Revealing the Complex Dynamics of Antimicrobial Resistance Genes Immigration from Sewers to Activated Sludge

Claire Gibson

Department of Civil Engineering and Applied Mechanics

McGill University, Montreal, Canada

February 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

TABLE OF CONTENTS

REVEALING THE COMPLEX DYNAMICS OF ANTIMICROBIAL RESISTANCE GENE IMMIGRATION FROM SEWERS TO ACTIVATED SLUDGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

AMR	Antimicrobial Resistance
ANOSIM	Analysis of Similarity
ARB	Antimicrobial Resistant Bacteria
ARG	Antimicrobial Resistance Genes
AS	Activated Sludge
ASV	Amplicon Sequencing Variants
AS-WWTP	Activated Sludge Wastewater Treatment Plant
CARD	Comprehensive Antibiotic Resistance Database
COD	Chemical Oxygen Demand
DbMEM	Distance based eigen vector maps
ddPCR	Droplet Digital PCR
DNA	Deoxyribonucleic Acid
GI	Genomic Islands
HGT	Horizontal Gene Transfer
HRT	Hydraulic Retention Time
MGE	Mobile Genetic Element
ML	Mixed liquor
MRG	Metal Resistance Genes
NA	Not Applicable
NCBI	National Center for Biotechnology Information
ND	Not detected
OTU	Operational Taxonomic Units
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction

RDA	Redundancy Analysis
RGI	Resistance Gene Identifier
RNA	Ribonucleic Acid
SRT	Solid Retention Time
TSS	Total Suspended Solids
VSS	Volatile Suspended Solids
WGS	Whole Genome Shotgun
WWTP	Wastewater treatment plant

Symbols

b_i	Decay rate of i-th taxon
$f_{16S,Inf,i}$	Proportion of the i-th taxon in the influent
$f_{16S,ML,i}$	Proportion of the i-th taxon in mixed liquor
$f_{OHO,Capt}$	Capture of the influent biomass by the mixed liquor solids
m_i	Immigration level of i-th taxon
$S_{i,inf}$	Concentration of substrates consumed by i-th taxon
μ_i	Specific growth rate of i-th taxon
$\mu_{net,i}$	Net growth rate of the i-th raxon
$X_{bio,i,inf}$	Biomass concentration of i-th taxon in influent
$X_{bio,i,ML}$	Biomass concentration of i-th taxon in mixed liquor
$X_{Tot,Inf}$	Total solids of the influent
$X_{tot,ML}$	Total solids of the mixed liquor
$Y_{i,inf}$	Biomass yield of i-th taxon
$\gamma_{DNA,Inf}$	DNA Extraction Yield for the influent
$\gamma_{DNA,ML}$	DNA extraction yield for the mixed liquor
θ	Hydraulic Retention Time
θ_x	Solid Retention Time

ABSTRACT

Antimicrobial resistance (AMR) is widely recognised as one of the greatest threats to public health worldwide. The United Nations Environment Assembly (UNEA-3) have recognised the importance of the environment in the development, spread and transmission of AMR to humans and animals (United Nations Environment Programme, 2022). Such recognition calls for wider surveillance of antimicrobial resistance genes (ARGs) in wastewater and other environmental reservoirs.

To monitor ARGs in environmental reservoirs, we developed a novel, low cost and sensitive multiplexed targeted amplicon sequencing approach for the detection of ARG sequence variants. The identification of variants provides information on the possible transfer of specific ARGs between reservoirs. It can also be interfaced with databases to explore the human and animal health risks associated with the population of ARGs detected. Using this amplicon sequencing approach, a diverse microbial community rich in ARG sequence variants was detected in influent wastewater samples across Quebec.

To understand the interactions between the diverse populations of ARG variants in the wastewater and the downstream microbial community of the treatment facility, laboratory-scale reactor experiments were used. Microbial community composition has increasingly emerged as a key determinant of ARG content. However, in activated sludge wastewater treatment plants (AS-WWTPs), a comprehensive understanding of the microbial community assembly process remains elusive. An important part of this process is the immigration dynamics between the influent and activated sludge. While the influent wastewater contains a plethora of ARGs, the persistence of a given ARG depends initially on the immigration success of its carrying population.

To investigate the fate of bacterial populations across the immigration interface, controlled manipulative laboratory-scale experiments were used. In full-scale wastewater systems community composition is dependent on the wastewater substrate composition. To account for this, the immigration dynamics of heterotrophs were analysed by decoupling them from the substrate landscape using a synthetic wastewater. The systematic control of factors influencing immigration outcomes between sewers and AS-WWTP allowed fundamental mechanisms to be revealed. Using this approach, immigrating influent populations were observed to contribute up to 25 % of the 16S rRNA gene amplicon sequencing reads in the activated sludge and caused an increase in the abundance of up to 70 % (11/15) of the ARGs quantified by digital droplet PCR. Further exploration of operational parameters influencing immigration demonstrated a positive correlation between relative reads introduced and solid retention time. The monitoring of ARG sequence variants at the immigration interface revealed various immigration patterns such as (i) suppression of the indigenous mixed liquor variant by the immigrant, or conversely (ii) complete immigration failure of the influent variant. These immigration profiles are reported for the first time here and highlight the crucial information that can be gained using this novel multiplex amplicon sequencing techniques.

The highly controlled experimental protocol developed herein will inform the design of future studies on microbial community dynamics in wastewater treatment plants. This information will be valuable in the optimization of processes to reduce antimicrobial resistance. In the future, multiplexed amplicon sequencing approaches should be used as an effective and inexpensive tool for routine ARG source tracking and risk assessment of environmental samples.

RÉSUMÉ

L’antibiorésistance est largement reconnue comme l'une des plus grandes menaces pour la santé publique dans le monde. L'Assemblée des Nations unies pour l'environnement (UNEA-3) a reconnu l'importance de l'environnement dans le développement, la propagation et la transmission d’antibiorésistances aux humains et aux animaux (Programme des Nations unies pour l'environnement, 2022). Une telle reconnaissance appelle à une surveillance plus large des gènes de résistance aux antimicrobiens (ARG de l’acronyme anglais) dans les eaux usées et autres réservoirs environnementaux.

Pour surveiller les ARG dans les réservoirs environnementaux, nous avons développé une nouvelle approche multiplexe de séquençage ciblé sur des amplicons. Cette approche pour le suivi des variants en séquence d’ARG est sensible et peu chère. L'identification des variantes fournit des informations spécifiques sur le transfert possible d'ARG entre les réservoirs. Elle peut également être interfacée avec des bases de données pour explorer les risques pour la santé humaine et animale associés à la population d'ARG détectées. En utilisant cette approche de séquençage d'amplicons, une communauté microbienne diversifiée riche en variant de séquence ARG a été détectée dans des échantillons d'eaux usées aux affluents d’usines de traitement à travers le Québec.

Pour comprendre les interactions entre les différentes populations de variants d'ARG dans les eaux usées et la communauté microbienne en aval dans les installations de traitement, des expériences en réacteurs à l'échelle du laboratoire ont été utilisées. La composition de la communauté microbienne apparaît dans la littérature de plus en plus comme un déterminant clé de la teneur en ARG. Cependant, dans les stations d'épuration par boues activées, une compréhension globale du

processus d'assemblage de la communauté microbienne reste insaisissable. Une partie importante de ce processus est la dynamique d'immigration entre l'affluent et les boues activées. Alors que les eaux usées entrantes contiennent une pléthore d'ARG, la persistance d'un ARG donné dépend initialement du succès du processus l'immigration de sa population porteuse.

Pour étudier le devenir des populations bactériennes après l'immigration, des expériences contrôlées de manipulation des conditions d'opération des réacteurs à l'échelle du laboratoire ont été utilisées. Dans les systèmes d'eaux usées à grande échelle, la composition de la communauté dépend de la composition du substrat des eaux usées. Pour en tenir compte, la dynamique d'immigration des hétérotrophes a été analysée en les dissociant du paysage de substrats à l'aide d'une eau usée synthétique. Le contrôle systématique des facteurs influençant les résultats de l'immigration entre les égouts et les réacteurs par boues activées a permis de mettre en évidence des mécanismes fondamentaux. En outre, il a été observé que les populations bactériennes de l'affluent ont contribué par leur immigration jusqu'à 25 % des séquences d'amplicons d'ARNr 16S dans les boues activées et ont provoqué une augmentation de l'abondance jusqu'à 70 % (11/15) des ARG quantifiés par qPCR. Une exploration plus approfondie des paramètres opérationnels influençant l'immigration a démontré une corrélation positive entre l'abondance relative des séquences introduites par l'immigration et le temps de rétention solide. Le suivi des variants en séquence d'ARG à l'interface d'immigration a révélé divers schémas d'immigration tels que (i) la suppression des variants indigènes de la liqueur mixte par celui immigrant, ou inversement (ii) l'échec complet de l'immigration du variant de l'affluent. Ces profils d'immigration sont rapportés pour la première fois ici et mettent en évidence les informations cruciales qui pourront être obtenues à l'aide de notre nouvelle technique de séquençage multiplexe d'amplicons d'ARG.

Le protocole expérimental hautement contrôlé utilisant des réacteurs pour lesquels tous les aspects d'opération (incluant les compositions microbiennes et biochimiques de l'affluent) que nous avons développé ici éclairera la conception d'études futures sur la dynamique des communautés microbiennes dans les stations d'épuration. Il contribuera ainsi à l'optimisation de ces processus pour réduire la résistance aux antimicrobiens. À l'avenir, notre approche de séquençage multiplexe d'amplicons d'ARG devrait être utilisées comme outil efficace et peu coûteux pour le suivi de routine des sources d'ARG et l'évaluation des risques des échantillons environnementaux.

ACKNOWLEDGMENTS

Throughout this PhD journey I have grown as both a researcher and person, this would not have been possible without the people around me. Firstly, I would like to express my sincere gratitude to my supervisor, Professor Dominic Frigon, for his guidance and encouragement throughout this process. I would also like to thank my co-supervisor, Dr. Bing Guo for her support and valuable contributions towards this research.

I would like to extend my gratitude to the technical and administrative staff in the Department of Civil Engineering. I thank the Benedek lab managers, Daina Brumelis and Dr. Pinar Ozcer for their valuable assistance in the lab. I sincerely appreciate the technical support received from Jorge Sayat, Dr. William Cook, John Bartczak and William Boyd Dumais, and the administrative help from Sun Chee Wong and Sandra Shewchuk-Boyd. I would also like to acknowledge the financial support received from the Faculty of Engineering and the Natural Sciences and Engineering Research Council of Canada (NSERC).

I wish to thank the past and present members of the Microbial Engineering Laboratory for their encouragement and friendship. A special thanks goes to Natalia Klimova and Dr. Susanne Kraemer for their invaluable help over the last two years, and to Chenxiao Liu for his continued support throughout. I am deeply appreciative of all the friends I have made in the Benedek Laboratory, who made office 569 a great working environment.

Lastly, my heartfelt thanks goes to my family. I will be forever grateful to my partner, Gabriel, for his support during the highs and lows of this PhD journey, and for keeping me laughing throughout. Finally, I thank my parents for their love and reassurance, without them, this would not have been possible.

CONTRIBUTIONS TO KNOWELDGE

The contributions of this thesis include the development of innovative methods to study both immigration in the context of microbial community assembly, and the specific fate of ARGs in the environment. The specific contributions of this work are as follows;

1. Developed a low cost and sensitive multiplexed amplicon sequencing method for the detection of ARG sequence variants in environmental samples. This method can be utilised in the future for the routine monitoring of ARG sequence variants. The data obtained can be used in both ARG risk assessment and source tracking.
    a. Of the 60 ARG and metal resistance genes detected in Quebec sewers, 16 were found to have more than one sequence variant present.
    b. Unique ARG sequence variants were identified in different sample types (influent and activated sludge) and in different locations, demonstrating the value of this approach in environmental surveillance.
2. Developed a reproducible reactor system to study the impact of specific variables on the microbial community of the activated sludge. This reactor protocol can be utilised in the future to test the impact of other wastewater variables and optimise WWTP design.
    a. It was shown that when fed with a synthetic wastewater (Syntho), reproducible microbial communities were formed irrespective of the mixed liquor inoculum used. This validated the reactor setup for future, comparable studies which use different sources of mixed liquor inoculum.
    b. For this experimental system, a method was established to intrinsically classify populations as resident or immigrants based on experimental observations, which goes beyond the simple comparison of influent and downstream mixed liquor taxa used in the literature.

3. Quantified the impact of immigration on the microbial community and ARGs in the activated sludge.
    a. Immigrant genera were found to typically exhibit negative net growth rates in the activated sludge
    b. Immigration was shown to cause significant increases in the abundance of ARGs.
    c. Despite an increase in the abundance of ARGs with immigration, it was demonstrated that this was not always due to the addition of an immigrating variant
4. Demonstrated that immigration was impacted by solids retention time (SRT)
    a. At lower SRTs, the activated sludge resident populations were found to be more similar that of the influent
    b. By intrinsically classifying the immigrating populations, it was noted that immigration had a greater impact on the activated sludge community at higher SRTs
    c. It was observed that the abundance of ARGs was impacted by immigration, which varied with SRT. By comparing the intrinsically defined resident and immigrant populations at each SRT and the ARG relative abundance, links could be drawn between the populations and ARG dynamics.

PREFACE

In accordance with the 'Guidelines for Thesis Preparation', this thesis is present in a Manuscript based format. In Chapter 1 and 2, a general introduction and literature review are provided. Chapter 3 to 6 include 4 research articles which will be submitted for publication in the journals *Environmental Science and Technology, Applied and Environmental Microbiology, Frontiers in Microbiology* and *Nature Communications*. In the final chapter, a general discussion of the results is provided and suggestions for future work. The author of this thesis is the first author in all manuscripts included. In all manuscripts, the co-authors have reviewed and approve of the following statements.

Chapter 3

Claire Gibson, Susanne A. Kraemer, Natalia Klimova, Laura Vanderweyen, Nouha Klai, Emmanuel Diaz Mendoza, Bing Guo, David Walsh & Dominic Frigon (2023). Multiplexed Amplicon Sequencing Reveals High Sequence Diversity of Antibiotic Resistance Genes in Québec Sewers. To submit to *Environmental Science and Technology.*

Authors Contributions:

Claire Gibson designed the study, prepared samples for sequencing, analysed the data and prepared the manuscript.

Susanne A. Kraemer developed the bioinformatics pipeline for analysis of amplicon sequencing data, processed the sequencing data and revised the manuscript.

Natalia Klimova optimised and analysed samples using inverse PCR method.

Laura Vanderweyen organised the sampling campaign, collected influent wastewater samples and mapped the sampling sites.

Nouha Klai helped to plan the sampling campaign and collect influent wastewater samples.

Emmanuel Diaz Mendoza helped with sample analysis.

Bing Guo helped with data analysis and revised manuscript.

David Walsh helped with the development of the bioinformatics pipeline and supervised the research.

Dominic Frigon obtained funding, supervised the research and revised the manuscript.

Chapter 4

Claire Gibson, Shameem Jauffur, Bing Guo & Dominic Frigon (2023). Activated Sludge Microbial Community Assembly: The Role of Influent Microbial Community Immigration. Submitted to *Applied and Environmental Microbiology.*

Authors' contributions:

Claire Gibson designed the study, operated two sets of reactors, sequenced samples, analysed data and prepared the manuscript.

Shameem Jauffur designed the study and operated one set of reactors.

Bing Guo helped with data analysis and reviewed the manuscript.

Dominic Frigon obtained funding, supervised the research and revised the manuscript.

Chapter 5

Claire Gibson, Susanne A. Kraemer, Natalia Klimova, Bing Guo & Dominic Frigon (2023). Revealing the Complex Dynamics of Antimicrobial Resistance Gene Immigration from Sewers to Activated Sludge. Accepted by *Frontiers in Microbiology.*

Authors' contributions:

Claire Gibson designed the study, conducted digital droplet PCR, developed multiplexed amplicon sequencing approach, analysed the data and prepared the manuscript.

Susanne A. Kraemer developed the multiplexed amplicon sequencing approach, developed the bioinformatics pipeline, processed the sequencing data and reviewed the manuscript.

Natalia Klimova developed the multiplexed amplicon sequencing approach, prepared multiplexed amplicon sequencing samples for sequencing.

Dominic Frigon obtained funding, supervised the research and revised the manuscript.

Chapter 6

Claire Gibson, Bing Guo & Dominic Frigon (2023). Solid Retention Time Impacts Immigration and Antimicrobial Resistance in Wastewater Treatment Plants. To be submitted to *Nature Communications*

Authors' contributions:

Claire Gibson designed the study, operated the reactors, analysed the samples, analysed the data and prepared the manuscript.

Bing Guo helped with data analysis, reviewed the manuscript.

Dominic Frigon obtained funding, supervised the research and revised the manuscript.

CHAPTER 1

Introduction and Research Objectives

1.1 Introduction

1.1.1 Antimicrobial Resistance in the Environment

Antimicrobial resistance (AMR) is considered to be one of the greatest threats to global health, food security and development today (1). It was estimated that in 2019 alone, there were 4.95 million deaths associated with AMR, and without action this number is predicted to rise to 10 million by 2050 (2). In the past, efforts to monitor and control the spread of AMR often focused on areas with high antimicrobial use such as clinical and agricultural settings. However, environmental reservoirs, such as wastewater treatment plants (WWTPs), have been demonstrated over recent years to play a critical role in the dissemination of AMR (3). Antimicrobial resistant genes (ARG) and bacteria (ARB) are consistently identified in wastewater effluent and waste biosolids produced during the treatment process (4–6). The discharge of effluent wastewater into surface waters and the use of waste biosolids as fertilisers offer clear pathways for the spread of AMR to the environment. Research is needed to improve our knowledge of factors affecting the persistence and dissemination of AMR within these environments with the aim of prolonging the efficacy of antimicrobials and protecting public health.

Sewers are the first urban structure in the pathway of AMR dissemination from human excreta into the environment (Figure 1.1). Influent wastewater has been found to contain an abundance of ARGs and ARBs, and sewers act as a direct passage of these resistance determinants into multiple environmental compartments. In Quebec, many of the sewer systems are combined and carry both municipal wastewater and storm water. Thus, heavy rainfall and seasonal snowmelt can result in combined sewer overflow (CSO) events, whereby untreated wastewater containing ARGs and ARB is discharged directly into the environment.

Little is known about the factors impacting the abundance and diversity of ARGs in Quebec sewers. Comparisons of the 16S rRNA gene amplicon sequences observed in faecal discharges and in sewer wastewater estimate that around 8 % of the microorganisms originate from the human microbiota (7). The sewers also develop their own microbial community, which is likely impacted by environmental conditions (wastewater composition and temperature) and the state of the sewer including its age, architecture and operation. Among different bacterial species, the fitness cost of harbouring ARGs varies (8–11). Therefore, it could be hypothesised that changes in the microbial community composition of the influent wastewater impact the abundance and diversity of ARG in sewers.

In the dissemination of AMR, it is unclear whether wastewater treatment plants (WWTPs) act a point of control or if they enhance the spread of AMR to the environment. Nonetheless, wastewater treatment processes do not achieve complete removal of ARG and ARB before effluent discharge and elevated levels of AMR have been identified in surface waters and sediments downstream of the discharge point (4,6). During the activated sludge treatment process (Figure 1.1), a diverse microbial community is raised in the plant to become very active at aerobically degrading the pollutants contained in the incoming wastewater. Subsequently, solids are removed by settling before discharge of treated wastewater to surface water bodies. The composition and diversity of the activated sludge microbial community has the largest impact on the performance and stability of the system (12), and the bacterial community composition has increasingly emerged as a primary determinant of ARG content (13). In addition, the dense microbial community provides ideal conditions for the dissemination of ARGs associated with mobile genetic elements (MGEs) (together referred to as AMR determinants), which can transfer by horizontal gene transfer (HGT) between species (14–16). To prevent the dissemination of AMR, efforts must be made to minimise

the release of ARGs from sewers and WWTPs into the environment. Our knowledge of conditions favourable for the persistence of AMR in wastewater and sludge is lacking. However, evidence suggests that the microbial community within these environments may dictate the fate of mobile ARGs in sewers and WWTPs (13,17,18).

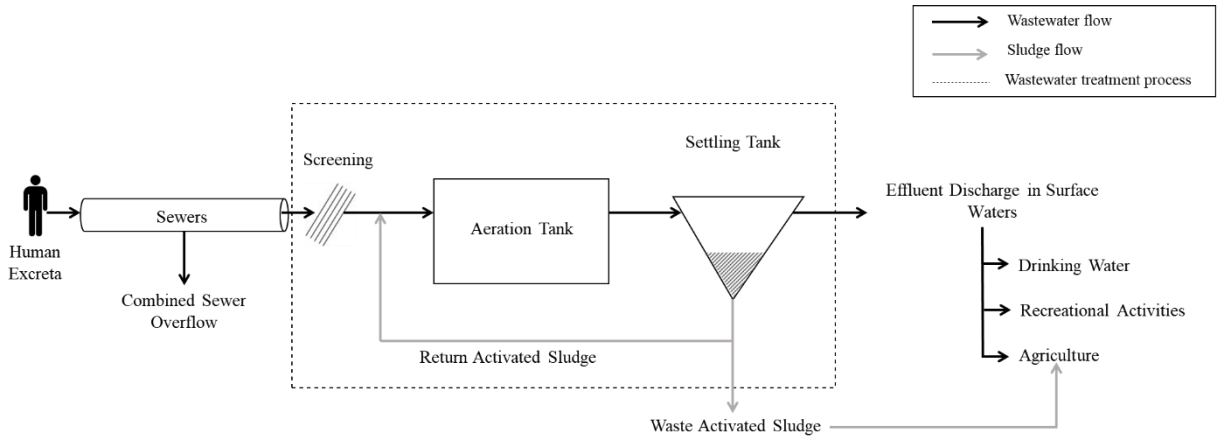


Figure 1.1- Wastewater collection and activated sludge biological treatment process. Wastewater enters the WWTP and is combined with activated sludge. The resulting mixed liquor is then aerated. Solids are removed in the settling tank before wastewater effluent is disinfected and discharged into surface waters. Residual solids are being increasingly applied to agricultural land.

1.1.2 The Role of Influent Immigration in Activated Sludge Microbial Community Assembly

The role of influent immigration on the microbial community of the activated sludge has drawn increasing attention over the past decade due to its importance in accurately modelling the wastewater treatment process and its role in maintaining specific populations (19). With sufficient developments in the field, influent immigration could play a vital role in WWTP bioaugmentation to achieve new treatment goals. However, the relative importance of influent immigration on the

activated sludge microbial community remains poorly understood. The fate of immigrating bacteria is influenced by a number of overarching mechanisms including niche processes whereby the microbial community composition is determined based upon the presence or absence of an organisms preferred environmental conditions or substrate (20), and neutral drift which states that organisms colonise an environment and randomly change within it through random birth, death, and dispersal events (21,22). The co-existence of niche and neutral mechanisms introduces additional complexity in determining how the microbial community will react under certain conditions. To further this, the variability in operational parameters and environmental conditions in WWTPs makes it challenging to conclusively determine the fate of immigrating taxa. Given the complex behaviour of microbial communities within sewers and WWTPs, a reproducible and highly controlled system is required to accurately investigate and quantify the impact of immigration on the activated sludge microbial community. By utilising such a system, the microbial community could be better defined into specific populations, and the intricate relationship between microbial community composition and ARG persistence could be better understood.

1.1.3 Approaches to Monitoring AMR in the Environment

To relate the dynamics of specific populations in the activated sludge to AMR, high resolution and sensitive approaches are required to monitor the changes in ARGs between the influent and activated sludge. Previous studies often relied upon techniques such as quantitative PCR to survey and quantify ARGs in the environment, but this technique provides little information on the sequence diversity of ARGs (23–25). Over recent years, studies have increasingly recognised the value of monitoring ARG sequence variants for source tracking approaches (26), with sequence variants of ARGs detected in different habitats including faeces, WWTP and pristine

environmental samples (27). However, current method to detect ARG sequence variants in the environment have various limitations. Shotgun metagenomics can be used to detect a large number of ARGs and mobile genetic elements, but is often limited by its poor detection limit (10^{-3} copy/genome equivalent). Whilst methods such as ResCap (an hybridisation-based nucleic acid fragment capture method) improve the detection limit of metagenomic sequencing (28), the increase in analytical cost makes it prohibitive for routine monitoring of environmental samples. To enable AMR data to be used for risk assessment and source tracking purposes, low cost and sensitive methods are required for the detection of ARG sequence variants in the environment. Furthermore, by utilising techniques for the detection of ARG sequence variants, the immigration dynamics between the influent and activated sludge could be studied in detail for the first tim

1.2 Research objectives

The overall aim of this thesis was to explore the relationship between microbial community assembly and antimicrobial resistance in activated sludge wastewater treatment plants. To achieve this, new approaches had to be developed which allowed the microbial community to be defined into specific resident or immigrant populations and ARG dynamics to be investigated following ARG variants across reservoirs. Based upon the relationship observed within numerous environmental settings, it was hypothesised that the microbial community composition would be a key determinant of antimicrobial resistance in the activated sludge. The specific objectives of this project were as follows:

Objective 1: Investigate the extent of ARG sequence diversity in Quebec sewers and the impact of sewer conditions on the distribution of variants

a) Develop a method to detect ARG sequence variants in environmental samples

b) Determine the impact of sewer conditions on ARG sequence diversity
c) Establish approaches to use ARG sequence variant information in antimicrobial risk assessment

Objective 2: Evaluate the impact of influent immigration on the microbial community assembly of the activated sludge

a) Develop a reproducible, controlled reactor system allowing the impact of specific operational parameters on the activated sludge process to be assessed
b) Investigate the impact of influent immigration on the activated sludge microbial community composition
c) Examine the reproducibility of immigration effects between different influent microbial communities and into different activated sludge communities

Objective 3: Study the impact of influent immigration on antimicrobial resistance gene dynamics in the activated sludge

a) Examine the impact of influent immigration on antimicrobial resistance gene abundance and diversity in the activated sludge
b) Study the dynamics of ARG sequence variants at the interface between the influent and activated sludge

Objective 4: Investigate the impact of operational solids retention time on influent immigration and antimicrobial resistance in wastewater treatment plants

a) Evaluate the impact of solids retention time on microbial immigration between the influent and activated sludge
b) Determine the impact of solids retention time on antimicrobial resistance in the activated sludge both with and without immigration
c) Develop specific microbial population definitions to allow community dynamics to be related to changes in antimicrobial resistance

1.3 Thesis Organisation

Following a literature review (Chapter 2), this thesis is divided into four main research chapters, each addressing a specific objective. A final chapter (Chapter 7) presenting general discussions and conclusions conclude the thesis.

Chapter 3: Multiplexed Amplicon Sequencing Reveals High Sequence Diversity of Antibiotic Resistance Genes in Québec Sewers (Objective 1)

In this chapter, a novel, low cost and highly sensitive method for the detection of ARG sequence variants is developed. To demonstrate the use of this tool, a survey of AMR in Quebec sewers was conducted. Data on ARG sequence variants in the samples was analysed in relation to the sewer characteristics and the microbial community composition. To demonstrate the applications of ARG sequence variant information the Comprehensive Antibiotic Resistance Gene Database was used to assess the risk of given variants based upon their mobility and occurrence among pathogens.

Chapter 4: Activated Sludge Microbial Community Assembly: The Role of Influent Microbial Community Immigration (Objective 2)

In this chapter, a reactor protocol was developed which allowed the impact of immigration to be studied with appropriate reproducibility and controls. Using this method, the specific impact of immigration on activated sludge microbial communities derived from three different full-scale plants was investigated independently of wastewater substrate composition. The reproducibility of the immigration effect was assessed by repeating the experiment with three different sources of influent microbial communities. Using this approach, influent immigration could be accurately quantified, and the net growth rate of specific populations was estimated.

Chapter 5: Revealing the Complex Dynamics of Antimicrobial Resistance Gene Immigration for Sewers to Activated Sludge (Objective 3)

In this chapter, the impact of influent immigration on AMR in the reactors operated in Chapter 4 was assessed. Using droplet digital PCR, the impact of immigration on the abundance of ARGs was evaluated. To study the dynamics of ARGs at the interface between the influent and activated sludge in greater resolution, ARG sequence variants were analysed using multiplexed amplicon sequencing. By carefully defining microbial populations, and ARG sequence variants within the reactors, parallels could be drawn between the microbial community composition and AMR in the activated sludge.

Chapter 6: Solids Retention Time Impacts Influent Immigration and Antimicrobial Resistance in Wastewater Treatment Plants (Objective 4)

In this chapter, the impact of operational solids retention time on immigration and AMR in the activated sludge was assessed. Based upon the protocol developed in Chapter 4, reactors were operated at five different solids retention times. The microbial community was analysed using 16S rRNA gene amplicon sequencing and categorised into specific resident or immigrant populations. ARGs were analysed by quantitative PCR array, and parallels were drawn between microbial population dynamics and ARG abundance. Finally, the limitations of current approaches to studying the impact of immigration are explored.

1.4 References

1. United Nations Environment Programme (UNEP). Environmental Dimensions of Antimicrobial Resistance: Summary for Policymakers. 2022. Available from: https://wedocs.unep.org/bitstream/handle/20.500.11822/38373/antimicrobial_R.pdf [Accessed on 2022 May 27].

2. Murray CJL, Shunji Ikuta K, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399:629–55.

3. Skandalis N, Maeusli M, Papafotis D, Miller S, Lee B, Theologidis I, et al. Environmental spread of antibiotic resistance. Antibiotics. 2021;10(6):640.

4. Quintela-Baluja M, Abouelnaga M, Romalde J, Su JQ, Yu Y, Gomez-Lopez M, et al. Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. Water Res. 2019;162:347–57.

5. Reichert G, Hilgert S, Alexander J, Rodrigues de Azevedo JC, Morck T, Fuchs S, et al. Determination of antibiotic resistance genes in a WWTP-impacted river in surface water, sediment, and biofilm: Influence of seasonality and water quality. Sci Total Environ. 2021;768:144526.

6. 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. 2011;45(2):681–93.

7. Guo B, Liu C, Gibson C, Frigon D. Wastewater microbial community structure and functional traits change over short timescales. Sci Total Environ. 2019;662:779–85.

8. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl. 2015;8:273–83.

9. Kassen R, Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet. 2006;38(4):484–8.

10. Hurdle JG, O'Neill AJ, Chopra I. The isoleucyl-tRNA synthetase mutation V588F conferring mupirocin resistance in glycopeptide-intermediate Staphylococcus aureus is not associated with a significant fitness burden. J Antimicrob Chemother. 2004;53(1):102–4.

11. Yokoyama M, Stevens E, Laabei M, Bacon L, Heesom K, Bayliss S, et al. Epistasis analysis uncovers hidden antibiotic resistance-associated fitness costs hampering the evolution of MRSA. Genome Biol. 2018;19:94.

12. Miura Y, Hiraiwa MN, Ito T, Itonaga T, Watanabe Y, Okabe S. Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance. Water Res. 2007;41:627–37.

13. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nat Lett. 2014;509:612–6.

14. Wang S, Li S, Du D, Abass OK, Nasir MS, Yan W. Stimulants and donors promote megaplasmid pND6-2 horizontal gene transfer in activated sludge. J Environ Sci. 2023;126:742–53.

15. Zhang T, Zhang XX, Ye L. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One. 2011;6(10):e26041.

16. Kim S, Yun Z, Ha U, Lee S, Park H, Kwon EE, et al. Transfer of antibiotic resistance plasmids in pure and activated sludge cultures in the presence of environmentally representative micro-contaminant concentrations. Sci Total Environ. 2014;468–469:813–20.

17. Wu D, Huang X-H, Sun J-Z, Graham DW, Xie B. Antibiotic resistance genes and associated microbial community conditions in aging landfill systems. Environ Sci Technol. 2017;51(21):12859–67.

18. Dunivin TK, Shade A. Community structure explains antibiotic resistance gene dynamics over a temperature gradient in soil. FEMS Microbiol Ecol. 2018;94(3):fiy016.

19. Jauffur S, Isazadeh S, Frigon D. Should activated sludge models consider influent seeding of nitrifiers? Field characterization of nitrifying bacteria. Water Sci Technol. 2014;70(9):1526–32.

20. Pholchan MK, de C. Baptista J, Davenport RJ, Sloan WT, Curtis TP. Microbial community assembly, theory and rare functions. Front Microbiol. 2013;4:68.

21. Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. The contribution of species richness and composition to bacterial services. Nature. 2005;436:1157–60.

22. Hubbell SP. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton: Princeton University Press; 2001.

23. Chen Q, An X, Li H, Su J, Ma Y, Zhu Y-G. Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. Environ Int. 2016;92–93:1–10.

24. Merlin C, Bonot S, Courtois S, Block J-C. Persistence and dissemination of the multiple-antibiotic-resistance plasmid pB10 in the microbial communities of wastewater sludge microcosms. Water Res. 2011;45:2897–905.

25. Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, Stedtfeld RD, et al. Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc Natl Acad Sci U S A. 2013;110(9):3435–40.

26. Li L-G, Yin X, Zhang T. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. Microbiome. 2018;6:93.

27. Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An Omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun. 2021;12:4765.

28. Lanza VF, Baquero F, Martínez JL, Ramos-Ruíz R, González-Zorn B, Andremont A, et al. In-depth resistome analysis by targeted metagenomics. Microbiome. 2018;6:11.

CHAPTER 2

Literature Review

2.1 The Global Emergence and Spread of Antimicrobial Resistance

The continued overuse of antimicrobials in human and animal medicine and food production has resulted in the emergence and spread of antimicrobial resistance (AMR) worldwide. Without action, it is estimated that by 2050 there could be 10 million deaths each year attributed to AMR, and common infections could once again kill (1). The World Health Organizations global action plan on AMR emphasises the need for increased surveillance and research to strengthen our knowledge base of AMR (2). In the past, efforts to combat the spread have primarily focused on AMR in health and agricultural sectors. However, over recent years there has been growing recognition of the importance of the environment in the development, transmission and spread of AMR (3). Based upon the close connection between humans, animals, plants and the wider environment a 'One Health' approach to tackle AMR was adopted, aiming to draw upon the expertise of multiple sectors to work together (4).

2.2 The Relationship Between the Resistome and Microbiome Composition

In environmental settings, the microbial community composition has emerged as an important determinant in the abundance and diversity of ARGs. Significant correlations have been observed between the microbiome and resistome. Studies on the microbial community in landfill and leachates found the abundance of most ARGs to be related to the abundance of specific populations within the phyla *Firmicutes* and *Actinobacteria* (5). Whilst in mariculture samples (water and sediments), populations within the phyla *Proteobacteria*, *Actinomycetes* and *Cyanobacteria* significantly positively correlated with ARGs (6). In different soils types and at various temperatures, the composition of distinct resistomes correlated with the microbial phylogenetic and taxonomic structure (7–9). Furthermore, in wastewater treatment plants the changes in

microbial community composition throughout the treatment process has been shown to be correlated with changes in ARG abundance (10).

The movement of mobile genetic elements (MGEs), which are crucial in the mobilisation of ARGs, is also associated with the microbial community composition. The host range of plasmids is variable, some are limited to specific species or genera, whilst others are broad range and are found in multiple phyla (11,12). Of the fully sequenced plasmids, the majority are found in the *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Cyanobacteria* indicating these phyla to be important in the dissemination of antibiotic resistance (12,13). The stability of MGEs has also been shown to vary at species level in *Pseudomonas*, demonstrating that MGE dynamics in natural communities are complex and driven by species capable of maintaining plasmids (14). Despite the importance of MGEs, few studies have considered the dynamics and affinity of these AMR determinants in diverse microbial communities such as those in wastewater and activated sludge.

Resistance mechanisms target important biological function and have been shown to exert fitness costs (15). The fitness cost associated with antibiotic resistance mechanisms can vary between species, thus resulting in differences in the prevalence of ARB based on microbial community structure (16). Some bacteria have been shown to harbour resistance mechanisms cost free (17,18), whilst others have displayed fitness costs which restrict the dissemination of resistance (19–21). Fitness costs have been shown to vary based on the genetic background of the mutation (15). Therefore, changes in the microbial community are likely to change the fitness cost associated with given mutations. The fitness cost is also impacted by the ARGs present. A study evaluating the fitness cost of a multi-resistance plasmid containing 13 ARGs demonstrated that the entire fitness cost of the plasmid was associated with the carriage of only two genes; the extended spectrum beta-lactamase *bla*CTX-M15 and the *tet*AR tetracycline resistance gene (22).To this day,

no clear patterns in the cost of resistance have been elucidated, with variations observed between Gram negative and positive organisms (15), and elevated fitness cost observed at higher minimum inhibitory concentrations (MIC) (23,24). Many studies have focused on the fitness cost of AMR in individual species. However, the persistence of AMR determinants in diverse and complex environments has not been considered in depth.

2.3 Sewer Microbial Community and Antimicrobial Resistance

The microbial community in sewers originates from a combination of sources including human faecal microorganisms, microorganisms indigenously growing in the sewer infrastructure and water infiltrations carrying soil and aquatic microorganisms (25). Understanding the composition and the role of the sewer microbial community has attracted increasing attention recently due to the implications on concrete degradation (26), the variations in water quality (25) and the prevalence of ARGs (27). Sewers are the foremost section of the wastewater cycle, and thus receive waste containing antibiotic residuals, ARGs and ARB. These influxes promote resistance among the resident bacterial population (27). Residual concentrations of antibiotics and heavy metals in wastewater and sewer sediments positively select for ARB thus affecting the structure of the microbial community and the prevalence of ARGs (28,29). MGEs and ARGs are abundant within sewers and wastewater and given the potential for positive selection of these AMR determinants it is crucial that factors affecting their prevalence and abundance are further investigated.

Early studies suggest that wastewater origin may influence the abundance of ARGs in wastewater and sewers. Grabow & Prozesky, 1973 showed that 26 % of coliforms in clinical sewage have transferable resistance compared to only 4 % in municipal waste suggesting clinical waste may have a greater influence on the propagation of AMR (30). ARGs and ARBs conferring resistance

to numerous classes of antimicrobials remain frequently detected in hospital wastewaters (31–33). In addition, higher concentrations of antibiotics such as fluoroquinolones are found in hospital waste, and correlations have been observed between ARG abundance and residual antibiotics thus suggesting hospital waste causes antibiotic dissemination (34). However, another study showed that the density of bacteria resistant to amoxicillin, tetracycline and ciprofloxacin in three WWTPs are not significantly affected by the wastewater origin (35). In addition, the impact of clinical waste on the prevalence of ARG in influent wastewater was found to be negligible compared to those without clinical waste (36,37). The effect of clinical waste on ARG abundance in wastewater could be minimal due to the relatively low flow of wastewater from clinical settings and the widespread presence of ARGs within non-clinical human populations or the environment.

The activity of microorganisms in the sewer is affected by other environmental factors, including temperature, attachment to the wall or presence of sediments, flow regime and pH (38,39). The operation and layout of sewer networks can also result in the formation of diverse habitats, including both aerobic and anaerobic environments. As a consequence, the sewer environment is home to diverse biochemical functions including anaerobic processes such as fermentation and methane production (39,40), and aerobic processes such as nitrification (41–43). Based on variations in metabolic fitness costs and affinity of AMR among species, it is conceivable that the prevalence of ARGs varies between communities growing in anaerobic and aerobic setting. In wastewater systems, many of the ARBs and mobile ARGs are thought to have originated from the human and animal gut (28). In the human gut, the majority of resistance genes are harboured by strict anaerobes (44), suggesting the plausibility that anaerobic environmental conditions favour the persistence of ARGs. Research on the use of anaerobic digestion for the treatment of waste biosolids produced during the wastewater treatment process has documented increases in ARG

concentration following treatment (45,46). Other studies on low energy treatments for domestic wastewater found that anaerobic-aerobic sequence bioreactors and aerobic systems were superior to anaerobic systems alone in reducing ARG concentration, suggesting ARGs to be less stable in aerobic environments (47,48).

Conversely, it was shown that mesophilic anaerobic processes can cause reductions in the most prevalent ARGs (49,50). In addition, *Escherichia coli* and *Pseudomonas aeruginosa* exhibit faster loss of tetracycline ARGs under anaerobic conditions (51). It was proposed that the less energetically favourable (fermentative) anaerobic environment caused bacteria to lose genes to alleviate the energetic burden associated with their maintenance (51). However, research showed that during aerobic wastewater treatment the concentration of ARGs such as *bla*TEM, *sul*I and *qnr*S, increase whilst others such as *erm*B and *tet*W decrease (34), suggesting that the aerobic environment is favourable for some ARGs, whilst unfavourable for others. The variability in the cost and affinity of mobile AMR between species highlights the potential for variability in the persistence of AMR in diverse microbial communities such as those in wastewater and sludge. This relationship may also be crucial in the persistence of ARGs in the wider environment and could significantly impact the risk to public health. To date, no studies have aimed to explore the relationship between the microbiome and resistome based on the sewer operational conditions.

2.4 Microbial Community Assembly in Activated Sludge Wastewater Treatment Systems

The activated sludge (AS) treatment process is one of the most widely used biological wastewater treatment processes worldwide (52) which relies primarily on the activity of the microbial community. The composition and diversity of the activated sludge microbial community has the largest impact on the performance and stability of the system (53). Understanding the processes that drive microbial community diversity and assembly is a key question in microbial ecology.

Numerous studies have focused on variations in the activated sludge microbial community composition due to temporal and spatial differences (54), wastewater characteristics (55), operational conditions (56) and geographical location (57–59). The key focus of many of these studies is to characterise the functional organisms in activated sludge with the aim of improving system performance and reducing the frequency of bulking and foaming problems (60). However, to date few studies have considered the microbial community assembly and composition in the activated sludge in relation to the persistence of ARGs.

Numerous factors are known to affect the microbial community of the activated sludge. Analysis of three full scale treatment plants with wastewaters of different origins (municipal, slaughterhouse and refinery sewage) concluded chemical composition rather than the bacterial composition to be the main factor in shaping the taxonomic structure of the activated sludge (61). Furthermore, when fed with a common food source, different communities converge to produce highly reproducible communities with similar family level structures (62,63) and isolates within the community are capable of utilising the by products of other organisms for growth (63). Beyond the wastewater substrate composition, the activated sludge community structure can be affected by operational parameters such as the solid retention time (64), pH and temperature (65). Changes in pH determine the fate of different bacterial populations either by enabling or inhibiting their growth (66,67). Similarly, nitrite-oxidizing bacteria in the activated sludge are affected by temperature, with a temperature-dependent shift in the population structure demonstrated by cultivation-based approaches (68,69). The factors affecting the structure of the activated sludge microbial community are numerous (70), given this, the study of microbial community assembly in full scale WWTPs is often challenging as there are numerous variables between sites.

The role of the influent microbial community is one of the greatest uncertainties associated with activated sludge microbial community assembly. The potential for influent microbial community immigration has been long debated with two conflicting theories being developed in the form of the ***niche*** and ***neutral*** models (71). The niche model, which was largely exemplified in the preceding paragraphs, proposes that the microbial community is determined based upon the presence or absence of an organisms preferred environmental conditions and substrates (i.e., niche) (72). The stochastic neutral model is based upon the theory that microbes colonise an environment and change randomly within it through random birth, death, dispersal and speciation events (73,74). However, the niche and neutral theory can co-exist. Under certain conditions it has been shown that co-occurrence of similar species is not incompatible with the niche theory, and niche relations can favour co-existence (75). The potential for co-existence introduces additional complexity when considering how a microbial community will react under different conditions.

Early studies show that the relative abundance of common taxa can be modelled using the neutral model alone (76–78). In addition, the activated sludge community remains stable over time despite changes in the composition of the influent community (61), suggesting that the overall effect of the influent community is negligible, and the neutral theory applies. However, further developments in the field suggest that the influent community may have an important role in maintaining the population. By comparing the OTUs shared between the mixed liquor and wastewater influent, it can be inferred that there are high immigration rates in full scale WWTPs as shared OTUs can be observed (43,79–81). Concurrent with the niche model for community assembly, Lee et al. 2015 identified that the relative abundance of OTUs in the influent is lower in the activated sludge, thus suggesting selection to be important in the assembly of the activated sludge community. However, it should be noted that in this study, diversity was assessed using

only 500 reads per sample, which is likely insufficient to represent the true diversity within the community. In lab scale studies of anaerobic membrane bioreactors, it was found that the continuous feeding with wastewater microbiota caused bacterial community shifts (82). However, despite immigration into the activated sludge microbial community, some populations have low growth rates or there may be no growth at all (80,83). The rate of immigration can also be affected by disturbance of the resident microbial community, for example caused by changes in operation of a system or environmental conditions (84,85). During disturbances, there is a greater probability of successful colonisation due to a greater availability of resources and space (85,86), and immigration is thought to play a critical role in community recovery (84). Thus far, the importance of the influent microbial community and wastewater on the assembly of the activated sludge and the related resistome is unclear. ARG and MGEs have been identified in WWTPs, although the relationship of these AMR determinants with the microbial community structure and assembly has not been investigated in depth.

2.5 Approaches to monitoring AMR in the environment

To control the spread of AMR, effective monitoring strategies are urgently required. In the past, efforts to study AMR commonly relied upon culture based techniques to study resistance phenotypes in numerous sample types. These methods remain commonly used in numerous sectors, for example in agricultural settings *E. coli* and *Salmonella* isolates are used to assess the presence of ARB (87). In addition, the Centers for Disease Control and Prevention is committed to establishing a bank of AMR isolates obtained from healthcare, food and community settings (88). Although this approach provides an assessment of phenotypic resistance, it is limited by the number of bacteria which can be cultured in the lab. In environmental microbial communities, it is though that fewer than 1 % of bacteria can be cultured (89), consequently using these techniques

alone likely underestimates the prevalence of AMR in the environment. To address this limitation, numerous studies use quantitative PCR approaches to detect multiple ARGs in mixed microbial communities. This method has been used to analyse numerous sample types including soil (90), wastewater sludge (91) and manure (92). Although allowing AMR to be monitored in diverse microbial communities, limitations remain on the practicality of this data for surveillance and risk assessment strategies.

Over recent years, the value of information on ARG sequence variants has emerged. In different sample types, such as feces, WWTPs and environments unaffected by anthropogenic activity ARG amino acid sequence variants have been observed (93). Furthermore, it was found that ARG risk can vary within the same gene family (93). By studying ARG sequence variants in the environment, source tracking and risk assessment could be drastically improved. However, methods currently available to do so have various limitations. Metagenomic approaches are being increasingly used to detect large numbers of ARGs in environmental samples (7,94,95). This approach is limited by its sensitivity, as ARGs are often present in populations which occur in low abundance, below the detection limit (96). Consequently, deep sequencing is required which can be costly, and impedes the number of samples that can be analysed which is a major drawback when routine monitoring is required. To improve the detection limit of metagenomic sequencing, novel targeted metagenomic approaches such as ResCap (resistome capture) have been developed which significantly increase the sensitivity and specificity of metagenomic sequencing, with improved recovery of target genes (97). However, this approach further increases the costs and workflow of metagenomic sequencing, which again limits the number of samples that can be analysed. In addition, others have demonstrated this approach to underestimate the abundance of

ARGs when compared to quantitative PCR approaches (98), likely due to DNA fragmentation which is an issue when relying upon longer sequence lengths.

Recent developments in the field of AMR surveillance have moved towards shorter length reads for the detection of ARGs. This method is less impacted by DNA fragmentation in different sample types, and lower cost allowing a greater number of samples to be routinely monitored. DARTE-QM is a method which can be used to detect multiple ARGs simultaneously in environmental samples (99). The method involves using custom multiplexed primers to perform PCR-based library preparation for the MiSeq sequencing platform. The first DARTE-QM panel includes 796 primer pairs, and has been demonstrated to detect ARGs with high accuracy and sensitivity in both mock communities and environmental samples, at a fraction of the cost of metagenomic sequencing. Such methods shows promise for routine use in monitoring of environmental samples, but require optimisation to address concerns relating to PCR inhibition in specific sample types and sequencing noise. Furthermore, development of an appropriate sequencing analysis pipeline is required to distinguish between true ARG sequence variants and sequencing artifacts.

2.6 Conclusions

WWTPs have been shown to be a major source of AMR in the environment. Efforts to optimise the treatment process to achieve greater removal of ARGs often produce variable results (100–103), which likely stems from our lack of understanding of factors impacting their persistence. To enable process optimisation, we first need a better understanding of the relative contribution of different variables to AMR in the environment. In numerous settings, the microbial community composition has emerged as a key determinant of AMR (7,8,10). However, despite the importance of the microbial community in the activated sludge process, the relationship between community assembly and AMR in this environment remains poorly understood. To specifically investigate

the relationship between microbial community assembly and AMR in WWTPs, highly controlled studies are required which include appropriate reproducibility and controls. Due to the multitude of variables between full scale wastewater treatment plants, these studies can not be conducted at full scale. Instead, a model system should be developed to allow community assembly mechanisms to be accurately assessed. By combining these model systems with high-resolution methods for tracking ARG sequence variants, the dynamic relationship between the microbiome and resistome in WWTPs could be explored for the first time.

2.7 References

1. O'Neill J. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. London: Wellcome Trust; 2014. Available from: https://wellcomecollection.org/works/rdpck35v/items. [Accessed on 2018 Nov 27].

2. World Health Organization. Global action plan on antimicrobial resistance. Geneva; 2015. https://apps.who.int/iris/bitstream/handle/10665/193736/9789241509763_eng.pdf [Accessed on 2018 Jan 05].

3. United Nations Environment Programme (UNEP). Environmental Dimensions of Antimicrobial Resistance: Summary for Policymakers. 2022. Available from: https://wedocs.unep.org/bitstream/handle/20.500.11822/38373/antimicrobial_R.pdf [Accessed on 2022 May 27].

4. World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), World Health Organization for Animal Health (OIE). Strategic Framework for collaboration on antimicrobial resistance- together for One Health. Geneva; 2022. Available from:https://www.who.int/publications/i/item/9789240045408 [Accessed on 2022 May 25].

5. Wu D, Huang X-H, Sun J-Z, Graham DW, Xie B. Antibiotic resistance genes and associated microbial community conditions in aging landfill systems. Environ Sci Technol. 2017;51(21):12859–67.

6. He LX, He LY, Gao FZ, Wu DL, Ye P, Cheng YX, et al. Antibiotics, antibiotic resistance genes and microbial community in grouper mariculture. Sci Total Environ. 2022;808:152042.

7. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nat Lett. 2014;509:612–6.

8. Dunivin TK, Shade A. Community structure explains antibiotic resistance gene dynamics over a temperature gradient in soil. FEMS Microbiol Ecol. 2018;94(3):fiy016.

9. Zhou Z-C, Zheng J, Wei Y-Y, Chen T, Dahlgren RA, Shang X, et al. Antibiotic resistance genes in an urban river as impacted by bacterial community and physicochemical parameters. Environ Sci Pollut Res. 2017;24:23753–62.

10. Ju F, Beck K, Yin X, Maccagnan A, Mcardell CS, Singer HP, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J. 2019;13:346–60.

11. van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP, Aarts HJM. Acquired antibiotic resistance genes: an overview. Front Microbiol. 2011;2:203.

12. Robertson J, Bessonov K, Schonfeld J, Nash JHE. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. Microb Genomics. 2020;6(10):mgen000435.

13. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015;6:242.

14. Kottara A, Hall JPJ, Harrison E, Brockhurst MA. Variable plasmid fitness effects and mobile genetic element dynamics across Pseudomonas species. FEMS Microbiol Ecol. 2018;94(1):fix172.

15. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl. 2015;8:273–83.

16. Kassen R, Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet. 2006;38(4):484–8.

17. Hurdle JG, O'Neill AJ, Chopra I. The isoleucyl-tRNA synthetase mutation V588F conferring mupirocin resistance in glycopeptide-intermediate Staphylococcus aureus is not associated with a significant fitness burden. J Antimicrob Chemother. 2004;53(1):102–4.

18. Knight GM, Budd EL, Lindsay JA. Large mobile genetic elements carrying resistance genes that do not confer a fitness burden in healthcare-associated meticillin-resistant Staphylococcus aureus. Microbiology. 2013;159(8):1661–72.

19. Collins J, Rudkin J, Recker M, Pozzi C, O'Gara JP, Massey RC. Offsetting virulence and antibiotic resistance costs by MRSA. ISME J. 2010;4(4):577–84.

20. Laabei M, Uhlemann A-C, Lowy FD, Austin ED, Yokoyama M, Ouadi K, et al. Evolutionary trade-offs underlie the multi-faceted virulence of Staphylococcus aureus. PLOS Biol. 2015;13(9):e1002229.

21. Yokoyama M, Stevens E, Laabei M, Bacon L, Heesom K, Bayliss S, et al. Epistasis analysis uncovers hidden antibiotic resistance-associated fitness costs hampering the evolution of MRSA. Genome Biol. 2018;19:94.

22. Rajer F, Sandegren L. The Role of antibiotic resistance genes in the fitness cost of multiresistance plasmids. MBio. 2022;13(1):e03552-21.

23. Ender M, McCallum N, Adhikari R, Berger-Bächi B. Fitness cost of SCCmec and methicillin resistance levels in Staphylococcus aureus. Antimicrob Agents Chemother. 2004;48(6):2295–7.

24. Hurdle JG, O'Neill AJ, Ingham E, Fishwick C, Chopra I. Analysis of mupirocin resistance and fitness in Staphylococcus aureus by molecular genetic and structural modeling techniques. Antimicrob Agents Chemother. 2004;48(11):4366–76.

25. McLellan SL, Fisher JC, Newton RJ. The microbiome of urban waters. Int Microbiol. 2015;18(3):141–9.

26. Grengg C, Mittermayr F, Ukrainczyk N, Koraimann G, Kienesberger S, Dietzel M. Advances in concrete materials for sewer systems affected by microbial induced concrete corrosion: A review. Water Res. 2018;134:341–52.

27. Auguet O, Pijuan M, Borrego CM, Rodriguez-Mozaz S, Triadó-Margarit X, Giustina SV Della, et al. Sewers as potential reservoirs of antibiotic resistance. Sci Total Environ. 2017;605–606:1047–54.

28. Kaeseberg T, Zhang J, Schubert S, Oertel R, Siedel H, Krebs P. Sewer sediment-bound antibiotics as a potential environmental risk: Adsorption and desorption affinity of 14 antibiotics and one metabolite. Environ Pollut. 2018;239:638–47.

29. Murray AK, Zhang L, Yin X, Zhang T, Buckling A, Snape J, et al. Novel insights into selection for antibiotic resistance in complex microbial communities. MBio. 2018;9(4):e00969-18.

30. Grabow WOK, Prozesky OW. Drug resistance of coliform bacteria in hospital and city sewage. Antimicrob Agents Chemother. 1973;3(2):175–80.

31. Perry MR, Lepper HC, McNally L, Wee BA, Munk P, Warr A, et al. Secrets of the hospital underbelly: Patterns of abundance of antimicrobial resistance genes in hospital

wastewater vary by specific antimicrobial and bacterial family. Front Microbiol. 2021;12:703560.

