagner, H. vegan: Community Ecology Package. R package version 2.5–6. 2019

Oravcova, V., Janecko, N., Ansorge, A., Masarikova, M., and Literak, I. (2014). First record of vancomycin-resistant Enterococcus faecium in Canadian wildlife. Environ Microbiol Rep *6*, 210–211.

Oravcova, V., Svec, P., and Literak, I. (2017). Vancomycin-resistant enterococci with vanA and vanB genes in Australian gulls. Environ Microbiol Rep *9*, 316–318.

Overbey, K.N., Hatcher, S.M., and Stewart, J.R. (2015). Water quality and antibiotic resistance at beaches of the Galápagos Islands. Front. Environ. Sci 64.

Pace, A., Dipineto, L., Fioretti, A., and Hochscheid, S. (2019). Loggerhead sea turtles as sentinels in the western Mediterranean: antibiotic resistance and environment-related modifications of Gram-negative bacteria. Marine Pollution Bulletin *149*, 110575.

Pacheco-Sandoval, A., Schramm, Y., Heckel, G., Brassea-Pérez, E., Martínez-Porchas, M., and Lago-Lestón, A. (2019). The Pacific harbor seal gut microbiota in Mexico: Its relationship with diet and functional inferences. PLoS One *14*, e0221770.

Páez-Rosas, D., Torres, J., Espinoza, E., Marchetti, A., Seim, H., and Riofrío-Lazo, M. (2021). Declines and recovery in endangered Galapagos pinnipeds during the El Niño event. Sci Rep *11*, 8785.

Paradis, E., et al. ape: Analyses of Phylogentetics and Evolution. R package version 5.5. 2021.

Pärnänen, K., Karkman, A., Hultman, J., Lyra, C., Bengtsson-Palme, J., Larsson, D.G.J., Rautava, S., Isolauri, E., Salminen, S., Kumar, H., et al. (2018). Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. Nature Communications *9*, 1–11.

Partridge, S.R. (2015). Resistance mechanisms in Enterobacteriaceae. Pathology *47*, 276–284.

Partridge, S.R., Kwong, S.M., Firth, N., and Jensen, S.O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev *31*, e00088-17.

Peña, C.G.-D. la, Garduño-Niño, E., Vaca-Paniagua, F., Díaz-Velásquez, C., Barrows, C.W., Gomez-Gil, B., and Valenzuela-Núñez, L.M. (2019). Comparison of the fecal bacterial microbiota composition between wild and captive bolson tortoises (Gopherus flavomarginatus). Herpetological Conservation and Biology *14*, 587–600.

Perilli, M., Felici, A., Franceschini, N., De Santis, A., Pagani, L., Luzzaro, F., Oratore, A., Rossolini, G.M., Knox, J.R., and Amicosante, G. (1997). Characterization of a new TEM-derived beta-lactamase produced in a Serratia marcescens strain. Antimicrob Agents Chemother *41*, 2374–2382.

Petrella, S., Clermont, D., Casin, I., Jarlier, V., and Sougakoff, W. (2001). Novel Class A β-Lactamase Sed-1 from Citrobacter sedlakii: Genetic Diversity of β-Lactamases within the Citrobacter Genus. Antimicrob Agents Chemother *45*, 2287–2298.

Petrella, S., Pernot, L., and Sougakoff, W. (2004). Crystallization and preliminary X-ray diffraction study of the class A beta-lactamase SED-1 and its mutant SED-G238C from Citrobacter sedlakii. Acta Crystallogr D Biol Crystallogr *60*, 125–128.

Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., and Keller, M. (2007). Targeted Access to the Genomes of Low-Abundance Organisms in Complex Microbial Communities. Appl Environ Microbiol *73*, 3205–3214.

Poirel, L., Figueiredo, S., Cattoir, V., Carattoli, A., and Nordmann, P. (2008). Acinetobacter radioresistens as a Silent Source of Carbapenem Resistance for Acinetobacter spp. Antimicrobial Agents and Chemotherapy *52*, 1252–1256.

