Multi-site sampling and risk prioritization of antibiotic resistance genes in sewage environments

by

Chengzhen L. Dai

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2019

Author: ……..
Department of Electrical Engineering and Computer Science
January 24, 2019

Certified by: ……..
Eric Alm, PhD
Department of Biological Engineering
Thesis Supervisor

Accepted by: ……..
Katrina LaCurts, PhD
Chairman, Department Committee on Graduate Theses

Multi-site sampling and risk prioritization of antibiotic resistance genes in sewage environments

by

Chengzhen L. Dai

Submitted to the Department of Electrical Engineering and Computer Science
on January 24, 2019,
in partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The spread of antibiotic resistance across human and environmental habitats is a global public health challenge. In this study, we investigate the public health relevance of antibiotic resistance found in wastewater by combining metagenomic sequencing of wastewater environments with risk prioritization of resistance genes. We find that many of the genes commonly found in wastewater are not readily present in humans. Ranking antibiotic resistance genes based on their potential pathogenicity and mobility reveals that most of the antibiotic resistance genes in wastewater are not directly clinically relevant. Residential sewage was found to be of greater risk to human health than wastewater treatment plants and can be as risky as hospital effluent. Across countries, we show that differences in antibiotic resistance can, in some cases, resemble differences in antibiotic drug consumption. Finally, we find that the flow of antibiotic resistance genes is influenced by geographical distance and environmental selection.

Thesis Supervisor: Eric Alm, PhD
Title: Professor of Biological Engineering

Acknowledgements

First and foremost, I want to thank Professor Eric Alm for his guidance, mentorship, and advising throughout the development of this research. His curiosity, enthusiasm, encouragement, and research expertise has helped me grow as a researcher. I would also like to thank Professor Carlo Ratti at the Senseable City Lab for providing me the resource and opportunity to be part of the Underworlds team. Many thanks to Claire Duvallet and Siavash Isazadeh for being the first people to welcome me into the Alm Lab and always encouraging me to learn and grow. Additional thanks to Shinkyu Park, whose support preceded the start of my time at MIT and who has been nothing but selfless in being a mentor. I will forever remember the countless hours we spent driving around MIT's campus and Massachusetts collecting samples and the bond we built during that time. To the members of the Alm Lab, thank you for being some of the best people I've ever worked with; every one of you have played an essential role in my growth as a researcher and as a person.

A second round of thanks goes to all of my friends who have truly made my time at MIT memorable, both during this past year and a half and during my undergraduate years. Their love and support have helped me become the person that I am today. Thank you to all the friends that I have been fortunate to lived with and thank you to Camp Kesem, Amphibious Achievement, and Potlucks for being communities I can always rely on.

It goes without saying that I would not be where I am today if it wasn't for all the love and support that my parents, Chengtao Li and Linhua Dai, have given me throughout my life. I am forever grateful for their selflessness. Thank you to my sister, Alyson Dai, for being my biggest cheerleader.

Table of Contents

List of Figures

Chapter 1

Introduction

The spread of antibiotic resistant bacteria and resistance genes is a global public health challenge.[1–3] Antibiotic resistance genes (ARGs) confer resistance to antibiotics and are present in both bacterial pathogens and the broader environment.[4–7] The transfer of genes between bacteria through horizontal gene transfer is of particular concern for antibiotic resistance, as it can facilitate the transmission of ARGs from environmental reservoirs to human pathogens.[6,8–11] While current efforts to monitor antibiotic resistance are largely limited to clinical settings,[12–14] government agencies have begun to recognize the importance of antibiotic resistance in the environment.[15,16] As health officials explore the need for environmental surveillance of antibiotic resistance, it is critical to understand the human health relevance of ARGs found in different environments.

Efforts to monitor antibiotic resistance in the environment face the challenge of interpreting the threat that environmental resistomes pose to human health. Environmental bacteria are difficult and costly to culture.[17,18] As such, the ability to identify the evolution, transmission, and host range of resistance genes is limited.[18] To overcome this challenge, studies have employed metagenomic sequencing to measure antibiotic resistance in the environment.[19–21] These studies often establish the presence of resistance genes in an environment as a risk to human health. However, antibiotic resistance genes do not all pose the same risk to human health and the presence of certain genes may be more indicative of the organisms harboring these genes than of a real public health threat.[21] In order for resistance genes to be of relevance to human health, they should reside on mobile genetic elements and be hosted by human bacterial pathogens.[21,22] Most resistance genes in the environment, however, are unlikely to be found in human-associated bacteria, as they often occupy different habitats and are

phylogenetically distant.[18,21,22] Therefore, studies need to move beyond reporting the presence and abundance of ARGs in an environment and also assess their relevance to human health.[21–24]

Wastewater has been proposed as an important source of environmental antibiotic resistance[25–28] as well as an environment for monitoring the prevalence of antibiotic resistance in humans.[16,29,30] Previous studies of antibiotic resistance in wastewater have primarily focused on hospital effluents and wastewater treatment plants (WWTPs), locations which are widely viewed as hotspots for antibiotic resistant bacteria and horizontal gene transfer.[20,27,28,31,32] WWTPs serves as the interface between human society and the environment, as sewage from various sources meet at WWTPs and undergo treatment processes before being released into the environment. Despite treatment processes that remove human-associated bacteria, ARGs are still prevalent in wastewater effluents.[27,28,33] Hospital wastewater has also been proposed as an important source of environmental antibiotic resistance due to the high levels of antibiotic use and resistant bacteria amongst hospital patients.[17,31] Abundances of ARGs in hospital effluents have been found to be higher than in downstream environments such as WWTPs and surface water.[19,20,27,34] However, hospitals contribute less than 1% of the total amount of municipal sewage at WWTPs and have been found to have little influence on the levels of antibiotic resistance observed at WWTPs.[31,32,34,35]

In this study, we combine multi-site sampling of upstream and downstream wastewater with risk prioritization of ARGs to evaluate the public health relevance of antibiotic resistance found in sewage environments. We show that many ARGs found in wastewater are not human-associated and most genes in wastewater are likely not of risk to human health. We also find that the abundance of antibiotic resistance genes in wastewater for certain classes of antibiotics mirrors antibiotic consumption across countries. Lastly, by comparing resistomes sampled within and across countries, we show that the diversity of resistance gene is shaped by environmental selection and geography. Taken together, this study provides direct evidence for the hypothesis that different genes

and environments pose different risks to human health and illustrates the complexities in interpreting the public health relevance of resistomes.

Chapter 2

Material and Methods

2.1 Wastewater collection and processing.

Grab wastewater samples (250 mL) were taken from the manholes at 13 different residential catchment sites in Boston, MA (3 neighborhoods); Cambridge, MA (4 neighborhoods); Seoul, South Korea (3 neighborhoods); and Kuwait City, Kuwait (3 neighborhoods). We also collected samples at pump stations preceding WWTPs in Boston, MA and Kuwait City, Kuwait. Samples from these pump stations represent the influent of WWTPs as these pump stations directly feed wastewater into the treatment plants. Samples from manholes were collected from the selected manhole with a commercial peristaltic pump (Boxer) sampling at a rate of 100 mL/min. Samples from WWTPs were collected using a sampling pole. 30 ml of collected sewage were filtered through 0.2-μm PTFE membrane filters. PTFE membrane filters were kept in RNAlater at -80 degrees Celsius until DNA extraction. The lab filtration system consisted of a Masterflex peristaltic pump (Pall), Masterflex PharMed BPT Tubing (Cole-Palmer), 47 mm PFA filter holders (Cole-Palmer) and 47mm PTFE Omnipore filter membranes (Millipore).

