MULTI-OMIC ANALYSIS OF MICROBIAL COMMUNITY ECOLOGY IN THE HUMAN GUT AND IN A HUMAN-IMPACTED STREAM

by

CORALIS DEL MAR RODRIGUEZ GARCIA

(Under the Direction of Elizabeth A. Ottesen)

ABSTRACT

Antibiotic resistance is a global public health threat, resulting in increasing numbers of infections that do not respond to conventional antibiotic therapies. While early studies focused primarily on hospital acquitted infections, it is becoming increasingly clear that commensal and environmental bacteria are playing a role in the spread of antibiotic resistance. When the human gut microbiome is expose to antibiotics, this selects for resistance in diverse bacteria that can then be transmitted to other community members through fecal-oral transmission or through the environment through numerous pathways. One such method of transmission is passage via wastewater into freshwater environments used for farm irrigation and recreation. As a result, microbial communities in surface water are essential to understand in order to predict transmission patterns and mitigate the spread of antibiotic resistance. This work presents three investigations spanning the human-environment continuum. In Chapter 2, we present a study of ESBL *Enterobacteriaceae* in the human gut, in which we found that 4.5% of the southeastern US population that participated in the study, asymptomatically carried ESBL-E bacteria and that 64% of those remain colonized after 3 months. In Chapter 3, we present an undergraduate teaching module to isolate and characterize ESBL-E from freshwater streams where students

develop microbiology skills and improve awareness of the antibiotic resistance in the environment. In Chapter 4, we present a study examining how the metabolic activity of bacterioplankton predict their abundance or success downstream the creek. Together, the work presented here aims to advance the understanding of microbial community ecology and the ecology of antibiotic resistance in the freshwater environments and the human gut.

INDEX WORDS: microbiome, ESBL, metagenomic, metatranscriptomics, single cell genomes, antibiotic resistance, microbial community, microbial activity

MULTI-OMIC ANALYSIS OF MICROBIAL COMMUNITY ECOLOGY IN THE HUMAN GUT AND IN A HUMAN-IMPACTED STREAM

by

CORALIS DEL MAR RODRIGUEZ GARCIA

B.S., University of Puerto Rico-Mayagüez, 2013

M.S., University of Puerto Rico-Mayagüez, 2016

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

MULTI-OMIC ANALYSIS OF MICROBIAL COMMUNITY ECOLOGY IN THE HUMAN GUT AND IN A HUMAN-IMPACTED STREAM

by

CORALIS DEL MAR RODRIGUEZ GARCIA

Major Professor:	Elizabeth A. Ottesen
Committee:	Vincent J. Starai
	Erin K. Lipp
	Jonathan G. Frye

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2024

DEDICATION

To my husband and best friend, Joshuam: por tu apoyo incondicional para alcanzar mis metas, por celebrar cada victoria y cada fracaso de este proceso, por sostenerme y mantenerme a flote cuando sola no podía y sobretodo por tu dedicación a nuestra familia para darme el espacio para llegar al final de este camino.

To my mom, Lesvia: por creerme capaz de lograrlo todo y no dejar que me rindiera nunca.

To my dad, Jaime: por ser mi primera inspiración científica y mi modelo a seguir.

To my sister and soulmate, Rubí: por escucharme y entenderme aún sin palabras y por siempre estar de mi lado.

Finally, **to my twin daughters Amalia and Laila**: aunque llegaron tarde en este capítulo fueron la fuerza y motivación que necesitaba para terminarlo. Esta y todas mis metas serán siempre para ustedes.

ACKNOWLEDGEMENTS

To my mentor and advisor, Liz, I am incredible grateful for all your support through this journey. You were there for me first as a committee member, then as an advisor and you have been a wonderful mentor in my scientific and professional journey and even as a twin mom! It has been an honor to grow these years by your side and I will always be grateful for your mentoring approach and guidance.

To my committee members, Vinny, Erin and Jonathan, for always being available to answer my questions and your feedback and advice to improve through these years. You are inspiring scientists, and it has been an honor to share this journey with all of you. I appreciate your kindness and support in the scientific and professional area. I would also like to thank Dr. Eric Stabb, for your mentoring and guidance during the first years of this degree. A heartful thanks to a previous advisor that continues mentoring and showing up for me even after my M.S. degree under his supervision, Dr. Rafael Montalvo.

To my community that supported me through these years and were my strength in this process, I feel so grateful to have you by my side and I appreciate your unconditional encouragement. First and most to my husband who has been there in every win and loss of this process and his support was essential in achieving this goal. Special thanks to my family, my in-laws, my lab mates: Helen Dukes, Rachel Dockman, Amber DePoy and Kevin Riedmuller, Katie Bellissimo, Macey Coppinger and to my friends: Devyn Rohlfs, Rafael Payán, Mara Cuebas, Lilliam Penchi, Melanie Rosado, Alexandra Medina, Joel Clavell, Abby Ward and Jamila Pham. You were a source of joy, fun and love in this journey.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Relevance

"Fresh water resources are most precious to life on earth and in actual fact water is the elixir to life" (1). Water as a vital resource is involved in daily lifestyle on multiple ways, including as a source of drinking water and recreational water, and as a recipient of wastes and runoff from wastewater treatment plants, manufacturing facilities and farming. The connections of these uses through the environment result in exchange of myriad chemical and biological contaminants. A critical contaminant to freshwater environments, specially from wastewater treatment plants, healthcare facilities and farming, are antibiotics residues and antibiotic resistant bacteria. The presence of antibiotics, ARGs and AR bacteria in widely-use freshwater environments is a public health and ecological issue (2) and a reason why the OneHealth initiative recognizes the environment as one of the critical components of antimicrobial resistance (AMR) worldwide (3, 4). As a potential connector of antimicrobial resistance to and from human, agricultural and industrial practices, it is crucial to better understand the dynamics of freshwater environments in order to present efficient strategies to mitigate the AMR issue. Studies incorporating analysis of antibiotic resistance that incorporate data from human, animals and environmental samples are needed to add a cohesive perspective on how each of these factors influence AMR flow. Similarly, education and awareness of antibiotic resistance in the environment is essential to alleviate human contributions thar increase AMR prevalence. The dissertation presented here elaborates on the topics of antibiotic resistance prevalence in the community and environment by presenting a study

of ESBL-*Enterobacteriaceae* carriage among asymptomatic humans as well as teaching module for upper-level undergraduate students on awareness of antibiotic resistance in freshwater environments. I also present work on the microbial community ecology of freshwater environments in the same community, which is not directly related to antibiotic resistance but lays the foundation to better understand the fate of antibiotic resistant pathogens within freshwater environments. This introductory chapter seeks to highlight how these topics are related and the relevance of considering human-environment connections when aiming to advance our understanding of antimicrobial resistance.

Antibiotic Resistance

Antibiotic resistance occurs ubiquitously in any microbial population through spontaneous mutations. In the presence of an antibiotic pressure, bacteria with mutations that confer resistance are selected for, resulting in an increased abundance of antibiotic resistant bacteria in the population. Many organisms naturally produce antimicrobial molecules that target competitors, directly or as a byproduct (5, 6). Some of those secondary metabolites can act as antibiotics (7, 8). When exposed to naturally produced antimicrobials bacteria evolve strategies to resist their antimicrobial mechanism and those strategies could then be used against therapeutic or artificially produced antibiotics, resulting in an antibiotic resistant bacterium without prior exposure to that antibiotic (9).

One of the most medically important antibiotic resistance mechanisms are the beta lactamases that act against beta-lactam antibiotics. Beta-lactam antibiotics target enzymes involved in peptidoglycan synthesis, interrupting cell wall formation. Beta-lactamases cleave the beta-lactam rings in commonly used antibiotics like penicillin, monobactams and cephalosporins. Beta-lactamase genes are naturally found in Gram-positive and Gram-negative pathogens like

Aeromonas, Acinetobacter, Burkholderia and *Pseudomonas, Staphylococcus aureus, Enterococcus, and Streptococcus pneumoniae* (10). More specifically, Gram-negative bacteria from the *Enterobacteriaceae* family like *Escherichia coli* and *Klebsiella pneumoniae* are commonly associated with beta-lactamases genes encoding for extended-spectrum beta-lactamases (ESBLs), carried in plasmids and mobile elements (11, 12). ESBLs are medically relevant as the causing agent of many infection complications. Part of my work focus on this group of beta-lactamases that especially relevant in the study of antibiotic resistance transmission and their rapid global spread.

Extended Spectrum Beta Lactamases (ESBLs)

ESBLs are defined by their ability to cleave multiple beta-lactams from the penicillin and cephalosporin groups including ampicillin, cefotaxime, ceftazidime, ceftriaxone, cephalothin, and occasionally cefepime, carbapenems or other newly develop drugs, resulting in ineffective treatment with first-line antibiotics and requiring more advanced and costly treatments (13). ESBLs are divided into the subtypes SHV, TEM, CTX, OXA, PER, GES and unusual subtypes like VEB-1 and BES-1, based on their genetic mutations and preferences in beta-lactams to hydrolyze. TEM enzymes were among the first plasmid-mediated beta-lactamases recognized, and prior to the emergence of CTX type enzymes were the most prevalent ESBL worldwide (14, 15). The ability of TEM beta-lactamase to be encoded in plasmids allowed the rapid dissemination of this enzyme, creating the first beta-lactamase epidemic. In recent years, the prevalence of TEM ESBLs has been replaced by CTX-M.

CTX-M enzymes are one of the most diverse extended-spectrum beta-lactamase subtypes. Their variation is connected to the mobilization of the genes, more than a collection of mutations. As they become mobile in insertion elements and plasmids, the continuous transfer between different

species created a variety of clusters (15). That diversity increased their spectrum of activity reaching more beta lactam antibiotics. Their association with mobile elements like IS6 and plasmids like IncF, that are known by their highly efficient transmission, present an ideal opportunity for CTX-M enzymes to disperse. On top of that, when these mobile elements are carried by highly disseminated clones, it creates the perfect storm for a CTX-M pandemic (16). In fact, CTX-M-15 is commonly found in the *E. coli* serotype ST131, a globally found isolate which has been considered a key player in the dissemination of ESBLs (17). The combination of versatility and mobility allowed CTX-M enzymes to become the most relevant and concerning ESBLs today.

A cornerstone of the spread of ESBLs has been high rates of community-acquired infections. Community-acquired ESBL infections occur in individuals without recent inpatient hospital exposure. This is a challenge, as many efforts to control the spread of antibiotic resistance have focused on hospital-acquired infections. Over the last few decades, community-acquired infections with ESBLs have escalated worldwide (18–20). CTX-M enzymes were found in more than 90% of ESBL-producing bacteria from community-acquired infections suggesting a key role of these enzymes in the community spread (21). In fact, a study by Lob et al. (22) found that the proportion of ESBL-*E. coli* isolates was lower in hospital-acquired UTI than in community-acquired UTIs (22).

ESBLs in the human gut

A major unanswered question are the sources of community acquired ESBL infections. A key candidate is the commensal gut microbiome. It's been shown that many of these infections are caused by extraintestinal pathogenic *E. coli* (23). Extra-intestinal pathogenic infections are often caused by *E. coli* that were initially in the patient's own gut and were transmitted to a new location

in the body (24). As a result, the presence of ESBL genes in commensal gut bacteria can play an important role in shaping infection rates. A recent meta-analysis reported that the global carriage of intestinal ESBL-*E.coli* has increased by 10-fold in the last 20 years (25). In fact, a recent metanalysis estimated the worldwide, pooled prevalence of ESBL-E colonization could be as high as 14% of individuals (14).

Even non-pathogenic isolates present in the commensal gut microbiome can play a role in the dissemination of antibiotic resistance when AR genes are carried in mobile genetic elements such as plasmids. Stephens et al. (26) found that around 80% of antibiotic resistance genes in commensal *E. coli* isolates from rectal swabs of healthy students were carried in plasmids. Similarly, Wang et al. (27) identified and tested plasmid replicons carrying ESBL genes in ESBL-E isolates from healthy human samples which supports the propagation of these genes by fecal carriage of commensal flora. The fact that those ESBL are typically carried in plasmids are a major concern as a source of community-acquired infections allowing the opportunity of a genetic exchange with possibly pathogenic bacteria.

ESBLs in the environment and in freshwater

Even though ESBLs are frequently found in the human gut, this does not fully explain their rapid dissemination and transmission between humans. Research on this topic suggest that there are environmental vectors contributing to antibiotic resistance dispersion. Iseppi et al. (28) investigated genes, antibiotic resistance patterns and genotypes in *Enterococcus* and *Enterobacteriaceae* strains isolated from fecal samples of humans, dogs and cats, and found that all the samples carried at least one ESBL or AmpC gene, highlighting the prevalence of antibiotic resistance in these sources. Similarly, ESBLs genes are frequently found in farm animals used for human consumption. The presence of ESBL genes in multiple environments demonstrate their

widespread through the community. It can be hypothesized that an important factor contributing to this rapid spread is their presence in surface waters, which serves as a link between the different environments. The work presented here focus on the freshwater environment as a source and/or vector of ESBL spread.

Antibiotic resistant (AR) bacteria in freshwater.

As a rendezvous place of treated sewage, farming activities, soil runoff and human activities, streams have been hypothesized to serve a critical role in the dispersion and exchange of antibiotic resistance genes and antibiotic resistant bacteria (24). Bacterial groups carrying multi-drug resistance enzymes like extended-spectrum beta-lactamases have been identified in freshwater environments globally including environmental water in Japan (29), surface water in Switzerland (30), a river basin in China (31), a river in Spain (32), recreational waters of Norway (33), farm waste and canals in Thailand (34), and locally in surface water in Georgia, USA (35). In our local (Athens, GA) watershed, Cho et al. (2023) reported high levels of ESBL-E (17.3 %) and CRE (7.7%) in surface water samples, while 41.7% of the wastewater treatment plant effluent samples had detectable ESBL-E and CRE (36). The presence of this pathogens in freshwater, represents a transmission risk regardless of their abundance levels in this environment. Chapter 3 discusses a curriculum module that provides training to microbiology undergraduates in key tools and techniques while creating student awareness of human-environment connections to the antibiotic resistance in the environment.

Even without external inputs from neighboring activities (farming, healthcare facilities, recreation) freshwater environments can contribute to the development of antibiotic resistance and that can later be transfer to bacteria outside this environment. Antibiotics can be found in water and soil environments at subinhibitory concentrations without specific providing sources (37–39). These

low concentrations contribute to the selection of antibiotic resistance bacteria and the evolution of previously sensitive bacteria becoming resistant. As a result, native freshwater populations can be vectors and carriers of ARGs. A study of microbes in a large river by Wang et al. (40) identified 1853 antibiotic-resistant bacterial species belonging to 22 phyla, while genera like *Limnohabitans*, *Acidimicrobium*, *Methylotenera* and *Flavobacterium* were the most abundant ARG hosts. Similarly, on a review on the methods to identify ARG hosts in the environment, Rice et al. (41) reported *Enterobacteriaceae*, *Comamonadaceae* and *Moraxellaceae* as frequent dominant hosts, most of them specific for freshwater environments, except for *Enterobacteriaceae*. As the mechanisms that freshwater bacteria use to become resistant to antibiotics are often similar to those of clinical isolates, although differences have been observed like the frequently use of RND efflux pumps by water isolates (42, 43). This demonstrates the importance of considering a range of taxa when investigating this topic.

In addition to antibiotic resistance developed by pelagic bacterial communities in streams, aquatic animals and practices associated with them could also be sources of antibiotic resistance genes. A study of aquaculture ponds in China identified a high abundance of antibiotics and antibiotic resistance genes in this setting (44). Similarly, a metagenomic study of freshwater shrimp's guts and aquaculture environments found ARGs, mostly for efflux pumps and target modification, significantly correlated with mobile genetic elements and with the genera *Aeromonas*, *Yersinia* and *Clostridium* (45).

The movement of extracellular DNA through streams may also play a role in AR dissemination. Mao et al. (46) found high concentrations of *sul1*, *sul2*, *tetW* and *tetT* antibiotic resistance genes in freshwater sediment, especially as extracellular DNA. Similarly, Dong et al. (47) found that

extracellular DNA in sludge samples from different sources carry a higher number of ARGs and more importantly, that the extracellular DNA was preferentially associated to competent cells. Environmental conditions like rainfall have also been shown to influence the presence of ARG and AR bacteria in rivers by shifting the microbial community and increasing the abundance of ARGs from dislodged soil and runoff from areas around the streams (48, 49). The complexity of sources of antibiotic resistance in streams connects this environment to the antibiotic resistance threat in more than one way and emphasizes the urgency of a better understanding of its dynamics.

Wastewater and the human-environment connection

It is increasingly clear that a major source of ESBL isolates in water may well be the human commensal gut microbiome. While it is reassuring to think that modern wastewater conveyance and treatment infrastructure eliminated such transmission, any system of that size cannot fully eliminate the human-environment connection. Wastewater treatment plants receive a direct input of ARGs like ESBL genes by the sewage collection. Treatment at these facilities is more focused on the removal of bacteria than genes associated with them. Multiple studies have identified high abundance of ESBL genes in wastewater samples suggest that wastewater treatment plants are contributors to the antibiotic resistance in the environment by providing ideal conditions between antibiotics and bacteria and increasing the chances of genetic exchange (50–52). The use of recycled water from effluent of wastewater treatment plants, used for irrigation or urban parks and agricultural practices has been shown to increase the abundance of ARGs and its potential transfer into the environment (53–55). In addition, untreated sewage can reach surface waters through the presence of leaks in wastewater transmission infrastructure. A recent study conducted locally reported high abundance (73%) of ARGs in water samples that was correlated with wastewater transmission infrastructure, suggesting an underexplored source for antibiotic resistance

contamination (56). Similarly, another study compared the microbial communities and antibiotic resistance from wastewater plant and its receiving lake and determined that the wastewater plant is not contributing to the antibiotic resistance gene pool in the lake, but that sewage was disposed in the lake through other sources (57). Overall, the wastewater conveyance and treatment system represents a significant link between the carriage of ESBL genes in humans and freshwater environments contributing to the ARG spread.

Mobile elements in freshwater

The acquisition of ARGs by freshwater bacteria that can then be transmitted to humans during exposure to recreational water represents a health concern primarily when these genes are carried on mobile genetic elements that could allow them to be transferred to commensal or pathogenic bacteria in the human gut. Characterized mobile genetic elements in water and sediment co-existing with multiple ARGs that provide resistance to different antibiotic classes such as quinolones and macrolides, suggest the impact of mobile elements in this environment (58).

Plasmids represent a key mediator for transfer of antibiotic resistance genes between freshwater bacteria, human pathogens and commensals. Plasmids are a major source in ARGs dissemination and a key factor in the success of pandemic clones. In multidrug resistant isolates from recreational waters in Athens, GA, antibiotic resistance genes were associated with specific replicon types (59). Similarly, Mao et al. (2014), identified ARGs in plasmids for over 20 weeks, from eDNA in sediment and water mesocosms, while chromosomally-encoded genes like 16S rRNA were not detected, which suggest a longer persistence of ARG in plasmids in river sediments (46).

The guts of stream-dwelling animals could also play a role in antibiotic resistance gene exchange. Fu et al. (2017) performed a ARG transfer model using zebrafish that showed the hindgut of those animals as a key region for gene exchange by upregulating the mRNA expression of regulatory

genes for the self-transmissible RP4-plasmid (60). In this model, 15% of fecal bacteria from the zebrafish obtained ARGs from the conjugal transfer of plasmid RP4. The model presented in this data informs of complex possibilities to take into consideration about ARGs exchange, including the gut microbiota members of aquatic animals and the similarities to the gut microbiota of humans, how bacterial groups from both environments can interact and a demonstration of these vents happening in streams environments.

Bacteriophages represent another commonly known mechanism for gene exchange. With a transduction rate between 10^{-8} to 10^{-5} transductions/PFU in freshwater settings it is expected to consider phages as a key driver of antibiotic resistance exchange (61). Although freshwater bacteriophages are mostly studied in the context of taxonomy classification, Moon et al. (62) identified polymyxin, multidrug efflux proteins and beta-lactamases in viral metagenomes from urban surface water. The same genes were also found in bacterial metagenomes from the same samples indicating the ARGs carriage by actively infecting phages. Similarly, Colombo et al. (63) examined the viromes of bacteriophages of a community exposed to tetracycline, compared to a community without exposure and showed an increased in ARG abundance in viromes from the microbial community exposed to tetracycline resistance genes compared to the viromes not exposed, suggesting the potential transfer of these genes to aquatic bacteria. Phage-like particles called gene transfer agents (GTAs) could also be important players in the dissemination of antibiotic resistance genes (64, 65). Although transduction mechanisms for gene exchange are vastly known, this exchange mechanisms in the context of freshwater bacterial communities are not always considered. The knowledge about bacteriophages and the studies connecting these vectors to ARGs carriage demonstrate that they are a relevant source of ARGs exchange in streams.

A mobilization hierarchy proposed by Cantón et al. (15) suggest that insertion sequences are used for diffusion of ARGs, in this case CTX-M. Integrons contribute to the maintenance of ARGs in the bacteria, while transposons and plasmids with integrons participate in the spread of those genes (15). It's been shown that freshwater isolates like *Aeromonas* can carry a class I integron with a similar frequency as *Enterobacteriaceae* members and that integron could carry beta-lactamase resistance genes (66). Studies performed in lakes in China, found a significant correlation between int1 and the antibiotic resistance genes *sul1*, *sul2*, bla_{TEM}, *tet, tetC* and *tetX* (47, 58). Similarly, Lin et al. (2016) (67) reported the presence of class I integrons in 79% of the strains, from eels and aquaculture ponds. Together, these mechanisms suggest that freshwater microbial communities could play a major role in the evolution and spread of antibiotic resistance genes.

Freshwater microbial communities

To comprehend the contribution of freshwater environments to the spread of antibiotic resistance, it is essential to understand microbial dynamics in this environment. Chapter 4 of this thesis is not specifically focused on antibiotic resistance but presents my work on microbial community assembly in freshwater systems. Freshwater communities are inhabited by a diverse microbial community. This community has a crucial role in shaping the spread of antibiotic resistance genes through the environment because they can become recipients and/or sources of AR genes through horizontal gene transfer and because they can occupy a niche interfering with colonization of AR bacteria in that environment. As an example, *E. coli* have been shown to survive longer in water environments when there are limited members of the river-associated taxa (68, 69). The introduction and persistence of opportunistic pathogens in freshwater environments, not only present a risk for ARG exchange but can also alter the dynamics of the community by displacing

important members playing key roles in the system. As a result, in order to understand antibiotic resistance spread through freshwater it is imperative to understand its microbial community.

Community Assembly

Riverine microbial communities perform essential environmental functions (70, 71). The success of these functions depends on a rigorous organization of members derived from different habitats within this environment. Freshwater habitats include the sediment-associated groups, biofilm groups, microbiota of aquatic flora and fauna and pelagic microbes including free-floating and motile bacteria. Gweon et al. (72) hypothesize that the migration of individuals through the water column and the stability of the different habitats within the stream favors heterogeneous microbial communities. This heterogeneity allows for a differentiated set of functions but also for competition for the abundant resources in one habitat but scarce in other. The niche separation and possible competition between the groups in multiple habitats creates a complex scenario to achieve community assembly, even without the addition of external factors, but this assembly dynamics are essential to comprehend a detailed picture of this communities.

Multiple studies have focus on the factors affecting microbial community assembly, although there is still limited knowledge about the precise conditions that rule those dynamics (73–77). Some studies argue that geographical distances and site location has a stronger impact on microbial assembly on freshwater streams (78, 79), while others argue that microbial community assembly and bacterial dynamics are driven by environmental stressors like redox gradients and nutrient availability (80–83). Lear et al. (84) demonstrate that microbial communities follow the same trends as macroorganisms when it comes to biogeography. These arguments are not necessarily independent from one another, as nutrients and chemical gradients are associated with the site conditions and the activities around that site with possibly discharge or overflow into near streams.

When investigating the microbial communities of streams, there should be a context of the surrounding of that stream as well as the conditions in its headwaters. Microbial communities at streams begin with headwater streams and soil bacteria, as well as bacteria from groundwater (85, 86). Teachey et al. (87) showed that stream ecosystems assemble rapidly at headwaters and diversity is decreased downstream. With that in mind, a snapshot of a downstream site to understand microbial community dynamics might miss relevant taxa in the assembly and succession process. Stadler and del Giorgio (88) used laboratory mesocosms to characterize the total and reactive (defined by RNA activity) taxa on different locations through an aquatic network and identified that reactive taxa was originated in soil or soil-water, highlighting the relevance upstream history when trying to understand community assembly. In a study from artic freshwater, comparing microbial taxa from soil, headwaters and receiving lake in Alaska found that 58% of bacteria and 43% of archaea found in upstream environments were also found in the lake, although there was a shift of dominant groups (85).

Human behavior and the built environment have major impacts on stream microbial communities. A study by Simonin et al. (89) investigated the changes in microbial assembly based on urbanization levels, showing that taxa in the family *Comamonadaceae* has the highest number of positive responders to the urban impact. They also showed that alpha diversity was not affected by the impacts, although the community structure changed with land cover. Similarly, Hosen et al. (90) identified the genus *Polynucleobacter* as a human associated taxa increased in compositional shifts from urbanization and the genus *Gallionella*, associated with corrosion of water systems. McLellan (91) describes the bacterial taxa associated with urban sources that represent prospective sources for microbial communities found in freshwater environments including *Rhodobacter* and *Zooglea* in treated effluent, *Enterobacteriaceae* in stormwater, and *Bacteroidaceae* and

Lachnospiraceae in combined sewer overflow. This demonstrate that urbanization and human impacts around freshwater environments are not only connected to a source of antibiotic resistance genes but as a driver of differences in main groups of microbial communities.

In the last decades, different groups have characterized a "core" freshwater bacterial community (92–96). Crump et al. (97) were one of the first groups to establish a resident group and a transient group of microbes in freshwater environments, based on their continued presence in the population and their similarity to other isolates commonly found in freshwater environments. Others have built on this model, with Fortunato et al. (2013) identifying key taxa that reassembled based on seasonal conditions with different environments creating a niche for dominant bacterial populations (98). Borton et al. (99) recently defined a core river microbiome based on genomes sequencing, contributing to the more commonly used strategies of 16S rRNA gene and metagenomics. These dominant taxa are suggested to play a role in driving the community assembly in each environment.

Members of *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Proteobacteria* and *Verrucomicrobia* are generally considered typical freshwater taxa (96, 100). Among the *Protetobacteria*, Betaproteobacteria are prominent in freshwater environments. Common motile freshwater bacteria include well-studied groups like *Aeromonas*, an opportunistic pathogen and prospective connection between freshwater and humans. The genus *Aeromonas* has also been reported as an important vector of ARG in water (101). Non-motile freshwater microbiome includes groups like *Limnohabitans*, *Flavobacterium*, *Polynucleobacter*. *Limnohabitans* and *Polynucleobacter* groups have shown abundance patterns associated with high and low pH, respectively (102). More specifically the *Polynucleobacter* genus encompass a species with genome diversification and great ecologic variability (103). Most of the studies characterized communities by focusing on

abundant or freshwater-associated taxa but to understand the community assembly process there is a need to consider the roles of low-abundant taxa.

As many studies have focused on "core" and "resident" taxa, key questions remain regarding the roles and functional activity of that transient taxa while in the environment. Yang et al. (104) found that abundant freshwater taxa had more interactions with rare taxa than with other abundant groups, which presents similar dynamics between microbial communities in other environments and emphasizes the roles of rare taxa in assembly and stability of this environment (38, 104–106). This transient group becomes even more important when they can represent ARG contributors into this environment or participate in genomic exchange with the native freshwater community.

Summary

The antibiotic resistance threat is one of many layers connecting human, animals, environment, and communities. Although antibiotic stewardship improvements are a big step in the right direction to overcome this public health issue, it is imperative to increase education of this topic to promote better uses and management strategies of freshwater resources. An optimized and conscious control of antibiotic use could reduce the antibiotic load in the environment and the rise of the AR bacterial populations. As shown here community-acquired infections are a growing contributor to this problem and the focus on this source can provide insight into predicting patterns of exchange and ecology of AR strains. Similarly, education at small and large scale will improve the understanding of antibiotic resistance spread and increase awareness of this issue. The work presented here provides deeper understanding to each of these topics with the aim of contributing to the growing body of work in these areas but more importantly, interconnecting them.

References

1. Shafi S, Kamili AN, Shah MA, Parray JA, Bandh SA. 2017. Aquatic bacterial diversity: Magnitude, dynamics, and controlling factors. Microb Pathog https://doi.org/10.1016/j.micpath.2017.01.016.

2. Samreen, Ahmad I, Malak HA, Abulreesh HH. 2021. Environmental antimicrobial resistance and its drivers: a potential threat to public health. J Glob Antimicrob Resist. Elsevier Ltd https://doi.org/10.1016/j.jgar.2021.08.001.

3. Velazquez-Meza ME, Galarde-López M, Carrillo-Quiróz B, Alpuche-Aranda CM. 2022. Antimicrobial resistance: One Health approach. Vet World 743–749.

4. WHO. 2014. Antimicrobial resistance. Global report on surveillance. World Health Organization 61.

5. Linares JF, Gustafsson I, Baquero F, Martinez JL. 2006. Antibiotics as intermicrobial signaling agents instead of weapons.

6. Fajardo A, Martínez-Martín N, Mercadillo M, Galán JC, Ghysels B, Matthijs S, Cornelis P, Wiehlmann L, Tümmler B, Baquero F, Martínez JL. 2008. The neglected intrinsic resistome of bacterial pathogens. PLoS One 3.

7. Terkina IA, Parfenova V V., Ahn TS. 2006. Antagonistic activity of actinomycetes of lake Baikal. Appl Biochem Microbiol 42.

8. Barros J, Becerra J, González C, Martínez M. 2013. Antibacterial metabolites synthesized by psychrotrophic bacteria isolated from cold-freshwater environments. Folia Microbiol (Praha) 58.

9. Alm EW, Zimbler D, Callahan E, Plomaritis E. 2014. Patterns and persistence of antibiotic resistance in faecal indicator bacteria from freshwater recreational beaches. J Appl Microbiol 117.

10. Bush K, Bradford PA. 2020. Epidemiology of B-Lactamase-Producing Pathogens. Clin Microbiol Rev 33:1–37.

11. Carattoli A. 2009. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother 53:2227–2238.

12. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. Journal of Antimicrobial Chemotherapy 73:1121–1137.

13. Shaikh S, Fatima J, Shakil S, Rizvi SMD, Kamal MA. 2015. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. Saudi J Biol Sci 22:90–101.

14. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E. 2016. Fecal Colonization with Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors among Healthy Individuals: A Systematic Review and Metaanalysis. Clinical Infectious Diseases 63:310–318.

15. Cantón R, González-Alba JM, Galán JC. 2012. CTX-M enzymes: Origin and diffusion. Front Microbiol 3.

16. Cantón R, Coque TM. 2006. The CTX-M β-lactamase pandemic. Curr Opin Microbiol 9:466–475.

17. Peirano G, Richardson D, Nigrin J, McGeer A, Loo V, Toye B, Alfa M, Pienaar C, Kibsey P, Pitout JDD. 2010. High prevalence of ST131 isolates producing CTX-M-15 and CTX-

M-14 among extended-spectrum-β-lactamase-producing Escherichia coli isolates from Canada. Antimicrob Agents Chemother 54:1327–1330.

18. Bezabih YM, Bezabih A, Dion M, Batard E, Teka S, Obole A, Dessalegn N, Enyew A, Roujeinikova A, Alamneh E, Mirkazemi C, Peterson GM, Bezabhe WM. 2022. Comparison of the global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli between healthcare and community settings: A systematic review and meta-Analysis. JAC Antimicrob Resist 4.

19. Doi Y, Adams J, O'Keefe A, Quereshi Z, Ewan L, Paterson DL. 2007. Community-acquired extended spectrum B-lactamase producers, United States.

20. CDC. 2019. Antibiotic Resistance Threats in The United States, 2019U.S. Department of Health and Human Services.

21. Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. Clin Microbiol Rev 26:744–758.

22. Lob SH, Nicolle LE, Hoban DJ, Kazmierczak KM, Badal RE, Sahm DF. 2016. Susceptibility patterns and ESBL rates of Escherichia coli from urinary tract infections in Canada and the United States, SMART 2010–2014. Diagn Microbiol Infect Dis 85:459–465.

23. Duffy N, Karlsson M, Reses HE, Campbell D, Daniels J, Stanton RA, Janelle SJ, Schutz K, Bamberg W, Rebolledo PA, Bower C, Blakney R, Jacob JT, Phipps EC, Flores KG, Dumyati G, Kopin H, Tsay R, Kainer MA, Muleta D, Byrd-Warner B, Grass JE, Lutgring JD, Rasheed JK, Elkins CA, Magill SS, See I. 2022. Epidemiology of extended-spectrum β-lactamase-producing Enterobacterales in five US sites participating in the Emerging Infections Program, 2017. Infect Control Hosp Epidemiol 43:1586–1594.

24. Escobar-Páramo P, Grenet K, Le Menac'h A, Rode L, Salgado E, Amorin C, Gouriou S, Picard B, Rahimy MC, Andremont A, Denamur E, Ruimy R. 2004. Large-scale population structure of human commensal Escherichia coli isolates. Appl Environ Microbiol 70.

25. Bezabih YM, Sabiiti W, Alamneh E, Bezabih A, Peterson GM, Bezabhe WM, Roujeinikova A. 2021. The global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli in the community. Journal of Antimicrobial Chemotherapy 76:22–29.

26. Stephens C, Arismendi T, Wright M, Hartman A, Gonzalez A, Gill M, Pandori M, Hess D. 2020. F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal Escherichia coli. mSphere 5.

27. Wang J, Stephan R, Karczmarczyk M, Yan Q, Hächler H, Fanning S. 2013. Molecular characterization of blaESBL-harboring conjugative plasmids identified in multi-drug resistant Escherichia coli isolated from food-producing animals and healthy humans. Front Microbiol 4:1–9.

28. Iseppi R, Di Cerbo A, Messi P, Sabia C. 2020. Antibiotic resistance and virulence traits in vancomycin-resistant enterococci (Vre) and extended-spectrum β-lactamase/ampc-producing (ESBL/ampc) enterobacteriaceae from humans and pets. Antibiotics 9.

29. Miyagi K, Hirai I, Sano K. 2016. Distribution of Aeromonas species in environmental water used in daily life in Okinawa Prefecture, Japan. Environ Health Prev Med 21.

30. Zurfluh K, Abgottspon H, Hac̈hler H, Nuësch-Inderbinen M, Stephan R. 2014. Quinolone resistance mechanisms among Extended- Spectrum Beta-Lactamase (ESBL) producing Escherichia coli isolated from rivers and lakes in Switzerland. PLoS One 9.

31. Chen H, Shu W, Chang X, Chen JA, Guo Y, Tan Y. 2010. The profile of antibiotics resistance and integrons of extended-spectrum beta-lactamase producing thermotolerant coliforms isolated from the Yangtze River basin in Chongqing. Environmental Pollution 158.

32. Piedra-Carrasco N, Fàbrega A, Calero-Cáceres W, Cornejo-Sánchez T, Brown-Jaque M, Mir-Cros A, Muniesa M, González-López JJ. 2017. Carbapenemase-producing enterobacteriaceae recovered from a Spanish river ecosystem. PLoS One 12.

33. Jørgensen SB, Søraas A V., Arnesen LS, Leegaard TM, Sundsfjord A, Jenum PA. 2017. A comparison of extended spectrum β-lactamase producing Escherichia coli from clinical, recreational water and wastewater samples associated in time and location. PLoS One 12:1–15.

34. Runcharoen C, Raven KE, Reuter S, Kallonen T, Paksanont S, Thammachote J, Anun S, Blane B, Parkhill J, Peacock SJ, Chantratita N. 2017. Whole genome sequencing of ESBL-producing Escherichia coli isolated from patients, farm waste and canals in Thailand. Genome Med 9.

35. Cho S, Nguyen HAT, McDonald JM, Woodley TA, Hiott LM, Barrett JB, Jackson CR, Frye JG. 2019. Genetic characterization of antimicrobial-resistant escherichia coli isolated from a mixed-use watershed in northeast Georgia, USA. Int J Environ Res Public Health 16:1–14.

36. Cho S, Hiott LM, Read QD, Damashek J, Westrich J, Edwards M, Seim RF, Glinski DA, Bateman Mcdonald JM, Ottesen EA, Lipp EK, Henderson WM, Jackson CR, Frye JG. 2023. Distribution of Antibiotic Resistance in a Mixed-Use Watershed and the Impact of Wastewater Treatment Plants on Antibiotic Resistance in Surface Water. Antibiotics 2023 12:1586.

37. Andersson DI, Hughes D. 2014. Microbiological effects of sublethal levels of antibiotics. Nat Rev Microbiol 12:465–478.

38. Martinez JL. 2009. Environmental pollution by antibiotics and by antibiotic resistance determinants. Environmental Pollution 157:2893–2902.

39. McArthur J V., Fletcher DE, Tuckfield RC, Baker-Austin C. 2016. Patterns of Multi-Antibiotic-Resistant Escherichia Coli from Streams with No History of Antimicrobial Inputs. Microb Ecol 72:840–850.

40. Wang Z, Chen Q, Zhang J, Zou Y, Huang Y, Yan H, Xu Z, Yan D, Li T, Liu C. 2023. Insights into antibiotic stewardship of lake-rivers-basin complex systems for resistance risk control. Water Res 228:119358.

41. Rice EW, Wang P, Smith AL, Stadler LB. 2020. Determining Hosts of Antibiotic Resistance Genes: A Review of Methodological Advances. Environ Sci Technol Lett 7:282–291.

42. Vaz-Moreira I, Nunes OC, Manaia CM. 2014. Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. FEMS Microbiol Rev 38:761–778.

43. Lupo A, Coyne S, Berendonk TU. 2012. Origin and Evolution of Antibiotic Resistance: The Common Mechanisms of Emergence and Spread in Water Bodies. Front Microbiol 3.

44. Xiong W, Sun Y, Zhang T, Ding X, Li Y, Wang M, Zeng Z. 2015. Antibiotics, Antibiotic Resistance Genes, and Bacterial Community Composition in Fresh Water Aquaculture Environment in China. Microb Ecol 70:425–432.

45. Zhao Y, Zhang X, Zhao Z, Duan C, Chen H, Wang M, Ren H, Yin Y, Ye L. 2018. Metagenomic analysis revealed the prevalence of antibiotic resistance genes in the gut and living environment of freshwater shrimp. J Hazard Mater 350:10–18.

46. Mao D, Luo Y, Mathieu J, Wang Q, Feng L, Mu Q, Feng C, Alvarez PJJ. 2014. Persistence of Extracellular DNA in River Sediment Facilitates Antibiotic Resistance Gene Propagation. Environ Sci Technol 48:71–78.

47. Dong P, Cui Q, Fang T, Huang Y, Wang H. 2019. Occurrence of antibiotic resistance genes and bacterial pathogens in water and sediment in urban recreational water. Journal of Environmental Sciences 77:65–74.

48. Di Cesare A, Eckert EM, Rogora M, Corno G. 2017. Rainfall increases the abundance of antibiotic resistance genes within a riverine microbial community. Environmental Pollution 226:473–478.

49. Garner E, Benitez R, von Wagoner E, Sawyer R, Schaberg E, Hession WC, Krometis L-AH, Badgley BD, Pruden A. 2017. Stormwater loadings of antibiotic resistance genes in an urban stream. Water Res 123:144–152.

50. Proia L, Anzil A, Borrego C, Farrè M, Llorca M, Sanchis J, Bogaerts P, Balcázar JL, Servais P. 2018. Occurrence and persistence of carbapenemases genes in hospital and wastewater treatment plants and propagation in the receiving river. J Hazard Mater 358:33–43.

51. Hsu J-T, Chen C-Y, Young C-W, Chao W-L, Li M-H, Liu Y-H, Lin C-M, Ying C. 2014. Prevalence of sulfonamide-resistant bacteria, resistance genes and integron-associated horizontal gene transfer in natural water bodies and soils adjacent to a swine feedlot in northern Taiwan. J Hazard Mater 277:34–43.

52. Singh R, Singh AP, Kumar S, Giri BS, Kim K-H. 2019. Antibiotic resistance in major rivers in the world: A systematic review on occurrence, emergence, and management strategies. J Clean Prod 234:1484–1505.

53. Limayem A, Wasson S, Mehta M, Pokhrel AR, Patil S, Nguyen M, Chen J, Nayak B. 2019. High-Throughput Detection of Bacterial Community and Its Drug-Resistance Profiling From Local Reclaimed Wastewater Plants. Front Cell Infect Microbiol 9.

54. Vassallo A, Kett S, Purchase D, Marvasi M. 2022. The Bacterial Urban Resistome: Recent Advances. Antibiotics 11:512.

55. O'Flaherty E, Cummins E. 2017. Antibiotic resistance in surface water ecosystems: Presence in the aquatic environment, prevention strategies, and risk assessment. Human and Ecological Risk Assessment: An International Journal 23:299–322.

56. Damashek J, Westrich JR, McDonald JMB, Teachey ME, Jackson CR, Frye JG, Lipp EK, Capps KA, Ottesen EA. 2022. Non-point source fecal contamination from aging wastewater infrastructure is a primary driver of antibiotic resistance in surface waters. Water Res 222:118853.

57. Bondarczuk K, Piotrowska-Seget Z. 2019. Microbial diversity and antibiotic resistance in a final effluent-receiving lake. Science of The Total Environment 650:2951–2961.

58. Wang Z, Han M, Li E, Liu X, Wei H, Yang C, Lu S, Ning K. 2020. Distribution of antibiotic resistance genes in an agriculturally disturbed lake in China: Their links with microbial communities, antibiotics, and water quality. J Hazard Mater 393:122426.

59. Cho S, Barrett JB, Frye JG, Jackson CR. 2020. Antimicrobial resistance gene detection and plasmid typing among multidrug resistant enterococci isolated from freshwater environment. Microorganisms 8:1–15.

60. Fu J, Yang D, Jin M, Liu W, Zhao X, Li C, Zhao T, Wang J, Gao Z, Shen Z, Qiu Z, Li J. 2017. Aquatic animals promote antibiotic resistance gene dissemination in water via conjugation:

Role of different regions within the zebra fish intestinal tract, and impact on fish intestinal microbiota. Mol Ecol 26:5318–5333.