Poirel, L., Naas, T., and Nordmann, P. (2010). Diversity, epidemiology, and genetics of class D beta-lactamases. Antimicrob Agents Chemother *54*, 24–38.

Poole, K. (2005). Aminoglycoside Resistance in Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy *49*, 479–487.

Poole, K. (2005). Efflux-mediated antimicrobial resistance. J. Antimicrob. Chemother. *56*, 20–51.

Pornruangwong, S., Hendriksen, R.S., Pulsrikarn, C., Bangstrakulnonth, A., Mikoleit, M., Davies, R.H., Aarestrup, F.M., and Garcia-Migura, L. (2011). Epidemiological investigation of Salmonella enterica serovar Kedougou in Thailand. Foodborne Pathog Dis *8*, 203–211.

Post, V., Recchia, G.D., and Hall, R.M. (2007). Detection of Gene Cassettes in Tn402-Like Class 1 Integrons. Antimicrobial Agents and Chemotherapy *51*, 3467–3468.

Potron, A., Poirel, L., and Nordmann, P. (2011). Origin of OXA-181, an emerging carbapenem-hydrolyzing oxacillinase, as a chromosomal gene in Shewanella xiamenensis. Antimicrob. Agents Chemother. *55*, 4405–4407.

Quan, J., Langelier, C., Kuchta, A., Batson, J., Teyssier, N., Lyden, A., Caldera, S., McGeever, A., Dimitrov, B., King, R., et al. (2019). FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. Nucleic Acids Res. *47*, e83.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590-596.

Rausch, P., Rühlemann, M., Hermes, B.M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., von Frieling, J., Hentschel, U., et al. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. Microbiome *7*, 133.

Rehman, Z.U., Fortunato, L., Cheng, T., and Leiknes, T. (2020). Metagenomic analysis of sludge and early-stage biofilm communities of a submerged membrane bioreactor. Science of The Total Environment *701*, 134682.

Rice, L.B. (2006). Antimicrobial Resistance in Gram-Positive Bacteria. The American Journal of Medicine *119*, S11–S19.

Robinson, D.A., Kearns, A.M., Holmes, A., Morrison, D., Grundmann, H., Edwards, G., O'Brien, F.G., Tenover, F.C., McDougal, L.K., Monk, A.B., et al. (2005). Re-emergence of early pandemic Staphylococcus aureus as a community-acquired meticillin-resistant clone. The Lancet *365*, 1256–1258.

Robinson, T.P., Bu, D.P., Carrique-Mas, J., Fèvre, E.M., Gilbert, M., Grace, D., Hay, S.I., Jiwakanon, J., Kakkar, M., Kariuki, S., et al. (2016). Antibiotic resistance is the quintessential One Health issue. Trans R Soc Trop Med Hyg *110*, 377–380.

Sandri, C., Correa, F., Spiezio, C., Trevisi, P., Luise, D., Modesto, M., Remy, S., Muzungaile, M.-M., Checcucci, A., Zaborra, C.A., et al. (2020). Fecal Microbiota Characterization of Seychelles Giant Tortoises (Aldabrachelys gigantea) Living in Both Wild and Controlled Environments. Front. Microbiol. *11*.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864.

Scott, K.P., Melville, C.M., Barbosa, T.M., and Flint, H.J. (2000). Occurrence of the New Tetracycline Resistance Gene tet(W) in Bacteria from the Human Gut. Antimicrob Agents Chemother *44*, 775–777.

Scott, L.C., Lee, N., and Aw, T.G. (2020). Antibiotic Resistance in Minimally Human-Impacted Environments. Int J Environ Res Public Health *17*, 3939.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods *9*, 811–814.

Seminati, C., Mejia, W., Mateu, E., and Martin, M. (2005). Mutations in the quinolone-resistance determining region (QRDR) of Salmonella strains isolated from pigs in Spain. Veterinary Microbiology *106*, 297–301.