2.2 DNA extraction and shotgun metagenomic sequencing.

0.2-μm filter membranes were thawed in ice. RNAlater was removed and filters were washed with phosphate-buffered saline (PBS) buffer twice. Metagenomic DNA was extracted from each filter with Power Water extraction kit (MO BIO Laboratories Inc.), according to manufacturer's instructions. Sample concentrations were quantified with the Quant-iT PicoGreen dsDNA Assay (Life Technologies) and normalized to equal

concentration. Paired-end libraries were prepared with 100-250 pg of DNA using the Illumina Nextera XT DNA Library Preparation Kit according to the manufacturer's instructions. Insert sizes and concentrations for each library were determined with an Agilent Bioanalyzer DNA 100 kit (Agilent Technologies). Libraries were finally sequenced on the Illumina NextSeq platform at the MIT Biomicro Center to generate 2 x 150 bp paired reads. Approximately 10 million reads were generated for each sample (~5 million for each pair).

2.3 Processing of shotgun metagenomic sequencing data.

Raw paired-end DNA sequences (FASTQ reads) were quality controlled prior to any analysis. Low-quality reads and adaptor sequences were removed using Trimmomatic with the ILLUMINACLIP parameters: 'NexteraPE-PE.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:50'.[36] To remove sequences resulting from human contamination, we used Bowtie2 in default mode to align reads to the human reference genome (GRCh37) and remove the mapped reads from downstream analysis.[37] The remaining sequences were used in further analysis.

2.4 Comparison of wastewater resistomes to published fecal and wastewater samples

For comparison with hospital effluent wastewater and human feces, whole metagenome shotgun reads were downloaded from ENA or SRA. Hospital effluent wastewater sequences was obtained from Rowe *et al.* (2017), ENA accession PRJEB12083; Human fecal samples were obtained from Lim *et al.* (2014), ENA accession ERP002391, for South Korea and Obregon-Tito *et al.* (2015), SRA accession PRJNA268964, for the United States.[19,38,39] Raw fastq reads from these samples were processed as above and all further analyses were performed together with sequences generated in this study.

2.5 Quantification of antibiotic resistance genes

To identify and quantify the abundance of antibiotic resistance genes in each quality-controlled metagenomic sample, we used ShortBRED.[40]S ShortBRED first clusters proteins based on shared amino acid identity and identifies unique marker sequences for each cluster that distinguish them from close homologues. ShortBRED then quantifies the abundance of each protein cluster by mapping reads to those unique markers. Since antibiotic resistance genes can share homology with genes of non-resistance functions, ShortBRED provides greater accuracy than mapping reads to the entire protein sequence.

We generated ShortBRED markers from the Structured ARG reference database (SARG). The SARG database contains 4,049 amino acid sequences and was constructed by integrating ARDB and CARD. We used 100% identity for clustering. We then mapped the clustered sequences to the Integrated Microbial Genomes database, version 3.5, for homology mapping and generation of unique markers. In total, ShortBRED generated 9,807 marker sequences. We used these marker sequences to quantify antibiotic resistance gene abundance in metagenomes by mapping paired fastq reads to them at 99% sequence identity. Abundances were normalized to reads per kilobase per million reads (RPKM).

2.6 Risk ranking of antibiotic resistance genes

We used ARG Ranker to prioritize antibiotic resistance genes based on their risk relevance to human health.[41] Building off of the criteria risk ranking outlined in Martinez *et al.* (2015), ARG Ranker takes the genes in the SARG database[42] and assess their prevalence in publicly available whole genome and metagenomic sequences. Since transferability is a major bottleneck that determines an ARG's risk potential, genes found on plasmids of previously-sequenced isolates of pathogens are ranked higher in risk potential (Rank 1-3) than genes not found in plasmids (Rank 4-5). Amongst mobile genes (Rank 1-3), those with a diverse range of host phylogeny are ranked higher (Rank 1-2) than those with a less phylogenetically diverse range of hosts (Rank 3). We consider the

genes in the top two ranks as clinically-relevant as they have the greatest potential for transfer between host and environments. Abundance in anthropogenic environments versus natural environments differentiate genes between the top two ranks, with genes having a higher abundance being ranked higher.

2.7 Diversity analysis

Beta diversity was calculated at gene level using the Jensen-Shannon distance (JSD). Nucleotide diversity was measured using Wright's F_{ST}, as implemented in metaSNV.[43] We first aligned metagenomic reads to a database of representative sequence. This database included a single nucleotide reference sequence for each gene found by ShortBRED. In total, 264 ARGs were assessed. We used the nucleotide sequences from the CARD database that shared > 99% similarity width the SARG database (N = 3136), as the SARG database only included amino acid sequences and was constructed using CARD.[44] We called and filtered variant with metaSNV using the default setting. Wright's F_{ST} was then calculated using the flags: -div. We only analyze genes present in all three countries. Statistical analyses were perform using the *scipy* Python package.

Chapter 3

Results and Discussion

3.1 The resistome of residential sewage is a complex mixture of human-associated and environmental ARGs

To understand patterns of antibiotic resistance across human-associated environments, we compared the metagenomes of residential sewage with those from similar wastewater environments and the human gut microbiome. We collected and sequenced wastewater from 13 residential manholes in Boston, MA; Cambridge, MA; Seoul, South Korea; and Kuwait City, Kuwait. We also collected wastewater from pump stations directly preceding WWTPs in Boston and Kuwait (Materials and Methods); these samples represent WWTP influents. To compare residential sewage with other wastewater environments and human stool, we downloaded and reprocessed publicly available metagenomic data of adult stool samples in the U.S. and South Korea and a hospital effluent in the United Kingdom. We used ShortBRED with the SARG database to quantify the abundance of antibiotic resistance genes (ARGs) from metagenomes (Material and Methods).

Residential sewage contains more ARGs than human stool and WWTP influent, but less than hospital effluent. The median number of antibiotic resistance genes in residential sewage was 161 (Figure 1A). In comparison, hospital effluent and pump station samples had median ARG counts of 238.5 and 125.5 genes, respectively (Figure 1A). All types of wastewater samples contained more ARGs than stool samples, which had medians of 50 genes in South Korea and 29.5 genes in the U.S. We found similar trends in abundance measurements (Figure 1B). Previous studies have implicated both hospital effluent and WWTPs as hotspots for antibiotic resistance, with hospital effluent harboring more antibiotic resistance than WWTPs.[19,20,28] Our results are consistent with these

expectations. Moreover, the higher levels of antibiotic resistance in residential wastewater than at WWTPs suggest that residential wastewater is also a major reservoir of resistance genes and is a source of antibiotic resistance for WWTPs.