61. Schneider CL. 2021. Bacteriophage-Mediated Horizontal Gene Transfer: Transduction, p. 151–192. *In* Harper, DR, Abedon, ST, Burrowes, BH, McConville, ML (eds.), . Springer International Publishing, Cham.

62. Moon K, Jeon JH, Kang I, Park KS, Lee K, Cha C-J, Lee SH, Cho J-C. 2020. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome 8:75.

63. Colombo S, Arioli S, Neri E, Della Scala G, Gargari G, Mora D. 2017. Viromes As Genetic Reservoir for the Microbial Communities in Aquatic Environments: A Focus on Antimicrobial-Resistance Genes. Front Microbiol 8.

64. Christensen S, Serbus LR. 2020. Gene Transfer Agents in Symbiotic Microbes, p. 25–76. *In* .

65. Rapp BJ, Wall JD. 1987. Genetic transfer in Desulfovibrio desulfuricans. Proceedings of the National Academy of Sciences 84:9128–9130.

66. Henriques IS, Fonseca F, Alves A, Saavedra MJ, Correia A. 2006. Occurrence and diversity of integrons and β-lactamase genes among ampicillin-resistant isolates from estuarine waters. Res Microbiol 157:938–947.

67. Lin M, Wu X, Yan Q, Ma Y, Huang L, Qin Y, Xu X. 2016. Incidence of antimicrobial-resistance genes and integrons in antibiotic-resistant bacteria isolated from eels and aquaculture ponds. Dis Aquat Organ 120:115–123.

68. Korajkic A, Wanjugi P, Harwood VJ. 2013. Indigenous Microbiota and Habitat Influence Escherichia coli Survival More than Sunlight in Simulated Aquatic Environments. Appl Environ Microbiol 79:5329–5337.

69. van Elsas JD, Semenov A V, Costa R, Trevors JT. 2011. Survival of Escherichia coli in the environment: fundamental and public health aspects. ISME J 5:173–183.

70. Morin S, Artigas J. 2023. Twenty Years of Research in Ecosystem Functions in Aquatic Microbial Ecotoxicology. Environ Toxicol Chem https://doi.org/10.1002/etc.5708.

71. Crump BC, Bowen JL. 2024. The Microbial Ecology of Estuarine Ecosystems. Ann Rev Mar Sci https://doi.org/10.1146/annurev-marine-022123-101845.

72. Gweon HS, Bowes MJ, Moorhouse HL, Oliver AE, Bailey MJ, Acreman MC, Read DS. 2021. Contrasting community assembly processes structure lotic bacteria metacommunities along the river continuum. Environ Microbiol 23:484–498.

73. Debray R, Herbert RA, Jaffe AL, Crits-Christoph A, Power ME, Koskella B. 2022. Priority effects in microbiome assembly. Nat Rev Microbiol https://doi.org/10.1038/s41579-021-00604-w.

74. Jackson CR, Millar JJ, Payne JT, Ochs CA. 2014. Free-living and particle-associated bacterioplankton in large rivers of the Mississippi River basin demonstrate biogeographic patterns. Appl Environ Microbiol 80.

75. Ruiz-González C, Niño-García JP, Lapierre JF, del Giorgio PA. 2015. The quality of organic matter shapes the functional biogeography of bacterioplankton across boreal freshwater ecosystems. Global Ecology and Biogeography 24.

76. Wisnoski NI, Lennon JT. 2021. Microbial community assembly in a multi-layer dendritic metacommunity. Oecologia 195.

77. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, Blaschke AP, Blöschl G, Mach RL, Kirschner AKT, Farnleitner AH, Eiler A. 2015. Bacterial diversity along a 2600km river continuum. Environ Microbiol 17:4994–5007.

78. Febria CM, Koch BJ, Palmer MA. 2015. Operationalizing an ecosystem services-based approach for managing river biodiversityWater Ecosystem Services: A Global Perspective.

79. de Paula M, da Costa TA, Silva, Soriano AAB, Lacorte GA. 2023. Spatial distribution of sediment bacterial communities from São Francisco River headwaters is influenced by human land-use activities and seasonal climate shifts. Brazilian Journal of Microbiology 54:3005–3019.

80. Wang Y, Wang Y, Shang J, Wang L, Li Y, Wang Z, Zou Y, Cai W, Wang L. 2024. Redox gradients drive microbial community assembly patterns and molecular ecological networks in the hyporheic zone of effluent-dominated rivers. Water Res 248:120900.

81. Niu X, Wang H, Wang T, Zhang P, Zhang H, Wang H, Kong X, Xie S, Xu J. 2024. The combination of multiple environmental stressors strongly alters microbial community assembly in aquatic ecosystems. J Environ Manage 350:119594.

82. Zeglin LH. 2015. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. Front Microbiol 6.

83. Adams HE, Crump BC, Kling GW. 2015. Isolating the effects of storm events on arctic aquatic bacteria: temperature, nutrients, and community composition as controls on bacterial productivity. Front Microbiol 06.

84. Lear G, Washington V, Neale M, Case B, Buckley H, Lewis G. 2013. The biogeography of stream bacteria. Global Ecology and Biogeography 22.

85. Crump BC, Amaral-Zettler LA, Kling GW. 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME Journal 6:1629–1639.

86. Sorensen JPR, Maurice L, Edwards FK, Lapworth DJ, Read DS, Allen D, Butcher AS, Newbold LK, Townsend BR, Williams PJ. 2013. Using Boreholes as Windows into Groundwater Ecosystems. PLoS One 8.

87. Teachey ME, McDonald JM, Ottesen EA. 2018. Rapid and stable microbial community assembly in the headwaters of third-order stream. Appl Environ Microbiol 85:1–15.

88. Stadler M, del Giorgio PA. 2022. Terrestrial connectivity, upstream aquatic history and seasonality shape bacterial community assembly within a large boreal aquatic network. ISME Journal 16.

89. Simonin M, Voss KA, Hassett BA, Rocca JD, Wang SY, Bier RL, Violin CR, Wright JP, Bernhardt ES. 2019. In search of microbial indicator taxa: shifts in stream bacterial communities along an urbanization gradient. Environ Microbiol 21.

90. Hosen JD, Febria CM, Crump BC, Palmer MA. 2017. Watershed Urbanization Linked to Differences in Stream Bacterial Community Composition. Front Microbiol 8.

91. McLellan SL, Fisher JC, Newton RJ. 2015. The microbiome of urban waters. International Microbiology https://doi.org/10.2436/20.1501.01.244.

92. Crump BC, Hobbie JE. 2005. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol Oceanogr 50.

93. Doherty M, Yager PL, Moran MA, Coles VJ, Fortunato CS, Krusche A V., Medeiros PM, Payet JP, Richey JE, Satinsky BM, Sawakuchi HO, Ward ND, Crump BC. 2017. Bacterial biogeography across the Amazon River-ocean continuum. Front Microbiol 8.

94. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK. 2002. Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquatic Microbial Ecology 28.

95. Warnecke F, Amann R, Pernthaler J. 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ Microbiol 6.

96. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews 75.

97. Crump BC, Kling GW, Bahr M, Hobbie JE. 2003. Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. Appl Environ Microbiol 69:2253–2268.

98. Fortunato CS, Eiler A, Herfort L, Needoba JA, Peterson TD, Crump BC. 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. ISME Journal 7:1899–1911.

99. Borton MA, McGivern BB, Willi KR, Woodcroft BJ, Mosier AC, Singleton DM, Bambakidis T, Pelly A, Liu F, Edirisinghe JN, Faria JP, Leleiwi I, Daly RA, Goldman AE, Wilkins MJ, Hall EK, Pennacchio C, Roux S, Eloe-Fadrosh E, Good SP, Sullivan MB, Henry CS, Wood-Charlson EM, Ross MRV, Miller CS, Crump BC, Stegen JC, Wrighton KC. A functional microbiome catalog crowdsourced from North American rivers.

100. Song LY, Wang YQ. 2015. Investigation of microbial community structure of a shallow lake after one season copper sulfate algaecide treatment. Microbiol Res 170.

101. Piotrowska M, Popowska M. 2014. The prevalence of antibiotic resistance genes among Aeromonas species in aquatic environments. Ann Microbiol 64:921–934.

102. Jezbera J, Jezberová J, Koll U, Horňák K, Šimek K, Hahn MW. 2012. Contrasting trends in distribution of four major planktonic betaproteobacterial groups along a pH gradient of epilimnia of 72 freshwater habitats. FEMS Microbiol Ecol 81.

103. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J. 2016. Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME Journal 10.

104. Yang Y, Cheng K, Li K, Jin Y, He X. 2022. Deciphering the diversity patterns and community assembly of rare and abundant bacterial communities in a wetland system. Science of The Total Environment 838:156334.

105. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Küsel K, Rillig MC, Rivett DW, Salles JF, Van Der Heijden MGA, Youssef NH, Zhang X, Wei Z, Hol GWH. 2017. Where less may be more: How the rare biosphere pulls ecosystems strings. ISME Journal https://doi.org/10.1038/ismej.2016.174.

106. Jiao S, Yang Y, Xu Y, Zhang J, Lu Y. 2020. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. ISME Journal 14.

CHAPTER 2:

LONG-TERM GUT COLONIZATION WITH ESBL-PRODUCING *ESCHERICHIA COLI* IN PARTICIPANTS WITHOUT KNOWN RISK FACTORS FROM THE SOUTHEASTERN UNITED STATES[1]

[1] Coralis Rodriguez Garcia, William A. Norfolk, Amanda H. Sullivan, Amanda L. Glatter, Megan S. Beaudry, Nicholas A. Mallis, Michael Welton, Travis C. Glenn, Erin K. Lipp and Elizabeth A. Ottesen. 2024. To be submitted to journal mSphere.

Abstract

We evaluated gut carriage of extended spectrum beta lactamase producing *Enterobacteriaceae* (ESBL-E) in southeastern U.S. residents without recent in-patient healthcare exposure. Study enrollment was January 2021-February 2022 in Athens, Georgia, U.S. and included a diverse population of 505 adults plus 50 child participants (age 0-5). Based on culture- screening of stool samples, 4.5% of 555 participants carried ESBL-Es. This is slightly higher than reported in studies conducted 2012-2015, which found carriage rates of 2.5-3.9% in healthy U.S. residents.

All ESBL-E confirmed isolates (n=25) were identified as *Escherichia coli*. Isolates belonged to 11 sequence types, with 48% classified as ST131. Ninety six percent of ESBL-E isolates carried a $bla_{CTX\text{-}M}$ gene. Isolated ESBL-Es frequently carried virulence genes as well as multiple classes of antibiotic resistance genes. Long-term colonization was common, with 64% of ESBL-E positive participants testing positive when rescreened three months later. One participant yielded isolates belonging to two different *E. coli* sequence types that carried $bla_{CTX\text{-}M\text{-}1}$ genes on near-identical plasmids, suggesting intra-gut plasmid transfer.

Isolation of *E. coli* on media without antibiotics revealed that ESBL-*E. coli* typically made up a minor fraction of the overall gut *E. coli* population, although in some cases they were the dominant strain. ESBL-E carriage was not associated with a significantly different stool microbiome composition. However, some microbial taxa were differentially abundant in ESBL-E carriers. Together, these results suggest that a small subpopulation of US residents are long-term, asymptomatic carriers of ESBL-Es, and may serve as an important reservoir for community spread of these ESBL genes.

Introduction

Antimicrobial resistance is a serious and growing public health threat worldwide (1, 2) that contributes to increased complication rates, as well as increased treatment costs (3, 4). Extended-spectrum beta-lactamase (ESBL) producing *Enterobacteriaceae* (ESBL-E) are listed as a serious threat by the U.S. Center for Disease Control and Prevention (CDC), contributing to 197,400 infections and 9,100 deaths in 2019 (2).

ESBL-E were first reported in 1983 and have since spread rapidly throughout the world (5–8). ESBL enzymes confer resistance to multiple antibiotics including penicillin, monobactams, and cephalosporins, commonly used to treat infections caused by Gram negative bacteria (9). There are multiple classes of ESBL enzymes including TEM, SHV, OXA and CTX-M (10–12). Recently, attention has been drawn to the CTX-M class ESBLs as it is currently the most common ESBL found worldwide (13–16). CTX-M-15 is typically associated with *E. coli* sequence type 131 (ST131), a frequent causative agent of extraintestinal infections and outbreaks (17, 18). The successful dissemination of this *E. coli* clone contributed to the wide spread of this ESBL enzyme (19). Another factor contributing to the rapid dissemination of ESBLs is their frequent association with mobile elements including plasmids (20–22). Microorganisms often carry ESBL genes on plasmids from the groups IncF, IncI1, IncA/C and IncHI2, facilitating horizontal gene transfer of antibiotic resistance (23). These plasmids also carry antibiotic resistance genes for other classes, easily resulting in multi-drug resistant organisms (24).

In a meta-analysis by Bezabih et al. (25), the global intestinal carriage rate of ESBL-*E. coli* in the community increased from 2.6% to 26.4% during 2001-2020, highlighting the significance of investigating community-associated ESBL-E. The global prevalence of ESBL-E fecal carriage in the community is highly variable throughout the globe, with substantially higher rates (76.3%) in

Tanzania and the lowest (1.9%) in Australia (25, 26). The exact causes of these geographic variations are unknown, although they have been associated to factors including the use of antibiotics in food animals and sanitation standards (27).

Few studies have examined ESBL-E carriage in healthy individuals in the U.S. Doi et al. (28, 29) reported a 3.9% prevalence of ESBL-*E. coli* in community-associated samples from outpatient clinics in 5 U.S. states collected in 2009-2010. Vento et al. (30) found that only 1 of 101 healthy U.S. military personnel based in the U.S. (May-June 2011) carried an ESBL-E. Weisenberg et al. (31), reported a 2.5% colonization rate for ESBL-E in 2012, among New York residents in participants at pre-travel or no international travel planned. Finally, Islam et al. (32) reported an ESBL-E carriage rate of 3.5% among stool samples from healthy children collected 2013-2015 in three U.S. cities. Overall, these studies suggest that carriage of ESBL-E in the U.S. is low. However, the steady increase of community-associated ESBL-E infections from 2013-2019 (2, 29) suggest a need for updated data on ESBL-E community carriage in the U.S.

The aim of this work, was an examination of the prevalence and risk factors for gut colonization of ESBL-E in the community among participants living in or near Athens, Georgia (GA), U.S. We also evaluated the frequency of long-term (~3 month) carriage of ESBL-E among positive participants. Finally, we performed an in-depth genomic analysis, including mobile elements, of ESBL-E isolated in this study and a comparison of gut microbiome community composition between carriers and non-carriers.

2.3 Materials and Methods

2.3.1 Specimen collection, metadata collection and analysis

Recruitment and consenting of participants were performed by the Clinical and Translational Research Unit (CTRU) at the University of Georgia. Inclusion criteria included: the ability and

willingness to answer an online survey regarding risk factors and to physically visit the CTRU to obtain and return the specimen collection kit, as well as age >18 years for adult participants. Adult participants with a child age ≤ 5 years in their household were invited to enroll the child in the study. Exclusion criteria included pregnancy and in-patient (overnight) hospitalization/health care in the last 12 months for reasons other than uncomplicated childbirth. Prospective participants who reported systemic (oral or intravenous) antibiotic use within the last 48 hours were asked to schedule specimen collection/drop-off for a later date, as were participants with active COVID-19 infections. Children who did not reside with participant parent for 5 days or more per week were also excluded. A signed consent was obtained from each participant or parent. To protect the confidentiality of personal data, all participants were assigned a unique, randomly generated identification number. All research activities involving human subject research were reviewed and approved by the Institutional Review Board (IRB) at University of Georgia, Athens.

Participants were provided a stool specimen container that was pre-labeled with their participant ID (Medline Industries, Cat. No. DYND36500) with a scheduled sample return date. Participants were asked to collect a stool sample using the provided collection kit and to complete an online questionnaire to collect demographic information as well as possible environmental risk factors for carriage of antibiotic resistant bacteria (see supplemental materials) as close in time as possible and no more than 12 hours before their scheduled return appointment. After collection, they were instructed to keep the stool specimen in a refrigerator or protected from heat until return to the facility. Upon return to the CTRU, stool specimens were stored at 4ºC until processing. Upon transfer to the study team, stool specimens were subdivided for processing, typically within 1-3 hours of receipt (max. 24 hours). For culture work, 200 mg of stool specimens were diluted in 1 mL of 1X PBS and immediately processed as described below. For DNA extraction and molecular

analyses, 200 mg sub-specimens were transferred to sterile cryovials and stored at -20°C until processing as described below.

Participant data was collected by online survey (Qualtrics) and matched to laboratory samples via alphanumeric participant IDs. Participant surveys consisted of a series of questions relating to each participant's demographics, lifestyle, socioeconomic status, environmental exposures, preexisting medical conditions, and travel experience. All child surveys were completed by a designated parent or guardian. Upon completion of sample collection, all data were downloaded, matched to laboratory results, and cleaned for analysis using R statistical analysis software with 'tidyverse', 'readxl', 'magrittr', 'reshape2' and 'rcompanion'. ESBL carriage frequencies and Fisher's test with Monte-Carlo simulations were used to calculate p-values using 'fisher.test' function in R. P-adjusted values were calculated using the 'p.adjust' function with the Benjamini-Hochberg procedure for False Discovery Rate (FDR) (33, 34). Statistical significance was considered from p-values less than 0.05.

2.3.2 Isolation, testing and confirmation of ESBL-E

200 mg of each stool specimen was diluted in 1 ml of 1X PBS. 100µl of the diluted aliquot was pre-enriched overnight in 5 ml tryptic soy broth (TSB) at 37ºC. 10 µl of the enriched culture was then spread onto CHROMagar™ ESBL plates and incubated at 37ºC overnight. Well-isolated colonies from CHROMagar ESBL plates were re-streaked for isolation and purification on the same medium and incubated under the same conditions. Presumptive ESBL-producing colonies were re-streaked on blood agar, identified by MALDI-TOF MS, and confirmed by antibiotic resistant profiling using Sensititre panel ESB1F at the UGA Veterinary Diagnostics Laboratory. An isolate was considered ESBL-positive if it was resistant to Cefotaxime and Ceftazidime but susceptible when clavulanic acid was added to each. Other antibiotics tested with Sensititre

included Ampicillin, Cefazolin, Cefepime, Cefoxitin, Cefpodoxime, Ceftriaxone, Cephalothin, Ciprofloxacin, Gentamicin, Imipenem, Meropenem and Piperacillin/Tazobactam. Confirmed ESBL-positive participants were requested to provide a second sample at least 90 days after the initial participation. 22 of 25 positive participants provided a second sample, the remaining 3 could not be contacted after multiple attempts. Second-visit samples were processed under the same conditions as the first samples.

2.3.3 Whole-genome sequencing and analysis

Isolated colonies from ESBL-confirmed participants were used for whole-genome analysis. Genomic DNA of isolates was extracted using Omega Biotek Bacterial DNA Kit. Purified DNA was arrayed in 96-well plates, normalized, and run on agarose minigels for QC. Genomic libraries were created using an NEBNext Ultra II FS DNA Library Prep Kit with custom primers and protocols (35) and an Opentrons OT-2 robot. Libraries for all ESBL-confirmed isolates were sequenced on a NovaSeq S4 6000 to obtain PE150 reads. Trimming and quality filtering was performed via Trimmomatic v0.39.

For long-read sequencing, DNA extracts were normalized to 50 ng per 9 µl. Oxford Nanopore Technologies (ONT) libraries were prepared following manufacturer's instructions for the rapid barcode 96 kit and cleaned with Ampure XP beads. A total of 75 µl of library was loaded onto an ONT MinION flow cell (9.3.1) for sequencing. Following sequencing, bases were called using super accuracy mode with Guppy v6.1.1.

De novo assembly was performed using Unicycler v0.4.7. using the short-reads first approach and default settings (36). A phylogenetic tree between ESBL-E genomes was constructed using maximum likelihood and core single nucleotide polymorphisms with Parsnp, with a randomly selected reference genome (AREA_ 490). Core genome single nucleotide polymorphisms (SNPs)

between isolates from first and second samples were calculated using CSI Phylogeny (37, 38). AMRFinder v3.9.8 was used to identify antimicrobial resistance genes in assembled genomes using the Plus genes database (--plus) and *E. coli* as the reference organism (-O) to identify point mutations (39). Mobile genetic elements were identified using MobileElementFinder v1.0.3 with antimicrobial resistance genes annotated (40). Tools from the Center for Genomic Epidemiology (http://www.genomicepidemiology.org/services/) that were used to analyze the hybrid assemblies including: MLST v2.0.9 with configuration for *Escherichia coli* #1 and a minimum depth of 5X; SeroTypeFinder v2.0.1 with 85% threshold ID and 60% minimum length, FimTyper v1.0 with 95% threshold ID, 95% threshold identity and 60% minimum coverage and VirulenceFinder v2.0.3 for *E. coli* with 90% threshold ID and 60% minimum length (41–44). Assemblies were classified by phylogroups using the ClermonTyping web server (http://clermontyping.iame-research.center)(45).

2.3.4 Microbial community characterization

Fecal *E. coli* were isolated without antibiotic selection from all stool samples using CHROMagar ECC. Up to 48 well-isolated colonies (if available) were re-streaked on LB plates for purity, then arrayed in 96-well plates to grow overnight with Mueller Hinton II broth media. 3 µl subsamples of the overnight culture were patched in Mueller Hinton II agar plates containing antibiotics at CLSI standards: Ampicillin (32 µg/ml), Ceftriaxone (32 µg/ml), Ciprofloxacin (4 µg/ml), Tetracycline (4 µg/ml) and Trimethoprim (8 µg/ml) to characterize the AR profile of the commensal *E. coli* community in the ESBL-E positive participants.

For 16S rRNA gene library sequencing, DNA was isolated from 200 mg of each stool sample with the Omega Biotek Stool DNA kit following the manufacturer protocol for pathogen detection (Omega Biotek, Norcross, GA, USA). The optional incubation with DS buffer and Proteinase K

was performed to improve recovery of Gram-positive bacteria. Extracted DNA was eluted in 100µl of elution buffer from the kit heated at 65ºC. DNA concentrations were measured with NanoDrop. Amplification of 16S rRNA gene was done using primers 515F and 806R as described previously (46), followed by library sequencing with Illumina MiSeq 250 x 250 bp at the Georgia Genomics Facility. Sequence analysis was performed using DADA2 (version 1.18) in R (version 4.0.2) including filtering, trimming, merge and taxonomy assignment with SILVA database, version 138 (47, 48). Data analyses were done in R using phyloseq, vegan, dplyr, tidyverse and DESeq packages with samples rarefied to a depth of 10,000 when needed (49, 50).

2.3.5 Data availability

Whole-genome assemblies and 16S rDNA sequences have been submitted to NCBI under the BioProject: PRJNA894544.

2.4 Results

We recruited a total of 555 participants including 505 adults and 50 children between January 2021 and November 2021 from the southeastern U.S. (mostly northeastern Georgia) (Fig. S2.1). Overall, the study population reflected the demographic composition of the study area as expected based on U.S. census data (Table S2.2), with some over-representation of participants who identified as white and non-Hispanic, were 18-39 years of age, and/or reported female sex. (Table S2.1).

2.4.1 Carriage of ESBL *Enterobacteriaceae* (ESBL-E)

From 555 samples, 25 participants (4.5%) were positive for stool carriage of an ESBL-E. One of the 25 ESBL-E positive participants was a child, whose parent also tested positive. Of the 25 ESBL-E positive participants, 22 provided a second sample at follow-up (97-176 days later),

including the parent-child dyad. 14 of those 22 samples (64%) were ESBL-E positive, whereas the remaining 8 were ESBL-E negative.

ESBL-E carriage was not identified as significantly associated with any of the demographic or socioeconomic risk factors examined (Table 2.1, Table S2.1). We observed slightly, but not significantly, increased incidence of ESBL-E in participants identified as biological males and Asian or Black/African American race identities. The full list of occupational, lifestyle, and environmental risk factors tested is available in supplemental table S2.1. A subset of the study population regularly experienced interaction with increased risk environments; however, these exposures were not significantly associated with ESBL-E carriage. ESBL-E carriage was also not associated with significant differences in self-reported gastrointestinal (GI) distress, urinary tract infection, or antibiotic usage. A review of the health information provided by ESBL-E positive participants did not suggest severe chronic health problems.

International travel has been widely associated with ESBL colonization (51). The proportion of ESBL-positive participants that lived internationally in the last 5 years is higher than the ESBL-negative participants (17% vs. 7%); however, it was not significantly associated with ESBL-carriage (fdr-corrected p-value 0.09). Only 8 participants (all ESBL-negative) reported international travel in the previous 30 days prior to sample collection.

2.4.2 Characteristics of ESBL-E isolates

Genome assemblies and phylogenetic relationships

All confirmed ESBL-E were identified as *Escherichia coli* by MALDI-TOF. High-quality draft genome sequences were obtained for the ESBL-E isolates of each 25 initial and 14 second visit samples (Table S2.3). Genome sizes ranged from 4,981,979 bp to 5,465,567 bp.

Isolates belonged to phylogroups B2 (17 isolates), D (4), A (2) B1 (2) and F (1). Of the B2 isolates, 12 belonged to the uropathogenic group ST131, including members with *fimH*41 and *fimH*30 alleles (Fig. S2.2). Other closely related groups, ST2279 (2 isolates), ST1193 (2) and ST636 (1) were also present (Fig. 2.1F). All the samples from second visits that yielded confirmed ESBL-E positive were near-clonal (3-55 SNPs) with the isolate of the initial visit from that participant with the sole exception of the isolates from participant 497 (Fig. 2.1E). The initial isolate from participant 497 was assigned to ST154 in phylogroup B1, whereas the isolate from this participant's second sample, 497R, was assigned to ST106 in phylogroup D (Fig. 2.1A-F). The parent-child dyad samples were both assigned to phylogroup A but had different sequence types: ST10 (parent) and ST305 (child) and 22,309 SNP differences.

Antibiotic resistance profile

All confirmed ESBL-E positive isolates were phenotypically resistant to Ampicillin, Cefotaxime, Ceftazidime, Cefpodoxime, Ceftriaxone, Cefazolin, and Cephalothin. They presented variable resistance to Cefoxitin and Cefepime as shown in Figure 2.1B. Of the 25 initial ESBL-E isolates, 18 had resistance to Ciprofloxacin (72%) and 6 were resistant to Gentamicin (24%). ESBL-E isolates from the second visit exhibited the same phenotypic antibiotic resistance profile as their original sample except for 491R, which showed a decreased resistance to Cefoxitin compared to the isolate 491. Some second-visit isolates (C483R and 487R) had resistance to Cefoxitin but were susceptible to other beta-lactams tested and were not confirmed as ESBL-E. None of the isolates were resistant to the carbapenems tested.

All isolates, except for the child sample, carried at least one $bla_{CTX\text{-}M}$ beta lactamase gene (Fig. 2.1D). The most predominant beta lactamase gene was $bla_{CTX\text{-}M\text{-}15}$ in 12 isolates, followed by $bla_{CTX\text{-}M\text{-}27}$ in 8 isolates, $bla_{CTX\text{-}M\text{-}14}$ in 3 isolates, and $bla_{CTX\text{-}M\text{-}1}$ in 2 isolates. $bla_{CTX\text{-}M\text{-}15}$ was located

on a plasmid in 2 isolates and located within a chromosomally encoded transposon in 2 isolates. *bla*$_{CTX\text{-}M\text{-}15}$ was chromosomally encoded without an obvious mobile genetic element in the remaining 8 isolates (Fig. 2.1D). All isolates with *bla*$_{CTX\text{-}M\text{-}27}$ carried the gene in IncF plasmids (Table 2.2).

Other beta lactamase genes identified in ESBL-E isolates included *blaEC*, *bla*$_{TEM\text{-}1}$, and *bla*$_{OXA\text{-}1}$. All ESBL isolates carried the *blaEC* gene that confers resistance to Ampicillin, except for isolates 504 and 486, which carried *blaEC5* (not shown). Beta lactamase genes found in lower prevalence were *bla*$_{TEM\text{-}1}$ (carried by 8 isolates) and *bla*$_{OXA\text{-}1}$ (carried by 6 isolates). Seven copies of *bla*$_{TEM\text{-}1}$ were carried by IncF plasmids, and one isolate had a copy in the chromosome. *bla*$_{OXA\text{-}1}$ was encoded in an IncF plasmid of one isolate, in transposons of two isolates and chromosomally in three isolates (Fig. 2.1D). Isolate 492 carried three copies of *bla*$_{TEM\text{-}1}$, two of them in two different plasmids and one chromosomally. No currently characterized ESBL beta lactamase genes were identified in the child sample, C483. It is not currently clear whether this isolate carries an uncharacterized gene or may have lost its ESBL gene following initial isolation on selective media. In addition to these beta lactamases, all isolates carried antimicrobial resistance genes in the efflux class, while 16 carried AR genes for Fosfomycin, 12 for sulfonamides, 12 for tetracycline, 10 for trimethoprim, 7 for macrolides, 8 for quaternary ammonium and 5 for phenicol (Fig. 2.1C). Interestingly, the child isolate C483 carried two operons (*pco* and *sil)* for resistance genes in metal classes like Copper and Silver, both in close association with transposase genes. Nearly all isolates (22 of 26, including 497R) carried quinolone resistance genes. All ST131 isolates had mutations on *gyrA* and *parE* genes and exhibit Ciprofloxacin resistant phenotype, except isolate 482 (Fig. S2.2). For participants with a confirmed ESBL-E isolate from their second visit, all but two encoded the same resistance genes on re-isolation. One of these is 497R, which as discussed

elsewhere, denoted a different sequence type from the initial visit isolate. The other, 493R, was identified as clonal (14 SNPs) but lacked erythromycin, trimethoprim, streptomycin, tetracycline, and sulfamethoxazole antibiotic resistance genes that were present in the original isolate, 493. Both isolates carried a near-identical IncF plasmid, but in 493 this plasmid contained an IS6 insertion element that encoded the resistance genes and is missing in the plasmid carried by 493R (possibly lost in a deletion event) (Fig. S2.3).

Virulence genes and Plasmids

Phylogroup B2 isolates carried a large variety of virulence genes associated with extraintestinal *E. coli* (ExPEC) and had three or more genes associated with sepsis-associated *E. coli* (SEPEC) (Fig. 2.2).

However, virulence genes were not limited to phylogroup B2 isolates. The *afa* genes, encoding for afimbrial adhesins, are also associated with diffusely adherent *E. coli* (DAEC) and were found in isolates 498 and 499 from phylogroup D and 492 from phylogroup B2. Six of 9 genes in the adhesin associated locus (*aal*) that encodes for the Coli surface antigen 23 (CS23) were present in isolate 498, which also carried the *eatA* gene, encoding for ETEC autotransporter A. Isolate 495 carried the epsilon subtype of intimin outer membrane protein gene, *eae*, associated with enteropathogenic *E. coli* (EPEC), hemorrhagic *E. coli* (EHEC) and Shiga-toxin *E. coli* (STEC). Isolate 488 has 3 P fimbriae genes (A, C, F) in the same transposon as bla_{OXA-1} in addition to other resistance genes (*aac(6')-Ib-cr5*, *ermD*, *mdtM*).

The most frequently identified and biologically significant plasmids identified belonged to the IncF group, with 24 of the 26 unique ESBL-E isolates carrying at least one plasmid in this group. Most beta lactamases genes located within plasmids were carried by IncF plasmids, except for isolates 497 and 497R which carried $bla_{CTX-M-1}$ in an IncI1 plasmid (Table 2.2). The only plasmid

carrying more than one beta-lactamase gene was the IncF in isolate 495, which carried genes *bla*$_{TEM-1}$, *bla*$_{CTX-M-15}$, and *bla*$_{OXA-1}$. Isolate 492 has 2 plasmids each carrying a copy of *bla*$_{TEM-1}$. Eight isolates carried IncF plasmids that encoded *bla*$_{CTX-M-27}$, all of which also carried the plasmid-encoded enterotoxin, *senB*. *bla*$_{CTX-M-27}$ was encoded in an IS6 transposon array frequently associated with up to 16 additional antibiotic resistance genes (Fig. S2.4). Two of these isolates, 486 and 504, were near-clonal (52 SNPs) and carried plasmids that were 99.98% identical, with the main difference being two 65Kb IS6 transposon arrays encoding (among other genes) 16 resistance genes that are present in the plasmid carried by isolate 504 but absent in 486 (Fig. S2.5). Three isolates carried an IncI1 plasmid, including 497 and 497R. Isolates 497 and 497R, from the same participant at different times and in different sequence groups, shared a near-identical (99.7% pairwise identity) plasmid encoding *bla*$_{CTX-M-1}$ (Fig. S2.6). Plasmid group IncY was identified in 2 isolates, while replicons for IncI2, IncH, IncB/O/K/Z and IncN were each present in one isolate accompanied by an IncF plasmid (Table 2.2).

2.4.3 Carriage of antibiotic resistant *E. coli* among ESBL-E positive participants

We also examined the overall prevalence of AR among commensal *E. coli* in ESBL-E positive participants. Up to 48 *E. coli* isolates were tested for resistance to ampicillin, ceftriaxone, ciprofloxacin, tetracycline, and trimethoprim. Participants 500 and 498 yielded no *E. coli* colonies on this medium. Six of 25 ESBL-E positive participants showed dominance of beta-lactam resistant strains in their commensal *E. coli* community, with more than 50% of the colonies resistant to both ceftriaxone and ampicillin (Fig. 2.3). Five participants carried commensal *E. coli* resistant to ciprofloxacin. High levels (>50% colonies) of tetracycline resistance were found in 10 of the 25 ESBL-E positive participants. Trimethoprim resistance was less common, with only six participants showing high resistance (>50% colonies) (Fig. 2.3). In the second samples provided,

eight participants had an increase in the proportion of *E. coli* colonies resistant to both ceftriaxone and ampicillin. Most of the second samples that were ESBL-E negative, had a decreased proportion of commensal *E. coli* colonies that were resistant to the antibiotics tested.

Gut microbial community composition in ESBL-E positive and ESBL-E negative participants

ESBL-E colonization was not associated with significantly different stool microbiome alpha diversity (Fig. S2.7). Stool microbiome composition as evaluated by weighted Bray-Curtis distances (Fig. 2.4A) resulted in a significant PERMANOVA p-value of 0.03. However, the R^2 value was 0.003, indicating that ESBL-E carriage explained only a very small proportion of community variance. Participants with negative second visit samples did not exhibit significantly larger between-sample shifts in stool microbiome composition than participants with continued colonization at re-sampling (Fig. 2.4B; Fig. S2.8).

The relative abundance of most microbial classes was similar between ESBL-E positive and ESBL-E negative samples (Fig. S2.9). However, DESeq2 analysis identified 21 amplicon sequence variants (ASVs) that were significantly enriched and 50 that were significantly depleted in samples from ESBL-E positive vs. ESBL-E negative participants (Fig. 2.5).

2.5 Discussion

A key goal of this work was to evaluate asymptomatic carriage of ESBL-E in U.S. residents without significant healthcare exposure. We observed a carriage rate of 4.5% in our study population, which was recruited from the vicinity of the city of Athens, in northeastern Georgia. This reflects a higher prevalence than previous reports from the U.S. that range between 1.7% and 3.5% (26, 30, 32), although this remains lower than carriage rates reported from European and African countries (52–54). One caveat is that our study participants were recruited from the

southeastern U.S., which has reported a higher rate (19.9%) of healthcare associated ESBL-E infections than the national average of 12.2% (55).

Our study did not identify any demographic, socioeconomic, environmental, or health-associated risk factors that were significantly associated with the carriage of ESBL-E. This contrasts with previous studies that have reported antibiotic usage (56) and recent international travel (57) as correlated to ESBL-E carriage. These results may have been impacted by the COVID-19 pandemic, which limited international travel during the study period (January 2021-February 2022). This lack of association with specific risk factors suggests that ESBL *E. coli* may be circulating, albeit at low rates, among the general population in the study area.

A large fraction (14 of 22) of ESBL-E positive participants remained positive when re-tested at least 3 months after their initial visit, which is similar to previous reports in Sweden (58). That study, as well as others, found that sequence type ST131 or strains classified in the B2 and D phylogroups were more likely to persist (58, 59). However, our persistent isolates were distributed across many *E. coli* phylotypes, suggesting that persistent colonization is not confined to these phylogroups (60).

Overall, the most prevalent group of ESBL *E. coli* was B2, which matches previous studies in North America (61–64). In terms of sequence types, 46% of the isolates were identified as belonging to ST131, a globally distributed uropathogenic clade that is widely associated with ESBL gene carriage and identified as the most frequent multidrug resistant extraintestinal pathogenic *E. coli* (65). ST131 subclades C1, C2 and C1-M27 (as classified by $bla_{CTX\text{-}M}$ and *fimH* (66)) were all present.

All but one isolate carried at least one gene encoding for an CTX-M type enzyme, for a total prevalence of 96%, compared to the 90% reported in a previous review of community isolates

from different geographical regions worldwide (67). This suggests that *bla*$_{CTX\text{-}M}$ carrying *E. coli* may be a key driver of community ESBL-E spread in the southeastern U.S. In fact, an *E. coli* carrying *bla*$_{CTX\text{-}M\text{-}15}$ has been isolated from a stream sample in Athens, GA (68). Among the isolates that carried *bla*$_{\text{CTX-M}}$ genes, *bla*$_{CTX\text{-}M\text{-}15}$ was the most predominant, followed by *bla*$_{CTX\text{-}M\text{-}27}$ and *bla*$_{CTX\text{-}M\text{-}14}$. This agrees with previous reports that *bla*$_{\text{CTX-M-15}}$ is the most abundant beta lactamase gene circulating in the U.S., closely followed by *bla*$_{\text{CTX-M-27}}$ (69, 70), although a third study of urinary tract infections from gram-negative pathogens isolated in hospitals of Canada and the U.S. (2010-2014) reported *bla*$_{CTX\text{-}M\text{-}14}$ as more abundant than *bla*$_{CTX\text{-}M\text{-}27}$ (71). *bla*$_{CTX\text{-}M\text{-}27}$ was primarily associated with ST131 C1-M27 and C1/H30R clades as previously reported (72) but was also present in ST131 clones with *fimH41*. In all recovered isolates with this gene, the *bla*$_{CTX\text{-}M\text{-}27}$ was borne on an IncF-type plasmid.

In addition to the resistance shown to beta-lactams antibiotics, 72% of the ESBL-E isolates showed resistance to Ciprofloxacin and 24% were resistant to Gentamicin. Ciprofloxacin resistance is widely associated with ST131 (17); however, 7 isolates from other groups also exhibited ciprofloxacin resistance suggesting a broader relationship. A previous study of extra-intestinal pathogenic ESBLs from hospitalized patients in India reported a slightly lower prevalence of 65% (73). It remains unclear whether this increased prevalence is because ciprofloxacin resistance is more common in gut isolates or whether this is due to the population studied.

The most abundant plasmid type found in our isolates was IncF, which also was the plasmid type carrying the most antibiotic resistance and virulence genes, as has been widely reported (74). While plasmids from the incompatibility group IncI1 are typically associated with transfer of beta lactamase genes (75), we only isolated ESBL-carrying IncI1 plasmids from two participants. The IncI1 plasmid from participant 497, is of particular interest because the ESBL-*E. coli* isolates

obtained from the first and second fecal samples provided by this participant carried a near-identical copies of this plasmid but belonged to two different lineages of *E. coli*. This suggests a recent transfer of this plasmid in the human gut. Similar events have been reported elsewhere (76–79) and have been used to argue for the potential role of the human gut as a key site of horizontal gene transfer of antibiotic resistance genes.

In agreement with previous studies (80), many of our isolates, particularly those belonging to the B2 phylogroup, carried virulence genes associated with extra-intestinal pathogenic *E. coli* including uropathogenic (UPEC) and sepsis associated (SEPEC) strains. The *fimH* gene was the most prevalent virulence factor, followed by UPEC-associated genes *chuA, fyuA, usp*, and *yfcV* and *kpsMII* (found in 84% of ESBL-E isolates). Other virulence genes frequently found in our study include *iutA*, *papA, papC, sat, vat, cnf1* and *hlyA*, as previously shown by Iseppi et. al (81), however in contrast to their results *sfa* was not found in any isolate and *afa* genes were only in 3 isolates. Remarkably, the *eae* gene, previously used to classify EPEC isolates (82), was found in the isolate from participant 495, who also carried the only plasmid in this study with 3 different beta-lactamase genes. The presence of virulence genes in plasmids could explain the hybrid classification into different pathotypes of *E. coli* and contribute to the evolution of these pathotypes by accumulating virulence factors within the commensal community (83–85).

We also examined the frequency of antibiotic resistance among the overall population of commensal *E. coli* present in stool samples from ESBL-E positive participants. More than half of ESBL-E positive participants carried ampicillin- (64%) and tetracycline- (56%) resistant *E. coli*. However, presumptive ESBL-E made up most commensal *E. coli* in initial samples from only 6 of 25 ESBL-positive participants. This agrees with previous studies that in most cases the ESBL-producing *E. coli* is not the dominant strain among gut *E. coli* (54). Interestingly, 8 of the 22 second

samples showed an increase in the fraction of presumptive ESBL-*E. coli* isolated which could indicate a highly dynamic population structure and/or possible genetic exchange of antibiotic resistance genes among commensal *E. coli.*

Alpha and beta diversity analysis suggests that ESBL-E carriage was not associated with substantially different gut microbiome composition or diversity, consistent with the results of other studies (86, 87). However, pairwise analysis identified multiple taxa with significantly different abundance in carriers and non-carriers. Similar to other studies, we observed enriched taxa belonging to *Prevotella* (87, 88) and depletion of taxa belonging to *Sellimonas* (89) and *Bacteroides uniformis* (90). On the other hand, some taxa that other groups found enriched in ESBL-E carriers were identified as depleted in our ESBL-E communities, including *Clostridiales* (87), *Erysipelotrichaceae*, *Lactococcus*, *Bilophila* and *Negativibacillus* (88). In addition, *Desulfovibrio* and *Oscillospira* genera were identified as enriched among ESBL-E carriers in our study but depleted in another (86). These differences could be explained by the species of ESBL-*Enterobacteriaceae* in each study, given that the microbial population differs based on the ESBL-E species (89), or other factors influencing the microbiome composition as discussed before (91). Notable limitations of our study include that it was restricted to a specific geographic range (Athens, GA and vicinity) and that the study commenced during the global COVID-19 pandemic, which caused multiple large-scale changes in behavior including limitations on international travel. In addition, there was a possible bias provided by the participant self-reporting of risk factors. However, overall our results suggest that a subset of southeast U.S. residents are likely asymptomatic carriers of ESBL-*E. coli.*

2.6 Conclusions

To our knowledge this is the first genomic analysis of community associated ESBL-E carriage in the southeastern U.S. including long-term colonization, as previous studies only focused on the prevalence of ESBL-E or on healthcare-associated isolates. If the results from this study can be extrapolated, it suggests the potential for a small but notable increase in ESBL-E carriage in the U.S. since 2015, consistent with reports that the frequency of community-acquired ESBL-E infections also increased over this period (25) and supporting the role community-associated isolates in the incidence of ESBL-E outbreaks. Long term (>3 months) colonization was common in the study population, which underscores the potential of the human gut microbiome to serve as a long-term reservoir of ESBL *Enterobacteriaceae*. Colonizing ESBL *Enterobacteriaceae* were all identified as *E. coli* including strains that are unlikely to be pathogenic and strains carrying virulence genes associated with extraintestinal pathogenic *E. coli.* Finally, many strains carried multidrug resistance plasmids and we observed at least one participant where the same plasmid was observed in isolates with different phylogenetic backgrounds, consistent with a role for the human gut as a hotspot for antibiotic resistance gene exchange (92). Overall, our study suggests that the human gut may represent an important but under-recognized reservoir of ESBL genes and ESBL-carrying *E. coli and* highlights the relevance and importance of understanding the role of gut commensals in mediating the spread of antibiotic resistance.