32. Wang C, Mantilla-Calderon D, Xiong Y, Alkahtani M, Bashawri YM, Al Qarni H, et al. Investigation of antibiotic resistome in hospital wastewater during the COVID-19 pandemic: Is the initial phase of the pandemic contributing to antimicrobial resistance? Environ Sci Technol. 2022;56:15007–18.

33. Galarde-López M, Velazquez-Meza ME, Bobadilla-Del-valle M, Carrillo-Quiroz BA, Cornejo-Juárez P, Ponce-De-león A, et al. Surveillance of antimicrobial resistance in hospital wastewater: Identification of carbapenemase-producing Klebsiella spp. Antibiotics. 2022;11:288.

34. Rodriguez-Mozaz S, Chamorro S, Marti E, Huerta B, Gros M, Sanchez-Melsio A, et al. Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. Water Res. 2015;69:234–42.

35. Novo A, Manaia CM. Factors influencing antibiotic resistance burden in municipal wastewater treatment plants. Appl Microbiol Biotechnol. 2010;87:1157–66.

36. Mao D, Yu S, Rysz M, Luo Y, Yang F, Li F, et al. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Res. 2015;85:458–66.

37. Buelow E, Bayjanov JR, Majoor E, Willems RJ, Bonten MJ, Schmitt H, et al. Limited influence of hospital wastewater on the microbiome and resistome of wastewater in a community sewerage system. FEMS Microbiol Ecol. 2018;94(7):fiy087.

38. Biggs CA, Olaleye OI, Jeanmeure LFC, Deines P, Jensen HS, Tait SJ, et al. Effect of temperature on the substrate utilization profiles of microbial communities in different sewer sediments. Environ Technol. 2011;32(1–2):133–44.

39. Jin P, Shi X, Sun G, Yang L, Cai Y, Wang XC. Co-Variation between distribution of microbial communities and biological metabolization of organics in urban sewer systems. Environ Sci Technol. 2018;52(3):1270–9.

40. Shi X, Ngo HH, Sang L, Jin P, Wang XC, Wang G. Functional evaluation of pollutant transformation in sediment from combined sewer system. Environ Pollut. 2018;238:85–93.

41. Shanks OC, Newton RJ, Kelty CA, Huse SM, Sogin ML, McLellan SL. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl Environ Microbiol. 2013;79(9):2906–13.

42. Wells GF, Wu CH, Piceno YM, Eggleston B, Brodie EL, DeSantis TZ, et al. Microbial biogeography across a full-scale wastewater treatment plant transect: Evidence for immigration between coupled processes. Appl Microbiol Biotechnol. 2014;98:4723–36.

43. Lee S-H, Kang H-J, Park H-D. Influence of influent wastewater communities on temporal variation of activated sludge communities. Water Res. 2015;73:132–44.

44. van Schaik W. The human gut resistome. Philos Trans R Soc London. 2015;370:20140087.

45. Tong J, Liu J, Zheng X, Zhang J, Ni X, Chen M, et al. Fate of antibiotic resistance bacteria and genes during enhanced anaerobic digestion of sewage sludge by microwave pretreatment. Bioresour Technol. 2016;217:37–43.

46. Tong J, Lu X, Zhang J, Sui Q, Wang R, Chen M, et al. Occurrence of antibiotic resistance genes and mobile genetic elements in enterococci and genomic DNA during anaerobic digestion of pharmaceutical waste sludge with different pretreatments. Bioresour Technol. 2017;235:316–24.

47. Christgen B, Yang Y, Ahammad SZ, Li B, Catalina Rodriquez D, Zhang T, et al. Metagenomics shows that low-energy anaerobic−aerobic treatment reactors reduce antibiotic resistance gene levels from domestic wastewater. Environ Sci Technol. 2015;49:2577–84.

48. Lamba M, Ahammad SZ. Performance comparison of secondary and tertiary treatment systems for treating antibiotic resistance. Water Res. 2017;127:172–82.

49. Sui Q, Zhang J, Chen M, Tong J, Wang R, Wei Y. Distribution of antibiotic resistance genes (ARGs) in anaerobic digestion and land application of swine wastewater. Environ Pollut. 2016;213:751–9.

50. Jang HM, Shin J, Choi S, Shin SG, Park KY, Cho J, et al. Fate of antibiotic resistance genes in mesophilic and thermophilic anaerobic digestion of chemically enhanced primary treatment (CEPT) sludge. Bioresour Technol. 2017;244(1):433–44.

51. Rysz M, Mansfield WR, Fortner JD, Alvarez PJJ. Tetracycline resistance gene maintenance under varying bacterial growth rate, substrate and oxygen availability, and tetracycline concentration. Environ Sci Technol. 2013;47(13):6995–7001.

52. Hreiz R, Latifi MA, Roche N. Optimal design and operation of activated sludge processes: State-of-the-art. Chem Eng J. 2015;281:900–20.

53. Miura Y, Hiraiwa MN, Ito T, Itonaga T, Watanabe Y, Okabe S. Bacterial community structures in MBRs treating municipal wastewater: Relationship between community stability and reactor performance. Water Res. 2007;41:627–37.

54. Xia S, Duan L, Song Y, Li J, Piceno YM, Andersen GL, et al. Bacterial community structure in geographically distributed biological wastewater treatment reactors. Environ Sci Technol. 2010;44(19):7391–6.

55. Hu Q-Y, Li M, Wang C, Ji M. Influence of powdered activated carbon addition on water quality, sludge properties, and microbial characteristics in the biological treatment of commingled industrial wastewater. J Hazard Mater. 2015;295:1–8.

56. Kim YM, Cho HU, Lee DS, Park D, Park JM. Influence of operational parameters on nitrogen removal efficiency and microbial communities in a full-scale activated sludge process. Water Res. 2011;45:5785–95.

57. Shu D, He Y, Yue H, Wang Q. Microbial structures and community functions of anaerobic sludge in six full-scale wastewater treatment plants as revealed by 454 high-throughput pyrosequencing. Bioresour Technol. 2015;186:163–72.

58. Gao P, Xu W, Sontag P, Li X, Xue G, Liu T, et al. Correlating microbial community compositions with environmental factors in activated sludge from four full-scale municipal wastewater treatment plants in Shanghai, China. Appl Microbiol Biotechnol. 2016;100(10):4663–73.

59. Wagner M, Loy A, Nogueira R, Purkhold U, Lee N, Daims H. Microbial community composition and function in wastewater treatment plants. Antonie Van Leeuwenhoek. 2002;81:665–80.

60. Ju F, Zhang T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. ISME J. 2015;9:683–95.

61. Shchegolkova NM, Krasnov GS, Belova AA, Dmitriev AA, Kharitonov SL, Klimina KM, et al. Microbial community structure of activated sludge in treatment plants with different wastewater compositions. Front Microbiol. 2016;7:90.

62. Wittebolle L, Van Vooren N, Verstraete W, Boon N. High reproducibility of ammonia-oxidizing bacterial communities in parallel sequential batch reactors. J Appl Microbiol. 2009;107(2):385–94.

63. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Microbiology. 2018;361:469–74.

64. Han H, Zhang Y, Cui C, Zheng S. Effect of COD level and HRT on microbial community in a yeast-predominant activated sludge system. Bioresour Technol. 2010;101(10):3463–5.

65. Sun H, Chang H, Tang W, Zhang X, Yang H, Zhang F, et al. Effects of influent immigration and environmental factors on bacterial assembly of activated sludge microbial communities. Environ Res. 2022;205:112426.

66. Cremer J, Arnoldini M, Hwa T. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. Proc Natl Acad Sci. 2017;114(25):6438–43.

67. Ratzke C, Gore J. Modifying and reacting to the environmental pH can drive bacterial interactions. PLOS Biol. 2018;16(3):e2004248.

68. Alawi M, Off S, Kaya M, Spieck E. Temperature influences the population structure of nitrite-oxidizing bacteria in activated sludge. Environ Microbiol Rep. 2009;1(3):184–90.

69. Karkman A, Mattila K, Tamminen M, Virta M. Cold temperature decreases bacterial species richness in nitrogen-removing bioreactors treating inorganic mine waters. Biotechnol Bioeng. 2011;108(12):2876–83.

70. Xu S, Yao J, Ainiwaer M, Hong Y, Zhang Y. Analysis of bacterial community structure of activated sludge from wastewater treatment plants in winter. Biomed Res Int. 2018;2018:8278970.

71. Chase JM, Myers JA. Disentangling the importance of ecological niches from stochastic processes across scales. Philos Trans R Soc London B. 2011;366:2351–63.

72. Pholchan MK, de C. Baptista J, Davenport RJ, Sloan WT, Curtis TP. Microbial community assembly, theory and rare functions. Front Microbiol. 2013;4:68.

73. Hubbell SP. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton: Princeton Univeristy Press; 2001.

74. Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. The contribution of species richness and composition to bacterial services. Nature. 2005;436:1157–60.

75. Leibold MA, McPeek MA. Coexistence of the Niche and Neutral Perspectives in Community Ecology. Ecology. 2006;87(6):1399–410.

76. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ Microbiol. 2006;8(4):732–40.

77. Akarsubasi AT, Eyice O, Miskin I, Head IM, Curtis TP. Effect of sludge age on the bacterial diversity of bench scale sequencing batch reactors. Environ Sci Technol. 2009 Apr 15;43(8):2950–6.

78. Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, et al. Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci U S A. 2010;107(35):15345–50.

79. Jauffur S, Isazadeh S, Frigon D. Should activated sludge models consider influent seeding of nitrifiers? Field characterization of nitrifying bacteria. Water Sci Technol. 2014;70(9):1526–32.

80. Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 2016;10:11–20.

81. Dottorini G, Michaelsen TY, Kucheryavskiy S, Andersen KS, Kristensen JM, Peces M, et al. Mass-immigration determines the assembly of activated sludge microbial communities. Proc Natl Acad Sci. 2021;118(27):e2021589118.

82. Seib MD, Berg KJ, Zitomer DH. Influent wastewater microbiota and temperature influence anaerobic membrane bioreactor microbial community. Bioresour Technol. 2016;216:446–52.

83. Mei R, Narihiro T, Nobu MK, Kuroda K, Liu W-T. Evaluating digestion efficiency in full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial activity. Sci Rep. 2016;6:34090.

84. Guo B, Sheng Z, Liu Y. Evaluation of influent microbial immigration to activated sludge is affected by different-sized community segregation. npj Clean Water. 2021;4(1).

85. Tilman D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. Proc Natl Acad Sci U S A. 2004;101(30):10854–61.

86. Vuono DC, Munakata-Marr J, Spear JR, Drewes JE. Disturbance opens recruitment sites for bacterial colonization in activated sludge. Environ Microbiol. 2016;18(1):87–99.

87. CIPARS. The Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) 2018 Executive Summary. 2018. Available from: https://www.canada.ca/content/dam/phac-aspc/documents/services/surveillance/canadian-integrated-program-antimicrobial-resistance-surveillance-cipars/cipars-reports/2018-annual-report-executive-summary/2018-annual-report-executive-summary.pdf [Accessed 2022 Dec 31].

88. Centers for Disease Control and Prevention. CDC & FDA Antibiotic Resistance Isolate Bank [Internet]. 2022 [cited 2022 Dec 31]. Available from: https://www.cdc.gov/drugresistance/resistance-bank/index.html

89. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ. 2013;447:345–60.

90. Chen Q, An X, Li H, Su J, Ma Y, Zhu Y-G. Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. Environ Int. 2016;92–93:1–10.

91. Merlin C, Bonot S, Courtois S, Block J-C. Persistence and dissemination of the multiple-antibiotic-resistance plasmid pB10 in the microbial communities of wastewater sludge microcosms. Water Res. 2011;45:2897–905.

92. Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, Stedtfeld RD, et al. Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc Natl Acad Sci U S A. 2013;110(9):3435–40.

93. Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An Omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun. 2021;12:4765.

94. D'Costa VM, McGrann KM, Hughes DW, Wright GD. Sampling the antibiotic resistome. Science (80-). 2006;311(5759):374–7.

95. Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E, Larsson DGJ. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. Front Microbiol. 2014;5:648.

96. Lynch MD, Neufeld JD. Ecology and exploration of the rare biosphere. Nat Rev Microbiol. 2015;13:217–29.

97. Lanza VF, Baquero F, Martínez JL, Ramos-Ruíz R, González-Zorn B, Andremont A, et al. In-depth resistome analysis by targeted metagenomics. Microbiome. 2018;6:11.

98. Macedo G, van Veelen HPJ, Hernandez-Leal L, van der Maas P, Heederik D, Mevius D, et al. Targeted metagenomics reveals inferior resilience of farm soil resistome compared to soil microbiome after manure application. Sci Total Environ. 2021;770:145399.

99. Smith SD, Choi J, Ricker N, Yang F, Hinsa-Leasure S, Soupir ML, et al. Diversity of Antibiotic Resistance Genes and Transfer Elements-Quantitative Monitoring (DARTE-QM): a method for detection of antimicrobial resistance in environmental samples. Commun Biol. 2022;5:216.

100. Zhu Y, Wang Y, Zhou S, Jiang X, Ma X, Liu C. Robust performance of a membrane bioreactor for removing antibiotic resistance genes exposed to antibiotics: Role of membrane foulants. Water Res. 2018;130:139–50.

101. Xia S, Jia R, Feng F, Xie K, Li H, Jing D, et al. Effect of solids retention time on antibiotics removal performance and microbial communities in an A/O-MBR process. Bioresour Technol. 2012;106:36–43.

102. Staley C, Kaiser T, Vaughn BP, Graiziger C, Hamilton MJ, Kabage AJ, et al. Durable long-term bacterial engraftment following encapsulated fecal microbiota transplantation to treat Clostridium difficile infection. MBio. 2019;10(4):e01586-19.

103. Liu SS, Qu HM, Yang D, Hu H, Liu WL, Qiu ZG, et al. Chlorine disinfection increases both intracellular and extracellular antibiotic resistance genes in a full-scale wastewater treatment plant. Water Res. 2018;136:131–6.

CHAPTER 3

Multiplexed Amplicon Sequencing Reveals High Sequence Diversity of Antibiotic Resistance Genes in Québec Sewers

Connecting text: The value of monitoring antimicrobial resistance gene (ARG) sequence diversity for source tracking and risk assessment purposes has been increasingly recognised over recent years. However, current methods for the detection of ARG sequence variants have various limitations, including their high cost and poor detection limit in diverse microbial communities. In this chapter a multiplexed PCR method for the detection of ARG amplicon sequence diversity in environmental samples was developed. To demonstrate the use of this tool and potential applications of the data, sixteen samples were collected from the inlet of wastewater treatment plants located in Quebec, Canada. Influent wastewater has been demonstrated to contain high concentrations of ARGs, which likely originated from numerous sources including domestic and clinical waste, industrial discharges and sewer biofilms. Therefore, it was hypothesised that ARGs in the sewer samples would display high ARG sequence diversity. The comprehensive antibiotic resistance database was used to explore the risk of different ARG amplicon sequence variants and the potential applications of this tool.

3.1 Introduction

Antimicrobial resistance (AMR) is widely recognised as one of the greatest threats to modern medicine of the twenty first century. Traditional AMR monitoring strategies focused on areas with high antibiotic usage such as clinical and agricultural settings. More recently, the importance of the environment in the development, spread and transmission of AMR to humans and animals has emerged (1,2). One of the most notable examples is the *bla*NDM gene which was first identified in 2008 (3), and was demonstrated to disseminate in the environment through wastewaters (4). Such recognition calls for wider surveillance of antimicrobial resistance genes (ARG) in wastewater and other environmental reservoirs.

Studies are increasingly recognising the value of monitoring ARG sequence diversity for source tracking purposes. Machine learning classification has been used to profile the sequences of ARGs from different environments and train models to determine potential source types (5,6). Both common and unique ARGs were identified across ecotypes highlighting the potential value of obtaining data on ARG sequence diversity (6). To improve the performance of source tracking models and allow the tracking of resistome development in specific areas, it has been suggested that regional studies of ARGs are required. However, this can be costly and the data handling cumbersome. Zhang et al., 2021 studied amino acid sequence variants of ARGs in different habitats including feces, WWTPs and environmental samples from non-anthropogenically impacted areas. Assessment of ARG risk revealed that ARGs in the same gene family can pose substantially different risks due to differing mobility, hosts and ecological distribution (7). Taken together, these studies demonstrate the valuable information that can be gained and potential applications of ARG sequence variant monitoring. Yet, to date these approaches are limited due to the high cost per sample and the lengthy analyses required.

Current methods for the detection of ARGs in environmental samples have various limitations. Shotgun metagenomics can detect a large number of ARGs and mobile genetic elements, but is less sensitive than qPCR (8) and the analysis requires high levels of training. In environmental samples, the sequencing depth of metagenomic sequencing is often insufficient to represent all diversity within these complex communities (9,10). This method can also be costly, often meaning that fewer samples are analysed, which in turn limits the construction of large datasets which are required to understand the connectivity between environmental reservoirs. Quantitative PCR techniques have a higher detection limit (10^{-5} to 10^{-7} copy/genome), but further characterisation of sequence diversity in genes can be challenging.

Here, we introduce a multiplex targeted amplicon sequencing approach for the detection of sequence diversity of ARGs in environmental samples. A selection of antibiotic and metal resistance genes were amplified in a multiplex PCR reaction, barcoded to enable sample pooling and sequenced on the Illumina Miseq platform. This method is similar to that developed independently by Smith et al., 2022 (11) and enables relatively fast and low-cost analysis of sequence diversity in multiple ARGs and samples in a single sequencing run. This protocol can also be easily adapted to target different genes depending on the users' aims, enabling versatility and exhaustive performance calibration in the analysis of the targeted resistome.

To demonstrate the use of multiplex targeted amplicon sequencing, we surveyed influent wastewater from 16 wastewater treatment plants located in the province of Quebec, Canada. Wastewater has been demonstrated to contain a plethora of ARGs and antimicrobial resistant bacteria (ARBs). Previous studies have identified genetic diversity in ARGs based on the sample type and location (6). ARGs in influent wastewater are thought to originate from numerous sources including domestic and clinical waste, industrial discharges, urban run-off and growth selection in

sewer biofilms. Consequently, it was hypothesised that ARG sequence diversity would be observed among the sewer samples collected using our novel amplicon sequencing approach. Influent wastewater offers many potential pathways for the environmental dissemination of antibiotic resistance, through combined sewer overflow whereby untreated wastewater is discharged directly into the environment or incomplete removal during the wastewater treatment process. Thus using this data, we explore the risk of different ARG amplicon sequence variants (ARG-ASVs) and the potential applications of this tool.

3.2 Materials and Methods

3.2.1 Sample Collection

Grab samples were collected from the influent wastewater inlet at 16 Québec wastewater treatment plants between June and August 2018. Wastewater samples were transported to McGill University on ice and processed within 24 hours of collection. Samples were centrifuged in 2 mL microcentrifuge tubes at 12,000 ×g for 5 min., the supernatant was then removed, and centrifuged solids were stored at –80 °C until further analysis. Wastewater treatment plant metadata for the sampling period was obtained through the Québec Data Partnership for wastewater treatment plants (12). Samples obtained were assigned a sample ID as reported in Table S3.1, ordered according to the longitude of the sampling location.

3.2.2 Mapping of Sampling Area

Maps were produced using QGIS software and georeferenced in WSG 84 EPSG:4326. The co-ordinates of the samples WWTP influents were retrieved from publicly available data from the Government of Quebec (13). The co-ordinates of ambulatory health care services, hospitals, and nursing and residential care facilities were retrieved from the open database of health care facilities available on the Government of Canada open database of healthcare facilities (14) and imported

to QGIS as a CSV file. Land cover data from 2015 was retrieved from the Government of Canada Website (15), the TIFF file was imported to QGIS and projected from EPSG 3978 NAD83/ Canada Atlas Lambert. The colour scheme for each land category was as presented in the 2015 North American Land cover map (16) and according to the colour scheme outlined in Supplementary Table S3.2. Cattle numbers were retrieved from the 2016 agricultural census data (17). The TIFF file was imported to QGIS and projected from EPSG:3347 – NAD83/ Statistics Canada Lambert. The census consolidated subdivision level was used with a graduated colour ramp of four quantile (equal count) colour classes.

3.2.3 Microbial Community Analysis

DNA was extracted from stored biomass samples using DNeasy PowerSoil Kit (Qiagen, Germantown, MD, USA), as per manufacturers instructions using a centrifugation speed of 10,000 x g. PCR of the 16S rRNA gene V4 region was conducted using the 515F and 806R modified Caporaso primers (18,19). The PCR conditions were as follows: 94 ℃ for 3 mins followed by 35 cycles of 94 ℃ for 45 secs, 50 ℃ for 60 secs, 72 ℃ for 90 secs. 72 ℃ for 10 mins and a final hold at 4 ℃. After amplification of the 16S rRNA gene, amplicons were barcoded in a second PCR reaction with the following reaction conditions: initial denaturation at 94 ℃ for 3 minutes, followed by 15 cycles of 94 ℃ for 30 secs, 59 ℃ for 20 secs, 68 ℃ for 45 secs and a final elongation at 68 ℃ for 5 mins. Amplicons were sequenced on the Illumina MiSeq PE250 platform at Génome Québec Innovation Centre (Montréal, QC, Canada). The number of sequences obtained per sample are reported in Table S3.1.

3.2.4 Multiplexed Amplicon Sequencing

Target Amplicon Sequencing was performed using a custom built accel-amplicon™ Panel designed in collaboration with Swift Biosciences (now Integrated DNA Technologies). This

method involved performing a highly multiplexed PCR reaction followed by barcoding and sequencing. The panel used included a total of 159 primer pair combinations targeting 114 ARGs (Table S3.3). Primers were designed to produce amplicons of similar length with an average of 275 base pairs to avoid preferential amplification of shorter fragments.

Samples were prepared according to the manufacturer's instructions. An input DNA concentration of 25 ng was used. Magnetic clean-up steps were conducted using SPRI Select Beads (Beckman Coulter, IN, USA). The reaction conditions for the multiplexing PCR were as follows: 98°C for 30s followed by 4 cycles of 98°C for 10s, 63°C for 5 mins, 65°C for 1 min, then 22 cycles of 98°C for 10s and 64°C 1 min, and terminated by a final elongation step of 65°C for 1 min and hold at 4°C. After multiplexing, barcoding and clean-up steps were completed as outlined in the accel-amplicon protocol. Samples were sequenced on the Illumina MiSeq PE250 platform at Génome Québec Innovation Centre (Montréal, QC, Canada).

3.2.5 Bioinformatics

16S rRNA amplicon sequencing data was processed using Qiime2 software (20) and R Software with ‘vegan’ packages (21). Raw sequences were quality filtered using DADA2 (22) and ASV tables extracted from QIIME2 after quality filtering for further analysis in R. Taxonomy was assigned using the MiDAS 2.0 reference database (23) and tabulated. Principal coordinate analysis was performed using R ‘vegan’ package (21) and Bray-Curtis Dissimilarity. A heatmap of the most abundant genera was generated using the function ‘heatmap.2’ in ‘gplots’ package in R (24).

Sequenced Accel-Amplicon samples were obtained from Genome Quebec and forward and reverse reads merged using pear v.0.9.11. Subsequently, the merged reads were assigned to a specific gene product based on their beginning and end (representing the forward and reverse primer sequence used) using a custom python script (available upon request) and all reads converted to fasta format.

Reads may have multiple assignments (resulting from degenerate primer sequences). In these cases, we kept the first assignment for each duplicated read. Reads were sorted by length and deduplicated using vsearch v.2.13.3 (25). Unique sequences detected for each target gene may represent real ASVs or result from PCR and sequencing errors. To filter out such errors, we assumed that the distribution of ASV frequencies under an assumption of pure error would follow a Poisson distribution centred around the mean of all counts, with many ASVs only having one or two reads (which might have been seeded by cross-contamination or misassignments of reads during sequencing). For each sample and gene, we tested whether we could detect ASVs whose frequencies represented outliers to a Poisson distribution (after correction for multiple testing using the Bonferroni method), representing higher abundance, real ASVs after PCR. If such ASVs were detected, we removed low count "noise" ASVs from the sample by finding the minimum of the distribution function and cutting all ASVs below the minimum. A custom R script for ASV filtering is available upon request. For products in which more than a single ASV was retained after filtering, OTU tables were constructed using the clusterfast method in vsearch with an id of 1. To ensure that the ASVs represented correct products, we used tblastx to compare all ASVs to a custom database of ARGs based on a dereplicated version of the CARD database with a minimum id of 70 % (blast v.2.12.0). ASVs that did not have hits to their respective target gene at this id level were removed.

3.2.6 Multiple Sequence Alignment of ARG Sequences

For multiple sequence alignment, ARG sequences were obtained from NCBI Blast based upon 100 % query coverage and 100 % identity with the ARG-ASVs observed. Results were filtered to include only entries with the gene name included in the description (*bla*TEM or *bla*OXA) and complete CDS only. Sequences were aligned using MAFFT and the fast Fourier transform method

(FFT-NS-2) (26) for multiple sequence alignment. Phylogenetic trees were generated using the average linkage (UPGMA) method.

3.2.7 Inverse PCR

Inverse PCR was used as described by Pärnänen et al., 2016 (27) to analyse the flanking regions surrounding the *bla*OXA ARG. Briefly, the unknown flanking regions were digested using restriction enzymes and then ligated together. Clones were then sequenced using Sanger sequencing at Genome Quebec, Montreal. The obtained forward and reverse sequences were analysed using NCBI BLAST (28).

3.2.8 Statistical Analysis

Redundancy analysis (RDA) was used to explain the variation in ARG-ASVs (response variables) due to the explanatory variables considered (influent characteristics, land use and geographical location) (13–15,17). Geographical data (longitude and latitude) was transformed into distance based eigen vector maps (DbMEM) and used as an explanatory variable in the RDA. DbMEM variables were constructed using function dbmem() of package 'adespatial' in R (29). The function give.thresh() of package 'adespatial' in R was used to select a matrix truncation value of 0.825 to retain only distances among close neighbours. RDA of Hellinger-transformed sequence data for each gene was computed using the 'rda' function in R 'vegan' package (21). To account for RDA bias which varies based on the number of explanatory variables included, the R^2 values obtained were adjusted using Ezekiel's formula. The adjusted R^2 obtained measured the unbiased explained variation. An unbiased amount of explained variation close to 0 indicates that the explanatory variable did not explain more of the variation in the response variable than random deviates would. The significance of the R^2 values was tested using function anova.cca in the R 'vegan' package (21).

3.3 Results and Discussion

3.3.1 Microbial Community of Sewer Samples

In this study, sixteen wastewater samples were analysed which were collected at the entrance of WWTPs distributed along the Ottawa and St. Lawrence rivers between Gatineau and Quebec City (Table S3.4). The microbial community composition in the wastewater samples was assessed using 16S rRNA gene sequencing. Principle coordinate analysis revealed clustering of the microbial communities (Figure 3.1a), which did not significantly correlate with operational parameters such as flow and pipe design (all gravity sewer versus presence of pressured lines) or the geographical proximity of samples (Table S3.4) as assessed using a permutation test of environmental variables (Table S3.5). Analysis of the overall top 20 most abundant genera in the influent wastewater samples revealed *Acinetobacter* to be dominant members of the microbial community in Clusters 1 and 2 (Figure 3.1b). This is consistent with previous studies which identify *Acinetobacter* as dominant members of the wastewater microbial community worldwide (30–32). Although many species are non-pathogenic, others such as *Acinetobacter baumannii*, exhibits high levels of AMR to many antibiotics and are a major cause of antimicrobial resistant infection worldwide (33). Taxa from the family *Comamonadaceae*, which are phenotypically diverse and commonly reported in soil and water environments (34), were also dominant in all influent wastewater samples. Another dominant member of the microbial community included *Trichococcus*, a pathogen commonly reported in influent wastewater (35,36), which has been linked to sludge bulking problems in WWTPs (37).

Cluster 2 communities (sampling sites: G, N, C, O and B) featured a higher relative abundance of *Arcobacter* than in the communities of Clusters 1 and 3. *Arcobacter* are frequently observed in high abundance in influent wastewater samples (38). *Arcobacter* have been isolated from numerous sources including humans, animals and wastewater and are considered emerging

pathogens which have been associated with both human and animal disease (39). Raw wastewater samples have been identified as a potential source of *Arcobacter* infection (40). Finally, Cluster 3 (sampling sites: P and L) were found to harbour communities containing lower relative abundances of *Acinetobacter* and *Arcobacter*. Their distinct microbial communities from the other samples was dominated by *Aeromonas*, which is widespread in the environment and causes a wide range of diseases in humans and animals (41). AMR strains of *Aeromonas* have been demonstrated to persist after the wastewater treatment process and in receiving surface waters (41). Considering the level of AMR in influent wastewater, and the high abundance of important potential pathogens such as *Acinetobacter* and *Arcobacter*, increased surveillance of these environments would be beneficial to manage the associated risk and prevent further dissemination of AMR.

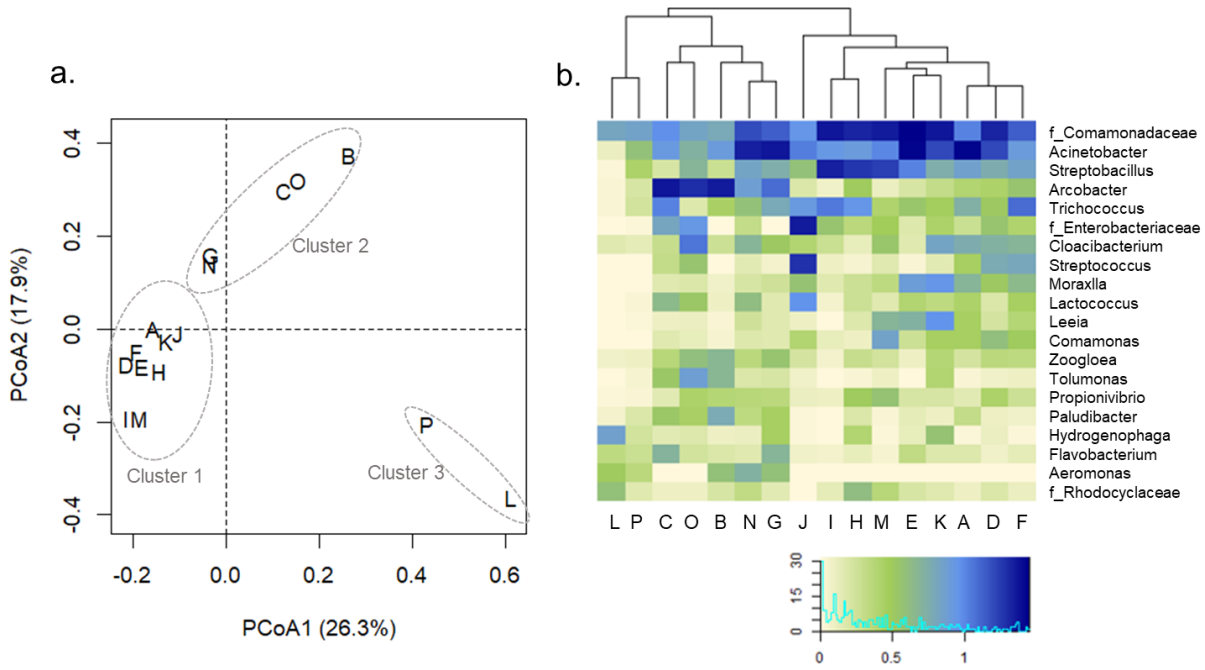


Figure 3.1- Bacterial community analysis based on amplicon sequencing of the V4 region of the 16S rRNA gene of wastewater samples collected from the inlet to 16 WWTPs in Quebec, Canada. a) Principal coordinate analysis with Bray-Curtis distance b) Heatmap of top 20 most abundant overall genera

3.3.2 High Sequence Diversity Observed in Sewer ARGs

Using targeted amplicon sequencing, a total of 60 ARGs and metal resistance genes (MRGs) were detected in the sewer samples, 16 of which displayed sequence diversity (Table 3.1). Of the sixteen genes, between 3 and 45 different ASVs were observed indicating high genetic diversity in sewer ARGs. Considering the stringent Poisson distribution-based filters used to identify real ASVs from sequencing errors, it is plausible that the ARG sequence diversity may have been underestimated. The high prevalence of ARG sequence diversity from a single sample type demonstrates the information that is missed when using traditional detection

methods such as quantitative PCR.

A redundancy analysis was conducted to partition the variance in ARG-ASV distributions among a number of explanatory variable groups (influent characteristics, land use, geographical location and microbial community composition; Table S3.4). A value of 0 indicates that the explanatory variables did not explain more of the variation than random deviations, whist a value of 1 explains 100 % of the variation. Negative values obtained indicate that the explanatory values explain less variance than a set of random normal deviates. Overall, the explanatory variables included in the redundancy analysis significantly explained only a small proportion of the variance in the distribution of ARG-ASVs . Influent factors (pipe design, influent flow and design population size) explained 30 % of the variance in the *mdt*G gene but did not significantly impact the diversity of other ARGs (Table 3.1**)**. The geographical location explained only 11 % of the variation in the *qac*L efflux pump and 19 % of the variance in the *sil*E gene. Finally, the variation in ARG diversity explained by land factors (land cover, cattle numbers and hospitals) and microbial community composition was not statistically significant (Table 3.1). Taken together, these results suggest that sequence diversity is influenced by other parameters not included in the redundancy analysis.

Table 3.1: Genes detected in sewer samples with sequence diversity and partitioning of variance

Gene	Resistance Mechanism	# Sewers Detected[a]	# Variants Observed	Unbiased Variation Explained			
				Influent Factors[b]	Land Factors[c]	Geographical Factors[d]	Microbial Community[e]
*ars*A	Metal resistance	11	16	0.00	0.12	−0.02	0.07
*bla*ACT	Beta-lactamase	13	45	0.07	−0.07	0.01	0.03
*bla*OXA	Beta-lactamase	16	4	0.58	−0.21	−0.13	0.07
*bla*SHV	Beta-lactamase	9	9	0.09	−0.03	−0.06	0.08
*bla*TEM	Beta-lactamase	16	3	0.20•	−0.07	0.07	−0.15
*cop*A	Metal resistance	11	16	0.08	0.18•	−0.04	0.06
*mdt*F	Efflux pump	7	7	−0.03	−0.13	0.01	−0.07
*mdt*G	Efflux pump	5	3	0.30*	0.13	−0.09	0.04
*mdt*H	Efflux pump	6	4	0.16	−0.15	0.011	0.15
*mdt*L	Efflux pump	5	3	−0.10	−0.20	−0.10	−0.08
*mef*A	Efflux pump	13	9	0.08	0.02	0.02	−0.07
*oqx*B	Efflux pump	13	17	−0.02	−0.04	0.06•	−0.12
*pmr*C	Target alteration	12	4	−0.05	−0.10	−0.02	0.1
*qac*L	Efflux pump	16	18	−0.12	0.32•	0.11*	0.22
*sil*E	Metal resistance	15	4	0.17	−0.04	0.19*	0.07
*tet*A	Efflux pump	15	3	0.11	−0.51	−0.07	−0.09

[a] Total number of samples = 16

[b] Influent explanatory factors = Flow (m^3/day) + Sewer pipe type (pressure main, gravity, both) + Design population size

[c] Land explanatory factors = land cover + cattle + hospitals

[d] Geographical explanatory factors = defined using PCNM eigenfunctions

[e] Microbial community defined by eigen values obtained from PCoA using Bray-Curtis dissimilarity

[f] Significance code- • $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p<0.001$, values without notation were not statistically significant

3.3.3 Patterns of Sequence Diversity and Reported Hosts

For the purpose of demonstrating the value of this tool, four ARGs were chosen for further characterisation (Figure 3.2). The *bla*TEM and *bla*OXA genes, which confer resistance to beta-lactam antibiotics, were selected due to their clinical importance and detection in all sewer samples. The final two genes selected were *mdt*H (multidrug efflux pump) and *pmr*C (conferring resistance to peptide antibiotics) to study the impact of other mechanisms of resistance and to

different antimicrobials. The ARG sequence diversity profiles between Gatineau and Quebec City (Figure 3.2a) for the remaining genes are reported in Figure S3.1.

Four ASVs of the *bla*OXA ARG were observed among the sewer samples (Figure 3.2b). The *bla*OXA Variants 1 and 4 were the most frequently detected, with a greater number of reads obtained for Variant 4 suggesting a higher abundance. The widespread detection of both Variant 1 and 4 suggest that the hosts are well adapted to the environmental conditions within the sewers or that there is a high influx of these variants from sources such as the human gut microbiome. Variant 2 and 3 were observed in only one sampling site each.

NCBI Blast (42) was used to gain more information about the previously reported hosts of each ASV. Based on the criteria that each ASV may differ by only a single nucleotide polymorphism (SNP), only 100 % identity matches were accepted in this analysis. Variant 1, was previously detected in several hosts such as genus *Acinetobacter, Aeromonas* and *Enterobacter* (Table S3.6). Variant 2 was reported in fewer hosts. However, those identified largely overlapped with previously reported hosts of Variant 1 (Table S3.6). Surprisingly Variant 4 (most abundant and most frequently detected) corresponded to no hits with 100 % identity in the NCBI Blast database, whilst Variant 3 had 100 % identity over 100 % of its sequence with positive hits in over 30 species (Table S3.6). The closest reported sequence to Variant 4 matched 99 % of the query and was reported in hosts from the genus *Enterobacter, Klebsiella* and *Pseudomonas*, which overlapped with the genera reported to contain Variant 3. The lack of previous observation of *bla*OXA Variant 4 could be due to sampling bias of the database, which primarily includes clinical isolates and much fewer environmental strains. Conversely, Variant 4 differs from Variant 3 in only the last 6 nucleotides. Specifically, the last nucleotide of Variant 4 is different from that in Variant 3.

Therefore, it is possible that this difference could be a sequencing/bioinformatics artefact, and Variant 3 and Variant 4 should be considered equivalent.

Among the sampled sewer communities, *bla*TEM Variant 2 was most prevalent (Figure 3.2c), whilst sub-dominant Variant 1 appeared to be present only in sampling locations typically far from Montreal. Both *bla*TEM Variant 1 and 2 have been reported in a large number of hosts (Table S3.6), suggesting that the gene is not restricted by phylogenetic barriers and possibly explaining the high occurrence of these ASVs among the sewer samples as some carriers are likely well adapted to this environment. *Bla*TEM Variant 3 has been reported only in synthetic constructs, which may justify its detection in only one wastewater samples and prevents any further hypothesis inference about its source. However, future sampling of different environments could enable tracking of such unique ASVs in different reservoirs to identify its origin and ecological vectors.

Four ASVs of the *pmr*C ARG were detected among the wastewater samples (Figure 3.2d). Variants 1 and 2 were the most frequently observed, and were reported in the same host species in the NCBI database (among genera *Escherichia*, *Salmonella* and *Shigella*; Table S3.6). Variant 3 and 4, which were identified in only one site each, have only been previously reported to be associated with *Escherichia coli*, suggesting a limited host range and consequently that they are less well adapted to persist in the sewer environment or the associated source microbiomes.

The *mdt*H gene was detected in fewer sewer samples, and the ASVs detected varied between sampling sites (Figure 3.2e). Variant 1 to 4 of the *mdt*H gene were all previously reported in taxa from the genera *Escherichia* and *Shigella*. The occurrence of all ASVs in similar hosts somewhat explains the seemingly random assembly of ASVs in different locations, as it is likely that all are similarly adapted to the sewer environment.

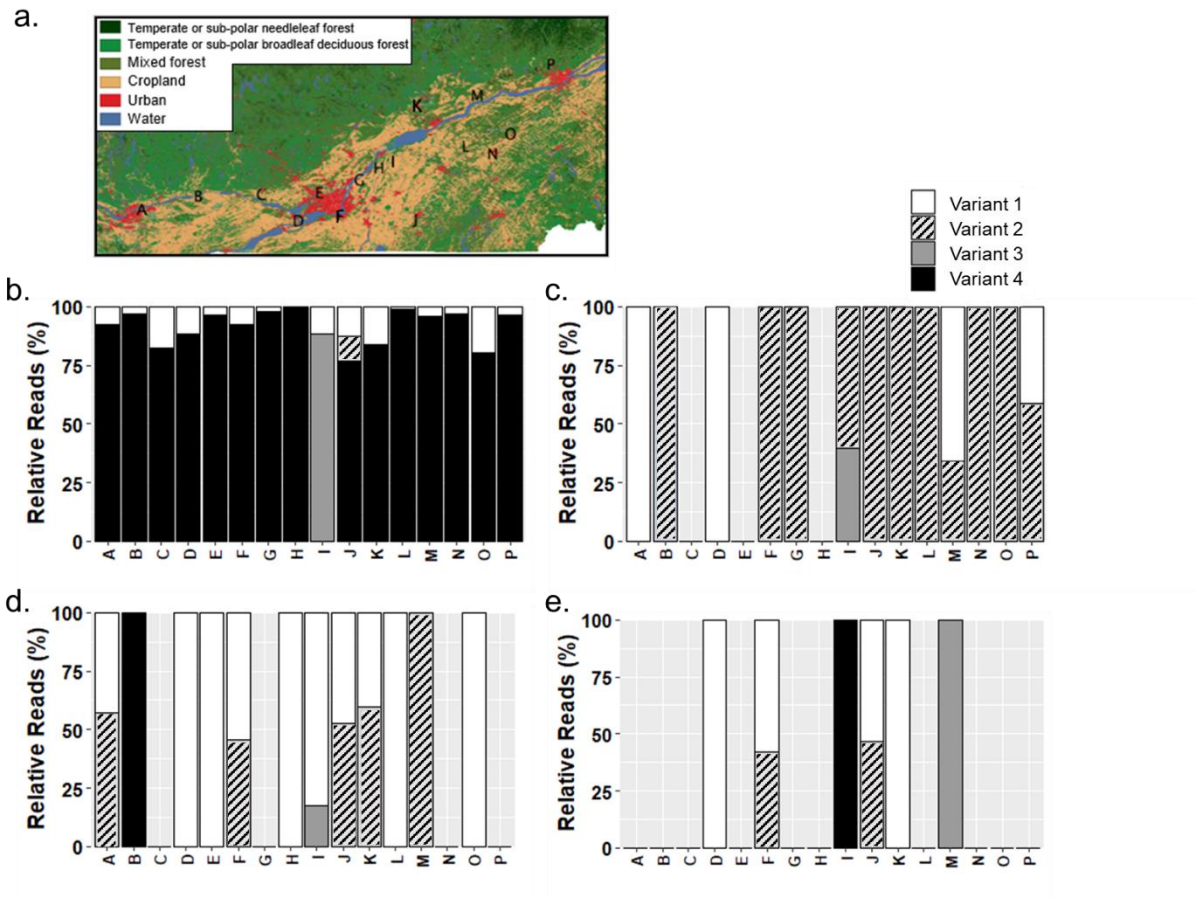


Figure 3.2 Sequence distribution for four genes detected in wastewater samples obtained in a transect along the Ottawa-St-Lawrence river basin between Gatineau (Point A) and Québec City (Point P). a) Sampling locations with respect to land coverage. Star indicates the island of Montréal. b) *bla*OXA c) *bla*TEM, d) *pmr*C, e) *mdt*H

Taken together, analysis of previously reported hosts using NCBI Blast suggest that the occurrence of a given ASV may be influenced by the number of potential hosts and their ability to grow and compete within the sewer environment. Interestingly, many dominant ASVs returned no 100 % matches in the NCBI database. This may be due to the sampling bias associated with the database, where a greater number of samples are obtained from clinical than environmental settings. Given this presumed bias, this result raises questions on the interpretation of qPCR-based monitoring of

AMR. These approaches provide little information on whether the quantified target is of any direct importance to the clinical outcomes in human or livestock, or if they merely represent general environmental circulation of genes in a given family. Such interrogations ask that we become increasingly precise in the identification of targets that are detected along with improving sample processing to expand the volume of environmental samples analysed.

3.3.4 Genetic context of ARG amplicon sequence variants

ARGs are common within the environment. However, their direct risk to public health is associated with the pathogenicity of their host and the likelihood of infection. This was aptly defined by Zhang et al., 2021, who outlined that the risk of a given ARG is largely dependent on 1) its enrichment in human associated environments 2) gene mobility and 3) its presence in pathogens. Mobile ARGs present in pathogens are considered to be current threats, whilst those that are mobile but not yet observed in pathogens are classified as future threats to public health.

To assess the mobility of each ARG-ASV, PLSDB (a plasmid database) was used to determine if they had been previously reported on a plasmid (Table 3.2) (43). The *bla*OXA Variants 1, 2 and 3 returned positive hits in the PLSDB, whilst no hits were obtained for Variants 4 (Table 3.2). Interestingly, plasmids containing *bla*OXA Variant 1 were detected in animals (dogs, pigs, chicken, duck), human and environmental (wastewater) samples, demonstrating that plasmids containing this variant have the ability to persist in a range of environments. Whereas Variant 2 and 3 were reported only on a plasmid obtained from a human sample.

Variants 1 and 2 of the *bla*TEM gene also returned hits in PLSDB. However, Variant 3 was not previously reported on a plasmid. This suggests that short length sequences may be used to distinguish between mobile and non-mobile variants of a given ARG. Variant 2 was observed in plasmids associated with domestic animals (dogs), livestock (chicken, turkey, cattle and swine)

and the environment (wastewater) again highlighting the potential for plasmids containing this ASV to persist in different environments. Conversely, plasmids with 100 % match to Variant 1 were only reported in soil and feed additives, suggesting more limited movement. Based on the mobility of the ASVs, it could be hypothesised that *bla*TEM Variants 1 and 2 pose a greater risk to public health than Variant 3 due to their potential ability to transfer to pathogens and cause infections. Further studies on ARG-ASVs associated with clinical infections and environmental strains could be used to expand our understanding of these patterns.

No 100 % hits in the PLSDB database were obtained for any of the *pmr*C ASVs, suggesting that this gene is likely less mobile than others and therefore poses less risk unless already present within pathogens. Of the *mdt*H ASVs, only Variant 1 returned positive hits in the PLSDB database. Variant 1 of the *mdt*H gene was the most frequently observed among the sewer samples, this could be related to the mobility of this gene and ability to transfer to hosts well adapted to survive within wastewater environments.

Considering the presence or absence of each ARG-ASV in the plasmid database alone, it can be observed that the variable mobility of a given ARG could be detected using short read sequencing data obtained using multiplexed amplicon sequencing approaches. Mobile ARG observed in human associated environments have the potential to transfer into human pathogens and present a high risk of contributing to new multi-drug pathogens in the future. Mobile ARGs originating from clinical environments also have the potential to transfer to environmental bacteria, which could aid in the dissemination of AMR in the environment. This demonstrates the need for ARG-ASV tracking, to monitor current and future high-risk sequence variants and develop effective control strategies to tackle the spread of AMR.

Table 3.2: Occurrence of ARG amplicon sequence variants among 263 clinically relevant pathogenic bacteria reported in CARD

Gene	Sequence Variant	PLSDB Database	Occurrence in Pathogens (n)			
			NCBI Chromosome	NCBI Plasmid	NCBI WGS	NCBI GI
*bla*OXA	1	Y	Y (5)	Y (16)	Y (30)	Y (2)
	2	Y	Y (1)	Y (1)	Y (2)	N
	3	Y	Y (8)	Y (7)	Y (30)	N
	4	NA	NA	NA	NA	NA
*bla*TEM	1	Y	Y (6)	Y (1)	Y (34)	Y (3)
	2	Y	Y (28)	Y (35)	Y (54)	Y (11)
	3	N	N	N	N	N
*pmr*C	1	N	Y (1)	N	Y (4)	N
	2	N	N	N	N	N
	3	N	N	N	N	N
	4	N	N	N	N	N
*mdt*H	1	Y	Y (2)	Y (1)	Y (6)	N
	2	N	N	N	N	N
	3	N	N	N	N	N
	4	N	N	N	N	N

n= number of pathogen species in which the variant has been detected (total pathogens= 263)

Y/N indicate if there were (Y) or were not (N) hits in the specified database for the variant.
NA represents genes which obtained no 100 % identity positive hits in any database.

3.3.5 High Risk ARG Amplicon Sequence Variants

The mobility of an ARG can be used as an indicator of the likelihood of future risk based on its ability to transfer into a pathogen. Whilst ARGs already present in pathogens present the greatest current risk due to the ability of the host to cause infection. The Comprehensive Antimicrobial Resistance Database (CARD) was used to determine the prevalence of the ARG-ASVs among the

sequenced genomes, plasmids, whole-genome shotgun (WGS) assemblies and predicted genomic islands (GI) of 263 important pathogens. Results from this analysis were aggregated to provide the percentage occurrence of a given gene within each species (full method available (44)). For this analysis, the 'perfect' paradigm of RGI was selected which tracked perfect matches of each ASV to reference sequences and mutations in the CARD database occurring within pathogens. This is consistent with the analysis conducted using NCBI Blast in Section 3.3.3 which considered only 100 % identity matches. This data was used to assess the prevalence and percentage occurrence of each ASV among 263 clinically relevant pathogens to determine whether each ARG variant is a current risk to public health based upon available data.

Analysis of the *bla*OXA gene ASVs in CARD revealed that Variants 1, 2 and 3 have been previously reported in pathogens. Variant 1 and 3 have been previously reported on plasmids in several different pathogens, whilst Variant 2 was only identified on a plasmid in *Pseudomonas aeruginosa* (Table 3.2). Furthermore, Variant 1 has also been identified in the predicted genomic islands of two pathogens. In the three databases consulted thus far (NCBI Blast, PLSDB and CARD), no 100% identity matches have been obtained for both Variant 4, suggesting that multiplexed amplicon sequencing can be used to detect novel sequence ASVs. Given the detection of Variants 1 to 3 in pathogens, it could be suggested that these forms of the *bla*OXA gene present a current risk to public health due to the ability of the hosts to cause infection. The presence of Variant 1 and 3 in a greater number of pathogens also suggests increased likelihood of infection, and thus it could be hypothesised that they pose a greater risk to public health than Variant 2.

Both *bla*TEM Variants 1 and 2 were identified in pathogens using CARD. Variant 2 was identified in a greater number of pathogenic hosts than Variant 1, many of which were associated with plasmids demonstrating the mobility of *bla*TEM genes with this sequence variant. Variant 3 was

not detected among the 263 pathogens included in CARD. Overall this alludes to *bla*TEM genes containing Variants 1 and 2 to being current threats to public health due to their previous associations with pathogens and their mobility. Conversely, Variant 3 has a lower chance of causing infection or transferring to pathogens because it was not previously reported in pathogens or on plasmids; thus, it should be considered a lower risk. These results highlight the different characteristics that can be inferred and monitored within a single ARG class and the need to study risk at sequence level.

To conclude this analysis, *pmr*C and *mdt*H were less frequently observed among the 263 pathogens in CARD. Variant 1 of the *pmr*C gene was detected in only four pathogens and was not associated with plasmids. Consequently, Variant 1 of the *pmr*C would be unlikely to transfer to other hosts via plasmid mediated horizontal gene transfer and thus is likely to remain associated with only a limited number of pathogens. Similarly, *mdt*H Variant 1 was observed in six pathogens. However only one report was on a plasmid. This suggests that the plasmid has recently transferred to pathogenic bacteria or limited mobility of the plasmid.

3.3.6 Wider Sequence Diversity in ARGs

Multiplexed amplicon sequencing was used to study the sequence diversity within a targeted area of each ARG (approximately 275 base pairs). To gain more information about possible sequence diversity in the remainder of the gene and to inform future target design, NCBI Blast was used to obtain the full gene sequences in which the ASVs have been previously reported. The obtained sequences were aligned to produce a phylogenetic tree enabling the diversity among the genes to be visualized.

Alignment of *bla*OXA sequences demonstrated that Variants 1 and 2 clustered together and belonged to the OXA-10 like subfamily of the gene, whilst Variant 3 and sequences similar to

Variant 4 (as no 100 % identity matches were obtained) belonged to the OXA-2 like subfamily. Previous studies have shown *bla*OXA-10 to have weak carbapenemase activity, but variants of this gene within the subfamily such as *bla*OXA-656 are more efficient carbapenemases (45). ARG Variant 1 was identified in numerous variants of the *bla*OXA-10 gene, many of which such as *bla*OXA-17 and *bla*OXA-74 display extended spectrum beta lactamase activity (46). Whilst *bla*OXA sequence Variant 2 was identified in only a small cluster of the *bla*OXA-10 subfamily, that included the *bla*OXA-101 and *bla*OXA-56 genes which display narrow spectrum beta lactamase activity (47). The *bla*OXA-2 like subfamily which clustered separately (containing Variants 3 and 4 of the current study), have been demonstrated to have more broad-spectrum activity. In the future, primer design could be optimized to target specific subgroups within the *bla*OXA-2 family and enable a more accurate assessment of ARG activity and hence risk.

Multiple sequence alignment of the *bla*TEM ARG showed the ASVs to be more widely distributed among different classes of *bla*TEM genes. The ASVs were often identified in the same *bla*TEM genes. For example, Variant 1 to 3 were present in *bla*TEM-1 and Variant 1 and 2 are both in *bla*TEM-181. Furthermore, the ASVs could not be used to distinguish between genes with narrow or extended spectrum activity. Among the *bla*TEM genes, extended spectrum beta-lactamases such as *bla*TEM-181 and *bla*TEM-166 are derived from genes with narrow spectrum beta-lactamase activity such as *bla*TEM-1 by mutations which alter the enzymes active site (48). In the future, primer optimization to target these mutations could improve risk assessment by providing additional information on differences in phenotypic resistance.