To identify common characteristics of antibiotic resistance in residential wastewater, we looked for genes shared across catchment sites and found a set of 42 genes present in at least one sample from each residential manhole (Figure 1A). These 42 core genes accounted for the majority of ARG abundance in residential sewage (Figure 1B). All of these core genes were also observed at high abundances in hospital effluent and WWTP influent (Figure 1A, 1B). In human fecal samples, these core sewage ARGs made up approximately 90% of the total ARG abundance (median = 96.2% and 90.8% in US and South Korean individuals, respectively), confirming that the majority of antibiotic resistance found in the human gut microbiome are also found in sewage (Figure 1B). Three South Korean fecal samples had high abundance of ARGs (>50%) that were not part of the core set, with many of these genes conferring multidrug or beta-lactam resistance.

However, not all of the core sewage genes are found in human stool, suggesting that many of the ARGs measured in sewage may not be directly relevant to human health. Individual stool samples only carried approximately half of the 42 core sewage genes at a time (median = 22 and 18 genes for South Korean and US samples, respectively; Figure 1C). Few core sewage genes were ubiquitously present in human stool, as only a fifth to a third of the core sewage genes were found in more than 90% of individuals (14 genes in more than 90% of South Korean stool samples; 9 in US stool). In fact, 10 of the 42 core genes were not observed in any human stool sample, suggesting that they are of environmental origin (Figure 1C). Since the transfer of genes from environmental bacteria to human pathogens is less frequent than between human-associated bacteria,[22] these 10 core genes likely do not pose an immediate threat to human health. Therefore, not all ARGs identified in wastewater are likely to have equal relevance to human health. Some may reflect genes carried by most humans while others are likely to be derived from the environment.

3.2 Most ARGs in sewage are not clinically-relevant, but upstream sewage captures more human-related resistomes

To better understand how the presence of ARGs in residential sewage relates to human health risks, we categorized genes based on their potential pathogenicity. We used the method described in Zhang *et al.* (in preparation), which ranks the risk of gene variants based on the variant's observed host pathogenicity, mobility, host range, and anthropogenic prevalence (Materials and Methods). In this method, variants into the top two ranks have been previously observed on plasmids and in a phylogenetically diverse range of hosts, including human pathogens. We therefore refer to the variants in these two ranks as clinically-relevant, as there exists published evidence of these mobile genes posing a substantial risk for the dissemination of resistance.[21]

Most of the ARGs found in sewage and humans are not clinically-relevant, and the majority of the core sewage genes are also not clinically-relevant. Using Zhang *et al.*'s approach, we found that clinically-relevant variants make up less than 50% of the total antibiotic resistance gene abundance in all of the wastewater and human samples we surveyed (Figure 2). These results also held when we looked at the top three ranks, which represents all mobile ARGs (Supplementary Figure 1). Amongst the 42 core sewage genes, clinically-relevant variants were found for only 11 of the core sewage genes and represented less than a 15% of the total ARG abundance in residential sewage (Figure 2, light red). These results directly support the hypothesis proposed by Martinez *et al.* (2015), in which the number of antibiotic resistance genes that are actually acquired by human pathogens and lead to clinical complications are low compared to the number of sequences classified as resistance genes in metagenomic studies. Thus, simply quantifying the presence and abundance ARGs in a given sample does not necessarily measure the health relevance of that sample's resistome.

All types of upstream wastewater harbor more clinically-relevant ARGs than the influent of WWTPs. We compared the presence and abundance of clinically-relevant variants in

residential sewage to the traditionally studied environment of wastewater treatment plants. In both countries (Kuwait and US) where we sampled upstream and downstream sewage, residential wastewater had higher abundances of clinically-relevant variants than the respective WWTP influent (Figure 2). At the same time, WWTP influent contained fewer variants found in human stool than residential wastewater (Supplementary Figure 2). Of the 29 clinically-relevant variants present in at least one healthy U.S. human stool sample, 90% (N = 26) were present in U.S. residential wastewater while only 76% (N = 22) were found in U.S. WWTP influent. More generally, 82% (N = 270) of all of the variants found in U.S. stool samples (N = 328) were observed in U.S. residential sewage whereas only 59% (N = 195) of them were identified downstream at the pump station. The microbial composition of human waste is known to decrease in similarity to the human fecal microbiota as it progresses through the sewage system.[11] Our results suggest this decrease also occurs for resistomes, both clinically-relevant and non-clinically-relevant ones, as upstream sampling better reflects the collection of microbes and antibiotic resistance in the contributing human population.

Although hospital effluent has more overall ARGs and is traditionally thought of as a hotspot for environmental antibiotic resistance, we found that residential sewage can harbor as many clinically-relevant variants as hospital effluent. U.K. hospital effluent had abundances of clinically-relevant variants ranging from 821 to 1555 RPKM (Figure 2). By comparison, the residential wastewater of South Korea and Kuwait had abundances of clinically-relevant variants ranging from 478 to 933 RPKM and 200 to 850 RPKM, respectively (Figure 2). Previous studies implicating hospital wastewater as a major source of environmental antibiotic resistance have largely compared hospital wastewater with environments further downstream such as wastewater treatment plants or surface water.[19,20,27,34] Our results thus challenge the prevailing hypothesis that hospital sewage represents the most concerning source of environmental antibiotic resistance genes and suggests that all upstream sewage may serve as important reservoirs of clinically-relevant genes.

3.3 Antibiotic resistance patterns across geography reflect human activity for some antibiotic classes

To assess whether ARGs in residential sewage reflect population-level antibiotic consumption, we compared the abundance of antibiotic resistance genes with available consumption data from the IQVIA MIDAS database.[45] We limited our analysis to South Korea and the U.S., where consumption data included both hospital and retail use.

For certain antibiotic classes, resistance across countries reflects antibiotic consumption patterns. Consumption of aminoglycoside and beta-lactam antibiotics is higher in South Korea than the U.S., as reflected by the median abundance of resistance genes to these antibiotics in the respective residential sewage samples (Figure 3A). South Korean samples also had higher abundances of chloramphenicol resistance than US samples (Figure 3A). Unlike the U.S., which phased out chloramphenicol use in the 1960s, South Korea continued its usage until 2013. The presence of chloramphenicol resistance may therefore be the result of persistent resistance, as previous studieshave found that resistance genes can persist for a long time after their introduction into the microbial flora.[46] However, this hypothesis needs further validation with better consumption data and direct antibiotic susceptibility testing. These resistance patterns were also observed in the human stool samples and were more evident in clinically-relevant genes than non-clinically-relevant ones (Supplementary Figure 3, 4).

Nonetheless, other antibiotic classes have resistance patterns which do not reflect known country-level differences in antibiotic consumption. Sulfonamide-trimethoprim consumption is higher in the U.S. than South Korea, but resistance to sulfonamide and trimethoprim, both separately and combined, were higher in South Korean samples (Figure 3B). Similarly, tetracycline consumption is higher in the U.S. than South Korea but median abundance of tetracycline resistance genes showed the opposite trend. These inconsistencies emphasize how multiple factors contribute to the antibiotic resistance observed in the environment.[47] Other factors, such as antibiotic use in agriculture and environmental contamination, can also drive resistance but data on them is limited.[48,49]

As efforts are made to fully understand the public health relevance of environmental antibiotic resistance, more comprehensive data on antibiotic use across sectors as well as better approaches to measuring antibiotic resistance in the environment are needed.