2.7 Acknowledgements

This work was supported by BAA contract no. 75D30120C09496 from the Centers for Disease Control and Prevention. Funders had no role in data collection and interpretation, or the decision to submit the publication. Participants recruitment and sample processing was performed with the help of the CTRU at University of Georgia, which is supported by the National Center for

Advancing Translational Sciences of the National Institutes of Health under award number: UL1TR002378. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.8 References

1. WHO Association Media Centre. 2015. Sixty-Eighth World Health AssemblyWorld Health Organization.

2. CDC. 2019. Antibiotic Resistance Threats in The United States, 2019U.S. Department of Health and Human Services.

3. Nelson RE, Slayton RB, Stevens VW, Jones MM, Khader K, Rubin MA, Jernigan JA, Samore MH. 2017. Attributable mortality of healthcare-associated infections due to multidrug-resistant gram-negative bacteria and methicillin-resistant staphylococcus aureus. Infect Control Hosp Epidemiol 38:848–856.

4. Neidell MJ, Cohen B, Furuya Y, Hill J, Jeon CY, Glied S, Larson EL. 2012. Costs of healthcare-and community-associated infections with antimicrobial-resistant versus antimicrobial-susceptible organisms. Clinical Infectious Diseases 55:807–815.

5. Knothe H, Shah P, Krcmery V, Antal M, Mitsuhashi S. 1983. Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of Klebsiella pneumoniae and Serratia marcescens. Infection 11:315–317.

6. Bush K, Bradford PA. 2020. Epidemiology of B-Lactamase-Producing Pathogens. Clin Microbiol Rev 33:1–37.

7. Rupp ME, Fey PD. 2003. Extended spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae: considerations for diagnosis, prevention and drug treatment. Drugs 63:353–365.

8. Whichard JM, Joyce K, Fey PD, Nelson JM, Angulo FJ, Barrett TJ. 2005. β-Lactam Resistance and Enterobacteriaceae, United States. Emerg Infect Dis 11:1464–1466.

9. Rawat D, Nair D. 2010. Extended-spectrum ß-lactamases in gram negative bacteria. J Glob Infect Dis 2:263.

10. Castanheira M, Simner PJ, Bradford PA. 2021. Extended-spectrum β-lactamases: An update on their characteristics, epidemiology and detection. JAC Antimicrob Resist 3.

11. Doi Y, Iovleva A, Bonomo RA. 2017. The ecology of extended-spectrum β-lactamases (ESBLs) in the developed world. J Travel Med 24:S44–S51.

12. Logan LK, Rispens JR, Medernach RL, Domitrovic TN, Hujer AM, Marshall SH, Rudin SD, Qureshi NK, Zheng X, Hayden MK, Weinstein RA, Bonomo RA. 2021. A Multicentered Study of the Clinical and Molecular Epidemiology of TEM- and SHV-type Extended-Spectrum Beta-Lactamase Producing Enterobacterales Infections in Children. Pediatric Infectious Disease Journal 40:39–43.

13. Cantón R, González-Alba JM, Galán JC. 2012. CTX-M enzymes: Origin and diffusion. Front Microbiol 3.

14. Cantón R, Coque TM. 2006. The CTX-M β-lactamase pandemic. Curr Opin Microbiol 9:466–475.

15. Logan LK, Hujer AM, Marshall SH, Domitrovic TN, Rudin SD, Zheng X, Qureshi NK, Hayden MK, Scaggs FA, Karadkhele A, Bonomo RA. 2016. Analysis of β-Lactamase Resistance Determinants in Enterobacteriaceae from Chicago Children: a Multicenter Survey. Antimicrob Agents Chemother 60:3462–3469.

16. Hu F, O'Hara JA, Rivera JI, Doi Y. 2014. Molecular Features of Community-Associated Extended-Spectrum-β-Lactamase-Producing Escherichia coli Strains in the United States. Antimicrob Agents Chemother 58:6953–6957.

17. Matsumura Y, Pitout JDD, Peirano G, Devinney R, Noguchi T, Yamamoto M, Gomi R, Matsuda T, Nakano S, Nagao M, Tanaka M, Ichiyama S. 2017. Rapid Identification of Different Escherichia coli Sequence Type 131 Clades. Antimicrob Agents Chemother 61:1–9.

18. Salamzade R, McElheny CL, Manson AL, Earl AM, Shaikh N, Doi Y. 2023. Genomic epidemiology and antibiotic susceptibility profiling of uropathogenic *Escherichia coli* among children in the United States. mSphere 8.

19. Nicolas-Chanoine MH, Bertrand X, Madec JY. 2014. Escherichia coli st131, an intriguing clonal group. Clin Microbiol Rev 27:543–574.

20. Shaikh S, Fatima J, Shakil S, Rizvi SMD, Kamal MA. 2015. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. Saudi J Biol Sci 22:90–101.

21. Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK. 2007. Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates. Appl Environ Microbiol 73:1976–1983.

22. Lukac PJ, Bonomo RA, Logan LK. 2015. Extended-Spectrum -Lactamase-Producing Enterobacteriaceae in Children: Old Foe, Emerging Threat. Clinical Infectious Diseases https://doi.org/10.1093/cid/civ020.

23. Carattoli A. 2011. Plasmids in Gram negatives: Molecular typing of resistance plasmids. International Journal of Medical Microbiology 301:654–658.

24. van Duin D, Paterson DL. 2020. Multidrug-Resistant Bacteria in the Community: An Update. Infect Dis Clin North Am 34:709–722.

25. Bezabih YM, Bezabih A, Dion M, Batard E, Teka S, Obole A, Dessalegn N, Enyew A, Roujeinikova A, Alamneh E, Mirkazemi C, Peterson GM, Bezabhe WM. 2022. Comparison of the global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli between healthcare and community settings: A systematic review and meta-Analysis. JAC Antimicrob Resist 4.

26. Bezabih YM, Sabiiti W, Alamneh E, Bezabih A, Peterson GM, Bezabhe WM, Roujeinikova A. 2021. The global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli in the community. Journal of Antimicrobial Chemotherapy 76:22–29.

27. Bevan ER, Jones AM, Hawkey PM. 2017. Global epidemiology of CTX-M β-lactamases: Temporal and geographical shifts in genotype. Journal of Antimicrobial Chemotherapy 72.

28. Doi Y, Park YS, Rivera JI, Adams-Haduch JM, Hingwe A, Sordillo EM, Lewis JS, Howard WJ, Johnson LE, Polsky B, Jorgensen JH, Richter SS, Shutt KA, Paterson DL. 2013. Community-associated extended-spectrum -lactamase-producing Escherichia coli infection in the United States. Clinical Infectious Diseases 56:641–648.

29. Doi Y, Adams J, O'Keefe A, Quereshi Z, Ewan L, Paterson DL. 2007. Community-acquired extended spectrum B-lactamase producers, United States.

30. Vento TJ, Cole DW, Mende K, Calvano TP, Rini EA, Tully CC, Zera WC, Guymon CH, Yu X, Cheatle KA, Akers KS, Beckius ML, Landrum ML, Murray CK. 2013. Multidrug-resistant gram-negative bacteria colonization of healthy US military personnel in the US and Afghanistan. BMC Infect Dis 13.

31. Weisenberg SA, Mediavilla JR, Chen L, Alexander EL, Rhee KY, Kreiswirth BN, Jenkins SG. 2012. Extended Spectrum Beta-Lactamase-Producing Enterobacteriaceae in International Travelers and Non-Travelers in New York City. PLoS One 7:2010–2013.

32. Islam S, Selvarangan R, Kanwar N, McHenry R, Chappell JD, Halasa N, Wikswo ME, Payne DC, Azimi PH, McDonald LC, Gomez-Duarte OG. 2018. Intestinal Carriage of Third-Generation Cephalosporin-Resistant and Extended-Spectrum β-Lactamase-Producing Enterobacteriaceae in Healthy US Children. J Pediatric Infect Dis Soc 7:234–240.

33. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300.

34. Jafari M, Ansari-Pour N. 2019. Why, when and how to adjust your P values? Cell J 20.

35. Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vásquez NJ, Finger JW, Pierson TW, Bentley KE, Hoffberg SL, Louha S, Garcia-De Leon FJ, Del Rio Portilla MA, Reed KD, Anderson JL, Meece JK, Aggrey SE, Rekaya R, Alabady M, Belanger M, Winker K, Faircloth BC. 2019. Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ 2019.

36. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:1–22.

37. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:1–15.

38. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS One 9:1–8.

39. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 11.

40. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. 2021. Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. Journal of Antimicrobial Chemotherapy 76:101–109.

41. Larsen M V., Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361.

42. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426.

43. Roer L, Tchesnokova V, Allesøe R, Muradova M, Chattopadhyay S, Ahrenfeldt J, Thomsen MCF, Lund O, Hansen F, Hammerum AM, Sokurenko E, Hasmana H. 2017. Development of Web Tool for Escherichia coli Subtyping Based on fimH Alleles. J Clin Microbiol 55:2538–2543.

44. Malberg Tetzschner AM, Johnson JR, Johnston BD, Lund O, Scheutzd F. 2020. In Silico Genotyping of Escherichia coli Isolates for Sequencing Data. J Clin Microbiol 58:e01269-20.

45. Beghain J, Bridier-Nahmias A, Nagard H Le, Denamur E, Clermont O. 2018. ClermonTyping: An easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genom 4:1–8.

46. Tinker KA, Ottesen EA. 2016. The core gut microbiome of the American cockroach, Periplaneta americana, is stable and resilient to dietary shifts. Appl Environ Microbiol 82:6603–6610.

47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.

48. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41:590–596.

49. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

50. McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8.

51. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E. 2016. Fecal Colonization with Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors among Healthy Individuals: A Systematic Review and Metaanalysis. Clinical Infectious Diseases 63:310–318.

52. Geser N, Stephan R, Korczak BM, Beutin L, Hächler H. 2012. Molecular identification of extended-spectrum-β-lactamase genes from Enterobacteriaceae isolated from healthy human carriers in Switzerland. Antimicrob Agents Chemother 56:1609–1612.

53. Mshana SE, Falgenhauer L, Mirambo MM, Mushi MF, Moremi N, Julius R, Seni J, Imirzalioglu C, Matee M, Chakraborty T. 2016. Predictors of blaCTX-M-15 in varieties of Escherichia coli genotypes from humans in community settings in Mwanza, Tanzania. BMC Infect Dis 16:1–9.

54. Nicolas-Chanoine MH, Gruson C, Bialek-Davenet S, Bertrand X, Thomas-Jean F, Bert F, Moyat M, Meiller E, Marcon E, Danchin N, Noussair L, Moreau R, Leflon-Guibout V. 2013. 10-fold increase (2006-11) in the rate of healthy subjects with extended-spectrum β-lactamase-producing Escherichia coli faecal carriage in a parisian check-up centre. Journal of Antimicrobial Chemotherapy 68:562–568.

55. Castanheira M, Farrell SE, Krause KM, Jones RN, Sader HS. 2014. Contemporary diversity of β-lactamases among enterobacteriaceae in the nine U.S. census regions and ceftazidime-avibactam activity tested against isolates producing the most prevalent β-lactamase groups. Antimicrob Agents Chemother 58:833–838.

56. van den Bunt G, van Pelt W, Hidalgo L, Scharringa J, de Greeff SC, Schürch AC, Mughini-Gras L, Bonten MJM, Fluit AC. 2019. Prevalence, risk factors and genetic characterisation of extended-spectrum beta-lactamase and carbapenemase-producing Enterobacteriaceae (ESBL-E and CPE): A community-based cross-sectional study, the Netherlands, 2014 to 2016. Eurosurveillance 24.

57. Mizuno Y, Miura Y, Yamaguchi T, Matsumoto T. 2016. Extended-spectrum β-lactamase-producing Enterobacteriaceae colonisation in long-term overseas business travellers. Travel Med Infect Dis 14:561–567.

58. Titelman E, Hasan CM, Iversen A, Nauclér P, Kais M, Kalin M, Giske CG. 2014. Faecal carriage of extended-spectrum β-lactamase-producing Enterobacteriaceae is common 12 months after infection and is related to strain factors. Clinical Microbiology and Infection 20.

59. van Duijkeren E, Wielders CCH, Dierikx CM, van Hoek AHAM, Hengeveld P, Veenman C, Florijn A, Lotterman A, Smit LAM, van Dissel JT, Maassen CBM, de Greeff SC. 2018. Long-

term Carriage of Extended-Spectrum β-Lactamase-Producing Escherichia coli and Klebsiella pneumoniae in the General Population in The Netherlands. Clin Infect Dis 66.

60. Rumball NA, Alm EW, McLellan SL. 2023. Genetic Determinants of Escherichia coli Survival in Beach Sand. Appl Environ Microbiol 89.

61. Logue CM, Doetkott C, Mangiamele P, Wannemuehler YM, Johnson TJ, Tivendale KA, Li G, Sherwood JS, Nolan LK. 2012. Genotypic and phenotypic traits that distinguish neonatal meningitis-associated Escherichia coli from fecal E. coli isolates of healthy human hosts. Appl Environ Microbiol 78.

62. Johnson JR, Owens K, Gajewski A, Kuskowski MA. 2005. Bacterial characteristics in relation to clinical source of Escherichia coli isolates from women with acute cystitis or pyelonephritis and uninfected women. J Clin Microbiol 43.

63. Zhang L, Foxman B, Marrs C. 2002. Both urinary and rectal Escherichia coli isolates are dominated by strains of phylogenetic group B2. J Clin Microbiol 40.

64. Sannes MR, Kuskowski MA, Owens K, Gajewski A, Johnson JR. 2004. Virulence factor profiles and phylogenetic background of Escherichia coli isolates from veterans with bacteremia and uninfected control subjects. Journal of Infectious Diseases 190.

65. Rogers BA, Doi Y. 2013. Who Is Leading This Dance? Understanding the Spread of Escherichia coli Sequence Type 131 . Infect Control Hosp Epidemiol 34.

66. Petty NK, Zakour NLB, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Moriel DG, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-Baño J, Pascual A, Pitout JDD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. 2014. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A 111.

67. Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. Clin Microbiol Rev 26:744–758.

68. Cho S, Nguyen HAT, McDonald JM, Woodley TA, Hiott LM, Barrett JB, Jackson CR, Frye JG. 2019. Genetic characterization of antimicrobial-resistant escherichia coli isolated from a mixed-use watershed in northeast Georgia, USA. Int J Environ Res Public Health 16:1–14.

69. Duffy N, Karlsson M, Reses HE, Campbell D, Daniels J, Stanton RA, Janelle SJ, Schutz K, Bamberg W, Rebolledo PA, Bower C, Blakney R, Jacob JT, Phipps EC, Flores KG, Dumyati G, Kopin H, Tsay R, Kainer MA, Muleta D, Byrd-Warner B, Grass JE, Lutgring JD, Rasheed JK, Elkins CA, Magill SS, See I. 2022. Epidemiology of extended-spectrum β-lactamase-producing Enterobacterales in five US sites participating in the Emerging Infections Program, 2017. Infect Control Hosp Epidemiol 43:1586–1594.

70. Mostafa HH, Cameron A, Taffner SM, Wang J, Malek A, Dumyati G, Hardy DJ, Pecora ND. 2020. Genomic Surveillance of Ceftriaxone-Resistant Escherichia coli in Western New York Suggests the Extended-Spectrum β-Lactamase blaCTX-M-27 Is Emerging on Distinct Plasmids in ST38. Front Microbiol 11.

71. Lob SH, Nicolle LE, Hoban DJ, Kazmierczak KM, Badal RE, Sahm DF. 2016. Susceptibility patterns and ESBL rates of Escherichia coli from urinary tract infections in Canada and the United States, SMART 2010–2014. Diagn Microbiol Infect Dis 85:459–465.

72. Matsumura Y, Pitout JDD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, De Vinney R, Bradford PA, Motyl MR, Tanaka M, Nagao M, Takakura S, Ichiyama S. 2016. Global Escherichia coli sequence type 131 clade with blaCTX-M-27 gene. Emerg Infect Dis 22:1900–1907.

73. Chakraborty A, Adhikari P, Shenoy S, Saralaya V. 2015. Clinical significance and phylogenetic background of extended spectrum β-lactamase producing Escherichia coli isolates from extra-intestinal infections. J Infect Public Health 8:248–253.

74. Carattoli A. 2009. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother 53:2227–2238.

75. Carattoli A, Villa L, Fortini D, García-Fernández A. 2021. Contemporary IncI1 plasmids involved in the transmission and spread of antimicrobial resistance in Enterobacteriaceae. Plasmid https://doi.org/10.1016/j.plasmid.2018.12.001.

76. Göttig S, Gruber TM, Stecher B, Wichelhaus TA, Kempf VAJ. 2015. In vivo horizontal gene transfer of the carbapenemase OXA-48 during a nosocomial outbreak. Clinical Infectious Diseases 60.

77. Trobos M, Lester CH, Olsen JE, Frimodt-Møller N, Hammerum AM. 2009. Natural transfer of sulphonamide and ampicillin resistance between Escherichia coli residing in the human intestine. Journal of Antimicrobial Chemotherapy 63.

78. Sandegren L, Linkevicius M, Lytsy B, Melhus Å, Andersson DI. 2012. Transfer of an Escherichia coli ST131 multiresistance cassette has created a Klebsiella pneumoniae-specific plasmid associated with a major nosocomial outbreak. Journal of Antimicrobial Chemotherapy 67.

79. Hazen TH, Mettus R, McElheny CL, Bowler SL, Nagaraj S, Doi Y, Rasko DA. 2018. Diversity among blaKPC-containing plasmids in Escherichia coli and other bacterial species isolated from the same patients. Sci Rep 8:10291.

80. Geurtsen J, de Been M, Weerdenburg E, Zomer A, McNally A, Poolman J. 2022. Genomics and pathotypes of the many faces of Escherichia coli. FEMS Microbiol Rev 46.

81. Iseppi R, Di Cerbo A, Messi P, Sabia C. 2020. Antibiotic resistance and virulence traits in vancomycin-resistant enterococci (Vre) and extended-spectrum β-lactamase/ampc-producing (ESBL/ampc) enterobacteriaceae from humans and pets. Antibiotics 9.

82. Chandra M, Cheng P, Rondeau G, Porwollik S, McClelland M. 2013. A single step multiplex PCR for identification of six diarrheagenic E. coli pathotypes and Salmonella. International Journal of Medical Microbiology 303:210–216.

83. Hazen TH, Michalski J, Luo Q, Shetty AC, Daugherty SC, Fleckenstein JM, Rasko DA. 2017. Comparative genomics and transcriptomics of Escherichia coli isolates carrying virulence factors of both enteropathogenic and enterotoxigenic E. coli. Sci Rep 7.

84. Yu F, Chen X, Zheng S, Han D, Wang Y, Wang R, Wang B, Chen Y. 2018. Prevalence and genetic diversity of human diarrheagenic Escherichia coli isolates by multilocus sequence typing. International Journal of Infectious Diseases 67:7–13.

85. Neuwirth C, Siebor E, Pechinot A, Duez JM, Pruneaux M, Garel F, Kazmierczak A, Labia R. 2001. Evidence of in vivo transfer of a plasmid encoding the extended-spectrum β-lactamase TEM-24 and other resistance factors among different members of the family Enterobacteriaceae. J Clin Microbiol 39:1985–1988.

86. Gosalbes MJ, Vázquez-Castellanos JF, Angebault C, Woerther PL, Ruppé E, Ferrús ML, Latorre A, Andremont A, Moya A. 2016. Carriage of enterobacteria producing extended-spectrum β-lactamases and composition of the gut microbiota in an Amerindian community. Antimicrob Agents Chemother 60:507–514.

87. Le Bastard Q, Chapelet G, Birgand G, Hillmann BM, Javaudin F, Hayatgheib N, Bourigault C, Bemer P, De Decker L, Batard E, Lepelletier D, Montassier E. 2020. Gut microbiome

signatures of nursing home residents carrying Enterobacteria producing extended-spectrum β-lactamases. Antimicrob Resist Infect Control 9.

88. Peng Y, Liang S, Poonsuk K, On H, Li SW, Pascal Maurin MM, Chan CH, Chan CL, Sin ZY, Tun HM. 2021. Role of gut microbiota in travel-related acquisition of extended spectrum β-lactamase-producing Enterobacteriaceae. J Travel Med 28:1–11.

89. Prevel R, Enaud R, Orieux A, Camino A, Sioniac P, M'Zali F, Dubois V, Berger P, Boyer A, Delhaes L, Gruson D. 2023. Bridging gut microbiota composition with extended-spectrum beta-lactamase Enterobacteriales faecal carriage in critically ill patients (microbe cohort study). Ann Intensive Care 13.

90. Piewngam P, Quiñones M, Thirakittiwatthana W, Yungyuen T, Otto M, Kiratisin P. 2019. Composition of the intestinal microbiota in extended-spectrum β-lactamase-producing Enterobacteriaceae carriers and non-carriers in Thailand. Int J Antimicrob Agents 53.

91. Stoppe N de C, Silva JS, Carlos C, Sato MIZ, Saraiva AM, Ottoboni LMM, Torres TT. 2017. Worldwide phylogenetic group patterns of Escherichia coli from commensal human and wastewater treatment plant isolates. Front Microbiol 8.

92. Nhu NTQ, Young VB. 2023. The Relationship Between the Microbiome and Antimicrobial Resistance. Clinical Infectious Diseases 77:S479–S486.

Table 2.1: Selected demographic characteristics, risk factors, and presence of ESBL-E among all adult participants. The demographic data presented on Table 1 are based on the questionnaire responses from the first visit of all adult participants. The child demographics were not included in the tables or statistical analyses of risk factors, as they were not considered independent of the parent participants.

Demographic or Risk Factor	N	ESBL-E positive N=24	ESBL-E negative N=481	p-value[1]/FDR p-adjusted[2]
Biological Sex				
Female	349	13 (54%)	336 (70%)	0.12/0.79
Male	156	11 (46%)	145 (30%)	
Race				
Asian	37	4 (17%)	33 (7%)	0.17/0.79
Black or African American	59	5 (21%)	54 (11%)	
White	377	15 (63%)	362 (75%)	
Age (years)				
18-29	234	13 (54%)	221 (46%)	0.17/.079
30-39	84	2 (8%)	82 (17%)	
40-49	53	3 (13%)	50 (10%)	
50-59	49	2 (8%)	47 (10%)	
60-69	45	1 (4%)	44 (9%)	
70-79	22	2 (8%)	20 (4%)	
80-89	5	0 (0%)	5 (1%)	
Antibiotic use in the past 1 month (oral, topical, or intravenous)				
Yes	47	2 (8%)	45 (9%)	1/1
No	454	22 (92%)	432 (90%)	
Gastrointestinal conditions or symptoms in the past 1 month				
Yes	177	8 (33%)	169 (35%)	1/1
No	327	16 (67%)	311 (65%)	
Urinary tract infection in the past 1 month (self-diagnosed or diagnosed by doctor)				
Yes	13	1 (4%)	12 (2%)	0.47/0.79
No	492	23 (96%)	469 (98%)	
International travel (past year)				
Yes	31	1 (4%)	30 (6%)	1/1
No	471	23 (96%)	448 (93%)	
Lived internationally in the last 5 years				
Yes	37	4 (17%)	33 (7%)	0.09/0.79
No	463	20 (83%)	443 (92%)	
Exposure to treated recreational water in the past month				
Yes	127	4 (17%)	123 (26%)	0.47/0.79
No	371	20 (83%)	351 (73%)	
Exposure to untreated recreational water in the past month				
Yes	107	3 (13%)	104 (22%)	0.44/0.79
No	391	21 (88%)	370 (77%)	

Table 2.2: Assembled plasmids carried by each isolate identified by PlasmidFinder with encoded antibiotic genes identified by AMRFinderPlus and encoded virulence genes identified by VirulenceFinder. Plasmid number in each isolate matches the numbers in panels C and D of Figure 2.1 and Figure 2.2.

Isolate[a]	Plasmid Type	Plasmid-associated Antibiotic Resistance Genes	Virulence Genes
497R (n=4)	1. IncI1 2. IncFII	1. *bla*$_{CTX\text{-}M\text{-}1}$	1. *cia* 2. *mcbA, traJ, traT*
498 (n=3)	1. IncFII 2. IncFIB	2. *bla*$_{CTX\text{-}M\text{-}27}$	1. *aap, eatA, faeF, traT* 2. *traJ, traT*
497 (n=2)	1. IncF 2. IncI1	1. *tetA, tetR* 2. *bla*$_{CTX\text{-}M\text{-}1}$	1. *anr, ompT* 2. *cia*
495 (n=2)	1. IncF 2. IncF	1. *bla*$_{CTX\text{-}M\text{-}15}$, *bla*$_{OXA\text{-}1}$, *bla*$_{TEM\text{-}1}$, *aph(3'')-Ib, aph(6)-Id, aac(6')-Ib-cr5, catB3, floR, qacEdelta1, qnrS1, sul1, sul2, terB, terC, terD, terE, tet(A), dfrA1*	1. *terC* 2. *anr, traJ, traT*
483 (n=3)	1. IncFII 2. IncB/O/Z/K	1. *bla*$_{CTX\text{-}M\text{-}15}$, *qnrS1*	1. *traT* 2. *ireA*
490 (n=3)	1. IncF	1. *tet(B)*	1, *anr, iucC, iutA, sitA, traT*
492 (n=6)	1. IncF 2. IncFII	1. *bla*$_{TEM\text{-}1}$, *aph(3'')-Ib, aph(6)-Id, sul2, tet(B)* 2. *bla*$_{TEM\text{-}1}$	1. *anr, iucC, iutA, sitA, traT* 2. *mcbA, traJ, traT*
504 (n=3)	1. IncF 2. IncFII	1. *bla*$_{CTX\text{-}M\text{-}27}$, *aadA5, aph(3'')-Ib, aph(6)-Id, erm(B), mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *senB* 2. *fyuA, traJ*
486 (n=3)	1. IncF	1. *bla*$_{CTX\text{-}M\text{-}27}$	1. *senB*
482 (n=3)	1. IncF 2. Col156		1. *traJ, traT* 2. *senB*
505 (n=1)	1. IncF	1. *bla*$_{CTX\text{-}M\text{-}27}$, *aadA5, aph(3'')-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *anr, senB, traT*
493 (n=2)	1. IncF 2. IncF	1. *bla*$_{CTX\text{-}M\text{-}27}$, *aadA5, aph(3'')-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *anr, senB, traT* 2. *traT*
496 (n=1)	1. IncF	1. *bla*$_{TEM\text{-}1}$, *aph(3'')-Ib, aph(6)-Id, qnrS1, sul2, dfrA14*	1. *anr, traJ, traT*
503 (n=1)	1. IncF	1. *tet(A)*	1. *iutA, senB*
485 (n=2)	1. IncF		1. *anr, iutA, senB, traJ, traT*
487 (n=2)	1. IncF	1. *aph(3'')-Ib, aph(6)-Id, sul2, tet(A)*	1. *anr, senB*
488 (n=3)	1. IncFI 2. IncFII	1. *bla*$_{TEM\text{-}1}$, *aadA5, aph(3'')-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *traJ, traT* 2. *mcbA, traT*
500 (n=3)	1. IncF	1. *bla*$_{TEM\text{-}1}$, *aac(3)-IId, aadA5, aph(3'')-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *senB, traJ, traT*
484 (n=3)	1. IncF 2. IncFII	1. *bla*$_{TEM\text{-}1}$, *aac(3)-IId, aph(3'')-Ib, aph(6)-Id, sul2, tet(B), dfrA17*	1. *anr* 2. *traT*
494 (n=2)	1. IncF	1. *bla*$_{TEM\text{-}1}$	1. *anr, senB, traT*
489 (n=2)	1. IncF	1. *bla*$_{CTX\text{-}M\text{-}27}$	1. *anr, senB, traT*
501 (n=3)	1. IncF 2. IncI1	1. *bla*$_{CTX\text{-}M\text{-}27}$	1. *anr, senB, traT* 2. *cia*
502 (n=2)	1. IncF	1. *bla*$_{CTX\text{-}M\text{-}27}$	1. *anr, senB*

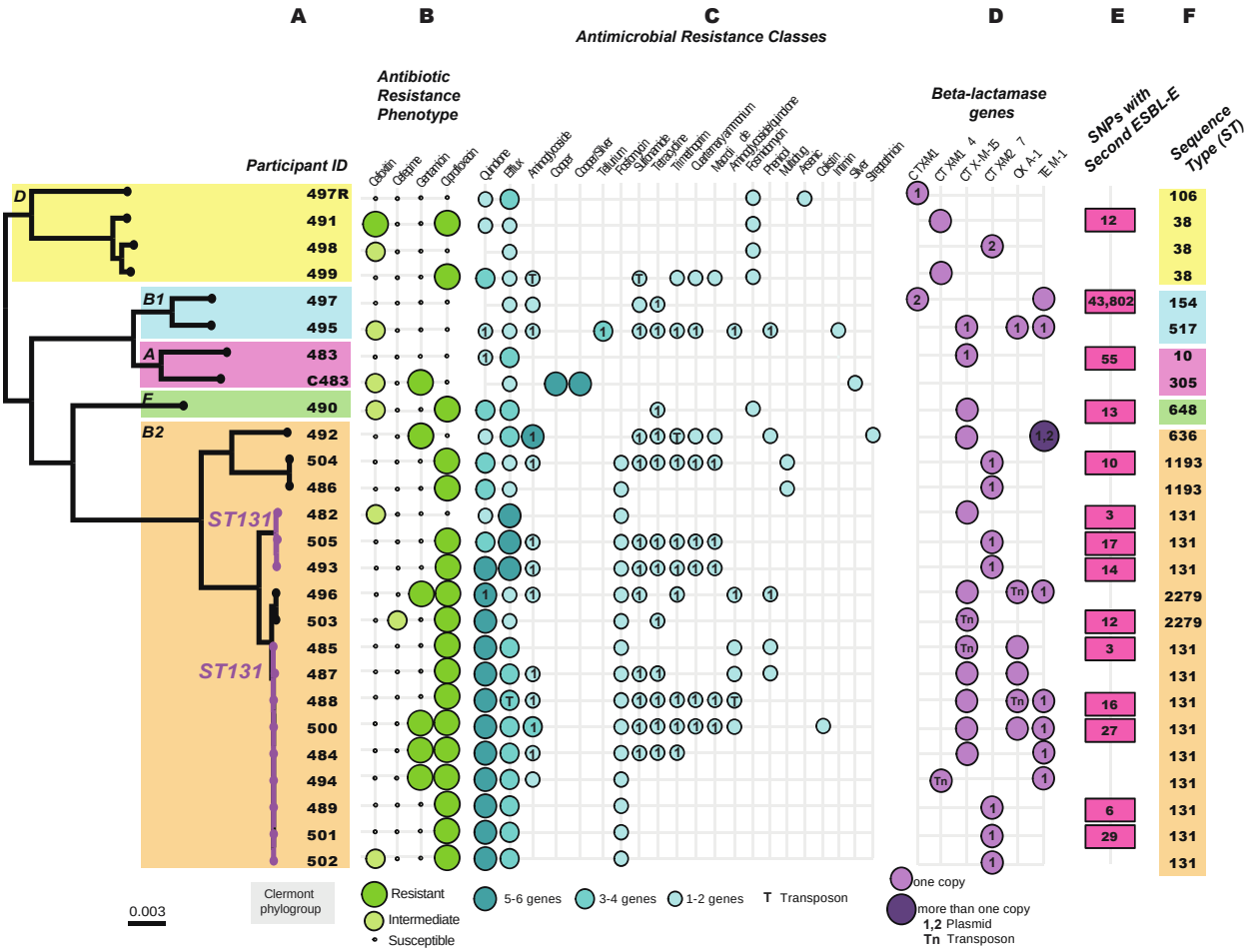

Figure 2.1: Characteristics and classification of ESBL-E isolates. A) Core-genome single nucleotide polymorphisms (SNPs) phylogenetic tree of ESBL-E isolates per participant colored by Clermont phylogroup, highlighting ST131 isolates. B) Phenotypic characterization of antibiotic resistance by isolate tested in Sensititre, after the confirmation as ESBL-E given their resistance to beta-lactams: Cefotaxime and Ceftazidime. C) Antimicrobial resistance genes found by AMRFinderPlus in the genome, transposons (T) or plasmids (1) of each isolate classified by antimicrobial class. D) Beta-lactamase genes found in each isolate, chromosomally or in mobile elements. Circles labelled with a number match plasmids described in Table S2.5 for each isolate. E) SNPs between first and second sample confirmed as ESBL-positive to determine clonality of isolates. F) Sequence type (ST) classification of each isolate, colored by phylogroup.

Second samples collected (R) are not included in the figure as they were the same sequence type (ST) and share almost the same genetic content as their original sample, with the exception of 497R.

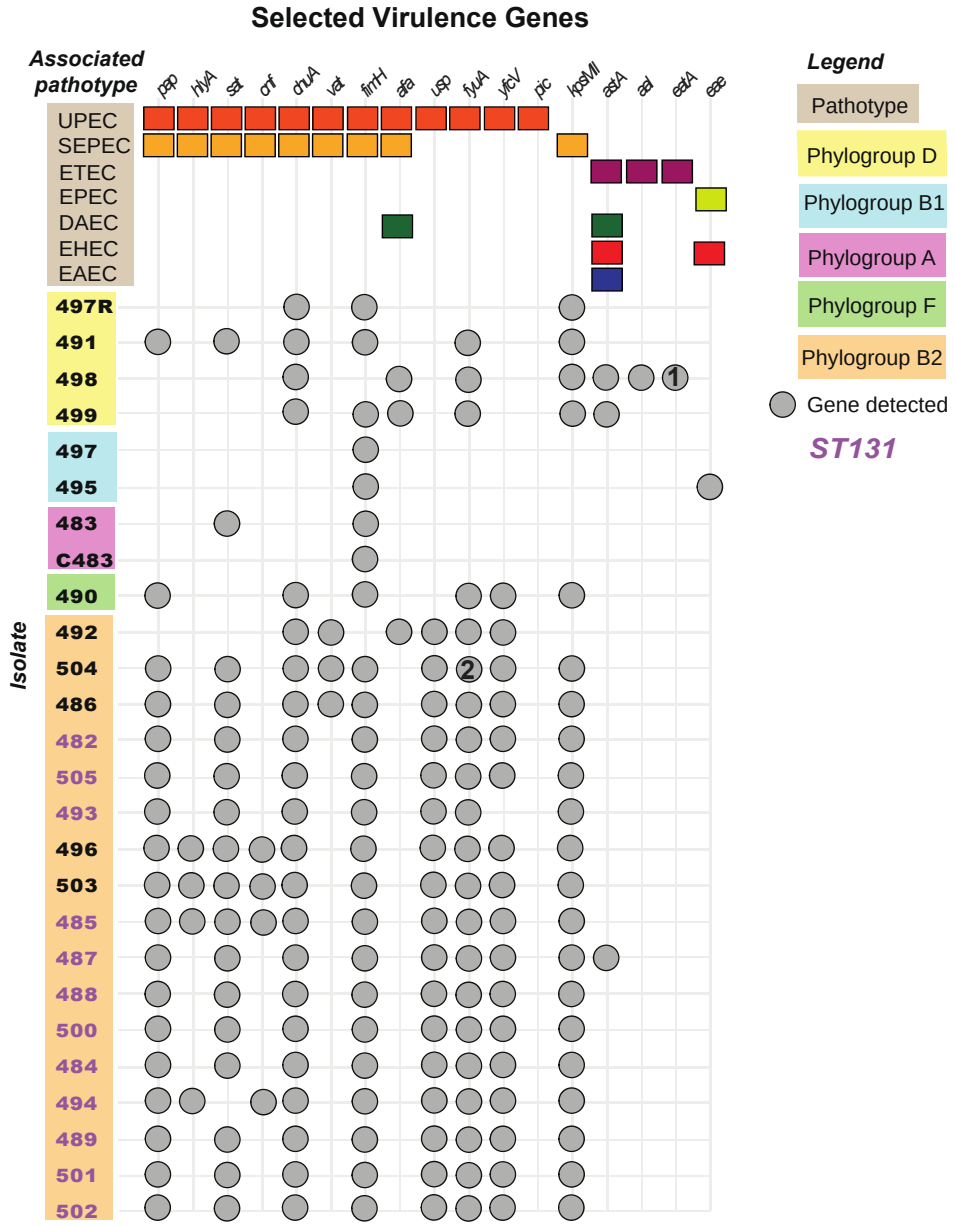

Figure 2.2: Selected virulence genes associated with pathogenic *E. coli* pathotypes. Virulence genes were identified by VirulenceFinder using the genome assemblies of each isolate. Only selected virulence genes are shown based on their associated pathotype. Colored boxes represent the genes associated with each pathotype and gray circles indicate the presence of that gene in each isolate. Isolates are clustered by

Clermont phylogroups matching Figure 2.1. Circles labelled with a number match plasmids described in Table S2.5 for each isolate.

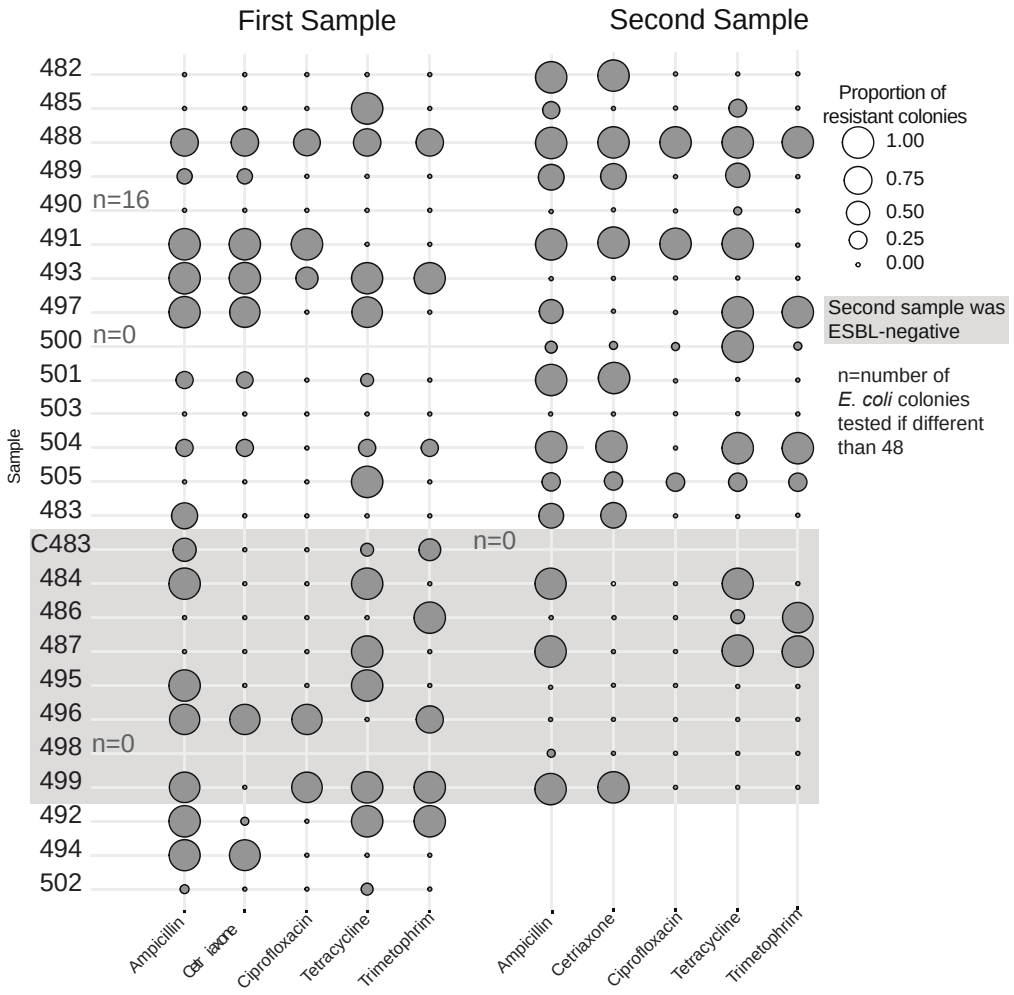

Figure 2.3: Antibiotic resistance phenotype of commensal *E. coli* isolated from ESBL-E positive participants. A range between 1-48 *E. coli* colonies isolated from ESBL-E positive participants were patched into Mueller Hinton II agar with antibiotics at CSLI standard levels. Each circle size and color intensity show the proportion of those colonies that were resistant to the antibiotic tested.