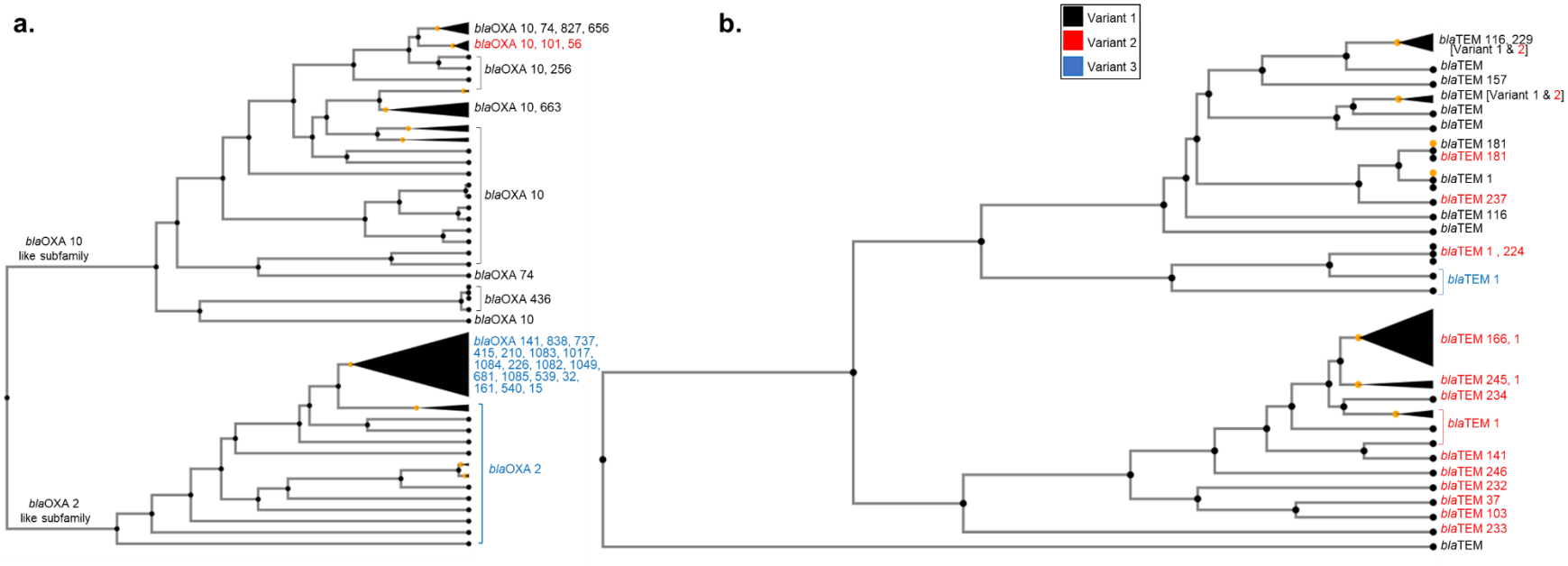


Figure 3.3- Multiple sequence alignment of ARGs containing each sequence variant in a) *bla*OXA and b) *bla*TEM

3.3.7 Genetic Context of ARGs in the Environment

The risk and genetic context analyses conducted thus far have relied upon data available in public databases, which are often bias towards clinically derived isolates. To further investigate the genetic context of the specific ASVs in the samples collected in the current study, an inverse PCR approach was used to obtain sequence information on the flanking regions of the ARG. For this analysis, we focused on the *bla*OXA-2 like ASVs (variant 3 and 4) as they were the most widely detected among the sewer samples. Clones obtained from 3 samples from along the transect of sampling sites (site A, F and P) were analyzed. Using this approach, it was found that all of the flanking regions sequenced were previously reported to be associated with mobile genetic elements including integrons such as *Int*I1, transposons such as tn21, and plasmids. This suggests that mobility may be crucial to the persistence of ARG-ASVs in the environment. Among the plasmids, some such as pALT31 and pALT33 were present in all three sampling sites. These plasmids were first reported in wastewater biosolids, and have been demonstrated to harbor ARGs to numerous classes of antimicrobials and also persist in the absence of antibiotics (49). In the future, inverse PCR approaches could be used in combination with multiplexed amplicon sequencing approaches to gain further information on the hosts of ARGs in the environment and confirm the findings of database focused analyses.

3.3.8 The Future of Multiplexed Amplicon Sequencing

Using targeted amplicon sequencing, this study identified novel ARG-ASVs not previously reported in NCBI, CARD or PLSDB databases. Information on ARG sequence diversity can be used to infer the risk of a given ARG-ASV based on its mobility and occurrence within pathogenic bacteria. As previously discussed by Zhang et al. 2021, we demonstrate that different ASVs within an ARG class present different risks. This information could be a valuable asset in environmental

surveillance and could be used to inform management strategies to control emergence and spread of AMR.

Targeted amplicon sequencing is a relatively low-cost technique, which allows multiple samples to be analysed in a single sequencing run with a detection limit much lower than achievable by other methods. This will allow a greater number of samples to be analysed and patterns in sequence diversity to be observed. In the future, targeted sampling of AMR hotspots could be used to build databases of unique and common ARG-ASVs observed in each environment. This data would be beneficial in environmental studies for AMR source tracking, and in clinical settings to determine the risk of a given ARG-ASVs or the mechanistic origin or recurrent nosocomial infections.

3.4 Supplementary Material

Table S3.1: Sample ID of Sewer Samples and 16S rRNA Gene Sequences Obtained Before Filtering

Sampling Location	Given ID	Sequences Obtained
Gatineau	A	92066
Granby	J	86674
La Prairie	F	69821
Lachute	C	82091
Papineauville	B	75985
Pincourt	D	104849
Plessisville	O	97617
Rosemere	E	101762
Shawinigan	K	92524
St Casimir	M	96397
St Eulalie	L	65885
St Ours	H	87912
St Robert	I	87473
Stoneham	P	92964
Verchères	G	63155
Victoriaville	N	82945

Table S3.2: Colour scheme applied for mapping land use

Value	Description	Colour Scheme
1	Temperate or sub-polar needleleaf forest	RGB 0 61 0
2	Sub-polar taiga needleleaf forest	RGB 148 156 112
3	Tropical or sub-tropical broadleaf evergreen forest	RGB 0 99 0
4	Tropical or sub-tropical broadleaf deciduous forest	RGB 30 171 5
5	Temperate or sub-polar broadleaf deciduous forest	RGB 20 140 61
6	Mixed forest	RGB 92 117 43
7	Tropical or sub-tropical shrubland	RGB 179 158 43
8	Temperate or sub-polar shrubland	RGB 179 138 51
9	Tropical or sub-tropical grassland	RGB 232 220 94
10	Temperate or sub-polar grassland	RGB 225 207 138
11	Sub-polar or polar shrubland-lichen-moss	RGB 156 117 84
12	Sub-polar or polar grassland-lichen-moss	RGB 186 212 143
13	Sub-polar or polar barren-lichen-moss	RGB 64 138 112
14	Wetland	RGB 107 163 138
15	Cropland	RGB 230 174 102
16	Barren lands	RGB 168 171 174
17	Urban	RGB 220 33 38
18	Water	RGB 76 112 163
19	Snow and Ice	RGB 255 250 255

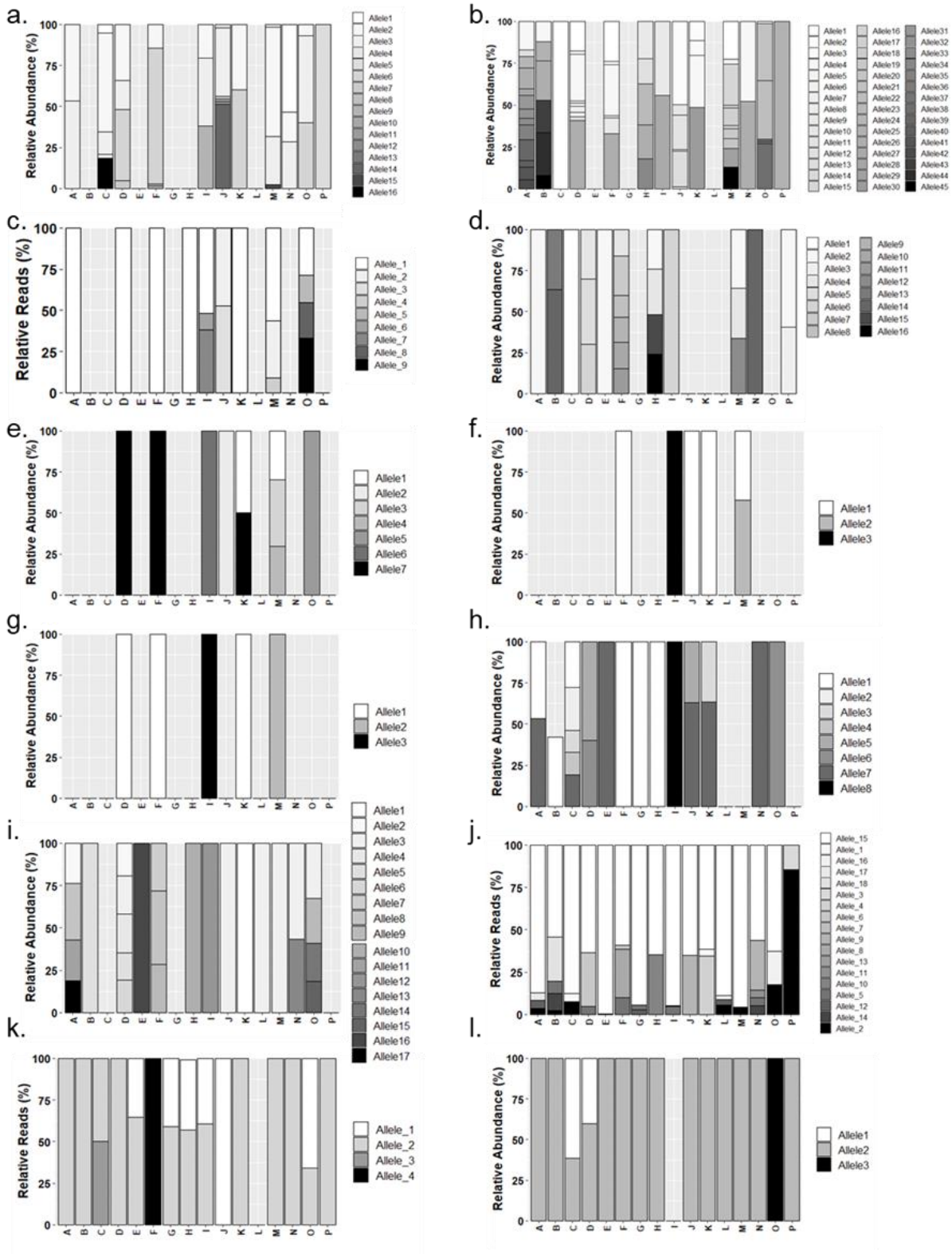


Figure S3.1- ARG sequence diversity of a) *ars*A b)*bla*ACT c) *bla*SHV d) *cop*A e) *mdt*F f) *mdt*G g) *mdt*L h) *mef*A i) *oqx*B j) *qac*L k) *sil*E l) *tet*A

Table S3.3 Multiplexed Amplicon Sequencing Primers

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
ampC	CGCCGTAATTCAGCCGCTGA	ampC	GGAGGTGTGGGTCGCCAAATTC
ampCblaDHA	TAAACCGCTGATGGCACAGCA	ampCblaDHA	GCACCCAGCACACCTGTGA
amra	GCAGCAGGACGTCGAGGTG	amra	TTTCATCTGCGTGGCCTCGAC
ant2D	CGTGTAACCCGCAAGCACGATGATA	ant2D	AGCCTGTAGGACTCTATGTGCTTTGTAGG
ant3D	TGGCAGCGTAAGGACATTCTTGC	ant3D	ATTGCCCAGTTGGCAACGA
ant6ia	GTTCGCCATGAGCTGCTGA	ant6ia	ATATCAGCGGCATATGTGCTATCC
ant9	AGGATATTGTCCCTTGGCATTTCC	ant9	AGATGCTGTCAGCCACATTCTG
aph2Id02aph901	ATTCATGCGCCCAATGGTTTCA	aph2Id02	TGCCCGCATGTATATCAGAATGAC
aph3ia	TGAAACATGGCAAAGGTAGCGTTGC	aph3ia	TCGTGATTGCGCCTGAGCGA
aph3via01	GCTTGCTATCTATAAGGAGGCACTCA	aph3via02	CACGAGTCTCGGTTAACTCATTCC
aph4ia	GCGTGGATATGTCCTGCGGGTAA	aph4ia	GACCGATTCCTTGCGGTCCGA
aph4ib	GTTCGTCCACGGCGACCTG	aph4ib	CAGAGGAACTGCGCCAGTTCC
aph6ic01	ACCAGGCGCTGCTGATGG	aph6ic01	GCTCGCCGAAGTCGAGCA
aph6ic02	GCTGCTCACCACCCACTCG	aph6ic02	CTGGAACCATTCCTGTAGCGGATG
ARR01	TTTGGCGATTGGTGACTTGCTAACCAC	ARR01	CCACGGCGCTTTAAGTCCTCCAAC
arsA01arsA03a	CGATGGCCGGGCTGGAAA	arsA01	AGTAATAGGGTATCTGTCGGTAGCTC
		arsA02	GTAGTAATAGGGTATCTGTCGGCAACTC
arsA02arsA03b	CCAATGGCCGGGCTGGAAA	arsA03aarsA03b	GGAGTAACAGGGTATCTGTCGGTAGCTC
betompK	ACCGGTTACGGCCAGTGGGA	betompK	GCAAAGTTCAGGCCGTCAACCAGA
bl1ec	GGTATGGCGGTAGCGGGAATTTATCA	bl1ec	TCGCTTGAGGATTTCACCTCATCCG
bl2akcc	CGTACTGCGCGACCTCGAC	bl2akcc	GCTCGCCGGAGTTCAGCTC
bl2anps	CGAGCCGTCTCGACCGCATC	bl2anps	CCGAGATGAACATGGTCGCGATCC
bl2bfona02	CAGCGCGATTCCCGGCGATA	bl2bfona02	CATCCTGCTGCGGCTGGGTAAA
bl2btle01	ATGTCGTTGGCCGAGTTGTG	bl2btle01	GGTCTGGTTGGCCCGCATCC
bl2btle02	CAACACCGCCGGCAACCTG	bl2btle02	CGGTCTTGTCGGCCGTGGTC

Table S3.3 Contunued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
bl2bula01	CCGCCATCCCTGGCGATGA	bl2bula01	TGCTGTGGCTGTGTAAAGTAGG
bl2bula02	TGTGCAGCACCAGCAAGGTGATG	bl2bula03	GCCGCCAAGATGCTCAAGGA
bl2bula03	TCAGCAGCAACTGTCTGAACTGG		
bl2cbro02	TGTGTTGGCCTCATCTGCACATC	bl2cbro02	GCCAGGTGGCACCCTTTCC
bl2dicrNPS01	ATCCTCGTGCGCAACAACGATAC	bl2dicrNPS01	GGAGAAAGGCAACCTGTTCTTCCG
bl2dicrNPS02	AAGAGCGCATTCAGGGTCAGCTG	bl2dicrNPS02	AGCCGCCCAACCGGTCTTG
blaACI	CCTGCGCGAAATGAACCTGA	blaACI	CGTGCGGCGCGACATTA
blaACT01	TGGCAGCCGCAGTGGAA	blaACT01	GGCGTAATGCGCCTCTTC
blaACT02	CCGGATGAGGTCACGGATAACGC	blaACT02	TCCCAGCCCAGACCCTGATACA
blaACT03	TGGCGCAGTCTCGCTACTGG	blaACT03	GCCACGTAGCTGCCAAACC
blaAMPH	TGATGCGCTGGATGCAGCAG	blaAMPH	CCAGGTCGTTGATGCCATCGC
blaBEL	ATGGCACAGACTGTGTCAAAGCTG	blaBEL	GACGTGCTAACTTTGCTACAGAAGCA
blabjp1	CGAGCTGGGAGATGACCGTCAAG	blabjp1	GCTTGGCGAGCTGCTTGTCG
blablaec	CGCCGAACAAACCGATACTGGA	blablaec	CGTCACACAGCAAACGTTGTTTGA
BlaCAU	CTGCACCAGCTGGACCAC	BlaCAU	TGTTGAAGGCGGCTTCGG
blaCMY01	CTAACTCCAGCATTGGTCTGTTTGG	blaCMY01	GGCCAGTTCAGCATCTCCCA
blaCMY02	ACTGGCAGCCGCAGTGGA		
blacphA	TGCTGGAGGTGATCAACAACAACTAC	blacphA	CTCCTTGAGGATGCAGTTGCCA
Blactxm01	GGAAGTGTGCCGCTGTATGC	Blactxm01	AAGCGCTCATCAGCACGATAAAG
Blactxm02	AAAGCTGGCGGCGCTGGA	Blactxm02	ATTGTCGCTGTACTGCAACGC
Blactxm03	CCTTCCGTCTGGACAGAACCG	Blactxm03	GAATGCTCGCGCTACCGGTA
blaFOX	GCCGATGATCAAGGAGTATCGG	blaFOX	GGGAACTGCAGCGGCAA
Blaimp01	CAGCACGGGCGGAATAGAGTG	Blaimp01	GCGGACTTTGGCCAAGCTTCTA
blaLAT	GAGTTACGAAGAGGCAATGACCA	blaLAT	ACATATCGCCAATACGCCAGTAG
blaLCR	GTGCAGAGAACCAGGCAATCGC	blaLCR	TGTCTAGAGTCTGGTCCTGGTTCCAA

Table S3.3 Continued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
blaMIR	AGGCCATTCCGGGTATGGC	blaMIR	GGTTGCCAGATCCAGCATGC
blaMOX	ACCAGCTCGGCGGATCTG	blaMOX	GAGCCGGTCTTGTTGAAGAGC
blaOXA03	AGGCACGATAGTTGTGGCAGAC	blaOXA03	GTAGAATTCCGCATTGCTGATCGC
blaOXA04	GACGGAAAGCCAAGAGCCATGAA	blaOXA04	CACCATGCGACACCAGGATTTGA
blaPAOPDC	TACAACCGGTGATGAAGGCCTATG	blaPAOPDC	GCGGTATAGGTCGCGAGGTC
blaPER	TCGCAAATGAAGCGCAGATGCA	blaPER	CTCTGGTCCTGTGGTGGTTTCGA
blaPSE	CGGACCTGCTGCAATTCATTGCC	blaPSE	ACACTACATAGGCGCCGAAACCG
blaSHV11LEN01	GCCGCCATTACCATGAGCGA	blaSHV11LEN01	GCACGGAGCGGATCAACGG
blaSHV11LEN02	GCCGCCATTACCGTGAGCGA	blaSHV11LEN02	AGCACGGAACGGATCAACGG
blaSHV11LEN03	CGCCGTCATTACCATGAGCGA	blaSHV11LEN03	GCACGCAGCGGATCAACGG
blaTEM	GAACCGGAGCTGAATGAAGCC	blaTEM	CGGGAGGGCTTACCATCTGG
blaVEB	CCATTTCCCGATGCAAAGCGTTA	blaVEB	GGGAATTCCTCTTTAATCGGACTCCA
blaVIM	TCTCCACGCACTTTCATGACGAC	blaVIM	GTCGGTCGAATGCGCAGCAC
catA1	CGACGATTTCCGGCAGTTTCTACACATA	catA1	CCCTGCCACTCATCGCAGTACTG
chlchiA	AGCTCTTGTTTGGACCGCTATCG	chlchiA	CATGCCCAAACGTAGAAACGCAAA
cmlA101	CAACGAGTTGCGGGCTTGCA	cmlA101	TGGCAAGCAATACTGCTCCAGCTA
cmxA	GCGTGCACTGTCGATCCTGCTC	cmxA	GCCAGGAAGGTGAATGCCGCAA
copA	GCGAGGTCGGTTGGGCTCA	copA	AGAACGGCACCTACTGGTACCAC
dfhr01	GTGCCAAAGGTGAACAGCTCCTG	dfhr01	CCACCACCTGAAACAATGACATGATCC
dfhr02dfrA1,5,14	GCGGTCCAGACATATCCTGGTC	dfhr02dhfr02	CCAGACACTATAACGTGACCGTTGA
dfrA1201	GGTAATCTCAAGCCAAGCTAACTACC	dfrA12-01,04,04a	ACGCGCATAAACGGAGTGG
dfrA1202	AATCTCACGCCAAGCTAACTACC	dfrA12-02,03,204b	AATGGAGTGGGTGTACGGAATTACAG
dfrA1203adfrA1203b	GGTAATCTCACACCAAGCTAACTACC		
dfrA1204adfrA1204b	TGGTAATCTCACGCCAAGCTAACTACC		
dfrA22	GTCGTTATGGGCCGCAAGACG	dfrA22	GTGCGTGTACGTAATTGTGGCTTGA

Table S3.3 Continued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
dfrA2501	GTGATCGGTTGCGGTCCACAC	dfrA2501	AGCGTAGAGGCCATGGGCAA
dfrAq	ACATACCCTGGTCCGCGAAAG	dfrAq	CGCCACCAGACACTATAACGTGA
dhfr01	CGGGAATGGCCCTGATATTCCATGGA	dhfr01	CACCACCTGAAACGATGACATGATCCG
emrD	GATCCCGGCGGCGTTCTTC	emrD	CGGCAGCATCGCCGAAAGC
ereA01	GAAACGGCTCGGCAGGAG	ereA02	ATTGTTGTGCGCCAGCAGA
ereA02	GGAAACGGCTCAGCAGGAG	ereA01	TGATTATTGTGCGCCAGCAGA
ereB02	CAGCTCATCGATCACCTCATGAAACCG	ereB02	CACGTACGGAAGTATCTCCCTCAA
lncWtrwB	TTGTGGCGGGCCTGCAATCG	lncWtrwB	AGGCGGTGAGGTCGGGCAA
macA	GACGCGCTGCGTAATGTTGA	macA	TTGCGTCTGTTTCACAGCCATC
marA	CCAGACGCAATACAGACGCTATTACC	marA	TTCGCCTTGCATATTGGTCATCCG
marR	CACAGTTTAAGGTGCTCTGCTCTATCC	marR	GCAAATACTCAAGTGTTGCCACTTCG
mcr101mcr102	GCTATTAATCATGGGCGCGGTGA	mcr101	GGCATGATCGGATTGACATACGAACGC
mdtF	TTTGGTTCCGAGTACGCCATGC	mdtF	GTTATAGCGCGCCACGGTGGA
mdtg	AAGAGATGCTGCATATGCGGGAAG	mdtg	TGGCGTCTGAACGTATGACATTGG
mdth	CTGCCGTTAAATGGATGTATGCCA	mdth	CCAATCGCCAGAACCAGACG
mdtL	CCAGCGAGGCGCAGTTGCATA	mdtL	GCTAACACCGGAATGATGCAGGTA
mefA	GATCTGCGATAGTCTTGTCTATGGC	mefA	ACTGACTATAGCCTGCACATTTCG
mexD	CTGGTGATGTTCCTGTTCCTACAG	mexD	GAAGGCCAGCGGCAGGAACAC
mexE	GACACGTCGTCCGGCACGA	mexE	GCTGCGTGCCGTTCACGAC
mgt	TCGGTTCGGCCTTCACCAA	mgt	GGCGTTGCCGAACTGGTC
mphA01	GCTGGCTCGACGACGATTC	mphA02	CCGAGTCGAGGGCGAAGA
mphE	AAGTGAGCAATTGGAAACCCGCTA	mphE	AGGCCGCTGCTCTTTCTAAAGTC
msrD01	GGCAAGCTAGGTGTTGAGCAATTAG	msrD01	CCTTCACGATCTAAATGGCTCGTA
msrD02	GCCTTATCGGCACAGGTTCATGG	msrD02	TTCCTTCACGGTCTAGATGACTGGTA
msrD03	GGCAAGCTAGGTGTTGAGCA	msrD03	TCCTTCACGGTCTAAATGGCTCGTA
		nimA01	TCCTGTCATGTGGTCGATGCA

Table S3.3 Continued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
nimA02	ACAAGCTGGACGCCATCGC	nimA02	AGCTCGATGGCCTCTTTGCC
obrj	CGCGCTGGACTACCAGGAAGC	obrj	AGGCTGATGCGCGGGAAGAAC
okp	AACACCTTGCCGACGGGATG	okp	ACCATCCACTGCAGCAGCTG
oprn	CGCTTCGAAAGCCTGTGGTGGAA	oprn	AGGGCGTCGCTGGACTCCA
oqxB	TGCTGGTGGTGCTGGTAGTGATC	oqxB	AAACGCCATCGGCACGAACAC
otra	GTACGGGTGCGCTTCGAC	otra	GACTGCGAACGCCGGTAC
PBRT	GCAGGAAGGCCGTCAGATCAC	PBRT	CTCATGGCGCTCTATCAAGTCGTC
pcoA	GCCATTCCGGTCTGCAGGA	pcoA	GGCCTGCCCGTTCATGAGA
pilDPA	CGACCTACAACCTGGTGCTG	pilDPA	GGTGATCGGCATCGATCAGG
PmrC01	ACGCCATCACCTTGCAGGTTA	PmrC01	GGCAGTTTCAGTACCGTGCATG
PmrC02	TTCGGCAGACAATAACCACCATAACC	PmrC02	TATTGTCATCGGTGAGCGTTCAGAAC
qacFH01	GTTGCAATCTTTGGCGAGGTCA	qacFH01qacFH03a	TTTAGAACGGCGACACCACTG
qacFH02	GCTGTTTCAATCCTTGGCGAGGTCA	qacFH02qacFH03b	CGCTGACCTTGGATAGCAGGTTTAGAAC
qacFH03aqacFH03b	CTGTTTCAATCTTTGGCGAGGTCA		
qnrB01	CGACCTGAGCGGCACTGAATTTA	qnrB01	GCTCGCCAGTCGAAAGTCGAA
QnrS1S3S5	CTGCAAGTTCATTGAACAGGGTGA	QnrS1S3S5	CACCTCGACTTAAGTCTGACTCTTTCAG
qnrS2	ATGCCAGCTTGCGATGGCAAA	qnrS2	GTGGCATAAATTAGCACCCTGTAGGC
QnrVC4VC5VC701	TGGTATCGAGTTCAGAGAATGCGA	QnrVC4VC5VC702	AACCATACAACTCCGAGTGGCTCAA
robA01	TCAAATGCGCGTGCAGTTCTGG	robA01	GTAGCGCTCAATATCCTGACCTTTAC
satG	CCGGGACTTACCGTGGGTTC	satG	CTTCCAGGCATCGGCATCTCA
silE	TCATGGGCAACCGCTGCTCTTTC	silE	GGGCCACTGAAACCGTGAATATCCATG
spcN	TGCTGCTGCGCGAGCA	spcN	CACGTCCAGCAGGCTCCAG
strA	CGGCAATTCCGGGAGTACCG	strA	CCAGTTCTCTTCGGCGTTAGCAA
sul1	GACGAGATTGTGCGGTTCTTCG	sul1	ATTTCGCGAGGGTTTCCGAGA

Table S3.3 Continued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
sul201	TATCGCGCCGGTGCTGGA	sul201sul202	GGACAAGGCGGTTGCGTTTG
sul202	GTATCGCGCCGGTACTGGA		
sulIII	TGATACAACTGAAGTGGGCGTTGTGGA	sulIII	CACGCTTTACACCAGCCTCAACTAAAGC
surA1AB	GCTGGCGATCATAACTATTCTGACTC	surA1AB	CGCTTGTTTAGCGTGTTCCCA
tcmA	CCATCGTGGCCATAGCCAAC	tcmA	CGATCGCCATGTTGAGCTTCTCG
tcr3	ATCGGCGTCGGCATCTTCGG	tcr3	GTCTGGGACAGGCCGATGCC
tet39	TGACTGAGAAGGTTCAGGAGCAATC	tet39	ACCAATTGCATCGCAAGCTATTCC
tet40	CTGCTGGGAAACATAAACCTGCAA	tet40	CCGAACCAAGTGACGCTGTTC
tetA	TCGGCGAGGATCGCTTTCACTG	tetA	TCCTCATCCACCTGCCTGGACA
tetB01	GCCAGTCTTGCCAACGTTATTACG	tetB01	ATCCCTGAAAGCAAACGGCCTAA
tetE	TGATTGCTGGACCAGTCATTGG	tetE	CCATACGAAGCGCTCTTCTCC
tetG	TTCAAGCCGGCTTGGAGAGC	tetG	ACAATCCAAACCCAACCGTTCCA
tetL01	CTGGGTGAACACAGCCTTTATGTTAACC	tetL01	GGCTATCATTCCAACAATCGCTGGAC
tetO	GCAGGGACAGAACTATTAGAGCCATATC	tetO	GCTAACTTGTGGAACATATGCCGAAC
tetQ	TGGATTGAAGACCCGTCTTTGTCC	tetQ	AGCAGGTGTACTTACCGGGCTATA
tetR	TGCTCGACGCCTTAGCCATTG}	tetR	GCGACTTGATGCTCTTGATCTTCCAA
tsnR	GCGGTGCAGCGGATCATCGA	tsnR	CCGTCGAGGACGACGACGTC
vanA	GTTAAGCCGGCGCGTTCAG	vanA	TTTGGTCCACCTCGCCAACA
vanSB01vanSB02	CATAGGGCGCACGGTTGC	vanSB02	TCGCGCCTGATCAATGCG
vanXA01	GTGGACGGCTACCTGGTGAACC	vanXA01	CCCATGTCGGCGAGCTCACC
vph01	GAGTTGTTCCCGCTCATGTCC	vph02	CCGAGGTCGCCATGGACCA
vph02	GCCGTACCTGGTGCTGAGC		
mfpa	CGTGAATCTGGCCGAGTCACAAC	mfpa	CGCAAGTCGGCGTCATCCAG
norA	ACCAGGGATTGGTGGATTTATGGC	norA	CGCCACCCGTAATAGCAATCGA
qacAB01	GGAATTGGAATTGCAGCCATTGGC	qacAB01qacAB02	CGCCCACTACAGATTCTTCAGCTAC

Table S3.3 Continued

Forward Primer Name	Forward Sequence	Reverse Primer Name	Reverse Primer Sequence
qacAB02	AATTGGAACTGCAGCCATTGGC		
qnrB01	CGACCTGAGCGGCACTGAATTTA	qnrB02	TCCCACAGCTCGCATTTCTCCA
qnrB02	AAACGGCAGCTTTATGCTGTGTGA	qnrB01	GCTCGCCAGTCGAAAGTCGAA
qnrD	GACAGGAATAGCTTGGAAGGGTGTG	qnrD	ACGGCGCCAGTTATCACAGTG
qnrVC01	GCTCAAACCTTCGAGATACACAGTTC	qnrVC01	CCTCGAAGATTTGCACCAATCCATC
qnrVC02	TGGTATCGAGTTCAGAGAGTGCGA	qnrVC02	CATACAACTCCGAGTGGCTCAAATCA
QnrVC1VC3VC6	GGGCACTAGAAGGGTGCGA	QnrVC1VC3VC6	CCTCGAAGATTTGCACCAATCCA
QnrVC4VC5VC701	TACAGCTCCGAGTGGCTCAA	QnrVC4VC5VC701	TGGTATCGAGTTCAGAGAATGCGA
QnrVC4VC5VC702	TTGGTATCGAGTTCAGAGAGTGCGA	QnrVC4VC5VC702	AACCATACAACTCCGAGTGGCTCAA

Table S3.4: Sample metadata used in partitioning of variance

ID	Sample	Longitude	Latitude	Design Pop.	Avg. Flow (m^3/day)	Cattle/Division[a]	Land Cover[b]	Sewer Pipe Type
A	Gatineau	-75.6047	45.4744	230000	143708.6	High	Urban	Gravity
B	Papineauville	-75.0217	45.6122	1700	1724.8	Medium	Forest/Cropland	Force
C	Lachute	-74.3672	45.6306	12809	10998.8	Medium	Forest/Cropland	Combined
D	Pincourt	-73.9872	45.3572	13545	7660.9	Low	Semi-Urban/Cropland	Force
E	Rosemère	-73.7694	45.6497	27000	22186.1	Low	Urban	Force
F	La Prairie	-73.5553	45.4153	64430	61073.3	Low	Urban/Cropland	Gravity
G	Verchères	-73.3536	45.7803	3600	4343.0	Low	Semi-Urban/Cropland	Force
H	St-Ours	-73.1511	45.9042	2055	1270.2	Low	Forest/Cropland	Force
I	St-Robert	-73.0066	45.9765	376	113.8	Low	Semi-Urban/Cropland	Gravity
J	Granby	-72.7736	45.3706	38400	56084.1	Medium	Forest/Semi-Urban/Cropland	Gravity
K	Shawinigan	-72.7561	46.5539	24925	19506.8	Low	Forest/Semi-Urban/Cropland	Force
L	St-Eulalie	-72.2556	46.1281	450	392.7	Low	Forest/Cropland	Force
M	St-Casimir	-72.1356	46.6572	1265	808.8	Low	Forest/Cropland	Force
N	Victoriaville	-71.9764	46.0511	34125	33942.6	Medium	Semi-Urban/Cropland	Force
O	Plessisville	-71.7897	46.2578	8000	5609.2	Medium	Forest/Semi-Urban/Cropland	Force
P	Stoneham	-71.3756	46.9775	3780	1288.4	Medium	Urban/Forest	Force

a: Number of cattle in division where low= 0-2001, medium=2001-5871 and high= 5871-543566

b: Dominant land use in the division

Table S3.5: Significance of environmental variables obtained using function envfit permutation test in R Vegan Package

Variable	r^2	Pr (>r)	Significance[1]
Longitude	0.0573	0.682	NS
Latitude	0.1175	0.449	NS
Population Design	0.0453	0.690	NS
Volume/year m3	0.0874	0.489	NS
Sewer Type (Combined)	0.1637	0.287	NS

[1] NS: Not significant

Table S3.6: Variant Sequence Data and NCBI Results

Variant	Sequence	Reported Hosts NCBI Blast		
blaTEM Variant 1	>blaTEM_Variant_1 ATACCAAACGACGAGCGTGAC ACCACGATGCCTGTAGCAATG GCAACAACGTTGCGCAAACTA TTAACTGGCGAACTACTTACT CTAGCTTCCCGGCAACAATTA ATAGACTGGATGGAGGCGGAT AAAGTTGCAGGACCACTTCTG CGCTCGGCCCTTCCGGCTGGC TGGTTTATTGCTGATAAATCT GGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTG GGG	*Achromobacter denitrificans* *Acinetobacter baumannii* *Acinetobacter haemolyticus* *Acinetobacter sp.* *Aeromonas hydrophila* *Alcaligenes sp.* *Achromobacter denitrificans* *Acinetobacter baumannii* *Acinetobacter haemolyticus* *Acinetobacter sp.* *Aeromonas hydrophila* *Alcaligenes sp.* *Arabidopsis thaliana* *Babesia bigemina* *Bacillus cereus* *Bacillus mycoides* *Bacillus subtilis* *Bacillus safensis* *Bacillus sp. BT-B158* *Bacillus subtilis* *Bacillus tropicus* *Bacillus velezebsus* *Bifidobacterium longum* *Burkholderia cepacian* *Burkholderia lata* *Burkholderia sp.*	*Chlamydia trachomatis* *Clostridioides difficile* *Clostridium botulinum* *Cronobacter sakazakii* *Escherichia coli* *Eimeria acervulina* *Eimeria maxima* *Enterobacter cloacae* *Enterococcus faecium* *Escherichia sp.* *Escherichia marmotae* *Faecalibacterium prausnitzii* *Flavobacterium columnare* *Francisella philomiragia* *Helicobacter pylori* *Klebsiella michiganensis* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Legionella pneumophila* *Macrococcus caseolyticus* *Morganella morganii* *Mycolicibacterium smegmatis* *Mycobacterium tuberculosis* *Mycoplasma mycoides* *Neisseria meningitidis*	*Phaffia rhodozyma* *Propionibacterium freudenreichii* *Pseudomonas aeruginosa* *Pseudomonas sp.* *Rhizobium leguminosarum* *Ruthenibacterium lactatiformans* *Saccharomyces cerevisiae* *Salmonella enterica* *Schizosaccharomyces pombe* *Serratia marcescens* *Solanum pennellii* *Staphylococcus aureus* *Staphylococcus hominis* *Staphylococcus saprophyticus* *Streptococcus agalactiae* *Streptococcus lutetiensis* *Streptococcus pneumoniae* *Cloning vectors* *Expression vectors* *Stenotrophomonas sp.* *Streptomyces sp.* *Trypanosoma brucei* *uncultured bacterium* *Vibrio parahaemolyticus* *Vibrio vulnificus*

Variant	Sequence	Reported Hosts NCBI Blast		
blaTEM Variant 2	>blaTEM Variant 2 ATACCAAACGACGAGCGTGAC ACCACGATGCCTGCAGCAATG GCAACAACGTTGCGCAAACTA TTAACTGGCGAACTACTTACT CTAGCTTCCCGGCAACAATTA ATAGACTGGATGGAGGCGGAT AAAGTTGCAGGACCACTTCTG CGCTCGGCCCTTCCGGCTGGC TGGTTTATTGCTGATAAATCT GGAGCCGGTGAGCGTGGGTCT CGCGGTATCATTGCAGCACTG GGG	*Acinetobacter baumannii* *Acinetobacter johnsonii* *Acinetobacter towneri* *Aeromonas caviae* *Aeromonas hydrophila* *Aeromonas media* *Aeromonas sp. ASNIH2* *Aeromonas veronii* *Atlantibacter hermannii* *Bacillus subtilis* *Bacteroides fragilis* *Chlamydia trachomatis* *Chryseobacterium gallinarum* *Citrobacter amalonaticus* *Citrobacter braakii* *Citrobacter farmeri* *Citrobacter freundii* *Citrobacter koseri* *Citrobacter portucalensis* *Citrobacter sp.* *Citrobacter werkmanii* *Citrobacter youngae* *Clostridioides difficile* *Cronobacter sakazakii* *Enterobacter asburiae* *Enterobacter bugandensis* *Enterobacter chengduensis* *Enterobacter cloacae* *Enterobacter hormaechei*	*Enterobacter kobei* *Enterobacter roggenkampii* *Enterococcus faecium* *Enterobacter sp.* *Enterobacteriaceae bacterium* *Escherichia albertii* *Escherichia coli* *Escherichia fergusonii* *Escherichia marmotae* *Haemophilus influenzae* *Haemophilus parainfluenzae* *Hafnia alvei* *Kingella kingae* *Klebsiella aerogenes* *Klebsiella huaxiensis* *Klebsiella michiganensis* *Klebsiella oxytoca* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Klebsiella sp.* *Klebsiella variicola* *Leclercia adecarboxylata* *Leclercia adecarboxylata* *Leclercia sp.* *Morganella morganii*	*Mycobacterium tuberculosis* *Neisseria gonorrhoeae* *Neisseria mucosa* *Proteus mirabilis* *Proteus terrae subsp. cibarius* *Proteus vulgaris* *Providencia huaxiensis* *Providencia rettgeri* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Raoultella ornithinolytica* *Raoultella planticola* *Salmonella enterica* *Salmonella sp.* *Serratia liquefaciens* *Serratia marcescens* *Serratia sp.* *Shigella boydii* *Shigella flexneri* *Shigella sonnei* *Shigella sp.* *Staphylococcus aureus* *Streptococcus pneumoniae* *Synthetic construct* *Uncultured bacterium* *Vibrio cholerae* *Vibrio parahaemolyticus* *Vibrio vulnificus* *Yersinia pseudotuberculosis*

Variant	Sequence	Reported Hosts NCBI Blast
blaTEM Variant 3	>blaTEM Variant 3 ATACCAAACGACGAGCGTGACA CCACGATGCCTGCAGCAATGGC AACAACGTTGCGCAAACTATTA ACTGGCGAACTACTTACTCTAG CTTCCCGGCAACAATTAATAGA CTGGATGGAGGCGGATAAAGTT GCAGGACCACTTCTGCGCTCGG CCCTTCCGGCTGGCTGGTTTAT TGCTGATAAATCTGGTGCCGGT GAGCGTGGGTCTCGCGGTATCA TTGCAGCACTGGGG	Synthetic construct
pmrC Variant1	>pmrC Variant 1 TTGATGTACTCTTCAAGCCCGT GGAACAGCACTTCGTCATAGCA TTCGCCGTTGATGCACTGATCA GGTAGATTCAGCGCGGTGACGT TCTGGTGAGGCACGCGGTCGCA GGCACCTTTACAGCCGCCATCG TTGTCATTCCACAGCACGTTGA TGCCCGCTCGCTGAATGATATC CAGCACGCCTTCCTGGTGCTGT GCCAGCTCTTCTTTGTAGTGCT CACGCGGCATATCCGAGAA	*Escherichia coli* *Escherichia sp.* *Salmonella sp.* *Shigella boydii* *Shigella flexneri* *Shigella sonnei* *Synthetic construct*
pmrC Variant2	>pmrC Variant 2 TTGATGTACTCTTCCAGCCCGT GGAACAGTACTTCGTCATAGCA TTCGCCGTTGATGCACTGACCA GGCAGATTCAGCGCGGTGACGT TCTGGTGAGGCACGCGGTCGCA GGCACCTTTACAGCCGCCATCG TTGTCATTCCACAGCACGTTGA TGCCCGCTCGCTGAATGATATC CAGCACGCCTTCCTGGTGCTGT GCCAGCTCTTCTTTGTAATGCT CACGCGGCATATCCGAGAA	*Escherichia coli* *Salmonella sp.* *Shigella flexneri* *Shigella sonnei* *Shigella sp.*

Variant	Sequence	Reported Hosts NCBI Blast		
pmrC Variant 3	>pmrC Variant 3 AAACGACGGATTGTTGTAATGGTT GCCAATACGCGGATCGAGACAAAC GCCTTCGATAATCTGGCGAGTATT CAGACCTAAACTTTCTGCATAGCT ATCCAGTTCATTAAAGTACGCTAC GCGCATCGCCAGATAGGTATTAGC GAAAAGTTTTATCGCTTCTGCTTC AGTGGAGTCGGTAAACAGGGTCGG GATATTTTGCTTAATCGCCCCTTC CTGTAACAACGCAGCAAAACGTTC AGCGC	*Escherichia coli*		
pmrC Variant 4	>pmrC Variant 4 TTGATGTACTCTTCCAGCCCGTGG AACAGCACTTCGTCATAGCATTCG CCGTTGATGCACTGACCAGGCAGG TTCAGCGCGGTGACGTTCTGGTGA GGTACGCGATCGCAAACGCCTTTA CAGCCGCCATCGTTGTCTTTCCAC AGCACGTTGATGCCCGCTCGCTGA ATGATATCCAGCACGCCTTCCTGG TGCTGTGCCAGCTCTTCTTTGTAG TGCTCACGCGGCATATCCGAGAA	*Escherichia coli*		
blaOXA Variant 1	>blaOXA Variant 1 GCAATGGGAAAGAGACTTGACCTT AAGAGGGGCAATACAAGTTTCAGC TGTTCCCGTATTTCAACAAATCGC CAGAGAAGTTGGCGAAGTAAGAAT GCAGAAATACCTTAAAAAATTTTC CTATGGCAACCAGAATATCAGTGG TGGCATTGACAAATTCTGGTTGGA AGGCCAGCTTAGAATTTCCGCAGT TAATCAAGTGGAGTTTCTAGAGTC TCTATATTTAAATAAATTGTCAGC ATCTAAAGAAAACCAGCTAATAGT AAAAGAGGCTTTGGTAACGGAGGC GGCACCTGAATATCTAGTGCATTC AAAAACTGGTTTTTCTGGTGTGGG AACTGAG	*Acinetobacter baumannii* *Acinetobacter johnsonii* *Aeromonas caviae* *Aeromonas hydrophila* *Aeromonas media* *Aeromonas salmonicida* *Aeromonas simiae* *Aeromonas sp.* *Aeromonas veronii* *Citrobacter braakii* *Citrobacter freundii* *Citrobacter sedlakii* *Citrobacter werkmanii* *Enterobacter aerogenes* *Enterobacter asburiae* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei*	*Escherichia coli* *Escherichia fergusonii* *Gallibacterium anatis* *Klebsiella grimontii* *Klebsiella michiganensis* *Klebsiella oxytoca* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Morganella morganii* *Pantoea agglomerans* *Proteus mirabilis* *Proteus sp.* *Proteus vulgaris* *Providencia heimbachae* *Providencia rettgeri* *Providencia sp.* *Providencia stuartii* *Pseudocitrobacter faecalis*	*Pseudomonas aeruginosa* *Pseudomonas fulva* *Pseudomonas monteilii* *Pseudomonas putida* *Pseudomonas shirazica* *Pseudomonas sp.* *Salmonella enterica* *Salmonella sp.* *Serratia marcescens* *Shewanella putrefaciens* *Shigella flexneri* *Shigella sonnei* *Stenotrophomonas maltophilia* *Vibrio cholerae* *Vibrio fluvialis* *Vibrio parahaemolyticus* *Yokenella regensburgei*

Variant	Sequence	Reported Hosts NCBI Blast		
blaOXA_ Variant2	>blaOXA Variant 2 ACAATGGGAAAGAGACTTGAGCTT AAGAGGGGCAATACAAGTTTCAGC GGTTCCCGTATTTCAACAAATCGC CAGAGAAGTTGGCGAAGTAAGAAT GCAGAAATATCTTAAAAAATTTTC ATATGGTAACCAGAATATCAGTGG TGGCATTGACAAATTCTGGTTGGA GGGTCAGCTTAGAATTTCCGCAGT TAATCAAGTGGAGTTTCTAGAGTC TCTATTTTTAAATAAATTGTCAGC ATCAAAAGAAAATCAGCTAATAGT AAAAGAGGCTTTGGTAACGGAGGC TGCGCCTGAATATCTTGTGCATTC AAAAACTGGTTTTTCTGGTGTGGG AACTGAG	*Aeromonas caviae* *Aeromonas hydrophila* *Alcaligenes faecalis* *Citrobacter freundii* *Klebsiella oxytoca* *Klebsiella michiganensis* *Proteus mirabilis* *Pseudomonas aeruginosa*		
blaOXA_ Variant3	>blaOXA Variant 3 GAACGCCAAGCGGATCGTGCCATG TTGGTTTTTGATCCTGTGCGATCG AAGAAACGCTACTCGCCTGCATCG ACATTCAAGATACCTCATACACTT TTTGCACTTGATGCAGGCGCTGTT CGTGATGAGTTCCAGATTTTTCGA TGGGACGGCGTTAACAGGGGCTTT GCAGGCCACAATCAAGACCAAGAT TTGCGATC	*Achromobacter denitrificans* *Acientobacter baumannii* *Aeromonas caviae* *Aeromonas sp.* *Alcaligenes faecalis* *Bordetella bronchiseptica* *Citrobacter freundii* *Corynebacterium asperum* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei*	*Escherichia coli* *Enterobacter aerogenes* *Klebsiella oxytoca* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Klebsiella sp.* *Mannheimia haemolytica* *Morganella morganii* *Pasteurella multocida* *Phytobacter ursingii* *Providencia rettgeri*	*Pseudomonas aeruginosa* *Pseudomonas baetica* *Pseudomonas monteilii* *Pseudomonas putida* *Salmonella enterica* *Salmonella typhimurium* *Serratia marcescens* *Staphylococcus aureus* *Stenotrophomonas maltophilia* *Uncultured bacterium* *Vibrio anguillarum* *Vibrio cholerae*
blaOXA_ Variant4	>blaOXA Variant 4 GAACGCCAAGCGGATCGTGCCATG TTGGTTTTTGATCCTGTGCGATCG AAGAAACGCTACTCGCCTGCATCG ACATTCAAGATACCTCATACACTT TTTGCACTTGATGCAGGCGCTGTT CGTGATGAGTTCCAGATTTTTCGA TGGGACGGCGTTAACAGGGGCTTT GCAGGCCACAATCAAGACCAAGAT TTT	No 100% similarity observed		

Variant	Sequence	Reported Hosts NCBI Blast
mdtH Variant1	>mdtH Variant 1 TTGAAGCGTGTCTGTCGTTAACGT TGCTCTACCCTATCGCCCGCTGGA GTGAAAAGCATTTTCGTCTGGAAC ACCGGTTG ATGGCTGGGCTGTTGATAATGTCA TTAAGCATGATGCCGGTGGGCATG GTCAGCGGCCTGCAACAACTTTTC ACCCTGAT TTGTCTGTTTTATATCGGGTCGAT CATTGCCGAGCCTGCGCGTGAAAC CTTAAGTGCTTCGCTGGCGGACGC AAGAGCTC GCGGCAGCTATATGGGGTTTAGC	*Escherichia coli* *Escherichia fergusonii* *Escherichia sp.* *Shigella boydii* *Shigella dysenteriae* *Shigella felxneri* *Shigella sonnei* *Synthetic Escherichia coli*
mdtH Variant2	>mdtH Variant 2 TTGAAGCGTGTCTGTCGTTAACGT TGCTCTATCCTATCGCCCGCTGGA GTGAAAAGCATTTTCGTCTGGAAC ACCGGTTGATGGCTGGGCTGTTGA TAATGTCATTAAGCATGATGCCGG TGGGCATGGTCAGCGGCCTGCAAC AACTTTTCACCCTGATTTGTCTGT TTTATATCGGGTCGATCATTGCCG AGCCTGCGCGTGAAACCTTAAGTG CTTCGCTGGCGGACGCAAGAGCTC GCGGCAGCTATATGGGGTTTAGC	*Escherichia coli* *Shigella boydii* *Shigella flexneri* *Shigella sp.*
mdtH Variant3	>mdtH Variant 3 TTGAAGCGTGTCTGTCGTTAACGT TGCTCTACCCTATCGCCCGCTGGA GTGAAAAGCATTTTCGTCTGGAAC ACCGGTTAATGGCTGGGCTGTTGA TAATGTCATTAAGCATGATGCCAG TAGGCATGGTCAGCGGCCTGCAAC AACTTTTCACCCTGATTTGTCTGT TTTATATCGGGTCGATCATTGCCG AGCCTGCGCGTGAAACCTTAAGTG CTTCGCTGGCAGACGCAAGAGCTC GCGGCAGCTATATGGGGTTTAGC	*Enterobacter hormaechei* *Escherichia coli* *Shigella flexneri*

Variant	Sequence	Reported Hosts NCBI Blast
mdtH Variant4	>mdtH Variant 4 TTGAAGCGTGTCTGTCGTTAAC GTTGCTCTATCCTATCGCCCGC TGGAGTGAAAAGCATTTTCGTC TGGAACATCGGTTGATGGCTGG GCTGTTGATAATGTCATTAAGC ATGATGCCAGTGGGCATGGTCA GCGGCCTGCAACAACTTTTCAC CCTGATTTGTCTGTTTTATATC GGGTCGATCATTGCCGAGCCTG CGCGTGAAACCTTAAGTGCTTC GCTGGCAGACGCAAGAGCTCGC GGCAGCTATATGGGGTTTAGC	*Escherichia coli* *Shigella dysenteriae*

3.5 References

1. United Nations Environment Programme (UNEP). Environmental Dimensions of Antimicrobial Resistance: Summary for Policymakers. 2022. Available from:https://wedocs.unep.org/bitstream/handle/20.500.11822/38373/antimicrobial_R.pdf. [Accessed on 2022 May 27].

2. Skandalis N, Maeusli M, Papafotis D, Miller S, Lee B, Theologidis I, et al. Environmental spread of antibiotic resistance. Antibiotics. 2021;10(6):640.

3. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, et al. Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother. 2009;53(12):5046–54.

4. Islam MA, Islam M, Hasan R, Hossain MI, Nabi A, Rahman M, et al. Environmental spread of New Delhi metallo-β-Lactamase-1-producing multidrug-resistant bacteria in Dhaka, Bangladesh. Appl Environ Microbiol. 2017;83(15):e00793-17.

5. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8(9):761–5.

6. Li L-G, Yin X, Zhang T. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. Microbiome. 2018;6:93.

7. Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An Omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun. 2021;12:4765.

8. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021 Jan 18;22(1):178–93.

9. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Reports 2018 81. 2018 Apr 12;8(1):1–11.

10. Jankowski P, Gan J, Le T, McKennitt M, Garcia A, Yanaç K, et al. Metagenomic community composition and resistome analysis in a full-scale cold climate wastewater treatment plant. Environ Microbiomes. 2022 Dec 1;17(1):1–20.

11. Smith SD, Choi J, Ricker N, Yang F, Hinsa-Leasure S, Soupir ML, et al. Diversity of Antibiotic Resistance Genes and Transfer Elements-Quantitative Monitoring (DARTE-QM): a method for detection of antimicrobial resistance in environmental samples. Commun Biol. 2022;5:216.

12. Government of Quebec. Municipal pressures - Wastewater discharges - Wastewater treatment plants - Wastewater discharges - Quebec data [Internet]. 2022 [cited 2022 Apr 20]. Available from: https://www.donneesquebec.ca/recherche/dataset/pressions-municipales-rejets-d-eaux-usees/resource/ae0509d7-466f-4bfc-90aa-b4d4fd82018a?inner_span=True

13. Government of Quebec. Wastewater treatment plants - Wastewater discharges [Internet]. 2021. 2021 [cited 2022 Sep 20]. Available from: https://www.donneesquebec.ca/recherche/dataset/pressions-municipales-rejets-d-eaux-usees/resource/ae0509d7-466f-4bfc-90aa-b4d4fd82018a?inner_span=True

14. Government of Canada. Open Database of Healthcare Facilities [Internet]. 2020-04-17. 2020 [cited 2022 Sep 20]. Available from: https://open.canada.ca/data/en/dataset/543fe07a-fd79-40e9-a829-ccd697526765

15. Government of Canada. 2015 Land Cover of Canada [Internet]. 2019 [cited 2022 Sep 20]. Available from: https://open.canada.ca/data/en/dataset/4e615eae-b90c-420b-adee-2ca35896caf6

16. Commission for Environmental Cooperation. Land cover, 2015 (Landsat and RapidEye, 30m) [Internet]. 2022 [cited 2022 Sep 20]. Available from: http://www.cec.org/fr/atlas-environnemental-de-lamerique-du-nord/couverture-terrestre-2015-landsat-et-rapideye-30m/

17. Statistics Canada. 2016 Agricultural Census [Internet]. Governement of Canada. 2018 [cited 2020 Sep 20]. Available from: https://open.canada.ca/data/en/dataset/f8c4f16d-ec09-4d23-b2f9-99ade21762d0

18. Apprill A, Mcnally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75:129–37.

19. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2015;18(5):1403–14.

20. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

21. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. Vegan: Community Ecology Package. 2022.

22. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–7.

23. McIlroy SJ, Kirkegaard RH, McIlroy B, Nierychlo M, Munk Kristensen J, Karst SM, et al. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017;1–9.

24. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data version 3.1.3 from CRAN. 2022.

25. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.

26. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.

27. Pärnänen K, Karkman A, Tamminen M, Lyra C, Hultman J, Paulin L, et al. Evaluating the mobility potential of antibiotic resistance genes in environmental resistomes without metagenomics. Sci Rep. 2016;6(35790).

28. NCBI Research Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2013;41:D8–20.

29. Dray S, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, et al. Package "adespatial." CRAN; 2018.

30. Numberger D, Ganzert L, Zoccarato L, Mühldorfer K, Sauer S, Grossart HP, et al. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. Sci Rep. 2019;9:9673.

31. Cai L, Ju F, Zhang T. Tracking human sewage microbiome in a municipal wastewater treatment plant. Appl Microbiol Biotechnol. 2014;98:3317–26.

32. Mclellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. Diversity and population structure of sewage derived microorganisms in wastewater treatment plant influent. Environ Microbiol. 2010;12(2):378–92.

33. Rice LB. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. J Infect Dis. 2008;197:1079–81.

34. Willems A, Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F. The Family Comamonadaceae: The Prokaryotes. The Prokaryotes: Alphaproteobacteria and Betaproteobacteria. Berlin, Heidelberg: Springer; 2014. 777–851 p.

35. VandeWalle JL, Goetz GW, Huse SM, Morrison HG, Sogin ML, Hoffmann RG, et al. Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. Environ Microbiol. 2012;14(9):2538–52.

36. Shchegolkova NM, Krasnov GS, Belova AA, Dmitriev AA, Kharitonov SL, Klimina KM, et al. Microbial community structure of activated sludge in treatment plants with different wastewater compositions. Front Microbiol. 2016;7:90.

37. Wang P, Yu Z, Qi R, Zhang H. Detailed comparison of bacterial communities during seasonal sludge bulking in a municipal wastewater treatment plant. Water Res. 2016;105:157–66.