3.4 The flow of antibiotic resistance gene varies at different geographic scales

The composition of resistomes differs between residential catchment sites within a city. Pairs of samples from different residential manholes in the same city had higher beta diversity than sample pairs from the same manhole (median JSD = 0.30 vs. 0.27, respectively; Figure 4a). Samples from the same manhole were collected one after another with ~5 minute breaks in between (Materials and Methods). Therefore, different sets of humans likely contributed to the ARGs in each sample, resulting in the observed differences between these samples.[50,51] Across manholes, however, the physical conditions of the wastewater environment are also different, with levels of oxygen and temperature often varying between sites.[17] These variations likely contribute to differences in microbial composition and consequently, ARG composition.[11,52,53] As expected, beta diversity between samples from different countries are higher than all within-city comparisons (median JSD = 0.53, Figure 4A). Thus, these differences in resistome composition may reflect selection resulting from different environmental conditions in individual manholes.

Despite differences in the overall resistome composition between different catchments, the nucleotide diversity of individual ARGs remain similar across manholes in the same city. We aligned metagenomic reads against a representative nucleotide sequence to identify single nucleotide polymorphisms for each antibiotic resistance gene (Materials and Method). We then calculated F_{ST} values to quantify genetic diversity between samples for each gene that had sufficient coverage and polymorphisms (N = 35; Materials and Method).[54] F_{ST} is a measure of genetic differentiation, with values ranging from 0 to 1 where 0 represents no substructure and 1 means completely different alleles between the subpopulations.[54] Overall, pairs of samples from different manholes within

the same city did not have significantly different F_{ST} values than pairs of samples from the same manhole (median F_{ST} = 0.08 versus 0.08, respectively). That is, variants of a gene present across multiple locations within a city were as similar to each other as those found in consecutive samples taken from one manhole. As expected, genes were dissimilar across countries, suggesting that there exist barriers to the distribution of ARGs across larger geographic distance (median F_{ST} = 0.14 for across country comparisons vs median F_{ST} = 0.08 for within country comparisons; Figure 4B). To understand how the composition of these 35 genes differ across catchment sites and across geography, we evaluated the beta diversity of these 35 genes between samples. Similar to the results for the overall resistome composition, beta diversity was highest between countries and was higher between different catchments in the same city than within the same catchment (Supplementary Figure 5). Taken together, these results suggesting that while individual manhole environments play a role in selecting for abundances of genes, specific gene variants themselves likely have few barriers to distribution at smaller geographical scales.

Chapter 4

Conclusion

Urban sewage systems likely play a major role in the dissemination of antibiotic resistance from humans to the environment. In this study, we evaluated the presence of antibiotic resistance genes across wastewater environments, assessing their relevance and risk to human health and identifying patterns across geography. We sampled upstream residential sewage, an understudied part of the sewage system that is close to the human waste sources, across multiple countries to highlight challenges in the evaluation of antibiotic resistance in sewage. We found that a substantial proportion of the antibiotic resistance genes commonly found in sewage are not present in human feces and do not pose an immediate threat to human health, suggesting that evaluating an environment's risk to human health should not rely solely on measuring the presence of antibiotic resistance genes in that environment. While WWTPs and hospital effluents are traditionally viewed as antibiotic resistance hotspots, we showed that residential sewage may also be a major source of antibiotic resistance, containing higher abundances of ARGs than at WWTP samples and at times reaching comparable levels of risk as hospital effluent. Although some classes of antibiotics exhibited similar patterns between consumption and resistance across countries, others did not, highlighting that the relationship between environmental antibiotic resistance and population-level antibiotic consumption is complex. Lastly, we demonstrated that despite some differences due to environmental selection between manholes, gene flow readily occurs within a city but larger geographical distances serve as a barrier to gene flow. Overall, our study highlights the challenges in evaluating the public health relevance of antibiotic resistance genes found in wastewater environments and provide insights on how to address these challenges. By targeting specific genes (e.g. human-associated genes, clinically-relevant genes) and by evaluating the diversity of genes at different scales (e.g. nucleotide,

composition), we are able to better understand the relationship between antibiotic resistance found in the environment and human health. As wastewater becomes a focal point in the efforts to monitor population-level antibiotic resistance and mitigate its spread to the environment, these challenges and insights should be considered in order to identify and evaluate suitable interventions.

Appendix A

Figures

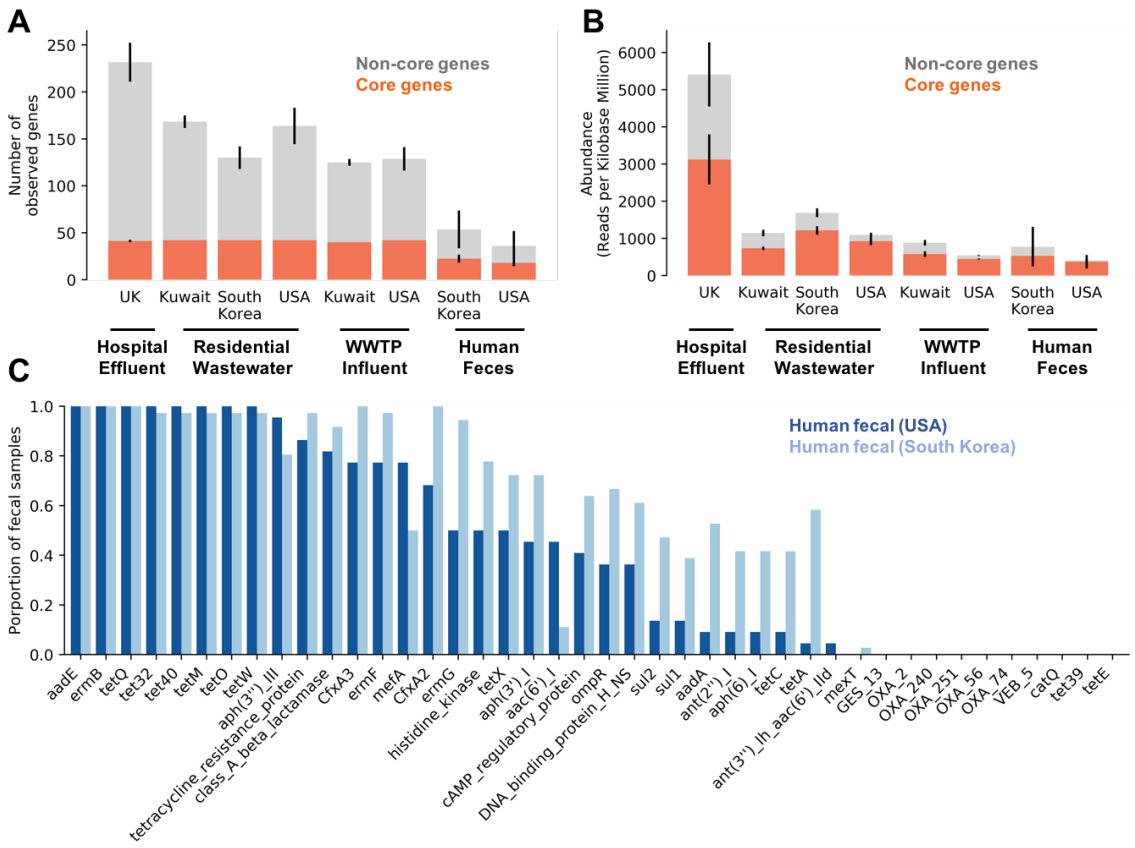

Figure A-1. Presence and abundance of antibiotic resistance genes across environments

(A) Number of antibiotic resistance genes observed per environment. Orange coloring represent the core residential sewage genes (n = 42) while grey colorings represent the remaining non-core genes. Error bars for each coloring (orange and grey) represent the standard deviations in the number of genes present for that category.