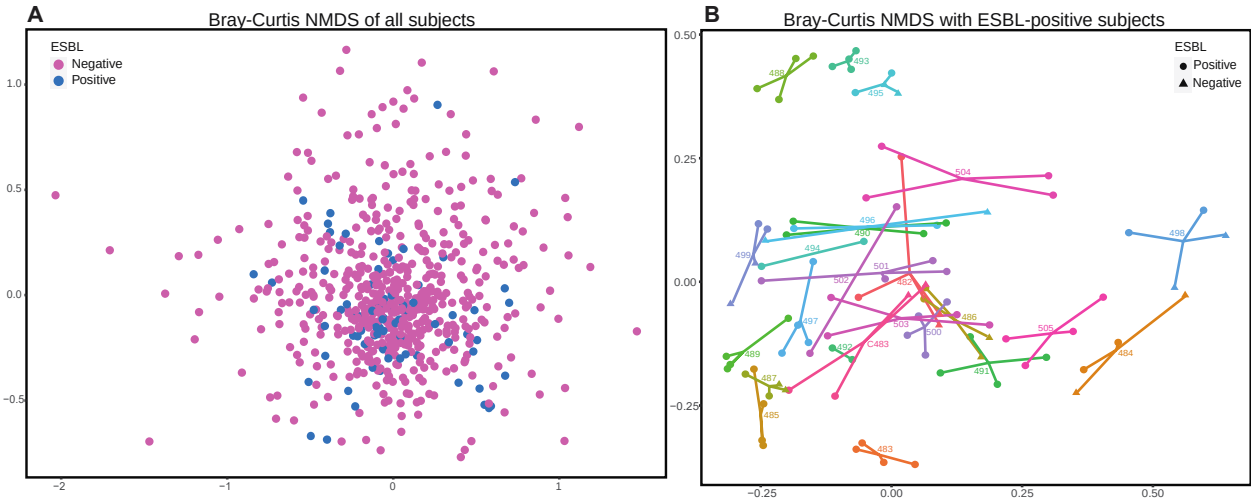

Figure 2.4: Microbial community similarities based on ESBL-E carriage. A) Bray-Curtis distances of metagenomic microbiome samples from all participants, organized in a non-metric multidimensional scaling (NMDS). Magenta dots represent participants that were ESBL-E negative while blue dots represent participants colonized with ESBL-E. B) Comparison of samples (first, duplicate of first, second and duplicate of the second sample) from the same ESBL-E positive participant, calculated by Bray-Curtis distances. Each number indicates the participant ID, while circles or triangles represent the ESBL-E status by sample.

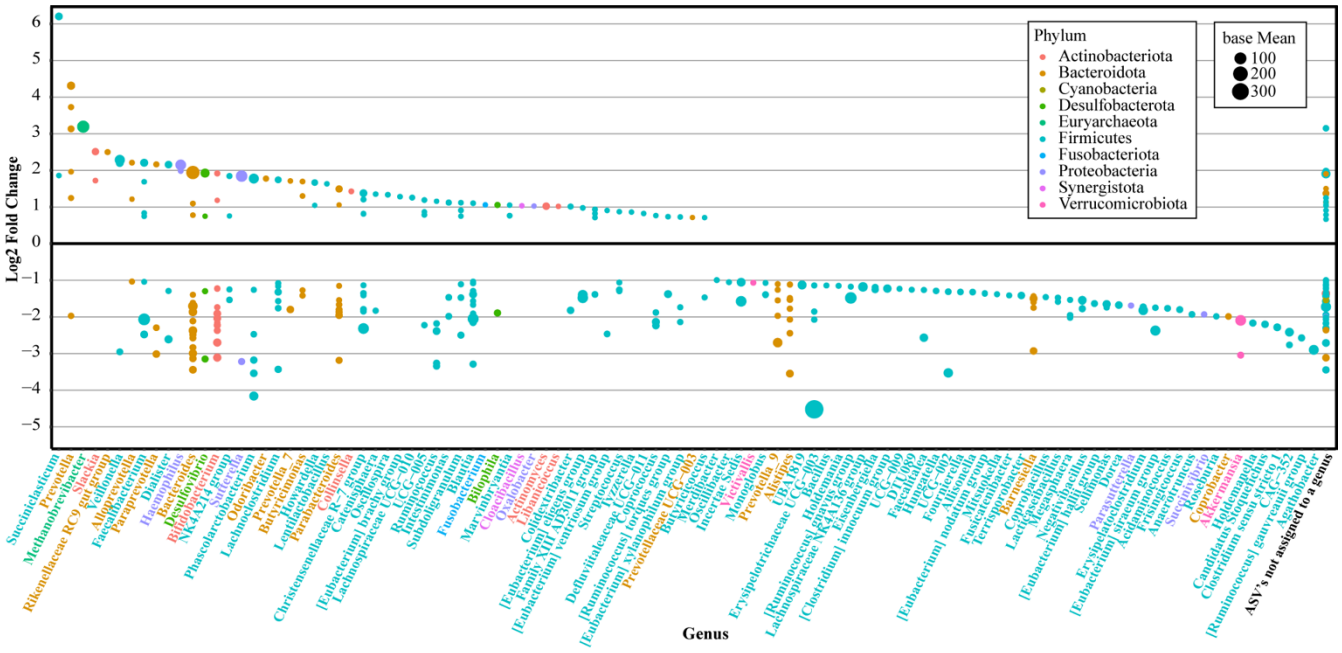

Figure 2.5: Genus assignments for ASVs identified as enriched or depleted in ESBL-E positive participants. ASVs identified as significantly enriched or depleted by DESeq2 analysis with local fit and Wald test identifying significantly different ASVs under an alpha=0.001. ASVs are colored by phylum and organized by genus. Point size shows normalized mean counts and y axis shows log2fold change in ESBL-E positive participants. ASVs with a positive log2fold change were enriched, while ASVs with a negative log2fold change were depleted in ESBL-E positive participants.

CHAPTER 3:

CHARACTERIZATION OF EXTENDED SPECTRUM BETA LACTAMASE-PRODUCING *ENTEROBACTERIACEAE* FROM RECREATIONAL WATER IN THE ATHENS, GA COMMUNITY USING AN UNDERGRADUATE LABORATORY MODULE COMPRISED OF FIVE EXPERIMENTS.[1]

[1]Coralis Rodriguez Garcia, Helen Dukes, Elizabeth A. Ottesen and Julie Grainy. To be submitted to Journal of Biology and Microbiology Education.

3.1 Abstract

We present a laboratory module aimed at introducing undergraduate students to basic microbiological culture-based and molecular techniques through an exploratory project using locally-collected stream water samples, while also educating them on the global public health threat of antibiotic resistant organisms. Through this eight laboratory session module students, with a basic background microbiology and aseptic technique, are involved in quality testing of water sources in their neighborhoods, followed by isolation of extended spectrum beta-lactamase producing Enterobacteriaceae. By the end of the module, students should be able to isolate *Enterobacteriaceae* from the environment using selective and differential media, identify the isolate using biochemical tests, characterize antibiotic resistance phenotypes using Kirby Bauer and MIC tests, and evaluate the presence of select beta-lactamase genes of interest using PCR. To complement the laboratory sessions, students participated in a weekly flipped classroom session with collaborative peer discussions and activities to reinforce concepts applied in the laboratory. Learning outcomes were measured during four semesters with concept checks, in-lecture activities, exams, and laboratory reports. We hypothesized that each of the module learning objectives were being achieved by more than 50% of the student population through the implementation of this authentic research laboratory module. Here we highlight specific questions used to assess each learning objective and demonstrate that 65-100% of the student population achieved each learning objective. We present a ready-to-adapt module with flexible resources that can be implemented in courses across disciplines in biology, microbiology, environmental sciences, and public health.

3.2 Introduction

Extended spectrum beta-lactamases (ESBLs) confer antibiotic resistance to multiple beta-lactam antibiotics frequently used in healthcare settings to treat infections. They are ubiquitously found

in environments from water sources, wastewater, and animals (1–3). The production of these enzymes is mostly associated with microorganisms from the *Enterobacteriaceae* family that contains commensal members of the human gut that can easily carry and exchange resistance genes through plasmids (4, 5). ESBL-producing-*Enterobacteriaceae* (ESBL-E) are one of the main drivers of the antibiotic resistance spread worldwide (6). Therefore, infections with ESBL-E have been listed by the CDC as a serious threat (7). The health impact and environmental distribution of ESBL-E highlights the importance of creating awareness about these enzymes responsible for the resistance.

In this laboratory module the students participate in hands-on experiments about ESBL-E in local water sources. The motivation for this module was to train students in basic microbiological techniques for selective isolation and biochemical characterization microbes as well as phenotypic and genotypic approaches to antimicrobial resistance typing. To encourage student engagement with the laboratory module, students will isolate and characterize novel organisms from unamended water samples collected locally. A secondary goal was to raise awareness of antimicrobial resistance as a global public health threat. The module presents a unique connection between water quality and multidrug-resistant pathogens in the community. Through 8 laboratory sessions (2 hours each) the students perform five laboratory exercises (Fig. 3.1) that include bacterial water quality testing from recreational water sources in their community, paired with isolation of ESBL-E through selective and differential media, species identification of those isolates with biochemical tests, phenotypic testing of antibiotic resistance profiles and molecular tests for the presence of selected beta-lactamase genes potentially responsible for that phenotype.

This module is currently implemented with a flipped classroom lecture (8, 9) scheduled prior to the laboratory exercises which provides students with online content of videos and laboratory instructions for the week. The 50 minute in-person lecture time involves a weekly meeting of students and instructor(s) during which all lab sections meet as a combined group to discuss previous results, test content knowledge, and work through problem-based learning activities in small groups while the instructors are available to moderate and clarify discussions. At the end of the module, the students present a guided project report to practice scientific writing and critical thinking.

3.2.1 Intended Audience

This module was designed for undergraduate students with a life science major, enrolled in an Introductory Microbiology Laboratory course. However, it could be used in any course that includes a lesson about water quality, antibiotic resistance, or characterization of environmental microorganisms (10–12). The exercises do not require any previous skills beyond familiarity with aseptic handling of bacterial cultures and BSL-2 safety precautions but are recommended to upper-level students with an introductory biology or microbiology knowledge. The number of students per class varies by semester but typically includes between 30-60 students, with laboratory sessions of 20 students or less, with students working in groups of 3-4 to complete the exercises but is scalable with appropriate facilities and equipment. The flipped classroom combined all lab sections for large group discussions and small breakout group activities of 3-4 students, different from their lab groups to compare results.

3.2.2 Learning Time

The current implementation of this module occurs over the course of 4 weeks with a weekly meeting for lecture (50-75 minutes) and two laboratory meetings per week (totaling 4 hours of lab

per week), although it could be adjusted to courses with laboratory sessions that only meet once a week (Fig. 3.1). The flipped classroom approach requires the students to spend approximately 60 minutes reviewing the material by themselves prior to class. During lecture time, the first ~5 minutes are spent reviewing the goals and learning objectives for the week, followed by a 5-minutes concept check about the topic of the week, from the material that students previously engaged with before class. After that, there is a ~15-minute student-led post-lab discussion. Finally, the remaining ~25 minutes are used for the learning activities completed in breakout groups. The distribution of time during laboratory sessions varies according to the exercise of the week but generally start with 10-15 minutes for interpretations of previous results, followed by a brief overview of the tasks for the day and then the experimental execution. The 50 minutes pre-lab lecture class meeting could be converted to be part of the lab sessions and/or an out-of-class pre-lab assignment if your course does not include a separate lecture class meeting.

3.2.3 Prerequisite student knowledge

In terms of technical knowledge, this module requires the basics of aseptic techniques, microbial growth, microscopy, and biosafety level 2 practices, usually accomplished from a previous introductory course in Biology or Microbiology or during the first few weeks of the course prior to starting the module. For conceptual knowledge, the students should have a prerequisite or corequisite Biology or Microbiology course to understand concepts like the metabolic properties for differential growth on selected media, the steps of a PCR, and molecular mechanisms of antibiotic resistance. In addition, the students need to review the weekly topic content (videos and readings) prior to starting each exercise.

3.2.4 Learning objectives

By the end of this module students should be able to 1) isolate a pure culture of environmental microbes, 2) determine the identity of their isolate with biochemical tests and 3) evaluate antibiotic resistance in their isolate. Objectives for Exercise 1 include to understand the principles of using fecal indicator bacteria for water safety monitoring, how selective and differential media can be used to isolate specific bacteria from the environment, the use of most probable number, and to determine water safety levels for recreation and consumption. Exercise 2 aims to fulfill the objectives of understanding the difference between ESBL, ESBL-E and *Enterobacteriaceae* and to describe why they are important, while Exercise 3 expects students to identify their isolate using biochemical properties and to understand how to read an EnteroPluri test tube with the variations in phenotype. Exercise 4 and 5 focus on characterization of antibiotic resistance. The learning objectives for Exercise 4 include to interpret antibiotic resistance phenotypes of their isolate using Kirby Bauer tests and MIC test strips with or without the presence of ESBL-inhibiting clavulanic acid. The learning objectives for Exercise 5 involve describing the molecular mechanisms of polymerase chain reaction (PCR), and gel electrophoresis, and applying the approach to the genotype characterization of their water isolate. The complete list of learning objectives for each exercise are included in the student and faculty instructions available upon request.

3.3 Methods

This module includes 5 laboratory exercises performed over 8 lab sessions (each 2 hours) as indicated in Figure 3.1. Specific methods for implementing the exercises are in student instructions and faculty notes available upon request.

3.3.1 Materials

The first exercise starts with an environmental water sample, collected from a stream that is likely to carry a high number of antibiotic resistance genes (1, 13). The dilution of this sample requires 1 ml sterile pipettes and pipettor as well as 9 ml dilution blanks. Students will test the water quality using tubes with Lauryl Tryptose Broth (LTB), Brilliant Green Lactose Bile (BGLB) broth and E. coli broth (EC) that will be used to calculate the Most Probable Number (MPN). Exercise 2 involves the filtration of the same environmental water sample to concentrate the bacteria for enrichment followed by isolation. Each group will use 47 mm glass fiber filters previously autoclaved, 0.5g of autoclaved cellulose pool-filter fiber in individual test tubes and 25 ml of peptone broth to prepare the pre-enrichment, 15 ml sterile H2O in test tubes, sterile forceps, 2 sterile 100 ml specimen cups with lids and 100 ml of sterile ddH2O with a filter apparatus described in the student and faculty instructions (available upon request). The following days of Exercise 2 require ESBL CHROMAgar plates and MacConkey agar plates, in addition to the pre-enrichment sample prepared on day 1 of this exercise. The following Exercise 3 uses the prospective ESBL-E isolate from Exercise 2 and requires oxidase reagent, sterile cotton swabs, Gram stain supplies and Entero-Pluri test with alpha-naphthol, KOH and Kovac's reagent, to characterize and identify species of the isolate. Exercise 4 aims to further characterize the same isolate by testing the antibiotic resistance phenotypically using Mueller-Hinton agar plates and antibiotic disks containing: ampicillin, cefotaxime, imipenem, meropenem and tetracycline. In addition to the antibiotic disks, the Minimal Inhibitory Concentration (MIC) and ESBL mechanism of resistance is tested with MIC strips of cefotaxime and clavulanic acid. The students will need a ruler to mark the correct spacing between antibiotic disks and to measure the results. On the last exercise, students will investigate the antibiotic resistance genotype of the ESBL-E isolate using

PCR. The PCR reagents required for this exercise include: GoTaq G2 Hot Start Master Mix 2X, 10µM primers, nuclease free water, ESBL-E isolated colonies, TE buffer, PCR tubes, microcentrifuge tubes, PCR tube rack, micropipettes and tips, ice bucket, thermocycler, 100 bp DNA ladder, 6X DNA loading dye, 50 ml pre-cast 1% agarose gel containing 1.5 µl Ethidium Bromide, 1X TAE running buffer, gel electrophoresis rig, electrophoresis power supply, and UV lamp and/or gel imager to visualize the DNA in gel.

3.3.2 Student Instructions

Students are provided with an electronic file grouped by exercise that they print and bring to lab in a three-ring binder easily able to be disinfected. The learning goals for the module are listed at the beginning of the lesson with the schedule of experiments in the module. Each exercise in the file includes background information on the topic, the materials available, and the procedure for each experiment. In addition, there is a designated space to record results and questions to guide an individual and group discussions about the results. Our most up-to-date version of the student instructions for this module are available upon request and can be modified for your own use.

3.3.3 Faculty Instructions

The faculty instructions correlate to the student instructions with the addition of material preparation guidance, tips and tricks, and possible modifications. To identify a prospective water collection site, we recommend small streams, running through older areas of a town or a densely populated area, past animal facilities or hospitals, or receiving wastewater treatment plan effluent. The techniques laid out here could also be used for isolation of organisms directly from wastewater treatment plant influent or effluent. We specifically targeted sites previously identified with a high antibiotic resistance gene content, including beta-lactamases (13, 14), although a recent implementation successfully incorporated student-collected samples. A specific faculty suggestion

for implementation of this module includes the preservation of a self-collected positive control. During the first semester executing this module, a positive control might not be available, but after the initial isolation and confirmation of an ESBL-E isolate from a student group, that isolate can be stored as a positive control for future semesters. Our most up-to-date version of the faculty instructions for this module are available upon request and can be modified for your own use.

Suggestions for determining student learning

The assessment techniques we use for this module include formative pre-lab concept checks, lecture activities, post-lab discussions, lab participation as well as summative final project reports and midterm and final exams. The assessment tools we present here were evaluated by points, although the final grade of the course is reported in the standard letter grade scale (A-F). Reports of the assessment included here correspond to activities completed during lecture and the protocol to collect these results was approved by the Institutional Review Board (PROJECT00007873) at the University of Georgia.

3.3.4 Sample data

A collection of sample results is shown in Figure 3.1. In more detail, during Exercise 1, most groups found MPN numbers in their water samples below the Georgia standard for water quality. Common misinterpretations we noticed were related to the distinction between total coliforms and confirmed coliforms. CHROMAgar-ESBL plates used in Exercise 2 yielded a high number of colored colonies indicating bacteria that were resistant to the antibiotic mix in this media. Students usually selected blue and beige colonies for further isolation and characterization. Results for EnteroPluri test tubes in Exercise 3 were a quick and efficient way to identify the isolates. As a note, we found that EnteroPluri results were most reliable when read immediately after incubation. For implementation in a teaching laboratory, we recommend storing the tubes at 4°C after initial

inoculation, then transferring to a 37°C incubator approximately 24 hours before the next class period. Some common organisms identified through the EnteroPluri tests were in the *Acinetobacter*, *Kluyvera*, *Serratia*, *Citrobacter*, *Enterobacter*, and *Escherichia* genera. In Exercise 4, Kirby Bauer ran smoothly, with most organisms tested showing resistance to at least ampicillin. Those that had resistance to cefotaxime in the Kirby Bauer test were also confirmed as Cefotaxime resistant in the MIC test strips. Some semesters resulted in many isolates showing Cefotaxime resistance through a non-ESBL mechanism, while other semesters show Cefotaxime-susceptible isolates in the Kirby Bauer and MIC strips. Lastly, PCR results from Exercise 5 were mixed, with some showing common ESBL-genes that confirm the phenotypic profile, while others did not show amplification of any of the genes tested, regardless of their antibiotic resistance profile observed with Kirby Bauer. We occasionally see instances where an amplified band is observed with gel electrophoresis that seems inconsistent with the Kirby Bauer sensitive phenotype, in which case we predict that those ESBL genes were present by not expressed or functional during the growing conditions tested.

3.3.5 Safety Issues

This module is focused on the isolation of environmental organisms and the use of possibly contaminated water sources. Given the pathogenic nature of the organisms intended to isolate, this module requires Biosafety Level 2 laboratory and training before the laboratory activities. The laboratory activities performed in this module comply with the ASM Guidelines for Biosafety in Teaching Laboratories and the standard laboratory practices for BSL-2 (15), including the use of long pants, closed shoes, lab coat, gloves and safety glasses/googles. Autoclaved wooden sticks and sterile swabs were used to avoid the use of flame or Bunsen burners to sterilize loops. We

intentionally designed the experiment to primarily grow cultures on agar plates, to minimize the risk of spilling liquid cultures of potential pathogens.

3.4 Results and Discussion

3.4.1 Field testing

This module has been successfully implemented for 5 semesters including: Fall 2021, Spring 2022, Fall 2022, Spring 2023 and Fall 2023. We assessed student responses to the activity via two different survey mechanisms, a one-time survey conducted immediately on completion of the course during the Fall semester of 2022 and a retrospective survey of course participants from all 5 semesters that was conducted in Fall 2023. During Fall 2022, 9 out of 27 students that completed an anonymous course evaluation selected this ESBL module as the lab activity they enjoyed the most, and when compared to the total 5-semester student population that completed a post-module survey, almost 60% agree or somewhat agree with a similar statement (Fig. 3.3H). From the cohort of Fall 2023, 14 of 21 students identified the ESBL module for which lab activity they learned the most from, which is agreed with by more than 60% of the overall students that participated in the additional post-module survey (Fig. 3.3G). Many students noted their appreciation of the multi-week experiments building within the same module. Moreover, students commented that they were captivated by the connection of the experiments to their everyday life and even described it as "depressing" when they learned about the prevalence of antibiotic resistance in their surroundings.

3.4.2 Evidence of student learning

In this study we present results collected from five semesters with a total of 70 students that provided consent to use their grades. The students that signed the consent form were also offered the option to complete a post-module survey, resulting in 36 responses among all semesters. Figure 3.2 presents the average scores on selected questions that reflect one learning objective from each

exercise in the module. Almost all assessment questions that we analyzed, except for exercise 4 in Fall 2023, reached an average grade higher than 60% indicating the achievement of the correlating learning objectives. Exercise 4 includes the interpretation of antibiotic resistance phenotype using Kirby Bauer and MIC test. The question used to test this learning objective (Table 3.1) also incorporates the ability to identify and/or define an ESBL producer. When asked in the post-module survey responses, most students acknowledged that they do not feel confident about their ability to read an MIC test accurately (Fig. 3.3). In contrast, more than 60% agreed they can differentiate an ESBL-producer accurately, which indicates that the lagging concept is the MIC strip interpretation. This could be caused by the frequent lack of classic ESBL-E patterns on the MIC test strips results, requiring a higher level of critical thinking to understand the resistance mechanisms. Moreover, the final written report about this module, had an average grade of 95%, while the lowest value was 73%, achieving a successful outcome in scientific data analysis and writing skills in this scaffolded project report.

On a post-module survey completed by 36 students, where they had the opportunity to opt-out of any question, more than 70% of the students agreed that this module exposed them to new techniques (Fig. 3.3D). The authentic research-based techniques students are exposed to in this module can easily fulfill the requirements of a course-based undergraduate research experience (CURE) (16). Close to 50% agreed that the module made them feel like a real-world scientist, which is a satisfactory outcome for undergraduates that have not had research exposure, but it also offers the opportunity for potential improvements in the science identity in undergraduate students on the course (Fig. 3.3J) (17, 18). Science identity plays an important role in the retention of students in the STEM fields, especially those within underrepresented communities (19, 20). Strategies to potentially improve the students' science identity include more leadership roles in the

experiments and assignments that highlight relatable scientists or even student peers (21–23). However, this module achieved the goal of increasing antibiotic resistance awareness as agreed by more than 70% of the students. This accomplishment contributes successful efforts to other similar initiatives towards this important topic (24–27).

3.4.3 Potential modifications

The implementation of this module could be modified according to the days available. Exercises 1 and 2 use the same environmental water sample but they are not dependent on each other. Exercises 3, 4 and 5 can be done independent of each other, but require the isolates generated after Exercise 2. Student instructions are designed to investigate the presence of CTX-M, TEM-1, and CMY-2 genes but could be modified to identify any number of ESBL genes, by choosing a different set of primers or even whole genome sequencing (28). In terms of assessment, students could achieve a deeper understanding of this module using the 2-stage exam evaluation (29). Other modification potentials are addressed in the faculty instructions available upon request.

3.5 Acknowledgements

We would like to thank the teaching assistants and peer-learning assistants for their helpful contribution in the collection of images and the successful implementation of this module in the laboratory, as well as the students of MIBO 3510L from the semesters of Fall 2023, Spring 2023, Fall 2022, Spring 2022 and Fall 2021. The authors declared no conflict of interest.

3.6 References

1. Cho S, Nguyen HAT, McDonald JM, Woodley TA, Hiott LM, Barrett JB, Jackson CR, Frye JG. 2019. Genetic characterization of antimicrobial-resistant escherichia coli isolated from a mixed-use watershed in northeast Georgia, USA. Int J Environ Res Public Health 16:1–14.

2. Jørgensen SB, Søraas A V., Arnesen LS, Leegaard TM, Sundsfjord A, Jenum PA. 2017. A comparison of extended spectrum β-lactamase producing Escherichia coli from clinical, recreational water and wastewater samples associated in time and location. PLoS One 12:1–15.

3. Iseppi R, Di Cerbo A, Messi P, Sabia C. 2020. Antibiotic resistance and virulence traits in vancomycin-resistant enterococci (Vre) and extended-spectrum β-lactamase/ampc-producing (ESBL/ampc) enterobacteriaceae from humans and pets. Antibiotics 9.

4. Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, Nolan LK. 2007. Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates. Appl Environ Microbiol 73:1976–1983.

5. Rawat D, Nair D. 2010. Extended-spectrum ß-lactamases in gram negative bacteria. J Glob Infect Dis 2:263.

6. Bush K, Bradford PA. 2020. Epidemiology of B-Lactamase-Producing Pathogens. Clin Microbiol Rev 33:1–37.

7. CDC. 2019. Antibiotic Resistance Threats in The United States, 2019U.S. Department of Health and Human Services.

8. Rodríguez G, Díez J, Pérez N, Baños JE, Carrió M. 2019. Flipped classroom: Fostering creative skills in undergraduate students of health sciences. Think Skills Creat 33.

9. Hotle SL, Garrow LA. 2016. Effects of the Traditional and Flipped Classrooms on Undergraduate Student Opinions and Success. Journal of Professional Issues in Engineering Education and Practice 142.

10. Foster-Hartnett D, Mwakalundwa G, Henry G, Bofenkamp L, Dvorak CMT, Lalande A, Johnson TJ. 2023. FUTIs: an In-Person or Online Graphing, Bioinformatics, and Scientific

Literacy Exercise That Explores the Presence of Antibiotic Resistance in Foodborne Urinary Tract Infections. J Microbiol Biol Educ https://doi.org/10.1128/jmbe.00045-23.

11. Fuhrmeister ER, Larson JR, Kleinschmit AJ, Kirby JE, Pickering AJ, Bascom-Slack CA. 2021. Combating Antimicrobial Resistance Through Student-Driven Research and Environmental Surveillance. Front Microbiol 12.

12. Weber CF. 2014. Hormones and Antibiotics in Nature: A Laboratory Module Designed to Broaden Undergraduate Perspectives on Typically Human-Centered Topics. J Microbiol Biol Educ 15.

13. Damashek J, Westrich JR, McDonald JMB, Teachey ME, Jackson CR, Frye JG, Lipp EK, Capps KA, Ottesen EA. 2022. Non-point source fecal contamination from aging wastewater infrastructure is a primary driver of antibiotic resistance in surface waters. Water Res 222:118853.

14. Cho S, Hiott LM, Barrett JB, McMillan EA, House SL, Humayoun SB, Adams ES, Jackson CR, Frye JG. 2018. Prevalence and characterization of Escherichia coli isolated from the upper oconee watershed in Northeast Georgia. PLoS One 13:1–15.

15. Byrd JJ, Emmert E, Maxwell R, Townsend H. 2019. Guidelines for Biosafety in Teaching Laboratories Version 2.0: A Revised and Updated Manual for 2019. J Microbiol Biol Educ 20.

16. Parks S, Joyner JL, Nusnbaum M. 2020. Reaching a Large Urban Undergraduate Population through Microbial Ecology Course-Based Research Experiences. J Microbiol Biol Educ 21.

17. Adkins-Jablonsky SJ, Akscyn R, Bennett BC, Roberts Q, Morris JJ. 2020. Is Community Relevance Enough? Civic and Science Identity Impact of Microbiology CUREs Focused on Community Environmental Justice. Front Microbiol 11.

18. Trujillo G, Tanner KD. 2014. Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. CBE Life Sci Educ 13.

19. Le PT, Doughty L, Thompson AN, Hartley LM. 2019. Investigating undergraduate biology students' science identity production. CBE Life Sci Educ 18.

20. McCartney M, Roddy AB, Geiger J, Piland NC, Ribeiro MM, Lainoff A. 2022. Seeing Yourself as a Scientist: Increasing Science Identity Using Professional Development Modules Designed for Undergraduate Students. J Microbiol Biol Educ 23.

21. Schinske JN, Perkins H, Snyder A, Wyer M. 2016. Scientist Spotlight Homework Assignments Shift Students' Stereotypes of Scientists and Enhance Science Identity in a Diverse Introductory Science Class. CBE—Life Sciences Education 15:ar47.

22. Musgrove MMC, Ko ME, Schinske JN, Corwin LA. 2024. Broadening Participation in Biology Education Research: A role for affinity groups in promoting social connectivity, self-efficacy, and belonging. CBE—Life Sciences Education 23.

23. Aranda ML, Diaz M, Mena LG, Ortiz JI, Rivera-Nolan C, Sanchez DC, Sanchez MJ, Upchurch AM, Williams CS, Boorstin SN, Cardoso LM, Dominguez M, Elias S, Lopez EE, Ramirez RE, Romero PJ, Tigress FN, Wilson JA, Winstead R, Cantley JT, Chen JC, Fuse M, Goldman MA, Govindan B, Ingmire P, Knight JD, Pasion SG, Pennings PS, Sehgal RNM, de Vera PT, Kelley L, Schinske JN, Riggs B, Burrus LW, Tanner KD. 2021. Student-Authored Scientist Spotlights: Investigating the Impacts of Engaging Undergraduates as Developers of Inclusive Curriculum through a Service-Learning Course. CBE Life Sci Educ 20.

24. Elmahi OKO, Uakkas S, Olalekan BY, Damilola IA, Adedeji OJ, Hasan MM, dos Santos Costa AC, Ahmad S, Essar MY, Thomson DJ. 2022. Antimicrobial resistance and one health in

the post COVID-19 era: What should health students learn? Antimicrob Resist Infect Control https://doi.org/10.1186/s13756-022-01099-7.

25. Genné-Bacon EA, Bascom-Slack CA. 2018. The PARE Project: A Short Course-Based Research Project for National Surveillance of Antibiotic-Resistant Microbes in Environmental Samples. J Microbiol Biol Educ 19.

26. Maicas S, Fouz B, Figàs-Segura À, Zueco J, Rico H, Navarro A, Carbó E, Segura-García J, Biosca EG. 2020. Implementation of Antibiotic Discovery by Student Crowdsourcing in the Valencian Community Through a Service Learning Strategy. Front Microbiol 11.

27. Alvarado M, Clemente-Casares P, Moreno DA, de Groot PWJ. 2020. MicroMundo Upside Down: Targeted Searching for Antibiotics-Producing Bacteria From Soil With Reverse Antibiosis Approaches. Front Microbiol 11.

28. Greenman NA, Jurgensen SK, Holmes CP, Kapsak CJ, Davis RE, Maza WM, Edemba D, Esser BA, Hise SM, Keen TN, Larson HG, Lockwood DJ, Wang B, Harsh JA, Herrick JB. 2021. Genomics of Environmental Salmonella: Engaging Students in the Microbiology and Bioinformatics of Foodborne Pathogens. Front Microbiol 12.

29. Meaders CL, Vega Y. 2023. Collaborative Two-Stage Exams Benefit Students in a Biology Laboratory Course. J Microbiol Biol Educ 24.

Table 3.1: Selected questions to measure the learning objectives associated with each exercise. The same questions were used in most semesters if available. Variations of the question between semester are included below. Average scores for each question are presented in Figure 3.2.

Ex. #	Question
1	You have an MPN code of 5-2-3, and want to calculate the MPN using the formula in your lab notes/lecture. What is the total volume of original sample (in mL) in negative tubes (Vn in the formula)? Do not round _____ Based on that formula, what is the MPN/100 mL? Round to the nearest integer (no decimal places).
2	Many Enterobacteriaceae carry their antibiotic resistance genes on mobile genetic elements that are easily exchanged with other Enterobacteriaceae. Based on this, explain why public health scientists find it worrying to find even non-pathogenic Enterobacteriaceae carrying ESBL genes in local surface waters.
3	You perform an EnteroPluri test and obtain a code of 64340. Using the following picture, which of the following properties is this organism positive for? Select all that are positive.
4	Your Kirby Bauer assay shows resistance to both ampicillin and cefotaxime, and your MIC test strip assay results in growth all the way along the strip. Version 1: Can you confirm that this organism is an ESBL-producer? Why or why not? Version 2: Is this organism resistant to cefotaxime? How do you know this? Can you confirm that this organism is carrying an ESBL gene? Why or why not?
5	If the TEM positive control lane of the agarose gel has a 1100 bp DNA fragment, the negative control lane has no band, and your experimental lane has a 300 bp DNA fragment, what does this mean?

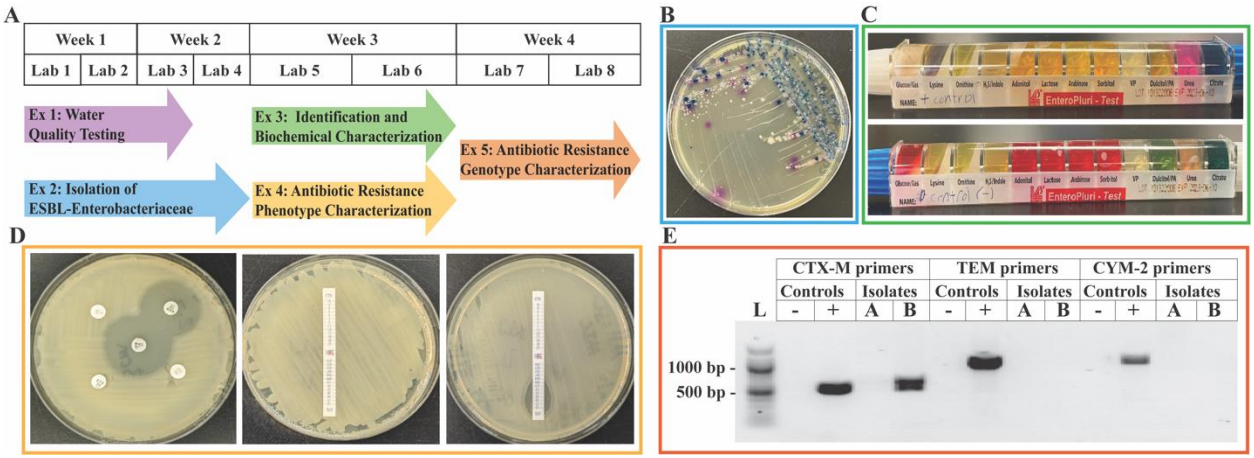

Figure 3.1: Module timeline and sample data collected from students' results and positive control. A) This module includes 5 different overlapping exercises that can be completed in 8 lab sessions (2 hours each). B) Bacterial colonies from an initial streak for isolation growing on CHROMAgar-ESBL. Students continue to re-streak until a pure culture is obtained. Color interpretations are listed in Student Instructions in Appendix 1. C) EnteroPluri test results of an ESBL-E positive control on top and negative control on bottom. Tubes were stored at 4ºC and then incubated at 37ºC for 24 hours prior to the reading of results. Picture taken before addition of VP and Indole reagents. D) Sample results from antibiotic susceptibility tests via the Kirby Bauer disk diffusion assay and MIC test strips. Antibiotic disks used include Ampicillin, Tetracycline, Cefotaxime, Imipenem and Meropenem. MIC strips contained Cefotaxime at a concentration gradient and Cefotaxime with Clavulanic Acid on the other side in a concentration gradient as well. The first MIC test example exhibits Cefotaxime resistance, but by a non-ESBL mechanism, while the last example shows a classic MIC result for an ESBL isolate with resistance on the Cefotaxime side (top of strip) and inhibition of that resistance with the addition of clavulanic acid (bottom of strip). E) Agarose gel electrophoresis with 100 bp DNA ladder, controls, and student isolates to investigate the presence or absence of CTX-M, TEM and CMY-2 genes.

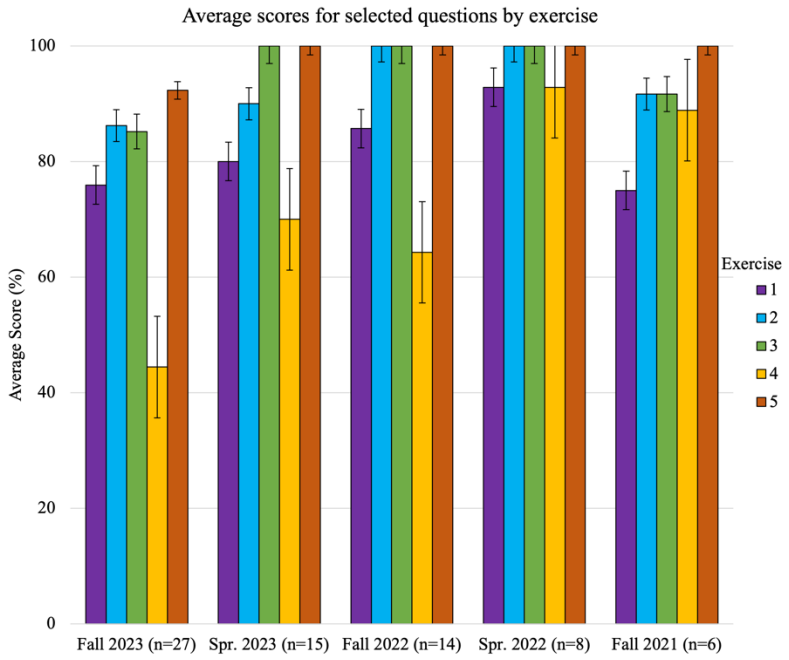

Figure 3.2: Average scores on selected assessment questions for each module exercise. Scores were collected from the same question or slight variation of the same question each semester. Question types include multiple selection, short answer, and written response. Colors indicate the corresponding exercise, while "n" indicates the number of students that participated by semester.

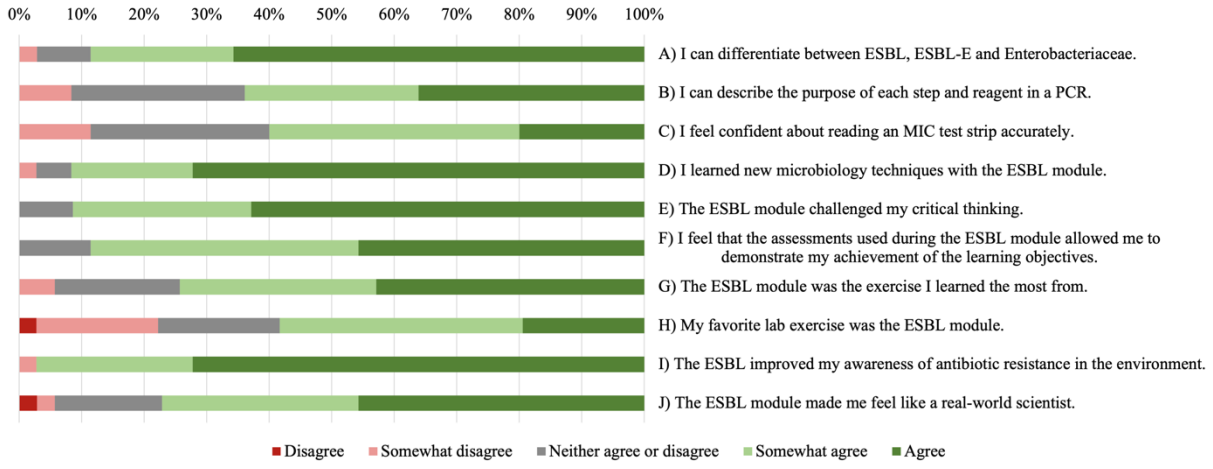

Figure 3.3: Student self-assessment of learning on a post-module survey. Among the 70 students that provided consent for this study, each survey question was answered by 35-36 participants, some of them with a gap of 2 or more years after participating in the module.

CHAPTER 4

MICROBIAL TRANSCRIPTIONAL ACTIVITY IN STREAM WATER DOES NOT PREDICT MICROBIAL ENRICHMENT OR DEPLETION ALONG THE LENGTH OF THE STREAM[1].

[1]Coralis Rodriguez Garcia, Morgan Teachey and Elizabeth A. Ottesen. To be submitted to PLOS ONE.

4.1 Abstract

Streams and rivers harbor microbial communities whose dynamics can affect essential ecosystem services. These impact critical human uses including sanitation, agriculture and recreation. Therefore, it is important that we understand the factors influencing the microbial community of these environments. Stream microbial communities are highly complex and dynamic, including both freshwater-associated taxa as well as diverse organisms originating from surrounding soil and sediment environments. As water moves through the watershed, the proportion of freshwater-associated taxa typically increases while the proportion of non-freshwater microbes typically decreases. However, how these selection dynamics happen and the degree to which these non-freshwater taxa are transcriptionally and metabolically active during their residence in the stream remains unclear. To address some of these questions, we used metagenomics, metatranscriptomics and single-cell genomics to characterize the abundance and activity of microbial communities in McNutt Creek. We found that metagenomic libraries included a large number of taxa that are not present in RNA samples, suggesting the presence of cell-free DNA. In addition, many microbes showed higher abundance in DNA than in RNA libraries, including typical freshwater taxa. We found that microbial transcriptional activity as denoted by RNA: DNA abundance ratios is not a predictor of enrichment or depletion along the creek and that transient taxa exhibit similar levels of transcriptional activity to those of freshwater associated microbes. These results contribute to the understanding of microbial ecology, assembly and functions in freshwater environments and highlights the relevance of omics approaches to assess microbial community dynamics in streams.

4.2 Introduction

Freshwater environments play an essential role in the ecosystem (Shafi et al. 2015; Morin and Artigas, 2023; Crump and Bowen, 2024). Freshwater is critically important for provision of

drinking water as well as for use in agriculture, industry, and recreation. Rivers and streams also serve as crucial ecological pathways that interconnect diverse terrestrial ecosystems as well as connecting terrestrial environments to estuarine and marine systems. Freshwater microbial communities are critically important to in-stream processes including nutrient cycling and remediation after natural and human-derived disturbances (1–3).

The nature of microbes found in freshwater environments has been subject to extensive debate. Early work suggested that the extensive interconnection between water, sediment, and soil environments precluded the development of a unique freshwater community (4). However, extensive work in lake systems has established that these are inhabited by a core of "typical freshwater taxa" that are characteristic of freshwater environments (5–7). These microbes show evidence of in-system selection in response to water characteristics and are believed to represent freshwater specialists (8).