38. Fisher JC, Levican A, Figueras MJ, McLellan SL. Population dynamics and ecology of Arcobacter in sewage. Front Microbiol. 2014;5:525.

39. Ghaju Shrestha R, Tanaka Y, Haramoto E. A Review on the prevalence of Arcobacter in aquatic environments. Water. 2022;14(8):1266.

40. Brückner V, Fiebiger U, Ignatius R, Friesen J, Eisenblätter M, Höck M, et al. Prevalence and antimicrobial susceptibility of Arcobacter species in human stool samples derived from out- and inpatients: the prospective German Arcobacter prevalence study Arcopath. Gut Pathog. 2020;12:21.

41. Skwor T, Stringer S, Haggerty J, Johnson J, Duhr S, Johnson M, et al. Prevalence of potentially pathogenic antibiotic-resistant aeromonas spp. in treated urban wastewater effluents versus recipient riverine populations: A 3-year comparative study. Appl Environ Microbiol. 2020;86(3):e02053-19.

42. Madden T. Chapter 16 - The BLAST Sequence Analysis Tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook . 2nd ed. National Center for Biotechnology Information (US); 2002.

43. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: A resource of complete bacterial plasmids. Nucleic Acids Res. 2019;47:D195-202.

44. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. NTO CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020 Jan 1;48(D1):D517–25.

45. Kotsakis SD, Flach CF, Razavi M, Larsson DGJ. Characterization of the first OXA-10 natural variant with increased carbapenemase activity. Antimicrob Agents Chemother. 2019;63(1):e01817-18.

46. Danel F, Hall LM, Duke B, Gur D, Livermore DM. OXA-17, a further extended-spectrum variant of OXA-10 beta-lactamase, isolated from Pseudomonas aeruginosa. Antimicrob Agents Chemother. 1999;43(6):1362–6.

47. Porto A, Ayala J, Gutkind G, Di Conza J. A novel OXA-10–like β-lactamase is present in different Enterobacteriaceae. Diagn Microbiol Infect Dis. 2010;66(2):228–9.

48. Bajpai T, Pandey M, Varma M, Bhatambare GS. Prevalence of TEM, SHV, and CTX-M Beta-Lactamase genes in the urinary isolates of a tertiary care hospital. Avicenna J Med. 2017;7(1):12–6.

49. Law A, Solano O, Brown CJ, Hunter SS, Fagnan M, Top EM, et al. Biosolids as a source of antibiotic resistance plasmids for commensal and pathogenic bacteria. Front Microbiol. 2021;12:606409.

CHAPTER 4

Activated Sludge Microbial Community Assembly: The Role of Influent Microbial Community Immigration

Connecting text: In Chapter 3, diverse microbial communities and multiple antimicrobial resistance genes were detected in the influent wastewater samples from Quebec, however little is known about their fate upon entry into the wastewater treatment plant (WWTP). Biological WWTPs are host to diverse microbial communities. Whilst they receive a constant influx of microbes from influent wastewater, the importance of immigrants in the microbial community assembly of the activated sludge (AS) remains unclear. To enable optimisation of the wastewater treatment process, we first need a better understanding of how the activated sludge microbial community is assembled and the impact of the sewer microbial community.

In Chapter 4, controlled manipulative experiments were utilised that decoupled the influent wastewater substrate composition from its microbial community structure to reveal the fundamental mechanisms involved in immigration between sewers and activated sludge. By developing a protocol to reproducibly grow AS-WWTP communities in the lab and establishing logical independent population definitions, the impact of influent immigration could be studied with appropriate reproducibility and controls for the first time.

4.1 Abstract

Wastewater treatment plants (WWTPs) are host to diverse microbial communities and receive a constant influx of microbes from influent wastewater, however the impact of immigrants on the structure and activities of the activated sludge (AS) microbial community remains unclear. To gain insight on this phenomenon known as perpetual community coalescence, the current study utilised controlled manipulative experiments that decoupled the influent wastewater composition from the microbial populations to reveal the fundamental mechanisms involved in immigration between sewers and AS-WWTP. The immigration dynamics of heterotrophs were analysed by harvesting wastewater biomass solids from 3 different sewer systems and adding to synthetic wastewater. Immigrating influent populations were observed to contribute up to 25 % of the sequencing reads in the AS. By modelling the net growth rate of taxa, it was revealed that immigrants primarily exhibited low or negative net growth rates. By developing a protocol to reproducibly grow AS-WWTP communities in the lab, we have laid down the foundational principals for the testing of operational factors creating community variations with low noise and appropriate replication. Understanding the processes that drive microbial community diversity and assembly is a key question in microbial ecology. In the future, this knowledge can be used to manipulate the structure of microbial communities and improve system performance in WWTPs.

4.2 Importance

In biological wastewater treatment processes, the microbial community composition is essential in the performance and stability of the system. To allow future process optimisation to meet new treatment goals, we need a better understanding of factors influencing the microbial community assembly in WWTPs. This study developed a

reproducible protocol to investigates the impact of influent immigration (or perpetual coalescence of the sewer and activated sludge communities) with appropriate reproducibility and controls. We demonstrate herein that influent immigration contributed up to 25 % of the sequencing reads in the activated sludge under the studied conditions, highlighting the need to consider this process in future WWTP modelling and design.

4.3 Introduction

In activated sludge wastewater treatment plants (AS-WWTPs), the microbial community composition is intricately related to the performance and stability of the system. However, a comprehensive understanding of the structuring process of the community remains elusive. During the activated sludge (AS) treatment process, a diverse heterotrophic microbial community is grown in the plant to become highly effective at degrading the organic pollutants contained in the incoming wastewater. In recent years, efforts have been made to modify these processes to improve performance and meet new objectives for wastewater management. Modifications often require control over specific functional populations within the diverse AS microbial community. For example, enhanced phosphate removal processes require the growth of Polyphosphate Accumulating Organisms (e.g., members of the genera *Candidatus* Accumulibacter and *Tetrasphaera*) (1), whilst to curtail operational problems a reduction in the abundance of detrimental bulking and foaming bacteria (e.g., genera *Gordonia* and *Thauera*) is desired (2). With a limited knowledge of community assembly mechanisms, modification and optimization of the AS process is challenging. A greater understanding of the underlying microbial mechanisms involved in the community assembly and maintenance is crucial to benefit the engineering and function of these systems.

Previous focus on microbial community assembly has been placed on the impact of temporal differences (3), wastewater characteristics (4), operational conditions (5,6) and geographical location (7) on the AS community. However, the role of the diverse and continuously immigrating influent microbial community is poorly understood. This is a special case of community coalescence with a perpetual occurrence which likely drive in part the community dynamics (8).

The community in municipal wastewater originates from numerous environments including human faeces, sewer biofilm and sediments, and soil runoffs (3). Studies have produced conflicting results on the relative importance of influent immigration on the assembly of the AS community. Some suggested that the overall effect of the influent community is negligible, as they observed that the AS microbial community remains stable over time despite changes in the composition of the influent wastewater community (4). The activity of immigrating taxa has also been questioned, with some studies showing these populations often have low or negative net growth rates (9,10).

Other literature reports argued influent immigration to be an important process in microbial community assembly. In lab-scale AS reactors operated at extremely low temperatures and solids retention times, immigration from sewers was found to be essential to maintain complete nitrification (11). The occurrence of shared operational taxonomic units (OTUs) between influent and AS communities also infers high immigration rates in full-scale AS-WWTPs (9,12). Furthermore, a recent study of 11 AS treatment plants in Denmark, selected based on their highly similar process design, identified the AS community composition to be strongly influenced by the influent wastewater community composition (13). However, it remains difficult to ascertain if the influent community composition is a factor independent from the substrate composition of the wastewater, which precludes strong conclusions.

With contrasting reports on the impact of immigration, it is challenging to understand the relative importance of this process. This variability can be somewhat explained by the limited methods used to study the immigration process, which lack reproducibility and assessment of the activity of immigrants. As reviewed by Mei & Lie (2019), commonly used approaches include counting shared microbial species between ecosystems (i.e., the influent and AS-WWTP), microbial source tracking and neutral community modelling (14). However, these approaches only imply a contribution from the immigrating source and provide little information about the fate and activity of these organisms. Furthermore, neutral modelling approaches to investigate immigration (such as the neutral theory of biodiversity) assume that organisms are functionally equivalent (15), this assumption is likely inaccurate considering the variety of substrates available to the heterotrophic populations (16) and the niche differences between the influent and AS. It has been proposed that immigrants should be classified as either rare diffusive immigrants, or time continuous high-rate mass flow immigrants (17). Heterotrophic mass flow immigrants appear to be heavily influenced by deterministic selection, suggesting that they should be divided into relevant functional guilds when assessing their assembly mechanisms. Consequently, we cannot rely on mathematical modelling approaches alone to resolve the relative importance of selection, immigration, and drifts in the assembly of the community. Furthermore, full-scale systems display too much variability in terms of wastewater composition, operation and community to allow appropriate reproducibility and provide an accurate assessment of the mechanisms at play during immigration.

Given the complex behaviour of microbial communities within wastewater, reproducible and highly controlled systems are required to accurately investigate and quantify the impact of immigration into AS-WWTPs and determine the fate of

immigrating bacteria. Because the presence of specialized populations in the influent wastewater is often correlated with the presence of specific substrates (18; Chapter 7), it is impossible to understand immigration independently from the wastewater substrate compositions in full-scale systems. Therefore, here we propose the development of a system where the substrate landscape and the immigrating community can be manipulated independently using highly controlled laboratory-scale reactors, to investigate fundamental questions on the quantitative impact of immigration and activity. To address the limitations observed in past immigration studies, this work aimed to decouple the influent substrates and microbial community, and specifically investigate the impact of immigration with appropriate reproducibility and controls. Using a series of highly controlled reactors (a total of 72 reactors were operated for 17-25 weeks), we manipulated the influent microbial community independently to identify the effect of constant immigration on the AS microbial community. The laboratory-scale AS reactors were inoculated with mixed liquors from three different AS-WWTPs, representing three of the most different municipal AS microbial communities among the AS-WWTPs located within 100 km of Montréal (Québec, Canada). The purpose of this was to determine whether the impact of immigration was dependent on the established community within the reactors. This also evaluated the reactor protocol for future use with different but similar starting communities. Finally, dynamics in the AS microbial community compositions were analysed by 16S rRNA gene amplicon sequencing. The long-term goal of the current work is to provide a reproducible protocol to conduct studies on the fate of immigrants and their activities in the AS mixed liquor microbial communities, and to develop an appropriate method to study this aspect of biological wastewater treatment systems.

4.4 Materials and Methods

4.4.1 Experimental Design and Reactor Setup

In total, 3 ***sets*** of reactors were operated (Set A, B and C), which differed based upon the source of influent solids used. Figure 4.1b shows a basic schematic of *one* set of reactors operated. For the first set of reactors (Set A), the reactors received La Prairie (Québec, Canada) influent solids and were operated from May until September 2016. For the second and third sets (Set B and Set C), the test reactors received Cowansville and Pincourt (Québec, Canada) influent solids, respectively, and were operated from May until November 2017. To investigate the impact of the starting community, each ***set*** of reactors were divided into 3 ***blocks*** (Block a, b and c), the reactors in each block were inoculated with mixed liquor obtained from either La Prairie, Cowansville or Pincourt AS-WWTPs, respectively (Figure 4.1b).

The three wastewater treatment plants were selected as a source of influent solids and mixed liquor inocula from among 8 AS-WWTPs within a diameter of approximately 100 km from Montréal (Québec, Canada) based on the results of a previous study. This study revealed that the compositions of the activated sludge mixed liquor microbial communities at the three selected plants were at the extremes of the composition distribution when visualized using a principal coordinate analysis plot in a previous study (19). In other words, the mixed liquor communities showed the greatest differences among the sites tested.

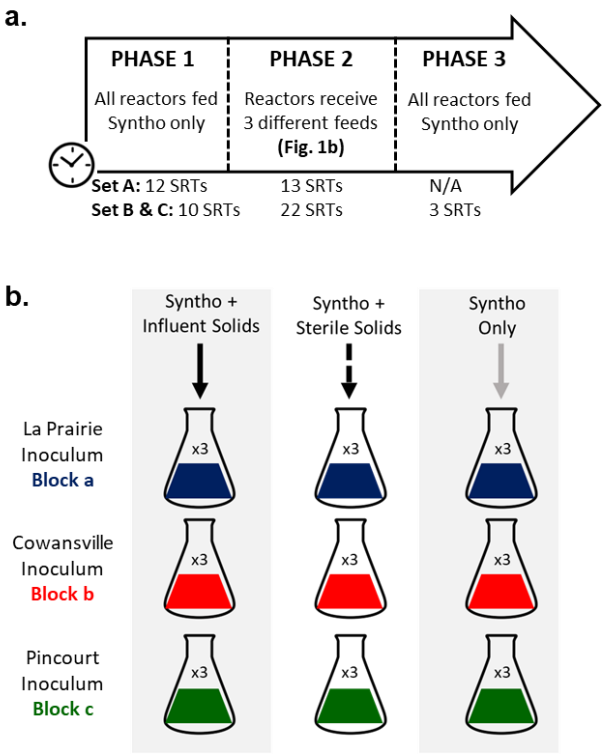


Figure 4.1- Experimental set up. (a) Schematic of one set of reactors (i.e., receiving one source of influent solids) in Phase 2. Three sets of reactors were operated in total. Set A received La Prairie influent solids; Set B, Cowansville influent solids; Set C, Pincourt influent solids. The block describes the inoculum mixed liquor source. (b) Timeline of reactor operation. Reactors were operated in three phases with feed altered.

The reactors were operated in three phases (Figure 4.1a), with the feeds altered to investigate the effect of immigration. During Phase 1, all reactors received Syntho (see section 4.4.4*)* only to develop a steady-state core microbial community adapted to the specific substrate composition. At the end of Phase 1, the mixed liquors of all 9 reactors that had received the same inoculum (i.e., 1 block of reactors; Figure 4.1b) were combined and split again to ensure a homogeneous stable community in each reactor as described by Kaewpipat and Grady, 2002 (20). During Phase 2, the 3 test reactors in each inoculum block received Syntho supplemented with influent solids (resulting in 9 test reactors per set) to simulate immigration into the AS. In addition, nine reactors (3 per block) were defined as substrate controls and received Syntho supplemented with

sterile influent solids (autoclaved at 121°C, 15 psi for 30 minutes). The final 9 reactors (3 per block) received Syntho only to act as a continuity control (9 per reactor set). In total, 72 reactors were operated comprising of 3 blocks of inoculum and 3 sources of influent solids, operated in triplicate resulting in 27 reactors per influent solids used. In 2017, there was only one set of Syntho only controls used because Sets B (Cowansville influent solids) and Set C (Pincourt influent solids) were operated simultaneously and their Syntho only control reactors were equivalent (which allowed the economy of 9 reactors). A detailed overview of the reactor experimental design is outlined in Table S4.2.

4.4.2 Reactor Operation

The reactors were inoculated with 90 mL of mixed liquor (2.7 g-VSS/L) obtained from La Prairie (block a), Cowansville (block b) and Pincourt (block c) AS-WWTPs. The AS mixed liquor (ML) samples from the 3 AS-WWTPs were collected within a period of 4 hours (block a; May 2016, block b and c; May 2017). Samples were transported on ice, stored at 4 ℃ and processed within 24 hours of collection. All reactors were incubated at 21 °C with shaking at 180 rpm to allow aeration and gentle mixing. The average reactor hydraulic retention time (HRT) was 1.8 days, and the solid retention time (SRT; the average time solids were maintained within the reactors) was 5 days. The HRT and SRT were controlled through daily feeding and wasting. Three times a week, the total reactor volume was settled in a 100-mL graduated cylinder for 45 min to allow solid-liquid separation to mimic the clarifier of an activated sludge wastewater treatment process, and the appropriate volume of the supernatant was removed.

4.4.3 Influent Solid Processing for Phase 2 Feed

Influent wastewater solids were concentrated at the AS-WWTP by settling influent wastewater for 15 min in 20-L buckets. Approximately 10-L of concentrated influent wastewater was transported to the lab on ice and processed immediately. To separate the influent wastewater and solid fractions, the wastewater was centrifuged in 50 mL aliquots at 21,100 × g for 10 min (Thermofisher Scientific model ST16R Centrifuge). The supernatant was removed, and the solids were re-suspended in Synthetic wastewater (Syntho; defined in section 4.4.4) to wash. The centrifugation and wash step were repeated three times, and the influent solids were resuspended in a final ~800 mL of synthetic wastewater. The volatile suspended solids (VSS) concentration of the stock influent solid solution was measured using the Standard Method 2540E (21) before dilution. The concentrated influent solids stock was stored at 4 ℃ for no more than two weeks before use.

4.4.4 Feed Preparation

A stock solution of influent solids was prepared as described in section *4.4.3*. Synthetic wastewater 'Syntho' was produced using a modified recipe as described by Boeije et al., 1999 (22), and autoclaved in 500 mL containers at 121 ℃ and 15 psi for 30 minutes. The 'Syntho + Influent Solids' feed was prepared by diluting the stock influent solid solution in Syntho to a concentration of 120 mg-VSS/L, which represents the average VSS concentration of influent wastewater reported in the three wastewater treatment plants studied (23). The 'Syntho + autoclaved solids' feed was prepared in the same way with the addition of autoclaving at 121 ℃/15 psi for 1 hour to sterilise.

4.4.5 Sample Analysis

Mixed liquor reactor samples were collected weekly for analysis of total suspended solids (TSS) and volatile suspended solids (VSS) using Standard Methods 2540B and 2540E respectively (21). Effluent wastewater (the supernatant after settling) samples were collected for the analysis of chemical oxygen demand (COD) using Standard Methods 5220D (21). Biomass samples for DNA extraction were centrifuged and stored ready for nucleic acid extraction at −80 ℃ until further use.

4.4.6 16S rRNA Gene Amplicon Sequencing

DNA was extracted from stored biomass samples using DNeasy PowerSoil Kit (Qiagen, Germantown, MD, USA) as per manufacturers instructions with a centrifugation speed of 10,000 x g. PCR of the 16S rRNA gene V4 region was conducted using 515F (5'- GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3') primers (24,25). PCR conditions were as follows: 94 ℃ for 3 min followed by 35 cycles of 94 ℃ for 45 sec, 50 ℃ for 60 sec, 72 ℃ for 90 sec. 72 ℃ for 10 min and a final hold at 4 ℃. Unique Uniprimer barcodes were added to the amplicons from each sample in a second PCR reaction to allow sample pooling. Reaction conditions were as follows; 94 ℃ for 3 min followed by 15 cycles of 94 ℃ for 30 sec, 59 ℃ for 20 sec, 68 ℃ for 45 sec. 68 ℃ for 5 min and a final hold at 4 ℃. Barcoded amplicons were purified using QIAquick PCR Purification Kit (Qiagen, Germantown, MD, USA) and pooled at equimolar concentrations. Amplicons were sequenced on the Illumina MiSeq PE250 platform at McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada).

4.4.7 Bioinformatic and Statistical Analyses

Sequencing data was analysed using QIIME 2 (26) and R Software (packages "vegan" and "ape"). Specifically, the raw sequences were denoised and errors corrected using DADA2 (27) in Qiime2 pipelines. Sequences were rarefied to 40,000 reads in R. ASV tables were exported directly from Qiime2 after quality filtering. Taxonomy was assigned using MiDAS 2.0 reference database (28) and tabulated. Microbial community diversity was analysed using the "vegan" package of R (29) and Jaccard dissimilarity.

Analysis of Similarities (ANOSIM) was conducted using the using the "vegan" package of R (30) to statistically test the significance of differences between groups. The ANOSIM test compared the mean of ranked dissimilarities between groups with the mean of ranked dissimilarity within groups. An ANOSIM R value close to '1' indicates dissimilarity between groups, whilst a value close to '0' indicated that the similarities between groups and within groups were on average the same (30).

An odds ratio test was used to assess whether the growth rate of immigrants impacted their fate during Phase 3 of reactor operation. Values were calculated by dividing the odds of the first group (genera with a positive net growth rate persisting at the end of Phase 3) by the odds of the second group (genera with a negative net growth rate persisting at the end of Phase 3).

4.4.8 Definitions of Population Categories

The ***core resident populations*** were present in the majority of reactors and followed the criteria outlined by Wu et. al 2019 for core communities (7). Taxa were filtered based on the overall abundant genera, by calculating the mean relative abundance of a given genera. Genera with a mean relative abundance of over 0.1 % were selected as overall abundant taxa. Ubiquitous genera between reactors were identified based on their

occurrence in over 80 % of the reactor set (where n=27; genera must have occurred in at least 22 reactors). Taxa that were present in reactors receiving either Syntho only or Syntho + influent solids but that appeared more sporadically (i.e., in fewer than 80% of reactors) or in lower abundance (below 0.1 %) were classified as the ***non-core resident populations***.

The populations of immigrating bacteria were determined and classified by comparing the reactors receiving influent solids with those receiving Syntho only (i.e., without active immigration). Bacteria present only in reactors receiving Syntho + influent solids (or sterile influent solids) were classified as influent immigrants, the presence of which was dependent upon immigration. Taxa occurring in both the reactors receiving influent solids and sterile influent solids were filtered based upon abundance; those occurring in equal abundance under both conditions were considered to be ***residual immigrants*** or residual DNA, whilst those occurring in higher abundance in the reactor receiving active solids were classified as ***growing immigrants*** (or ***actively growing immigration-dependent populations***).

4.4.9 Quantifying Net Growth Rate with Steady-State Modelling

The impact of immigration on the average growth of taxa in activated sludge systems can be visualized using a scatter plot of their relative abundances in the mixed liquor vs. the influent wastewater. A steady-state mass balance on biomass and substrates can be used to develop a model quantifying the levels of immigration and the growth rate and mapping them on the relative abundance scatter plot. The model used here was partially developed by Grady et al., 2011 (31) and it is an extension of the model presented by Mei et al., 2019 (14). A detailed development is available in Guo, 2022 and its key elements are presented here for convenience (32).

With a mass balance on the biomass of the i-th taxon for an activated sludge system assuming completely mixed biomass, it is possible to show that the i-th taxon specific growth rate (μ_i) is given by eq 1.

$$\mu_i = \left(\frac{1}{\theta_x} + b_i\right)(1 - m_i) \qquad \text{(eq. 1)}$$

Where θ_x is the solid retention time, b_i is the decay rate of the i-th taxon, and m_i is defined as the immigration level of the i-th taxon defined by eq. 2.

$$m_i = \frac{X_{Bio,i,Inf}}{X_{Bio,i,Inf} + Y_i S_{i,Inf}} = \frac{X_{Bio,i,Inf}}{X_{Bio,i,ML}(\theta/\theta_x)(1 + b_i\theta_x)} \qquad \text{(eq. 2)}$$

Where $X_{Bio,i,Inf}$ and $X_{Bio,i,ML}$ are the biomass concentrations of the i-th taxa in the influent and the mixed liquor, respectively, $S_{i,Inf}$ is the concentration of substrates consumed by the i-th taxon, Y_i is the biomass yield of the i-th taxon on the substrate consumed, and θ is the hydraulic retention time.

To map the growth rates and immigration levels of the relative abundance scatter plots, eq. 1 and eq. 2 can be developed considering the capture of the influent biomass of the i-th taxon by the mixed liquor solids ($f_{OHO,Capt}$), the DNA extraction yields for the influent and mixed liquor ($\gamma_{DNA,Inf}$, $\gamma_{DNA,ML}$) and the total solids of the influent and the mixed liquor ($X_{Tot,Inf}$, $X_{Tot,ML}$). With these considerations, the relationship between the proportions of the i-th taxon in the influent and mixed liquor as determined by 16S rRNA gene amplicon sequencing ($f_{16S,Inf,i}$, $f_{16S,ML,i}$) is given by eq. 3.

$$f_{16S,i,ML} = \frac{\theta_x}{(1 - \mu_{net,i}\theta_x)} \cdot \frac{f_{OHO,Capt}}{\theta} \cdot \frac{X_{Tot,Inf}\gamma_{DNA,Inf}}{X_{Tot,ML}\gamma_{DNA,ML}} \cdot f_{16S,i,Inf} \qquad \text{(eq. 3)}$$

Where $\mu_{net,i}$ is the net specific growth rate of the i-th taxon defined as specific growth rate minus specific decay rate ($\mu_{net,i} = \mu_i - b_i$).

The log-log version of eq. 3 (eq. 4) shows that taxa appearing on the same 45°-line (i.e., 1:1 line) in the log-log scatter plot of relative abundances have the same net growth rate and the same immigration level, and that the y-intercept (bold term between square brackets) of this 45°-line is a function of $\mu_{net,i}$.

$$\log\left(f_{16S,i,ML}\right) = \left[\log\left(\frac{\theta_x}{\theta}\cdot\frac{f_{OHO,Capt}X_{Tot,Inf}\gamma_{DNA,Inf}}{X_{Tot,ML}\gamma_{DNA,ML}}\right) - \log\left(1-\mu_{net,i}\theta_x\right)\right]_{intercept} + 1_{slope}\log\left(f_{16S,i,Inf}\right)$$

(eq. 4)

4.5 Results

4.5.1 Variations in Overall Community Compositions During Experiment

Determining the specific impact of influent immigration on the structure of the AS community in full-scale systems presents a significant challenge because of the numerous spatial and temporal variations between locations. The current study utilised controlled reactor experiments to investigate the impact of immigration on a naturally assembled, diverse microbial community. During Phase 1, all reactors received Syntho (a synthetic wastewater; 22) to develop a steady-state core microbial community adapted to a specific substrate composition. This procedure also minimized the biomass of *residual immigrants* (immigrant populations not consuming resources as defined in section 4.4.8) from the mixed liquor inocula remaining in the system prior to the test period (i.e., Phase 2).

During Phase 1 when fed with Syntho only, the microbial communities of the reactors became more similar to one another when compared to the differences among starting inocula (Figure 4.2), and the amplicon sequencing variant (ASV) richness decreased (Figure S4.1). Irrespective of the inoculum, 46.2 $\pm$ 4.7 % ($\pm$ indicates standard

deviation) of the observed genera at the end of Phase 1 were shared between the 27 resulting AS reactor communities forming the *core resident populations* of Syntho (as defined in section 4.4.8), whilst at the beginning of Phase 1, only 27.1 ± 4.7 % of the observed genera were shared among the 3 inocula. Nonetheless, the majority of AS reactor communities at the end of Phase 1 remained significantly clustered based on the inoculum received according to both Jaccard and Bray-Curtis dissimilarities. The ANalysis Of SIMilarity (ANOSIM) can be used to express in a common quantity the differences between clusters, with ANOSIM R values closer to 1.0 indicative of higher dissimilarity between groups, whilst those closer to 0.0 indicating higher similarity. Comparing Jaccard and Bray-Curtis dissimilarities among communities at the end of Phase 1 revealed a higher difference measured by Jaccard dissimilarities (ANOSIM R = 0.88 ± 0.05) than Bray-Curtis dissimilarity (ANOSIM R = 0.61 ± 0.04; Figure 4.2). The Jaccard dissimilarity is more sensitive to the least abundant populations than the Bray-Curtis dissimilarity, this suggests that the main difference between the communities assembled on Syntho during Phase 1 was in low abundance populations. These observations show that Syntho selected for a *core resident populations* comprising of the most abundant populations. Whilst populations specific to the inoculum community were in lower abundance. As the experiment progressed, the microbial community of most reactors receiving the same feed remained clustered based on the inoculum (Table S4.3) with the exception of a few outliers as visualised in Figure S4.1.

During Phase 2, reactors received three different feeds. The test reactors received a feed comprising of Syntho supplemented with influent solids to determine the impact of immigration. The remaining reactors received Syntho and sterile solids or Syntho only, which acted as substrate and continuity controls, respectively. At the end of Phase 2,

principal coordinate analysis using Jaccard dissimilarity (Figure 4.3) revealed that immigration caused the resulting communities to become more similar to the influent community, as shown by the points representing these communities being located closer to the influent community along the PCoA1 axis. The microbial community of the reactors receiving sterile solids also moved away from the starting position, likely due to the additional substrate and residual DNA in the autoclaved influent biomass (Figure 4.3).

Unlike with Jaccard dissimilarity, when visualising the community composition data using Bray-Curtis dissimilarity (Figure 4.3), the impact of immigration on the overall community was less apparent, particularly in reactors of Set A and B. Given the higher sensitivity of the Jaccard dissimilarity than the Bray-Curtis dissimilarity for the presence of low abundance population, it appears that immigration had more of an impact on these populations that are likely maintained by their continuous influx in the mixed liquor from the influent.

During Phase 2, the ASV richness of reactor Set A (test reactors receiving La Prairie influent solids) and B (test reactors receiving Cowansville solids) increased (Figure S4.2). This increase was not observed among the flasks receiving sterile influent solids nor Syntho only throughout, thus indicating that the immigrant populations are associated with the changes. The impact of immigration on the richness of reactor Set C was more variable. The original richness of the reactors was not restored to that of the starting community (Figure S4.2), likely due to the relative substrate-composition simplicity of Syntho compared to actual wastewater and the greater homogeneity of laboratory-scale flask reactors than of full-scale AS-WWTPs.

Phase 3 of reactor operation was introduced to determine whether the impact observed with immigration could be sustained over time without continuous seeding of influent solids. During this phase, influent solids were removed from feeding, and all reactors received Syntho only as in Phase 1. Within 3 SRTs, the communities in reactors which previously received active influent solids during Phase 2, moved back towards the starting position (start Phase 2) and became more similar to the communities in reactors which received sterile inactive influent solids throughout (average Jaccard distance between reactors with immigration and sterile controls reduced from 0.56 to 0.47 in Set B and 0.59 to 0.49 in Set C indicating increased similarity; Figure 4.3) demonstrating that the full impact of immigration could not be maintained over time without continuous seeding.

4.5.2 Classification of Genera

To provide a more detailed analysis of the impact of immigrating taxa, the genera forming the microbial community at the end of Phase 2 were classified in four categories according to their presence in reactors receiving influent solids. This analysis was conducted at genus level as ASVs within the same genus are likely to have similar ecological functions, and because direct comparisons of ASVs between the influents and the mixed liquors were often unreliable due to their low abundances in one of the two compartments. The four categories assigned are outlined in section 4.4.8 and are briefly recalled here. First, a genus was defined as a ***core resident population*** if it occurred in at least 80 % of all the reactors within a set with a relative abundance of at least 0.1 %. Second, ***non-core resident genera*** appeared in reactor without immigration, but in fewer than 80 % of these reactors or at an average abundance below 0.1 %. Third, ***growing immigrant genera*** were only present in the reactors receiving influent solids and were at a higher abundance than in the reactors receiving autoclaved

solids in the same set (i.e., same influent solids source). These genera were presumed to be actively growing. Fourth, ***residual immigrant genera*** were those present only in the reactors receiving influent solids and occurring in the reactors with immigration in equal or lower abundance to the sterile autoclaved control group within the same set.

The resident community in all reactors consisted of between 50 and 73 genera depending on the set of reactors (i.e., influent community), and these genera accounted for between 69 and 76 % of the total reads observed in the reactor set (Table 4.1). Conversely, the immigrant populations showed a higher diversity (118 to 206 genera) but accounted for much fewer reads from the reactor set (4 % to 14 %). This high richness and low abundance of the immigrant populations is in line with the impact of immigration being more visible when using the Jaccard dissimilarity, but not when using the Bray-Curtis dissimilarity (Figure 4.3). Therefore, operation and the Syntho substrate composition appeared to have determined the resident abundant members of the communities, whilst immigration contributed towards low abundance members of the communities.

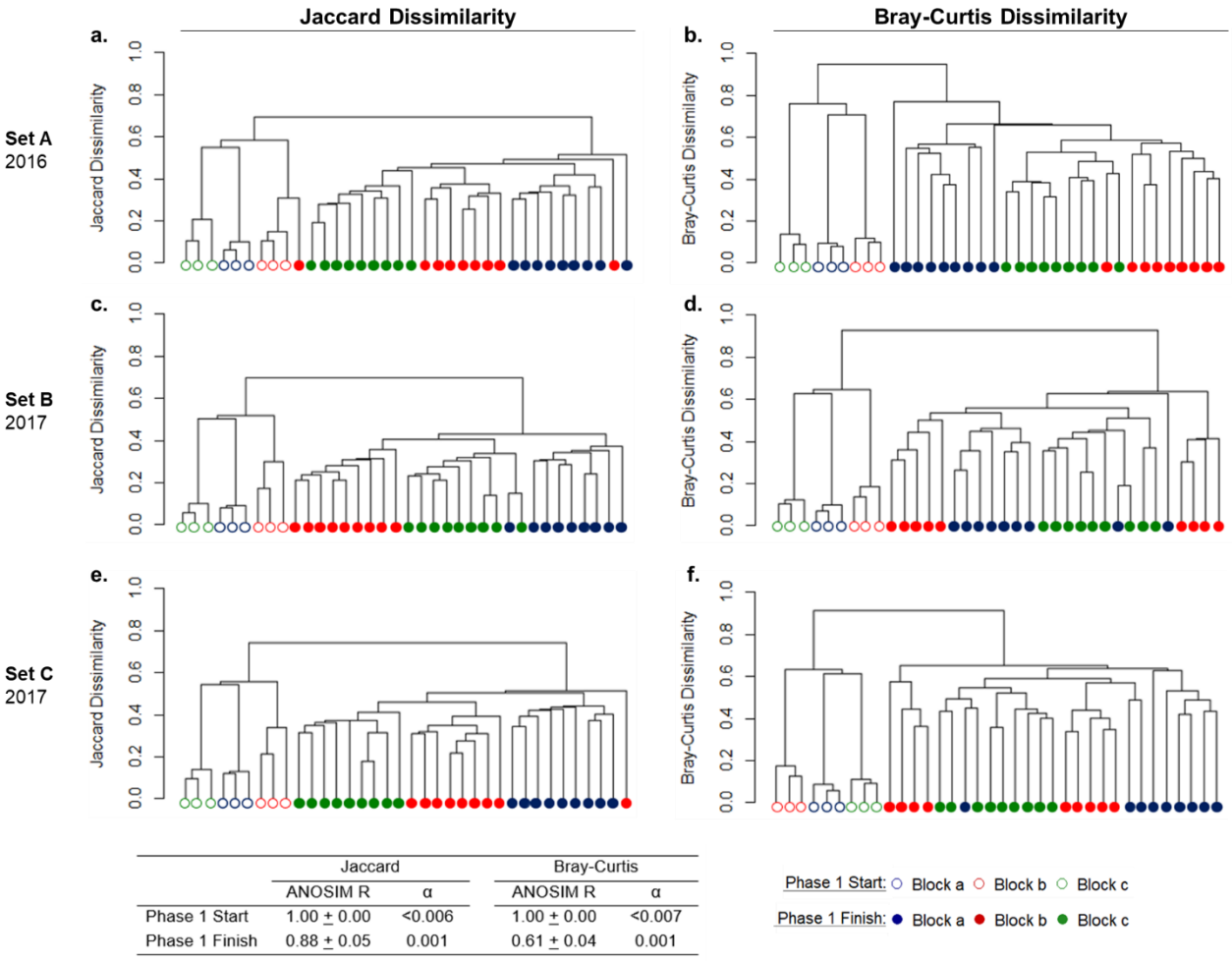


Figure 4.2- Tree dendrogram of microbial community at the start (hollow symbols) and end (filled symbols) of Phase 1 using UPGMA clustering method and the Jaccard dissimilarity (a, c, e) or the Bray-Curtis dissimilarity (b, d, f). Set A (a, b) were operated in 2016, whilst Sets B (c, d) and Set C (e, f) were operated in 2017. The Block indicates the source of inoculum: Block a-La Prairie mixed liquor, Block b-Cowansville mixed liquor, Block c-Pincourt mixed liquor. Inoculum communities were the same for Set B and C as they were operated in tandem. Each inoculum (hollow symbols) represents a single biological sample that was extracted and sequenced 3 times, whilst the end of Phase 1 (filled symbols) are communities raised in independent reactors.

Table 4.1: Percentage of overall reads which form the resident and immigrant communities at the end of Phase 2 for reactor sets receiving different influent solids

	Set A: La Prairie		Set B: Cowansville		Set C: Pincourt	
	Genera[1]	Reads (%)[2]	Genera	Reads (%)	Genera	Reads (%)
Core Residents	73	75.3 ± 5.4	54	76.0 ± 4.9	50	69.3 ± 11.5
Non-Core Residents	133	15.7 ± 4.3	121	7.9 ± 1.7	122	22.4 ± 12.8
Growing Immigrants	127	8.2 ± 2.0	206	14.3 ± 4.9	118	4.2 ± 1.4
Residual Immigrants	7	0.3 ± 0.2	5	2.1 ± 1.6	12	2.9 ± 3.2

[1]Number of genera within the category

[2]Percentage of overall reads ± standard deviation

The immigrant population varied based upon the source of influent solids (Figure 4.4). In reactor Set A, which received influent solids from La Prairie wastewater treatment plant, the genus Spb280 (genus midas_g_81, family *Comamonadaceae*; 33), and an uncultured genus of the family *Synergistaceae* were the dominant immigrants. Whilst in Set B, members of the genus *Aquabacterium* (family *Comamonadaceae*) were dominant, and in Set C, the most abundant immigrant genus was *Dechloromonas* (family *Rhodocyclaceae*). In addition to unique immigrant communities based on the source of influent solids, there also appeared to be some inoculum effect (between Blocks). Within the immigrant population of Set B reactors, those which were inoculated with La Prairie mixed liquor (Block A) were abundant in the genus SipK9 (genera *midas_g_1719 and midas_g_2835,* family *Rhodobacteraceae*; 33) which accounted for up to 11.7 % of the reads and explained some of the variability in percentage reads contributed though immigration reported in Table 4.2. Whilst in the other Set B reactors inoculated with Cowansville and Pincourt (Block b and c,

respectively), SipK9 was not observed with the same high abundance. The function of SipK9 is currently unknown but given the abundance in the reactors may warrant further investigation.

Table 4.2: Percentage of overall reads which form the resident and immigrant communities at the end of Phase 3

	Set B		Set C	
	Genera	Reads (%)	Genera	Reads (%)
Core Residents	54	85.7 ± 5.9	50	78.3 ± 15.2
Non-Core Residents	124	10.4 ± 5.0	121	19.4 ± 13.4
Growing Immigrants	65	3.5 ± 2.1	29	1.8 ± 2.7
Residual Immigrants	4	0.1 ± 0.1	9	0.2 ± 0.3

[1]Number of Genera within the category

[2]Percentage of overall reads ± Standard Deviation

During Phase 3, influent solids were removed, and all reactors received Syntho Only as in Phase 1 to determine if immigration had a lasting effect on the microbial community. Phase 3 was only conducted for reactor Set B and C in 2017. At the end of Phase 3, the number of immigrants remaining in the reactors had reduced in both reactors Set B and C (Table 4.2). In reactor Set C, immigrants remaining included taxa from the families *Rhodospirillaceae*, *Rhodocyclaceae*, *Nitrospiraceae* and *Flavobacteriaceae*. In reactor Set B, the families *Rhodocyclaceae*, *Gallionellaceae* (genus *Candidatus* Nitrotoga), *Comamonadaceae* (genus *Rhodoferax*) and *Xanthomonadaceae* (genus sipK9) were among the most abundant growing immigrants remaining. To further investigate factors

influencing the persistence of immigrants in the reactors at the end of Phase 3, further analysis of the net growth rate was conducted.

4.5.3 Assessing the Growth Rate of Genera

The impact of immigration on the various genera can be visualized with a log-log scatter plot of their relative abundances in the mixed liquor and in the influent microbial communities (Figure 4.5). Assuming that the ASV composition of each genera is the same in the influent and in the mixed liquor, the log-log scatter plot of the relative abundances of the genera can be understood as representing their net growth rate by using a mass balance model (eq. 4). Based on the model in eq. 4, genera with different relative abundances but falling on the same 45°-lines on Figure 4.5 have the same net growth rate. As a reference, a 45°-line was drawn on the scatter plot of Figure 4.5 for net growth rates (μ_{net}) equal to 0 (i.e., the point where the growth rate of the organisms is equal to its decay rate). Any points falling above the zero net growth rate line ($\mu_{net} = 0$) display a positive net growth rate, whilst those below this line have an overall negative net growth rate. According to reactor theory, organisms displaying positive net growth rates should be maintained in the reactor without immigration assuming no other factors such as competition occur, whilst organisms displaying negative net growth rates would be washed out if immigration were stopped.

Genera classified in the core resident population by our criteria typically exhibited a positive net growth rate (Figure 4.5), whilst genera classified in the growing immigrant population typically had a negative net growth rate, particularly those from Cowansville and Pincourt influents (Set B and C; Figure 4.5b and 4.5c). Selection against certain genera present in the influent at high relative abundant (up to 1 % of the reads) was also observed as these genera remained undetected in the mixed liquor (Figure 4.5; shown

as ‘ND’ in ML). Interestingly, some genera classified as immigrants were not detected (ND) in the influent, suggesting that they occurred below the detection limit in the influent solids, and increased to detectable levels within the reactor communities.

In Figure 4.5, it was observed that a selection of immigrants displayed a positive net growth rate. Based on reactor theory, it would be expected that these genera could be maintained within the reactors without continuous immigration. At the end of Phase 3, although reduced, a number of immigrants remained in the reactors (Table 4.2). Analysis of the net growth rate of these taxa (Figure 4.6), showed that of those remaining at the end of Phase 3 between 75 and 77 % had displayed an overall positive net growth rate during Phase 2. However, many other immigrants that had a positive net growth rate during Phase 2, were not detected at the end of Phase 3. An odds ratio test of the taxa remaining at the end of Phase 3, and their associated net growth rate evaluated in Phase 2 supported that taxa with a positive net growth rate were between 5.6 and 7.4 times more likely to persist at the end of Phase 3 than those with a negative net growth rate.

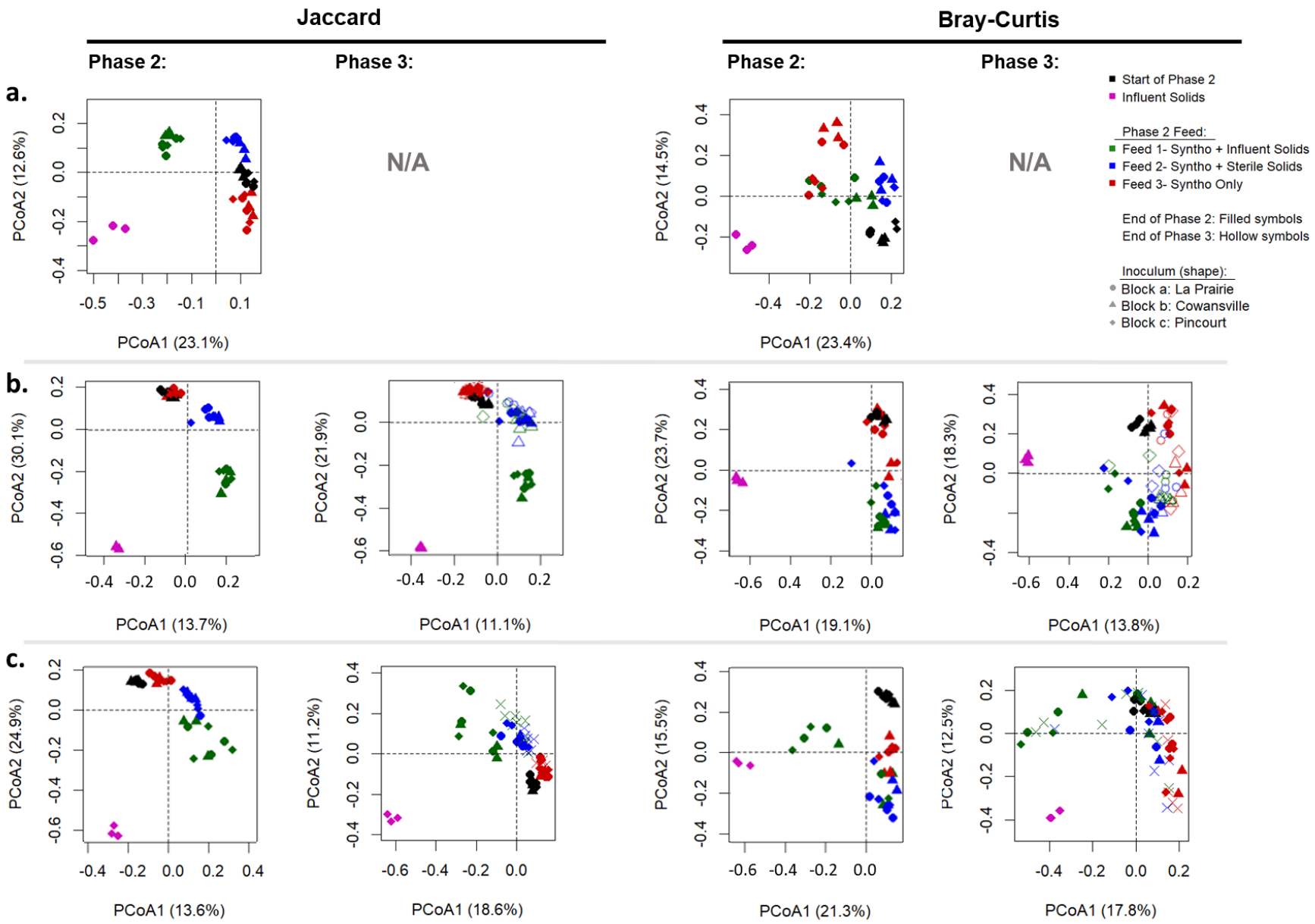


Figure 4.3- Principal coordinate analysis of reactors communities at the end of Phase 2 and 3, visualised at ASV level using Jaccard and Bray-Curtis Dissimilarity. a) Set A- La Prairie Influent Solids b) Set B- Cowansville influent solids c) Set C- Pincourt influent solids. Filled symbols represent samples taken at the end of Phase 2 and hollow symbols were samples collected at the end of Phase 3.

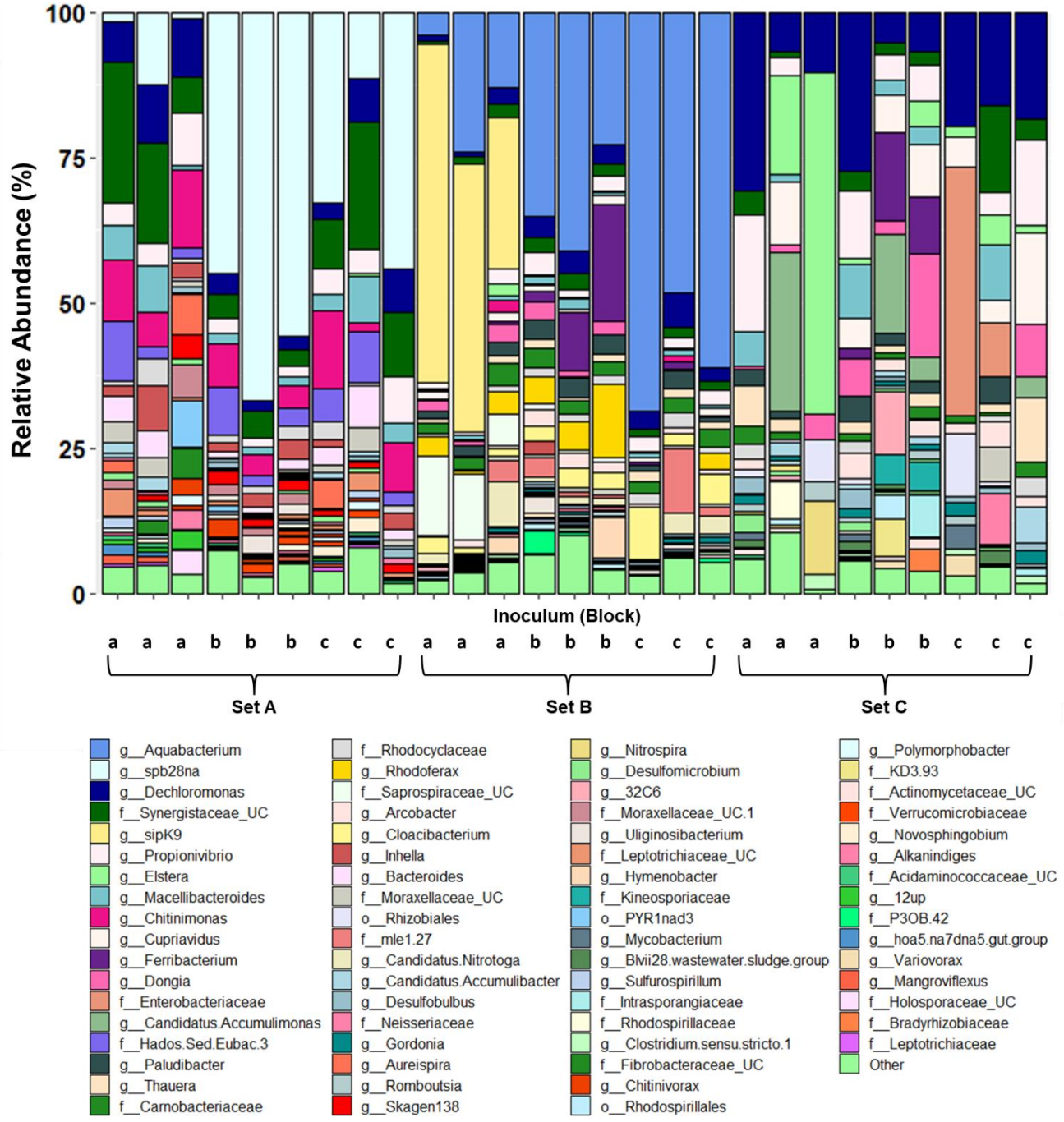


Figure 4.4- Top 70 immigrating genera, normalised to total immigrants per reactor. Inoculum A- La Prairie ML, B- Cowansville ML, C- Pincourt ML. Set A- La Prairie influent solids, B- Cowansville influent solids, C- Pincourt influent solids. UC in the legend indicates uncultured strains of a given taxa.

The remainder of the immigrants present at end of Phase 3 were classified as having a negative net growth rate (Figure 4.6). The detection of these taxa at the end of Phase 3 was unexpected, however it was noted that these genera typically occurred in higher abundance at the end of Phase 2. During Phase 3, the abundance typically reduced by at least 85 % suggesting that washout would likely occur over time. When excluding these taxa, an odds ratio test showed that taxa with a positive net growth rate were 14.0 to 15.2 times more likely to persist at the end of Phase 3 than those with a negative net growth rate. Other factors, such as the influent solids as an additional food source also likely influenced the persistence or loss of these taxa during Phase 3.

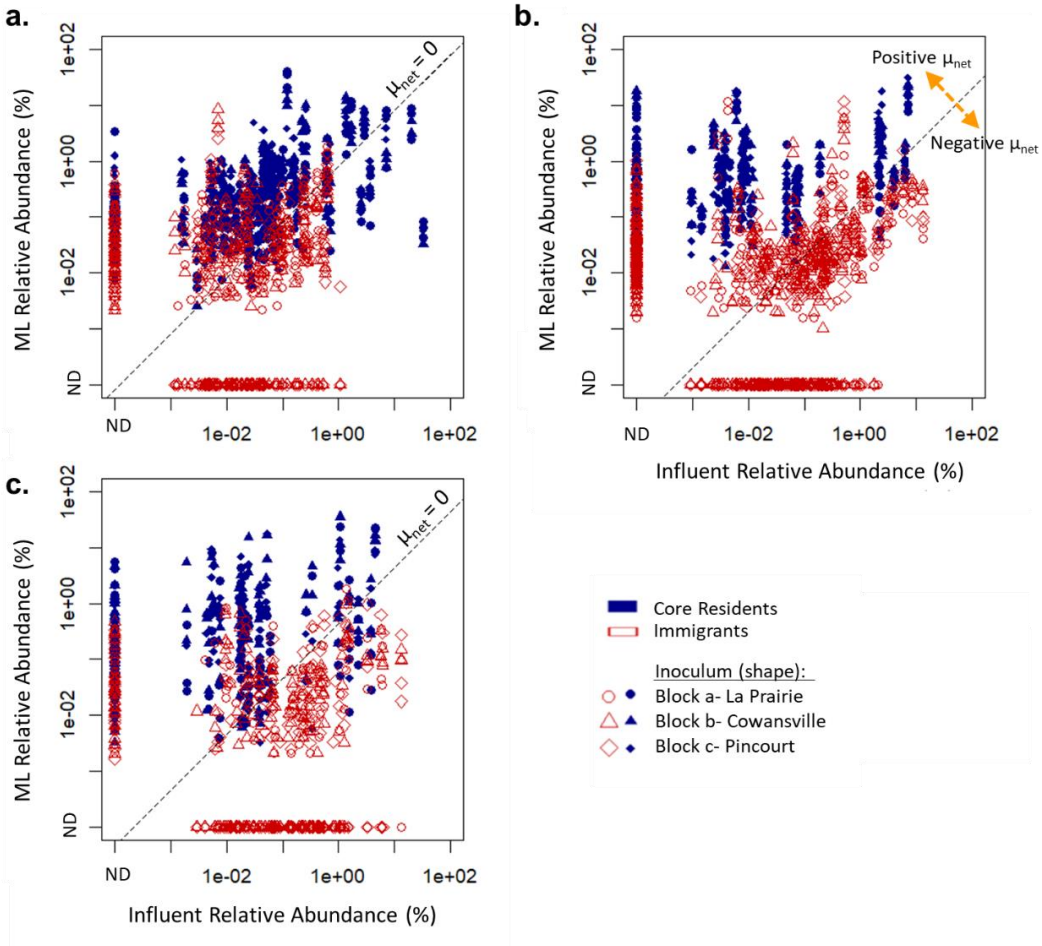


Figure 4.5- Comparison of influent and mixed liquor (ML) relative abundance of immigrant and resident genera. 'ND': Not Detected indicated genera that were below the detection limit. Reactors received influent solids of different sources: a) Set A- La Prairie influent solids b) Set B- Cowansville influent solids c) Set C- Pincourt. Dashed line represents a net growth rate of 0. Genera on the same 45°-lines have the same net growth rates (μ_{net})

4.5.4 Impact of Immigration on Core Resident Genera

The quantitative assessment of immigration reported thus far focused exclusively on immigration dependent communities, i.e., those which are introduced into the system and only present with continuous immigration. Core resident genera were present under all reactor conditions, both with and without immigration, and accounted for up to 76 % of sequencing reads (Table 4.1). It was also noted these genera were often detected within the influent wastewater, with differences in the ASVs observed in the influent solids and the reactors receiving only Syntho. It could be hypothesised that although the core resident population remains relatively stable over time in all reactors irrespective of feed received, immigration may contribute to diversity of these genera at ASV level.