(B) Abundance of antibiotic resistance genes observed per environment in reads per kilobase millions. Coloring and error bars represent the same as in (A).

(C) Proportion of the human fecal samples in South Korea (n = 36) and the U.S. (n = 22) with each core residential sewage gene.

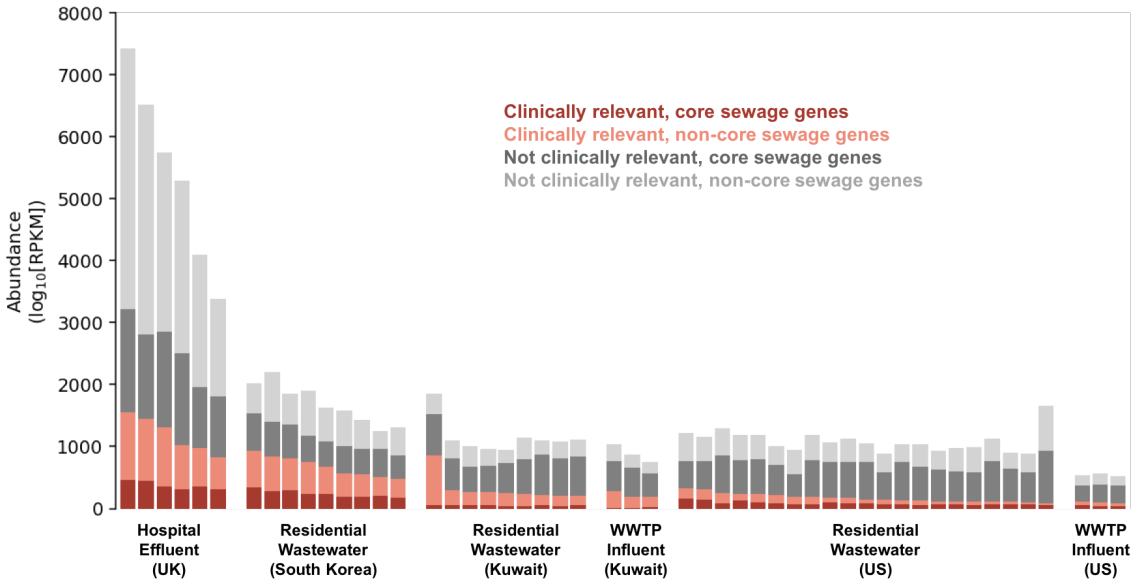

Figure A-2. Risk prioritization of antibiotic resistance genes

Abundance of antibiotic resistance gene in each wastewater environment, categorized by risk. Proportions in shades of red represent abundances of clinically-relevant antibiotic resistance genes while shades of grey represent non-clinically-relevant genes. Within each coloring (red and grey), genes in the core set of residential sewage genes are shaded darker (e.g. dark red = genes which are clinically relevant and core; dark grey = genes which are not clinically relevant and core). Samples within each environment are ranked in descending order based on total abundance of clinically relevant genes. Genes are defined as clinically-relevant if they have previously been observed on plasmids and have a diverse range of hosts. RPKM = Reads per Kilobase Million.

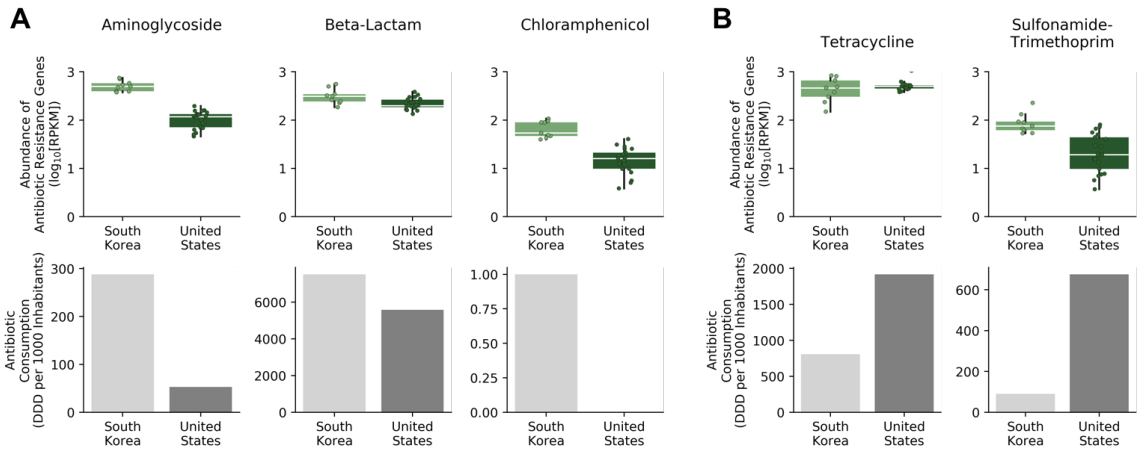

Figure 3. Antibiotic resistance and antibiotic consumption across geography by drug class

Patterns of antibiotic resistance in residential sewage and human feces (top panels) and patterns antibiotic consumption (bottom panels) between South Korea and the United States. Antibiotic resistance and antibiotic consumption pattern show concordance for some classes of antibiotics (A) but not for others (B). Antibiotic resistance is represent in log abundances of reads per kilobase million while antibiotic consumption is represented in defined daily dose (DDD) per 1000 inhabitants. Each point represents one sample. Data for antibiotic consumption (2015) was obtained from the IQVIA MIDAS database (https://resistancemap.cddep.org/AntibioticUse.php).

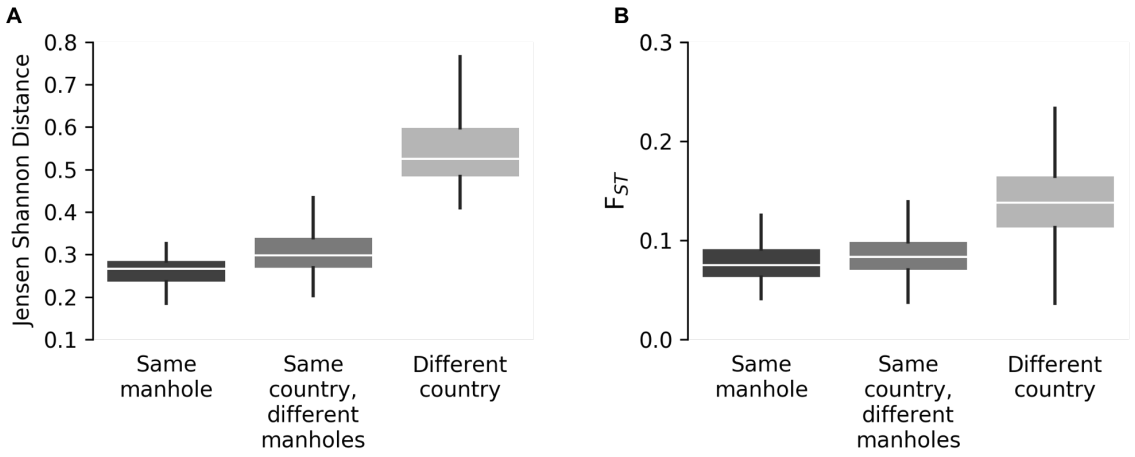

Figure 4. Beta diversity and nucleotide diversity across different geographic scales

(A) Beta diversity in Jensen Shannon Distance of resistomes between samples at different scales of geographical comparisons. Error bars represent standard deviations while center bars represent medians.