Shao, Y., Forster, S.C., Tsaliki, E., Vervier, K., Strang, A., Simpson, N., Kumar, N., Stares, M.D., Rodger, A., Brocklehurst, P., et al. (2019). Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. Nature *574*, 117–121.

Simões, R.R., Poirel, L., Da Costa, P.M., and Nordmann, P. (2010). Seagulls and beaches as reservoirs for multidrug-resistant Escherichia coli. Emerging Infect. Dis. *16*, 110–112.

Skurnik, D., Le Menac'h, A., Zurakowski, D., Mazel, D., Courvalin, P., Denamur, E., Andremont, A., and Ruimy, R. (2005). Integron-Associated Antibiotic Resistance and Phylogenetic Grouping of Escherichia coli Isolates from Healthy Subjects Free of Recent Antibiotic Exposure. Antimicrob Agents Chemother *49*, 3062–3065.

Skurnik, D., Ruimy, R., Andremont, A., Amorin, C., Rouquet, P., Picard, B., and Denamur, E. (2006). Effect of human vicinity on antimicrobial resistance and integrons in animal faecal Escherichia coli. J Antimicrob Chemother *57*, 1215–1219.

Smalla, K., Cook, K., Djordjevic, S.P., Klümper, U., and Gillings, M. (2018). Environmental dimensions of antibiotic resistance: assessment of basic science gaps. FEMS Microbiol. Ecol. *94*.

Smith, D.L., Dushoff, J., and Morris, J.G. (2005). Agricultural Antibiotics and Human Health. PLoS Med *2*.

Soberón-Chávez, G., Alcaraz, L.D., Morales, E., Ponce-Soto, G.Y., and Servín-González, L. (2017). The Transcriptional Regulators of the CRP Family Regulate Different Essential Bacterial Functions and Can Be Inherited Vertically and Horizontally. Front. Microbiol. *8*.

Song, C., Zhang, C., Fan, L., Qiu, L., Wu, W., Meng, S., Hu, G., Kamira, B., and Chen, J. (2016). Occurrence of antibiotics and their impacts to primary productivity in fishponds around Tai Lake, China. Chemosphere *161*, 127–135.

Song, K., Li, L., and Zhang, G. (2016). Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. Scientific Reports *6*, 35736.

Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A.W., Brito, I.L., Weitz, D.A., Pitkänen, L.K., Vigneault, F., Virta, M.Pj., et al. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. The ISME Journal *10*, 427–436.

Stokes, H.W., Nesbø, C.L., Holley, M., Bahl, M.I., Gillings, M.R., and Boucher, Y. (2006). Class 1 Integrons Potentially Predating the Association with Tn402-Like Transposition Genes Are Present in a Sediment Microbial Community. J Bacteriol *188*, 5722–5730.

Taft, D.H., Liu, J., Maldonado-Gomez, M.X., Akre, S., Huda, M.N., Ahmad, S.M., Stephensen, C.B., and Mills, D.A. (2018). Bifidobacterial Dominance of the Gut in Early Life and Acquisition of Antimicrobial Resistance. MSphere *3*, e00441-18.

Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R. (2007). Initiation of actinorhodin export in Streptomyces coelicolor. Mol. Microbiol. *63*, 951–961.

Tao, W., Zhang, X.-X., Zhao, F., Huang, K., Ma, H., Wang, Z., Ye, L., and Ren, H. (2016). High Levels of Antibiotic Resistance Genes and Their Correlations with Bacterial Community and Mobile Genetic Elements in Pharmaceutical Wastewater Treatment Bioreactors. PLOS ONE *11*, e0156854.

Tapiainen, T., Koivusaari, P., Brinkac, L., Lorenzi, H.A., Salo, J., Renko, M., Pruikkonen, H., Pokka, T., Li, W., Nelson, K., et al. (2019). Impact of intrapartum and postnatal antibiotics on the gut microbiome and emergence of antimicrobial resistance in infants. Sci Rep *9*, 10635.