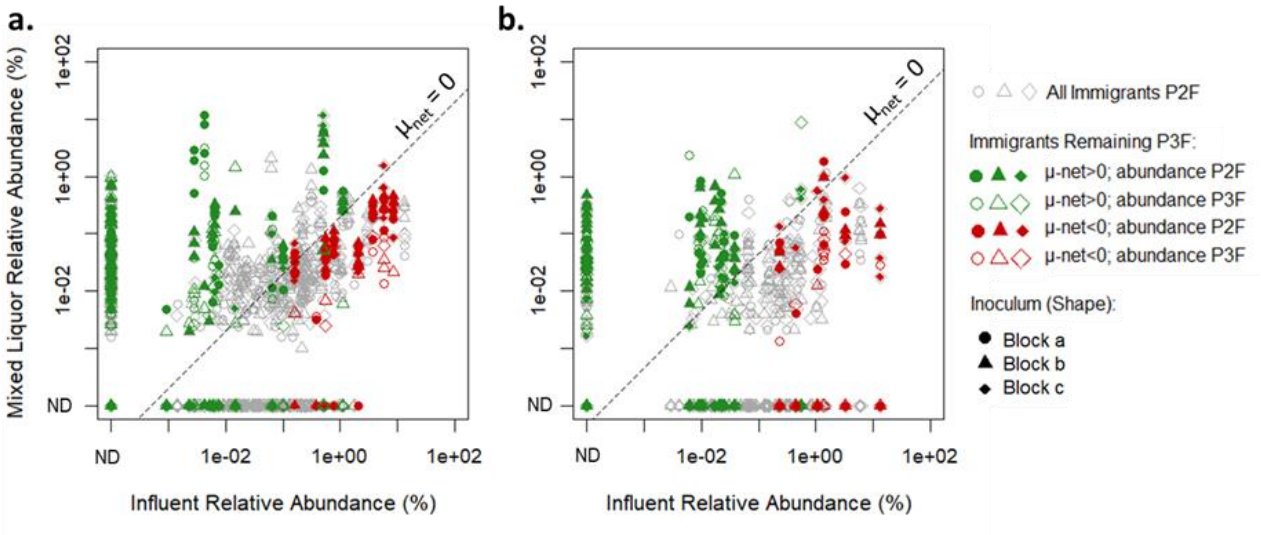
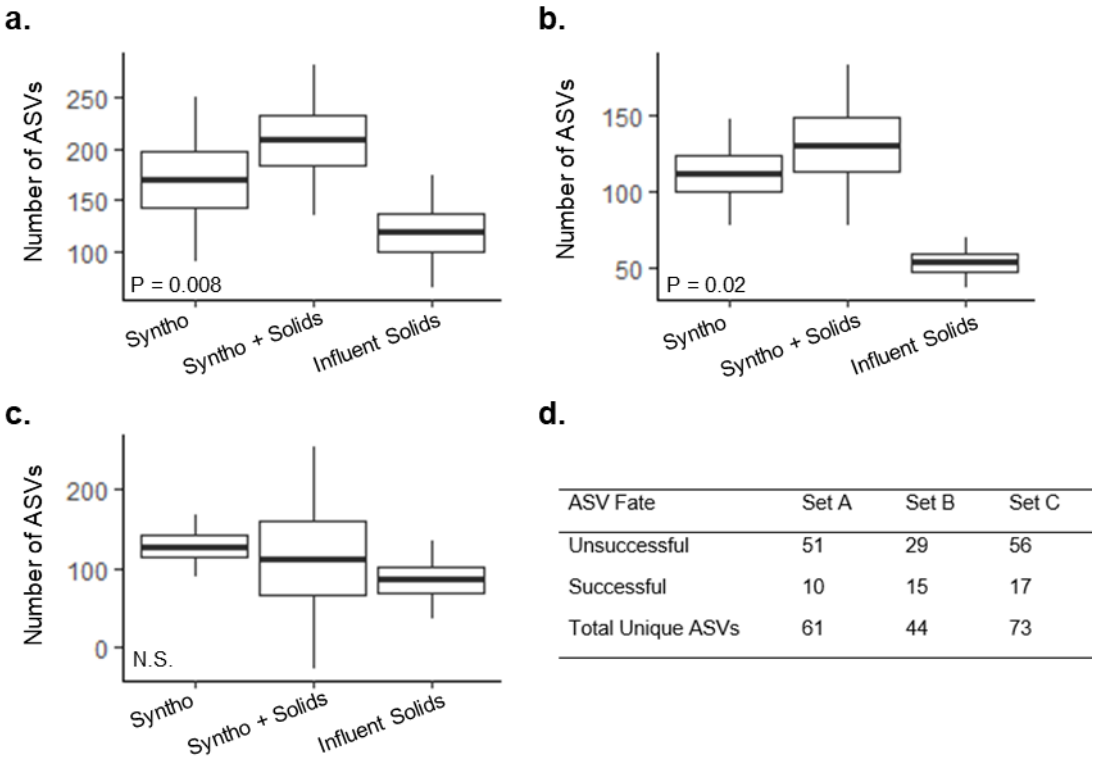


Figure 4.6- Comparison of growing immigrant genera at the end of phase 2 (P2F) and phase 3 (P3F). 'ND': Not detected indicated genera which were below the detection limit. a) Reactor Set B; received Cowansville Influent Solids b) Reactor Set C received Pincourt Influent Solids. μnet was modelled as described in section 4.4.9 and Table S4.3.

Analysis of the ASV richness of core resident genera at the end of Phase 2 (Figure 4.7) showed a significant increase with immigration in Set A and Set B of the reactors (Figure 4.7a and 4.7b; unpaired t-test P = 0.008 and P= 0.02 respectively). A significant increase was not observed in Set C of the reactors (Figure 4.7c), where the impact of immigration on ASV diversity appeared to be more variable. In total between 44 and 73 unique ASVs belonging to genera classified as core residents were identified in the influent microbial community (Figure 4.7d), demonstrating that immigration had an impact on the core resident community at ASV level. However, between only 10 to 17 of these ASVs were detected in the reactors receiving influent solids, indicating selection to occur between these settings. Of those successfully immigrating were ASVs from the genera *Acinetobacter*, *Pseudomonas* and *Zoogloea*. The genus *Acinetobacter*, which was highly diverse without immigration (up to 15 ASVs) had 3 additional ASVs introduced. However, selection was also observed, and several *Acinetobacter* influent ASVs were not detected in the reactors. Multiple sequence alignment was used to study the sequence similarity between the ASVs (Figure S4.3a) and it was observed that ASVs which formed the core *Acinetobacter* population and successful immigrants typically clustered together.

Similarly, up to 4 additional *Pseudomonas* ASVs were introduced with immigration whilst other influent ASVs did not successfully immigrate. Multiple sequence alignment of the *Pseudomonas* ASVs did not show clear clustering based on the fate of a given ASV (Figure S4.3b). In the genus *Zoogloea* all influent ASVs were detected in the reactors with immigration. Multiple sequence analysis was performed to look at the sequence similarity between each ASV and to further investigate patterns of selection (Figure S4.3c). Among the *Zoogloea* ASVs, no clustering was observed based upon their occurrence in the core resident populations or influent solids.



ASV Fate	Set A	Set B	Set C
Unsuccessful	51	29	56
Successful	10	15	17
Total Unique ASVs	61	44	73

Figure 4.7- Impact of immigration on ASV richness of the core resident genera. a) Reactor Set A received La Prairie influent solids, b) Reactor Set B received Cowansville influent solids, c) Reactor Set C received Pincourt influent solids. d) The fate of unique influent ASVs in the reactors; 'Unsuccessful' were those detected in the influent, but not in the reactors, 'Successful' were those which were detected in the reactors with immigration.

4.6 Discussion

The importance of influent immigration has long been debated among technical professionals, theorists, and modellers alike. One factor obscuring this debate in studies of full-scale wastewater treatment plants is the impossibility to differentiate the impact of influent substrate landscape from the immigration of microbial populations. To achieve this differentiation, the current study aimed to develop a reproducible and controlled system to investigate the impact of immigration and its relative contribution to the wastewater treatment plant community. To determine the fate of immigrating

bacteria both substrate and continuity controls were included and to assess the impact of the AS-WWTP community itself, three different inoculums were used.

4.6.1 Wastewater Substrate Composition and Operation Cause Communities to Become More Similar

The impact of wastewater substrate composition on the activated sludge community is well documented in the literature (34,35). Thus, it was as expected that during Phase 1 the introduction of a common wastewater source and identical operational conditions caused the inoculum communities to become more similar (Figure 4.2). As discussed, comparing the Jaccard and Bray-Curtis dissimilarities of the communities at the end of Phase 1 revealed a higher difference measured by Jaccard dissimilarities (ANOSIM R = 0.88 $\pm$ 0.05) than Bray-Curtis dissimilarity (ANOSIM R = 0.61 $\pm$ 0.04; Figure 4.2). Taken together this suggested that the influent's substrate composition selects for a core resident microbial community, whilst features of the inoculum were associated with lower abundance genera.

At the end of Phase 2, reactor Set A and B remained significantly clustered based upon the inoculum received (Figure S4.1 and Table S4.3; Jaccard distance, ANOSIM R= 0.77 $\pm$ 0.13). However, in reactor Set C, clustering was not observed in the test reactors which received influent solids, suggesting that features of the starting community were impacted by immigration (Table S4.3).

On a larger scale, with an increase in the popularity of microbial transplantation research over recent years to improve feed efficiency in livestock (36), or to restore function of commensal gut microbiota and reduce the prevalence of multidrug resistant organisms (37), the plausibility of activated sludge bioaugmentation has been raised. This concept would involve taking a desired activated sludge microbial community

from one location and transplanting it into a AS-WWTP requiring optimisation. The current study suggests activated sludge bioaugmentation may have variable results, as the activated sludge community is largely dependent on the wastewater substrate composition and not the inoculum alone.

4.6.2 Immigration Impacted the Microbial Community of the Activated Sludge

Previous studies have inferred the impact of immigration based upon shared taxa between the influent and AS-WWTP, which provides limited information about the microbial activity of these organisms within the AS-WWTP. The inclusion of a sterile substrate control in this study allowed us to quantitively assess the impact of immigration, enabling the active and inactive portions of the immigrating community to be distinguished experimentally. Active immigration-dependent genera were defined as those only present in reactors receiving influent solids, and in a higher abundance than in the sterile control, indicating growth of the taxa. Actively growing immigration dependent taxa were found to account for between 4 to 14 % of the total reads in the mixed liquor, contributing a significant proportion of the community (Table 4.2). This result was consistent with previous studies that reported 10 % of OTUs to be present primarily due to immigration (9). Considering the significant proportion of activated sludge reads contributed through immigration, this source should not be neglected in process design and optimisation. Nonetheless, due to the low abundance of immigrant populations, analysis at ASV level was not possible. Future immigration studies should consider performing deeper sequencing strategies to enable analysis of ASV diversity in taxa occurring at low abundance.

Beyond immigration-dependent genera, immigration also impacted core resident genera, which were defined based upon their presence in at least 80 % of reactors under

all conditions (with and without immigration) (Table 4.1). It was observed that genera classified as core residents were often present in the influent wastewater, and the ASV composition of these genera was different between the two environments. Analysis showed a significant increase in diversity of the core resident population with immigration in two of the three sets of reactors (reactor Set A and B; Figure 4.7a & 4.7b). Successful immigration was observed between the influent and mixed liquor with between 1 and 6 % of reads contributed by immigrating ASVs. As a result, the steady-state model used in this study may somewhat inaccurately estimate the growth rate of some core resident genera receiving immigrant ASVs.

Selection was an important factor filtering the immigrating populations (Figure 4.7d), as the majority of unique ASVs identified in the influent solids were not detected in the reactors with immigration. Overall, between only 16 and 34 % of ASVs unique to the influent wastewater successfully immigrated into the reactors. This is consistent with previous studies which show that heterotrophs are strongly selected between influent wastewater and the activated sludge (12). This could be due to niche differences between the environments, or competition for resources with the well-established, metabolically similar reactor core resident community. The influence of selection varied at genus level among the resident core genera (Figure S4.3). For example, ASV selection was observed in genus *Acinetobacter* but not in genus *Zoogloea.* This supports the proposal that heterotrophs should not be considered as a whole, and instead should be divided into relevant functional guilds for proper dynamic analyses.

4.6.3 Growth Rate is a Key Determinant in the Fate of Immigrants

Analysis of the resident and immigrant populations (Figure 4.5) showed differences in the net growth rates. Resident genera typically exhibited positive net growth rates,

explaining how they are present without constant seeding. Conversely, immigrant genera typically displayed low or negative net growth rates, particularly in reactor Set B and C. Based upon the terminology proposed by Frigon and Wells (2019) the majority of immigrants could be classified as mass-flow immigrants (i.e., continuous immigration at a high rate), as they displayed low or negative net growth rates in the mixed liquor (17). This could be attributed to differences in niche availability between the influent wastewater and the reactor communities. This result is consistent with that predicted by studies using a mass balance approach, which showed a large proportion of taxa shared between the influent and AS-WWTP to have a negative net growth rate (9). It was concluded that this fraction of the immigrant community was inactive and did not contribute to the metabolism of the activated sludge. However, despite these findings, we empirically demonstrated that although displaying a negative net growth rate, active immigration-dependent genera are in fact growing and consuming substrates within the activated sludge because their abundance was higher than in reactors receiving fresh solids than in reactors receiving autoclaved solids. This observation is in agreement with previous studies that demonstrated the rescue of complete nitrification in activated sludge bioreactors by nitrifiers in the influent stream (11). Consequently, observations of shared OTUs (ASVs or taxa) do not accurately predict activity, and more work should be done to quantify the metabolic contribution of immigrants with net negative growth rates.

During Phase 2, continuous influent immigration prevented the competitive exclusion of the immigrating bacteria, which were not well adapted to the reactor systems. This was confirmed during Phase 3 (reactor Set B and C only), when reactors received synthetic wastewater only and immigration was removed. During this phase, a large proportion of immigration-dependent genera were washed out within 3 SRTs (Table

4.3). Of those remaining, between 75 and 77 % exhibited a positive net growth rate during Phase 2, which is consistent with reactor theory that these genera could be maintained within the reactors without the need for continuous immigration. However, other genera classified as actively growing immigration-dependent genera which displayed a positive net growth rate during Phase 2 (Figure 4.5) were not detected at the end of Phase 3. The discrepancy of these counter observations with reactor theory could be due to the imprecisions and inaccuracies in the quantification of growth rates. Nonetheless, they could also be explained by various abiotic and biotic phenomena. Abiotic phenomena include the possibility that influent solids provided an additional food sources or other beneficial niche markers for these genera which were removed during Phase 3. Biotic explanations may come from the co-selection of immigration-dependent populations as a single unit, but with some unit members exhibiting positive net-growth rates and others exhibiting negative net-growth rate (38). The reproducibility of the experimental system presented here would allow these different mechanisms to be explored in detail in future experiments.

The remainder of immigrants present at the end of Phase 3 displayed a negative overall net growth rate (Figure 4.6). Generally, these taxa had a higher relative abundance (0.1 % or above) at the end of Phase 2, whilst those with lower abundance were washed out by the end of Phase 3. A notable reduction (85 % and above) in the abundance of these genera was observed within 3 SRTs, suggesting that washout would likely occur with prolonged reactor operation and no immigration.

4.6.4 Resident Core Genera

Detailed studies of the microbial community composition of AS-WWTP samples collected worldwide has enabled core taxa to be identified based upon their occurrence

in different locations and their relative abundance (33). The core resident genera of the reactor communities overlapped with those previously reported in full-scale wastewater treatment plants. For example, some of the most abundant resident core genera such as *Zoogloea*, *Haliangium* and *Flavobacterium* were previously classified as strict core genera, which were observed in at least 80 % of the AS-WWTPs with a relative abundance >0.1 % (33). Whilst other abundant reactor core residents such as *Aeromonas* and *Thermomonas* were classified as 'general core' taxa, which occurred in 50 % of full-scale AS-WWTP at an abundance of >0.1 %. Taken together, these results demonstrate that the core community produced during this reactor study is representative of that in full-scale AS-WWTPs.

4.6.5 Growing Immigrant Genera

The community of immigrating bacteria varied based upon the source of influent solids received. Among reactor Set A and B, the genera Spb280 and *Aquabacterium* respectively (both family *Comamonadaceae*) were dominant immigrants (Figure 4.4) and found exclusively within these reactor sets. Members of the *Comamonadaceae* family have previously been identified as habitat generalists capable of surviving in both influent wastewater and activated sludge (17). Thus, the presence of these immigrants is consistent with results from full-scale AS-WWTP immigration studies.

Among reactor Set C, which received Pincourt influent solids, the genera in the immigrant population were more variable (Figure 4.4). Whilst some abundant genera such as *Dechloromonas* appeared in all reactors, others such as *Elstera* and *Candidatus* Accumulimonas appeared more sporadically. This could be due to differences in niche availability in reactors or co-selection. Limitations associated with small-scale reactor experiments should also be considered. Given the influent solid particle size and the

small reactor volume, there may have been a slight variability in the microbial community of the influent solids received by each reactor that exacerbated the observed community drifts.

Among the reactors, few examples of inoculum effect were observed. This validates this reactor protocol to be suitable for future immigration (or perpetual coalescence) studies, regardless of the mixed liquor inoculum source. In reactor Set B, one example where inoculum appeared to have an impact, was in the genus SipK9 (family *Xanthomonadaceae*), which was dominant only among reactors receiving inoculum A (La Prairie) and accounted for between 2.5-11.7 % of overall reads. SipK9 is an uncultured genus which has been previously reported in Antarctic mineral soils (39) and activated sludge AS-WWTPs as documented in the MiDAS ecosystem specific reference database (33). In the reactors inoculated with Pincourt and Cowansville ML, the maximum abundance of SipK9 was 0.008 % indicating that there were differences in niche availability or co-selection between the reactors based on the inoculum received. The function of this population is currently unknown. However, its high abundance warrants further investigation.

Control over specific populations may be key to both process optimisation and mitigation of operational problems in activated sludge. In the immigrating community specialised bacteria such as the phosphorus accumulating organisms *Dechloromonas*, *Candidatus* Accumulimonas, *Candidatus* Accumulibacter and *Tetrasphaera* were identified (Figure 4.4). Given the importance of these genera in enhanced biological phosphorous removal, immigration may be key in process optimisation. Unfavourable taxa such as *Gordonia*, a filamentous bacteria, and *Thauera,* which cause bulking and dewatering problems in activated sludge treatment plants, were also identified (40). By

accounting for the degree of immigration in the design of solutions, operational problems in AS-WWTPs could be managed more efficiently.

4.7 Acknowledgments

We would like to thank Julia Qi, Carlos Vasquez Ochoa, Nouha Klai and Zeinab Bakhshijooybari for their assistance with reactor operation. We are also indebted to the operators and staff at Cowansville, La Prairie and Pincourt wastewater treatment plants for access to the facilities and help with sampling.

This work was funded by NSERC through a Discovery grant (NSERC RGPIN-2016-06498) and a Strategic Program Grant (NSERC STPGP 521349-18). NSERC had no role in the design of this study.

Claire Gibson, Shameem Jauffur, and Bing Guo were partly funded by the McGill University Engineering Doctoral Award.

4.8 Supplementary Material

Table S4.1- Phases of reactor operation

Phase	Reactor Feed	Duration (SRTs) [1]	
		2016	2017
1	All receive Syntho only	12	10
2	Reactors receive three different feeds[2]	13	22
3	All receive Syntho Only	NA[3]	3

[1]One SRT represents 5 days of reactor operation

[2]Three feeds; Syntho and influent solids, Syntho and sterile solids and Syntho Only

NA: not available; reactors in 2016 were not operated with a third phase

Table S4.2: Overview of the Distribution of Reactors in the Experimental Design (**Total Reactors:** 72)

	Syntho + Influent Solids			**Syntho + Sterile Solids**			**Syntho Only**		
	Block[1]			Block			Block		
Set[2]	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**
A	3	3	3	3	3	3	3	3	3
B	3	3	3	3	3	3	3	3	3
C	3	3	3	3	3	3	Same as Set B[3]		

[1] Where Block a, b and c represent the different sources of the inoculum

5 Where Set A, B and C represent the different sources of influent solids

[3] Syntho Only Control was shared for set B and C as the experiments were conducted at the same time

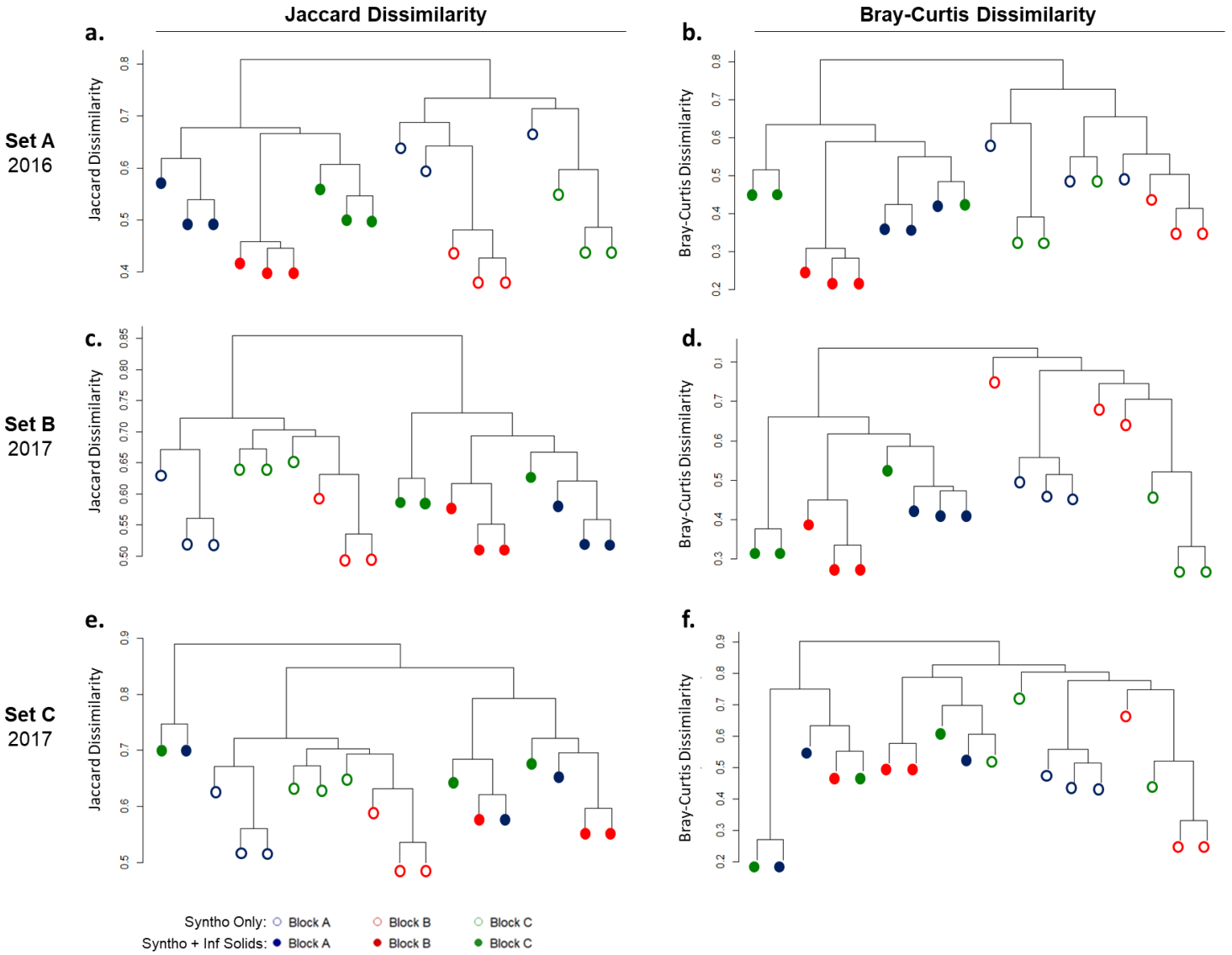


Figure S4.1- Tree dendrogram of microbial community at the end of Phase 2 using UPGMA clustering method and the Jaccard dissimilarity (a, c, e) or the Bray-Curtis dissimilarity (b, d, f). Filled symbols indicate test reactors which received influent solids during Phase 2. Hollow symbols indicate samples which received Syntho only throughout. Set A (a, b) were operated in 2016, whilst Sets B (c, d) and Set C (e, f) were operated in 2017. The Block indicates the source of inoculum: Block a-La Prairie mixed liquor, Block b-Cowansville mixed liquor, Block c-Pincourt mixed liquor. Inoculum communities were the same for Set B and C as they were operated in tandem.

Table S4.3- ANOSIM Analysis of Microbial Community at the End of Phase 2 (Figure S1)

		Jaccard Dissimilarity		Bray-Curtis Dissimilarity	
		ANOSIM R	α[1]	ANOSIM R	α
Set A	Syntho + Influent Solids	0.96	0.005	0.84	0.008
	Syntho Only	0.61	0.013	0.53	0.017
Set B	Syntho + Influent Solids	0.75	0.003	0.75	0.007
	Syntho Only	0.77	0.001	0.46	0.02
Set C	Syntho + Influent Solids	NS[2]	NS	NS	NS
	Syntho Only	0.77	0.003	0.46	0.017

[1] α – Significance level
[2] NS- Not significant

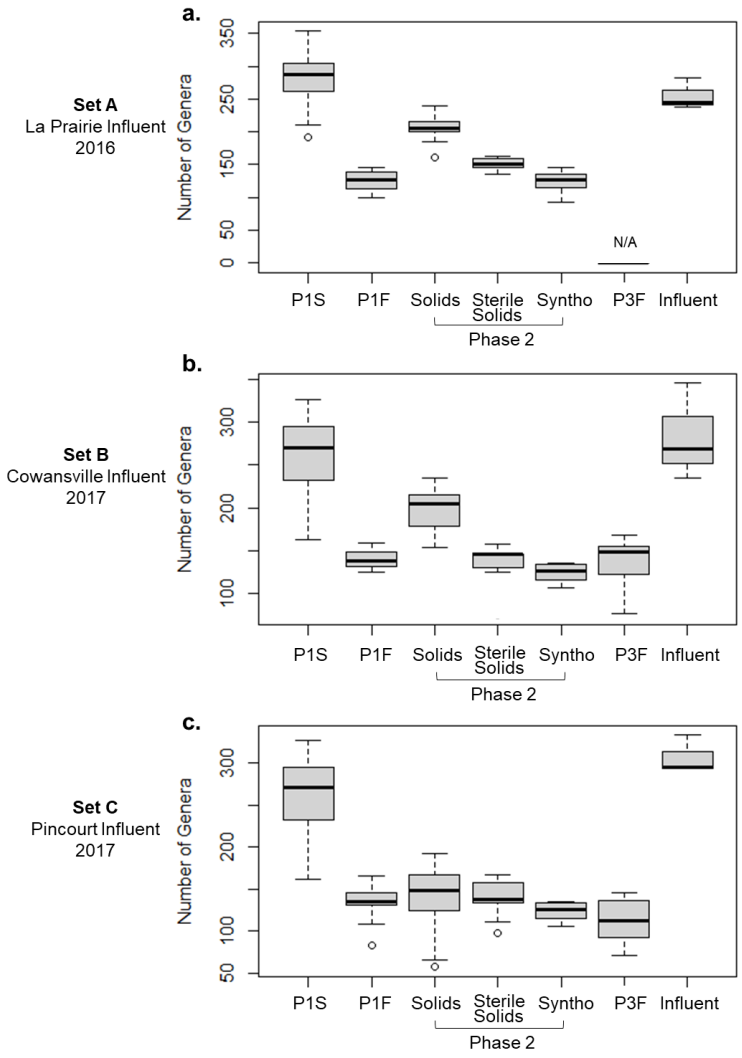


Figure S4.2- Genus richness boxplot of reactor communities in the inoculum at the beginning of the experiment (P1S; n = 9), at the end of Phase 1 (P1F; n = 27), end of Phase 2 (Solids: fed Syntho + influent solids, Sterile Solids: Syntho + autoclaved influent solids, and Syntho only; n = 9 each), end of Phase 3 for reactors receiving Syntho + influent solids (P3F; n = 9), and Influent Solids (n = 3). Phase 3 was completed for only

Table S4.4: Values used in the calculation of the relationship between $f_{16S,i,ML}$, $f_{16S,i,inf}$, and $\mu_{net,i}$

Parameter	**Symbol**	**Reactor Set**		
		A- La Prairie	B- Cowansville	C- Pincourt
Solid Retention Time (days)	θ_x	5.0	5.0	5.0
Hydraulic Retention Time (days)	θ	1.8 ± 0.28	1.8 ± 0.28	1.8 ± 0.28
Fraction of influent captured by ML solids	$f_{OHO,Capt}$	1	1	1
Total VSS of the Influent (mg-VSS/L)	$X_{Tot,Inf}$	120	120	120
Total VSS of the ML (mg-VSS/L)[1]	$X_{Tot,ML}$	470.9[2]	685.3 ± 94.22	644.4 ± 73.54
DNA extraction yield of Influent (µg-DNA/mg-VSS)[3]	$\gamma_{DNA,Inf}$	1.73 ± 0.15	0.72 ± 0.07	0.94 ± 0.13
DNA extraction yield of ML (µg-DNA/mg-VSS)[4]	$\gamma_{DNA,ML}$	1.51 ± 0.09	1.75 ± 0.004	1.06 ± 0.09

[1]Where n=18 representing the average VSS of the nine reactors receiving influent solids (with immigration) over a two week

[2]Calculated based upon COD/VSS ratio and validated with experimental data obtained during Phase 1 as no experimental were obtained for this reactor set during Phase 2

[3]Influent DNA yield where n=3 and samples taken from three different time points

[4]Mixed liquor DNA yield where n=3 for each reactor set (1 reactor per block within each set). Samples obtained from the final

± indicates the standard deviation of reported measurements

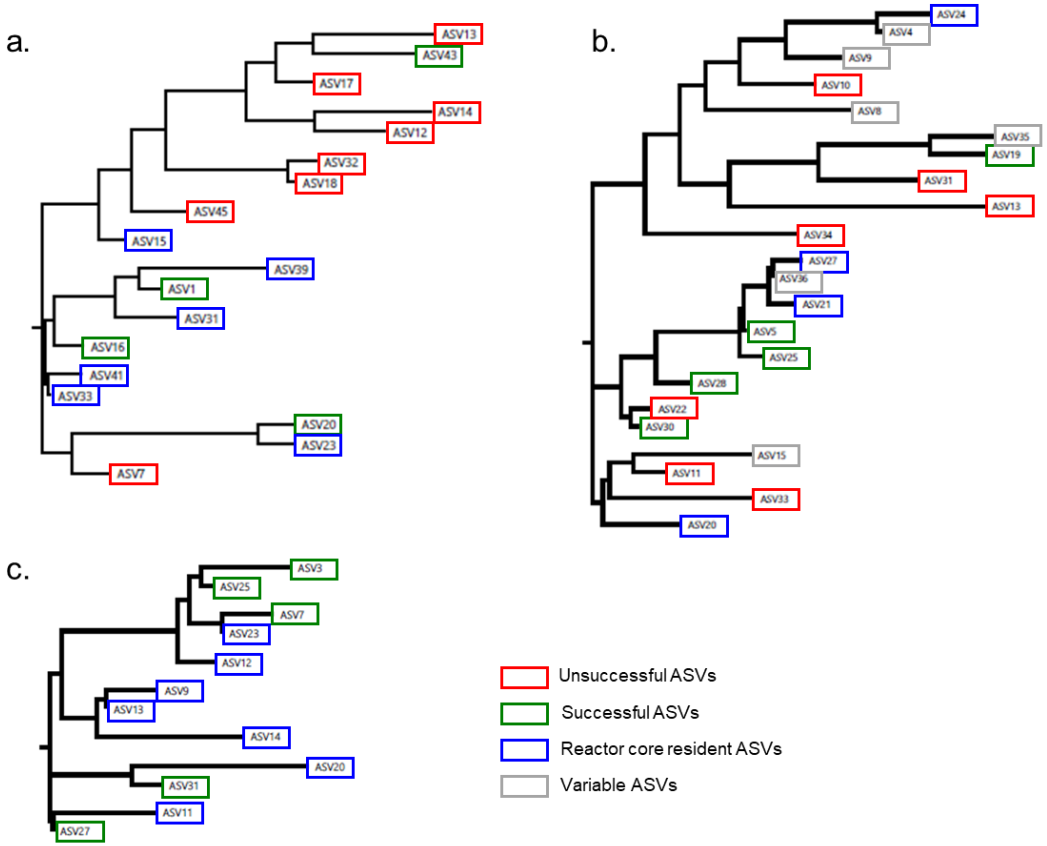


Figure S4.3- Multiple sequence alignment tree generated using BIONJ neighbour joining algorithm. a) Acinetobacter ASVs b) Pseudomonas ASVs c) Zoogloea ASVs. Unsuccessful ASVs are those that were detected in the influent but not in reactors with immigration. Successful ASVs are those detected in the influent and in the reactors with immigration only. Reactor core resident ASVs were those present in all reactors regardless of feed. Variable ASVs are those which had different fates in each reactor set (e.g. successful in one reactor set and unsuccessful in another).

4.9 References

1. Dorofeev AG, Nikolaev YA, Mardanov A V., Pimenov N V. Role of phosphate-accumulating bacteria in biological phosphorus removal from wastewater. Appl Biochem Microbiol. 2020;56(1):1–14.
2. Jiang X-T, Guo F, Zhang T. Population dynamics of bulking and foaming bacteria in a full-scale wastewater treatment plant over five years. Sci Rep. 2016;6:24180.
3. Guo B, Liu C, Gibson C, Frigon D. Wastewater microbial community structure and functional traits change over short timescales. Sci Total Environ. 2019;662:779–85.
4. Shchegolkova NM, Krasnov GS, Belova AA, Dmitriev AA, Kharitonov SL, Klimina KM, et al. Microbial community structure of activated sludge in treatment plants with different wastewater compositions. Front Microbiol. 2016;7:90.
5. Han H, Zhang Y, Cui C, Zheng S. Effect of COD level and HRT on microbial community in a yeast-predominant activated sludge system. Bioresour Technol. 2010;101(10):3463–5.
6. Alawi M, Off S, Kaya M, Spieck E. Temperature influences the population structure of nitrite-oxidizing bacteria in activated sludge. Environ Microbiol Rep. 2009;1(3):184–90.
7. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat Microbiol. 2019;4:1183–95.
8. Rocca JD, Muscarella ME, Peralta AL, Izabel-Shen D, Simonin M. Guided by microbes: Applying community coalescence principles for predictive microbiome engineering. mSystems. 2021;6(4):e00538-21.
9. Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 2016;10:11–20.
10. Mei R, Narihiro T, Nobu MK, Kuroda K, Liu W-T. Evaluating digestion efficiency in full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial activity. Sci Rep. 2016;6:34090.
11. Jauffur S, Isazadeh S, Frigon D. Should activated sludge models consider influent seeding of nitrifiers? Field characterization of nitrifying bacteria. Water Sci Technol. 2014;70(9):1526–32.
12. Lee S-H, Kang H-J, Park H-D. Influence of influent wastewater communities on temporal variation of activated sludge communities. Water Res. 2015;73:132–44.
13. Dottorini G, Michaelsen TY, Kucheryavskiy S, Andersen KS, Kristensen JM, Peces M, et al. Mass-immigration determines the assembly of activated sludge microbial communities. Proc Natl Acad Sci. 2021;118(27):e2021589118.

14. Mei R, Liu W-T. Quantifying the contribution of microbial immigration in engineered water systems. Microbiome. 2019;7:144.

15. Harris K, Parsons TL, Ijaz UZ, Lahti L, Holmes I, Quince C. Linking statistical and ecological theory : Hubbell's unified neutral theory of biodiversity as a hierarchical 131irichlet process. Proc IEEE. 2017;105(3):516–29.

16. Guo B, Manchester M, Luby T, Frigon D. Composition of heterotrophic specialized sub-guilds defined by a positive RNA and polyhydroxyalkanoate correlation in activated sludge. Water Res. 2018;144:561–71.

17. Frigon D, Wells G. Microbial immigration in wastewater treatment systems: analytical considerations and process implications. Curr Opin Biotechnol. 2019;57:151–9.

18. Guo B, Frigon D. Cellular metabolic markers and growth dynamics definition of functional groups in activated sludge wastewater treatment heterotrophic population. McGill University, Canada; 2019.

19. Isazadeh S, Jauffur S, Frigon D. Bacterial community assembly in activated sludge: mapping beta diversity across environmental variables. Microbiol Open. 2016;5:1050–60.

20. Kaewpipat K, Grady CPL. Microbial population dynamics in laboratory-scale activated sludge reactors. Water Sci Technol. 2002;46:19–27.

21. American Public Health Association. Standard methods for the examination of water and wastewater. 22nd ed. Rice EW, Baird RB, Eaton AD, Clesceri LS, editors. Washington DC; 2012.

22. Boeije G, Corstanje R, Rottiers A, Schowanek D. Adaptation of the CAS Test System and Synthetic Sewage for Biological Nutrient Removal. Chemosphere. 1999;38(4):699–709.

23. Isazadeh S, Frigon D. Biosolids minimization by partial ozonation of return activated sludge: Model development and bacterial population dynamics. McGill University, Canada; 2014.

24. Apprill A, Mcnally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75:129–37.

25. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2015;18(5):1403–14.

26. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

27. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–7.

28. McIlroy SJ, Kirkegaard RH, McIlroy B, Nierychlo M, Munk Kristensen J, Karst SM, et al. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017;1–9.

29. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. Vegan: Community Ecology Package. 2022.

30. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol. 1993;18(1):117–43.

31. Grady CPL, Daigger GT, Love NG, Filipe CDM. Biological Wastewater Treatment. 3rd ed. London: CRC Press; 2011.

32. Guo B, Liu C, Gibson C, Klai N, Lin X, Frigon D. Wastewater influent microbial immigration and contribution to resource consumption in activated sludge using taxon-specific mass-flow immigration model. bioRxiv. 2022;504022.

33. Dueholm MKD, Nierychlo M, Andersen KS, Rudkjøbing V, Knutsson S, Arriaga S, et al. MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. Nat Commun. 2022;13:1908.

34. Wittebolle L, Van Vooren N, Verstraete W, Boon N. High reproducibility of ammonia-oxidizing bacterial communities in parallel sequential batch reactors. J Appl Microbiol. 2009;107(2):385–94.

35. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Microbiology. 2018;361:469–74.

36. McCormack UM, Curião T, Metzler-Zebeli BU, Wilkinson T, Reyer H, Crispie F, et al. Seeking to improve feed efficiency in pigs through microbial modulation via fecal microbiota transplantation in sows and dietary supplementation of offspring with inulin. Appl Environ Microbiol. 2019;85(22):e01255-19.

37. Yoon YK, Suh JW, Kang E-J, Kim JY. Efficacy and safety of fecal microbiota transplantation for decolonization of intestinal multidrug-resistant microorganism carriage: beyond Clostridioides difficile infection. Ann Med. 2019;51:379–89.

38. Diaz-Colunga J, Lu N, Sanchez-Gorostiaga A, Chang CY, Cai HS, Goldford JE, et al. Top-down and bottom-up cohesiveness in microbial community coalescence. Proc Natl Acad Sci. 2022;119(6):e2111261119.

39. Smith JJ, Tow LA, Stafford W, Cary C, Cowan DA. Bacterial Diversity in Three Different Antarctic Cold Desert Mineral Soils. Microb Ecol. 2006;51(4):413–21.

40. Prombutara P, Allen MS. Flocculation-related gene identification by whole-genome sequencing of Thauera aminoaromatica MZ1T floc-defective mutants. Appl Environ Microbiol. 2016;82(6):1646–52.

CHAPTER 5

Antimicrobial Resistance Gene Sequencing is Necessary to Reveal the Complex Dynamics of Immigration from Sewers to Activated Sludge

Connecting text: In Chapter 3, influent wastewater was shown to contain several antimicrobial resistance genes (ARGs) amplicon sequence variants, which were likely associated with different microbial populations or mobile genetic elements (MGEs). In Chapter 4, influent immigration was found to contribute up to 25 % of the reads in the activated sludge, with immigrating genera typically exhibited relatively low specific net growth rates. Whilst it is known that the influent wastewater contains a plethora of ARGs, in Chapter 5 the fate and persistence of a given gene upon entry into the wastewater treatment plant (WWTP) is investigated.

The dynamics of ARGs at the interface between the influent and activated sludge are explored in Chapter 5 using samples obtained from the reactor set-ups described in Chapter 4, and the multiplexed amplicon sequencing approach developed in Chapter 3. Furthermore, droplet digital PCR was used to accurately quantify gene abundances. The results suggest that increases in the abundance of ARGs with immigration are not necessarily related to direct immigration of genes between the influent to the activated sludge. These results highlight the need for detailed studies on the ecological context of ARGs to better understand their fate and persistence in the environment.

5.1 Abstract

Microbial community composition has increasingly emerged as a key determinant of antibiotic resistance gene (ARG) content. However, in activated sludge wastewater treatment plants (AS-WWTPs), a comprehensive understanding of the microbial community assembly process and its impact on the persistence of antimicrobial resistance (AMR) remains elusive. An important part of this process is the immigration dynamics (or community coalescence) between the influent and activated sludge. While the influent wastewater contains a plethora of ARGs, the persistence of a given ARG depends initially on the immigration success of the carrying population, and the possible horizontal transfer to indigenously resident populations of the WWTP.

The current study utilised controlled manipulative experiments that decoupled the influent wastewater composition from the influent microbial populations to reveal the fundamental mechanisms involved in ARG immigration between sewers and AS-WWTP. A novel multiplexed amplicon sequencing approach was used to track different ARG amplicon sequence variants across the immigration interface, and droplet digital PCR was used to quantify the impact of immigration on the abundance of the targeted ARGs. Immigration caused an increase in the abundance of over 70 % of the quantified ARGs. However, monitoring of ARG amplicon sequence variants (ARG-ASVs) at the immigration interface revealed various immigration patterns such as (i) suppression of the indigenous mixed liquor ARG-ASV by the immigrant, or conversely (ii) complete immigration failure of the influent ARG-ASV . These immigration profiles are reported for the first time here and highlight the crucial information that can be gained using our novel multiplex amplicon sequencing techniques. Future studies aiming to reduce AMR in WWTPs should consider the impact of influent immigration in process optimisation and design.

5.2 Introduction

Antimicrobial resistance (AMR) is recognized as one of the greatest threats to public health worldwide (1). Each year 700,000 deaths are attributed to AMR and without action this number is predicted to rise to 10 million by 2050 (2). The United Nations Environment Assembly (UNEA-3) have recognised the importance of the environment in the development, spread and transmission of AMR to humans and animals (3). Of particular interest are wastewater treatment plants (WWTPs), which have been identified as hotspots of AMR (4) and gateways to the environmental spread. Although a reduction in the load is observed, the wastewater treatment process does not effectively remove all phylogenetically mobile antimicrobial resistance genes (ARGs) and antimicrobial resistant bacteria (ARB) before release into the environment (5). Consequently, effluent wastewater has been shown to contribute to antibiotic resistance in surface waters and sediments downstream of effluent discharge points (6,7). ARGs are also disseminated in the waste biosolids produced during the treatment process (8,9), which are often applied to agricultural land as fertilisers which creates another route of AMR dissemination. To minimize the spread of AMR, wastewater treatment plant design requires urgent optimization for the removal of ARB and ARGs.

With increasing knowledge of emerging contaminants in wastewater and the benefits of water resource recovery, the need for new wastewater treatment technologies is widely recognised. In this context, studies have aimed to minimise ARG release into the environment with novel design. However results are often variable. For example, some studies of anoxic-aerobic membrane bioreactor observed a reduction in the abundance of ARGs (10,11), whilst others found the abundance of genes such as *tet*C to increase (12). The use of ozonation to reduce ARG loads resulted in removal efficiencies which varied between ARG classes (13). Similarly, some found

chlorination to cause large reductions in the abundance of ARGs (14) and increases in the abundance of cell free ARGs (15), whilst others reported no impact on genes such as *bla*TEM-1 (16). These contradictory results exemplify our lack of understanding of the drivers in the persistence and proliferation of ARBs and ARGs in WWTPs, which remains one of the greatest hurdles in developing appropriate treatment strategies to reduce AMR.

Influent wastewater contains a plethora of ARB which harbour several ARGs with specific sequence amplicon sequence variants (ASVs) often associated with the genetic context of the gene (17). Each ARG-ASV is likely to obey different elimination or persistence mechanisms following their immigration into the biological treatment process from the sewer. The characterization of these mechanisms, however, requires the identification and quantification of ARG amplicon sequence variants across the influent and activated sludge (AS-WWTP) interface with high resolutions and sensitivity (17,18). Although quantitative PCR is sensitive, it is uninformative with regards to ARG sequence variants. Conversely, metagenomic shotgun sequencing can provide information on the genetic context of the most abundant variants, but it has limited capabilities in the identification of variants occurring in low abundance, and is less sensitive than qPCR for ARG detection (19). In environmental samples such as soil and activated sludge, metagenomic sequencing depth is insufficient to represent all diversity within these complex communities (20,21). Numerous studies have used data on ARG occurrence in the influent and activated sludge to infer the origin of AMR. However, more recent studies demonstrate that ARGs can vary at the sequence level based upon their origin (22). The development of specific approaches to address this knowledge gap is the main goal of the current study.

The fate of immigrating ARB and ARGs is related to the complex ecological dynamics occurring at the interface between the influent and AS-WWTP. Firstly, the persistence of a given ARG

depends on the immigration and success of its carrying population in the downstream community (23). Some ARGs may persist because the carrying population is already an indigenous resident of the AS-WWTP community. In this case, the ARG may not even be present in the influent community. Secondly, the persistence of a given ARG may also be impacted by its phylogenetic mobility. Horizontal gene transfer has been demonstrated to play a pivotal role in the spread of ARG across species (24). Although a given ARB may be unsuccessful in the AS-WWTP community, processes such as conjugation, transformation and transduction enable the movement of ARGs and increase the likelihood of ARG persistence by transferring to other, better adapted, hosts. For example, conjugative plasmids have been shown to play a significant role in facilitating the persistence of multidrug resistant ARB in WWTPs (25). To determine the contributions of these mechanisms, ARG amplicon sequence variants (ARG-ASVs) need to be tracked across the immigration interface between the influent wastewater and the activated sludge. The current study employed a novel multiplex amplicon sequencing approach (17,18) to study ARG amplicon sequence variant dynamics with the required sensitivity and resolution. Such a technique will likely provide valuable insights into the contradictory results on the persistence of ARGs.

The compounded impact of complex immigration dynamics, community composition drifts and horizontal gene transfers on the fate of ARGs are better studied in highly controlled and replicated reactor experiments. In full-scale WWTP, the influent substrate compositions, bacterial population, and ARGs vary on a daily basis (26,27). These variations, even when small, limit our ability to accurately assess the exact mechanisms behind influent immigration and AMR in AS-WWTPs. Therefore, this study utilised triplicated lab-scale reactors fed with a synthetic wastewater, which allowed for a high level of control on the immigration process and the possibility of reproducing similar conditions in subsequent manipulative experiments. Among the

operated reactors, the only varying factor was immigration, which was simulated by the addition of naturally occurring microbial communities (i.e., suspended solids) harvested in municipal wastewaters to the synthetic wastewater.

The ARG analyses presented here expand upon previous observations on the population dynamics and community assembly obtained with the same experimental set-up and published previously (23). Droplet digital PCR was used to quantify the variation in abundance of the ARGs across the immigration interface. Due to influent wastewater containing ARGs originating from several sources (e.g., clinical waste, sewer biofilms and agricultural runoff (26,28)), it was hypothesised that ARG-ASVs from the influent wastewater would be different than the ones occurring endogenously in the activate sludge. Thus, multiplex amplicon sequencing was used for ARG source tracking.

5.3 Methods

5.3.1 Samples

Samples for analysis were collected from reactors previously operated as described by Gibson et al. (Gibson et al., 2023). Briefly, small scale activated sludge (AS) reactors were inoculated with mixed liquor taken from three full scale AS-WWTPs (Figure 5.1: Block A-La Prairie AS, Block B-Cowansville AS and Block C- Pincourt AS). Reactors were operated with a hydraulic retention time of 1.8 days and a solid retention time of 5 days, and fed with a synthetic wastewater (Syntho; Boeije et al., 1999). A synthetic wastewater was utilised to ensure a stable wastewater composition was maintained over time, this could not be achieved using wastewater obtained from a full-scale WWTP which is impacted by numerous factors. Synthetic wastewater (Syntho) was produced in the laboratory using the recipe developed by Boeije et al.(1999) and autoclaved at 121 °C and 5 PSI for 30 minutes to sterilise. The main carbon sources within Syntho were sodium acetate, starch,

glycerol and milk powder. During Phase 1 of reactor operation, all reactors received Syntho only to ensure that a stable community was formed and any existing immigrants were removed. In Phase 2, to investigate the impact of immigration, in one third of the reactors the synthetic wastewater was supplemented with influent solids (i.e., an influent wastewater community) taken from a full-scale AS-WWTP. Two control groups were also included, the first was a substrate control, that received Syntho supplemented with sterile influent solids (autoclaved at 121°C, 15 psi for 30 minutes). The final 9 reactors received Syntho only to act as a continuity control. Finally, during Phase 3 of reactor operation, influent solids were removed from the feeds to determine if the impact of immigration was maintained over time.

This experiment was repeated with the use of three different sources of influent wastewater community. It was observed that the impact of immigration was similar in each set of reactors, thus only one set of reactors was selected for ARG analysis (27 reactors). The chosen reactors received influent solids from Cowansville AS-WWTP (Set B; Gibson et al., 2023**)**. For ARG analysis, samples from each Block (with the same inoculum; Block A-La Prairie AS, Block B-Cowansville AS and Block C- Pincourt AS), receiving the same feed were pooled (3 biological replicates pooled). A total of 12 samples collected during Phase 2 of reactor operation were analysed: 3 groups with immigration, 3 groups with sterilised influent solids, 3 with synthetic wastewater only, and 3 influent wastewater samples as shown in Figure 5.1.

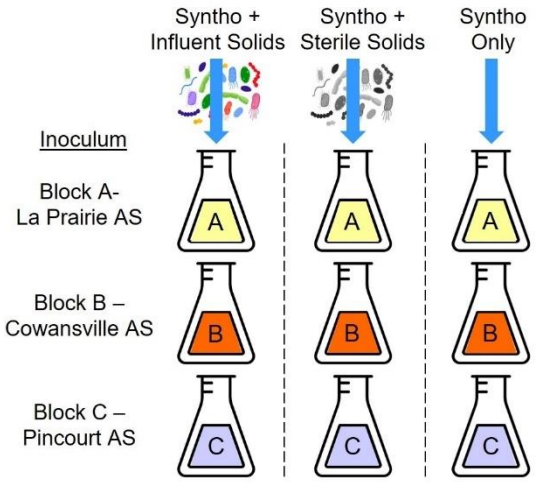


Figure 5.1- One set of Reactors which received one source of influent solids (27 reactors). Reactors were inoculated with three different activated sludge communities (Block A-C). One third of the reactors received synthetic wastewater (Syntho) supplemented with influent solid which was the group with immigration (9 reactors). Another third of the reactors received Syntho and sterile solids which acted as a substrate control (9 reactors). The final third of reactors received Syntho only and acted as a continuity control (9 reactors).

5.3.2 Sample Collection and Processing

Mixed liquor biomass samples were collected from the reactors and centrifuged in 2 mL microcentrifuge tubes for 5 mins at 16,000 × g. The supernatant was discarded and biomass was stored immediately at −80 ℃ until nucleic acid extraction (approximately 3 months). Three aliquots were stored per reactor sample. DNA was extracted from approximately 0.25 grams of each stored biomass sample using DNeasy PowerSoil Kit (Qiagen, Germantown, MD, USA) following the manufacturer's instructions with a final elution volume of 100 µL and all centrifugation steps conducted at 10,000 X g. One blank was included per batch of extractions (approximately 24 samples). Samples were extracted in duplicate, and DNA aliquots were immediately stored at −80 ℃ until future use.

The quality of nucleic acids was assessed using the ratio of absorbance at 260 nm and 280 nm (260/280) obtained using NanoDrop™ One. DNA extracts with a 260/280 value between 1.8-2.0 were considered to have good purity. Quant-iT™ PicoGreen™ dsDNA Assay kit (Invitrogen, Ontario) was used to accurately quantify the double-stranded DNA concentration of the extracts prior to droplet digital PCR.

5.3.3 Droplet Digital PCR

A total of 15 different antibiotic resistance genes were analysed using droplet digital PCR and multiplexed amplicon sequencing (Table S5.1). Clinically relevant ARGs were selected which displayed resistance to five classes of antimicrobials; beta-lactams (*bla*MOX, *bla*TEM, *bla*OXA), fluoroquinolones (*qnr*S, *qnr*B), macrolides (*dfr*A, *mph*E, *ere*A) and tetracyclines (*tet*O, *tet*Q, *tet*E) to investigate whether the impact of immigration varied between ARG classes. In addition, four multidrug resistance genes were included (*rob*A, *msr*D, *qac*L, *mar*R). Primers were designed and in silico verification was conducted by Swift Biosciences, Ann Arbor and manufactured by Integrated DNA Technologies, USA. Appropriate DNA dilutions were determined using a standard quantitative PCR reaction (Powerup SYBR green master mix) performed using three DNA dilutions to check for PCR inhibition. The fold dilution for digital droplet PCR was calculated using $Eq.1$. Samples were diluted to the appropriate concentration using DNase/RNase free water.

$$\boldsymbol{Dilution\ (fold) = 2^{24-Cq}} \qquad Eq.1$$

A 20 uL reaction mix was prepared for each droplet digital PCR reaction, which included 1 x QX200™ ddPCR™ Evagreen® Supermix (Bio-Rad, CA, USA; catalogue number 1864033), 100 nM of the forward and reverse primers, template DNA as determined using equation 1, and RNase/DNase-free water. For quality control purposes, a positive control obtained from

wastewater samples was used in each run to ensure the matrix was representative of the actual samples to be analysed. In addition, a no template control was included. Droplets were generated using the Bio-Rad QX200™ Droplet Generator and DG8™ Cartridges with 20 µL of reaction mix and 70 µL of Droplet Generation Oil for Evagreen. Following droplet generation, droplets were transferred to a 96 well plate and sealed with foil using a Bio-Rad PX1 PCR Plate Sealer. A Bio-Rad C100 Touch Thermal Cycler was used for thermal cycling of all plates. Reaction conditions used were as follows; 95 °C for 5 min., followed by 50 cycles of 95 °C for 30 sec., *x* °C for 60 sec. and 72 °C for 30 sec. whereby *x* is the annealing temperature (Table S5.1). Followed by 4 °C for 5 min., 90 °C for 5 min. and a hold at 12 °C. Annealing temperatures were optimised using a gradient PCR method on a pooled DNA sample. The optimal annealing temperature was selected based upon the separation between positive and negative droplets, and the lowest 'rain' i.e. droplets defined as neither positive or negative. The primer concentration was optimised by conducting a test assay on a representative pooled DNA sample, at various primer concentrations ranging from 100 – 250 nM. A final primer concentration of 100 nm was selected for the assays. To test for inhibitor and matrix effects, 10 fold dilutions of pooled DNA samples were quantified.