(B) Nucleotide diversity of resistomes between samples at different scales of geographical comparisons. Nucleotide diversity between each pair of samples was measured in terms of the average F_{ST} of genes with sufficient coverage and polymorphism (n = 35; Material and Methods). Error bars represent standard deviations while center bars represent medians.

Appendix B

Supplementary Figures

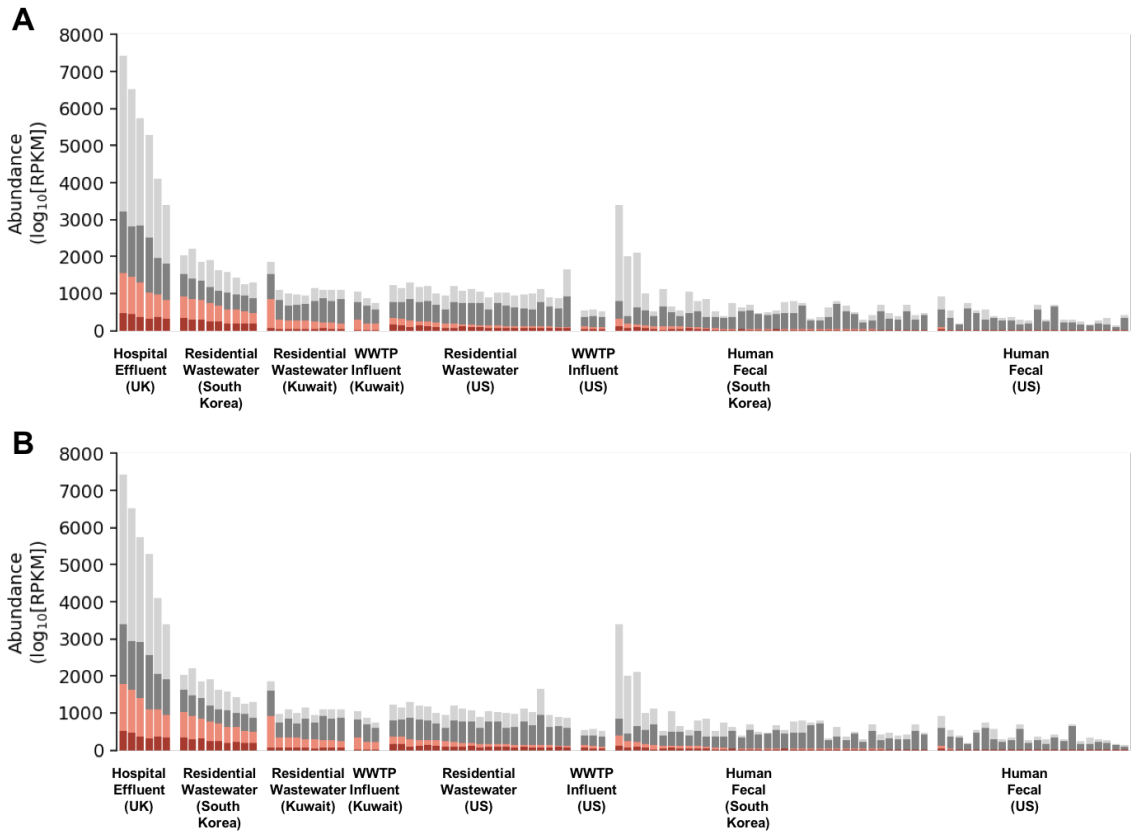

Figure B-1. Risk prioritization of antibiotic resistance genes

(A) Abundance of antibiotic resistance gene in each wastewater environment and human environment, categorized by risk. Similar to Figure 2, proportions in shades of red represent abundances of clinically-relevant antibiotic resistance genes while shades of grey represent non-clinically-relevant genes.

(B) Abundance of antibiotic resistance gene in each wastewater environment and human environment, categorized by risk. Similar to Figure 2, except proportions in shades of red represent abundances of mobile antibiotic resistance genes while shades of grey represent mobile genes.

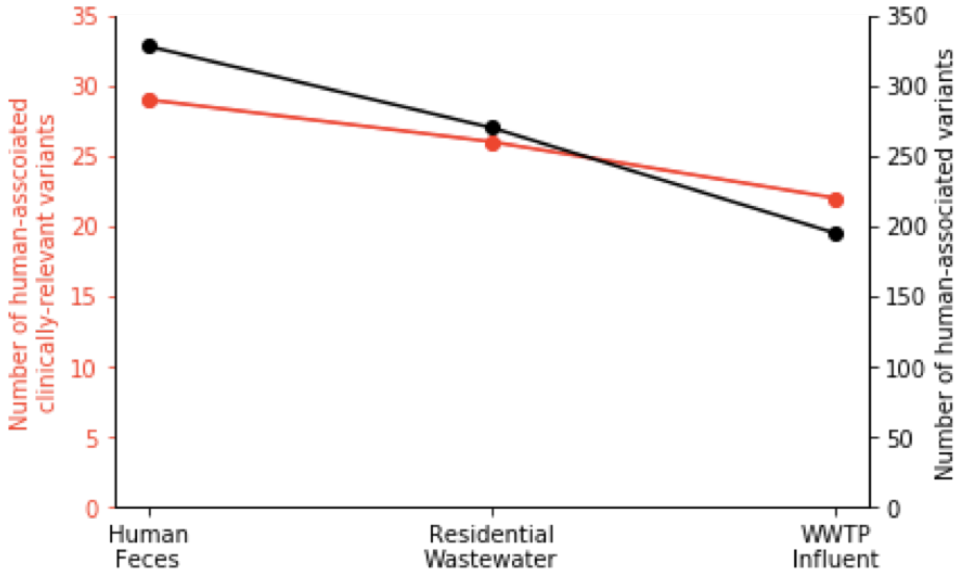

Figure B-2. Presence of human-associate antibiotic resistant gene variants by environment

The observed count of antibiotic resistance gene variants found in human fecal samples in sewage environments. Human-associated gene variants become less present further downstream.