Tavío, M.M., Aquili, V.D., Poveda, J.B., Antunes, N.T., Sánchez-Céspedes, J., and Vila, J. (2010). Quorum-sensing regulator sdiA and marA overexpression is involved in in vitro-selected multidrug resistance of Escherichia coli. J Antimicrob Chemother *65*, 1178–1186.

Teo, J.W.P., Tan, T.M.C., and Poh, C.L. (2002). Genetic Determinants of Tetracycline Resistance in Vibrio harveyi. Antimicrob Agents Chemother *46*, 1038–1045.

Thaller, M.C., Migliore, L., Marquez, C., Tapia, W., Cedeño, V., Rossolini, G.M., and Gentile, G. (2010). Tracking Acquired Antibiotic Resistance in Commensal Bacteria of Galápagos Land Iguanas: No Man, No Resistance. PLOS ONE *5*, e8989.

Thomasen, C.A., Mullen, N., Belden, L.K., May, M., and Hawley, D.M. (2017). Resident Microbiome Disruption with Antibiotics Enhances Virulence of a Colonizing Pathogen. Scientific Reports *7*, 16177.

Thompson, A.L., Houck, K.M., and Jahnke, J.R. (2019). Pathways linking Caesarean delivery to early health in a dual burden context: immune development and the gut microbiome in infants and children from Galápagos, Ecuador. Am J Hum Biol e23219.

Tiedje, J.M., Wang, F., Manaia, C.M., Virta, M., Sheng, H., Ma, L., Zhang, T., and Topp, E. (2019). Antibiotic Resistance Genes in the Human-Impacted Environment: A One Health Perspective. Pedosphere *29*, 273–282.

Vaz-Moreira, I., Nunes, O.C., and Manaia, C.M. (2017). Ubiquitous and persistent Proteobacteria and other Gram-negative bacteria in drinking water. Sci Total Environ *586*, 1141–1149.

Venables, W.N., and Ripley, B.D. (2002). Generalized Linear Models. In Modern Applied Statistics with S, W.N. Venables, and B.D. Ripley, eds. (New York, NY: Springer), pp. 183–210.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science *304*, 66–74.

Villacis, B., and Carrillo, D. (2013). The Socioeconomic Paradox of Galapagos. In Science and Conservation in the Galapagos Islands: Frameworks & Perspectives, S.J. Walsh, and C.F. Mena, eds. (New York, NY: Springer New York), pp. 69–85.

Vittecoq, M., Godreuil, S., Prugnolle, F., Durand, P., Brazier, L., Renaud, N., Arnal, A., Aberkane, S., Jean-Pierre, H., Gauthier-Clerc, M., et al. (2016). Antimicrobial resistance in wildlife. Journal of Applied Ecology *53*, 519–529.

Waldron, L.S., and Gillings, M.R. (2015). Screening Foodstuffs for Class 1 Integrons and Gene Cassettes. J Vis Exp e52889.

Walsh, S.J., and Mena, C.F. (2013). Perspectives for the Study of the Galapagos Islands: Complex Systems and Human–Environment Interactions. In Science and Conservation in the Galapagos Islands: Frameworks & Perspectives, S.J. Walsh, and C.F. Mena, eds. (New York, NY: Springer), pp. 49–67.

Walsh, S.J., McCleary, A.L., Heumann, B.W., Brewington, L., Raczkowski, E.J., and Mena, C.F. (2010). Community Expansion and Infrastructure Development: Implications for Human Health and Environmental Quality in the Galápagos Islands of Ecuador. Journal of Latin American Geography *9*, 137–159.

Walsh, T.R., Weeks, J., Livermore, D.M., and Toleman, M.A. (2011). Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. The Lancet Infectious Diseases *11*, 355–362.

Walther-Rasmussen, J., and Høiby, N. (2006). OXA-type carbapenemases. J Antimicrob Chemother *57*, 373–383.

Wang, F.-H., Qiao, M., Su, J.-Q., Chen, Z., Zhou, X., and Zhu, Y.-G. (2014). High Throughput Profiling of Antibiotic Resistance Genes in Urban Park Soils with Reclaimed Water Irrigation. Environ. Sci. Technol. *48*, 9079–9085.