However, the question of where these freshwater taxa come from and what happens to soil and sediment taxa when they are washed into the watershed remains unsettled. In particular, stream and riverine environments host large gradients in microbial diversity linked to position within the watershed, with headwaters showing the highest overall microbial alpha and beta diversity, and downstream environments showing decreased richness and variability in community composition. In a pioneering study, Crump et al (9) showed that the microbial communities of headwater streams are strongly influenced by soil bacteria that flow into the stream, resembling soil water communities more strongly than they do downstream lake communities. As the water moves downstream the soil influence decreases and freshwater-associated taxa increase and eventually predominate (9). A wide variety of other works have similarly identified that many riverine and watershed environments are inhabited by both transient taxa that appear to be environmentally

derived and resident, or core, taxa that show stronger evidence of in-stream or in-lake selection (5, 8–14).

It is not clear if the microbial activity is dominated by core taxa, even in headwater environments where they are present in relatively low abundance or if the transient/soil-associated microbes contribute to the activity and stream function, while present. To answer these questions, we used a collection of 'omic approaches including single-cell genomics, metagenomics and metatranscriptomics to evaluate transcriptional activity among different microbial taxa, and how it changes along the length of a single stream. Our sampling site, McNutt Creek, is part of a mixed-use watershed in Athens, GA, USA. Previous studies of this watershed (15, 16) revealed that the headwater streams are populated with highly diverse groups of soil-associated bacteria that decrease as water moves through the watershed. In this study, we aimed to understand whether and how these longitudinal trends in stream microbial community composition translate into changes in microbial activity.

4.3 Methods

Site description and sample collection

Surface water samples were collected from McNutt Creek in Athens, GA, USA (Fig. 4.1A). The distribution of sampling sites along the creek includes a mixed land use between residential and agricultural areas. This creek is part of the Upper Oconee Watershed, which supplies water to Athens and other cities. Sampling site locations are listed in Table 1.

Water samples for RNA and DNA extraction were collected on August 29, 2017. Cubitainers for water collection were pre-washed with 3x 12 hour soaks with 1% HCl followed by 3 rinses and 1 overnight soak with double-distilled water. Water was collected mid-stream and mid-depth, then filtered on site using MasterFlex L/S modular peristaltic pumps (Cole Palmer) and silicone tubing

(Fisher). Before each sample, pump lines were rinsed with approximately 250 ml of 2% bleach solution, followed by at least 500 mL of ddH2O. Water samples were prefiltered using a 47 mm 5 µm Durapore PVDF membrane pre-filter (MilliporeSigma), before capture of microbial samples on a 0.22 µm Durapore PVDF Sterivex cartridge filter (MilliporeSigma). Filtering times varied from 10- 15 minutes. The volume of water filtered was measured using a graduated cylinder. After filtration was complete, each Sterivex was detached from the line and shaken to remove excess water. A volume of 0.3 mL RNAlater was then added immediately to filters intended for future RNA extraction. Filters were store on ice for transport to the laboratory. After all sites were sampled, pre-filters and Sterivex were returned to the lab and stored at -80ºC, typically within 4 hours of collection.

Stream water samples for Single Cell Particle Assembly were collected on July 13, 2017 using 50 ml conical tubes rinsed 3 times with stream water. (max. rainfall on previous 8 days: 0.15 combined from 2 days) from the same locations as site 1, 4, and 6 with the addition of site after the merge of 2 main rivers in this watershed (Middle Oconee and North Oconee). Tubes were stored on ice and transported to the lab. Following Bigelow protocol, 1 ml of sample was transferred to each cryovial and 100 µl of glycerol-TE added. Sample was mixed gently and incubated for 1 minute at room temperature. Cryovials were stored at -80ºC until shipment.

Single Particle Sort and Sequencing

Cryo-preserved samples for Single Cell Genomes were sorted using the FACs Cell Sorting technique at the Joint Genome Institute (Stepanauskas et al. 2007). Sorted particles were then sequenced with Illumina NextSeq HO, also at JGI. SCGs were assembled by SPAdes (v3.10.1; phred-offset 33 -t 16 -m 115 -sc -k 25,55,95 -12). Then 200bp was trimmed from all contig ends and contigs discarded if the length was <2kbp or read coverage was less than 2 (BBMap: nodisk

ambig, filterbycoverage.sh: mincov). Genome completeness and contamination was calculated by CheckM. Taxonomic assignments were performed in KBase with the Genome Taxonomy Database Toolkit (GTDB-Tk).

RNA and DNA Extraction protocol

DNA extraction with Sterivex filters was performed according to the methods in Teachey et al. (16), followed by additional steps with the Omega BioTek Water DNA kit (step 13 to completion) to improve DNA yield. Extracted DNA was quantified by NanoDrop, pooled by sample site and used for metagenomic reads.

RNA extraction utilized an adaptation of the Omega BioTek HP Total RNA Kit protocol. Filters were thawed and the RNAlater removed. Sterivex filters were then crushed inside a heavy Ziplock bag until the bottom portion of the filter separated from the top. A pair of tweezers were cleaned with 5% bleach, ddH_2O, then ethanol and flame, then used to remove the filter out of the Sterivex casing. The filter was transferred into a Whirl-Pak bag with 250µl of 30mg/ml lysozyme in buffer and incubated at 30ºC for 10 min. with occasional movement of the lysozyme around the bag. The filter and lysate were then transferred into a 50 mL conical tube containing 1 g of 0.1mm zirconia/silica beads (previously baked at 180ºC for at least 2 hours to inactivate RNAses) and 2040 µl of GTC lysis buffer (with added beta-mercaptoethanol as per manufactuerer instructions). Tubes placed in a bead-beater for 30 seconds, followed by 10 seconds on ice. This step was repeated in 10 cycles of bead-beating and ice until reaching 5 minutes. The lysate was transferred into a 5 mL tube without the beads. The lysate was then passed through the RNA homogenizer in 700µl aliquots as per manufacturer instructions. Following this step, the filtrate was mixed with 1X volume of 70% ethanol and transferred into a HiBind RNA mini column in 700µl aliquots. Remaining clean-up steps followed kit instructions until elution with 50µl of 55ºC water. RNA

concentration and quality was measured by NanoDrop. Aliquots from both filters corresponding to the same site were pooled, stored at -80ºC and shipped to JGI on dry ice for library preparation and sequencing.

Genomic sequencing and analysis

Metagenomes and metatranscriptomes were generated at the DOE Joint Genome Institute (JGI). Meta-transcriptome libraries were prepared following JGI protocols for Illumina Ultra-Low Input RNASeq w/rRNA Depletion. Metagenomic libraries were prepared using the Joint Genome Institute protocols for Illumina Low Input Fragment 300bp. All libraries were sequenced using Illumina HiSeq-2000 (2x151bp reads). All Illumina sequence data were processed by their SOP 1056 and filtered with BBTools. As part of their pipeline, they used BBDuk (version 37.64) to filter out reads with adapters, reads with quality at 0, reads with 4 or more "N", reads shorter than 51 bp or reads with an average quality score less than 3.

Processing of metagenome and metatranscriptome reads

Quality-filtered metagenome and metatranscriptome reads were merged using BBMerge from BBMap v 38.90. Unmerged reads were recovered and 'merged' by inserting a series of 9 Ns between the forward read and the reverse complement of the reverse read. This protocol allowed the unmerged reads to be treated similarly to merged reads for annotation by homology searches, with the inserted Ns blocking false alignment over the read gap. Merged metatranscriptome reads were run through SortMeRNA to remove ribosomal RNA sequences.

Merged metagenome and metatranscriptome reads were mapped using DIAMOND 2.0.9 (17) to both the RefSeq protein database (release 205) and a custom database of all open reading frames from our single cell genomes. These results were combined, and top hits extracted, defined as the top-scoring hit (by bitscore) across both databases, as well as any hits with equal score. Each read

was then assigned to a single, unique hit via a custom script that first counted unambiguous hits (reads with only one top-scoring hit), then assigned reads with multiple, equal-scoring hits to a single hit based on hit abundance across the full dataset.

Analysis of metagenome and metatranscriptome taxonomic composition

To count taxon abundance, the resultant single top hit files were matched to the NCBI taxonomy database based on species name and counted accorded to taxID number. Reads mapping to proteins from our single cell genomes were assigned to taxIDs generated based on the taxonomy assignment we generated as described above. The taxonkit script, provided by NCBI, was used to extract the full taxonomic hierarchy of each hit. A hit count file with the number of reads from each file that was assigned to each unique taxonomy database entry was used for further analysis and visualizations in R using packages dplyr (18), tidyverse (19), phyloseq (20), ggplot2 (21), VennDiagram (22), MicEco (23) and MicrobiotaProcess (24).

Classified MG and MT samples were rarefied to a sample size of 2,814,029, then split between groups by dataset (MG or MT), to analyze the taxonomic distribution, community richness and beta diversity. Alpha diversity was calculated using the Shannon index while beta diversity was calculated using Bray-Curtis distances. Most abundant phyla across all samples were selected by filtering the minimum relative abundance of 1%. Abundant taxa were selected by filtering taxa present in more than 2 samples with a minimum number of reads of 20,000. Correlation coefficients were calculated using the two-sided Spearman method comparing their abundance in the metagenome to the distance in the creek (sites 1 to 6).

4.4 Results

Single Cell-amplified Genomes

Water was collected from 4 sites on July 13, 2017 for single particle sorting and genome sequencing. These sequencing efforts resulted in 294 single-cell amplified genomes (SCGs), from which 276 had between 10-88% of completion by CheckM (supplemental Table S4.1). The most abundant phylum identified in the SCG was *Pseudomonadota* with 91 genomes including common freshwater members like *Limnohabitans* (42 genomes), *Polynucleobacter* (16 genomes), and *Rhodoferax* (1 genome). The second most abundant phylum in SCG was *Actinomycetota* with 60 genomes, including some assigned to *Planktophila* (31 genomes) and *Rhodoluna* (5 genomes). A high number of SCGs (43 genomes) were classified in the Candidate Phyla *Patescibacteria*.

Metagenomic and metatranscriptomic sequencing

Water was collected for metagenomic and metatranscriptomic sequencing from 6 sites along the length of McNutts Creek in Athens, GA on August 29, 2017. Metadata collected included water temperature, pH, dissolved oxygen, and conductivity (Table S4.2). Water temperature varied by 1.89ºC across sites, and with the highest temperatures measured at downstream sites. The pH range was from 6.88 (site 2) to 7.41 (site 5). Dissolved oxygen concentrations ranked from 9.09 mg/L (site 1) to 11.62 mg/L (site 6) and conductivity ranged from 0.06 (site 6) to 0.113 (site 5).

Metagenomes and metatranscriptomes ranged from 26,966,978 filtered reads in site 1 of the MT and 118,485,616 filtered reads in site 3 of the MG. Approximately 10% (in MT) to 39% (in MG) of reads were assigned to a single unique taxonID (Supplemental Table 2). The taxonomic composition of metagenome and metatranscriptomes were broadly similar, with *Pseudomonadota*, *Actinomycetota* and *Bacteroidota* dominating the recovered taxa (Fig. 4.1B). Metatranscriptomes typically showed very high overlap in taxonomic composition with their paired metagenomes,

while metagenomes often included a large number of taxa absent in their paired metatranscriptome (Fig. 4.1C). This aligns well with alpha diversity measures, with metagenome samples also having higher Shannon Entropy indices than their paired metatranscriptome samples in most cases, except site 4 and 5 where metatranscriptome samples showed slightly higher alpha diversity (Fig. 4.1D). Principal coordinate analysis (PCoA) clustered samples primarily according to their position in the creek along the first principal component and then by dataset type (metatranscriptome vs. metagenome) along the second principal component (Fig. 4.1D). This was corroborated by a non-metric multidimensional analysis with envfit test using the vegan package in R (25) that resulted in an R-squared of 0.86 (p-value= 0.001) for distance in the creek, while the differences in dataset types resulted in an R-squared value of 0.20 (p-value = 0.12).

Metabolic activity by site

Among shared taxa, we observed a strong linear relationship between metatranscriptome and metagenome abundance (Figure 4.2a). The slope of that relationship was typically close to 1 (0.68-0.88) but the intercept was always negative, indicating that taxa were nearly always more abundant in the metagenome than the metatranscriptome. For most sites, the R-squared values ranged between 0.67 and 0.83, with the exception of site 2 at 0.47. This suggest that these two factors are typically related.

Figure 4.2b shows the overall distribution of RNA to DNA ratios for individual taxa. In general, the mean RNA to DNA ratio was below zero, but increased appreciably at site 6 with a slope of 0.29 and a R-squared of 0.18. Reflecting the greater variability seen in Figure 4.2a, site 2 had the greatest dispersion.

Patterns of abundant taxa

We used Spearman correlation to identify taxa that were significantly increasing and decreasing in metagenome abundance (as a proxy for cellular abundance) along the length of the stream. For this analysis, we first filtered the dataset to identify a set of 213 taxa that were present in more than 2 samples and had >20,000 assigned reads. These top 213 taxa were represented approximately 29% to 65% of the relative abundance from each site. Positive correlation (>0.5) coefficients indicate taxa that increase in abundance with increasing distance along the length of the watershed. Taxa showed a range of relationships between relative abundance in the metagenome and position in the stream (Fig. 4.3A), with similar numbers of taxa showing a negative correlation (rho < -0.5; n=88) vs. a positive correlation (rho > 0.5; n=78). The 78 taxa assigned to the positively correlated group were primarily assigned a small number of freshwater genera including *Limnohabitans* and *Polynucleobacter*, although it also unexpectedly included the opportunistic pathogen, *Enterococcus* (Fig. 4.3B). Taxa in the negatively correlated group were more diverse, including multiple genera assigned to the *Commamonadaceae* family along with a few groups groups more associated with sediments such as *Zoogloea* and *Opitutus*.

We then evaluated the RNA to DNA abundance ratios of taxa that increased or decreased along the length of the stream (Fig. 4.3C). The mean distribution of ratios of both groups followed the same pattern as the general taxa group with a slightly increased with distance and site 6 having the highest mean. Interestingly, negatively correlated taxa typically exhibited higher mean RNA:DNA ratios than positive, particularly in upstream sites, with a larger number of outliers above 1 (Fig. 4.3C).

Overall, RNA to DNA ratios for individual taxa were relatively flat across sites, with a slight increase of mean RNA:DNA ratios at site 6 (Fig. 4.4A). RNA to DNA ratios at either site 1 (the

headwaters) or at site 6 did not predict whether a taxon would increase or decrease in abundance along the length of the stream (Fig. 4.4B). Together, these results suggest that RNA transcript abundance in the headwaters does not predict microbial enrichment/depletion along the length of the stream, and that the difference in mean RNA:DNA ratios across sites observed in Figure 4.2 is related to the overall variation in the RNA:DNA ratios of individual microbes rather than increasing abundance of microbes with higher RNA:DNA abundances.

One surprise in the metatranscriptome was that many of the microbes with high RNA:DNA ratios represented genera associated with opportunistic pathogens that were not expected to be highly active in freshwater environments. To look further into this result, we identified and evaluated the relative abundance in metagenomes and metatranscriptomes of the *Acinetobacter, Enterococcus, Escherichia, Klebsiella* and *Staphylococcus* genera each dataset with the highest recruiting taxon identified as: *A. baumanni, A. pittii, A. indicus, A. nosocomialis* and A. *baylyi*, *E. faecium, E. coli, K. pneumoniae*, and *S. aureus* and *S. haemolyticus*. Taxa belonging to the *Escherichia* and *Enterococcus* genera in particular show high abundance in the metatranscriptome indicating high transcriptional activity (Fig. 4.5). On the other hand, Staphylococcus were present at substantially higher abundance in the metagenome than in the metatranscriptome, suggesting that these microbes may be in a low-activity state in this environment.

4.5 Discussion

We collected freshwater stream samples at 6 points through the McNutt Creek in Athens, GA to evaluate the connection between high metabolic content and the success in abundance through the creek as microbial community assembly happens. McNutt Creek is part of a mixed-use watershed that includes chicken farms, recreational lakes and wastewater treatment plants. Previous studies in this creek showed a rapid assembly at headwater and stabilizes with increased in distance

through the creek (15, 16). However, the factors driving these stability and abundance patterns are unknown. Here we used metagenomics and metatranscriptomics to evaluate the relationship between microbial transcriptional activity as defined by increased RNA (MT) and cellular abundance as defined by increased levels of DNA (MG). At a broad level, the taxonomic composition observed in our metagenome and metranscriptome analyses mirrored observations from our 16S rRNA gene analyses. These communities were dominated by typical freshwater taxa including *Actinomycetota*, *Pseudomonadota*, *Bacteroidota* and *Cyanobacteriota* (10, 16, 26). For both metagenome and metatranscriptome samples, diversity was highest in upstream samples, particularly site 1, and decreased with increasing distance travelled. This aligns well with previous observations of the same community, as well as literature trends illustrated in the introduction. (7, 27–30). In addition, beta diversity analysis shows that samples cluster more strongly by site/position along the stream than by sample type (metatranscriptome vs. metagenome), suggesting similar longitudinal trends in both datasets.

A major goal of this work was to explore the extent to which patterns of microbial activity reflected observed gradients in microbial diversity. The large number of taxa that were present in the metagenome but not the metatranscriptome suggests that there are diverse microbes present in these samples that are not reflected in the metatranscriptome. This could be caused by presence in the metagenome of cell-free DNA, sometimes termed environmental DNA, that can be recovered in the filtering process. DNA is substantially more stable in the environment than RNA (31) and can therefore potentially persist even after the death of the originating cell (32–34). This is a well-documented phenomena and has been used in a variety of contexts (35–37).

We were interested in examining metatranscriptomics to understand the extent to which transient microbes might be contributing to community activity and/or the extent to which microbes that are

present in upstream but not downstream environments might nonetheless be contributing to community function while present. RNA: DNA abundance ratios have been widely used as a measure of microbial activity (38–41). Previous work on a freshwater lake in Georgia, USA, found that taxa that were in low abundance in metagenome samples were typically present in even lower abundance in the metatranscriptome as a result of their transient status in that environment (28). This group also found that members of 2 freshwater phyla (*Verrucomicrobia* and *Planctomycetes*) typically had lower than average RNA to DNA ratios.

When initiating this work, we predicted that organisms with higher RNA:DNA ratios might be more likely to succeed in downstream environments, while microbes that decreased in abundance along the length of the stream might be poorly adapted to that environment and therefore show low activity even in the upstream environments. The results presented here reject this prediction by showing that RNA:DNA ratios are not predictive of enrichment/depletion along the length of the stream. If anything, microbes that were enriched along the length of the stream exhibited slightly lower mean RNA:DNA ratios, particularly in upstream sites. One factor affecting this may be that many of the organisms that survive in freshwater environments are slow-growing, oligotrophic organisms (42) and could have lower RNA content per cell (43).

We did observe a clear pattern where the RNA:DNA ratio for individual organisms was higher at site 6 than site 1. As our RNA and DNA ratios are based on relative abundance instead of absolute abundance, this may simply reflect the decreasing diversity of downstream datasets. The river continuum concept has long pointed out increasing availability of autocthonous, more labile carbon in downstream riverine environments (44), and multiple studies have backed this up with observations of increasing bacterial respiration rates in downstream environments (45, 46). While our data does not show increasing numbers of phototrophic taxa as might be predicted in

accordance with that model, this may be because our use of a prefilter would exclude eukaryotic algae, and by sampling only pelagic water we will not capture any photosynthetic biofilms or microbial mats.

A particularly interesting case were opportunistic taxa that are not typically considered native inhabitants of freshwater environments, but rather contaminants. For example, *Enterococci* and *Escherichia coli* are widely used as fecal indicator bacteria due to their believed origin primarily in animal feces (47–49). Their presence in this watershed environment is supported by the work of Cho et al (50). *Acinetobacter* groups have also previously identified in freshwater (51–53). While their presence in the metagenome was expected, we expected that these organisms, which are not thought to reproduce in freshwater environments, would have low metabolic activity in this environment and therefore low RNA abundance. Instead, these opportunistic pathogens were found to have some of the highest RNA:DNA ratios of any organism which suggests that when present, they could be highly active. This was based in part on prior work by Suttner et al. (54), who showed that enteric isolates presented a higher RNA to DNA ratio and could be interpreted as a signal of recent fecal pollution. The overall relative abundance found from these pathogenic groups is a public health concern. These organisms are known to carry plasmids and mobile elements with antibiotic resistance genes, contributing to the dissemination of antibiotic resistance through this environment.

Overall, this work provides new insight into microbial community ecology of streams. We showed that in general, metagenomic and metatranscriptomic samples captured similar patterns of microbial diversity and composition across sites, and that paired genomic and transcriptomic samples clustered more closely by location than by sample type. However, we also found that DNA samples captured a large number of low-abundance organisms that were not present in RNA

samples, suggesting that metagenomic sequencing techniques may be accounting for some cell-free DNA. Further, we found that RNA:DNA ratio in upstream sites was not predictive of whether that taxon would increase or decrease in abundance along the length of the stream, suggesting that transient stream community members should still be considered as active members of microbial dynamics in the stream. Overall, this work shows the power of multi-omic approaches to shed new light on stream microbial ecology.

4.6 References

1. Rodríguez-Ramos JA, Borton MA, McGivern BB, Smith GJ, Solden LM, Shaffer M, Daly RA, Purvine SO, Nicora CD, Eder EK, Lipton M, Hoyt DW, Stegen JC, Wrighton KC. 2022. Genome-Resolved Metaproteomics Decodes the Microbial and Viral Contributions to Coupled Carbon and Nitrogen Cycling in River Sediments. mSystems 7.

2. Battin TJ, Lauerwald R, Bernhardt ES, Bertuzzo E, Gener LG, Hall RO, Hotchkiss ER, Maavara T, Pavelsky TM, Ran L, Raymond P, Rosentreter JA, Regnier P. 2023. River ecosystem metabolism and carbon biogeochemistry in a changing world. Nature 613:449–459.

3. Hilderbrand RH, Bambakidis T, Crump BC. 2023. The Roles of Microbes in Stream Restorations. Microb Ecol 85:853–861.

4. Lindström ES, Leskinen E. 2002. Do neighboring lakes share common taxa of bacterioplankton? Comparison of 16S rDNA fingerprints and sequences from three geographic regions. Microb Ecol 44:1–9.

5. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK. 2002. Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquatic Microbial Ecology 28.

6. Warnecke F, Amann R, Pernthaler J. 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ Microbiol 6.

7. Ghai R, Rodríguez-Valera F, McMahon KD, Toyama D, Rinke R, de Oliveira TCS, Garcia JW, de Miranda FP, Henrique-Silva F. 2011. Metagenomics of the water column in the pristine upper course of the Amazon river. PLoS One 6.

8. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews 75.

9. Crump BC, Kling GW, Bahr M, Hobbie JE. 2003. Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. Appl Environ Microbiol 69:2253–2268.

10. Newton RJ, McLellan SL. 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front Microbiol 6.

11. McLellan SL, Fisher JC, Newton RJ. 2015. The microbiome of urban waters. International Microbiology https://doi.org/10.2436/20.1501.01.244.

12. Crump BC, Hobbie JE. 2005. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol Oceanogr 50.

13. Crump BC, Bowen JL. 2024. The Microbial Ecology of Estuarine Ecosystems. Ann Rev Mar Sci https://doi.org/10.1146/annurev-marine-022123-101845.

14. Crump BC, Amaral-Zettler LA, Kling GW. 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME Journal 6:1629–1639.

15. Hassell N, Tinker KA, Moore T, Ottesen EA. 2018. Temporal and spatial dynamics in microbial community composition within a temperate stream network. Environ Microbiol 20:3560–3572.

16. Teachey ME, McDonald JM, Ottesen EA. 2018. Rapid and stable microbial community assembly in the headwaters of third-order stream. Appl Environ Microbiol 85:1–15.

17. Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366–368.

18. Wickham H, François R, Henry L, Müller K. 2019. dplyr: A Grammar of Data Manipulation. R package version. Media.

19. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.

20. McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8.

21. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

22. Chen H. 2022. VennDiagram: generate high-resolution venn and euler plots. R package version 173.

23. Russel J. 2021. Russel88/MicEco: v0.9.15. 0.9.15. Zenodo.

24. Xu S, Zhan L, Tang W, Wang Q, Dai Z, Zhou L, Feng T, Chen M, Wu T, Hu E, Yu G. 2023. MicrobiotaProcess: A comprehensive R package for deep mining microbiome. Innovation 4.

25. Oksanen J, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. 2020. Vegan: Community Ecology Package. R package version 2.5-7. Community ecology package 10.

26. Vega L, Jaimes J, Morales D, Martínez D, Cruz-Saavedra L, Muñoz M, Ramírez JD. 2021. Microbial Communities' Characterization in Urban Recreational Surface Waters Using Next Generation Sequencing. Microb Ecol 81:847–863.

27. Fortunato CS, Eiler A, Herfort L, Needoba JA, Peterson TD, Crump BC. 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. ISME Journal 7:1899–1911.

28. Tsementzi D, Poretsky R, Rodriguez-R LM, Luo C, Konstantinidis KT. 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environ Microbiol Rep 6:640–655.

29. Febria CM, Koch BJ, Palmer MA. 2015. Operationalizing an ecosystem services-based approach for managing river biodiversityWater Ecosystem Services: A Global Perspective.

30. Satinsky BM, Gifford SM, Crump BC, Moran MA. 2013. Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis, p. 237–250. *In* .

31. Mao D, Luo Y, Mathieu J, Wang Q, Feng L, Mu Q, Feng C, Alvarez PJJ. 2014. Persistence of Extracellular DNA in River Sediment Facilitates Antibiotic Resistance Gene Propagation. Environ Sci Technol 48:71–78.

32. Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, Miaud C. 2011. Persistence of Environmental DNA in Freshwater Ecosystems. PLoS One 6:e23398.

33. Alvarez AJ, Yumet GM, Santiago CL, Toranzos GA. 1996. Stability of manipulated plasmid DNA in aquatic environments. Environ Toxicol Water Qual 11:129–135.

34. Huang S, Yoshitake K, Watabe S, Asakawa S. 2022. Environmental DNA study on aquatic ecosystem monitoring and management: Recent advances and prospects. J Environ Manage 323:116310.

35. Lee Y, Jung M, Wee J, Kim Y, Lee D, Lee D, Kim T, Cho K, Sim C. 2022. Species-specific environmental DNA analysis of the index species in soil ecosystem, *Allonychiurus kimi* (Collembola: Onychiuridae). Ecol Evol 12.

36. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, Martin L, Neff NF, Okamoto J, Wong RJ, Kharbanda S, El-Sayed Y, Blumenfeld Y, Stevenson DK, Shaw GM, Wolfe ND, Quake SR. 2017. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. Proceedings of the National Academy of Sciences 114:9623–9628.

37. Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. Biol Lett 4:423–425.

38. Orellana LH, Hatt JK, Iyer R, Chourey K, Hettich RL, Spain JC, Yang WH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. 2019. Comparing DNA, RNA and protein levels for measuring microbial dynamics in soil microcosms amended with nitrogen fertilizer. Sci Rep 9:17630.

39. Bowsher AW, Kearns PJ, Shade A. 2019. 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. mSystems 4.

40. DeAngelis KM, Silver WL, Thompson AW, Firestone MK. 2010. Microbial communities acclimate to recurring changes in soil redox potential status. Environ Microbiol 12.

41. Loeppmann S, Semenov M, Kuzyakov Y, Blagodatskaya E. 2018. Shift from dormancy to microbial growth revealed by RNA:DNA ratio. Ecol Indic 85.

42. Westoby M, Nielsen DA, Gillings MR, Litchman E, Madin JS, Paulsen IT, Tetu SG. 2021. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. Ecol Evol 11:3956–3976.

43. Hausmann B, Pelikan C, Rattei T, Loy A, Pester M. 2019. Long-Term Transcriptional Activity at Zero Growth of a Cosmopolitan Rare Biosphere Member. mBio 10.

44. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130–137.

45. del Giorgio PA, Cole JJ, Cimbleris A. 1997. Respiration rates in bacteria exceed phytoplankton production in unproductive aquatic systems. Nature 385:148–151.

46. Pollard PC. 2013. In situ rapid measures of total respiration rate capture the super labile DOC bacterial substrates of freshwater. Limnol Oceanogr Methods 11:584–593.

47. Lebreton F, Willems RJL, Gilmore MS. 2014. Enterococcus Diversity, Origins in Nature, and Gut ColonizationEnterococci: From Commensals to Leading Causes of Drug Resistant Infection.

48. Hammerum AM. 2012. Enterococci of animal origin and their significance for public health. Clinical Microbiology and Infection https://doi.org/10.1111/j.1469-0691.2012.03829.x.

49. Torres C, Alonso CA, Ruiz-Ripa L, León-Sampedro R, Del Campo R, Coque TM. 2018. Antimicrobial Resistance in Enterococcus spp. of animal origin . Microbiol Spectr 6.

50. Cho S, Barrett JB, Frye JG, Jackson CR. 2020. Antimicrobial resistance gene detection and plasmid typing among multidrug resistant enterococci isolated from freshwater environment. Microorganisms 8:1–15.

51. Cateau E, Verdon J, Fernandez B, Hechard Y, Rodier MH. 2011. Acanthamoeba sp. promotes the survival and growth of Acinetobacter baumanii. FEMS Microbiol Lett https://doi.org/10.1111/j.1574-6968.2011.02261.x.

52. Jung J, Baek JH, Park W. 2010. Complete genome sequence of the diesel-degrading Acinetobacter sp. strain DR1. J Bacteriol https://doi.org/10.1128/JB.00722-10.

53. Peleg AY, Seifert H, Paterson DL. 2008. Acinetobacter baumannii: Emergence of a successful pathogen. Clin Microbiol Rev https://doi.org/10.1128/CMR.00058-07.

54. Suttner B, Kim M, Johnston ER, Orellana LH, Ruiz-Perez CA, Rodriguez-R LM, Hatt JK, Brown J, Santo Domingo JW, Konstantinidis KT. 2021. Transcriptomic and rRNA:rDNA Signatures of Environmental versus Enteric Enterococcus faecalis Isolates under Oligotrophic Freshwater Conditions. Microbiol Spectr 9.

Table 4.1: Coordinates of sampling sites for metagenomes, metatranscriptomes and single cell genomes

Site	Coordinates
1	33.940642, -83.560772
2	33.936394, -83.557319
3 (MIDO 609)	33.9301, -83.5517
4 (MIDO 613)	33.9255, -83.5226
5 (MIDO 608)	33.9267, -83.4612
6 (MIDO 612)	33.9194, -83.4159
Site for SCGs (after MIDO and NORO merge)	33.8906, -83.3765

Table 4.2: Site conditions measured during sampling of metagenome and metatranscriptome samples

		Volume pumped (ml)						Dissolved Oxygen (DO)		
Site	Time sampled (am EST)	RNA1 filter	RNA2 filter	DNA1 filter	DNA2 filter	Temp. (°C)	pH	%	mg/L	conductivity
1	11:26	275	240	240	460	19.74	6.92	99.5	9.09	0.09
2	11:08	435	300	380	455	18.51	6.88	110.9	10.39	0.092
3	10:36	460	500	380	500	19.09	7.23	113.9	10.54	0.108
4	11:52	400	400	325	420	19.78	6.98	113.2	10.32	0.113
5	9:08	478	560	500	620	20.04	7.41	128	11.62	0.06
6	8:27	710	540	530	540	20.40	7.12	117.4	10.66	0.06

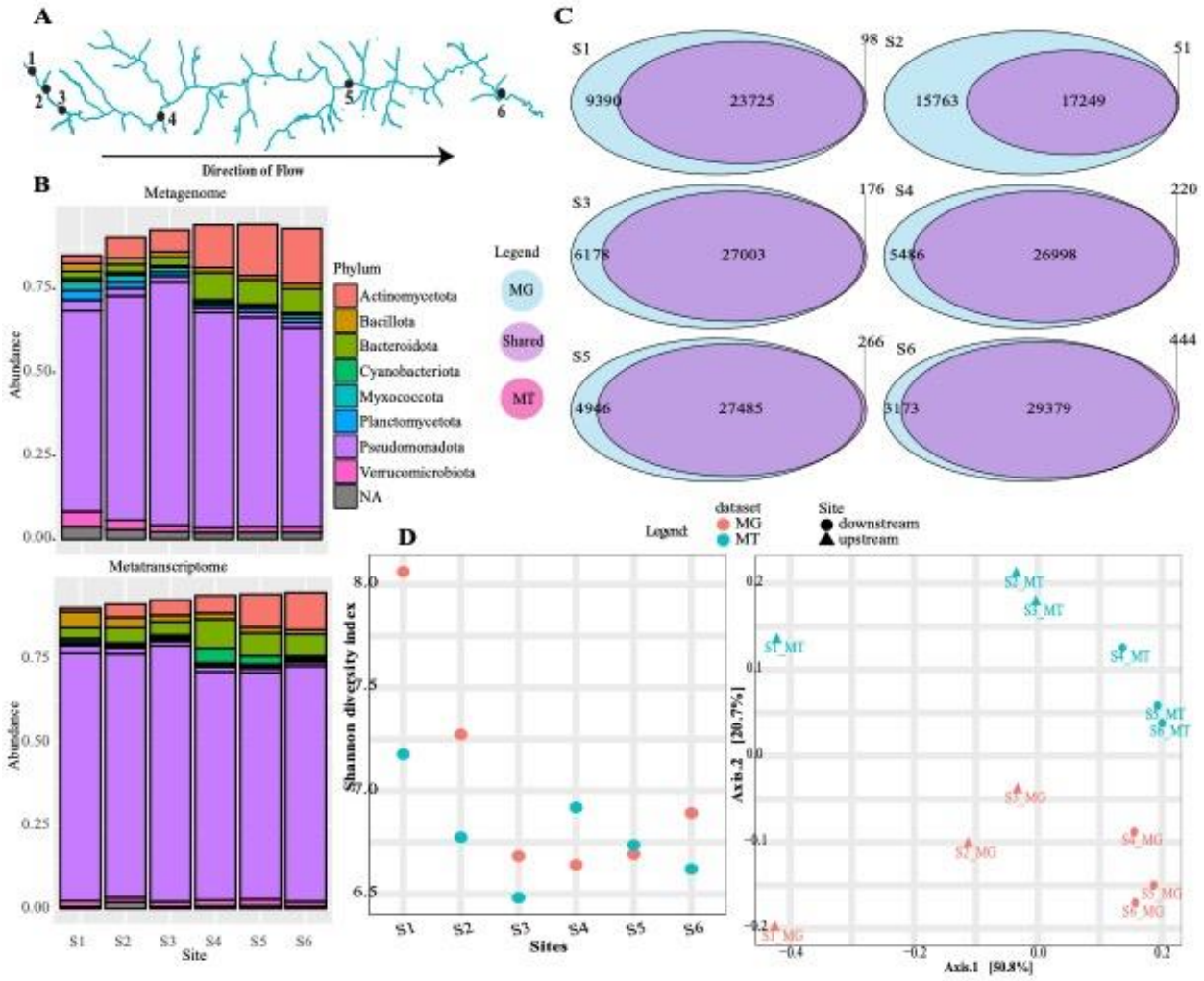

Figure 4.1: Description and comparison between dataset types. A) Map of sampling sites through McNutt Creek with headwaters at site 1. B) Top 1% of abundant phyla among all samples, presented by site. C) Venn Diagram comparing unique and shared taxa between each dataset type on each site. D) Alpha and beta diversity results calculates with Shannon index and Bray Curtis distances, respectively. Bray dissimilarity matrix was used to plot the Principal Coordinate Analysis (PCoA) clustering the samples by their differences in microbial composition.

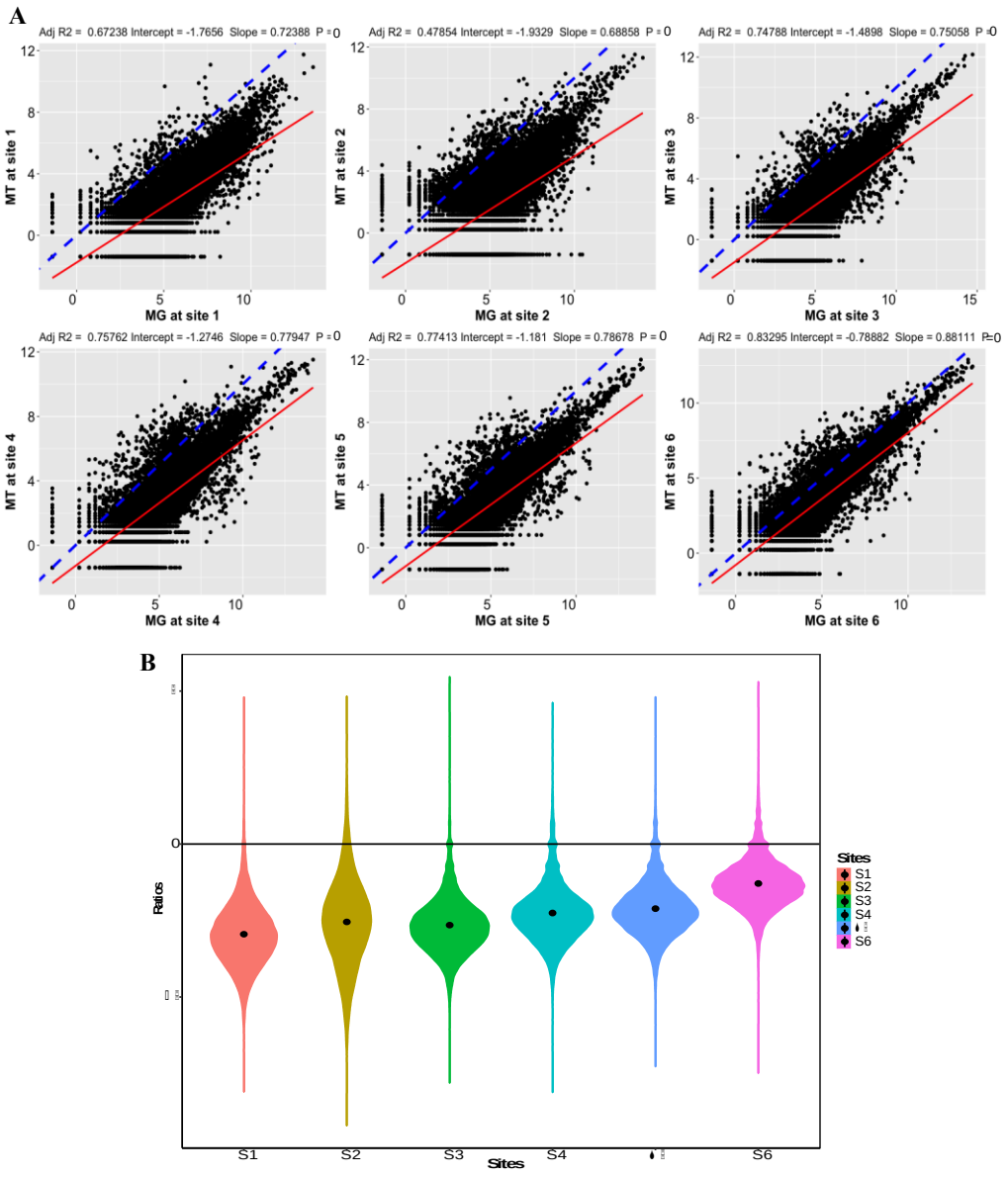

Figure 4.2: Association between metagenome and metatranscriptomes among all taxa. A) Relationship of metatranscriptome counts (pseudo log=0.25) against the metagenome counts (pseudo log=0.25) for each taxon. Blue line indicates a 1:1 relationship while the red line shows the observed trend in the relationship. B) Distributions of the RNA:DNA ratios of all taxa on each site.

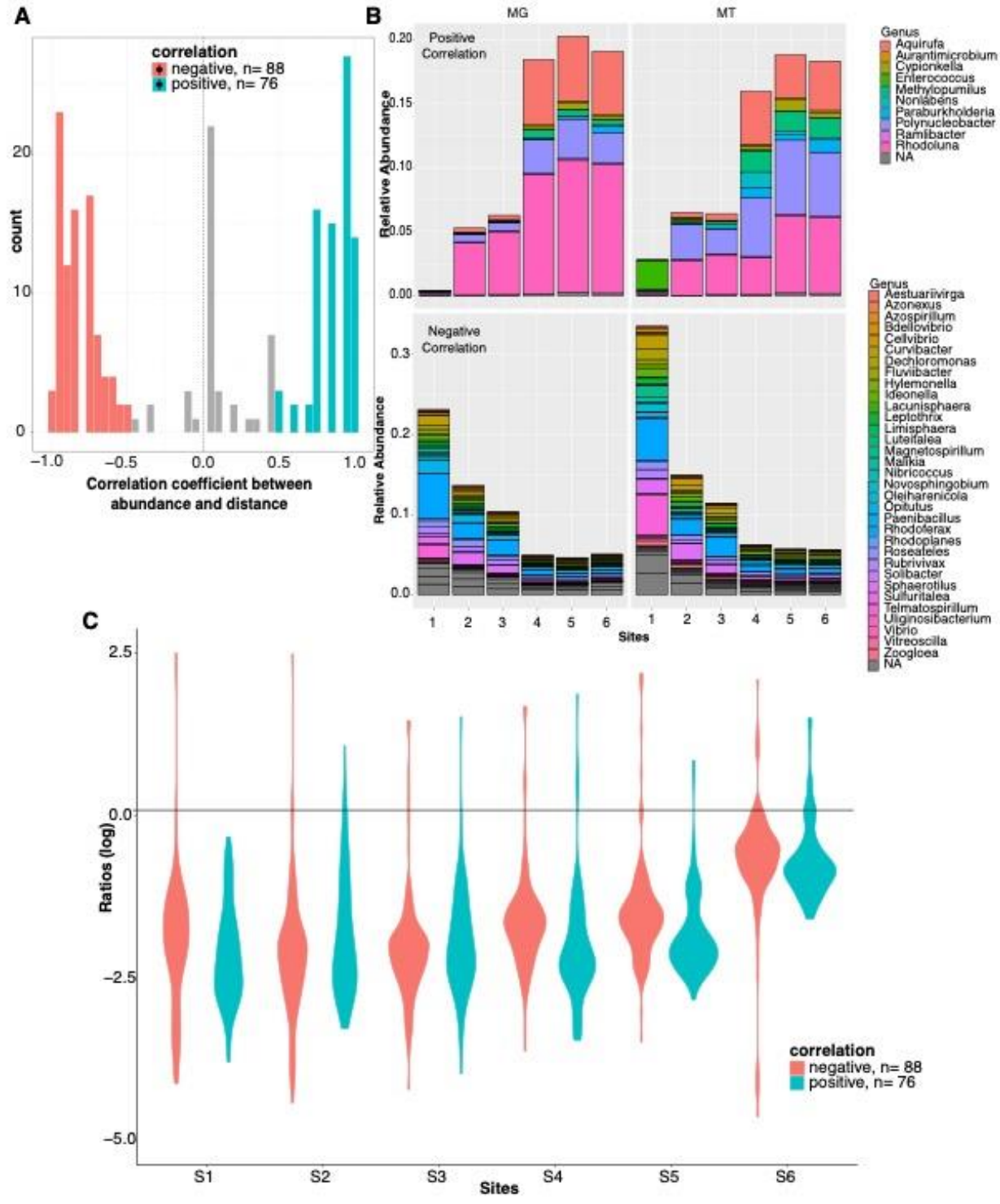

Figure 4.3: Activity and abundance patterns among taxa that increase or decrease along the length of the stream. A) Histogram of counts of abundant taxa that resulted in the respective correlation coefficient column. Colors represent the groups with highest and lowest coefficients.