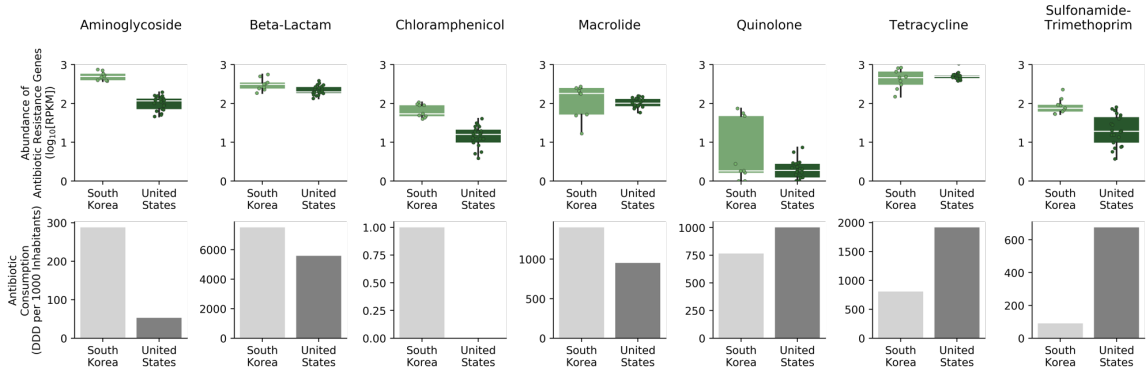

Figure B-3. Antibiotic resistance and antibiotic consumption across geography by drug class for all drug classes

Similar to Figure 3, except all classes of drug where antibiotic consumption and resistance data were available

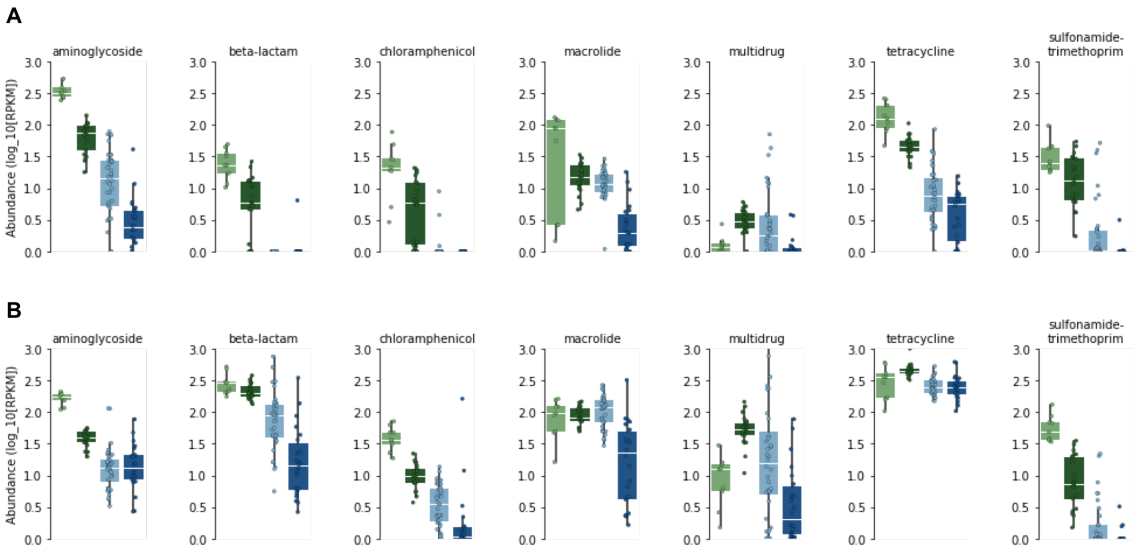

Figure B-4. Antibiotic resistance in the residential sewage sample and the human stool samples from U.S. and South Korea, grouped by drug class

Abundances of antibiotic resistance for each drug class, separated by clinically relevant genes (A) and non-clinically-relevant genes (B).

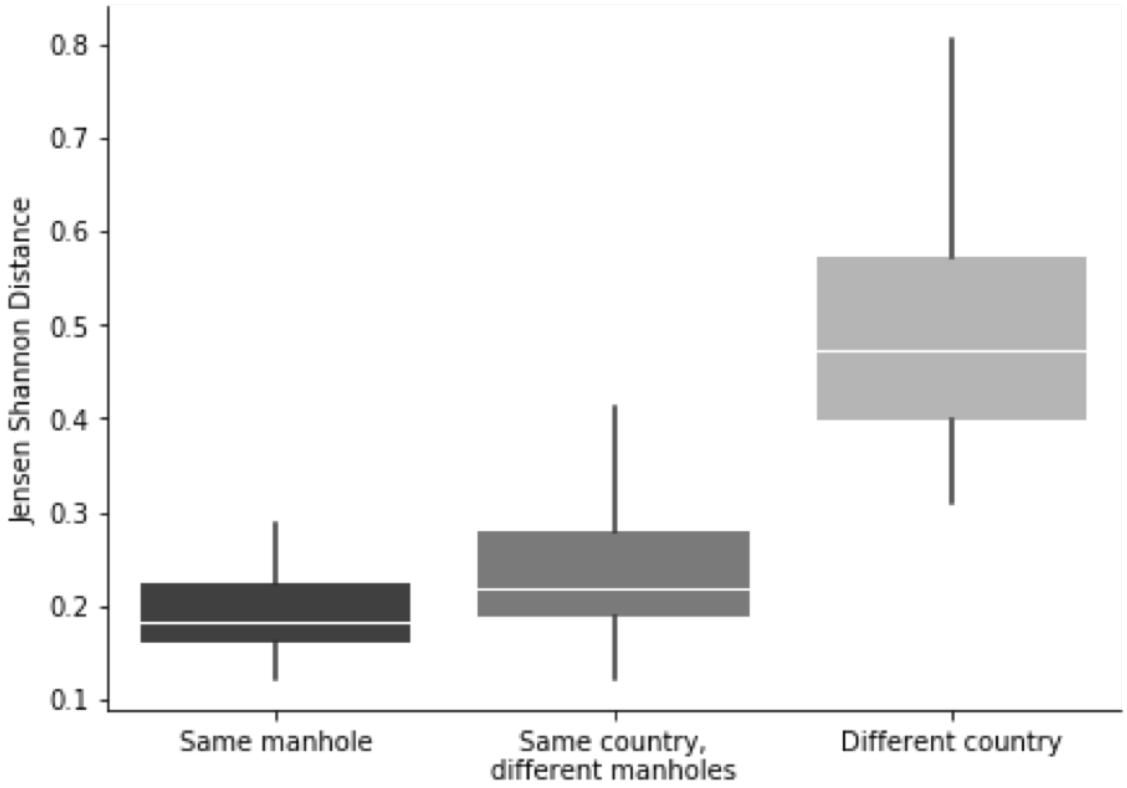

Figure B-5. Beta diversity of the 35 genes with F_{ST} measurements across different geographic scales

Beta diversity in Jensen Shannon Distance of resistomes that met the cutoff criteria for F_{ST} analysis. Comparisons were made between samples at different scales of geographical comparisons. Error bars represent standard deviations while center bars represent medians.