Droplets were analysed using the Biorad QX200™ Droplet Reader and the absolute quantification (ABS) experimental setup. Results were visualised using the QuantaSoft™ Software (Bio-Rad, version 1.7.4) to provide an absolute quantification of the ARG per nanogram of DNA. Reactions producing fewer than 10,000 droplets were excluded and the droplet digital PCR reaction was repeated. The threshold was defined using the auto-select function in the QuantaSoft™ software (Bio-Rad, version 1.7.4). In cases where an appropriate threshold was not automatically selected, a threshold was manually chosen. An example of a positive and negative result are provided in Figure S5.1a. All samples were analysed in duplicate in different PCR runs to confirm the technical

reproducibility. Copies of each ARG/ng-DNA were normalised to the copies of 16S rRNA gene/ng-DNA to provide a relative ARG copy number/16S rRNA gene. The limit of blank (LOB) was calculated using $Eq.2$, and $Eq.3$ (30) was used to determine the limit of detection (LOD) of samples. To confirm the calculated LOD values, assays were performed using six dilutions of a positive sample (Figure S5.1b), and the 95 % LOD was determined based upon the probability of detection at each concentration.

$$LOB = Mean_{Blank} + 1.645(SD_{Blank}) \qquad Eq.2$$

$$LOD = LOB + 1.645\left(SD_{Lowest\ sample}\right) \qquad Eq.3$$

5.3.4 Sequence diversity of ARGs

A multiplexed amplicon sequencing approach was used to analyse sequence variant diversity among the ARGs detected in the reactors. Details of method development are included in Appendix 1. ARGs were amplified using a multiplex PCR kit (Qiagen) with the following reaction conditions; 95 °C for 15 min. followed by 25 cycles of 94 °C for 30 sec., 60 °C for 90 sec. and 72 °C for 60 sec., and a final extension step at 60 °C for 30 min. Amplicons obtained from individual PCR reactions (using one set of primers per reaction) were pooled and used as a positive control. The products of the multiplex PCR reaction were purified using SPRI Select beads (Beckman Coulter) to remove primer dimers at a sample to bead ratio of 0.8. Samples were barcoded and pooled at equimolar concentration. The pooled samples were sequenced on the Illumina MiSeq PE250 platform at McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada).

ARG Amplicon sequence variants (ARG-ASVs) were defined as sequences which differed by at least one single nucleotide variant (SNP) within the amplified region of the targeted gene (around

275 base pairs). To distinguish between sequencing errors and actual sequence variants, a stringent filtering process was applied as described in Section 5.2.5.

5.3.5 Bioinformatics

Reads from each sample were split according to their forward and reverse primer sequences allowing for one sequencing error per primer sequence using a custom script available upon request. Read pairs where both or one of the reads was missing a primer sequence at the beginning or where the primer sequences found did not correspond to each other were removed. Subsequently, we used Trimmomatic v. 0.39 to trim all reads using the following settings: LEADING: 3 TRAILING: 3, SLIDINGWINDOW:4:15 and MINLEN:36. Then we used vsearch v2.13.3 to merge the trimmed read pairs before converting them to fasta format. The merged reads where then sorted and clustered with an identity of 1 and a minimum length of 100 bases while keeping track of the cluster size using the -sizeout option.

PCR and sequencing errors created many unique or very low count amplicon sequence variants that we removed using a custom R script. First, we tested if the distribution of read counts for each amplicon sequence variant produced by an individual primer pair followed a Poisson distribution (centered around the mean of all counts) and corrected the resulting p-values (subtracted from 1 such that amplicon sequence variants with high read counts had low p-values) using the Bonferroni correction. If none of the amplicon sequence variant have a corrected p-value below 0.5, we removed the whole dataset, as this indicated that we are not able to distinguish real amplicon sequence variants from those generated by PCR or sequencing errors (noise sequence variants). If any amplicon sequence variants survived the filtering, we aimed to remove the noise sequence variants from the datasets by finding the minimum of a density function based on the total data for each gene (including noise sequence variants) and removing all amplicon sequence variants with

counts less than the minimum found. By comparing with other filtering approaches (e.g., DADA2 (31) or minimum absolute value), it was found that the filtering adopted here was the most stringent, but was still able to recover the diversity of synthetic mixes of sequence variants.

We subsequently combined all reads corresponding to a specific resistance gene (including multiple primers amplifying the same gene). Files were re-replicated using a custom python script available upon request (using the Pandas module), before constructing a table of allelic distribution across samples equivalent to an OTU table for each resistance gene using vsearch v2.13.3's cluster_fast option with the otuout option. Unique consensus sequence variants were obtained using thecentroids option. To ensure that the resulting amplicon sequence variants were the products of amplification of the target ARG, we blasted all consensus sequences to a custom ARG database using the diamond (v0.9.32.133) blastx option with an e-value of 3 and a minimum identity of 80. Consensus sequences that were not similar to any entries in the ARG database were removed from the analysis.

5.3.6 Statistics

The Mann-Whitney U test was used to determine if there was a statistically significant increase in the abundance of each ARG with immigration. Samples taken from the reactors with immigration (receiving Syntho and influent solids) were compared to the control groups at a significance level of $p<0.05$.

Procrustes analysis was performed using the R Vegan package (32) to determine whether the microbial community of the reactors correlated with the ARG profile. Procrustes was used to compare the ARG and microbial community ordinations. ARG data was standardised to means of 0 and standard deviation of 1. Using the R Vegan package, Principal Component Analysis (PCA) was conducted for the ARG dataset based on the Euclidean distances. Principle coordinate analysis

(PCoA) was performed using Jaccard distance on the microbial community data, as this distance was previously shown to most effectively display the impact of immigration on the community in the current experiment (Gibson et al., 2023). The function “Protest” was used with 999 permutations to test the significance between the configurations.

5.3.7 Database Analysis of Amplicon Sequence Variants

The NCBI database (33) was used to search for previously reported hosts of the detected ASVs. To evaluate the mobility of each ASV, PLSDB (34) was used to determine whether they had been previously reported within plasmids. In this analysis, only 100 % identity and coverage results were considered as ASVs could vary by as little as a single SNP.

5.4 Results

5.4.1 Impact of Immigration on Abundance of ARGs in the Activated Sludge

Reactors were operated under highly controlled conditions to test the impact of immigration on antibiotic resistance in the activated sludge. All 27 reactors in the three Blocks (3 different inoculum) received a synthetic wastewater (Syntho) feed, which allowed the influent wastewater composition to be carefully controlled throughout reactor operation. To ensure the reactors were operating at steady state, the chemical oxygen demand and suspended solids concentrations were monitored over time and observed to be relatively constant (steady-state). After operating the reactors for 12 SRTs and ensuring steady state was reached, during Phase 2 in each Block (9 reactors in total; Block A-La Prairie AS, Block B-Cowansville AS and Block C- Pincourt AS), three reactors received synthetic wastewater with added influent solids to simulate the impact of immigration. Another three of the reactors received autoclaved influent solids and acted as a substrate control. The final three reactors received Syntho only as a continuity control to establish a baseline for the level of AMR without immigration.

The impact of influent immigration on the activated sludge microbial community is discussed at length in Gibson et al., 2023. Briefly, immigration impacted the microbial community composition of the activated sludge, and reactors with immigration became more similar to the microbial community of the influent solids. Up to 25 % of sequencing reads were observed to be contributed through influent immigration, representing a significant proportion of the activated sludge community. Using a mass balance approach, it was observed that the growing immigrant population typically exhibited a lower and often negative net growth rates in the activated sludge, when compared to the core resident genera which typically displayed a positive net growth rate. In Gibson et al., 2023, focus was placed on the impact of immigration on the microbial community composition of the activated sludge alone, whilst in this publication the impact on AMR is explored.

To determine the impact of immigration on AMR in the AS, digital droplet PCR was used to quantitatively assess the concentration of fifteen ARGs in the reactor samples and influent wastewater. Results showed that immigration caused a significant increase in the relative abundance of eleven of the fifteen ARGs in the activated sludge when compared to the control groups (sterile control and no immigration control together) (Figure 5.2; Mann Whitney U Test; $p<0.05$). Genes observed to increase in abundance were distributed in several classes of antimicrobial resistance such as ARGs against fluoroquinolones (*qnr*B and *qnr*S), beta-lactams (*bla*TEM and *bla*MOX), macrolides (*dfr*A and *mph*E), and tetracyclines (*tet*Q and *tet*O). Three out of four of the efflux pump-associated multi-antimicrobial resistance genes quantified showed a significant increase in concentration with immigration (*mar*R, *msr*D and *rob*A). Genes such as *bla*MOX were present in the reactors with and without immigration, whilst others such as *mar*R and *qnr*B were detected in the activated sludge mixed liquor only with immigration.

The relative abundances of over 70 % (11/15) of quantified ARGs increased in the activated sludge with immigration. However, they typically remained lower than those in the influent solids. To determine the impact of immigration on the overall ARG load, the absolute ARG concentrations were calculated (Table S5.2) using information on reactor operation including the volatile suspended solid concentration, yield of DNA/g-VSS, 16S rRNA gene/ng-DNA and gene abundance (16S rRNA gene) (Table S5.4, Gibson et al., 2023). It was observed that the load of genes such as *bla*OXA increased by 0.42 log gene copies/L between the influent and mixed liquor. Whilst others such as *rob*A reduced by 1.14 logs of gene copies/L between the influent and activated sludge. Consequently, despite the reduction in ARG relative abundance (copies/16S rRNA gene) between the influent solids and activated sludge, high absolute ARG concentrations remained.

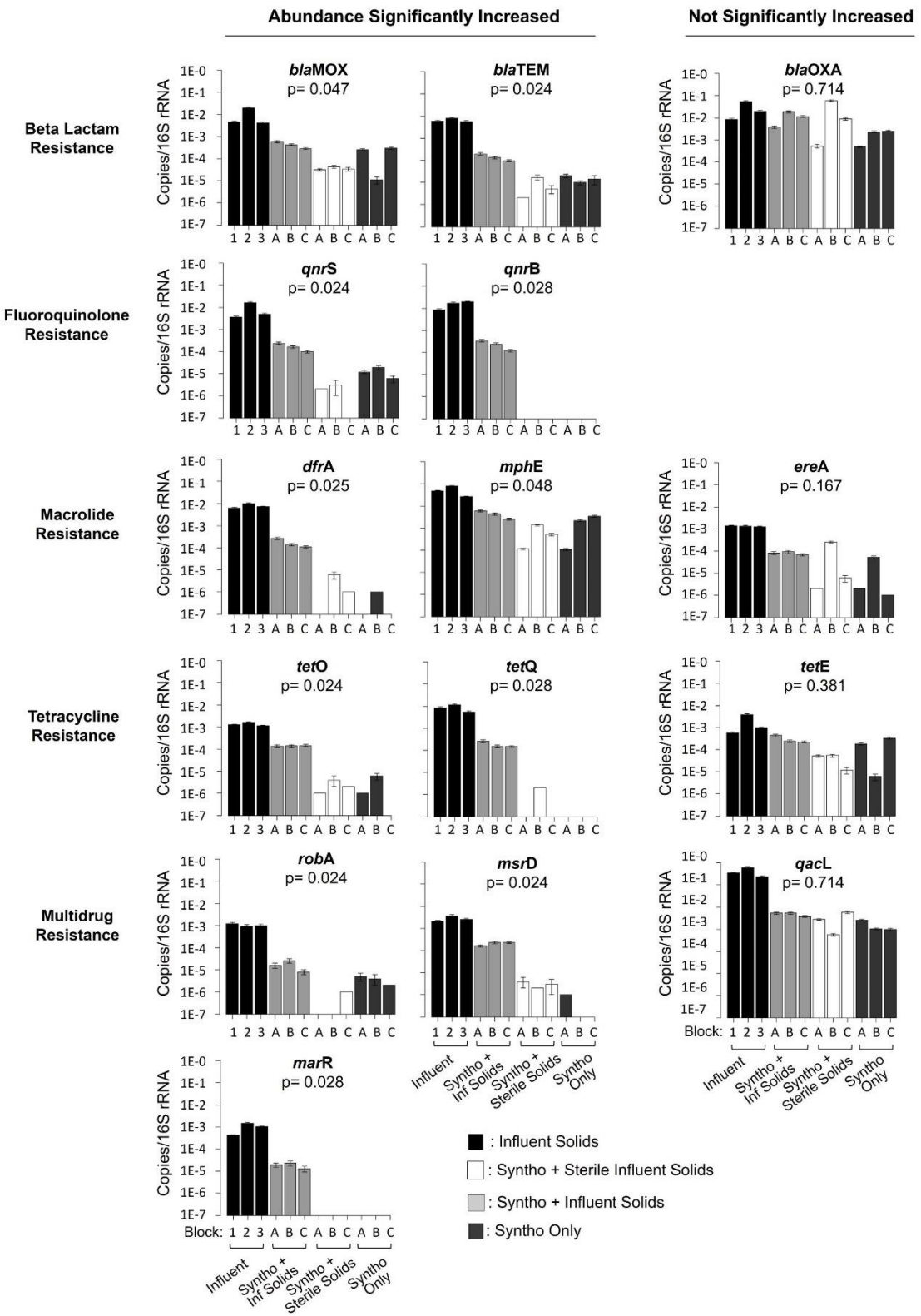


Figure 5.2- Quantification of antibiotic resistance genes using droplet digital PCR. Reactors with immigration received Syntho and influent solids. Reactors receiving Syntho and sterile influent solids acted as a substrate control. Reactors receiving Syntho only acted as a no immigration continuity control. The Block represents the inoculum source of the reactors; A- La Prairie ML, B- Cowansville ML, C- Pincourt ML, Influent samples are numbered based upon the order received by the reactors during Phase 2. Each bar represents a composite sample of three biological replicates. The Error bars represent the standard error between duplicates. The Mann-Whitney U Test was used to assess the statistical significance of the change in abundance of each ARG with immigration when compared to the sterile and no immigration control combined.

5.4.2 Correlation between the Microbial Community Composition and ARG Profile

In numerous studies, the microbial community composition has increasingly emerged as a key determinant of ARG content (35–37). In continuity with these observations, a Procrustes analysis was performed to determine whether the observed changes in the abundance of ARGs in the reactors correlated with changes in the microbial community composition with immigration. Procrustes analysis using the Jaccard distance matrix revealed a significant correlation (Procrustes $M^2 = 0.60$, $p = 0.015$) between the microbial community compositions of the reactor activated sludge and the profiles and concentrations of ARGs they carried (Figure 5.3b). However, this correlation was not significant (Procrustes $M^2 = 0.73$, $p = 0.290$; note that lower M^2 means a better correlation between the datasets) when using the Bray-Curtis dissimilarity (Figure 5.3d).

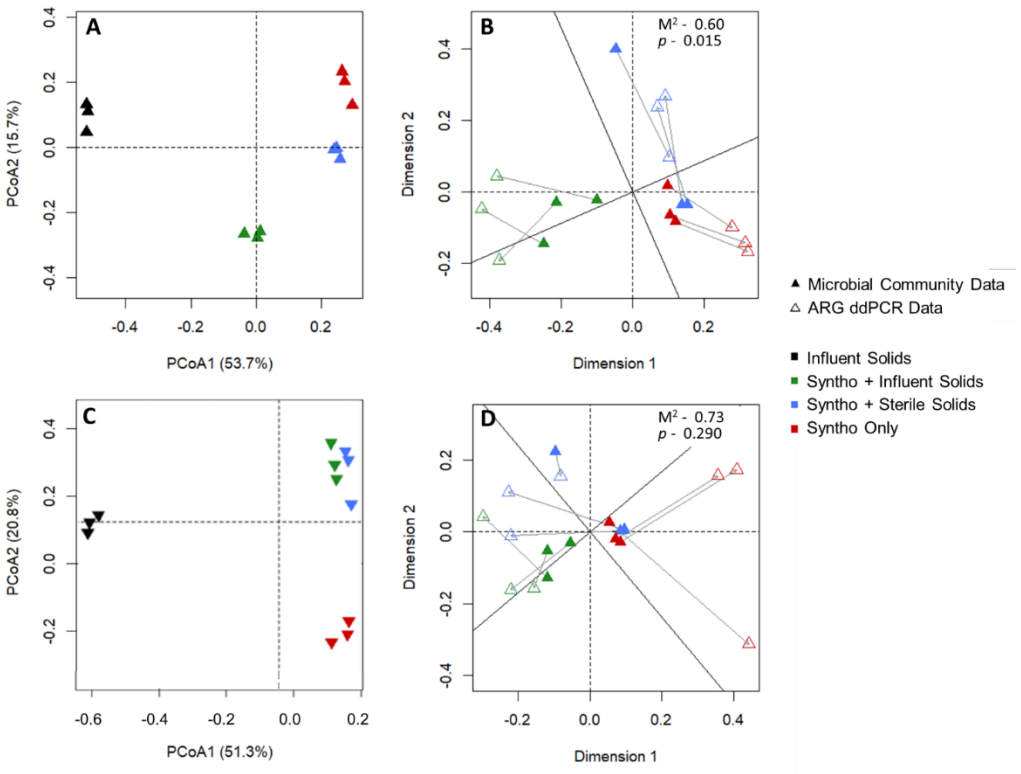


Figure 5.3- Procrustes Analysis to investigate the correlation between the microbial community and ARG composition. a) Microbial Community of the reactors visualised using Jaccard dissimilarity as determined by 16S rRNA gene sequencing b) Procrustes analysis of microbial communtiy and ARG profile with Jaccard dissimilarity c) Microbial Community of the reactors visualised using Bray Curtis dissimilarity as determined by 16S rRNA gene sequencing d) Procrustes analysis of microbial communtiy and ARG profile with Bray Curtis distance.

5.4.3 Analysis of ARG Amplicon Sequence Variants

Digital droplet PCR showed that the relative abundance of over 70 % (11 out of 15) of the ARGs investigated significantly increased with immigration. However, questions remained about the exact dynamics occurring at the interface between the influent and activated sludge. Chief among them was whether the increase in abundance of an ARG was due to the introduction of genes originating from the influent. In other words, were the observed changes in the activated sludge likely due to direct ARG immigration or other changes induced by the presence of influent solids? To investigate the immigration dynamics at greater depth, a multiplexed amplicon sequencing approach was used to detect ARG-ASVs within the influent and reactor samples. By utilising short read sequences, sequence diversity within a specific region of each ARG (around 275 base pairs) could be used as a marker to track the movement of genes between the influent and activated sludge.After stringent filtering based on the Poisson distribution, ARG-ASV information was obtained for eleven of the fifteen ARGs analysed (Figure 5.4). The four remaining targets (*msr*D, *mar*R, *qnr*B and *qnr*S) were undetected after the filtering process either because they did not survive the stringent filtering process (*msr*D and *qnr*B), or insufficient reads were obtained resulting in no detection (*qnr*S and *mar*R). Amplicon sequencing of the eleven ARGs revealed different ASV distributions or concurrence profiles between the influent and the activated sludge mixed liquors (Figure 5.4). In several cases, the ARG-ASVs were specific to either the influent or reactors, with few examples of concurrence in the reactors and the received influent solids. This demonstrates that PCR based methods over-simplified the ARG dynamics at the interface between the influent and activated sludge.

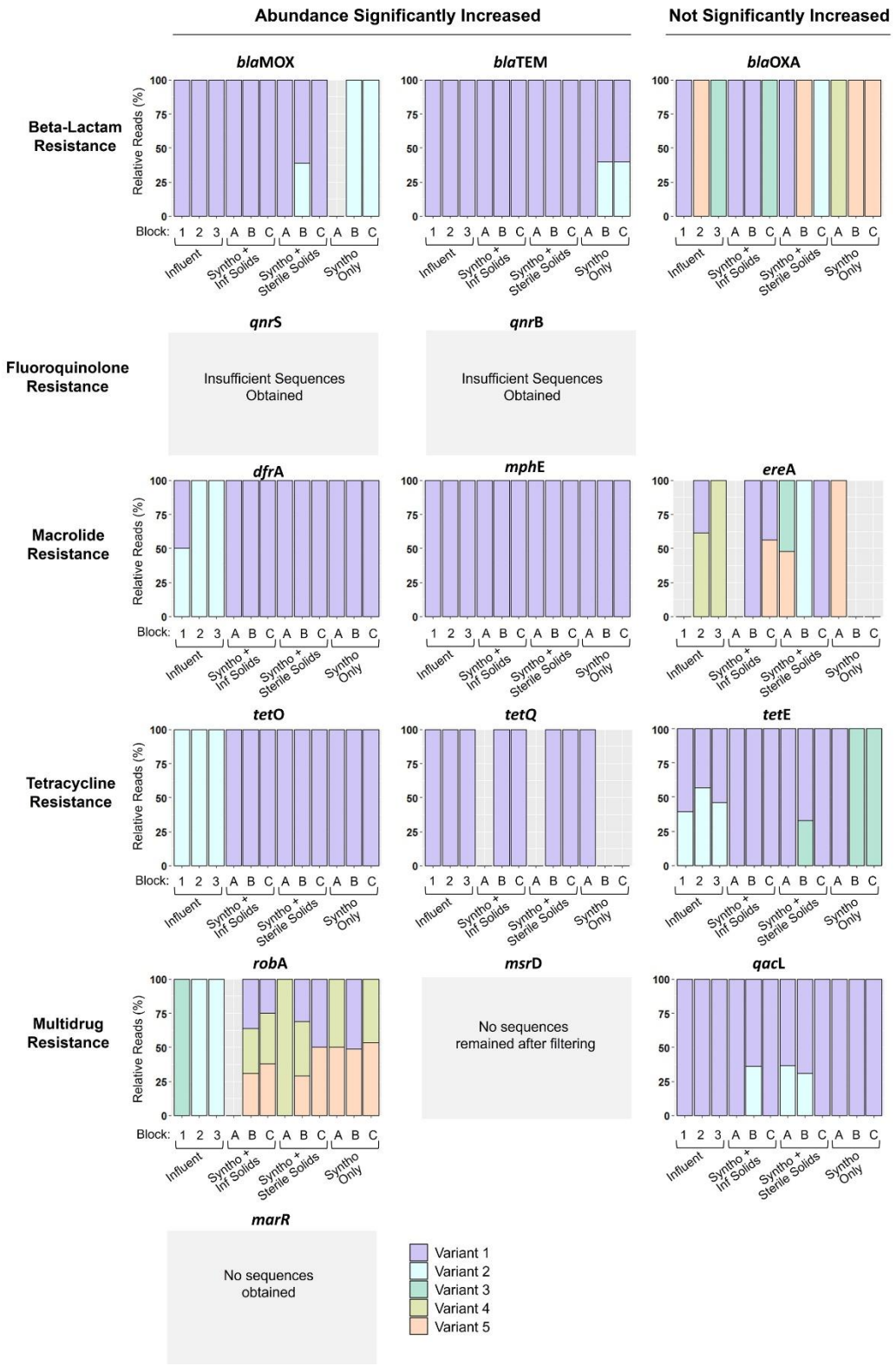


Figure 5.4 - ARG sequence variants detected using multiplexed amplicon sequencing. Reactors with immigration received Syntho and Influent (Inf) Solids. Reactors receiving Syntho and Sterile Influent Solids acted as a substrate control. Reactors with Syntho only acted as a no immigration control. Each sample represents a pool of biological triplicates. Block represents the inoculum of the reactor A- La Prairie Mixed Liquor, B- Cowansville Mixed Liquor, C-Pincourt Mixed Liquor. Influent 1, 2 and 3 represent the order in which they were fed to the reactors during Phase 2, with Influent 3 received for the final SRT. All reactor samples analysed were obtained from the final day of Phase 2.

The concurrence profiles between the influent and the activated sludge reactors could be classified into 6 similarity groups that were typically confined either to the ARGs that significantly increased in abundance with influent immigration (Groups 1 to 3) or to ARGs with abundances that statistically remained unchanged (Groups 4 to 6; Figure 5.4). As summarised in Table 5.1, *Group 1* consists of ARGs for which the same and single ASV was observed in all samples (genes *mph*E, and *tet*Q). This concurrence profile is consistent with direct immigration causing the significant increases in the abundances of these genes. *Group 2* collates ARGs with an ASV observed only in the control reactor samples, but not in the live immigration reactor or the influent samples (gene *bla*MOX and *bla*TEM). This concurrence profile is also consistent with direct immigration overwhelming the detection of the ARG-ASV observed in the control reactor. *Group 3* concurrence profiles are characterized by ARG-ASVs observed only in the influent, but not in the reactors (genes *dfr*A, *tet*O, *rob*A). This profile appears to follow a counter selection dynamic, whereby one or more of the influent sequence ASVs did not successfully immigrate between the influent and activated sludge. This demonstrated that the presence of a given ARG-ASV in the influent and the significant increase of an ARG with immigration, does not definitively predict the presence of the ASV within the activated sludge as could have been inferred using ddPCR results alone.

The *Group 4* profile which was only observed for *tet*E appeared to be a mixture of Group 2 and 3 profiles, suggesting that these two mechanisms could be at interplay. Group 3 and 4 ARGs exemplify the complexity of interactions at the interface between the influent and activated sludge. Considering the Procrustes results (Figure 5.3), the increase in the abundances of these ARGs could be associated with the increase in relative abundance of the bacterial populations that carry these ARGs instead (Group 3) or simultaneously (Group 4) to direct ARG immigration.

The last two profiles of concurrences were observed for ARGs that did not significantly increase in abundance (Figure 5.4). *Group 5* profile was only observed for *qac*L and showed ASVs that were only observed in some samples from reactors receiving influent solids (both live and autoclaved). These ASVs were in low abundances in these samples. Thus, they may have been below the detection limit in the influent or control reactors, and immigration made them detectable in some circumstances possibly due to small variations in community compositions as in Groups 3 and 4. Finally, the genes of *Group 6* (*bla*$_{\text{OXA}}$ and *ere*A) harboured 5 ASVs each present seemingly at random in samples from the influent or the reactors under different experimental conditions. Consequently, no clear profiles could be identified, and immigration did not appear to affect these genes.

5.4.4 Database Information on the Genetic Context of ASVs

To investigate the genetic context of ARG-ASVs, the NCBI database and PLSDB (34) were utilised to gain information about previously reported hosts and gene occurrence on plasmids (Table S5.3). To exemplify this approach and establish hypotheses for future work, genes from group 2-5 (Table 5.1) were analysed, as unique ASVs were detected in either the reactors or influent samples.

In group 2, analysis of the *bla*TEM ARG using the NCBI database revealed the ARG-ASVs to have previously been reported in similar bacterial hosts (Table S5.3), which included taxa from the genus *Acinetobacter*, *Enterobacter* and *Klebsiella* to name a few. The *bla*TEM ASV-1 was present in numerous variants of the *bla*TEM gene (over 60) including TEM-29, TEM-169 and TEM-26 whilst ASV-2 was present in only TEM-229, TEM-116, TEM-162, TEM-181 and TEM-157. Analysis of the two ARG-ASVs in PLSDB revealed both to have been previously reported on plasmids. Plasmids containing ARG-ASV 1 have been detected in numerous settings including

in humans, animals and the environment (water). Whilst plasmids containing *bla*TEM ASV-2 have been reported in fewer settings, which included soil, feed additives and *Orcytes gigas* (rhinoceros beetle).

Table 5.1: Concurrence profiles of ARG sequence variants in the influent and reactor samples

Group	Genes	Description	Conclusions
1	*mph*E, *tet*Q	Same single variant observed in all samples	Consistent with direct immigration
2	*bla*MOX, *bla*TEM	Unique variant present in control reactors that is not detected with immigration or in the influent solids	Direct immigration and suppression dynamics
3	*dfr*A, *tet*O, *rob*A	Unique variants present in the influent solids only	Counter Selection
4	*tet*E	Both 2 and 3: Unique variants present in the control groups and influent solids	Mixture of direct immigration, counter selection and suppression dynamics
5	*qac*L	Unique variant only present in reactors receiving influent solids (live and autoclaved)	Possible direct immigration of variants occurring below detection limit
6	*bla*OXA, *ere*A	Numerous variants distributed throughout the influent and reactor samples	Variants randomly distributed

In group 3, analysis of the ASVs of the *dfr*A gene found that *dfr*A ASV-1 corresponded to the *dfr*A-5 gene, and ASV-2 the *dfr*A-14 variant of the gene (38). ASV-1 (*dfr*A-5) and ASV-2 (*dfr*A-14) are phylogenetically closely related (39) and are commonly observed within integrons and on plasmids. As with the *bla*TEM ASVs, both *dfr*A ASVs have been previously reported in similar host including taxa from the genus *Aeromomas, Enterobacter* and *Citrobacter*.

In group 4, similar selection patterns were observed with the *tet*E ARG (Figure 5.5c). The influent solids contained two ASVs of the *tet*E ARG. ASV-1 was detected in the reactors even without

immigration, whilst ASV-2 did not successfully immigrate. The NCBI database returned no matches with 100% identity for sequence variant 1. The hosts of ASV-2 appeared to somewhat overlap with those of ASV-3 that was detected within the reactors. Interestingly, it was observed that ASV-2 was primarily reported to occur on plasmids in the NCBI database (33), whilst ASV-3 was more commonly chromosomal.

5.5 Discussion

5.5.1 Immigration Impacted ARG Relative Abundance in the Activated Sludge

Droplet digital PCR revealed that the relative abundance of 11 of the 15 ARGs quantified increased with immigration. The previous analysis of the community composition data of the current reactor experiment demonstrated that the impact of immigration was better visualised using the Jaccard dissimilarity than the Bray-Curtis dissimilarity, suggesting that immigration impacts the composition of lower-abundance taxa in the mixed liquor (Gibson et al., 2023).Consistent with this result, Procrustes analysis revealed a significant correlation between the microbial community composition and ARG profiles only when visualised with Jaccard dissimilarity. Taken together, this suggests that the changes in the ARG profiles observed is due to immigration of taxa mainly occurring at low abundances within the activated sludge.

Whilst the concentrations of the majority of ARGs increased with immigration, others such as *bla*OXA and *qac*L did not significantly change (Figure 5.2). The presence of these genes in similar abundance under all reactor conditions suggests that they are likely associated with the core resident populations, which were defined as taxa that occurred under all reactor conditions (i.e., independent of immigration).

5.5.2 Dynamics of ARG Amplicon Sequence Variants between Influent and Activated Sludge

Immigration impacted the concentration and diversity of ARGs in the activated sludge. However, it remained to be seen whether the increase in the abundance of ARGs with immigration was due to direct ARG transport or other changes associated with influent immigration. Recent studies have demonstrated ARG-ASVs to occur in samples originating from different sources (22). Therefore, it was hypothesised that the influent wastewater, which is strongly influenced by anthropogenic activities (26), could contain different ARG-ASVs compared to the activated sludge. The detection of unique ARG-ASVs would enable a more detailed analysis of the immigration dynamics between the influent and activated sludge.

Based upon the neutral model and the assumption of ecological equivalence (i.e., equal fitness of competitors), ASVs detected in the influent would be expected to immigrate into the activated sludge (40). In both group 3 and 4, unique ASVs were observed in the influent, which could have been introduced into the reactors with immigration. However, in only one instance was ARG-ASV immigration unequivocally observed in the current study (Figure 5.4). Among the *bla*MOX gene, a new ASV was introduced into the reactors with immigration (Figure 5.4). With the introduction of this sequence ASV, the indigenous *bla*MOX sequence ASV-2, which was present without immigration, was no longer detected, suggesting either competition among the hosts of these ARG-ASVs or simply that the low concentration of the indigenous resident ASV was overwhelmed by the immigrant ASV. The *bla*MOX ASV-1 was also detected in the reactors receiving sterile influent solids, suggesting that the gene persisted in the influent solids and activated sludge reactors after autoclaving. Counter selection of ARG-ASVs between the influent and activated sludge was more frequently observed, suggesting that ARG immigration is likely not dictated

solely by neutral processes and is impacted by other factors. This conclusion is supported by recent studies which have demonstrated that in WWTP samples obtained from five different countries, the resistome composition varies between the influent and activated sludge (41).

Counter selection of ARG ASVs was frequently observed between the influent and activated sludge, which suggested that direct immigration of ARGs was not occurring between the two environments. Despite this, significant increases in the relative abundances of many of these ARGs were observed in the activated sludge with immigration. Supported by the results from Procrustes analysis which showed a positive correlation between the microbial community and ARGs, this suggests that the relative abundance was instead impacted by the microbial community changes caused by immigration. This is consistent with previous studies that have identified the microbial community composition to be the main driver of AMR content in numerous environments (42,43).

The results from ARG amplicon sequencing demonstrate the complexity of ARG dynamics at the interface between the influent and activated sludge. Given the frequent observation of counter selection of ARG-ASVs, PCR based approaches alone cannot be used to accurately predict the AMR the activated sludge or infer the origin of ARGs. These data also support the contrasting results reported in the literature, which are likely influenced by the methods used and targets selected. Future studies should consider the use of amplicon sequencing approaches to enable a more accurate assessment of ARG dynamics and source tracking.

5.5.3 Genetic Context of ARG Amplicon Sequence Variants

Given the differences in the environmental conditions and microbial community between the influent and mixed liquor, counter selection of ARG-ASVs was not surprising. The successful immigration of a given ARG-ASV is likely associated with the ability of the host to adapt and establish itself under different environmental conditions, or the mobility of the gene that may be

an indicator of the likelihood of transfer to hosts better adapted to the reactor environment. To further investigate these factors, ARG-ASVs were analysed using the NCBI database and PLSDB (34).

In group 2, analysis of the *bla*TEM sequences revealed the two ARG-ASVs to have been previously reported in similar bacterial hosts and both to occur in plasmids (Table S5.3).Within the influent and reactor samples *bla*TEM ASV-1 was the most frequently observed. Plasmids containing this ASV have been previously detected in a diverse range of environments (34). Given that both *bla*TEM ASVs were detected among similar hosts, it could be suggested that the persistence of ASV-1 is linked to the ability of plasmids containing this ASV to move between environments.

In group 3, using multiplexed amplicon sequencing two ASVs of the *dfr*A gene were identified. The *dfr*A ASV-1 corresponded to the *dfr*A-5 gene, and ASV-2 the *dfr*A-14 variant of the gene. A 2011 review of class 1 and 2 integrons in pathogenic Gram-negative bacteria identified *dfr*A-5 (ASV-1) in chicken and pig samples, as well as estuarine and carriage water. In the same study, the dfrA-14 gene was reported in pig, cattle and chicken samples, but not in aquatic settings (44). More recently, *dfr*A-14 has been observed in surface waters (45) and WWTP effluent (46). The widespread detection of both ASV-1 and ASV-2 in clinical, agricultural, and environmental settings does not support the theory of selection due to niche differences between the influent and reactors. However, both ASV-1 and 2 have been observed in similar hosts, suggesting competition may occur resulting in the loss of ASV-2 (Table S5.3).

In group 4, *tet*E ASV-2 appeared most frequently in plasmids, ASV-3 was more commonly reported to be chromosomal. Previous studies have demonstrated chromosomal mutations to carry a larger fitness cost than plasmid acquired resistance (47), however, ASV-3 was found to persist

in the reactors and ASV-2 was undetected. ASV-2 has been associated with numerous plasmids found in water environments including PC1579, a conjugative plasmid recently reported to carry a novel Metallo-β-lactamase gene (48), and plasmid pWLK-NDM, which was identified in environmental isolates carrying *bla*NDM-1 and *bla*KPC-2 resistance genes (49). Given the previous reports of ASV-2 in aquatic environments, it would be expected that ARB carrying this gene would be capable of growing within the reactor environment. However, ASV-2 remained undetected in the activated sludge suggesting again that competition for resources may be occurring.

5.5.4 Development of Multiplexed Amplicon Sequencing

Multiplexed amplicon sequencing of ARGs is a relatively new technique requiring optimisation to ensure all primer pairs work successfully in tandem (17,18). Overall, eleven of the fifteen ARG targets produced data on ASV. Of the primers not successfully producing ARG-ASV information, this was often due to low levels of amplification. Considering the stringent Poisson distribution-based filters applied to the amplicon sequencing data, low count ASVs may also have been excluded during data processing. Future method developments should optimise primer design to ensure sufficient sequencing reads are obtained for all positive targets to improve resolution.

Future development of the multiplexed amplicon sequencing approach should also carefully consider which region of the ARG sequence to target. Among group 1 of concurrence profiles, low sequence diversity was found in the targeted region of the gene. The multiplexed amplicon sequence targets used herein with Illumina MiSeq protocols include only a small section of the ARG sequence (around 275 bp). Consequently, sequence diversity in other regions may have been missed. By considering regions of high genetic diversity in future design, this technique could be optimised to cover mutations of concern for example those impacting the phenotypic resistance.

5.5.5 Future Application of ARG Sequence Diversity Analysis

The monitoring of ARG ASVs at the immigration interface revealed various immigration patterns such as (i) suppression of the indigenous activated sludge ASV by the immigrant, or conversely (ii) counter selection and complete immigration failure of the influent ASV. These immigration profiles are reported for the first time here and highlight the crucial information that can be gained using our multiplex amplicon sequencing techniques.

Unique ASVs were observed among the influent and reactor samples, which highlights the potential for amplicon sequencing approaches to be utilised in the future for ARG source tracking purposes. To enable this, widespread sampling of reservoirs of antimicrobial resistance should be considered to identify ARG ASV markers present in different environments. Compared to techniques such as full metagenome sequencing, the novel amplicon sequencing approach applied here is relatively low cost and produces data that is more easily managed. In the future, such techniques could be applied for source tracking of ARG contamination in the environment and assessment of potential ARG mobility.

5.6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5.7 Author Contributions

Claire Gibson designed the study, conducted the digital droplet PCR, developed the multiplexed amplicon sequencing approach, analysed the data and prepared the manuscript. Susanne A. Kraemer developed the multiplexed amplicon sequencing approach, developed the bioinformatics pipeline, processed the sequencing data and reviewed the manuscript. Natalia Klimova developed

the multiplexed amplicon sequencing approach and prepared samples for sequencing. Dominic Frigon obtained funding, supervised the research and revised the manuscript.

5.8 Acknowledgments

We would like to thank Sophie Zhang, Julia Qi, Carlos Vasquez Ochoa, Nouha Klai, Zeinab Bakhshijooybari and Shameem Jauffur for their assistance with reactor operation. We would also like to acknowledge the assistance of the operators and staff at Cowansville, La Prairie and Pincourt wastewater treatment plants for access to the facilities and help with sampling. Finally, we are indebted to Biorad for the training provided and loan of the digital droplet PCR equipment.

5.9 Supplementary Material

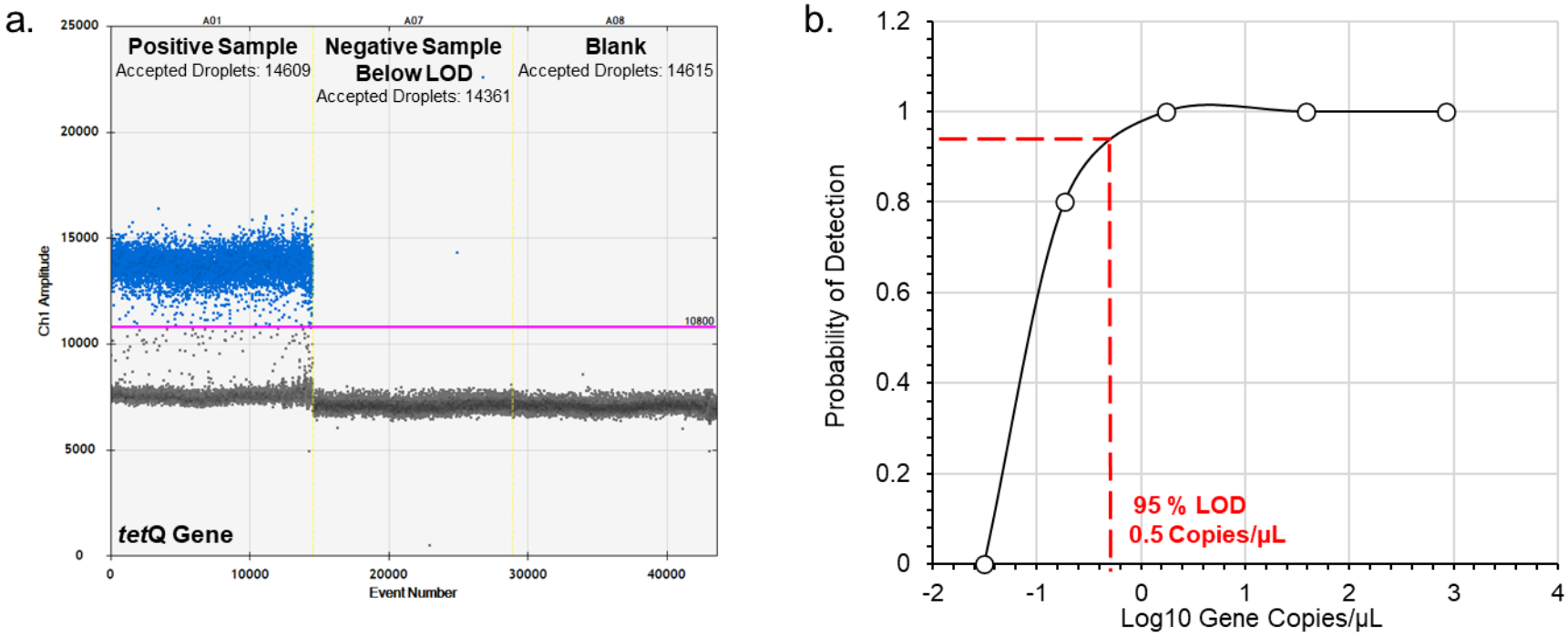


Figure S5.1- Digital droplet PCR for ARG detection a) Example of QuantaSoft™ droplet fluorescence analysis demonstrating a positive result, a result below the limit of detection and a no template control (blank) b) Determining the 95 % limit of detection for the *tet*Q ARG. Each sample was analysed 6 times in 2 different PCR runs (3 samples per run).

Table S5.1- Primers used for droplet digital PCR

Target	Forward Sequence (5'-3')	Reverse Sequence (5'-3')	Annealing Temp. (°C)	Resistance Drug Class[1]
*bla*MOX	ACCAGCTCGGCGGATCTG	GAGCCGGTCTTGTTGAAGAGC	65	Beta-Lactam
*bla*OXA	AGGCACGATAGTTGTGGCAGAC	GTAGAATTCCGCATTGCTGATCGC	65	Beta-Lactam
*bla*TEM	GAACCGGAGCTGAATGAAGCC	CGGGAGGGCTTACCATCTGG	65	Beta-Lactam
*dfr*A	ACATACCCTGGTCCGCGAAAG	CGCCACCAGACACTATAACGTGA	65	Diaminopyrimidines
*ere*B	CAGCTCATCGATCACCTCATGAAACCG	CACGTACGGAAGTATCTCCCTCAA	64	Macrolide
*mar*R	CACAGTTTAAGGTGCTCTGCTCTATCC	GCAAATACTCAAGTGTTGCCACTTCG	63	Multi-drug
*mph*E	AAGTGAGCAATTGGAAACCCGCTA	AGGCCGCTGCTCTTTCTAAAGTC	65	Macrolide
*msr*D	GGCAAGCTAGGTGTTGAGCAATTAG	TCCTTCACGGTCTAAATGGCTCGTA	65	Multi-drug
*qac*L	GTTGCAATCTTTGGCGAGGTCA	CGCTGACCTTGGATAGCAGGTTTAGAAC	63	Multi-drug
*qnr*B	CGACCTGAGCGGCACTGAATTTA	GCTCGCCAGTCGAAAGTCGAA	65	Fluoroquinolones
*qnr*S	ATGCCAGCTTGCGATGGCAAA	GTGGCATAAATTAGCACCCTGTAGGC	65	Fluoroquinolones
*rob*A	TCAAATGCGCGTGCAGTTCTGG	GTAGCGCTCAATATCCTGACCTTTAC	63	Multi-drug
*tet*E	TGATTGCTGGACCAGTCATTGG	CCATACGAAGCGCTCTTCTCC	64	Tetracycline
*tet*O	GCAGGGACAGAACTATTAGAGCCATATC	GCTAACTTGTGGAACATATGCCGAAC	64	Tetracycline
*tet*Q	TGGATTGAAGACCCGTCTTTGTCC	AGCAGGTGTACTTACCGGGCTATA	65	Tetracycline

Table S5.2- Log change in ARG between the influent and mixed liquor

Gene	Log Copies/16S rRNA		Log Copies/L		Log change
	Influent	Reactors	Influent	Reactors	
*bla*MOX	-2.14 + 0.31	-3.38 + 0.13	7.43	6.91	-0.52
*bla*TEM	-2.20 + 0.06	-3.87 + 0.13	7.26	6.41	-0.85
*bla*OXA	-1.68 + 0.41	-2.03 + 0.29	7.89	8.31	0.42
*qnr*S	-2.18 + 0.32	-3.81 + 0.16	7.37	6.48	-0.89
*qnr*B	-1.86 + 0.15	-3.67 +0.18	7.62	6.63	-0.99
*dfr*A	-2.11 + 0.10	-3.79 + 0.16	7.35	6.50	-0.85
*mph*E	-1.34 + 0.15	-2.39 + 0.15	8.16	7.90	-0.26
*ere*A	-2.88 + 0.01	-4.09 + 0.05	6.58	6.18	-0.40
*tet*O	-2.88 + 0.05	-3.86 + 0.01	6.58	6.41	-0.17
*tet*Q	-2.09 + 0.07	-3.74 + 0.11	7.38	6.53	-0.84
*tet*E	-2.89 + 0.41	-3.54 + 0.13	6.71	6.74	0.04
*rob*A	-2.99 + 0.07	-4.95 + 0.13	6.47	5.33	-1.14
*msr*D	-2.57 + 0.12	-3.69 + 0.07	6.89	6.58	-0.31
*qac*L	-0.43 + 0.12	-2.32 + 0.07	9.06	7.92	-1.11
*mar*R	-3.06 + 0.27	-4.75 + 0.10	6.44	5.52	-0.92

Log copies calculated by multiplying copies 16S rRNA gene/ng-DNA, Yield DNA/VSS, reactor VSS and relative ARG concentration. These values are available in Table S4 of Gibson et al., 2023.

Table S5.3- Sequence Variants Obtained from Targeted Amplicon Sequencing (Figure 4 & 5)

Gene	Allele Sequence	NCBI BLAST/CARD Result
*bla*MOX Variant 1	>blaMOX_1 CTGAAGTTTGTCGGCGCCAACATGA CAGGCACCGGGGACGAGGCGATGC AGCAGGCGATTGCCCTGACCCACAA GGGGGTTTACTCGGTGGGTGCCATG ACTCAGGGGCTCGGCTGGGAGAGTT ACGCCTATCCCGTGACCGAAGAGAC CTTGCTTGCAGGCAACTCGGGCAAG GTGATCCTCGAGGCCAACCCGACGG CGCCCGCCTCCAACGAGACGGGTAG CCAGGT	*Aeromonas media* Uncultured bacterium
Variant 2	> blaMOX _2 CTGCGCTTTGTGAAGGCCAACATCA GCGGGGTGGATAATGCGGCCATGCA GCAGGCCATCGATCTGACTCACCAG GGCCAGTATGCGGTGGGGGAGATG ACCCAGGGACTGGGCTGGGAGCGTT ACCCCTATCCCGTCAGCGAGCAGAC GCTGCTGGCGGGCAACTCCCCGGCG ATGATTTACAATGCCAACCCGGCGG CGCCCGCGCCCGCTGCGGCAGGGCA CCCTGT	No 100% identity found Highest percentage identity reported: 99.13 % Query Cover: 100 %
*tet*E Variant 1	>tetE_1 TGGTTTCCGATCTTGATTGCTGGACC AGTCATTGGTGGTTTTGCAGGTCAA CTTTCGGTACAGGCACCGTTTATGTT CGCTGCTGCCATTAACGGGCTGGCA TTTCTGGTCTCCCTATTCATTTTACA TGAGACCCATAATGCTAATCAGGTT AGTGACGAGATAAAGAATGAAACA ATCAATGAAACCACATCCTCCATAC GCGAGATGATCTCCCCATTATCGGG ATTGCTAGTTGTCTTTTTCATCATTC AATTGATTGGCCAAATCCCTGCAAC ATTATGGGTTTTATTC	No 100% identity found Highest percentage identity reported: 100 % Query Cover: 95 %
Variant 2	> tetE_2 TGGTTTTGCAGGTCAACTTTCGGTA CAGGCACCGTTTATGTTCGCTGCTG CTATTAACGGGCTGGCATTTCTGGT CTCCCTATTCATTTTACATGAGACCC ATAATGCTAATCAGGTTAGTGACGA GATAAAGAATGAAACAATCAATGA AACCACATCCTCCATACGCGAGATG ATCTCCCCATTATCGGGATTGCTAG TTGTCTTTTTCATCATTCAATTGATT GGCCAAATCCCTGCAACATTATGGG TTTTATTC	*Aeromonas caviae* *Aeromonas hydrophilia* *Aeromonas media* *Aeromonas salmonicida* *Aeromonas veronii* *Raoultella ornithinolytica* *Vibrio alginolyticus* *Vibrio parahaemolyticus*

Gene	Allele Sequence	NCBI BLAST/CARD Result
*tet*E (continued) Variant 3	>tetE _3 TGGTTTTGCCGGTCAACTTTCGGTAC AGGCACCGTTTATGTTCGCTGCTGC TATTAACGGGCTGGCATTTCTGGTC TCCCTATTCATTTTACATGAGACCCA TAATGCTAATCAGGTTAGTGACGAG TTAAAGAATGAAACAATCAATGAA ACCACATCCTCCATACGCGAGATGA TCTCCCCATTATCGGGATTGTTAGTT GTCTTTTTCATCATTCAATTGATTGG CCAAATCCCCGCAACATTATGGGTT TTATTC	*Aeromonas caviae* *Aeromonas media* *Aeromonas veronii* *Aeromonas sp.* *Enterobacter cloacae* *Yersinia ruckeri* *Escherichia coli* *Aeromonas hydrophila* *Aeromonas dhakensis*
*bla*TEM Variant 1	>blaTEM_1 ATACCAAACGACGAGCGTGACACC ACGATGCCTGCAGCAATGGCAACA ACGTTGCGCAAACTATTAACTGGCG AACTACTTACTCTAGCTTCCCGGCA ACAATTAATAGACTGGATGGAGGC GGATAAAGTTGCAGGACCACTTCTG CGCTCGGCCCTTCCGGCTGGCTGGT TTATTGCTGATAAATCTGGAGCCGG TGAGCGTGGGTCTCGCGGTATCATT GCAGCACTGGGG	*Acinetobacter baumannii* *Acinetobacter johnsonii* *Acinetobacter towneri* *Aeromonas hydrophila* *Aeromonas veronii* *Bacillus subtilis* *Bacteroides fragilis* *Chlamydia trachomatis* *Citrobacter amalonaticus* *Citrobacter freundii* *Citrobacter koseri* *Citrobacter portucalensis* *Citrobacter werkmanii* *Citrobacter youngae* *Clostridioides difficile* *Cronobacter sakazakii* *Enterobacter asburiae* *Enterobacter chengduensis* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei* *Enterobacter roggenkampii* *Enterobacter spp.* *Enterobacteriaceae* *Enterococcus faecium* *Escherichia albertii* *Escherichia coli* *Escherichia fergusonii* *Escherichia marmotae* *Haemophilus influenzae* *Haemophilus parainfluenzae* *Kingella kingaegi* *Klebsiella aerogenes* *Klebsiella huaxiensis* *Klebsiella michiganensis* *Klebsiella oxytoca* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Leclercia adecarboxylata* *Morganella morganii* *Mycobacterium tuberculosis* *Neisseria gonorrhoeae* *Pasteurella multocida* *Proteus mirabilis* *Proteus vulgaris* *Providencia rettgeri* *Providencia stuartii* *Pseudomonas aeruginosa* *Pseudomonas putida* *Raoultella planticola* *Salmonella enterica* *Serratia marcescens* *Shigella boydii* *Shigella dysenteriae* *Shigella flexneri* *Shigella sonnei* *Staphylococcus aureus* *Streptococcus suis* *Vibrio cholerae* *Vibrio parahaemolyticus*

Gene	Allele Sequence	NCBI BLAST/CARD Result	
*bla*TEM (continued) Variant 2	>blaTEM _2 ATACCAAACGACGAGCGTGACACC ACGATGCCTGTAGCAATGGCAACAA CGTTGCGCAAACTATTAACTGGCGA ACTACTTACTCTAGCTTCCCGGCAA CAATTAATAGACTGGATGGAGGCGG ATAAAGTTGCAGGACCACTTCTGCG CTCGGCCCTTCCGGCTGGCTGGTTT ATTGCTGATAAATCTGGAGCCGGTG AGCGTGGGTCTCGCGGTATCATTGC AGCACTGGGG	*Acinetobacter baumannii* *Acinetobacter haemolyticus* *Bacillus cereus* *Bacillus subtilis* *Bacillus velezensis* *Bifidobacterium longum* *Burkholderia cepacia* *Burkholderia lata* *Chlamydia trachomatis* *Clostridioides difficile* *Clostridium botulinum* *Cronobacter sakazakii* *Enterobacter cloaceae* *Enterococcus faecium* *Escherichia coli* *Faecalibacterium prausnitzii* *Helicobacter pylori* *Klebsiella michiganensis* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae*	*Leclercia adecarboxylata* *Legionella pneumophila* *Mycobacterium tuberculosis* *Mycoplasma mycoides* *Neisseria meningitidis* *Propionibacterium freudenreich* *Pseudomonas aeruginosa* *Rhizobium leguminosarum* *Ruthenibacterium lactatiforman* *Salmonella enterica* *Staphylococcus aureus* *Staphylococcus hominis* *Staphylococcus saprophyticus* *Streptococcus agalactiae* *Streptococcus lutetiensis* *Streptococcus pneumoniae* *Vibrio parahaemolyticus*
*dfr*A Variant 1	>blaTEM _1 GGAGAGCAGCTACTCTTTAAAGCCT TGACGTACAACCAGTGGCTTTTGGT GGGCCGCAAGACGTTCGAATCTATG GGAGCACTCCCTAATAGGAAATACG CGGTCGTTACTCGCTCAGCCTGGAC GGCCGATAATGACAACGTAATAGTA TTCCCGTCGATCGAAGAGGCCATGT ACGGGCTGGCTGAACTCACCGA	*Acinetobacter baumannii* *Aeromonas caviae* *Citrobacter koseri* *Comamonas testosteroni* *Enterobacter hormaechei* *Enterobacter roggenkampii* *Escherichia albertii* *Escherichia coli* *Escherichia fergusonii* *Klebsiella michiganensis* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae*	*Proteus mirabilis* *Proteus vulgaris* *Vibrio cholerae* *Pseudomonas aeruginosa* *Salmonella enterica* *Serratia marcescens* *Shigella boydii* *Shigella flexneri* *Shigella sonnei*
Variant 2	>blaTEM _2 GGGGAGCAGCTACTTTTTAAAGCAT TGACCTACAATCAGTGGCTTCTGGT GGGTCGCAAGACGTTTGAATCTATG GGCGCACTCCCCAATAGGAAATACG CGGTCGTTACCCGCTCAGGTTGGAC ATCAAATGATGACAATGTAGTTGTA TTTCAGTCAATCGAAGAGGCCATGG ACAGGCTAGCTGAATTCACCGG	*Aeromonas hydrophila* *Aeromonas veronii* *Citrobacter amalonaticus* *Citrobacter freundii* *Citrobacter portucalensis* *Citrobacter werkmanii* *Citrobacter youngae* *Enterobacter asburiae* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei* *Enterobacter roggenkampii* *Escherichia albertii* *Escherichia coli* *Escherichia fergusonii* *Klebsiella michiganensis* *Klebsiella grimontii* *Klebsiella pneumoniae* *Klebsiella quasipneumoniae*	*Morganella morganii* *Proteus mirabilis* *Proteus vulgaris* *Providencia rettgeri* *Providencia stuartii* *Pseudomonas aeruginosa* *Raoultella planticola* *Salmonella enterica* *Serratia marcescens* *Shewanella putrefaciens* *Shigella boydii* *Shigella dysenteriae* *Shigella flexneri* *Shigella sonnei* *Vibrio cholerae*