B) Taxonomic composition of positive and negative correlated taxa from the abundant group. C) Distribution of RNA:DNA ratios of positive and negative correlated taxa on each site.

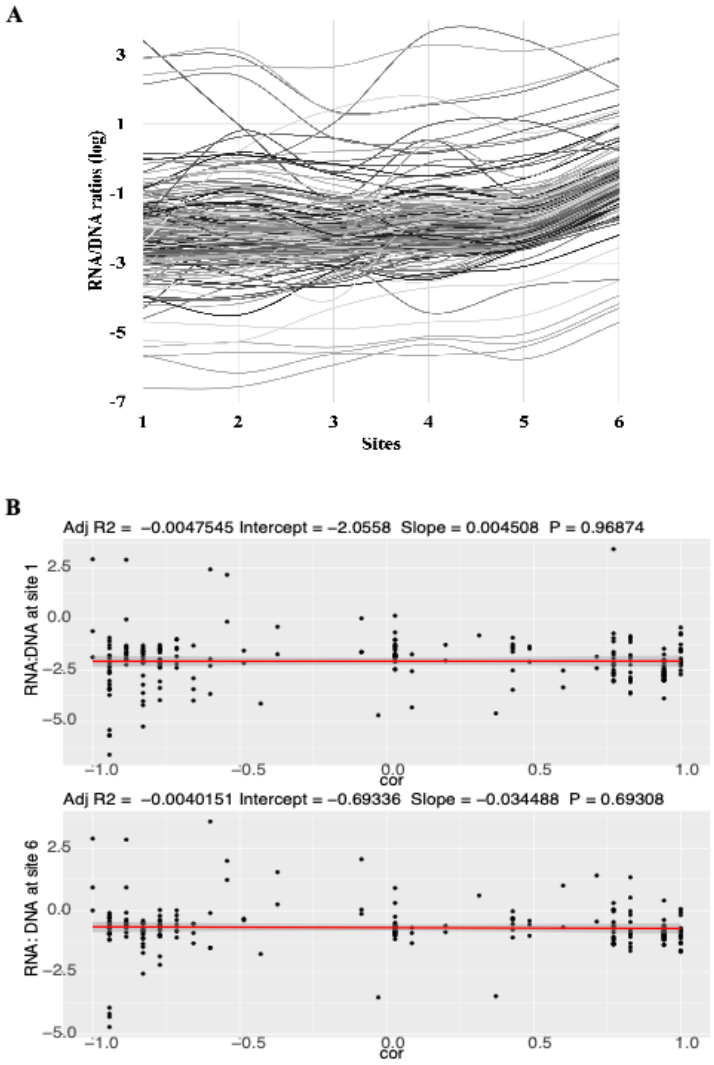

Figure 4.4: RNA to DNA abundance ratios is not predictive of growth along the length of the stream. A) RNA:DNA ratios of each abundant taxon displayed by site. B) Relationship between increased in abundance with distance and their RNA:DNA ratios at headwater (site 1) or at downstream (site 6).

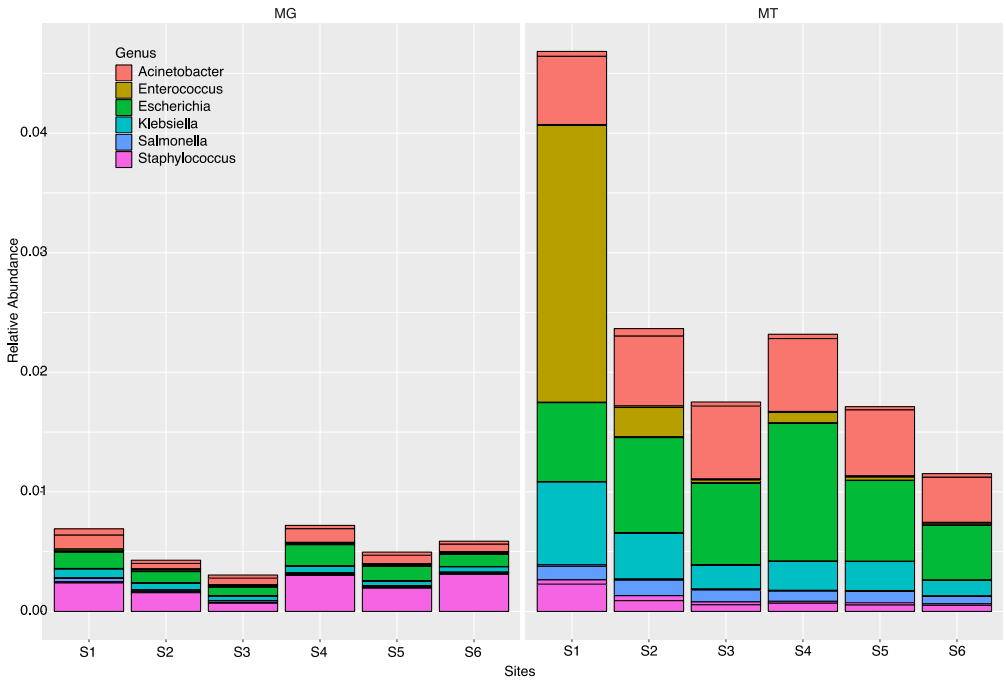

Figure 4.5: DNA and RNA abundance of selected opportunistic pathogens. The relative abundance of selected genera that represent common opportunistic human pathogens in metagenomic and metatranscriptomic datasets.

CHAPTER 5

CONCLUSIONS

5.1 Concluding remarks

Microbial communities play key roles in the health of the environment and of our own human bodies. One way in which microbial ecosystem processes can impact human health is through the spread of antibiotic resistance (1, 2). The human gut is an important platform for antibiotic resistance (AR) genes as the microbiome harbors AR bacteria that can become opportunistic pathogens and/or exchange AR genes with human pathogens. Therefore, the unrestricted use of antimicrobial agents and indiscriminative approaches to antibiotic stewardship are major contributors to antimicrobial resistance (3). Human interactions with the environment create bidirectional exchange of ARGs and AR bacteria. Human waste enters the environment as a result of leaks in wastewater conveyance infrastructure or via evasion of wastewater treatment processes. AR bacteria present in the environment can re-enter the human population either through direct exposure during recreation or indirect exposure as a result of agricultural practices, completing the circle. Freshwater environments play an essential role in this cycle, and their microbial communities are likely contributors to the movement and exchange of AR genes. The work presented here contributes to the microbial community ecology in both the human gut and stream environments to advance understanding into the dynamics involved in the antibiotic resistance issue.

In chapter 2, we presented work seeking to characterize extended spectrum beta-lactamase producing *Enterobacteriaceae* (ESBL-E) (4) present in the human gut of individuals without recent exposure to in-patient settings. The prevalence of community-acquired ESBL-E in

asymptomatic individuals was 4.5%, which is slightly higher than other areas of the U.S (5, 6). In the US, it is known that the south-Atlantic and mid-Atlantic area has the highest ESBL rates (7–9) which supports the prevalence found here. Furthermore, we found that this colonization was persistent for at least 3 months. The ESBL-E isolated from these individuals were all identified as *E. coli*, most of them carrying plasmids frequently associated with antibiotic resistance gene exchange (10, 11). This highlights the relevance of the human gut as a reservoir of ARGs exchange. More importantly, we found that most of our ESBL-E isolates carried the $bla_{CTX\text{-}M\text{-}15}$, a key beta lactamase in the global dissemination of ESBLs (12–15). The stool microbial community of these individuals did not present major changes in composition and assembly compared to individuals without ESBL-E colonization. However, the commensal *E. coli* on ESBL-E positive communities showed a dynamic antibiotic resistance profile suggesting a potential effect of ESBL-E isolates influencing the antibiotic resistance in the gut. We also found 2 different ESBL-*E.coli* isolates from the same individual carrying a nearly identical plasmid with antibiotic resistance genes, suggesting gene exchange dynamics in the gut. Together these results contribute a more informative perspective of the role of the human gut in antibiotic resistance dissemination. A future direction planned for this project is to assess the commensal microbiota more broadly to characterize their role in the antibiotic resistance spread, especially compared to these ESBL-E isolates of the same individuals. Future characterization of the commensal microbiota includes a phylogenetic analysis to determine the prevalence and diversity of *E. coli* types within the gut of the same individuals. Additionally, a comparative analysis of ARGs and mobile elements in these isolates would inform strategies of antibiotic resistance dynamics in the gut contributing to antibiotic resistance spread (16).

Antibiotic resistance spread through the environment is one of the most difficult challenges of this issue. Animal waste, hospital sewage and human fecal pollution are contributing sources of antibiotic resistance in the environment (17–20). A strategy to mitigate the human contribution to this environmental concern is through improving antibiotic resistance awareness. In Chapter 3, we presented a laboratory module for undergraduate students to isolate and characterize ESBL-E from freshwater streams in the community. This module exposed students to basic microbiology techniques as well as more advanced water quality monitoring techniques. In addition, students analyzed antibiotic resistance profiles phenotypically and genotypically. This resulted in a better understanding of the antibiotic resistance in the environment and possible gene exchange dynamics in the community. Students reported that this module was one of their favorite activities in the course and one of the exercises where they learned the most. Students reported struggles to interpret some of the most advanced tests to characterize antibiotic resistance (MIC test strip) suggesting the need of better strategies for training in this topic. Although this exercise provided research techniques commonly used in research laboratories on this topic, only a low percent of students agreed that this module made them feel like a scientist. The science identity is especially important in the retention of minorities in STEAM areas presenting an opportunity to improve this perception with activities targeting this issue (21, 22). On the other hand, a high proportion of students (>70%) agreed that this module increased their awareness of antibiotic resistance in the environment. This is especially important in efforts to mitigate the antibiotic resistance spread in the environment, more specifically in freshwater streams (23–25).

Riverine environments are a connection between different terrestrial environments and between estuarine and marine communities, playing an essential role in dissemination of pathogens and antibiotic resistance genes. To better understand the dissemination of ARGs in these environments

it is crucial to understand the dynamics of its microbial communities. The stream microbiome is assembled with the influence of soil and sediment microbes and rapidly assembles into a more specialized freshwater taxa group (26, 27). However, there is minimal understanding of the role of this transient soil-associated taxa while present in this community. Some groups have investigated water characteristics driving the microbial assembly and stability in streams, but it is still not clear which are the factors influencing this selection (28–32). In Chapter 4, we present work aimed at understanding how microbial activity (as evaluated by metatranscriptomic analysis) is related to microbial abundance (as evaluated by metagenomic abundance) across the length of a single freshwater stream. Overall, our metagenomic results align well with patterns of community assembly previously reported for this watershed (33, 34). Analysis of metatranscriptomes showed that RNA:DNA ratios for individual taxa was not correlated with enrichment/depletion along the course of the stream. Most of the microbial community showed RNA: DNA ratios below 1, although the ratio distribution increased with increased distance in the creek. In addition, we found that important opportunistic pathogens were present at high abundance in the metagenome and the metatranscriptome. Some of these pathogens include *Enterobacteriaceae*, frequently associated with ESBLs and antibiotic resistance exchange. The high RNA abundance of these pathogens could reflect high activity in these environments, which has important public health implications. Future work in this project should include a detailed analysis of the active genes in these populations as well as in the other taxa that are not considered part of the freshwater microbiome but could be involved in the assembly and other dynamics of these communities.

5.2 Future Directions

Overall, the work presented here advances a research area essential for public health as it integrates microbial communities from the environment and human gut. Prospective future investigations

could include a comparative analysis between ARGs in mobile elements present in the human commensal gut community and the ones present in mobile elements in the environment. The high abundance of opportunistic pathogens found in the freshwater dataset offers a unique opportunity to explore the gene content and activity of these microbes while present in these environments. This could elucidate antibiotic resistance gene exchange patterns that are currently unclear. Similarly, comparing the ARG content in mobile elements could predict pathways of ARG exchange and confirm the role of freshwater environments in ARG dissemination. Additionally, mesocosms studies could be performed to measure ARG transmission among opportunistic pathogens under environmental conditions, as well as ARG transmission between opportunistic pathogens and freshwater taxa. A major limitation to these approaches could be the identification of mobile elements like plasmids from short reads like the metagenome and metatranscriptome. However, their association with ARG and bacterial hosts could be predicted using replicon genes, a specific plasmid database for microbial metagenomes and methods like Hi-C previously used in wastewater bacteria carrying ARGs in plasmids (35, 36). As the antibiotic resistance threat remains, better understanding of gene exchange dynamics is needed. This dissertation offers the opportunity to explore these dynamics by assessing antibiotic resistance genes in humans and in the environment.

5.3 References

1. CDC. 2019. Antibiotic Resistance Threats in The United States, 2019. U.S. Department of Health and Human Services.

2. Velazquez-Meza ME, Galarde-López M, Carrillo-Quiróz B, Alpuche-Aranda CM. 2022. Antimicrobial resistance: One Health approach. Vet World 743–749.

3. WHO. 2014. Antimicrobial resistance. Global report on surveillance. World Health Organization 61.

4. Bush K, Bradford PA. 2020. Epidemiology of B-Lactamase-Producing Pathogens. Clin Microbiol Rev 33:1–37.

5. Bezabih YM, Sabiiti W, Alamneh E, Bezabih A, Peterson GM, Bezabhe WM, Roujeinikova A. 2021. The global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli in the community. Journal of Antimicrobial Chemotherapy 76:22–29.

6. Bezabih YM, Bezabih A, Dion M, Batard E, Teka S, Obole A, Dessalegn N, Enyew A, Roujeinikova A, Alamneh E, Mirkazemi C, Peterson GM, Bezabhe WM. 2022. Comparison of the global prevalence and trend of human intestinal carriage of ESBL-producing Escherichia coli between healthcare and community settings: A systematic review and meta-Analysis. JAC Antimicrob Resist 4.

7. Castanheira M, Farrell SE, Krause KM, Jones RN, Sader HS. 2014. Contemporary diversity of β-lactamases among enterobacteriaceae in the nine U.S. census regions and ceftazidime-avibactam activity tested against isolates producing the most prevalent β-lactamase groups. Antimicrob Agents Chemother 58:833–838.

8. Castanheira M, Simner PJ, Bradford PA. 2021. Extended-spectrum β-lactamases: An update on their characteristics, epidemiology and detection. JAC Antimicrob Resist 3.

9. Kaye KS, Gupta V, Mulgirigama A, Joshi A V., Scangarella-Oman NE, Yu K, Ye G, Mitrani-Gold FS. 2021. Antimicrobial Resistance Trends in Urine Escherichia coli Isolates From Adult and Adolescent Females in the United States From 2011 to 2019: Rising ESBL Strains and Impact on Patient Management. Clinical Infectious Diseases 73:1992–1999.

10. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. Journal of Antimicrobial Chemotherapy 73:1121–1137.

11. Carattoli A. 2009. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother 53:2227–2238.

12. Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. Clin Microbiol Rev 26:744–758.

13. Cantón R, Coque TM. 2006. The CTX-M β-lactamase pandemic. Curr Opin Microbiol 9:466–475.

14. Cantón R, González-Alba JM, Galán JC. 2012. CTX-M enzymes: Origin and diffusion. Front Microbiol 3.

15. Peirano G, Richardson D, Nigrin J, McGeer A, Loo V, Toye B, Alfa M, Pienaar C, Kibsey P, Pitout JDD. 2010. High prevalence of ST131 isolates producing CTX-M-15 and CTX-M-14 among extended-spectrum-β-lactamase-producing Escherichia coli isolates from Canada. Antimicrob Agents Chemother 54:1327–1330.

16. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal Escherichia coli. Nat Rev Microbiol 8:207–217.

17. Iseppi R, Di Cerbo A, Messi P, Sabia C. 2020. Antibiotic resistance and virulence traits in vancomycin-resistant enterococci (Vre) and extended-spectrum β-lactamase/ampc-producing (ESBL/ampc) enterobacteriaceae from humans and pets. Antibiotics 9.

18. Subramanya SH, Bairy I, Metok Y, Baral BP, Gautam D, Nayak N. 2021. Detection and characterization of ESBL-producing Enterobacteriaceae from the gut of subsistence farmers, their livestock, and the surrounding environment in rural Nepal. Sci Rep 11.

19. Cho S, Jackson CR, Frye JG. 2023. Freshwater environment as a reservoir of extended-spectrum β-lactamase-producing Enterobacteriaceae. J Appl Microbiol 134:1–22.

20. Cho S, Hiott LM, Read QD, Damashek J, Westrich J, Edwards M, Seim RF, Glinski DA, Bateman Mcdonald JM, Ottesen EA, Lipp EK, Henderson WM, Jackson CR, Frye JG. 2023. Distribution of Antibiotic Resistance in a Mixed-Use Watershed and the Impact of Wastewater Treatment Plants on Antibiotic Resistance in Surface Water. Antibiotics 2023 12:1586.

21. Le PT, Doughty L, Thompson AN, Hartley LM. 2019. Investigating undergraduate biology students' science identity production. CBE Life Sci Educ 18.

22. McCartney M, Roddy AB, Geiger J, Piland NC, Ribeiro MM, Lainoff A. 2022. Seeing Yourself as a Scientist: Increasing Science Identity Using Professional Development Modules Designed for Undergraduate Students. J Microbiol Biol Educ 23.

23. Elmahi OKO, Uakkas S, Olalekan BY, Damilola IA, Adedeji OJ, Hasan MM, dos Santos Costa AC, Ahmad S, Essar MY, Thomson DJ. 2022. Antimicrobial resistance and one health in the post COVID-19 era: What should health students learn? Antimicrob Resist Infect Control https://doi.org/10.1186/s13756-022-01099-7.

24. Genné-Bacon EA, Bascom-Slack CA. 2018. The PARE Project: A Short Course-Based Research Project for National Surveillance of Antibiotic-Resistant Microbes in Environmental Samples. J Microbiol Biol Educ 19.

25. Maicas S, Fouz B, Figàs-Segura À, Zueco J, Rico H, Navarro A, Carbó E, Segura-García J, Biosca EG. 2020. Implementation of Antibiotic Discovery by Student Crowdsourcing in the Valencian Community Through a Service Learning Strategy. Front Microbiol 11.

26. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews 75.

27. Crump BC, Kling GW, Bahr M, Hobbie JE. 2003. Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. Appl Environ Microbiol 69:2253–2268.

28. Debray R, Herbert RA, Jaffe AL, Crits-Christoph A, Power ME, Koskella B. 2022. Priority effects in microbiome assembly. Nat Rev Microbiol https://doi.org/10.1038/s41579-021-00604-w.

29. Jackson CR, Millar JJ, Payne JT, Ochs CA. 2014. Free-living and particle-associated bacterioplankton in large rivers of the Mississippi River basin demonstrate biogeographic patterns. Appl Environ Microbiol 80.

30. Ruiz-González C, Niño-García JP, Lapierre JF, del Giorgio PA. 2015. The quality of organic matter shapes the functional biogeography of bacterioplankton across boreal freshwater ecosystems. Global Ecology and Biogeography 24.

31. Wisnoski NI, Lennon JT. 2021. Microbial community assembly in a multi-layer dendritic metacommunity. Oecologia 195.

32. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, Blaschke AP, Blöschl G, Mach RL, Kirschner AKT, Farnleitner AH, Eiler A. 2015. Bacterial diversity along a 2600km river continuum. Environ Microbiol 17:4994–5007.

33. Hassell N, Tinker KA, Moore T, Ottesen EA. 2018. Temporal and spatial dynamics in microbial community composition within a temperate stream network. Environ Microbiol 20:3560–3572.

34. Teachey ME, McDonald JM, Ottesen EA. 2018. Rapid and stable microbial community assembly in the headwaters of third-order stream. Appl Environ Microbiol 85:1–15.

35. Camargo AP, Call L, Roux S, Nayfach S, Huntemann M, Palaniappan K, Ratner A, Chu K, Mukherjeep S, Reddy TBK, Chen I-MA, Ivanova NN, Eloe-Fadrosh EA, Woyke T, Baltrus DA, Castañeda-Barba S, de la Cruz F, Funnell BE, Hall JPJ, Mukhopadhyay A, Rocha EPC, Stalder T, Top E, Kyrpides NC. 2024. IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. Nucleic Acids Res 52:D164–D173.

36. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. ISME J 13:2437–2446.

APPENDIX A

DECIPHERING THE ROLES AND REGULATORY CONNECTIONS OF THE *VIBRIO FISCHERI* PHEROMONE-SIGNALING SYSTEM

Introduction

Vibrio fischeri is best known for its ability to produce bioluminescence. This metabolic property is known to play a key role in the symbiotic relationship with the *Euprymna scolopes* Hawaiian bobtail squid (1–3). Moreover, control of bioluminescence phenomenon has been studied and used as a model system to understand genetic regulation. Bioluminescence in *V. fischeri* is controlled by the lux system (4, 5). This system includes the *lux* operon (*luxICDABEG*), encoding genes for seven proteins, and the divergently transcribed *luxR* gene. LuxI produces the *N*-3-oxohexanoyl homoserine lactone (3OC6-HSL). This pheromone can diffuse in and out of cells, and at a critical concentration it attaches to LuxR, which then activates transcription of the *lux* operon by binding to the promoter region in the "*lux* box" (6, 7). Because the *lux* operon contains *luxI*, which encodes the 3OC6-HSL synthase, this activation creates a positive feedback loop. LuxA and LuxB form the luciferase that produces light, while the remaining operon genes (*luxC*, *luxD*, *luxE* and *luxG*) encode the proteins that regenerate intermediates and substrates in the reaction (8–10).

In addition to the *lux* operon, *V. fischeri* has two pheromone signaling systems with their own synthases and receptors (AinS-AinR, LuxS-LuxPQ) (11). LuxS has a minor role in symbiosis and culture conditions (12), while AinS produces *N*-octanoyl-homoserine lactone (C8-HSL) that can also impact bioluminescence indirectly by participating in a regulatory cascade that affects multiple genes including *luxR* (13–15). C8-HSL also binds to LuxR which connects the Ain and

Lux pheromone signaling systems (16, 17). LuxI/LuxR and AinS/AinR pheromones signaling systems are controlled by environmental conditions and positive feedback loops.

Typically, pheromone signaling is associated with cell density and the response is typically termed "quorum sensing". Multiple bacteria are known to carry these systems and use them to control production of virulence factors and biofilms (7, 18–21). However, *V. fischeri* pheromone signaling systems can respond to multiple environmental factors affecting luminescence production without directly interacting with the Lux proteins directly responsible for bioluminescence (22–25). A major environmental regulator of luminescence in *V. fischeri* is the ArcAB system (23). The sensor kinase, ArcB, phosphorylates ArcA in reducing conditions, and ArcA-P then binds to the promoter region of the *lux* box and repress transcription, possibly by obstructing the binding of the 3OC6-HSL and LuxR complex. Without *arcA* repression, the bioluminescence production is ~ 500-fold higher (23). This response is affected by a positive feedback loop as shown by an experiment using an *arcA* and *luxI* mutant (26). This same study showed that if ArcA repression is removed in a subpopulation, those cells can produce 3O6C-HSL and activate the lux operon response in cells that are still repressed by ArcA.

My work aimed to understand if positive feedback was a tool to activate a population response to environmental factors that are only experienced by a subpopulation, as may happens inside the crypt spaces colonized by *V. fischeri* in the squid light organ (27, 28). The goal of testing the positive feedback response of the lux system in an heterogenous environment requires a controlled activation of the system. To achieve this control, we sought to mimic the ArcA regulation mechanism using a regulator that is more easily controlled than the redox conditions that govern the ArcB/ArcA response.

Methods

Strain construction

To artificially control the lux operon regulation, we replaced the ArcA binding site upstream the "lux box" with a lac operator site. This modification was performed in a *V. fischeri* strain, KV6576, that encodes a *lacI*^{q} chromosomal insertion between *yeiR* and *glmS* (29). This chromosomally encoding LacI allowed to control the lux operon with isopropyl β-d-1-thiogalactopyranoside (IPTG). In the absence of lactose, LacI represses gene expression. In contrast, when IPTG is added, as an allolactose mimic, it removes the LacI repression allowing transcription of the operon. The resulting strain of this modification, CR2, was constructed using site directed mutagenesis (Figure 1). Primers CRG1 and CRG2 were used to amplify the plasmid pEVS151 with the 3' end of the primer containing an altered sequence to replace the *arcA* binding site with the *lac* operator (Figure 1). This newly constructed plasmid, pCRG7, was used as a donor for allelic exchange with *V. fischeri* strain KV6576 as recipient and the conjugative helper plasmid pEVS104. The triparental mating strategy used to create all strains in this project has been previously described (30).

Figure 1: Visualization of target sequence for insertion of first LacI operator by site directed mutagenesis. A) Annealing regions of primers CRG1 and CRG2 into original sequence. B) Mutated sequence as a result of the site directed mutagenesis to add LacI operator.

To increase repression of the lux operon, simulating the degree of repression created by ArcA, a second *lac* operator was added downstream the "*lux* box", between the -35 and -10 region. To create this new site directed mutagenesis, the initial plasmid, pCRG7, containing one *lacI*^{q} operator, was reverse amplified with primers CRG16 and CRG17, with an altered 3' end to insert the second operator, resulting in plasmid pCRG19. To enhance the efficiency allelic exchange, this new plasmid was fused with plasmid carrying the ColE1 origin of replication, pBluescript, creating pCRG21, a high copy number plasmid with two LacI operators (Figure 2). After the confirmation of the genetic construction, by restriction digestion and Sanger sequencing, pCRG21 was used as a donor plasmid with KV6576 to create CR4.

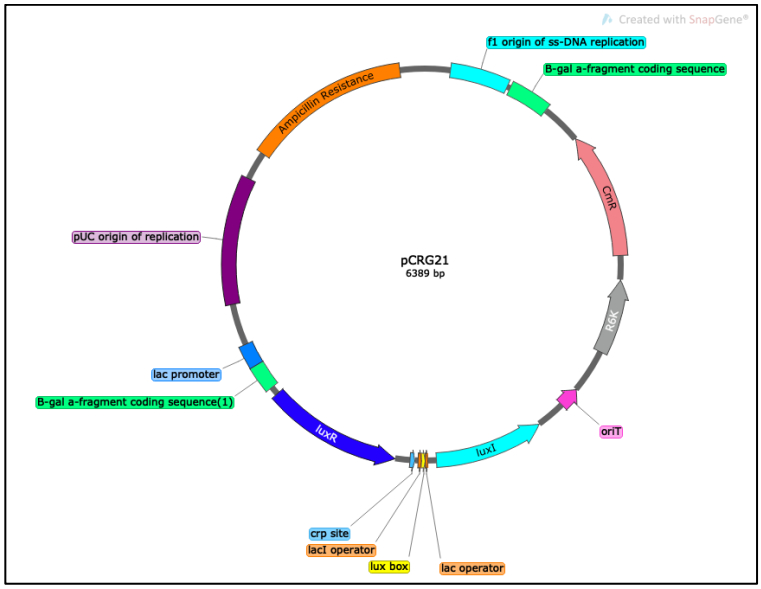

Figure 2: Map of pCRG21 plasmid, created in SnapGene. This plasmid was created as a fusion between pCRG19, providing the genes from *luxR* to CmR, and the pBluescript plasmid, providing the genes within the fragment of the split B-gal gene.

To create a visual signal to the response by positive feedback, a green fluorescent protein (GFP) reporter gene was inserted in the *lux* operon, downstream the *luxI* gene. This GFP variant has a shorter half-life that allows a dynamic signal (28). The GFP variant gene was amplified with reverse primer JLSssrA-ASV to generate the mutation in the 3' end of the sequence and forward primer CRG3 to ensure compatibility with the recipient plasmid, pJLB72, previously carrying a wild-type copy of the GFP downstream *luxI*. This combination resulted in pCRG4, the donor plasmid used with KV6576 ($lacI^q$) as a parent strain to generate CR5. The same conjugation was performed using CR4 as parent strain, resulting in strain CR7.

To compare these strains with a variant of a non-functional LuxI, a plasmid similar to pCRG4 was used. pCRG4 contained two restriction sites for BglII enzyme, including one within the *luxI* gene so the additional restriction site outside *luxI* was removed by restriction digestion and self-ligation, resulting in pCRG17, containing only one BglII restriction site in the *luxI*. Later, pCRG17 was digested with BglII and single stranded overhangs resulting from that digestion were filled with a Klenow fragment, and the blunt ends ligated together, to create a frameshift mutation in the *luxI* gene and generating plasmid pCRG23. Through the same triparental mating strategy this plasmid was used a donor, with CR5 as parent, to create CR6 (31). Strain CR6 has a destabilized GFP reporter in the mutated *luxI* and can receive the 3OC6-HSL from other cells but cannot produce this signal, therefore interrupting the positive feedback loop. After testing the *luxI* non-functional mutation of CR6 by fluorescence and luminescence phenotype as well as confirming the genotype, that strain was used as a parent conjugated with pCRG21, the same plasmid used to insert the two lac operators in strain CR4. The resulting strain, CR9, now has a destabilized GFP reporter for the *lux* operon with a non-functional *luxI* controlled by 2 $lacI^q$ operators.

Table 1: Primers designed and used in this study to modify plasmids and strains.

Primer name	Sequence (5'-3')	Description
CRG1	TTATCCGCTCACAATTTCCACCAAGATGCATAAATGTGA	Insertion of LacI operator
CRG2	TTGTGAGCGGATAACAAAGCTGTAGGATGGTACAGGT	Insertion of LacI operator
CRG3	ATTAGCGGCCGCGAAGGAGATATACATATGGCTAGCAAAGGAGAAGAAC	Forward primer for destabilized GFP
CRG16	GAGCGGATAACAACAGTTTGTTATGATCATGTA	Second lacI operator
CRG17	TTGTTATCCGCTCACCTGTACCATCCTACAGCT	Second lacI operator

Growth conditions and Bioluminescence Experiments

Wild-type and engineered strains were grown in artificial seawater medium (SWTO) (32) with the addition of 2 mM IPTG and/or 1 mM 3OC6-HSL when relevant. Luminescence experiments were performed in black-walled clear-bottom 96-well plates using a Synergy2 plate reader. Each well contained 200 µl of SWTO inoculated with 2 µl of overnight culture. Plates were incubated at 24ºC, shaking continuously at medium speed, over 18 hours with measurements every 30 minutes. Measurements included luminescence and optical density (OD) at 600 nm. Specific luminescence was calculated using the luminescence value over the OD value closest to 1.0. Fluorescence experiments were performed under the same plate reader conditions with the addition of a fluorescence read using an excitation filter of 485/20 and an emission filter of 528/20. Specific fluorescence was calculated using the maximum fluorescence units over the OD at the respective timepoint and then subtracting the fluorescence produced by the medium without bacteria, used as blank. All luminescence and fluorescence experiments were performed with biological triplicates. Fluorescence imaging was obtained in a fluorescence microscope, after the addition of 2 mM IPTG or 1 mM 3OC6-HSL.

Results

To better understand the positive feedback loop in *V. fischeri*, six newly genetically modified strains were created. Each strain modification was confirmed by restriction digestion, PCR and Sanger sequencing. Each strain growth rate was compared to their respective parental strains (data not shown). Engineered strains shown in Figure 3 were named in the order they were created.

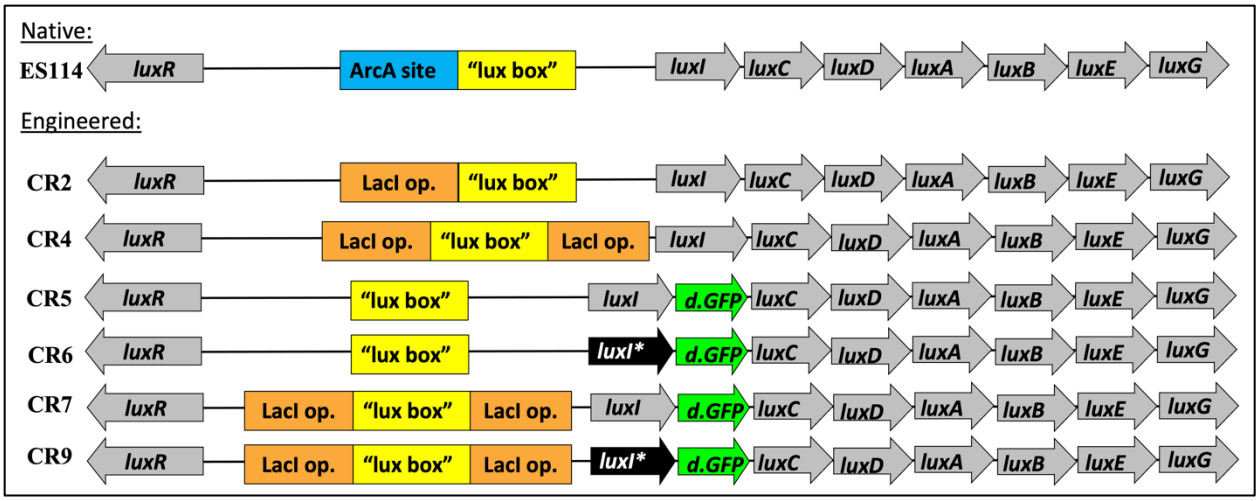

Figure 3: Genetic constructions in *V. fischeri* strains using KV6576 as parental strain and pEVS104 as helper plasmid. ES114 represents the native lux operon region while engineered strains show the *lacI* operators inserted around the "*lux* box". Strains CR5 to CR9 also contained the destabilized variant of the GFP reporter. Strains CR6 and CR9 contain a non-functional *luxI* allele.

Specific luminescence experiments used to test the LacI repression of the *lux* operon were performed with the first strain created with this capability, CR2. The repression levels were compared to the parent strain, KV6576, containing the native ArcA repression system and with a mutant strain, AMJ2 that lacks the *arcA* gene. The specific luminescence between those two strains was more than 200-fold difference, so our goal for the IPTG-controlled repression was similar. However, the specific luminescence difference in CR2 with and without repression was less than

100-fold (Figure 4). After adding a second LacI operator, the specific luminescence in CR4 strain with and without IPTG was ~300-fold difference, closer to the initial goal to replicate the native repression created by ArcA.

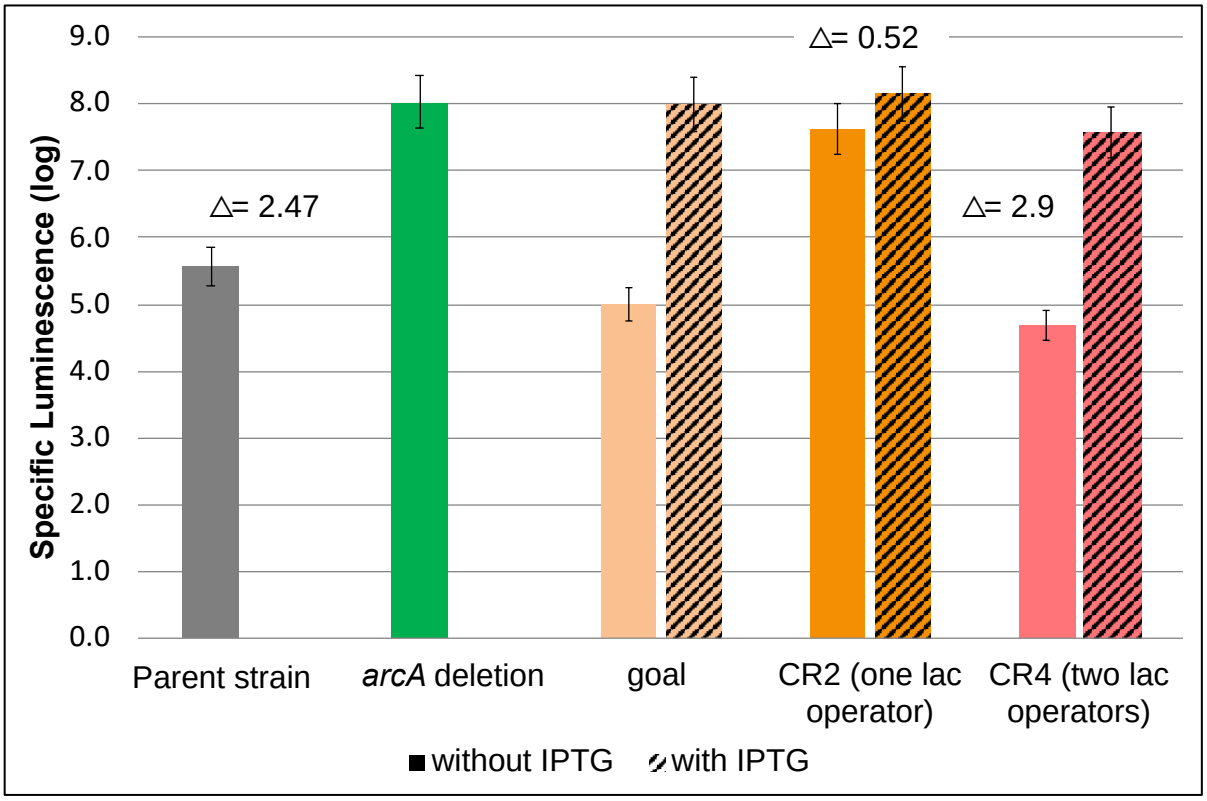

Figure 4: Specific Luminescence of CR2 and CR4 genetically engineered strains with and without the addition of IPTG repression compared to the native repression created by ArcA. The grey bar represents the strain KV6576, the green bar with *arcA* deletion represents strain AMJ2, while the goal bars are artificially created for comparison purposes. The delta numbers represent the difference in specific luminescence with the addition of IPTG.

Specific luminescence of CR6 with a non-functional *luxI* gene was also tested to confirm the expected phenotype. This strain can receive an activation signal, but it is unable to produce additional 3OC6-HSL, interrupting the positive feedback loop. Therefore, CR6 produces lower

luminescence values than its parent strain, CR5 (Figure 5). Luminescence decreased when cells are not able to produce 3OC6-HSL.

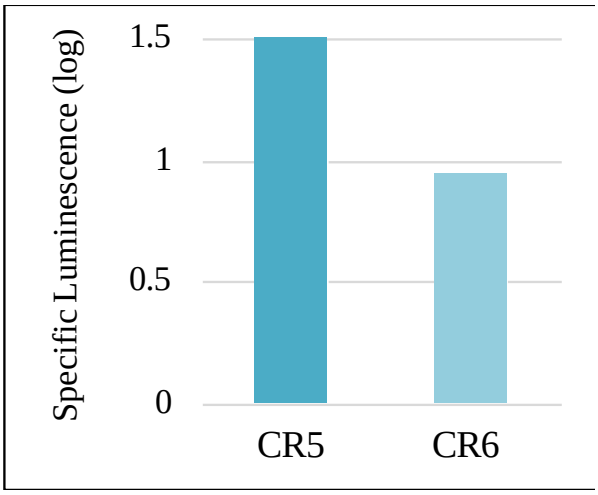

Figure 5: Specific luminescence of CR6 (*luxI* point mutation) compared to its parent strain, CR5.

In a similar way, CR7 was used to test the phenotype of the destabilized GFP reporter inserted downstream *luxI*. Specific fluorescence values were calculated at the maximum fluorescence value, compared to the parent strain, CR4, that do not contain a GFP reporter. As expected, the strain carrying the GFP reporter produces fluorescence above the autofluorescence of the parent strain (Figure 6).

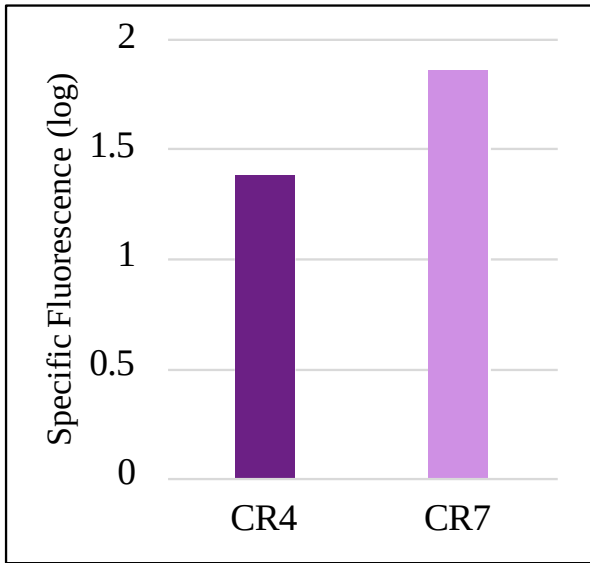

Figure 6: Fluorescence levels on strains with IPTG inducible promoter.

Lastly, the phenotypes of CR7 and CR9 were tested by luminescence and fluorescence experiments. Since both strains contained two LacI operators, the *lux* operon expression was controlled with IPTG. Also, they both carry a destabilized variant of the GFP reporter. In both strains, specific luminescence and specific fluorescence increased with IPTG addition, as a response of the LacI operators. Compared to each other, CR9 showed a lower luminescence and fluorescence, as an effect of its non-functional *luxI* allele (Figure 7).