Bibliography

1. Zaman, S. B. *et al.* A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **9**, e1403 (2017).
2. The Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. *The Review on Antimicrobial Resistance* Available at: https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf. (Accessed: 11th January 2019)
3. *Antimicrobial resistance: global report on surveillance*. (World Health Organization, 2014).
4. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
5. Baquero, F., Martínez, J.-L. & Cantón, R. Antibiotics and antibiotic resistance in water environments. *Curr. Opin. Biotechnol.* **19**, 260–265 (2008).
6. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251–259 (2010).
7. Lee, J. H., Park, K. S., Jeon, J. H. & Lee, S. H. Antibiotic resistance in soil. *Lancet Infect. Dis.* **18**, 1306–1307 (2018).
8. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
9. Cantón, R. Antibiotic resistance genes from the environment: a perspective through newly identified antibiotic resistance mechanisms in the clinical setting. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **15 Suppl 1**, 20–25 (2009).
10. Martínez, J. L. Ecology and Evolution of Chromosomal Gene Transfer between Environmental Microorganisms and Pathogens. *Microbiol. Spectr.* **6**, (2018).
11. Pehrsson, E. C. *et al.* Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
12. GLASS | Global Antimicrobial Resistance Surveillance System (GLASS). *WHO* Available at: http://www.who.int/glass/en/. (Accessed: 11th January 2019)
13. National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS) | NARMS | CDC. (2018). Available at: https://www.cdc.gov/narms/index.html. (Accessed: 11th January 2019)
14. European Antimicrobial Resistance Surveillance Network (EARS-Net). *European Centre for Disease Prevention and Control* Available at: http://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net. (Accessed: 11th January 2019)

15. Jones, C., Patel, J. & Jinks, T. Statement on International Environmental AMR Forum. *Wellcome Trust* Available at: https://wellcome.ac.uk/sites/default/files/international-environmental-amr-forum-joint-statement.pdf. (Accessed: 11th January 2019)
16. Gaze, W. & Depledge, M. Antimicrobial Resistance: Investigating the Environmental Dimension. *United Nations Environment Programme* Available at: http://wedocs.unep.org/bitstream/handle/20.500.11822/22263/Frontiers_2017_CH1_EN.pdf?sequence=1&isAllowed=y. (Accessed: 11th January 2019)
17. Wellcome Trust, Center for Disease Control & UK Science & Innovation Network. Initiatives for Addressing Antimicrobial Resistance in the Environment Current: Situation and Challenges. *Wellcome Trust* Available at: https://wellcome.ac.uk/sites/default/files/antimicrobial-resistance-environment-report.pdf. (Accessed: 7th January 2019)
18. Larsson, D. G. J. *et al.* Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environ. Int.* **117**, 132–138 (2018).
19. Rowe, W. P. M. *et al.* Overexpression of antibiotic resistance genes in hospital effluents over time. *J. Antimicrob. Chemother.* **72**, 1617–1623 (2017).
20. Ng, C. *et al.* Characterization of Metagenomes in Urban Aquatic Compartments Reveals High Prevalence of Clinically Relevant Antibiotic Resistance Genes in Wastewaters. *Front. Microbiol.* **8**, (2017).
21. Martínez, J. L., Coque, T. M. & Baquero, F. What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.* **13**, 116–123 (2015).
22. Bengtsson-Palme, J. & Larsson, D. G. J. Antibiotic resistance genes in the environment: prioritizing risks. *Nat. Rev. Microbiol.* **13**, 396 (2015).
23. Huijbers, P. M. C. *et al.* Role of the Environment in the Transmission of Antimicrobial Resistance to Humans: A Review. *Environ. Sci. Technol.* **49**, 11993–12004 (2015).
24. Topp, E., Larsson, D. G. J., Miller, D. N., Van den Eede, C. & Virta, M. P. J. Antimicrobial resistance and the environment: assessment of advances, gaps and recommendations for agriculture, aquaculture and pharmaceutical manufacturing. *FEMS Microbiol. Ecol.* **94**, (2018).
25. Munir, M., Wong, K. & Xagoraraki, I. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. *Water Res.* **45**, 681–693 (2011).
26. Marti, E., Jofre, J. & Balcazar, J. L. Prevalence of Antibiotic Resistance Genes and Bacterial Community Composition in a River Influenced by a Wastewater Treatment Plant. *PLOS ONE* **8**, e78906 (2013).
27. Rodriguez-Mozaz, S. *et al.* Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water Res.* **69**, 234–242 (2015).

28. Rizzo, L. *et al.* Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Sci. Total Environ.* **447**, 345–360 (2013).
29. Fahrenfeld, N. & J. Bisceglia, K. Emerging investigators series: sewer surveillance for monitoring antibiotic use and prevalence of antibiotic resistance: urban sewer epidemiology. *Environ. Sci. Water Res. Technol.* **2**, 788–799 (2016).
30. Global Sewage Surveillance Project - Compare Europe. *https://www.compare-europe.eu* Available at: https://www.compare-europe.eu/Library/Global-Sewage-Surveillance-Project. (Accessed: 11th January 2019)
31. Harris, S., Morris, C., Morris, D., Cormican, M. & Cummins, E. Antimicrobial resistant Escherichia coli in the municipal wastewater system: Effect of hospital effluent and environmental fate. *Sci. Total Environ.* **468**–**469**, 1078–1085 (2014).
32. Varela, A. R., André, S., Nunes, O. C. & Manaia, C. M. Insights into the relationship between antimicrobial residues and bacterial populations in a hospital-urban wastewater treatment plant system. *Water Res.* **54**, 327–336 (2014).
33. Ju, F. *et al.* Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* 1 (2018). doi:10.1038/s41396-018-0277-8
34. Buelow, E. *et al.* Limited influence of hospital wastewater on the microbiome and resistome of wastewater in a community sewerage system. *FEMS Microbiol. Ecol.* **94**, (2018).
35. Kümmerer, K. Resistance in the environment. *J. Antimicrob. Chemother.* **54**, 311–320 (2004).
36. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
38. Lim, M. Y. *et al.* Stability of Gut Enterotypes in Korean Monozygotic Twins and Their Association with Biomarkers and Diet. *Sci. Rep.* **4**, 7348 (2014).
39. Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **6**, 6505 (2015).
40. Kaminski, J. *et al.* High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *PLOS Comput. Biol.* **11**, e1004557 (2015).
41. Zhang, A. ni, Li, L.-G., Tiedje, J., Alm, E. J. & Zhang, T. Whom to Fight: Top Risk Antibiotic Resistances for Global Action. *Prep.*
42. Yang, Y. *et al.* ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. *Bioinforma. Oxf. Engl.* **32**, 2346–2351 (2016).
43. Costea, P. I. *et al.* metaSNV: A tool for metagenomic strain level analysis. *PLOS ONE* **12**, e0182392 (2017).

44. The Comprehensive Antibiotic Resistance Database. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3697360/. (Accessed: 24th January 2019)
45. Klein, E. Y. *et al.* Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3463–E3470 (2018).
46. Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* **23**, 1163–1169 (2013).
47. Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Health* **2**, e398–e405 (2018).
48. Hoelzer, K. *et al.* Antimicrobial drug use in food-producing animals and associated human health risks: what, and how strong, is the evidence? *BMC Vet. Res.* **13**, (2017).
49. Berendonk, T. U. *et al.* Tackling antibiotic resistance: the environmental framework. *Nat. Rev. Microbiol.* **13**, 310–317 (2015).
50. Ort, C., Schaffner, C., Giger, W. & Gujer, W. Modeling stochastic load variations in sewer systems. *Water Sci. Technol.* **52**, 113–122 (2005).
51. Ort, C., Lawrence, M. G., Reungoat, J. & Mueller, J. F. Sampling for PPCPs in Wastewater Systems: Comparison of Different Sampling Modes and Optimization Strategies. *Environ. Sci. Technol.* **44**, 6289–6296 (2010).
52. Huisman, J. L. *et al.* Quantification of oxygen fluxes in a long gravity sewer. *Water Res.* **38**, 1237–1247 (2004).
53. Vollertsen, J. & Hvitved-Jacobsen, T. Stoichiometric and kinetic model parameters for microbial transformations of suspended solids in combined sewer systems. *Water Res.* **33**, 3127–3141 (1999).
54. Wright, S. The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).