Gene	Allele Sequence	NCBI BLAST/CARD Result	
*bla*OXA Variant 1	>blaOXA_1 GAACGCCAAGCGGATCGTGCCATGT TGGTTTTTGATCCTGTGCGATCGAA GAAACGCTACTCGCCTGCATCGACA TTCAAGATACCTCATACACTTTTTGC ACTTGATGCAGGCGCTGTTCGTGAT GAGTTCCAGATTTTTCGATGGGACG GCGTTAACAGGGGCTTTGCAGGCCA CAATCAAGACCAAGATTTGCGATCA GCAATGCGGAATTCTACAGATCGGA AGAGCACACGTCT	*Acinetobacter baumannii* *Aeromonas caviae* *Alcaligenes faecalis* *Burkholderia cenocepacia* *Citrobacter freundii* *Citrobacter koseri* *Citrobacter portucalensis* *Citrobacter werkmanii* *Enterobacter asburiae* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei* *Enterobacter roggenkampii* *Escherichia coli* *Klebsiella michiganensis* *Klebsiella oxytoca*	*Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Pasteurella multocida* *Proteus mirabilis* *Providencia rettgeri* *Providencia stuartii* *Pseudomonas aeruginosa* *Pseudomonas monteilii* *Pseudomonas putida* *Pseudomonas stutzeri* *Salmonella enterica* *Serratia marcescens* *Shigella sonnei* *Stenotrophomonas maltophilia* *Vibrio cholerae*
Variant 2	>blaOXA _2 GAACGCCAAGCGGATCGTGCCATGT TGGTTTTTGATCCTGTGCGATCGAA GAAACGCTACTCGCCTGCATCGACA TTCAAGATACCTCATACACTTTTTGC ACTTGATGCAGGCGCTGTTCGTGAT GAGTTCCAGATTTTTCGATGGGACG GCGTTAACAGGGGCTTTGCAGGCCA CAATCAAGACCAAGATTTGCGACCA GCAATGCGGAATTCTACAGATCGGA AGAGCACAC	No 100% identity found Highest percentage identity reported: 99.54 % Query Cover: 92 %	
Variant 3	>blaOXA _3 GAACGCCAAGCGGATCGTGCCATGT TGGTTTTTGATCCTGTGCGATCGAA GAAACGCTACTCGCCTGCATCGACA TTCAAGATACCTCATACACTTTTTGC ACTTGATGCAGGCGCTGTTCGTGAT GAGTTCCAGATTTTTCGATGGGACG GCGTTAACAGGGGCTTTGCAGGCCA CAATCAAGACCAAGATTTGCGATCA GCAATGCGGAATTTAC	No 100% identity found Highest percentage identity reported: 99.54 % Query Cover: 100 %	
Variant 4	>blaOXA _4 GAACGCCAAGCGGATCGTGCCATGT TGGTTTTTGATCCTGTGCGATCGAA GAAACGCTACTCGCCTGCATCGACA TTCAAGATACCTCATACACTTTTTGC ACTTGATGCAGGCGCTGTTCGTGAT GAGTTCCAGATTTTTCGATGGGACG GCGTTAACAGGGGCTTTGCAGGCCA CAATCAAGACCAAGATTTGCGATCA GCGAT	*Pseudomonas aeruginosa* *Pseudomonas putida* *Stenotrophomonas maltophilia*	

Gene	Allele Sequence	NCBI BLAST/CARD Result	
*bla*OXA (continued Variant 5	>blaOXA _5 GAACGCCAAGCGGATCGTGCCATGT TGGTTTTTGATCCTGTGCGATCGAA GAAACGCTACTCGCCTGCATCGACA TTCAAGATACCTCATACACTTTTTGC ACTTGATGCAGGCGCTGTTCGTGAT GAGTTCCAGATTTTTCGATGGGACG GCGTTAACAGGGGCTTTGCAGGCCA CAATCAAGACCAAGATTTT	*Acinetobacter baumannii* *Aeromonas caviae* *Alcaligenes faecalis* *Burkholderia cenocepacia* *Citrobacter freundii* *Citrobacter koseri* *Citrobacter portucalensis* *Citrobacter werkmanii* *Enterobacter asburiae* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter kobei* *Enterobacter roggenkampii* *Escherichia coli* *Klebsiella michiganensis* *Klebsiella oxytoca*	*Klebsiella pneumoniae* *Klebsiella quasipneumoniae* *Pasteurella multocida* *Proteus mirabilis* *Providencia rettgeri* *Providencia stuartii* *Pseudomonas aeruginosa* *Pseudomonas monteilii* *Pseudomonas putida* *Salmonella enterica* *Serratia marcescens* *Shigella sonnei* *Stenotrophomonas maltophilia* *Vibrio cholerae*
*qac*L Variant 1	>qacL_1 CACGACGCTCTTCCGATCTCTGTTTC AATCTTTGGCGCGGTCATCGCAACT TCCGCACTGAAGTCTAGCCATGGAT TCACTAGGTTAGTTCCTTCCGTTGTA GTTGTGGCTGGCTACGGGCTTGCGT TCTATTTCTTGTCTCTCGCGCTCAAG TCCATTCCGGTCGGTATTGCTTACGC TGTATGGGCTGGGCTTGGCATCGTG CTTGTGGCAGCTATTGCTTGGATTTT CCATGGCCAAAAACTAGACTTCTGG GCGTTCATTGGCATGGGACTTAT	No 100% identity found Highest percentage identity reported: 99.61 % Query Cover: 93 %	
Variant 2	>qacL _2 TCGCAACTTCCGCACTGAAGTCTAG CCATGGATTCACTAGGTTAGTTCCTT CCGTTGTAGTTGTGGCTGGCTACGG GCTTGCGTTCTATTTCTTGTCTCTCG CGGTCAAGTCCATTCCGGTCGGTAT TGCTTACGCTGTATGGGCTGGGCTT GGCATCGTGCTTGTGGCAGCTATTG CTTGGATTTTCCATGGCCAAAAACT AGACTTCTGGGCGTTCATTGGCATG GGACTTAT	Uncultured prokaryote	
*mar*R Variant 1	>marR _1 GCTGTGCGGGCTGCATTACCCCGGT TGAACTGAAAAAAGTGTTGTCTGTC GATCTCGGCGCCTTAACGCGCATGC TCGAGCGTCTGGTCTGCAAAGGCTG GATTGACAGACTGCCTAACCCACAT GACAAACGCGGTGTGCTGGTGAAAC TCACCGAACACGGCGCGGCAATTTG TGAGCAATGTCATCAATTAGTAGGA CAAGACCTGCACCAGGAATTAACAA AAAACTTAACGGCGGA	*Citrobacter* *Citrobacter sp.* *Citrobacter amalonaticus*	

Gene	Allele Sequence	NCBI BLAST/CARD Result	
*tet*Q Variant 1	>tetQ_1 CTTTTCCATAAACTCATATAGTGAT GAATTGGAAATCTCGTTATATGGTT TGACCCAAAAGGAAATCATACAGA CATTGCTGGAAGAACGATTTTCCGT AAAGGTCCATTTTGATGAGATCAAG ACTATCTACAAAGAACGACCTATAA AAAAGGTCAATAAGATTATTCAGAT CGAAGTACCACCCAACCCTTACTGG GCCACAATAGGGCTGACTCTTGAAC CCTTACCGTTAGGGGCAGGGTTGCA AATCGAAAGTGACATCTCCTATGGT TATCTGAACCATTCTTTTCAAAATGC CGTTTTTGAAGGGATTCGTATGTCTT GCCAATCTGGTTTACATGGATGGGA AGTGACAGATCTGAAAGTAACTTTT ACTCAAGCCGAGTAT	*Alistipes communis* *Alistipes onderdonkii* *Bacteroides caccae* *Bacteroides caecimuris* *Bacteroides cellulosilyticus* *Bacteroides dorei* *Bacteroides eggerthii* *Bacteroides fragilis* *Bacteroides ovatus* *Bacteroides salyersiae* *Bacteroides sp.* *Bacteroides thetaiotaomicron* *Bacteroides uniformis* *Bactroides xylanisolvens* *Bacteroides zhangwenhongii* *Butyricimonas faecalis* *Coprobacter secundus* *Odoribacteraceae bacterium*	*Parabacteroides distasonis* *Parabacteroides goldsteinii* *Parabacteroides johnsonii* *Parabacteroides merdae* *Parabacteroides sp.* *Paraprevotella xylaniphila* *Phocaeicola dorei* *Phocaeicola vulgatus* *Prevotella buccalis* *Prevotella intermedia* *Prevotella melaninogenica* *Prevotella ruminocola* *Prevotella sp.* *Pseudoprevotella muciniphila* *Riemerella anatipestifer* *Sodaliphilus pleomorphus* *Uncultured bacterium*
*mph*E Variant 1	>mphE_1 CAGAAAATGGTTGGATAATGATGTT CTATGGGCAGATTTCACCCAATTTA TACATGGCGATTTATATGCTGGGCA TGTACTAGCTTCAAAGGATGGAGCT GTTTCAGGCGTTATTGATTGGTCAA CAGCCCATATAGATGACCCAGCGAT TGATTTTGCTGGGCATGTAACTTTGT TTGGAGAAGAAAGCCTCAAAACTCT AATCATCGAGTATGAAAAACTAGGG GGTAAAGTTTGGAATAAACTATATG AACAGACTTTAGAAAGAGCAGCGG CCTAGATCGGAA	No 100 % identity found Highest percentage identity reported: 100 % Query Cover: 96 %	
*ere*A Variant 1	>ereA_1 CACGTTGATATGCTGACTCACTTGTT GGCGTCCATTGATGGCCAGTCGGCG GTTATTTCATCGGCAAAATGGGGGG AGCTAGAAACGGCTCGGCAGGAGA AAGCTATCTCAGGGGTAACCAGATT GAAGCTCCGCTTGGCGTCGCTTGCC CCTGTCCTGAAAAAACACGTCAACA GCGATTTGTTCCGAAAAGCCTCTGA TCGAATAGAGTCGATAGAGTATACG TTGGAAACCTTGCGTATAATGAAAA CTTTCTTCGATGGTACCTCTC	*Providencia huaxiensis* *Pandoraea sp.* *Comamonas sp.* *Vogesella fluminis* *Uncultured bacterium* *Vogesella perlucida* *Comamonas koreensis*	

Gene	Allele Sequence	NCBI BLAST/CARD Result
*ere*A Variant 2	>ereA_2 CACGTTGATATGCTGACTCACTTGTT GGCGTCCATTGATGGCCAGTCGGCG GTTATTTCATCGGCAAAATGGGGGG AGCTAGAAACGGCTCGGCAGGAGA AAGCTATCTCAGGGGTAACCAGATT GAAGCTCCGCTTGGCATCGCTTGCC CCTGTCCTGAAAAAACACGTCAACA GCGATTTGTTCCGAAAAGCCTCTGA TCGAATAGAGTCGATAGAGTATACG TTGGAAACCTTGCGTATAATGAAAA CTTTCTTCGATGGTACCTCTC	No 100% identity found Highest percentage identity reported: 99.63 % Query Cover: 100 %
Variant 3	>ereA _3 CACGTTGATATGTTGACTCACTTGTT GGCGTCCATTGATGGCCAGTCGGCG GTTATTTCATCGGCAAAATGGGGGG AGCTAGAAACGGCTCGGCAGGAGA AAGCTATCTCAGGGGTAACCAGATT GAAGCTCCGCTTGGCGTCGCTTGCC CCCGTCCTGAAAAAACACGTCAACA GCGATTTGTTCCGAAAAGCCTCTGA TCGAATAGAGTCGATAGAGTATACG TTGGAAACCTTGCGTATAATGAAAA CTTTCTTCGATGGTACCTCTC	*Achromobacter denitrificans* *Enterobacter hormaechei* *Escherichia coli* *Helicobacter pylori* *Klebsiella pneumoniae* *Pseudomonas aeruginosa* *Salmonella enterica*
Variant 4	>ereA _4 CACGTTGATATGCTGACTCACTTGTT GGCGTCCATTGATGGCCAGTCGGCG GTTATTTCATCGGCAAAATGGGGGG AGCTAGAAACGGCTCGGCAGGAGA AAGCTATCTCAGGGGTAACCAGATT GAAGCTCCGCTTGGCGTCGCTTGCC CCTGTACTGAAAAAACACGTCAACA GCGATTTGTTCCGAAAAGCCTCTGA TCGAATAGAGTCGATAGAGTATACG TTGGAAACCTTGCGTATAATGAAAA CTTTCTTCGATGGTACCTCTC	*Escherichia coli* *Pseudomonas aeruginosa* *Salmonella enterica* *Serratia marcescens* *Thauera humireducens* *Uncultured bacterium*
*ere*A Variant 5	>ereA _5 CACGTTGATATGCTGACTCACTTGTT GGCGTCCATTGATGGCCAGTCGGCG GTTATTTCATCGGCAAAATGGGGGG AGCTAGAAACGGCTCGGCAGGAGA AAGCTATCTCAGGGGTAACCAGATT GAAGCTCCGCTTGGCGTCGCTTGCC CCTGTACTGAAAAAACACGTCAACA GCGATTTGTTCCGAAAAGCCTCTGA TCGAATAGAATCGATAGAGTATACG TTGGAAACCTTGCGTATAATGAAAA CTTTCTTCGATGGTACCTCTC	*Aeromonas hydrophila* *Escherichia coli* *Klebsiella pneumoniae* *Klebsiella oxytoca* *Laribacter hongkongensis* *Proteus mirabilis* *Proteus terrae* *Proteus vulgaris* *Providencia rettgeri* *Pseudoalteromonas sp.* *Salmonella enterica* *Vibrio alginolyticus* *Vogesella perlucida*

Gene	Allele Sequence	NCBI BLAST/CARD Result
*rob*A	>robA_1	
Variant 1	CACGATTTTCTCGGCAACGCGCCGA CCATTCCGCCAGTGCTCTACGGCCT GAATGAAACGCGTCCGAGTCAGGAT AAAGACGACGAACAAGAGGTATTC TATACCACCGCGTTAGCCCAGGATC AGGCAGATGGCTATGTACTGACGGG GCATCCGGTGATGCTGCAGGGCGGC GAATATGTGATGTTTACCTATGAAG GTCTGGGAACCGGCGTGCAGGAGTT TATCCTGACGGTATACGGAACGTGC ATGCCAATGCTCAACCTGACGCGCC	*Escherichia coli* *Shigella dysenteriae* *Shigella flexneri*
Variant 2	>robA_2 CGCGACTTCCTGAGCCATGCCCCGG CGATCCCGCCTATCCTGTATGGTCTC AACGAAACGCATCCAAGCCAGGAA AAAGACGACGAGCAGGAGGTGTTC TACACCACCGCGTTAACGCCGGAAA TGGCCAATGGCTATATTCAGGGCTC TAAACCGGTCGTGCTGGAAGGCGGA GAGTACGTGATGTTCTCCTACGAAG GGCTGGGAACGGGCGTACAGGAAT TCATCCTGACCGTTTACGGAACATG CATGCCTATGCTGAACCTGAATCGC C	*Enterobacter cloacae*
Variant 3	>robA_3 CGCGACTTCCTGAGCCACGCACCGG CGATCCCGCCTATTCTGTATGGTCTC AACGAAACGCACCCGAGCCAGGAA AAGGACGACGAGCAGGAGGTGTTC TACACCACCGCGCTGACGCCAGAGA TGGCCAATGGCTACATTCAGGGTTC AAAACCTGTCGTGCTGGAAGGCGGT GAATACGTGATGTTCGCCTATGAAG GGCTGGGAACGGGCGTTCAGGAGTT CATCCTGACCGTTTACGGAACCTGC ATGCCGATGCTGAATCTGAATCGCC	*Enterobacter cloacae*
Variant 4	>robA_4 CACGATTTTCTCGGCAACGCGCCGA CCATTCCGCCAGTGCTCTACGGCCT GAATGAAACGCGTCCGAGTCAGGAT AAAGACGACGAACAAGAGGTATTC TATACCACTGCGTTAGCCCAGGATC AGGCAGATGGCTATGTACTGACGGG GCATCCGGTGATGCTGCAGGGCGGC GAATATGTGATGTTTACCTATGAAG GTCTGGGAACCGGCGTGCAGGAGTT TATCCTGACAGTATACGGAACGTGC ATGCCAATGCTCAATCTGACGCGCC	*Escherichia coli*

Gene	Allele Sequence	NCBI BLAST/CARD Result	
Variant 5	>robA_5 GCACGATTTTCTCGGCAACGCGCCG ACCATTCCGCCGGTGCTCTACGGCC TGAACGAAACGCGTCCGAGTCAGG ATAAAGACGACGAACAAGAGGTAT TCTATACCACCGCGTTAGCCCAGGA TCAGGCAGATGGCTATGTACTGACG GGGCATCCGGTGATGCTGCAGGGCG GCGAATATGTGATGTTTACCTATGA AGGTCTGGGAACCGGCGTGCAGGA GTTTATCCTGACGGTATACGGAACG TGCATGCCAATGCTCAACCTGACGC GCC	*Escherichia coli*	
*tet*O Variant 1	>tetO_1 TCCACTTTGAAATTTATGCACCGCAGGA ATATCTCTCACGGGCGTATCATGATGCT CCAAGGTATTGTGCAGATATTGTA AGTACTCAGATAAAGAATGACGAGGTCA TTCTGAAAGGAGAAATCCCTGCTAGATG TATTCAAGAATACAGGAACGATTT AACTAATTTCACAAATGGGCAGGGAGTC TGCTTGACAGAGTTAAAAGGATACCAGC CAGCTATTGGTAAATTTATTTGCC AACCCCGCCGCCCGAATAGCCGTATAGA TAAG	*Bifidobacterium thermophilum* *Campylobacter jejuni* *Lactobacillus johnsonii* *Riemerella anatipestifer* *Streptococcus galloyticus* *Streptococcus phage* *Streptococcus suis* *Uncultured bacterium*	
Variant 2	>tetO_2 TCCACTTTGAAATTTATGCACCGCAGGA ATATCTCTCACGGGCGTATCATGATGCT CCAAGGTATTGTGCAGATATTGTA AGTACTCAGATAAAGAATGACGAGGTCA TTCTGAAAGGAGAAATCCCTGCTAGATG TATTCAAGAATACAGGAACGATTT AACTTATTTCACAAATGGGCAGGGAGTC TGCTTGACAGAGTTAAAAGGATACCAGC CAGCTATTGGTAAATTTATTTGCC AACCCCGCCGCCCGAATAGCCGTATAGA TAAG	*Actinobacillus pleuropneumoniae* *Anaerostipes rhamnosivorans* *Bifidobacterium breve* *Bifidobacterium pseudocatenulatum* *Blautia massiliensis* *Blautia sp.* *Campylobacter coli* *Campylobacter jejuni* *Clostridiales genomosp,* *Clostridioides difficile* *Coprococcus comes* *Eggerthella lenta* *Emergencia timonensis* *Enterocloster bolteae* *Enterocloster clostridioformis* *Enterococcus cecorum* *Enterococcus faecalis* *Enterococcus gallinarum*	*Erysipelotrichaceae bacterium* *Eubacterium sp.* *Faecalibacterium prausnitzii* *Glaesserella parasuis* *Lachnospiraceae bacterium* *Roseburia intestinalis* *Streptococcus dysgalactiae* *Streptococcus porcinus* *Streptococcus pyogenes* *Streptococcus suis* *Uncultured Blautia sp.* *Uncultured Dorea sp.* *Uncultured Eubacteriales bacterium* *Clostridium hylemonae* *Clostridium innocuum* *Clostridium scindens* *Clostridium symbiosum*

5.10 References

1. Abadii ATB, Rizvanov AA, Haertlé T, Blatt NL. World Health Organization report: Current crisis of antibiotic resistance. Bionanoscience. 2019;9:778–88.

2. O'Neill J. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. London: Wellcome Trust; 2014. Available from: https://wellcomecollection.org/works/rdpck35v/items. [Accessed on 2018 Nov 27].

3. United Nations Environment Programme (UNEP). Environmental Dimensions of Antimicrobial Resistance: Summary for Policymakers. 2022. Available from: https://wedocs.unep.org/bitstream/handle/20.500.11822/38373/antimicrobial_R.pdf. [Accessed on 2022 May 27]

4. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ. 2013;447:345–60.

5. Lapara TM, Burch TR, McNamara PJ, Tan DT, Yan M, Eichmiller JJ. Tertiary-treated municipal wastewater is a significant point source of antibiotic resistance genes into Duluth-Superior Harbor. Environ Sci Technol. 2011;45(22):9543–9.

6. Quintela-Baluja M, Abouelnaga M, Romalde J, Su JQ, Yu Y, Gomez-Lopez M, et al. Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. Water Res. 2019;162:347–57.

7. Reichert G, Hilgert S, Alexander J, Rodrigues de Azevedo JC, Morck T, Fuchs S, et al. Determination of antibiotic resistance genes in a WWTP-impacted river in surface water, sediment, and biofilm: Influence of seasonality and water quality. Sci Total Environ. 2021;768:144526.

8. Gao P, Munir M, Xagoraraki I. Correlation of Tetracycline and Sulfonamide antibiotics with corresponding resistance genes and resistant bacteria in a conventional municipal wastewater treatment plant. Sci Total Environ. 2012;421–422:173–83.

9. 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. 2011;45(2):681–93.

10. Zhu Y, Wang Y, Zhou S, Jiang X, Ma X, Liu C. Robust performance of a membrane bioreactor for removing antibiotic resistance genes exposed to antibiotics: Role of membrane foulants. Water Res. 2018;130:139–50.

11. Le TH, Ng C, Tran NH, Chen H, Gin KYH. Removal of antibiotic residues, antibiotic resistant bacteria and antibiotic resistance genes in municipal wastewater by membrane bioreactor systems. Water Res. 2018;145:498–508.

12. Xia S, Jia R, Feng F, Xie K, Li H, Jing D, et al. Effect of solids retention time on antibiotics removal performance and microbial communities in an A/O-MBR process. Bioresour Technol. 2012;106:36–43.

13. Staley C, Kaiser T, Vaughn BP, Graiziger C, Hamilton MJ, Kabage AJ, et al. Durable long-term bacterial engraftment following encapsulated fecal microbiota transplantation to treat Clostridium difficile infection. MBio. 2019;10(4):e01586-19.

14. Zhang M, Chen S, Yu X, Vikesland P, Pruden A. Degradation of extracellular genomic, plasmid DNA and specific antibiotic resistance genes by chlorination. Front Environ Sci Eng. 2019;13(3):38.

15. Liu SS, Qu HM, Yang D, Hu H, Liu WL, Qiu ZG, et al. Chlorine disinfection increases both intracellular and extracellular antibiotic resistance genes in a full-scale wastewater treatment plant. Water Res. 2018;136:131–6.

16. Pang Y, Huang J, Xi J, Hu H, Zhu Y. Effect of ultraviolet irradiation and chlorination on ampicillin-resistant Escherichia coli and its ampicillin resistance gene. Front Environ Sci Eng. 2016;10(3):522–30.

17. Gibson C, Kraemer SA, Klimova N, Vanderweyen L, Klai N, Guo B, et al. Multiplexed amplicon sequencing reveals high allelic diversity of antibiotic resistance genes in Québec sewers. In preparation, 2023.

18. Smith SD, Choi J, Ricker N, Yang F, Hinsa-Leasure S, Soupir ML, et al. Diversity of Antibiotic Resistance Genes and Transfer Elements-Quantitative Monitoring (DARTE-QM): a method for detection of antimicrobial resistance in environmental samples. Commun Biol. 2022;5:216.

19. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021 Jan 18;22(1):178–93.

20. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Reports 2018 81. 2018 Apr 12;8(1):1–11.

21. Jankowski P, Gan J, Le T, McKennitt M, Garcia A, Yanaç K, et al. Metagenomic community composition and resistome analysis in a full-scale cold climate wastewater treatment plant. Environ Microbiomes. 2022 Dec 1;17(1):1–20.

22. Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, Groussin M, et al. An Omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat Commun. 2021;12:4765.

23. Gibson C, Jauffur S, Guo B, Frigon D. Activated sludge microbial community assembly: The role of influent microbial community immigration. bioRxiv. 2023;525574.

24. Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB, et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front Microbiol. 2016;7:173.

25. Che Y, Xia Y, Liu L, Li AD, Yang Y, Zhang T. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome. 2019;7:44.

26. Guo B, Liu C, Gibson C, Frigon D. Wastewater microbial community structure and functional traits change over short timescales. Sci Total Environ. 2019;662:779–85.

27. Sun S, Geng J, Li B, Ma L, Sun X, Meng F, et al. Temporal variations of antibiotic resistance genes in influents and effluents of a WWTP in cold regions. J Clean Prod. 2021;328:129632.

28. Qin K, Wei L, Li J, Lai B, Zhu F, Yu H, et al. A review of ARGs in WWTPs: Sources, stressors and elimination. Chinese Chem Lett. 2020;31(10):2603–13.

29. Boeije G, Corstanje R, Rottiers A, Schowanek D. Adaptation of the CAS Test System and Synthetic Sewage for Biological Nutrient Removal. Chemosphere. 1999;38(4):699–709.

30. Armbruster DA, Pry T. Limit of Blank, Limit of Detection and Limit of Quantitation. Clin Biochem Rev. 2008; 29:S49.

31. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–7.

32. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. Vegan: Community Ecology Package. 2022.

33. NCBI Research Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2013;41:D8–20.

34. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: A resource of complete bacterial plasmids. Nucleic Acids Res. 2019;47:D195-202.

35. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nat Lett. 2014;509:612–6.

36. Wu D, Huang X-H, Sun J-Z, Graham DW, Xie B. Antibiotic resistance genes and associated microbial community conditions in aging landfill systems. Environ Sci Technol. 2017;51(21):12859–67.

37. Zhou Z-C, Zheng J, Wei Y-Y, Chen T, Dahlgren RA, Shang X, et al. Antibiotic resistance genes in an urban river as impacted by bacterial community and physicochemical parameters. Environ Sci Pollut Res. 2017;24:23753–62.

38. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48:D517-525.

39. Sánchez-Osuna M, Cortés P, Llagostera M, Barbé J, Erill I. Exploration into the origins and mobilization of di-hydrofolate reductase genes and the emergence of clinical resistance to Trimethoprim. Microb Genomics. 2020;6(11):mgen000440.

40. Harris K, Parsons TL, Ijaz UZ, Lahti L, Holmes I, Quince C. Linking statistical and ecological theory : Hubbell's unified neutral theory of biodiversity as a hierarchical dirichlet process. Proc IEEE. 2017;105(3):516–29.

41. Dai D, Brown C, Bürgmann H, Larsson DGJ, Nambi I, Zhang T, et al. Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. Microbiome. 2022;10:20.

42. Luo G, Li B, Li L-G, Zhang T, Angelidaki I. Antibiotic resistance genes and correlations with microbial community and metal resistance genes in full-scale biogas reactors as revealed by metagenomic analysis. Environ Sci Technol. 2017;51:4069–80.

43. Munck C, Albertsen M, Telke A, Ellabaan M, Nielsen PH, Sommer MOA. Limited dissemination of the wastewater treatment plant core resistome. Nat Commun. 2015;6:8452.

44. Stokes HW, Gillings MR. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into gram-negative pathogens. FEMS Microbiol Rev. 2011;35:790–819.

45. Kohler P, Tijet N, Kim HC, Johnstone J, Edge T, Patel SN, et al. Dissemination of Verona Integron-encoded Metallo-β-lactamase among clinical and environmental Enterobacteriaceae isolates in Ontario, Canada. Sci Rep. 2020;10:18580.

46. Che Y, Xu X, Yang Y, Břinda K, Hanage W, Yang C, et al. High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of resistance between WWTP- and human/animal-associated bacteria. Microbiome. 2022;10:16.

47. Vogwill T, Maclean RC. The genetic basis of the fitness costs of antimicrobial resistance: A meta-analysis approach. Evol Appl. 2015;8:284–95.

48. Cheng Q, Zheng Z, Ye L, Chen S. Identification of a novel metallo-β-Lactamase, VAM-1, in a foodborne Vibrio alginolyticus isolate from China. Antimicrob Agents Chemother. 2021;65(11):e0112921.

49. Dang B, Zhang H, Li Z, Ma S, Xu Z. Coexistence of the blaNDM-1-carrying plasmid pWLK-NDM and the blaKPC-2-carrying plasmid pWLK-KPC in a Raoultella ornithinolytica isolate. Sci Rep. 2020;10:2360.

CHAPTER 6

Solids Retention Time Impacts Influent Immigration and Antimicrobial Resistance in Wastewater Treatment Plants

Connecting text: Chapter 4 and 5 described the impact of immigration on the microbial community composition and antimicrobial resistance genes (ARGs) in the activated sludge. These experiments were conducted at a short to intermediate solids retention time (SRT) of 5 days. SRT is the main operational parameter used to control the performance and stability of activated sludge reactors. It is also known to modulate the composition of the activated sludge microbial community. Expanding upon the findings of Chapter 4 and 5, in Chapter 6 the impact of SRT on influent immigration and AMR was explored by repeating similar reactor experiments at five different SRTs between 1.8 and 14 days. By including no immigration controls, resident and immigrant populations of the activated sludge could be defined at each SRT. This approach revealed parallels between the ARG abundance and immigrant and resident populations. The results in this chapter also bring to question the accuracy of previous methods used to study immigration which rely simply on comparing shared populations between influent and activated sludge communities.

6.1 Introduction

Wastewater treatment plants (WWTPs) are reservoirs of emerging contaminants such as antibiotic resistance genes (ARGs) (1) and xenobiotics (2). Within these systems, the occurrence of ARGs is thought to be influenced by numerous factors such as the community assembly process (3), residual concentrations of antimicrobials (4), horizontal gene transfer within the treatment community (5) and the influx of genes from upstream locations (6,7). With the large number of factors likely influencing antimicrobial resistance (AMR) in full-scale WWTPs, it is challenging to delineate the relative contribution of each variable. In turn, efforts to reduce AMR in activated sludge WWTPs (AS-WWTPs) using novel technologies often produce conflicting results (8–13). To enable the development of new strategies to minimise the spread of AMR, we first need to understand the factors impacting the persistence of ARGs.

Influent wastewater is known to contain an abundance of antimicrobial resistant bacteria (ARB) and ARGs from municipal, clinical and agricultural settings (6,7). In the past, studies have identified shared ARGs between the influent and activated sludge communities which suggested that these ARGs immigrated into the WWTP. More recently, a study using highly controlled reactors confirmed that influent immigration alone causes a significant increase in the abundance and diversity of ARGs (14). However, high-throughput amplicon sequencing demonstrated that these ARG increases were not only attributed direct ARG immigration, but were associated with changes in the microbial community composition. With immigration occurring continuously in WWTPs, controlling this process could play an important role in reducing the emergence and spread of AMR in the environment.

In AS-WWTPs, both the solid retention time (SRT) and immigration are intricately related and play integral roles in shaping the microbial community composition. Within a WWTP, the resident

populations that are not augmented by immigrant populations must have average growth rates equal to the reciprocal of the solid retention time to be maintained within the system. However, immigration from the influent wastewater can rescue populations with lower growth rates allowing them to be maintained within the system due to constant bioaugmentation (15). In full-scale WWTPs, high rate activated sludge systems (with a low SRT) have been shown to be more susceptible to sewer immigration, with up to 62 % of OTUs shared between the influent and mixed liquor (16). This could be due to mass immigration effects or the high disturbance caused by low solid retention times (SRT). Similarly, Vuono et al. 2016 found that the disturbance caused at lower SRTs resulted in increased colonisation success in the activated sludge (17). However, these studies are often based upon the assumption that shared OTUs between the influent and activated sludge can be used as an indicator of immigration, which may not be entirely accurate. Considering the complex dynamics of ARGs previously observed between the influent and activated sludge (14), it remains to be seen whether the changes in the microbial community with SRT also impact the persistence of AMR in the activated sludge.

The current study aimed to investigate the relationship between immigration, SRT and AMR using a series of highly controlled reactors. Based upon the relationship between immigration and ARG abundance (14), it was hypothesised that AMR in the activated sludge would increase with the impact of immigration. Given the multitude of variables between full-scale AS-WWTPs, and the complexity of ARG dynamics between the influent and activated sludge, highly controlled lab-scale reactors were used in this study to ensure variables were controlled throughout and the established populations could be accurately defined.

6.2 Materials and Methods

6.2.1 Reactor Experimental Design

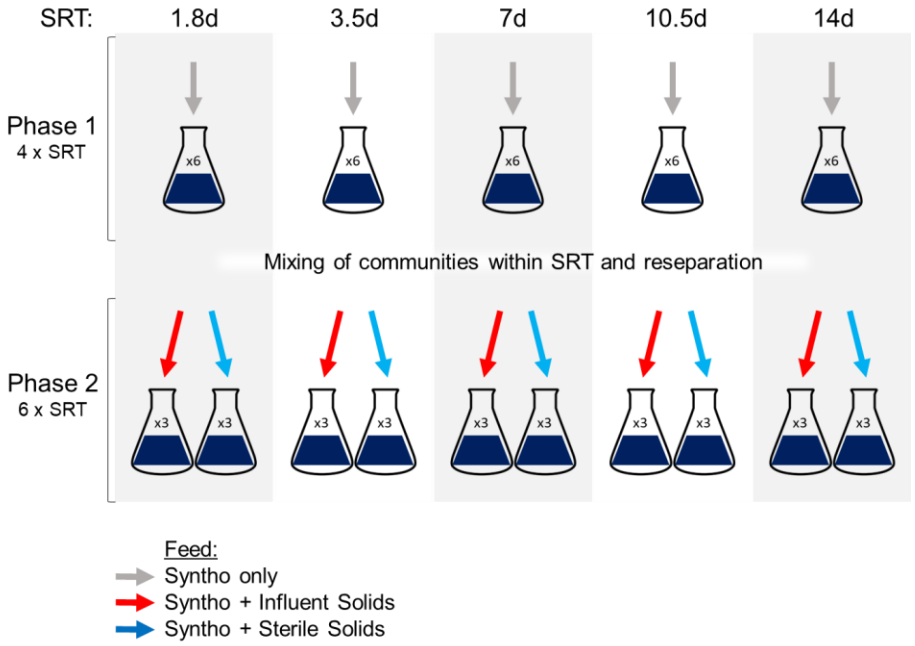


Figure 6.1- Schematic of Reactor Operation. Five sets of reactors were operated with different solid retention times (SRTs), ranging from 1.8 - 14 days. During Phase 1, all reactors were fed with synthetic wastewater only for 4 SRTs. During Phase 2 half of the reactors received synthetic wastewater supplemented with influent solids, and the remaining half received synthetic wastewater supplemented with sterile influent solids for 5 SRTs.

In total, 30 reactors were operated which comprised of 5 sets of reactors which differed based upon the SRT of operation as shown in Figure 6.1 (SRTs; 1.8, 3.5, 7, 10.5 and 14 days). Reactors were operated between January and June 2020. The beginning of operation was staggered to enable all reactors to be operated for the same number of SRTs, while finishing operation on the same date to ensure all reactors received the same sample of influent solids for the final SRT. As shown in Figure 6.1, the reactors were operated in two phases with the first aiming to establish a stable

reactor microbial community and the second phase to test the impact of immigration. During Phase 1, all reactors were fed with a synthetic wastewater (Syntho; as defined in Section 6.2.5) only for four SRTs to ensure a steady-state core microbial community was established which was adapted to the specific substrate composition and any existing immigration dependent taxa were removed. Measurements of TSS, VSS and soluble COD over time were used to ensure reactors were operating at steady-state. At the end of Phase 1, the reactors were mixed and split again to ensure a homogeneous stable community was present in each reactor as described by Kaewpipat and Grady, 2002 (18). During Phase 2 of reactor operation, half of the reactors from each set (3 reactors per SRT) received Syntho supplemented with influent solids harvested from a full-scale WWTP (Cowansville, Quebec, Canada) to simulate immigration. The remaining 3 reactors per SRT were fed with fed with Syntho supplemented with influent solids which had been sterilised by gamma irradiation as described in Section 6.2.4, and acted as a substrate control to distinguish between active immigration and the impact of the influent solids as an additional substrate. All reactors were operated for five SRTs during Phase 2.

6.2.2 Reactor Operation

Reactors were inoculated with activated sludge grab samples collected from La Prairie wastewater treatment plant located in Quebec, Canada one day before operation commenced. All inoculum grab samples were collected from the same region of the aeration tank, and transported to McGill University on ice, stored at 4 ℃ and processed within 24 hours of collection. Reactors were inoculated with 160-mL of 2700 mg-VSS/L sludge and incubated at 21 °C with shaking at 180 rpm to allow aeration and gentle mixing. The hydraulic retention time (HRT) and SRT were controlled through feeding and wasting six times per week. To mimic the clarifier of an activated sludge wastewater treatment process, three times per week the reactors were settled in 100-mL

graduated cylinders for 45 minutes to allow solid-liquid separation to occur, and the required volume of supernatant was removed. The average HRT was 1.8 days, and the solid retention time for each reactor set is as noted in Figure 6.1 ranging from 1.8 to 14 days.

6.2.3 Influent Solids Collection

Influent wastewater solids for use during Phase 2 were harvested from the influent stream at Cowansville wastewater treatment plant located in Quebec, Canada. On site, influent wastewater was concentrated by gravity settling using 20-L buckets. The concentrated influent solids were transported to McGill University on ice and processed immediately. The concentrated wastewater was centrifuged in 50-mL aliquots at 21,100 x g for 10 minutes (Thermofisher Scientific model ST16R Centrifuge) to separate the influent wastewater and solid fractions. The supernatant was removed and the influent solids were resuspended in Syntho (synthetic wastewater; as described in section 6.2.5). This step was repeated three times to wash the influent solids. After washing, the concentrated influent solids were resuspended in a final volume of ~ 600-mL of Syntho. The volatile suspended solids (VSS) concentration of the stock solution was measured using the Standard Method 2540E (19). The concentrated influent solids were stored at 4 ℃ for no more than two weeks, and diluted to a working concentration of 120 mg/L before use.

6.2.4 Preparation of Influent Solids for Substrate Control

To determine whether immigrating populations were actively growing in the activated sludge and distinguish between the impact of substrate addition, sterilised influent solids were used as a negative control. Influent solids were collected as described in Section 6.2.3, and gamma irradiation with Cobalt-60 was performed at Nordion™, Laval, Canada on a concentrated stock solution. This method inactivated microorganisms whilst minimising any change to the structure

of the influent solids. A final dose of 18.0 kGy was delivered to a 1.8-L volume of influent solid suspension held in a 2-L polyethylene bottle. Upon return to the lab, 50-mL aliquots of irradiated influent solids were stored to prevent contamination during use and diluted as required to a concentration of 120 mg/L. To assess the performance of the irradiation process, measurements were taken of faecal coliforms using Standard Method 9222D (19) with membrane fecal coliform medium and incubating at 44 °C for 24 hours. Total coliforms were quantified using Standard Method 9222B (19) using LES Endo Agar and incubating at 35 °C for 24 hours. The total and volatile suspended solids were measured using Standard Methods 2540D and 2540E, respectively (19). No significant change in the suspended solids concentration or chemical oxygen demand were observed after irradiation. In addition, no growth of microorganisms or spores was achieved within a 48-hour incubation period.

6.2.5 Reactor Feed Preparation

Throughout operation, all reactors were fed with a synthetic wastewater (Syntho) to ensure the wastewater composition remained stable over time. Syntho was produced as described by Gibson et al., 2023 (15) and autoclaved in 500-mL bottles at 121 ℃ and 15 psi for 30 minutes before use. During Phase 2, reactors received feeds supplemented with either active or gamma irradiated influent solids, as described in Section 6.2.3 and 6.2.4. To prepare these feeds, a stock solution of the influent solids was diluted in Syntho to a concentration of 120 mg-VSS/L, which represented the average VSS concentration of influent wastewater from three WWTPs located in Quebec, and to allow comparison with previous studies (20).

6.2.6 Sample Analysis

Effluent wastewater samples were analysed for chemical oxygen demand (COD) using Standard Methods 5220D (19). Mixed liquor samples were collected once per SRT for the analysis of total suspended solids (TSS) and volatile suspended solids (VSS) using Standard Methods 2540B and 2540E respectively (19). Biomass samples for DNA extraction and downstream sequencing were centrifuged at 16,000 x g for 5 minutes, the supernatant was removed and samples were stored at −80 ℃ for future use.

6.2.7 Microbial Community Analysis using 16S rRNA Gene Sequencing

DNA was extracted from biomass samples using DNeasy PowerSoil Pro Kit (Qiagen, Germantown, MD, USA), as per manufacturers instructions using a centrifugation speed of 10,000 x g. PCR of the 16S rRNA gene V4 region was conducted using the modified Caporaso primers; 515F and 806R primers (21,22). PCR cycling conditions were as follows: 94 ℃ for 3 mins followed by 35 cycles of 94 ℃ for 45 secs, 50 ℃ for 60 secs, 72 ℃ for 90 secs, 72 ℃ for 10 mins and a final hold at 4 ℃. Amplicons were sequenced on the Illumina MiSeq PE250 platform at McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada).

Sequencing data was analysed using QIIME 2 (23) and R Software (packages “vegan” and “ape”) (24). Specifically, the raw sequences were quality-filtered using DADA2 (25) in Qiime2 pipelines. Sequences were rarefied to 70,000 reads in R. ASV tables were exported directly from Qiime2 after quality filtering. Taxonomy was assigned using MiDAS 2.0 reference database (26) and tabulated. Microbial community diversity was analysed using R “vegan” package (24) and Jaccard dissimilarity.

6.2.8 Population Definitions

To enable the immigration dynamics to be studied in greater detail, the microbial community was divided into various populations. The overall core resident genera were defined based upon their presence within at least 80 % of reactors within a set (SRT) with an average relative abundance of at least 0.1 %. Genera not meeting these criteria, but present at the end of Phase 1 were considered to be *non-core residents* of the community. Within the overall core residents, the population was divided into two groups, firstly the *Universal Core Residents* were defined as those which were members of the core resident community in all reactors, regardless of the SRT of operation (as shown in Figure 6.2). Whilst the *SRT core residents* were genera which appeared to be members of the core resident community only at selected SRTs.

Genera belonging to the immigrant populations were defined as those which were not present at the beginning of Phase 2 (i.e. without immigration), but appeared at the end of Phase 2 upon addition of influent solids to the feed. To be considered to be a growing immigrants the average relative abundance of the genera had to be greater than reported in the sterilised substrate control group demonstrating growth. Immigrants occurring in equal or lower abundance in the group with immigration and sterile control were considered to be *residual immigrants.* As with the core resident population, the growing immigrant population was divided into sub-groups based upon their occurrence within each reactor set. The *universal immigrants* were those identified as growing immigrants at all SRTs. Whilst the *SRT immigrants* were detected only at selected SRTs. Finally, *reintroduced immigrants* were those which were members of the core resident population at some SRTs, but introduced as immigrants in others.

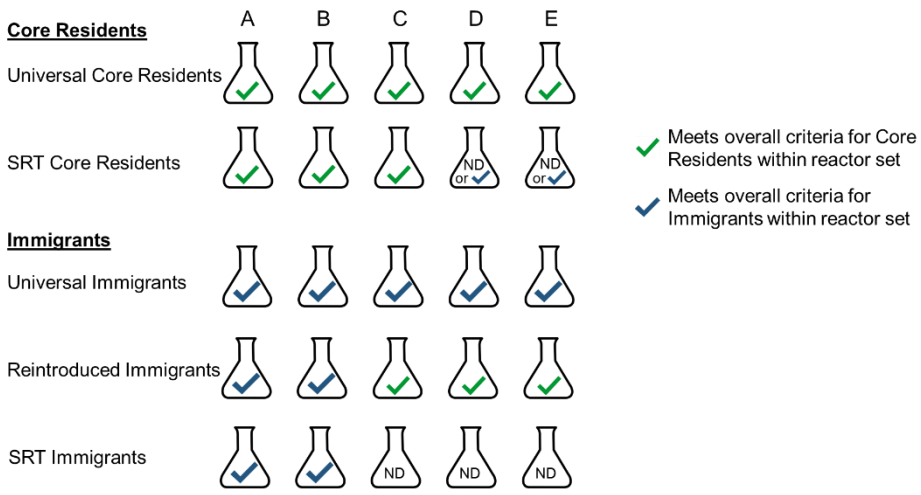


Figure 6.2- Reactor population definitions. Core residents were genera identified to in at least 80 % of reactors with an average relative abundance of at least 0.1 %. Immigrant populations were identified as genera not present at the beginning of Phase 2, but introduced with the addition of influent solids during Phase 2. 'ND' represents reactor sets in which a given genera did not meet the criterion for a core residents genus or immigrant.

6.2.9 Quantifying the Net Growth Rate with Steady-State Modelling

To visualise the impact of immigration on the average net growth rate within the activated sludge a steady state mass balance approach was used as previously described by Guo et al 2022 (7) and in Chapter 4. The growth rate of the i-th taxon can be defined as SRT^{-1} plus the decay rate (b_i) multiplied by 1 minus the immigration rate (m_i) as shown in equation 1. The net growth rate of the i-th taxon can be summarised as the growth rate (μ_i) minus the decay rate b_i (eq. 2). Considering a net growth rate of 0 d^{-1}, equation 3 could be used to determine the fraction of a given genera expected within the ML based upon its abundance in the influent, where θ_x represents the SRT, θ represents the HRT, $f_{OHO,Capt}$ the fraction of the influent biomass of the i-th taxon captured by the mixed liquor solids fraction, $X_{Tot,Inf,}$ and $X_{Tot,ML}$ represent the total solids of the influent and mixed liquor respectively, $\gamma_{DNA,Inf}$ and $\gamma_{DNA,ML}$ repesent the DNA extraction yields for the

influent and mixed liquor and $f_{16S,Inf,i}, f_{16S,ML,i}$ represent the relative abundance of the i-th taxon in the influent and mixed liquor as determined by 16S rRNA gene sequencing.

$$\mu_i = \left(\frac{1}{SRT} + b_i\right)(1 - m_i) \qquad \text{(eq. 1)}$$

$$\mu_{net} = (\mu_i - b_i) \qquad \text{(eq. 2)}$$

$$f_{16S,i,ML} = \frac{\theta_x}{(1-\mu_{net,i}\theta_x)} \cdot \frac{f_{OHO,Capt}}{\theta} \cdot \frac{X_{Tot,Inf}Y_{DNA,Inf}}{X_{Tot,ML}Y_{DNA,ML}} \cdot f_{16S,i,Inf} \qquad \text{(eq. 3)}$$

6.2.10 Quantification of ARGs by Quantitative PCR Array

Antimicrobial resistance genes in the reactor and influent samples were analysed by Resistomap, Finland using SmartChip qPCR. For this analysis, DNA was extracted from each reactor and triplicates were pooled. In total, 5 samples were analysed from the beginning of Phase 2 (1 per SRT), 10 from the end of Phase 2 (1 with immigration and 1 sterile control per SRT) and three influent solid samples from different time points of reactor operation. All pooled samples were analysed in triplicate by SmartChip qPCR.

In total, 96 targets were analysed as listed in supplementary Table S6.1, which comprised of 67 antimicrobial resistance genes, 15 mobile genetic elements and integrons, 6 taxonomic markers and 8 metal resistance genes and other resistance determinants. ARG abundance was normalised to copies per 16S rRNA gene.

6.2.11 Statistical Analysis of AMR Data

A test of multiple regression was carried out using the Tidyverse package (27) in R to determine the influence of SRT and reactor conditions on ARG abundance. Briefly, the reactor variables were converted into two binary dummy variables, with the first coded to distinguish against reactors

with solids added (both live and gamma irradiated) and the start of Phase 2 (before immigration was introduced). The second dummy variable was used to distinguish between the reactors receiving active influent solids and those receiving sterilised gamma-irradiated influent solids. Results were filtered based upon the F-test and only genes with a p-value < 0.05 were included in further analysis. The linear relationship between the variables was considered to be significant if the ρ-value was less that 0.05 and the null hypothesis rejected.

6.3 Results

6.3.1 SRT Selects for Unique Reactor Communities

During Phase 1 of reactor operation, all reactors received synthetic wastewater only (Syntho) to allow a steady-state community to be established. This phase of reactor operation also served to remove any existing immigrants from the system before the test phase with immigration began. After 4 SRTs, based upon reactor theory it was assumed that a steady-state microbial community had formed, however this was not confirmed experimentally. Analysis of the microbial community composition at the end of Phase 1 revealed that distinct microbial communities were assembled based upon the SRT of operation (Figure 6.3a). Principle coordinate analysis using Bray-Curtis dissimilarity showed clustering of the reactor microbial communities into high (7 days and above) and low (3.5 days and below) SRTs along the PCoA2 axis. When visualised with the Jaccard dissimilarity, the microbial communities further separated suggesting that each SRT has unique genera occurring at lower abundances, which have higher weights in the Jaccard dissimilarity.

Based upon the classifications previously established by Wu et al., 2019 (28), a core resident population was defined based upon the occurrence of a given genus at an average relative abundance of $\geq$ 0.1 % and detection in $\geq$ 80 % (5/6) of reactors within each set. Genera identified to belong to the core resident population at all SRTs were further classified as *Universal core*

residents based on their ability to persist under all reactor conditions. Genera which formed part of the core resident population at only selected SRTs were further classified as *SRT core residents* (section 6.2.8, above).

At each SRT, the reactors had unique universal and SRT core resident populations (Figure 6.3b and 6.3d). The abundance of universal core resident genera varied with SRT, for example, the genus *Acinetobacter* (family *Moraxelleaceae*) was more abundant at higher SRTs, whilst the genus *Zoogloea* (family *Rhodocyclaceae*) decreased in abundance with increasing SRTs. These different community structures likely have implications on niche availability and how the wastewater treatment process can be optimised.

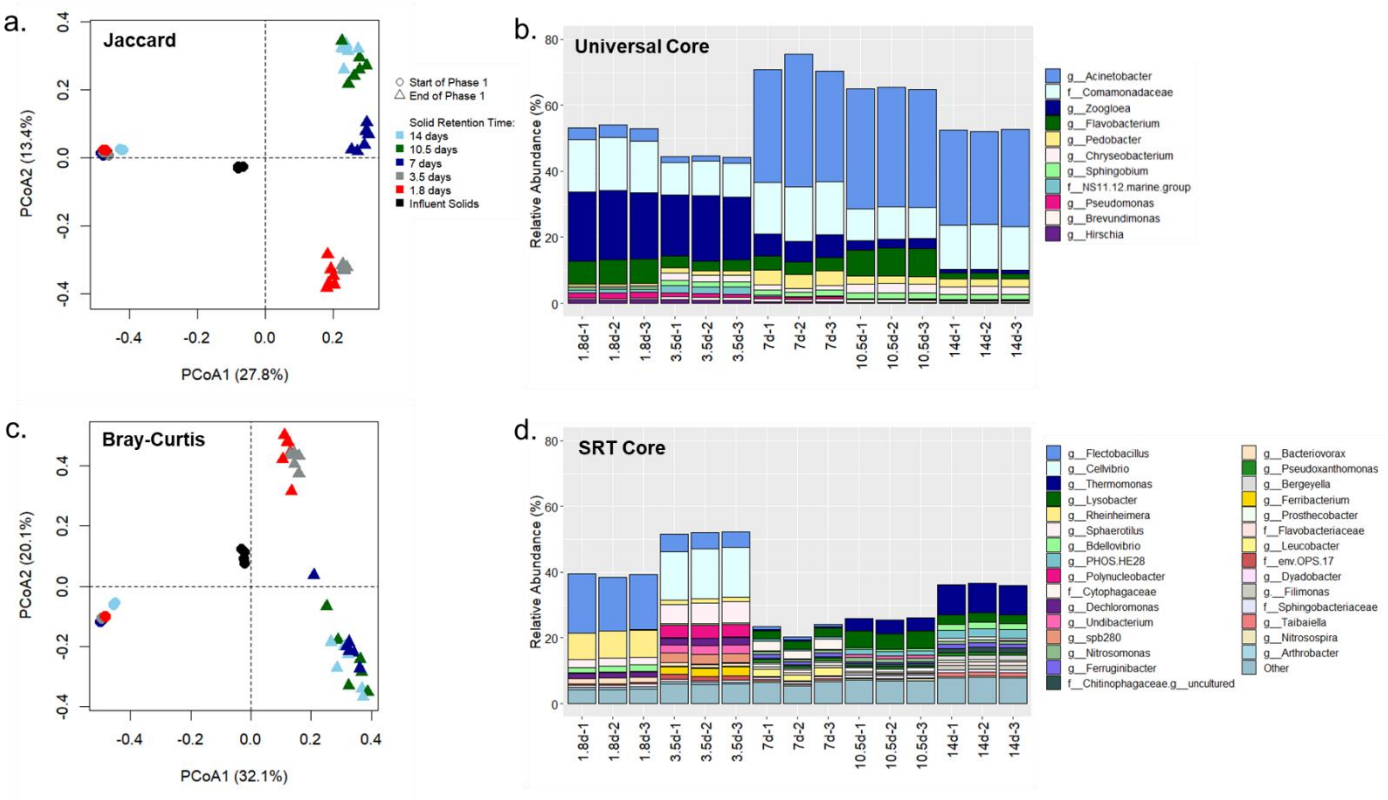


Figure 6.3 - Microbial community analysis at the end of Phase 1, after all reactors were fed synthetic wastewater only for four SRTs. a) Principle coordinate analysis using Jaccard dissimilarity showing unique clustering of communities based on SRT of operation b) universal core resident population including genera present at all SRTs c) Principle coordinate analysis using Bray-Curtis dissimilarity showing clustering of reactor communities into high ($\geq$ 7d) and low ($\leq$ 3.5d) SRTs. d) Top 30 SRT reactor core residents. Relative abundance given as a percentage of the total reads.

The SRT core resident populations included genera which were present only at specific SRTs. At higher SRT, the genera *Thermomonas* and *Lysobacter* (both in family *Xanthomonadaceae*) were among the dominant populations, whilst at lower SRT the genera *Flectobacillus* (family *Flexibacteraceae*), *Cellvibrio* (family *Pseudomonadaceae)* and *Rheiheimera* (family *Chromatiaceae)* were of the most abundant ones. Given the unique microbial communities formed at the end of Phase 1, it is likely that niche availability in the reactors differed which could consequently impact the immigration outcomes during Phase 2.