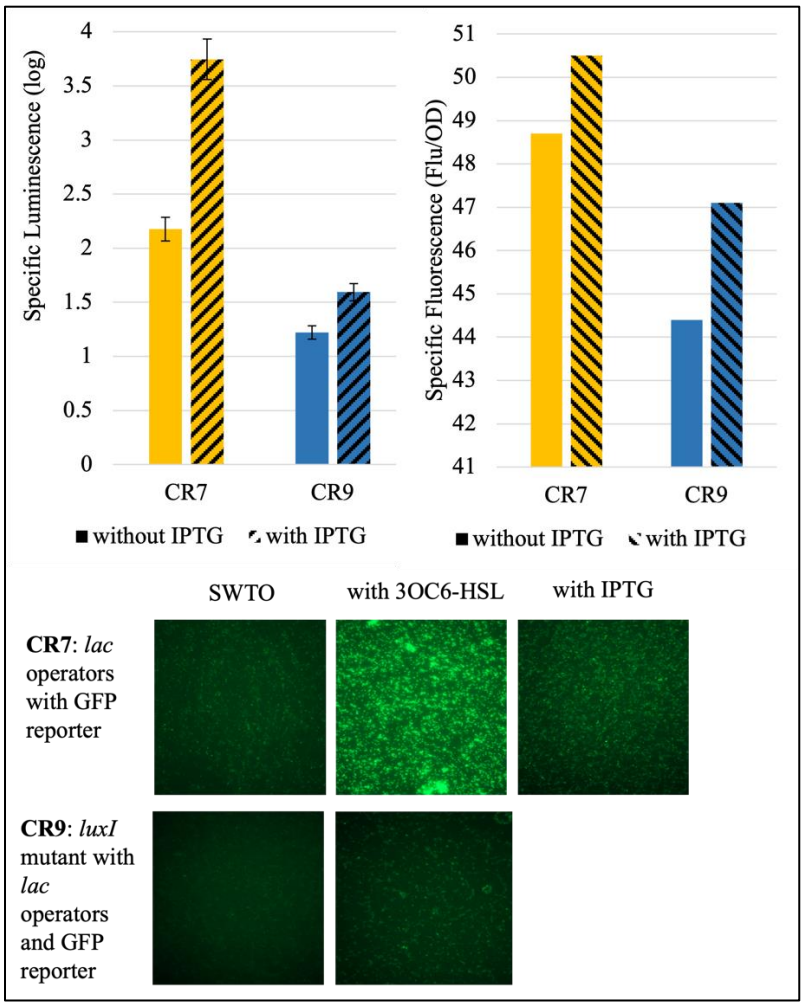

Figure 7: Comparison of luminescence and fluorescence of strains CR7 and CR9 in the presence or absence of IPTG to remove the repression of lacI. Fluorescence images of CR7 and CR9 exposed to 3OC6-HSL and IPTG.

Discussion

This work aimed to create genetically engineered *V. fischeri* strains to artificially control the *lux* operon and gain a better understanding of the LuxI/LuxR-mediated positive feedback loop. *V. fischeri* has been widely use as model system given its relatively straightforward capacity for genetic manipulation. Specifically genetic reporters are a convenient tool to obtain information about activity of specific genes and pathways. GFP is a widely used reporter given its effectiveness

at providing a visual signal as a responsive of the associated gene. However, this long-lasting GFP signal is not suitable when investigating dynamic environments. For this study, the destabilized GFP variant with a half-life of 81 minutes was a better alternative to use as a reporter for the lux operon activity during positive feedback (33, 34). Fluorescence was increased in experiments performed with the genetically modified strains under an activation signal of the *lux* operon with 3OC6-HSL, indicating the efficacy of this reporter. A *V. fischeri* strain not reported here but constructed using this GFP variant as a reporter for the lux operon was used to measure the statistical distributions of cell responses on the two main quorum sensing systems in *V. fischeri* as a response to their respective input signals (35).

In a similar way, the LacI operator system has been widely used as a genetic tool to manipulate transcription of genes, including genetic constructions in *V, fischeri* (36), although not to directly control the *lux* operon. The approach used in this study of site directed mutagenesis successfully created a LacI-controlled lux operon. However, when the activity was compared to the native regulator of this system, the repression was lower than expected. Although the reasons for this difference in regulation are not well understood, we predicted that the differences in structure in the DNA strands during the binding of LacI compared to ArcA did not entirely restrict the interaction of the LuxI-LuxR complex with the "*lux* box". The addition of a second LacI operator, downstream of the "*lux* box", provided a second binding site for LacI, and increased the repression of the *lux* operon to a similar level of regulation by the native regulator, ArcA.

Another component of the genetic manipulation of these strains was the genetic exchange of the *luxI* gene with a mutant allele that renders the gene non-functional. A similar manipulation was previously done in *V. fischeri* and a similar approach was used here. The newly created strains

showed a decreased in luminescence when compared to their parent strains carrying a functional copy of the gene.

Together, the strains created in this study provide practical genetic tools to perform experiments to understand the positive feedback loop in *V. fischeri.* In fact, some of these strains were used in quorum sensing experiments with a localized stimulus in still media (31) and to test wavelike dynamics from mathematical models (37). Deeper knowledge of this process will contribute a better understanding of the dynamics in symbiosis between *Euprymna scolopes* and *V. fischeri*, as well as other symbiosis systems that include microorganisms participating of positive feedback loops. Similarly, the engineered strains from this study could be useful in experiments aiming to comprehend pathways and dynamics of pathogenic members of the Vibrionaceae family.

References

1. Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG. 2000. Vibrio fischeri lux Genes Play an Important Role in Colonization and Development of the Host Light Organ. J Bacteriol 182:4578–4586.

2. Bose JL, Rosenberg CS, Stabb E V. 2008. Effects of luxCDABEG induction in Vibrio fischeri: enhancement of symbiotic colonization and conditional attenuation of growth in culture. Arch Microbiol 190:169–183.

3. Koch EJ, Miyashiro T, McFall-Ngai MJ, Ruby EG. 2014. Features governing symbiont persistence in the squid–vibrio association. Mol Ecol 23:1624–1634.

4. Engebrecht J, Nealson K, Silverman M. 1983. Bacterial bioluminescence: Isolation and genetic analysis of functions from Vibrio fischeri. Cell 32:773–781.

5. Engebrecht J, SIlverman M. 1984. Identification of genes and gene products necessary for bacterial bioluminescence. Proceedings of the National Academy of Sciences 81:4154–4158.

6. Egland KA, Greenberg EP. 1999. Quorum sensing in Vibrio fischeri: Elements of the luxl promoter. Mol Microbiol 31:1197–1204.

7. Fuqua WC, Winans SC. 1994. A LuxR-LuxI type regulatory system activates Agrobacterium Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. J Bacteriol 176:2796–2806.

8. Hastings JW, Nealson KH. 1977. Bacterial Luminescence. Annual Reviews in Microbiology 31:549–595.

9. Boylan M, Miyamoto C, Wall L, Grahan A, Meighen E. 1989. LuxC, D and E genes of the Vibrio fischeri luminescence operon code fot the reductase, transferase and synthase enzymes involved in aldehyde biosynthesis. Photochem Photobiol 49:681–688.

10. Nijvipakul S, Wongratana J, Suadee C, Entsch B, Ballou DP, Chaiyen P. 2008. LuxG is a functioning flavin reductase for bacterial luminescence. J Bacteriol 190:1531–1538.

11. Kuo A, Blough N V., Dunlap P V. 1994. Multiple N-acyl-L-homoserine lactone autoinducers of luminescence in the marine symbiotic bacterium Vibrio fischeri. J Bacteriol 176:7558–7565.

12. Lupp C, Ruby EG. 2004. Vibrio fischeri LuxS and AinS: Comparative Study of two signal synthases. J Bacteriol 186:3873–3881.

13. Fidopiastis PM, Miyamoto CM, Jobling MG, Meighen EA, Ruby EG. 2002. LitR, a new transcriptional activator in Vibrio fischeri, regulates luminescence and symbiotic light organ colonization. Mol Microbiol 45:131–143.

14. Miyamoto CM, Lin YH, Meighen EA. 2000. Control of bioluminescence in Vibrio fischeri by the LuxO signal response regulator. Mol Microbiol 36:594–607.

15. Miyashiro T, Wollenberg MS, Cao X, Oehlert D, Ruby EG. 2010. A single qrr gene is necessary and sufficient for LuxO-mediated regulation in Vibrio fischeri. Mol Microbiol 77:1556–1567.

16. Schaefer AMYL, Hanzelka BL, Eberhard A, Greenberg EP. 1996. Quorum Sensing in Vibrio fischeri: Probing Autoinducer-LuxR interactions with autoinducer analogs. J Bacteriol 178:2897–2901.

17. Egland KA, Greenberg EP. 2000. Conversion of the Vibrio fischeri transcriptional activator, LuxR, to a repressor. J Bacteriol 182:805–811.

18. Lewenza S, Sokol PA. 2001. Regulation of Ornibactin biosynthesis and N-Acyl-L-Homoserine Lactone production by CepR in Burkholderia cepacia. J Bacteriol 183:2212–2218.

19. Stauff DL, Bassler BL. 2011. Quorum sensing in chromobacterium violaceum: DNA recognition and gene regulation by the CviR receptor. J Bacteriol 193:3871–3878.

20. Selin C, Dilantha Fernando WG, de Kievit T. 2012. The PhzI/PhzR quorum-sensing system is required for pyrrolnitrin and phenazine production, and exhibits cross-regulation with RpoS in Pseudomonas chlororaphis PA23. Microbiology (N Y) 158:896–907.

21. Milton DL, Hardman A, Camara M, Chhabra SR, Bycroft BW, Stewart GSAB, Williams P. 1997. Quorum sensing in Vibrio anguillarum: Characterization of the vanI/vanR locus and identification of the autoinducer N-(3-Oxodecanoyl)-L-Homoserine lactone. J Bacteriol 179:3004–3012.

22. Lyell NL, Dunn AK, Bose JL, Stabb E V. 2010. Bright mutants of Vibrio fischeri ES114 reveal conditions and regulators that control bioluminescence and expression of the lux operon. J Bacteriol 192:5103–5114.

23. Bose JL, Kim U, Bartkowski W, Gunsalus RP, Overley AM, Lyell NL, Visick KL, Stabb E V. 2007. Bioluminescence in Vibrio fischeri is controlled by the redox-responsive regulator ArcA. Mol Microbiol 65:538–553.

24. Septer AN, Lyell NL, Stabb E V. 2013. The iron-dependent regulator fur controls pheromone signaling systems and luminescence in the squid symbiont Vibrio fischeri ES114. Appl Environ Microbiol 79:1826–1834.

25. Septer AN, Bose JL, Dunn AK, Stabb E V. 2010. FNR-mediated regulation of bioluminescence and anaerobic respiration in the light-organ symbiont Vibrio fischeri. FEMS Microbiol Lett 306:72–81.

26. Septer AN, Stabb E V. 2012. Coordination of the Arc Regulatory System and Pheromone-Mediated Positive Feedback in Controlling the Vibrio fischeri lux Operon. PLoS One 7:1–11.

27. Sycuro LK, Ruby EG, Mcfall-Ngai M. 2006. Confocal Microscopy of the Light Organ Crypts in Juvenile Euprymna scolopes Reveals Their Morphological Complexity and Dynamic Function in Symbiosis. J Morphol 267:555–568.

28. Stoudenmire JL, Essock-Burns T, Weathers EN, Solaimanpour S, Mrázek J, Stabb E V. 2018. An Iterative , Synthetic Approach To Engineer a High- Performance PhoB-Specific Reporter. Appl Environ Microbiol 84:e00603-18.

29. Ondrey JM, Visick KL. 2014. Engineering Vibrio fischeri for inducible gene expression. Open Microbiol J 8:122–129.

30. Stabb E V, Ruby EG. 2002. RP4-Based Plasmids for Conjugation between Escherichia coli and members of the Vibrionaceae. Methods Enzymol 358:413–426.

31. Patel K, Rodriguez C, Stabb E V., Hagen SJ. 2020. Spatially propagating activation of quorum sensing in Vibrio fischeri and the transition to low population density. Phys Rev E 101:062421.

32. Lyell NL, Colton DM, Bose JL, Tumen-Velasquez MP, Kimbrough JH, Stabb E V. 2013. Cyclic AMP Receptor Protein Regulates Pheromone-Mediated Bioluminescence at Multiple Levels in Vibrio fischeri ES114. J Bacteriol 195:5051–5063.

33. Andersen JB, Sternberg C, Poulsen LK, Bjørn SP, Givskov M, Molin S. 1998. New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. Appl Environ Microbiol 64:2240–2246.

34. Stoudenmire JL, Essock-Burns T, Weathers EN, Solaimanpour S, Mrázek J, Stabb E V. 2018. An Iterative, Synthetic Approach To Engineer a High-Performance PhoB-Specific Reporter. Appl Environ Microbiol 84.

35. Miller T, Patel K, Rodriguez C, Stabb E V., Hagen SJ. 2021. Dimension-reduction simplifies the analysis of signal crosstalk in a bacterial quorum sensing pathway. Sci Rep 11:19719.

36. Visick KL, Hodge-Hanson KM, Tischler AH, Bennett AK, Mastrodomenico V. 2018. Tools for Rapid Genetic Engineering of Vibrio fischeri. Appl Environ Microbiol 84.

37. Patel K, Rodriguez C, Stabb E V, Hagen SJ. 2021. Wavelike propagation of quorum activation through a spatially distributed bacterial population under natural regulation. Phys Biol 18:046008.

APPENDIX B

Supplemental Material for Chapter 2

Table S2.1: Observed and expected values for ESBL-E carriage on selected demographic information collected from the questionnaire filled by all adult participants during their first sample provided.

Demographics or Risk Factors	N	ESBL-E Positive (N=24)	ESBL-E Negative (N=481)	p-value/ FDR p-adjusted
Biological Sex				
Female	349	13 (4%)	336 (96%)	0.12/0.79
Male	156	11 (7%)	145 (93%)	
Race				
Asian	37	4 (11%)	33 (89%)	0.17/0.79
Black or African American	59	5 (8%)	54 (92%)	
White	377	15 (4%)	362 (96%)	
Age				
18-29	234	13 (6%)	221 (94%)	0.17/0.79
30-39	84	2 (2%)	82 (98%)	
40-49	53	3 (6%)	50 (94%)	
50-59	49	2 (4%)	47 (96%)	
60-69	45	1 (2%)	44 (98%)	
70-79	22	2 (9%)	20 (91%)	
80-89	5	0 (0%)	5 (100%)	
Currently Employed				
Yes	340	12 (4%)	328 (96%)	0.08/0.79
No	165	12 (7%)	153 (93%)	
Ethnicity				
Hispanic	42	1 (2%)	41 (98%)	0.84/1
Non-Hispanic	450	23 (5%)	427 (95%)	
Highest Schooling				
Less Than High School	5	0 (0%)	5 (100%)	0.14/0.79
High School/GED	54	2 (4%)	52 (96%)	
Some College	138	11 (8%)	127 (92%)	
Associate degree	39	3 (8%)	36 (92%)	
Bachelor's Degree	149	5 (3%)	144 (97%)	
Master's Degree	97	1 (1%)	96 (99%)	
Ph.D., MD, or JD	23	2 (9%)	21 (91%)	
Annual Household Income				
Less than $20,000	96	5 (5%)	91 (95%)	0.45/0.79
$20,000 to $34,999	72	6 (8%)	66 (92%)	
$35,000 to $49,999	51	3 (6%)	48 (94%)	
$50,000 to $74,999	84	2 (2%)	82 (98%)	
$75,000 to $99,999	63	4 (6%)	59 (94%)	

Over $100,000	132	4 (3%)	128 (97%)	
Residence Type				
Single Family Home	292	11 (4%)	281 (96%)	0.25/0.79
Apartment/Multi Family	209	13 (6%)	196 (94%)	
Homeless	1	0 (0%)	1 (100%)	
Rural, Urban or Suburban				
Urban	130	5 (4%)	125 (96%)	0.68/0.99
Rural	91	3 (3%)	88 (97%)	
Suburban	282	16 (6%)	266 (94%)	
Primary Water Source				
Town/Municipal Water	441	21 (5%)	420 (95%)	0.18/0.79
Private Well	49	1 (2%)	48 (98%)	
Other	15	2 (13%)	13 (87%)	
Exposure to animals at home, farm, or animal care facility[1]				
Dogs and/or Cats	251	9 (4%)	242 (96%)	0.42/0.79
Other than dog or cat	16	0 (0%)	16 (100%)	
Multiple Types	24	1 (24%)	23 (96%)	
None	214	14 (7%)	200 (93%)	
Regular livestock exposure				
Yes	36	0 (0%)	36 (100%)	0.40/0.79
No	468	23 (5%)	445 (95%)	
Number of risk environments exposed to[2]				
1	85	3 (4%)	82 (96%)	0.89/1
2	44	0 (0%)	44 (100%)	
3	20	2 (10%)	18 (90%)	
4	7	0 (0%)	7 (100%)	
5	6	0 (0%)	6 (100%)	
6	1	0 (0%)	1 (100%)	
7	1	0 (0%)	1 (100%)	
None	333	17 (5%)	316 (95%)	
Exposure to healthcare facilities (occupational or visitor)				
Yes	81	4 (5%)	77 (95%)	0.78/0.99
No	412	19 (5%)	393 (95%)	
Food poisoning in last 30 days				
Yes	5	0 (0%)	5 (100%)	1/1
No	500	24 (5%)	476 (95%)	
Gastrointestinal conditions or symptoms in the past 1 month				
Yes	177	8 (5%)	169 (95%)	1/1
No	327	16 (5%)	311 (95%)	

Multiple antibiotics taken this year				
Yes	40	3 (8%)	37 (92%)	0.43/0.79
No	465	21 (5%)	444 (95%)	
Antibiotics use in the past 1 month (oral, topical, or intravenous)				
Yes	47	2 (4%)	45 (96%)	1/1
No	454	22 (5%)	432 (95%)	
Regular medications or supplements taken in last 30 days				
Yes	363	14 (3%)	348 (96%)	0/0.01
No	40	9 (2%)	31 (40%)	
Urinary tract infection in the past 1 month (self-diagnosed or diagnosed by doctor)				
Yes	13	1 (8%)	12 (92%)	0.47/0/79
No	492	23 (5%)	469 (95%)	
International travel (past year)				
Yes	31	1 (3%)	30 (97%)	1/1
No	471	23 (5%)	448 (95%)	
Lived internationally in the last 5 years				
Yes	37	4 (11%)	33 (89%)	0.09/0.79
No	463	20 (4%)	443 (96%)	
Exposure to treated recreational water in the past month				
Yes	127	4 (3%)	123 (97%)	0.47/0.79
No	371	20 (5%)	351 (95%)	
Exposure to untreated recreational water in the past month				
Yes	107	3 (3%)	104 (97%)	0.44/0.79
No	391	21 (5%)	370 (95%)	
Additional house members				
0	57	3 (5%)	54 (95%)	0.74/0.99
>1	447	21 (5%)	426 (95%)	
Additional house members younger than 5 years old				
0	400	19 (5%)	381 (95%)	1/1
>1	92	4 (4%)	88 (96%)	
Eat poultry in the last week				
Yes	421	21 (5%)	400 (95%)	0.78/0.99
No	83	3 (4%)	80 (96%)	
Eat pork or beef in the last week				
Yes	385	16 (4%)	369 (96%)	0.32/0.79
No	119	8 (7%)	111 (93%)	
Eat fish or shellfish in the last week				
Yes	254	10 (4%)	244 (96%)	0.41/0.79
No	248	14 (6%)	234 (94%)	

Eat dairy in the last week				
Yes	473	24 (5%)	449 (95%)	0.39/0.79
No	32	0 (0%)	32 (100%)	
Eat raw fruit or vegetables in the last week				
Yes	439	19 (4%)	420 (96%)	0.22/0.79
No	65	5 (8%)	60 (92%)	

[1]Animals reported in 'Other than dog or cat' include reptiles, birds, rodents/small mammals. If participant reported dog and/or cat in addition to other types, then it was classified in 'Multiple.' [2]Questionnaire offered multiple choices as environmental risk exposure including animal waste, human waste, companion animals, pesticides/herbicides, childcare facilities, K-12 schools, raw meat/poultry, poultry, livestock, veterinary facilities and correctional facilities, data reported includes the number of environments selected.

Table S2.2: Demographic distribution from our study compared to Athens-Clarke County, the state of Georgia and over the U.S. based on data collected from census.gov (2021).

Demographics	Distribution of adults on this study (N=505)	Distribution in Athens Clarke County	Distribution in Georgia	Distribution in the U.S.A*
Asian	7.3%	4.0%	4.8%	5.8%
Black or African American	11.7%	27.8%	33.1%	12.1%
White	74.6%	60.5%	59.0%	61.2%
Mixed	1.8%	4.4%	2.4%	12.6%
Other	2.4%	-	-	-
Hispanic	8.3%			18.8%
Non-Hispanic	89.1%			81.2%
Identified as Female	69.1%			50.5%
Identified as Male	30.8%			49.5%

*From American Community Survey 2-21: ACS-1 year estimates data profiles

Table S2.3: Genome quality assessment using CheckM lineage workflow

Isolate	Genome Size (bp)	Completeness (%)	Contamination (%)	Total of N50 contigs	Longest contig (bp)	Mean contig length (bp)
AREA_482	4,981,979	99.07	0.52	157,133	398,159	25,418
AREA_483	5,438,138	99.97	0.07	5,106,887	5,106,887	543,814
AREA_484	5,299,331	99.97	0.45	5,126,651	5,126,651	1,059,866
AREA_485	5,461,929	99.97	0.33	5,310,577	5,310,577	1,820,643
AREA_486	5,172,013	99.97	0.39	5,101,202	5,101,202	1,724,004
AREA_487	5,270,828	99.97	0.33	5,189,749	5,189,749	1,756,943
AREA_488	5,225,697	99.97	0.36	1,341,054	1,703,105	348,380
AREA_489	5,235,109	99.97	0.33	3,263,269	3,263,269	193,893
AREA_490	5,444,513	99.97	0.60	3,725,720	3,725,720	494,956
AREA_491	5,303,945	99.82	0.35	678,206	728,890	88,399
AREA_492	5,154,401	99.67	0.72	4,934,235	4,934,235	644,300
AREA_493	5,141,066	99.97	0.39	4,722,299	4,722,299	1,028,213
AREA_494	5,373,309	99.97	0.33	5,164,941	5,164,941	1,791,103
AREA_495	5,260,164	99.07	0.49	822,129	2,523,629	328,760
AREA_496	5,335,487	99.97	0.43	5,183,188	5,183,188	1,778,496
AREA_497	4,880,542	99.93	0.10	4,699,031	4,699,031	1,626,847
AREA_498	5,427,885	99.97	1.33	4,679,123	4,679,123	904,648
AREA_499	5,203,150	99.97	0.39	5,203,150	5,203,150	5,203,150
AREA_500	5,327,239	99.97	0.36	5,108,113	5,108,113	197,305
AREA_501	5,210,053	99.37	0.33	687,543	2,330,037	89,829
AREA_502	5,141,719	99.97	0.33	5,039,263	5,039,263	1,713,906

AREA_503	5,274,630	99.97	0.33	5,150,049	5,150,049	1,758,210
AREA_504	5,275,050	99.97	0.45	5,081,105	5,081,105	1,318,763
AREA_505	5,039,223	99.97	0.33	4,916,587	4,916,587	2,519,612
AREA_C483	5,074,575	99.93	1.24	5,074,575	5,074,575	5,074,575
AREA_R482	5,103,194	99.67	0.37	323,367	522,390	60,038
AREA_R483	5,436,943	99.91	0.17	1,148,558	1,580,857	84,952
AREA_R485	5,465,567	99.97	0.48	5,315,488	5,315,488	2,732,784
AREA_R488	5,193,622	99.97	0.36	5,059,100	5,059,100	78,691
AREA_R489	5,224,086	99.97	0.33	3,260,324	3,260,324	200,926
AREA_R490	5,356,116	99.07	0.51	691,421	2,118,936	198,375
AREA_R491	5,358,613	99.97	0.29	5,236,147	5,236,147	1,071,723
AREA_R493	5,112,865	99.97	0.39	1,013,536	2,468,508	340,858
AREA_R497	5,047,689	99.97	0.06	3,122,400	3,122,400	630,961
AREA_R500	5,279,145	99.79	0.51	4,847,616	4,847,616	55,570
AREA_R501	5,229,413	99.97	0.33	3,849,269	3,849,269	116,209
AREA_R503	5,274,177	99.97	0.33	3,889,283	3,889,283	1,758,059
AREA_R504	5,286,355	99.97	0.45	5,082,024	5,082,024	881,059
AREA_R505	5,040,783	99.97	0.33	4,918,147	4,918,147	2,520,392

Table S2.4: Assembly results by isolate performed by Unicycler and visualized in Bandage. Some plasmids were assembled and circularized by Unicycler but not identified in PlasmidFinder. Complete and incomplete assemblies are available in NCBI.

Isolate	Scaffold length	Scaffold status	Scaffold ID	NCBI BioSample Accession Number
AREA_482	4,901,008 93,130	Incomplete Incomplete	Chromosome Plasmid	SAMN33912483
AREA_R482	4,900,803 111,841 94,195	Incomplete Incomplete Incomplete	Chromosome Plasmid Plasmid	SAMN33912484
AREA_483	5,106,888 85,875 83,837 95,437 59,126 6,989	Complete Complete Complete Incomplete Complete Complete	Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912485
AREA_R483	3,960,029 1,148,558 95,437 85,876 83,837 59,124 6,989	Incomplete Incomplete Complete Complete Complete Complete Complete	Chromosome Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912486
AREA_484	5,126,947 93,541 70,016 8,828	Incomplete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912487
AREA_485	5,310,580 149,251 2,101	Incomplete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912488
AREA_R485	5,315,468 150,079	Complete Complete	Chromosome Plasmid	SAMN33912489
AREA_486	5,101,199 65,601 5,210	Complete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912490
AREA_487	5,189,868 79,530 1,549	Complete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912491
AREA_488	5,058,538 96,059 68,388 3,257	Incomplete Incomplete Complete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912492
AREA_R488	5,059,104 135,828 3,257	Complete Incomplete Complete	Chromosome Plasmid Plasmid	SAMN33912493
AREA_489	5,091,059 107,575	Incomplete Complete	Chromosome Plasmid	SAMN33912494

	30,657 7,939	Complete Complete	Plasmid Plasmid	
AREA_R489	5,088,296 107,566 30,649	Incomplete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912495
AREA_490	5,257,305 123,216 47,546 7,939 5,269 3,256	Incomplete Complete Complete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912496
AREA_R490	2,119,052 1,403,314 1,653,205 123,216 47,546 5,269 4,063 1,115	Incomplete Incomplete Incomplete Complete Complete Complete Complete Incomplete	Chromosome Chromosome Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912497
AREA_491	5,186,701 116,786 4,076	Incomplete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912498
AREA_R491	5,242,716 115,977	Incomplete Complete	Chromosome Plasmid	SAMN33912499
AREA_492	4,934,455 123,608 71,888 7,939 5,430 5,166 4,073 2,077	Complete Complete Complete Complete Complete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912500
AREA_493	4,961,184 109,539 70,343	Incomplete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912501
AREA_R493	5,009,157 70,343 34,615	Incomplete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912502
AREA_494	5,164,944 110,786 97,582	Complete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912503
AREA_495	4,852,083 309,561 76,447 22,159	Incomplete Complete Complete Incomplete	Chromosome Plasmid Plasmid Plasmid	SAMN33912504
AREA_496	5,238,477 96,998	Incomplete Complete	Chromosome Plasmid	SAMN33912505
AREA_497	4,699,025 92,616 88,895	Complete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912506

AREA_497R	4,824,639 88,895 66,868 62,520 5,167	Incomplete Complete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid Plasmid	SAMN33912507
AREA_498	4,975,257 211,174 129,755 111,678	Incomplete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912508
AREA_499	5,203,158	Complete	Chromosome	SAMN33912509
AREA_500	5,108,101 143,170 68,037 5,165 3,257	Complete Complete Incomplete Complete Complete	Chromosome Plasmid Chromosome Plasmid Plasmid	SAMN33912510
AREA_R500	4,988,230 105,899 185,370 5,165	Incomplete Incomplete Incomplete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912511
AREA_501	5,003,752 120,161 86,612 2,112	Incomplete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912512
AREA_R501	3,849,274 1,176,107 120,235 86,643	Incomplete Incomplete Complete Complete	Chromosome Chromosome Plasmid Plasmid	SAMN33912513
AREA_502	5,039,267 97,288 5,167	Complete Complete Complete	Chromosome Plasmid Plasmid	SAMN33912514
AREA_503	5,171,793 102,829	Incomplete Complete	Chromosome Plasmid	SAMN33912515
AREA_R503	5,172,055 102,819	Incomplete Complete	Chromosome Plasmid	SAMN33912516
AREA_504	5,081,104 114,903 73,392 5,631	Complete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid	SAMN33912517
AREA_R504	5,082,014 114,914 73,403 5,631 5,214 5,165	Complete Complete Complete Complete Complete Complete	Chromosome Plasmid Plasmid Plasmid Plasmid Plasmid	SAMN33912518
AREA_505	4,916,592 122,636	Complete Complete	Chromosome Plasmid	SAMN33912519
AREA_R505	4,918,152 122,636	Complete Complete	Chromosome Plasmid	SAMN33912520
AREA_C483	5,074,572	Complete	Chromosome	SAMN33912521

Table S2.5: Assembled plasmids carried by each isolate identified by PlasmidFinder with encoded antibiotic genes identified by AMRFinderPlus and encoded virulence genes identified by VirulenceFinder. Plasmid number in each isolate matches the numbers in panels C and D of Figure 1 and Figure 2.

Isolate	Plasmid Type	Size (bp)	Antibiotic Resistance Genes	Virulence Genes
497R	1. IncI1 2. IncFII 3. IncFII 4. Col156	1. 88,895 2. 66,870 3. 62,521 4. 5,167	1. *bla*$_{CTX-M-1}$	1. *cia* 2. *mcbA, traJ, traT*
491	1. IncF 2. IncF	1. 116,786 2. 4,076		
498	1. IncFII 2. IncFIB 3. IncFIB (Phage)	1. 211,173 2. 129,758 3. 111,678	1. – 2. *bla*$_{CTX-M-27}$	1. *aap, eatA, faeF, traT* 2. *traJ, traT*
499	-	-		
497	1. IncF 2. IncI1	1. 92,616 2. 88,895	1. *tetA, tetR* 2. *bla*$_{CTX-M-1}$	1. *anr, ompT* 2. *cia*
495	1. IncF 2. IncF	1. 309,561 2. 76,447	1. *bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$, *bla*$_{TEM-1}$, *aph(3'')-Ib, aph(6)-Id, aac(6')-Ib-cr5, catB3, floR, qacEdelta1, qnrS1, sul1, sul2, terB, terC, terD, terE, tet(A), dfrA1*	1. *terC* 2. *anr, traJ, traT*
483	1. IncFII 2. IncB/O/Z/K 3. IncI2	1. 83,837 2. 85,878 3. 59,126	1. *bla*$_{CTX-M-15}$, *qnrS1*	1. *traT* 2. i*reA*
C483	-	-		
490	1. IncF 2. ColRNAI 3. Col440I	1. 123,216 2. 7,939 3. 3,256	1. *tet(B)*	1, *anr, iucC, iutA, sitA, traT*
492	1. IncF 2. IncFII 3. ColRNAI 4. Col440II 5. Col156 6. Col (BS512)	1. 123,593 2. 71,888 3. 7,939 4. 5,430 5. 5,166 6. 2,077	1. *bla*$_{TEM-1}$, *aph(3'')-Ib, aph(6)-Id, sul2, tet(B)* 2. *bla*$_{TEM-1}$	1. *anr, iucC, iutA, sitA, traT* 2. *mcbA, traJ, traT*
504	1. IncF 2. IncFII 3. Col156	1. 114,922 2. 73,392 3. 5,631	1. *bla*$_{CTX-M-27}$, *aadA5, aph(3'')-Ib, aph(6)-Id, erm(B), mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *senB* 2. *fyuA, traJ*
486	1. IncF 2. Col156	1. 65,601 2. 5,210	1. *bla*$_{CTX-M-27}$	1. *senB*
482	1. IncF 2. Col156 3. IncY	1. 74,522 2. 10,068 3. 88,765		1. *traJ, traT* 2. *senB* 3. -
505	1. IncF	1. 122,636	1. *bla*$_{CTX-M-27}$, *aadA5, aph(3'')-Ib, aph(6)-Id,*	1. *anr, senB, traT*

			mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17	
493	1. IncF 2. IncF	1. 109,542 2. 70,343	1. *$bla_{CTX-M-27}$, aadA5, aph(3")-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *anr, senB, traT* 2. *traT*
496	1. IncF	1. 96,998	1. *bla_{TEM-1}, aph(3")-Ib, aph(6)-Id, qnrS1, sul2, dfrA14*	1. *anr, traJ, traT*
503	1. IncF	1. 102,829	1. *tet(A)*	1. *iutA, senB*
485	1. IncF 2. Col (BS512)	1. 149,251 2. 2,101		1. *anr, iutA, senB, traJ, traT*
487	1. IncF 2. Col (MG828)	1. 79,530 2. 1,549	1. *aph(3")-Ib, aph(6)-Id, sul2, tet(A)*	1. *anr, senB*
488	1. IncFI 2. IncFII 3. Col440I	1. 96,050 2. 68,388 3. 3,257	1. *bla_{TEM-1}, aadA5, aph(3")-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *traJ, traT* 2. *mcbA, traT*
500	1. IncF 2. Col156 3. Col440I	1. 143,188 2. 5,165 3. 3,257	1. *bla_{TEM-1}, aac(3)-IId, aadA5, aph(3")-Ib, aph(6)-Id, mph(A), qacEdelta1, sul1, sul2, tet(A), dfrA17*	1. *senB, traJ, traT*
484	1. IncF 2. IncFII 3. IncQ1	1. 93,541 2. 70,016 3. 8,828	1. *bla_{TEM-1}, aac(3)-IId, aph(3")-Ib, aph(6)-Id, sul2, tet(B), dfrA17*	1. *anr* 2. *traT*
494	1. IncF 2. IncY	1. 110,786 2. 97,582	1. *bla_{TEM-1}*	1. *anr, senB, traT*
489	1. IncF 2. ColRNAI	1. 107,575 2. 7,939	1. *$bla_{CTX-M-27}$*	1. *anr, senB, traT*
501	1. IncF 2. IncI1 3. Col (BS512)	1. 120,161 2. 86,606 3. 2,112	1. *$bla_{CTX-M-27}$*	1. *anr, senB, traT* 2. *cia*
502	1. IncF 2. Col156	1. 97,289 2. 5,167	1. *$bla_{CTX-M-27}$*	1. *anr, senB*

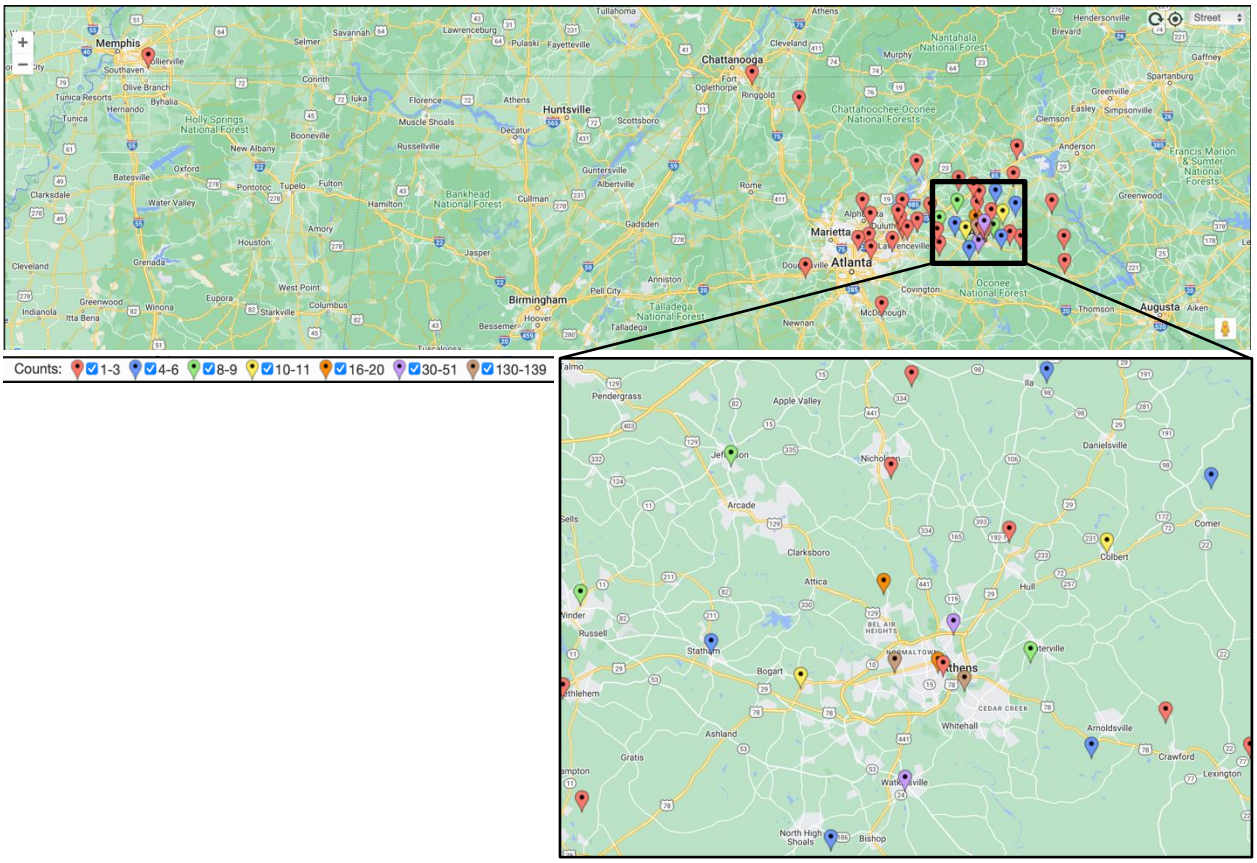

Figure S2.1: Map of residence zip codes reported by adult participants. The number of participants in each zip code is represented by color groups. Zip codes with more than four participants are limited to the Athens, GA area and vicinity. Map generated by EasyMapMaker.com.

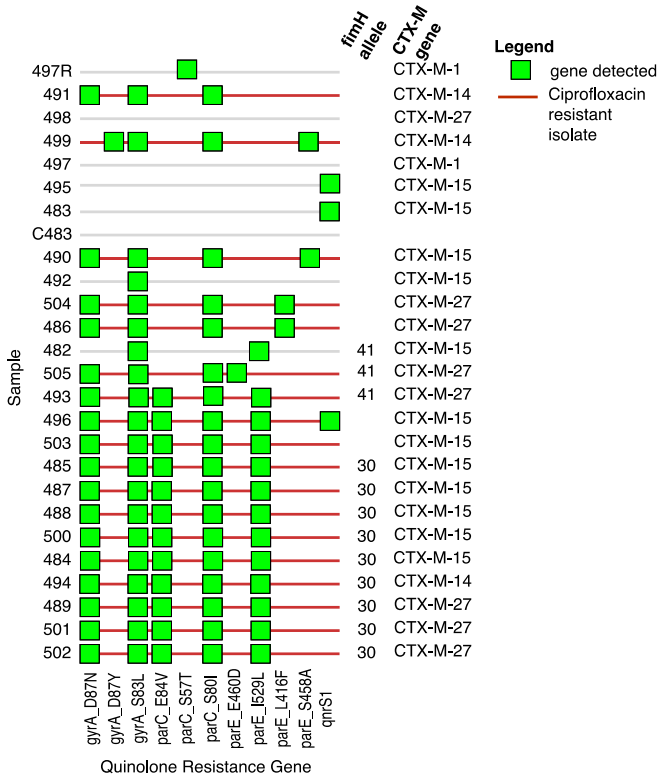

Figure S2.2: Quinolone resistance phenotype and genotype of ESBL-E isolates. Antimicrobial resistance genes from the quinolone class identified by AMRFinder on ESBL-E isolates are shown in green boxes. Samples that exhibited a Ciprofloxacin resistant phenotype are highlighted by red lines. For ESBL-E classified as ST131, their *fimH* allele is indicated, as well as their $bla_{CTX\text{-}M}$ allele if present for comparison of ST131 subclades.

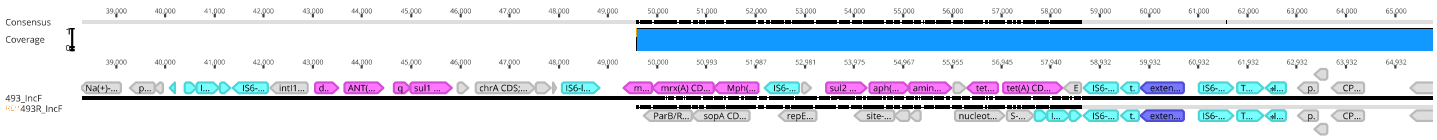

Figure S2.3: Partial alignment of IncF plasmid in isolates 493 and 493R. Alignment region showing insertion sequences (light blue) present in isolate 493 but missing in 493R, encoding multiple antibiotic resistance genes (pink). Beta-lactamase CTX-M-27 gene is shown in purple, also surrounded by IS6 sequences.

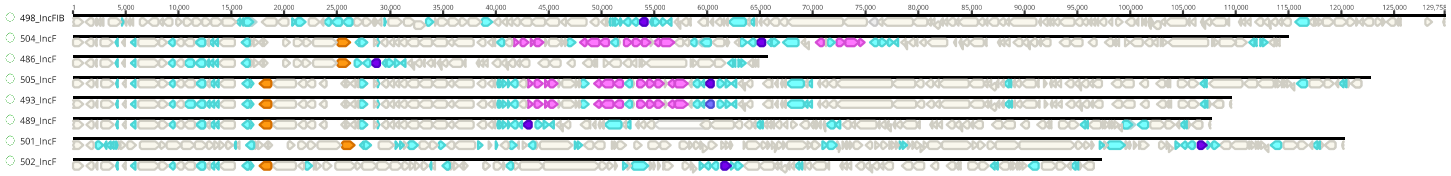

Figure S2.4: Overview of IncF plasmids carrying beta-lactamase CTX-M-27. Transposase and insertion sequences are shown in light blue, antibiotic resistance genes in pink, beta-lactamase CTX-M-27 gene is shown in purple, and plasmid-encoded enterotoxin is colored in orange. Specific antibiotic resistance genes can be found in Table 3. Sequences are not aligned by similarity but organized in the same order as the phylogenetic tree of Figure 1.

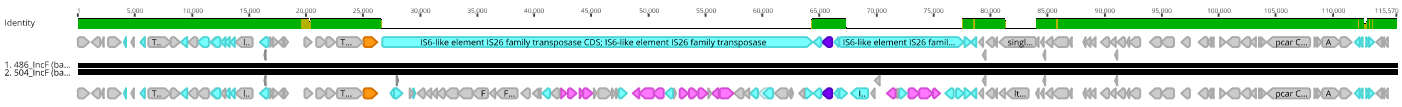

Figure S2.5: Alignment of IncF plasmid in isolates 486 and 504. Alignment region showing insertion sequences (light blue) lacking antibiotic resistance genes that are present in 504 (pink). Beta-lactamase $bla_{CTX\text{-}M\text{-}27}$ gene is shown in purple, also surrounded by IS6 sequences.