Wang, Q., Wang, P., and Yang, Q. (2018). Occurrence and diversity of antibiotic resistance in untreated hospital wastewater. Science of The Total Environment *621*, 990–999.

Wang, X., Gu, J., Gao, H., Qian, X., and Li, H. (2018). Abundances of Clinically Relevant Antibiotic Resistance Genes and Bacterial Community Diversity in the Weihe River, China. Int J Environ Res Public Health *15*.

Watkins, G., Cruz, F. (2007). Galapagos at Risk: A Socioeconomic Analysis. Puerto Ayora, Santa Cruz: Charles Darwin Research Foundation.

Wheeler, E., Hong, P.-Y., Bedon, L.C., and Mackie, R.I. (2012). Carriage of antibiotic-resistant enteric bacteria varies among sites in Galapagos reptiles. J. Wildl. Dis. *48*, 56–67.

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

World Health Organization (2017). WHO publishes list of bacteria for which new antibiotics are urgently needed. **https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed**

World Health Organization (2021). WHO integrated global surveillance on ESBL-producing E. coli using a "One Health" approach: implementation and opportunities. Geneva, Switzerland. Licence: CC BY-NC-SA 3.0 IGO.

Wright, G.D. (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Micro *5*, 175–186.

Xu, Y., Sim, S.-H., Song, S., Piao, S., Kim, H.-M., Jin, X.L., Lee, K., and Ha, N.-C. (2010). The tip region of the MacA alpha-hairpin is important for the binding to TolC to the Escherichia coli MacAB-TolC pump. Biochem Biophys Res Commun *394*, 962–965.

Yang, Y., Zhang, A.-N., Che, Y., Liu, L., Deng, Y., and Zhang, T. (2021). Underrepresented high diversity of class 1 integrons in the environment uncovered by PacBio sequencing using a new primer. Science of The Total Environment *787*, 147611.

Yin, X., Deng, Y., Ma, L., Wang, Y., Chan, L.Y.L., and Zhang, T. (2019). Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. Environment International *133*, 105270.

Yin, X., Jiang, X.-T., Chai, B., Li, L., Yang, Y., Cole, J.R., Tiedje, J.M., and Zhang, T. (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.

Yong, D., Toleman, M.A., Giske, C.G., Cho, H.S., Sundman, K., Lee, K., and Walsh, T.R. (2009). Characterization of a New Metallo-β-Lactamase Gene, blaNDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic Structure in Klebsiella pneumoniae Sequence Type 14 from India. Antimicrob Agents Chemother *53*, 5046–5054.

Yuan, M.L., Dean, S.H., Longo, A.V., Rothermel, B.B., Tuberville, T.D., and Zamudio, K.R. (2015). Kinship, inbreeding and fine-scale spatial structure influence gut microbiota in a hindgut-fermenting tortoise. Mol Ecol *24*, 2521–2536.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., and Larsen, M.V. (2012). Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother *67*, 2640–2644.

Zgurskaya, H.I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. Front Microbiol *2*.

Zhang, A.N., Li, L.-G., Ma, L., Gillings, M.R., Tiedje, J.M., and Zhang, T. (2018). Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection. Microbiome *6*, 130.

Zhang, T., Zhang, X.-X., and Ye, L. (2011). Plasmid Metagenome Reveals High Levels of Antibiotic Resistance Genes and Mobile Genetic Elements in Activated Sludge. PLOS ONE *6*, e26041.

Zhao, R., Feng, J., Yin, X., Liu, J., Fu, W., Berendonk, T.U., Zhang, T., Li, X., and Li, B. (2018). Antibiotic resistome in landfill leachate from different cities of China deciphered by metagenomic analysis. Water Res. *134*, 126–139.

Zheng, W., Huyan, J., Tian, Z., Zhang, Y., and Wen, X. (2020). 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. Environment International *135*, 105372.