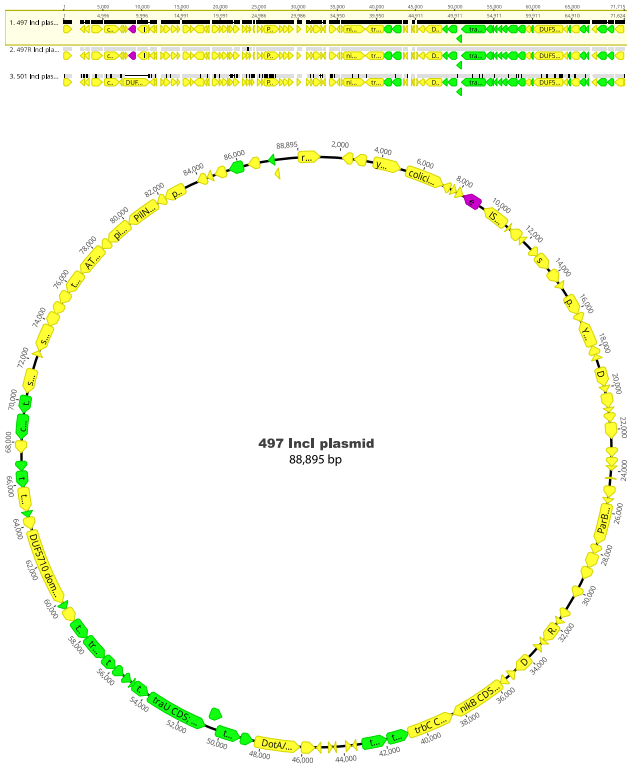

Figure S2.6: Map and alignment of plasmid from the IncI1 group shared between different *E. coli* strains of the same participant. The plasmid replicon of IncI1 was identified by PlasmidFinder. Conjugation-related genes encoded in the plasmid are highlighted in green. Extended-spectrum beta-lactamase gene $bla_{CTX\text{-}M\text{-}1}$ is shown in purple. Above, there is an alignment between all IncI1 plasmids found in this study, including plasmids encoded in different *E. coli* isolates from the same participant (497 and 497R).

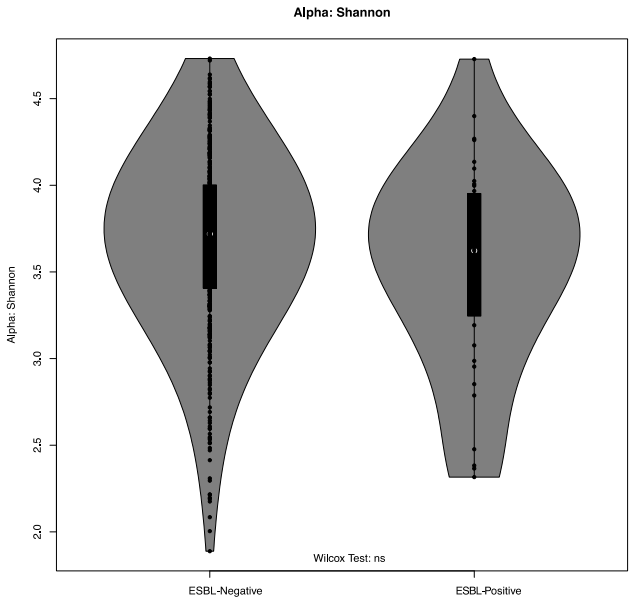

Figure S2.7: Alpha diversity by Shannon index. ASVs found in the initial samples of each participant (555) were tested for alpha diversity comparing ESBL positive and ESBL negative communities. The test was performed using Shannon index in the vegan package of R and confirmed with Wilcoxon test.

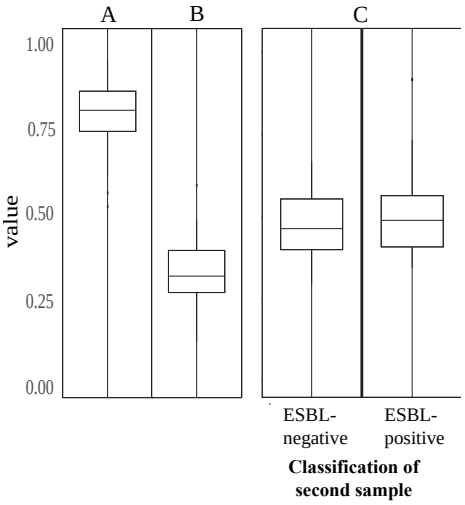

Figure S2.8: Bray-Curtis distances between ESBL-positive samples. Dissimilarity distances calculated between initial samples from each ESBL-positive individual (A), between initial samples and their duplicates (B) and, between initial and second sample from the same individual, grouped by ESBL status of second sample.

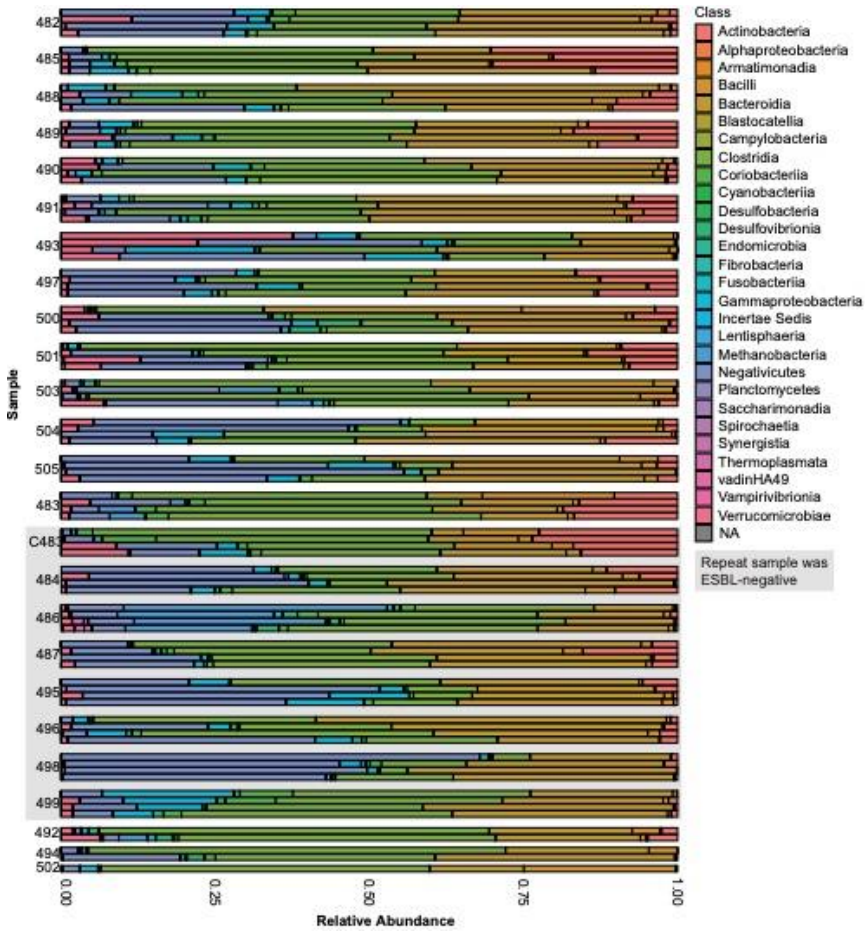

Figure S2.9: Microbiota composition of ESBL-positive individuals. Taxonomic bar plot displaying the relative abundance of Classes in ESBL-positive samples. The number assigned to each cluster represents an individual that was ESBL-positive in their first sample. Each bar cluster includes (from top to bottom): original sample, duplicate of original, repeated sample and duplicate of repeat sample when available. Clusters on the gray box represent individuals that were ESBL-negative on their repeated sample.

APPENDIX C

Supplemental Material for Chapter 4

Table S4.1: Identification of Single-Cell Amplified Genomes

contig ID	Taxa	completeness (%)	Contamination (%)	NCBI Bioproject	IMG files number	Genome Size (bp)
Ga0222834	NA	0%	0	PRJNA467388	181038	36919
Ga0222822	NA	0%	0	PRJNA467377	181045	69440
Ga0222824	NA	0%	0	PRJNA467379	181048	192121
Ga0222831		0%	0	PRJNA467386	181053	29105
Ga0222818	NA	0%	0	PRJNA467374	181056	225976
Ga0222821	NA	0%	0	PRJNA467376	181069	67563
Ga0222814	NA	0%	0	PRJNA467370	181076	174805
Ga0222763	NA	0%	0	PRJNA467320	181418	94142
Ga0222773	NA	0%	0	PRJNA467330	181425	101656
Ga0222768	NA	0%	0	PRJNA467325	181618	60653
Ga0222806	NA	0%	0	PRJNA467362	181627	63088
Ga0222781	NA	0%	0	PRJNA467338	181633	64980
Ga0222794	NA	0%	0	PRJNA467350	181635	162620
Ga0222771	NA	0%	0	PRJNA467328	181646	75021
Ga0222811	NA	0%	0	PRJNA467367	181655	293054
Ga0222897	NA	0%	0	PRJNA467452	181669	366606
Ga0222884	NA	0%	0	PRJNA467438	181685	280128
Ga0223030	NA	0%	0	PRJNA467587	182455	63400
Ga0223029	NA	2%	0	PRJNA467586	182423	220780
Ga0223037	NA	2%	0	PRJNA467594	182443	613916
Ga0222825	NA	3%	0	PRJNA467380	181050	258915
Ga0222807	NA	8%	0	PRJNA467363	181645	217246

Ga0223023	d__Bacteria;p__Bdellovibrionota;c__Oligoflexia;o__Oligoflexales;f__Oligoflexaceae;g__Oligoflexus;	9%	0	PRJNA467580	182467	1536353
Ga0222813	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Curtissbacterales;f__GWA2-41-24;g__;s__	9%	0	PRJNA467369	181071	193334
Ga0222766	Unclassified Bacteria	10%	0	PRJNA467323	181423	202566
Ga0223036	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__UBA5976;g__UBA5976;s__	10%	0	PRJNA467593	182438	494494
Ga0223017	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__Limnohabitans sp002347905	11%	0	PRJNA467575	182464	657060
Ga0222767	d__Bacteria;p__Bdellovibrionota;c__JAFGOL01;o__JAFGOL01;f__JAFGOL01;g__;s__	11%	0	PRJNA467324	181427	663713
Ga0223022	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	11%	0	PRJNA467579	182462	422202
Ga0222793	d__Bacteria;p__Patescibacteria;c__ABY1;o__BM507;f__GWC2-42-12;g__;s__	11%	0	PRJNA467349	181638	283419
Ga0222888	d__Bacteria;p__Bdellovibrionota;c__UBA2428;o__UBA2428;f__UBA2428;g__;s__	12%	1.724137931	PRJNA467442	181617	1385963
Ga0223055	d__Bacteria;p__Bdellovibrionota;c__UBA1018;o__CAIPTA01;f__CAIPTA01;g__JAKFYM01;s__	12%	0	PRJNA467619	182466	1042730
Ga0223027	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	13%	0	PRJNA467584	182440	278407
Ga0223019	Unclassified Bacteria	13%	0	PRJNA467577	182458	473167
Ga0223035	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	13%	0	PRJNA467592	182431	603011
Ga0222810	d__Bacteria;p__Acidobacteriota;c__Thermoanaerobaculia;o__UBA5066;f__Gp7-AA6;g__Gp7-AA6;s__	13%	0	PRJNA467366	181649	1457681
Ga0223008	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus_A;s__	14%	0	PRJNA467566	182160	402877
Ga0222782	d__Bacteria;p__Omnitrophota;c__Koll11;o__GIF10;f__Profunditerraquicolaceae;g__;s__	14%	0	PRJNA467339	181630	505180
Ga0222802	d__Bacteria;p__Bdellovibrionota;c__Bacteriovoracia;o__Bacteriovoracales;f__Bacteriovoracaceae;g__Bacteriovorax;s__	15%	0	PRJNA467358	181625	1084151

Ga0222835	d__Bacteria;p__Patescibacteria;c__Gracilibacteria;o__UBA1369;f__UBA1369;g__PALSA-1335;s__	15%	0	PRJNA467389	181067	201616
Ga0222941	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__;f__;g__;s__	15%	0	PRJNA467496	181228	633296
Ga0222804	d__Bacteria;p__Elusimicrobiota;c__Elusimicrobia;o__UBA1565;f__JACRDR01;g__;s__	15%	0	PRJNA467360	181619	689005
Ga0222817	d__Bacteria;p__Planctomycetota;c__SZUA-567;o__H5-PLA8;f__H5-PLA8;g__JACQFI01;s__	16%	0	PRJNA467373	181040	3323775
Ga0223006	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	16%	0	PRJNA467564	182184	750600
Ga0223061	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	17%	0	PRJNA467625	182426	317482
Ga0222902	d__Bacteria;p__Bdellovibrionota;c__UBA1018;o__UBA1018;f__UBA1018;g__;s__	17%	0	PRJNA467457	181704	1156634
Ga0222786	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__PALSA-1336;s__	18%	1.724137931	PRJNA467343	181623	570942
Ga0222997	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	19%	0	PRJNA467555	182156	812441
Ga0223012	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	19%	0	PRJNA467570	182164	1044564
Ga0222819	d__Bacteria;p__Patescibacteria;c__ABY1;o__SBBC01;f__CAIJZX01;g__JACQBO01;s__	20%	0	PRJNA467375	181061	326900
Ga0223048	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__CAMDFR01;g__;s__	20%	0	PRJNA467605	182439	620516
Ga0223025	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	21%	0	PRJNA467582	182450	355191
Ga0222830	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Prevotella;s__	21%	0	PRJNA467385	181064	591895
Ga0222960	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhizobiales;f__Kaistiaceae;g__;s__	21%	0	PRJNA467515	181211	1055590
Ga0222999	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	21%	0	PRJNA467557	182141	1095763
Ga0223058	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	21%	0	PRJNA467622	182434	628052

Ga0222949	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	21%	0	PRJNA467504	181193	723796
Ga0223044	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus;s__Methylopumilus sp903849185	23%	0	PRJNA467601	182436	257849
Ga0222959	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	23%	0.751380629	PRJNA467514	181204	711322
Ga0222832	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Gemmatimonadales;f__GWC2-71-9;g__JACOVF01;s__	23%	1.089901099	PRJNA467387	181037	873342
Ga0222816	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Moranbacterales;f__GWC2-37-73;g__;s__	23%	0	PRJNA467372	181081	498919
Ga0223034	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__PALSA-747;s__	23%	0	PRJNA467591	182446	337680
Ga0222876	d__Bacteria;p__Bdellovibrionota;c__Bacteriovoracia;o__Bacteriovoracales;f__Bacteriovoracaceae;g__;s__	24%	0	PRJNA467430	181695	1161949
Ga0222953	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	24%	0	PRJNA467508	181206	287218
Ga0222938	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__UBA5976;g__ATZT02;s__	24%	0	PRJNA467493	181207	474360
Ga0223031	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	25%	0	PRJNA467588	182463	421722
Ga0222972	d__Bacteria;p__Omnitrophota;c__Koll11;o__GIF10;f__Profundiferraquicolaceae;g__;s__	25%	0	PRJNA467530	182177	747408
Ga0223052	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	25%	0	PRJNA467616	182444	606174
Ga0222779	d__Archaea;p__Micrarchaeota;c__Micrarchaeia;o__Anstonellales;f__Anstonellaceae;g__1-14-0-10-45-29;s__	25%	0	PRJNA467336	181622	520788
Ga0223016	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__Sediminibacterium sp017853235	25%	0.2	PRJNA467574	182449	693268
Ga0222987	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__;s__	26%	0	PRJNA467545	182157	493548
Ga0222774	d__Bacteria;p__Elusimicrobiota;c__Elusimicrobia;o__2-01-FULL-59-12;f__2-01-FULL-59-12;g__;s__	26%	0	PRJNA467331	181650	1165900

Ga0223005	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	26%	0	PRJNA467563	182140	667977
Ga0222983	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	26%	0	PRJNA467541	182149	1295494
Ga0223026	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	26%	0.1	PRJNA467583	182457	792867
Ga0222979	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__Planktophila sp000372185	26%	0	PRJNA467537	182161	577376
Ga0222856	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	27%	0	PRJNA467410	181078	1099461
Ga0222957	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Micropepsales;f__Micropepsaceae;g__Rhizomicrobium;s__	27%	0	PRJNA467512	181191	2050216
Ga0222789	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__UBA1547;g__UBA1547;s__	27%	0.862068966	PRJNA467346	181639	309034
Ga0223056	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	27%	0	PRJNA467620	182456	978782
Ga0222934	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	27%	0	PRJNA467489	181182	690700
Ga0222937	d__Bacteria;p__Bdellovibrionota;c__Bacteriovoracia;o__Bacteriovoracales;f__Bacteriovoracaceae;g__Bacteriovorax;s__	27%	0	PRJNA467492	181199	1235099
Ga0222764	d__Bacteria;p__Patescibacteria;c__ABY1;o__SG8-24;f__GWF2-40-263;g__;s__	27%	0	PRJNA467321	181419	441631
Ga0222791	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	28%	0	PRJNA467347	181654	1122562
Ga0222978	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Micropepsales;f__Micropepsaceae;g__Rhizomicrobium;s__	28%	0	PRJNA467536	182167	1816603
Ga0223018	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus;s__	28%	0	PRJNA467576	182460	343466
Ga0223004	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__;s__	28%	0	PRJNA467562	182154	533515
Ga0222956	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	28%	0	PRJNA467511	181229	1382311

Ga0222950	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	29%	0	PRJNA467505	181181	608656
Ga0222800	d__Bacteria;p__Elusimicrobiota;c__Elusimicrobia;o__UBA1565;f__UBA9628;g__GWA2-66-18;s__	29%	0	PRJNA467356	181631	956531
Ga0223045	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__MAG-120802;s__	29%	0	PRJNA467602	182445	445203
Ga0222815	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__GCA-2401735;s__	30%	0	PRJNA467371	181044	726740
Ga0222783	d__Bacteria;p__Elusimicrobiota;c__Elusimicrobia;o__FEN-1173;f__;g__;s__	30%	0	PRJNA467340	181647	566903
Ga0222853	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	30%	0	PRJNA467407	181065	1560384
Ga0222958	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	30%	0.15576324	PRJNA467513	181197	830605
Ga0222998	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__NS11-12g;f__UBA955;g__UBA955;s__UBA955 sp002347985	30%	0	PRJNA467556	182174	920403
Ga0222981	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Chthoniobacterales;f__UBA6821;g__UBA6821;s__	30%	0	PRJNA467539	182151	764337
Ga0223032	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Hylemonella;s__	31%	0	PRJNA467589	182422	978462
Ga0222954	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	31%	0	PRJNA467509	181214	959088
Ga0223010	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus_A;s__	31%	0	PRJNA467568	182176	721509
Ga0223047	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	31%	0	PRJNA467604	182453	674145
Ga0223028	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	31%	0.5	PRJNA467585	182447	368326
Ga0222962	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	31%	0	PRJNA467517	181226	997236
Ga0222879	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	32%	0	PRJNA467433	181659	715121

Ga0223057	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__Polynucleobacter sp018687515	32%	0	PRJNA467621	182432	647416
Ga0222955	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	32%	0	PRJNA467510	181221	1096575
Ga0222993	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__Sediminibacterium sp024640255	32%	0	PRJNA467551	182181	865689
Ga0223060	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	32%	0	PRJNA467624	182430	742793
Ga0222935	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	33%	0	PRJNA467490	181185	798066
Ga0222887	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	33%	0	PRJNA467441	181687	688541
Ga0223051	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__	33%	0	PRJNA467615	182433	523326
Ga0222798	d__Bacteria;p__Patescibacteria;c__JABMPQ01;o__;f__;g__;s__	33%	0.925925926	PRJNA467354	181621	248171
Ga0222770	d__Bacteria;p__Nitrospirota;c__Nitrospiria;o__JACQBW01;f__JACQBW01;g__JACQBW01;s__	33%	0	PRJNA467327	181637	594256
Ga0223024	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	34%	0	PRJNA467581	182429	569257
Ga0222788	d__Bacteria;p__Myxococcota;c__Polyangia;o__Fen-1088;f__Fen-1088;g__;s__	34%	0.430107527	PRJNA467345	181632	1977780
Ga0222886	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	34%	0	PRJNA467440	181672	1041648
Ga0223062	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	34%	0	PRJNA467626	182459	537890
Ga0222826	d__Bacteria;p__Patescibacteria;c__Gracilibacteria;o__Peribacterales;f__Peribacteraceae;g__JACRIG01;s__	35%	1.913875598	PRJNA467381	181079	610167
Ga0222778	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Levybacterales;f__UBA12049;g__;s__	35%	0	PRJNA467335	181426	434279
Ga0223000	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	35%	0	PRJNA467558	182173	832148

Ga0222868	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	35%	0	PRJNA467422	181682	1038511
Ga0222942	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__	36%	0	PRJNA467497	181188	538187
Ga0222894	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__Nanoperiomorbaceae;g__;s__	36%	0	PRJNA467449	181663	435569
Ga0222881	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	36%	0.46728972	PRJNA467435	181665	989726
Ga0222927	d__Bacteria;p__Patescibacteria;c__ABY1;o__Magasanikbacterales;f__GWA2-37-8;g__JAKANM01;s__	36%	0	PRJNA467482	181227	837622
Ga0223007	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__ZLKRG08;s__	36%	0	PRJNA467565	182166	663446
Ga0222996	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__	36%	0	PRJNA467554	182142	401626
Ga0223046	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__MAG-120802;s__	37%	0	PRJNA467603	182442	557127
Ga0222882	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__SYFN01;s__	37%	0.434625136	PRJNA467436	181670	1140287
Ga0223015	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	37%	0	PRJNA467573	182454	520177
Ga0222890	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	38%	0	PRJNA467444	181674	1391559
Ga0223049	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__UBA952;s__	38%	0	PRJNA467606	182448	976075
Ga0222829	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Moranbacterales;f__GWC2-37-73;g__;s__	38%	0	PRJNA467384	181052	414778
Ga0222910	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	38%	0	PRJNA467465	181694	846643
Ga0222869	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Rhodocyclaceae;g__;s__	39%	0	PRJNA467423	181688	1367732
Ga0222896	d__Bacteria;p__Patescibacteria;c__ABY1;o__BM507;f__JAKAJF01;g__;s__	39%	0.99009901	PRJNA467451	181681	797663
Ga0222769	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Daviesbacterales;f__UBA10151;g__JACPZA01;s__	39%	0	PRJNA467326	181428	549449

Ga0222796	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Daviesbacterales;f__UBA10151;g__UBA10151;s__	39%	0.925925926	PRJNA467352	181634	443207
Ga0222787	d__Bacteria;p__Patescibacteria;c__ABY1;o__BM507;f__UBA917;g__CAIUCZ01;s__	39%	0	PRJNA467344	181643	711126
Ga0222799	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rickettsiales;f__Deianiraeaceae;g__CAMCQG01;s__	40%	0.473933649	PRJNA467355	181626	415524
Ga0222863	d__Bacteria;p__Dependentiae;c__Babeliae;o__Babeliales;f__Babeliaceae;g__;s__	40%	0.710900474	PRJNA467417	181657	509028
Ga0222945	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Pacearchaeales;f__GW2011-AR1;g__CABMGE01;s__	40%	0	PRJNA467500	181209	511757
Ga0222866	d__Bacteria;p__Cyanobacteriota;c__Vampirovibrionia;o__Vampirovibrionales;f__;g__;s__	40%	0.854700855	PRJNA467420	181662	1045674
Ga0222805	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Moranbacterales;f__UBA2206;g__JAKLFF01;s__	40%	0	PRJNA467361	181651	502730
Ga0222872	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	40%	0.909090909	PRJNA467426	181616	1119005
Ga0222812	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__UBA1406;f__GWC2-37-13;g__2-01-FULL-40-42;s__	40%	0	PRJNA467368	181642	414131
Ga0222966	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Levybacterales;f__UBA12049;g__GWB1-37-8;s__	40%	0	PRJNA467521	182150	530232
Ga0223040	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	41%	0	PRJNA467597	182451	1288559
Ga0222878	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	41%	0.029205607	PRJNA467432	181673	1045043
Ga0222795	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__UBA2100;g__UBA10103;s__	41%	0	PRJNA467351	181636	886944
Ga0223039	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Moranbacterales;f__SCTJ01;g__SCTJ01;s__	42%	0	PRJNA467596	182465	708563
Ga0222940	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhizobiales;f__Beijerinckiaceae;g__Methylocystis;s__	42%	0	PRJNA467495	181220	1346777
Ga0223014	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__UBA7398;s__UBA7398 sp009927575	42%	0.720720721	PRJNA467572	182421	677666

Ga0223021	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__MAG-120802;s__	42%	0	PRJNA467578	182461	630542
Ga0222898	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	42%	0	PRJNA467453	181698	740787
Ga0222862	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	42%	0	PRJNA467416	181054	660089
Ga0223042	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoluna;s__	43%	0	PRJNA467599	182435	628280
Ga0222883	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoluna;s__	43%	0	PRJNA467437	181676	675188
Ga0222971	d__Bacteria;p__Bdellovibrionota;c__Bdellovibrionia;o__Bdellovibrionales;f__Bdellovibrionaceae;g__;s__	43%	0	PRJNA467526	182159	1791770
Ga0223009	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	43%	0	PRJNA467567	182137	1180262
Ga0222885	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__ZLKRG08;s__	43%	0	PRJNA467439	181666	884717
Ga0222911	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	43%	0	PRJNA467466	181660	638404
Ga0222995	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	43%	0	PRJNA467553	182148	682444
Ga0222854	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	43%	0	PRJNA467408	181068	1339349
Ga0222765	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Micropepsales;f__Micropepsaceae;g__SZUA-430;s__	43%	0	PRJNA467322	181424	2034632
Ga0222780	d__Bacteria;p__Patescibacteria;c__ABY1;o__Veblenbacterales;f__UBA10138;g__;s__	43%	0	PRJNA467337	181629	532020
Ga0222907	d__Bacteria;p__Bdellovibrionota;c__Bdellovibrionia;o__Bdellovibrionales;f__UBA6776;g__JALHOQ01;s__	43%	0.892857143	PRJNA467462	181697	1519589
Ga0223041	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__Polynucleobacter sp009928245	43%	0.3	PRJNA467598	182452	892929
Ga0222772	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Moranbacterales;f__GWC2-37-73;g__;s__	44%	1.080108011	PRJNA467329	181652	536797

Ga0222784	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__JAAFIP01;s__	44%	0.11682243	PRJNA467341	181640	1171661
Ga0223059	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	44%	0	PRJNA467623	182427	518821
Ga0222943	d__Bacteria;p__Desulfobacterota_B;c__Binatia;o__JACPRU01;f__JACPRU01;g__;s__	44%	0	PRJNA467498	181196	1551821
Ga0222929	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__Megaira;s__Megaira sp005791235	44%	3.55450237	PRJNA467484	181195	785564
Ga0222992	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__Sediminibacterium sp017856545	44%	0	PRJNA467550	182138	818833
Ga0222908	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__Nanopelagicus sp010030895	45%	0	PRJNA467463	181661	603055
Ga0222994	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium;s__	45%	0	PRJNA467552	182178	1232084
Ga0222939	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	45%	0.15576324	PRJNA467494	181213	1316252
Ga0222797	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__UBA583;g__;s__	45%	0	PRJNA467353	181628	829825
Ga0222986	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA6257;f__2-01-FULL-56-20;g__JACRFR01;s__	46%	0	PRJNA467544	182135	364572
Ga0223033	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Aquisediminimonas;s__	46%	0	PRJNA467590	182441	1123400
Ga0223013	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__IMCC26077;s__	46%	0	PRJNA467571	182424	831374
Ga0222947	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	47%	0.496495327	PRJNA467502	181224	1498584
Ga0222980	d__Bacteria;p__Nitrospirota;c__UBA9217;o__UBA9217;f__UBA9217;g__JALNZF01;s__	47%	1.818181818	PRJNA467538	182170	1857796
Ga0222874	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	47%	0.46728972	PRJNA467428	181675	1474744

Ga0222928	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__2-02-FULL-39-11;f__JACQJU01;g__;s__	47%	0	PRJNA467483	181189	793994
Ga0222852	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	47%	0.024594196	PRJNA467406	181049	1702644
Ga0223053	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	47%	0	PRJNA467617	182425	677245
Ga0222948	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	48%	0	PRJNA467503	181186	1198248
Ga0223043	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus;s__	48%	0	PRJNA467600	182428	665839
Ga0222952	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	48%	0	PRJNA467507	181184	620216
Ga0222936	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	48%	0.15576324	PRJNA467491	181192	1291392
Ga0222968	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	48%	0	PRJNA467523	182153	1217543
Ga0222965	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__UBA2103;g__C7867-001;s__	49%	0	PRJNA467520	182163	667611
Ga0223054	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	49%	0.4	PRJNA467618	182420	787722
Ga0222828	d__Bacteria;p__Patescibacteria;c__Saccharimonadia;o__Saccharimonadales;f__UBA10212_A;g__;s__	49%	0	PRJNA467383	181059	764279
Ga0222905	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__MAG-120802;s__	49%	0	PRJNA467460	181693	949831
Ga0222944	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	49%	0.029205607	PRJNA467499	181203	1427893
Ga0222985	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA9219;f__UBA9219;g__JAFALY01;s__	49%	0	PRJNA467543	182175	1217198
Ga0222871	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	49%	0.029205607	PRJNA467425	181705	1751579
Ga0223011	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	49%	0	PRJNA467569	182158	813813
Ga0222916	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__;s__	50%	0	PRJNA467471	181194	804809

Ga0222931	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	50%	0	PRJNA467486	181210	1006572
Ga0222906	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__AAA044-D11;s__	50%	2.543859649	PRJNA467461	181658	1291973
Ga0222861	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	50%	0	PRJNA467415	181063	855597
Ga0222990	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	50%	0	PRJNA467548	182183	1319148
Ga0222845	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	50%	0.303030303	PRJNA467399	181066	1426335
Ga0222857	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	50%	0.526315789	PRJNA467411	181083	814500
Ga0222864	d__Bacteria;p__Patescibacteria;c__JAEDAM01;o__Absconditabacterales;f__Absconditicoccaceae;g__UM-FILTER-40-9;s__	50%	0	PRJNA467418	181701	858115
Ga0223001	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__Sediminibacterium sp024640255	50%	0	PRJNA467559	182180	1279321
Ga0222961	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	50%	0.046728972	PRJNA467516	181218	1436936
Ga0222892	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	50%	0.404040404	PRJNA467447	181686	1515927
Ga0222899	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	51%	0	PRJNA467454	181668	1266087
Ga0222803	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__UBA2100;g__UBA10103;s__	51%	0	PRJNA467359	181641	804132
Ga0222850	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__PALSA-747;s__	51%	0	PRJNA467404	181055	962396
Ga0222991	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__Polynucleobacter sp009928245	51%	0.1	PRJNA467549	182136	984151
Ga0222922	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	51%	0.205761317	PRJNA467477	181190	1344043
Ga0222973	d__Bacteria;p__Patescibacteria;c__Gracilibacteria;o__Peribacterales;f__Peribacteraceae;g__;s__	52%	0	PRJNA467531	182144	933663

Ga0222909	d__Bacteria;p__Myxococcota;c__Polyangia;o__Polyangiales;f__Polyangiaceae;g__PMG-095;s__	52%	2.741935484	PRJNA467464	181679	6440431
Ga0222912	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__SG8-39;g__SZAS-79;s__	52%	0.501812099	PRJNA467467	181700	1874703
Ga0222923	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__JAJYJE01;s__	52%	0	PRJNA467478	181198	685008
Ga0222900	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__SYFN01;s__	52%	0.042480884	PRJNA467455	181692	1433797
Ga0222880	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans_A;s__	53%	0.856697819	PRJNA467434	181699	1665280
Ga0222975	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoluna;s__	53%	0.097465887	PRJNA467533	182165	800080
Ga0222873	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Hylemonella;s__	53%	0	PRJNA467427	181677	1647988
Ga0222976	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__AAA044-D11;s__	53%	0	PRJNA467534	182152	952254
Ga0223003	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Bradyrhizobium;s__	53%	0	PRJNA467561	182168	2278690
Ga0222836	d__Bacteria;p__UBA10199;c__UBA10199;o__SPLOWO2-01-44-7;f__SPLOWO2-01-44-7;g__;s__	55%	2.100840336	PRJNA467390	181072	1633101
Ga0222984	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	55%	0	PRJNA467542	182179	856183
Ga0222919	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	55%	0.303030303	PRJNA467474	181215	1524333
Ga0222870	d__Bacteria;p__Patescibacteria;c__ABY1;o__SG8-24;f__GWF2-40-263;g__XYB2-FULL-45-11;s__	56%	0	PRJNA467424	181696	660511
Ga0222974	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Fonsibacter;s__	57%	0	PRJNA467532	182147	702743
Ga0222917	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__M0103;s__	57%	0	PRJNA467472	181201	1814102
Ga0222842	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	57%	0	PRJNA467396	181042	950819
Ga0222930	d__Bacteria;p__Armatimonadota;c__Fimbriimonadia;o__Fimbriimonadales;f__Fimbriimonadaceae;g__JAEVZK01;s__	57%	0	PRJNA467485	181202	3214256

Ga0222858	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__Aquirufa sp024642485	57%	0.404040404	PRJNA467412	181051	1625864
Ga0222982	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	57%	0.321261682	PRJNA467540	182146	1493115
Ga0222851	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__	57%	1.578947368	PRJNA467405	181060	780544
Ga0222977	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus;s__	58%	0	PRJNA467535	182169	799444
Ga0222967	d__Bacteria;p__Patescibacteria;c__WWE3;o__UBA101185;f__UBA10185;g__;s__	58%	1.724137931	PRJNA467522	182139	791437
Ga0222970	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__Planktophila sp000372185	58%	0	PRJNA467525	182162	823779
Ga0222877	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	58%	0.091510903	PRJNA467431	181689	1613123
Ga0222847	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	58%	0.029205607	PRJNA467401	181075	1866790
Ga0222837	d__Bacteria;p__Cyanobacteriota;c__Vampirovibrionia;o__Gastranaerophilales;f__Gastranaerophilaceae;g__CAIPUE01;s__	58%	1.709401709	PRJNA467391	181077	1251817
Ga0222925	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__Aquirufa sp024642485	60%	0.060606061	PRJNA467480	181212	1745497
Ga0222989	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	60%	0.205761317	PRJNA467547	182172	1605408
Ga0222932	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__	60%	0.588235294	PRJNA467487	181216	902533
Ga0222859	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__Polynucleobacter sp018687515	61%	0	PRJNA467413	181058	1196583
Ga0222933	d__Bacteria;p__UBP7;c__UBA6624;o__;f__;g__;s__	61%	1.282051282	PRJNA467488	181223	2087468
Ga0222776	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__UBA2103;g__1-14-0-10-45-20;s__	63%	1.123595506	PRJNA467333	181420	675008
Ga0222915	d__Bacteria;p__Patescibacteria;c__ABY1;o__BM507;f__UBA917;g__UBA919;s__	63%	0.990099901	PRJNA467470	181183	850566

Ga0222920	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__NS11-12g;f__UBA955;g__JABMMZ01;s__	63%	0.714285714	PRJNA467475	181222	1556582
Ga0222914	d__Bacteria;p__Patescibacteria;c__ABY1;o__UBA1558;f__UBA10009;g__2-01-FULL-45-10-B;s__	64%	0	PRJNA467469	181187	711577
Ga0223002	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	64%	0	PRJNA467560	182182	1017056
Ga0222792	d__Bacteria;p__Bdellovibrionota;c__Bdellovibrionia;o__Bdellovibrionales;f__UBA6776;g__RBG-16-40-8;s__	64%	0	PRJNA467348	181648	1722359
Ga0222901	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhizobiales;f__Beijerinckiaceae;g__CAIUPE01;s__CAIUPE01 sp009921785	65%	0	PRJNA467456	181690	1893358
Ga0222893	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA9339;f__UBA9339;g__;s__	65%	0.775193798	PRJNA467448	181671	1265797
Ga0222969	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__PALSA-1336;s__	65%	0	PRJNA467524	182143	918625
Ga0222895	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	65%	1.99376947	PRJNA467450	181678	2061066
Ga0222775	d__Bacteria;p__Patescibacteria;c__ABY1;o__BM507;f__GWE2-39-37;g__GWE2-39-37;s__	66%	0	PRJNA467332	181624	890315
Ga0222855	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	66%	0.909090909	PRJNA467409	181073	1718240
Ga0222903	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	66%	0.46728972	PRJNA467458	181683	1879743
Ga0222951	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Rhodoferax_C;s__Rhodoferax_C sp009924455	67%	0	PRJNA467506	181200	1924952
Ga0222904	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__;s__	67%	0.473933649	PRJNA467459	181680	725622
Ga0222924	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rickettsiales;f__;g__;s__	67%	0	PRJNA467479	181205	897338
Ga0222823	d__Bacteria;p__Bdellovibrionota;c__Bacteriovoracia;o__Bacteriovoracales;f__Bacteriovoracaceae;g__;s__	67%	2.232142857	PRJNA467378	181074	3065250
Ga0222777	d__Bacteria;p__Patescibacteria;c__Microgenomatia;o__Levybacterales;f__UBA12049;g__GWB1-37-8;s__	67%	2.777777778	PRJNA467334	181653	942487

Ga0222963	d__Bacteria;p__Spirochaetota;c__Leptospirae;o__Leptospirales;f__Leptospiraceae;g__;s__	67%	0	PRJNA467518	182155	2962049
Ga0222889	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	67%	0	PRJNA467443	181684	1729105
Ga0222913	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f__Chitinophagaceae;g__Sediminibacterium;s__	68%	0.492610837	PRJNA467468	181225	2199041
Ga0222926	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Nanopelagicus;s__Nanopelagicus sp009705675	68%	0.588235294	PRJNA467481	181219	925497
Ga0222849	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	68%	0.769869159	PRJNA467403	181080	2048544
Ga0222848	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__	68%	0.693333333	PRJNA467402	181047	1285026
Ga0222808	d__Bacteria;p__Patescibacteria;c__ABY1;o__SBBC01;f__CAIJZX01;g__CAIJZX01;s__	68%	0	PRJNA467364	181656	965778
Ga0222867	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	69%	0.058411215	PRJNA467421	181702	1884330
Ga0222946	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__Planktophila;s__	69%	0	PRJNA467501	181217	1104738
Ga0222918	d__Bacteria;p__Patescibacteria;c__Doudnabacteria;o__UBA920;f__UBA920;g__PALSA-1336;s__	70%	0	PRJNA467473	181208	999675
Ga0222785	d__Bacteria;p__Patescibacteria;c__Gracilibacteria;o__UBA1369;f__UBA1369;g__PALSA-1335;s__	70%	0.925925926	PRJNA467342	181620	1088118
Ga0222844	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Spirosomaceae;g__Aquirufa;s__	71%	0.303030303	PRJNA467398	181043	1908888
Ga0222891	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	71%	0.075934579	PRJNA467445	181691	2043667
Ga0222838	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Berkiellales;f__Berkiellaceae;g__;s__	71%	0.191570881	PRJNA467392	181046	1886969
Ga0222860	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylopumilus;s__Methylopumilus sp903849185	72%	0	PRJNA467414	181041	954870
Ga0222841	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	72%	0.934579439	PRJNA467395	181057	2007071

Ga0222865	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	73%	0.046728972	PRJNA467419	181664	1888952
Ga0222827	d__Bacteria;p__Cyanobacteriota;c__Vampirovibrionia;o__Gastranaerophilales;f__Gastranaerophilaceae;g__CAIPUE01;s__	73%	0	PRJNA467382	181084	1522784
Ga0222801	d__Bacteria;p__Omnitrophota;c__Koll11;o__2-01-FULL-45-10;f__2-01-FULL-45-10;g__2-01-FULL-45-10;s__	75%	3.225806452	PRJNA467357	181644	1329806
Ga0222921	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae_B;g__Limnohabitans;s__	77%	0.340732087	PRJNA467476	181230	2448911
Ga0223038	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Chthoniobacterales;f__Terrimicrobiaceae;g__UBA967;s__	78%	0.7	PRJNA467595	182437	2547393
Ga0222875	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoluna;s__	79%	0.584795322	PRJNA467429	181703	1394984
Ga0222964	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Chthoniobacterales;f__Terrimicrobiaceae;g__UBA967;s__	79%	1.351351351	PRJNA467519	182145	2269530
Ga0222840	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__UBA7398;s__UBA7398 sp009927575	81%	0	PRJNA467394	181039	1370840
Ga0222839	d__Bacteria;p__Omnitrophota;c__Koll11;o__GIF10;f__Profunditerraquicolaceae;g__;s__	82%	1.075268817	PRJNA467393	181082	1676519
Ga0222843	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__Nanopelagicaceae;g__ZLKRG08;s__	86%	0.701754386	PRJNA467397	181062	1302617
Ga0222846	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Rhodoluna;s__	88%	27.21247563	PRJNA467400	181070	2393559
Ga0222988	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter;s__Polynucleobacter sp009928245	89%	0.2	PRJNA467546	182171	1823971

Table S4.2: Summary of processed read counts per sample

Sample (site_dataset)	Sequenced reads	Filtered reads	Merged reads	Non-RNA reads	Mapped reads
S1_MG	93,011,342	92,580,044	46,505,671	NA	31,193,206
S2_MG	92,234,456	91,171,522	46,117,228	NA	35,679,559
S3_MG	118,746,752	118,485,616	59,373,376	NA	46,951,370
S4_MG	107,382,280	107,166,292	53,691,140	NA	41,899,714
S5_MG	95,387,236	95,151,634	47,693,618	NA	37,807,391
S6_MG	94,392,454	94,183,688	47,196,227	NA	37,173,512
S1_MT	84,506,930	26,966,978		7,416,086	2,820,687
S2_MT	80,339,152	34,877,490		8,889,765	3,557,583
S3_MT	117,012,470	37,060,100		9,199,782	4,480,327
S4_MT	142,382,978	44,146,232		11,894,714	5,296,866
S5_MT	121,687,856	38,864,856		11,036,143	5,561,296
S6_MT	104,530,082	61,989,162		25,358,614	15,510,143