6.3.2 Impact of Immigration Increases with SRT

After establishing a stable microbial community during Phase 1, immigration was introduced into the reactors in Phase 2. During this time, in addition of Syntho, half of the reactors (3 per SRT; 15 reactors total) received influent solids collected from a full-scale wastewater treatment plant, while the remaining 15 reactors received gamma-irradiated influent solids which acted as a substrate control. The substrate control was included to allow the impact of active immigration to be distinguished from the impact of the influent solids as an additional source of carbon and energy.

Principle coordinate analysis was used to visualise the overall community dynamics of the reactors at the end of Phase 2 (Figure 6.4). When visualised using Jaccard dissimilarity, the impact of both SRT and immigration on the microbial community could be clearly identified (Figure 6.4a). Reactors separated along the PCoA1 axis based upon the SRT of operation, and with immigration were found to move along the PCoA2 axis becoming more similar to that of the influent solids. Immigration impacted the microbial community under all reactor conditions, however at higher SRTs, the sterile control also displayed changes in the microbial community likely due to the introduction of residual DNA remaining after sterilisation. At lower SRTs, only slight drift was observed in the reactors receiving sterilised influent solids when compared to the starting position

(start Phase 2) suggesting residual DNA accumulation to have less of an effect at lower SRTs. Using Bray-Curtis dissimilarity, similar trends were observed (Figure 6.4b). At higher SRTs, particularly in the 14d-SRT reactor set, the sterile control drifted from the starting position becoming more similar to the influent solids (along the PCoA2, Figure 6.4b), suggesting a greater impact of immigration on the community assembly at high SRTs.

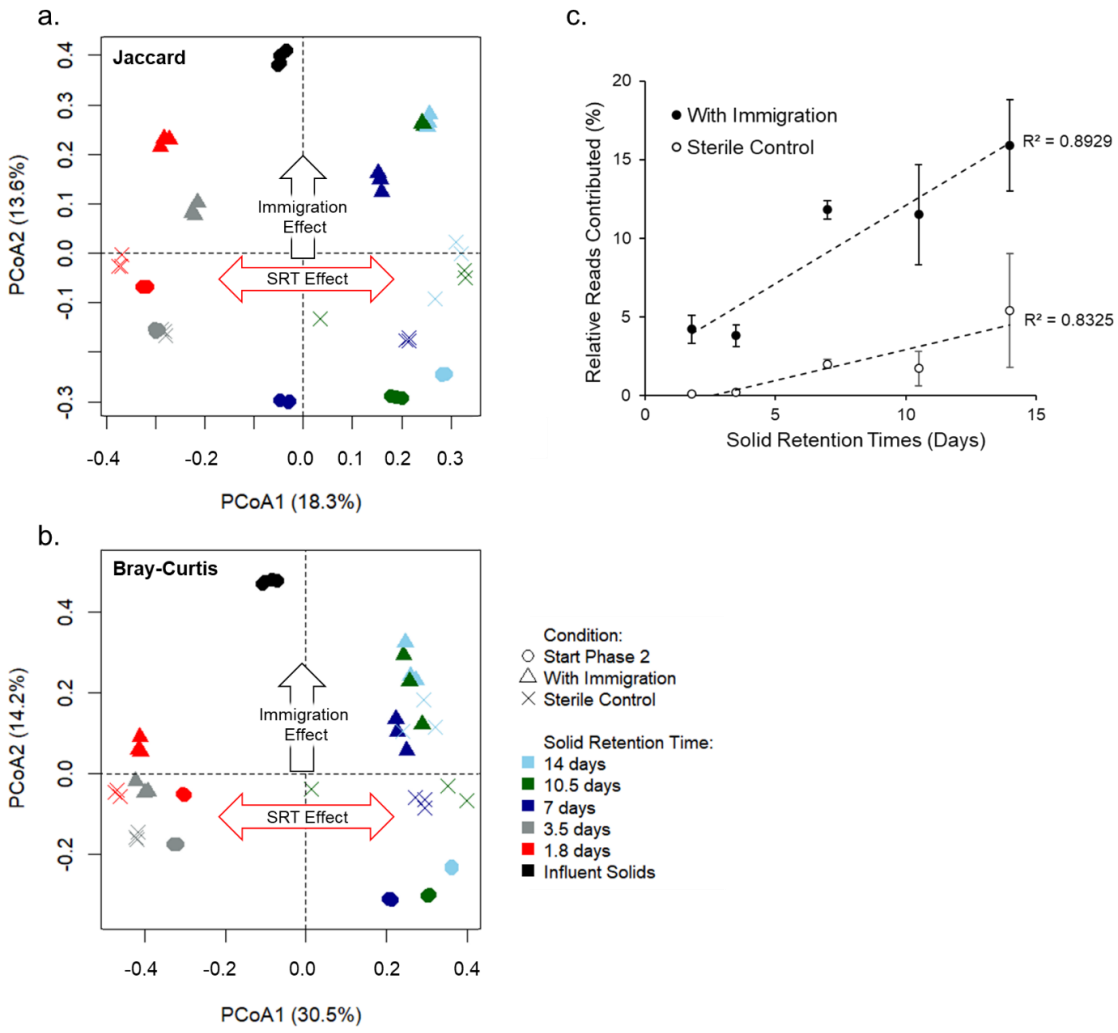


Figure 6.4- Analysis of the microbial community composition at the end of Phase 2. Principal coordinate analysis of reactor community composition visualised at ASV level using a) Jaccard dissimilarity, b) Bray-Curtis dissimilarity and c) Reads contributed through immigration and accumulation in sterile control.

6.3.3 Quantifying the Impact of Immigration at Different SRTs

The relative impact of immigration at different SRTs was further investigated by classifying the *growing immigrant* genera into different categories, and disregarding the *non-growing immigrant* genera. To be considered a growing immigrant, a genus had to occur only in reactors receiving influent solids (at the end on Phase 2), and at higher abundance than in reactors receiving irradiated influent solids. In this way, genera with abundances only driven by direct immigration or by growth on new substrates were excluded. The growing immigrants were further classified as: *Universal Immigrants* if they immigrated at all SRTs, *SRT immigrants* if they were only introduced at specific SRTs, and *Reintroduced immigrants* if genera were previously classified as SRT core residents at specific SRTs and were growing immigrants at others at the end of Phase 2. The population definitions established could not be achieved without the highly controlled reactor set-up and control groups included.

The proportion of 16S rRNA gene amplicon sequencing reads contributed by growing immigrant genera increased with SRT (Table 6.1) making the high-SRT activated sludge community structures drift closer to the influent community structures (Figure 6.4a and 6.4b). Intriguingly, the sterile influent control also accumulated a higher proportion of reads from genera classified as growing immigrants with increasing SRT (Figure 6.4c). The contribution of each growing immigrant sub-group was quantified in terms of relative reads of the overall community (Table 6.1). The proportions of the amplicon reads contributed by the universal immigrant genera appeared to be similar at all SRTs, whilst the relative abundance of reintroduced immigrant population followed a unimodal distribution centred at the 7-day SRT and < 2 % at other SRTs. Finally, the increased proportion of growing immigrants with SRT (Figure 6.4b) was due to the SRT immigrant genera which were typically the most abundant group of immigrants (Table 6.1).

Table 6.1: Reads contributed (%) by reactor populations

SRT	Core Residents		Non-Core Residents	Immigrant Population			Residual DNA
	Universal	SRT		Universal	Reintroduced	SRT	
1.8-day	21.1 ± 5.0	69.5 ± 6.8	3.2 ± 0.3	1.9 ± 0.1	0.2 ± 0.1	2.2 ± 0.1	2.0 ± 0.8
3.5-day	23.9 ± 4.9	58.0 ± 5.6	11.7 ± 1.0	1.5 ± 0.3	0.3 ± 0.2	1.9 ± 0.1	2.6 ± 0.6
7-day	45.7 ± 1.5	33.2 ± 0.6	8.9 ± 1.5	2.3 ± 0.3	5.1 ± 1.2	4.4 ± 0.6	0.4 ± 0.1
10.5-day	32.8 ± 13.4	28.4 ± 1.8	27.0 ± 15.2	2.3 ± 0.4	1.3 ± 0.2	7.9 ± 2.7	0.2 ± 0.1
14-day	15.7 ± 3.1	62.3 ± 6.6	5.7 ± 0.8	1.5 ± 0.2	1.5 ± 0.3	13.0 ± 2.5	0.3 ± 0.1

Where ± represents the standard deviation in the percentage reads detected between the three reactors

In previous studies, the impact of influent immigration on the assembly of the activated sludge microbial community was quantified by proportion of shared OTUs (ASVs or other taxonomic levels) between the influent and activated sludge mixed liquor (17,29,30). The premise of this analysis was tested here by comparing the proportion of shared ASVs between the influent and activated sludge before and after the introduction of immigration (i.e. start and end of Phase 2, respectively). This analysis revealed that, before immigration, a significant number of the ASVs resident in activated sludge communities were already shared with the influent communities despite immigration having not occurred (Table 6.2). Linear regression found ASVs and the reads shared between the influent and mixed liquor to be significantly negatively correlated with SRT (ASVs; $R^2 = 0.88$ and $\rho < 0.001$, Reads; $R^2 = 0.97$ and $\rho < 0.001$; Table S6.2). Without appropriate no-immigration controls, this result could be incorrectly assumed to represent increased immigration at lower SRTs. However, when immigration was actually introduced during Phase 2, the reads contributed by ASVs which were newly introduced significantly increased with SRT ($R^2 = 0.68$, $\rho < 0.001$; Table S6.2) consistent with the results from Table 6.1. If the total shared ASVs and associated reads were assumed to be immigrants, the fraction shared between the influent and

activated sludge would be similar at all SRTs (Table 6.2), despite the level of immigration actually having been shown to increase at higher SRTs (Table 6.1). After immigration, when compared to those shared at the beginning of Phase 2, the newly introduced ASVs displayed higher diversity but typically contributed fewer reads. Taken together, this suggests that newly introduced ASVs contribute more towards diversity in low abundance genera, which supports the patterns observed using principle coordinate analysis in Figure 6.4.

Table 6.2: The proportion of mixed liquor ASVs and reads shared with influent

SRT (Days)	Before Immigration[a]		After Immigration				Total Immigrants	
			Shared with P2S[b]		Newly introduced			
	ASVs (%)	Reads (%)	ASVs (%)	Reads (%)	ASV (%)	Reads (%)	ASVs (%)	Reads (%)
1.8	28.3 + 0.9	56.4 + 0.5	21.9 + 2.0	53.4 + 3.7	34.8 + 1.8	6.4 + 0.9	56.7 + 0.4	59.8 + 3.2
3.5	28.4 + 0.7	53.4 + 0.4	12.5 + 0.9	42.4 + 5.9	29.0 + 0.0	5.2 + 1.1	41.5 + 1.0	47.6 + 4.9
7	18.9 + 1.4	45.0 + 2.4	12.1 + 0.4	19.0 + 3.6	41.1 + 1.1	18.4 + 0.7	53.2 + 1.3	37.4 + 3.2
10.5	16.7 + 0.4	41.1 + 0.2	10.6 + 0.7	30.7 + 10.7	43.5 + 2.1	16.8 + 7.8	54.0 + 1.5	47.5 + 3.4
14	14.5 + 0.0	32.7 + 0.2	7.5 + 0.6	21.0 + 7.2	45.2 + 1.9	9.0 + 1.2	52.7 + 2.5	30.0 + 6.7

[a]Before Immigration indicates samples taken at the start of Phase 2 (after mixing)

[b]ASV and reads present at the end of Phase 2 (after immigration), which were also present at the beginning of Phase 2 (before immigration). Thus, they are considered resident (core or non-core) ASV.

6.3.4 Immigrating Populations vary With SRT

Analysis of the genera in the immigrant sub-groups revealed unique immigrant populations at each SRT. Among the SRT immigrants, at low SRT the genus *Paludibacter* (family *Paludibacteraceae*) was dominant, whilst at higher SRTs the genus *Comamonas* (family *Comamonadaceae*) and the family *Pseudomonadaceae* were in higher abundance (Figure 6.5). At 10.5-day and 14-day SRT the genus *Dechloromonas* (family *Rhodocyclaceae)* was reintroduced into the reactors whilst at lower SRTs, it was a member of the SRT core residents. Among the universal immigrants, species

from the genus *Macellibacteroides (*family *Porphyromonadaceae)*, *Propionivibrio* (family *Rhodocyclaceae*) and *Tolumonas* (family *Aeromonadaceae*) were dominant immigrants at all SRTs. These results demonstrate that SRT selects for unique core resident and immigrant populations. Efforts to optimise and control the microbial community in WWTPs in the future should carefully consider SRT as an important operational parameter.

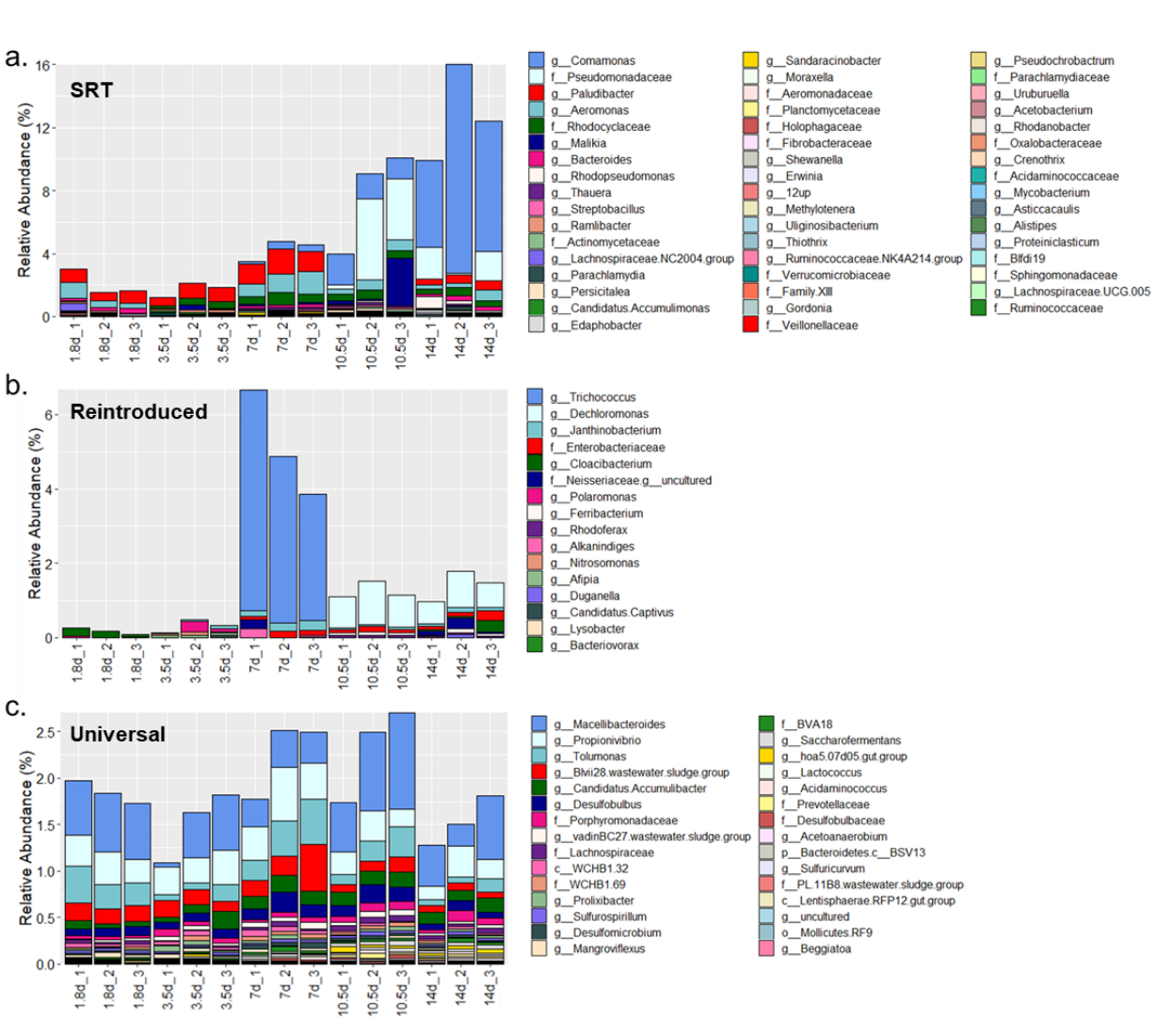


Figure 6.5- Reactor immigrant populations. Immigrants were defined as genera not observed at the beginning of Phase 2, which were introduced into reactors with the addition of influent solids. a) SRT immigrant population which consists of immigrants which appeared only at selected SRTs, b) Reintroduced immigrant population which includes genera which were classified as core residents at some SRTs, but introduced as immigrants in others c) Universal immigrant population which were found at all SRTs.

6.3.5 Growth Rates of Core Resident and Immigrating Populations

Using a mass balance approach (Section 6.2.9), a log-log scatter plot of the relative abundance of genera in the mixed liquor and influent was used to visualize the net growth rates of the different reactor populations (Figure 6.6). In this log-log scatter plot, genera with different relative abundances which are plotted on the same 45°-line have the same net growth rate (Eq. 3). On Figure 6.6, a line was drawn representing the expected relative abundance considering a net growth rate of 0 d^{-1} (i.e. the growth rate necessary to maintain the population without immigration, Eq. 2). Any populations plotted above the $\mu_{net} = 0$ line had positive net growth rates, whilst those below had negative net growth rates and were presumably maintained in the activated sludge by continuous immigration.

Core resident genera (universal and SRT) typically exhibited positive net growth rates at all SRTs, whilst the net growth rate of the immigrant genera was lower and often negative (Figure 6.6), both results were as expected based on the classifications. Counter selection of immigrant populations was clearly observed, with genera occurring in high abundance (> 0.1 %) within the influent solids remaining undetected within the mixed liquor after immigration (ND in Figure 6.6). Conversely, some genera classified as immigrants in the mixed liquor were not detected in the influent, suggesting that they occurred below the detection limit in either the influent or the mixed liquor during Phase 1, and increased in abundance with the addition of influent solids.

Various trends in the growth rate of immigrant populations were observed, which indicates the existence of underlying relationships between the reactors SRT, the absolute immigration rates, and niche occupied by these population. Universal immigrants typically appeared to be present in higher abundance in the influent than the SRT immigrant population (Figure 6.6). This suggests that the influent abundance somewhat influences the fate of immigrating genera under different

reactor conditions. At lower SRTs, the SRT immigrant populations often exhibited positive net growth rates (Figure 6.6). Although SRT immigrants contributed fewer reads to the overall population at lower SRTs, it was observed that in the 1.8-day and 3.5-day SRT reactors, the majority exhibited positive net growth rates (Figure S6.1). With increasing SRT, the proportion of SRT immigrants with a positive net growth rate reduced and the majority displayed overall negative net growth rates, suggesting that immigrants are better established within the reactor microbial communities at lower SRTs.

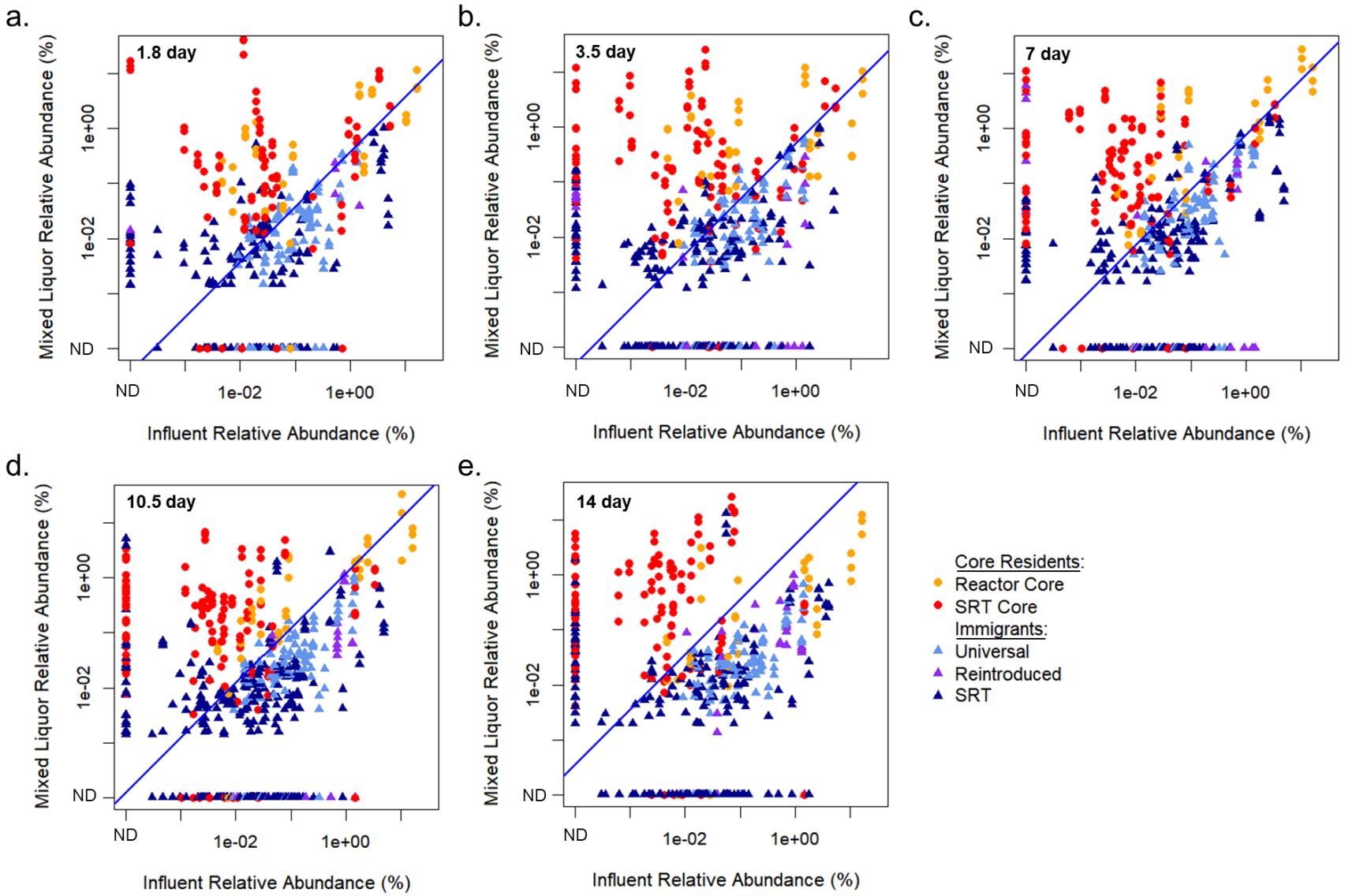


Figure 6.6- Scatter plot comparing the relative abundance of core resident and immigrant genera in the influent and mixed liquor (ML). ND: Not detected, which indicates genera that were below the detection limit in either the influent or ML. Reactors were operated at different solid retention times: a) 1.8 day, b) 3.5 days, c) 7 days, d) 10.5 days, e) 14 days. The μnet = 0 line represents the expected relative abundance of genera considering a net growth rate of 0d-1. Genera with different relative abundances but plotted on the same 45° line have the same net growth rate.

6.3.6 Impact of Immigration and SRT on Reactor AMR

Quantitative PCR array was used to assess the impact of immigration on the diversity and abundance of ARGs in the reactors. Among the samples analysed, the influent solids had the highest richness of ARGs (71.0 $\pm$ 1.4 genes). Before immigration was introduced, reactors operated at lower SRTs (1.8 and 3.5-days) typically had a higher number of ARGs detected (1.8-3.5-days; 65.5 $\pm$ 0.5 genes, 7-14 days; 59.3 $\pm$ 0.9 genes). Given the reported greater similarity between the influent solids and reactors at lower SRTs (Table 6.2), this could be related to the microbial community composition.

With the introduction of immigration during Phase 2, a significant increase in ARG richness in the activated sludge mixed liquor was observed (from 61.6 $\pm$ 3.7 to 67.4 $\pm$ 1.3 genes; $\rho = 0.01$) (Table S6.3). The richness of mobile genetic elements was not significantly impacted by immigration. Genes introduced into the reactors with immigration included *aph*A1/7 (aminoglycoside resistance), InuC (MLSB resistance) and *bla*GES (beta-lactam resistance) (Table S6.4). In other cases, counter selection was observed between the influent and mixed liquor. For example, the *aac*A/*aph*D (aminoglycoside resistance) and *bla*CTX-M (beta-lactam resistance) genes remained undetected in the activated sludge reactors with immigration despite being present within the influent solids (Table S6.4).

Among the genes detected many displayed an increase in abundance during Phase 2 of reactor operation. Overall, of the 89 genes detected using quantitative PCR, 44 $\pm$ 6.1 ARGs and 8.4 $\pm$ 3.0 MGEs displayed an increase in abundance with immigration at each reactor SRT (Figure 6.7). The MLSB ARGs were most often increased in abundance within the reactors receiving influent solids, whilst the impact of immigration appeared to vary with SRT for the ARGs in other classes. Remarkably, the abundance of MGEs was also impacted by immigration during Phase 2. Genes

such as TnpA_7 increased in abundance in all reactors and at all SRTs, whilst others such as ISAba3 increased only in the reactors receiving live influent solids and at some SRTs. These results suggest that ARG mobility is likely influenced by the dynamics at the immigration interface.

The impacts of immigration and SRT on the abundance of ARGs were statistically assessed using a multiple regression approach, and grouped into 7 overall trends (Figure 6.8; remaining genes displayed in Figure S6.2). Of the 89 ARGs and MGEs detected, 27 were found to be significantly correlated with SRT or immigration (raw data available in supplementary file 1). Group 1 and 2 genes exhibited relative abundances to be, respectively, positively and negatively correlated with the SRT, but not significantly affected by the addition of influent solids (Figure 6.8a and 6.8b). While 7 out of the 27 significantly changing ARGs and MGEs were classified in Group 1, Group 2 comprised of a single gene (*int*1_2). This demonstrates that although the gene richness may be higher at lower SRTs, the relative abundance is more frequently lower.

The most frequently observed trend among the analysed genes (8 out of 27) is reported in Group 3, in which the gene abundance increased with the addition of influent solids but the slope of the abundances with SRT did not significantly change (Figure 6.8c). Group 4 genes (only two instances) exhibited the reverse trends in that their abundance decreased with the addition of influent solids, but the slope of abundances with SRT remained statistically unchanged (Figure 6.8d). This trend was observed in the mercury resistance gene (*mer*A) and the IS21-ISAs29 insertion sequence. Intriguingly, although a significant decrease in the intercept with respect to SRT was observed for IS21-ISAs29, the resulting abundances only decreased at lower SRTs and increased at higher SRTs.

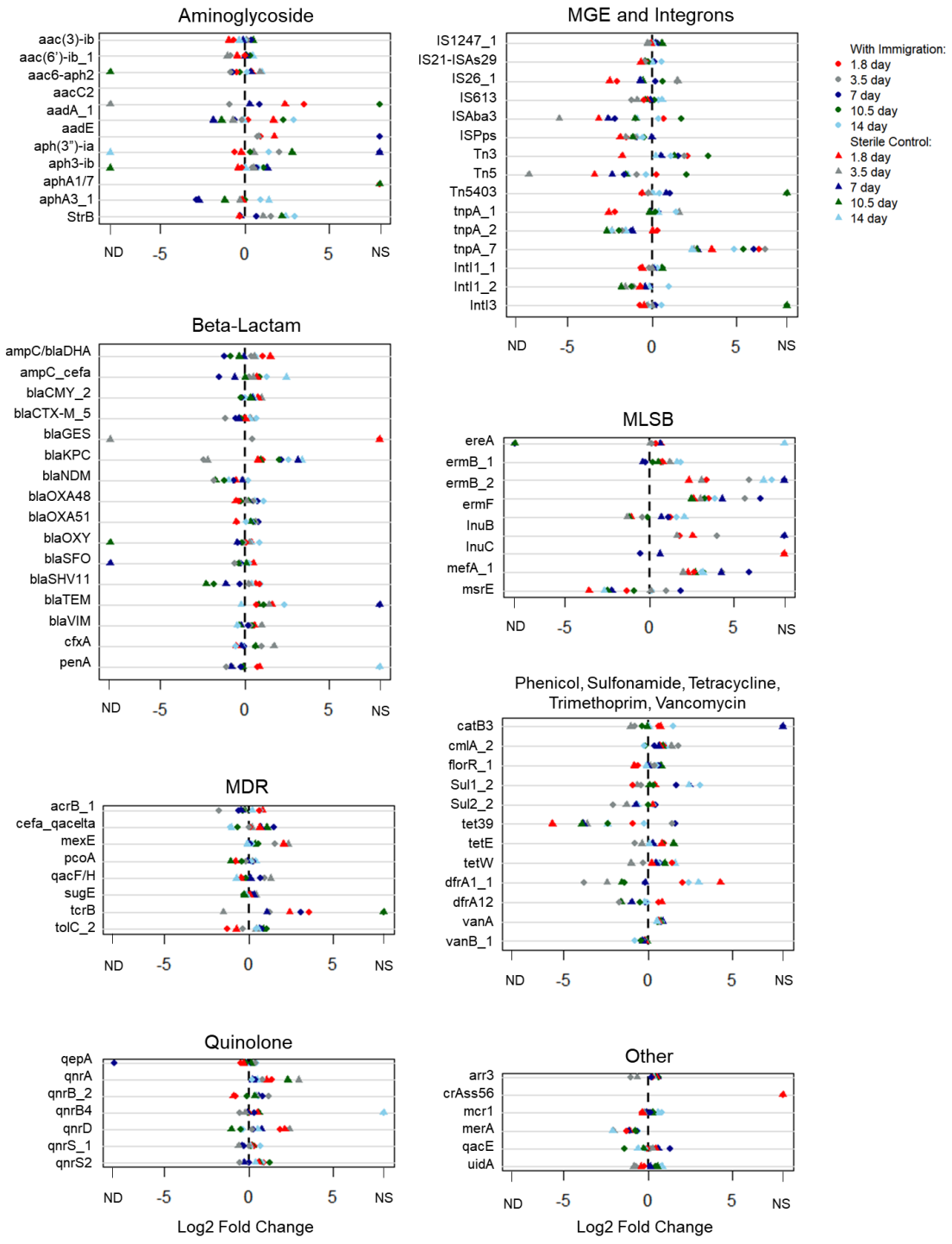


Figure 6.7- Fold change in gene abundance between the start and end of Phase 2. Different colours represent the SRT of operation. Circular points indicate reactors receiving live influent solids during Phase 2. Triangular points represent reactors receiving sterilised influent solids. 'ND' represents reactors in which the gene was not detected at the end of Phase 2 and 'NS' represent reactors in which the gene was not detected at the start of Phase 2.

In group 5 and 6 (Figure 6.8e and 6.8f) gene abundances were significantly impacted by both the SRT and the addition of influent solids. The abundances of genes in Group 5 (6 genes including *str*B [streptomycin resistance], *bla*TEM [beta-lactam resistance], and crAss56 [faecal marker bacteriophage infecting *Bateroidetes* spp.]) increased with the addition of influent solids resulting in a positive correlation with SRT (Figure 6.8e). Whereas the abundance of the genes in Group 6 (only *mex*E and *qnr*B) increased at lower SRTs after receiving influent solids resulting in a negative correlation with SRT (Figure 6.8f). Finally, Group 7 trends (Figure 6.8g) described a single MGE (insertion sequence ISPps) that decreased in abundance with immigration, and the slope of the abundances with SRT became less negative.

By using multiple regression to analyse the specific impact of SRT and immigration on gene abundance, the complexity of AMR dynamics at the immigration interface can be recognised, even when studied using a simplified WWTP system. Without highly controlled and reproducible reactor experiments which focus on the impact of specific operational parameters, the factors impacting the persistence and fate of ARGs in the activated sludge cannot be accurately assessed.

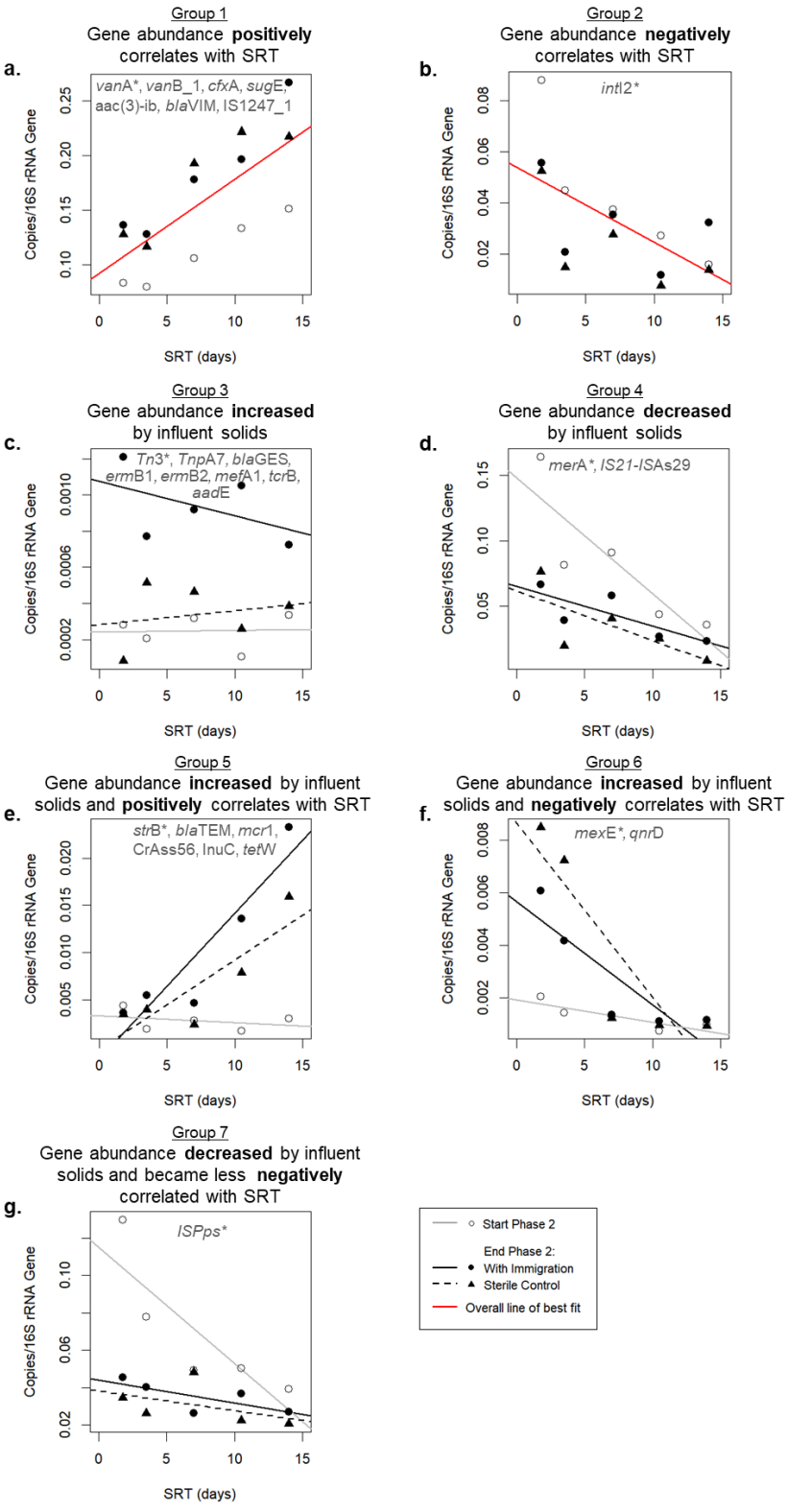


Figure 6.8- Impact of influent solids and SRT on gene abundance in the mixed liquor. Categories were established using multiple linear regression analysis. The genes marked with asterisk (*) are those plotted.

6.4 Discussion

6.4.1 Impact of Immigration varied with SRT

The SRT played a critical role in the formation of unique microbial communities throughout reactor operation. At the end of Phase 1, the SRT core resident population, which reflected differences caused by SRT, accounted for a significant proportion of reads (Figure 6.3). Even among the universal core residents, there were notable variations in the relative abundance of genera. During Phase 2, the impact of SRT on the microbial community composition remained and was reflected along the PCoA1 axis of Figure 6.4. Among the immigrants, the SRT portion accounted for the greatest number of reads demonstrating the dynamics at the immigration interface also to be heavily influenced by SRT (Table 6.1). These findings demonstrate SRT to be a dominant factor in determining both the core resident population and the immigration dynamics between the influent and activated sludge.

Immigration had a greater impact on the microbial community at higher SRTs (Table 6.1). This result contradicts previous studies, which report increased immigration at lower SRTs (16). The disparity between the results could be associated with the lack of controls in full-scale WWTPs. In immigration studies, a common metric used to assess the transfer of taxa between the influent and mixed liquor is the proportion of OTUs (or ASVs and genera) shared between the two environments. Calculating this metric in the current study revealed that even before immigration was introduced, reactors operated at lower SRTs had a greater percentage of ASVs shared with the influent (Table 6.2). In full-scale studies, where no immigration controls are unlikely, this result could be wrongly interpreted as increased immigration at lower SRTs. However when immigration was introduced during Phase 2, the impact was proven to be positively correlated with SRT. The positive correlation between SRT and immigration could be associated with numerous factors

including the lower growth rate required to be maintained within the system, integration over a longer period of time or variations in niche availability.

In processes dominated by neutral assembly mechanisms where equal fitness is assumed, the distribution of genera in the influent wastewater would be expected to be similar to that of the immigrant population (31). However, unique immigrant populations were identified at each SRT showing selection to occur between the influent and activated sludge. This is consistent with previous studies which identified microbial community shifts between the influent and activated sludge in several full-scale treatment plants (32) and distinct microbial communities at different SRTs (33). The differences in the immigrant population at each SRT shows that immigration is influenced by deterministic factors. These results demonstrate that immigration is not absolute and the fate of immigrating genera should be modelled in relation to SRT. Within the immigrant population, the functions and properties of genera vary and likely impact the performance of the wastewater treatment process which highlights the need to consider immigration in process optimisation and design.

6.4.2 Growth Rate of Core Resident and Immigrant Populations varies with SRT

Using a mass balance approach, core residents genera were found to exhibit mainly positive net growth rate in the activated sludge, whilst immigrant population typically had low and often negative net growth rates (Figure 6.6). This is consistent with results from a previous study that also utilised bench-scale flask reactors fed with synthetic wastewater which were operated at a 5-day SRT (15). Based upon reactor theory, organisms with a positive net growth rate would be expected to be maintained within the reactors over time, whilst those with a negative net growth rate would likely be washed out if immigration ceased. Therefore, core resident populations and immigrants which display a positive net growth rate would be expected to persist over time, whilst

the immigrant populations are likely dependent on continuous immigration to be maintained within the system.

The majority of core residents exhibited positive net growth rates, however at higher SRTs some universal core residents exhibited lower net growth rates. Analysing the ASV composition of a selection of these genera, it was revealed that at higher SRTs, there was neutral replacement of ASVs when immigration was introduced. Among the genera *Pseudomonas* and *Zoogloea*, after immigration was introduced, the ASV composition resembled that of the influent solids (Figure S6.4). Whilst at lower SRTs, the immigration dynamics at the ASV level appeared more complex. In genus *Pseudomonas*, negative or low ($< 0.005\ d^{-1}$) net growth rates were observed in the reactors operated at 7-14 day SRTs which all contained similar ASVs, suggesting that those introduced with the influent solids were not well adapted to grow in the activated sludge. Whereas in *Zoogloea,* the ASV composition was similar even in reactors exhibiting positive net growth rates suggesting the low net growth rate to be associated with other factors such as competition due to increased biomass concentration at higher SRTs (Figure S6.3) or the greater proportion of immigrants introduced. Taken together, these results suggest that WWTPs operated at higher SRT may be more impacted by neutral processes at the ASV level.

Universal immigrants, which were detected at all SRTs, typically occurred in higher abundance in the influent solids than the SRT immigrant population. This suggests that genera occurring in high relative abundance in the influent wastewater are better adapted to immigrate under different conditions. However, selection was still observed between the influent and activated sludge among these genera, consequently influent abundance cannot be used as a definitive predictor of the fate of a given genus.

In the SRT immigrant population, a greater proportion of genera had positive net growth rate at lower SRTs (Figure S6.1), this could be an indication of more successful colonisation as these genera would be more likely to persist in the absence of immigration. This is consistent with previous studies, which demonstrate the success of colonisation to increase at lower SRTs (17). Another possibility is the selection of specific metabolisms among the SRT immigrants. The MiDAS database (34) was used to assess the metabolic capabilities of these genera. Those not included in the database were excluded from this analysis to ensure the data used was consistent. At all SRTs other than 1.8-days, the majority of SRT immigrants displayed aerobic heterotrophic activity. At a 1.8-days SRT, a greater proportion of fermenters were present. Given that the reactors were operated under aerobic conditions this result was unexpected. However, it should be noted that this analysis was restricted by data availability, as only 40 % (9 genera) of the SRT immigrants could be classified within the 1.8-day reactor set. Consequently, vital information on the metabolic capabilities and the related niche availability could not be studied in depth.

6.4.3 AMR Dynamics at the Immigration Interface

ARGs and MGEs were present within the reactors prior to immigration being introduced (Phase 2 start; supplementary Table S6.4). These genes conferred resistance to numerous classes of antimicrobials including Beta-lactams, Aminoglycosides, Quinolones and MLSB's demonstrating that they are capable of persisting in the absence of selective pressures such as antimicrobials. During Phase 2, immigration was introduced into the reactors. Overall, the abundance of 52 $\pm$ 8 of the 89 genes detected increased in each reactor set with immigration (Figure 6.7) and slight increases in gene diversity were observed. This result is comparable to a previous study of reactor which also received Syntho supplemented with influent solids. In this study, the abundances of over 70% (11 out of 15) of the ARGs investigated increased with immigration (14).

The fold change of ARGs and MGEs was impacted by ARG class, gene type and reactor conditions (Figure 6.7). Among the MLSB ARGs, the majority displayed a positive log change with the addition of influent solids. This suggests that hosts carrying these genes, which have been found to belong to numerous phyla (32), are well adapted to move between the influent and activated sludge environments. In other cases, the fold change differed between reactor sets (at different SRTs), for example, the *tet*39 and *erm*B_2 resistance genes showed highly variable fold changes between the reactors, whilst with others such as *qep*A and *qnr*B4 the fold change remained similar in all reactors. These results demonstrate that studies should consider numerous genes when assessing the impact of operational parameters and immigration on AMR, as the persistence and dynamics of ARGs can be highly variable based upon a number of factors.

Within the MGEs, the abundance of genes such as TnpA_7 and Tn3 increased with immigration, whilst others such as TnpA_2 and IntI1_2 reduced (Figure 6.7). Mobile genetic elements have been demonstrated to play an important role in the adaptation of bacteria to local conditions (35). Studies demonstrating a change in the hosts of ARGs between the influent wastewater and activated sludge have suggested MGEs to be crucial in persistence and enable gene transfer to capable hosts (36). Thus, the changes in the abundance of MGEs with immigration could impact the fate and persistence of ARGs in the activated sludge.

6.4.4 The Relationship Between SRT, Immigration and AMR

The relationship between microbial community composition and AMR has been widely reported in the literature (37–39). However, without a thorough understanding of the community assembly process and the populations driving AMR, effective mitigation strategies cannot be developed. The controlled conditions applied during this study enabled the microbial community to be further defined into specific populations, with the origins of each clearly understood. In full-scale WWTPs

this cannot be achieved due to the lack of appropriate controls, and the multitude of variables impacting the microbial community at any one time. By managing these limitations using a series of highly controlled reactors, the microbial community could be accurately divided into populations and seminal parallels with ARG patterns could be drawn.

The core resident microbial community was defined as genera which were present in the reactors under all conditions and were not dependent upon immigration. Within this group, genera either belonged to the universal core residents, meaning that they occurred in all reactors regardless of SRT, or the SRT core resident genera which appeared at only selected SRT. In ARGs which are not impacted by immigration, little change in the abundance would be expected after Phase 2, however patterns may be observed based upon the core resident populations. This was as shown in group 1 (Figure 6.8a), where gene abundance increased with SRT but was not significantly impacted by immigration, suggesting these genes to be associated with the core resident populations. Furthermore, the fitness cost of ARG carriage could also be an influencing factor which has been demonstrated to negatively impact the growth rate of ARB which is likely to be more detrimental at lower SRTs where a higher growth rate is required. Similarly in group 2 (Figure 6.8b.), the negative correlation between SRT and gene abundance could be associated with changes in the abundance of host genera in the core resident population.

Among the immigrant populations, universal immigrants were genera which were detected in all reactors with immigration and typically accounted for a similar number of reads at all SRTs (Table 6.1). In ARGs associated with this population, an increase in the abundance would be expected with immigration which would likely be similar at all SRTs. This was as found in group 3 (Figure 6.8c) whereby the gene abundance increased with the addition of influent solids, but no changes were found with SRT.

Among the SRT immigrant populations, which included genera which were only detected at selected SRTs, the reads contributed positively correlated with the SRT. In ARGs thought to be associated with this population, an increase in abundance would be expected with immigration which positively correlates with SRT as shown in group 5 (Figure 6.8c). Among the genes in group 5 showing a significant increase in abundance with immigration and SRT was CrAss56, which is a bacteriophage which infects *Bacteriodes* that is used a marker of human faecal contamination. The abundance of this gene positively correlated with SRT, and may suggest the increase in abundance of ARGs in this group could be directly related to an increase in the influence of faecal matter. Although the SRT immigrant population contributed a greater number of reads at higher SRTs, there were also selected genera within this population which were exclusive to lower SRTs such as the genus *Paludibacter* (family *Paludibacteraceae*). In ARGs associated with this fraction of the population, a negative correlation between gene abundance and SRT could be expected as exemplified in group 6.

Previous studies in full-scale WWTPs have shown changes in ARG profiles between the influent and activated sludge to significantly correlate with the changes in the microbial community composition (32). However, without a thorough understanding of this relationship appropriate control measures cannot be developed. In this study for the first time, the role of the community assembly process and specific populations in the persistence of ARGs in the activated sludge could be defined.

6.5 Implications and Future Considerations

The results obtained from this study demonstrate the limitations associated with previous studies which define the impact of immigration based upon shared taxa between the influent and activated sludge. The similarity of the activated sludge microbial community with that of the influent is likely to vary between locations, and is not an indication of the relative impact of immigration. In this study, by including no immigration controls, immigration was shown to have a greater impact on the microbial community of the activated sludge at higher SRTs. Future studies aiming to optimise the microbial community in the activated sludge should consider the contribution of influent immigration as an important source of diversity and the impact of operational parameters on this process.

By controlling reactor conditions and developing appropriate population definitions, the dynamics between the microbial community composition and AMR could be studied in depth in relation to the community assembly process. Numerous patterns were observed for different ARG and MGEs which highlighted the complexity of AMR in the environment and the difficulty in establishing appropriate treatment strategies. In the future, studies aiming to reduce AMR in wastewater treatment plants should carefully consider which ARGs to prioritise and the factors impacting their persistence within this environment.

6.6 Supplementary Material

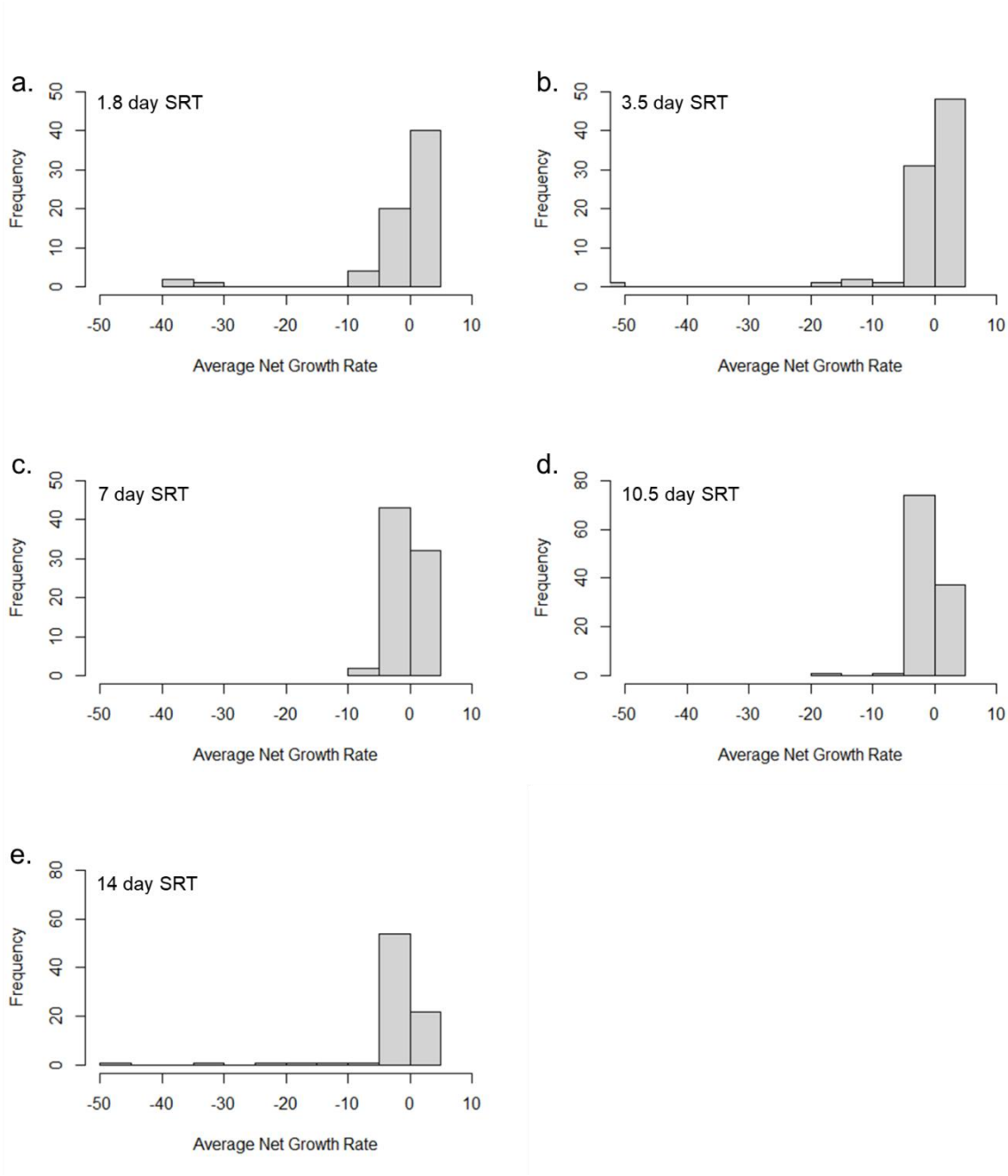


Figure S6.1- Histogram of the growth rate (μ) of genera classified as SRT immigrants at the end of Phase 2. The frequency represents the number of genera observed to have each net growth rate. Reactors were operated with a) 1.8 day b) 3.5 day c) 7 day d) 10.5 day e) 14 day SRT

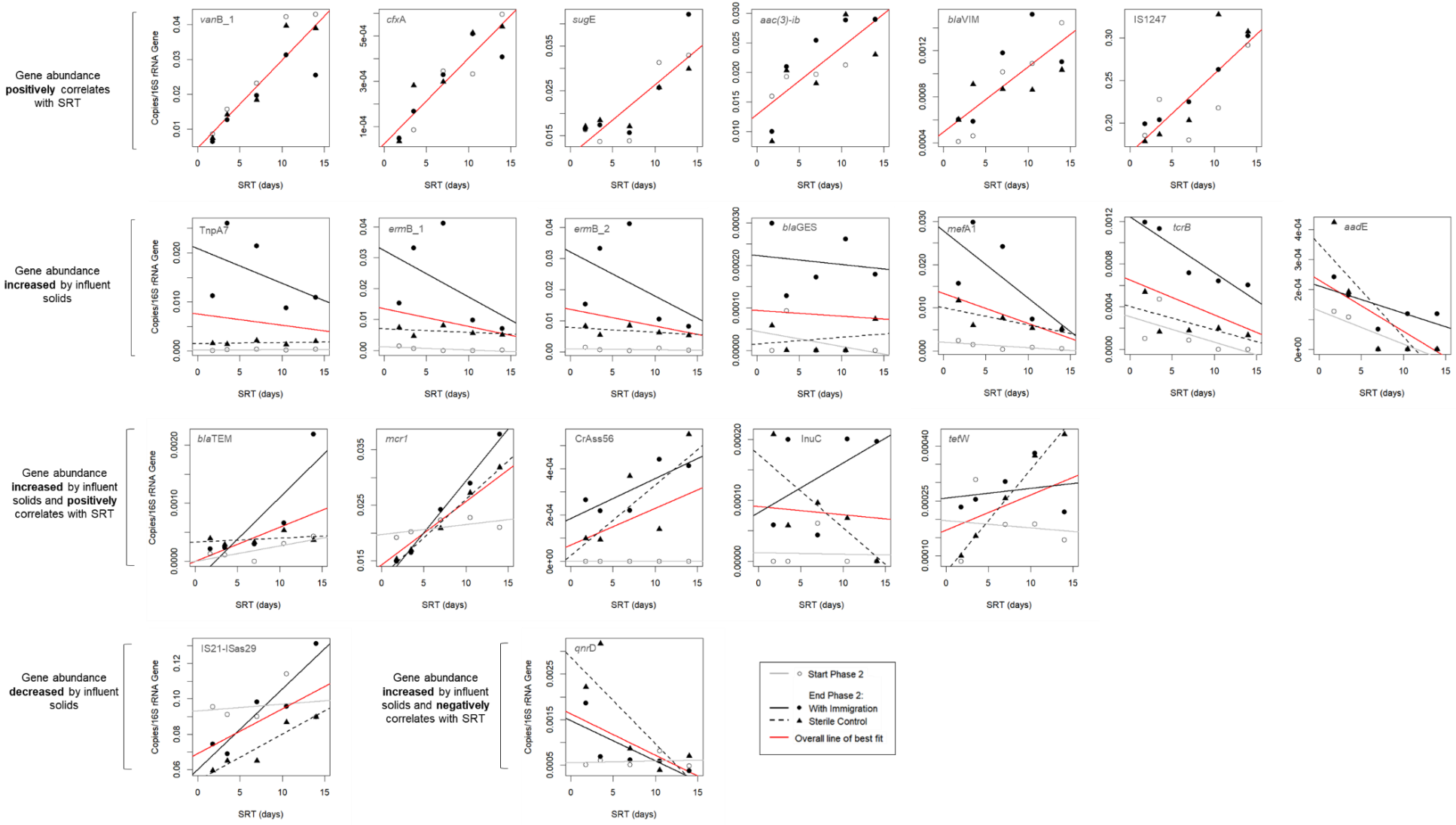


Figure S6.2- Genes significantly impacted by SRT and immigration. Analysed samples were a pool of the triplicate reactors receiving influent solids and operated at the same SRT. Quantitative PCR of each pool was conducted in triplicate.

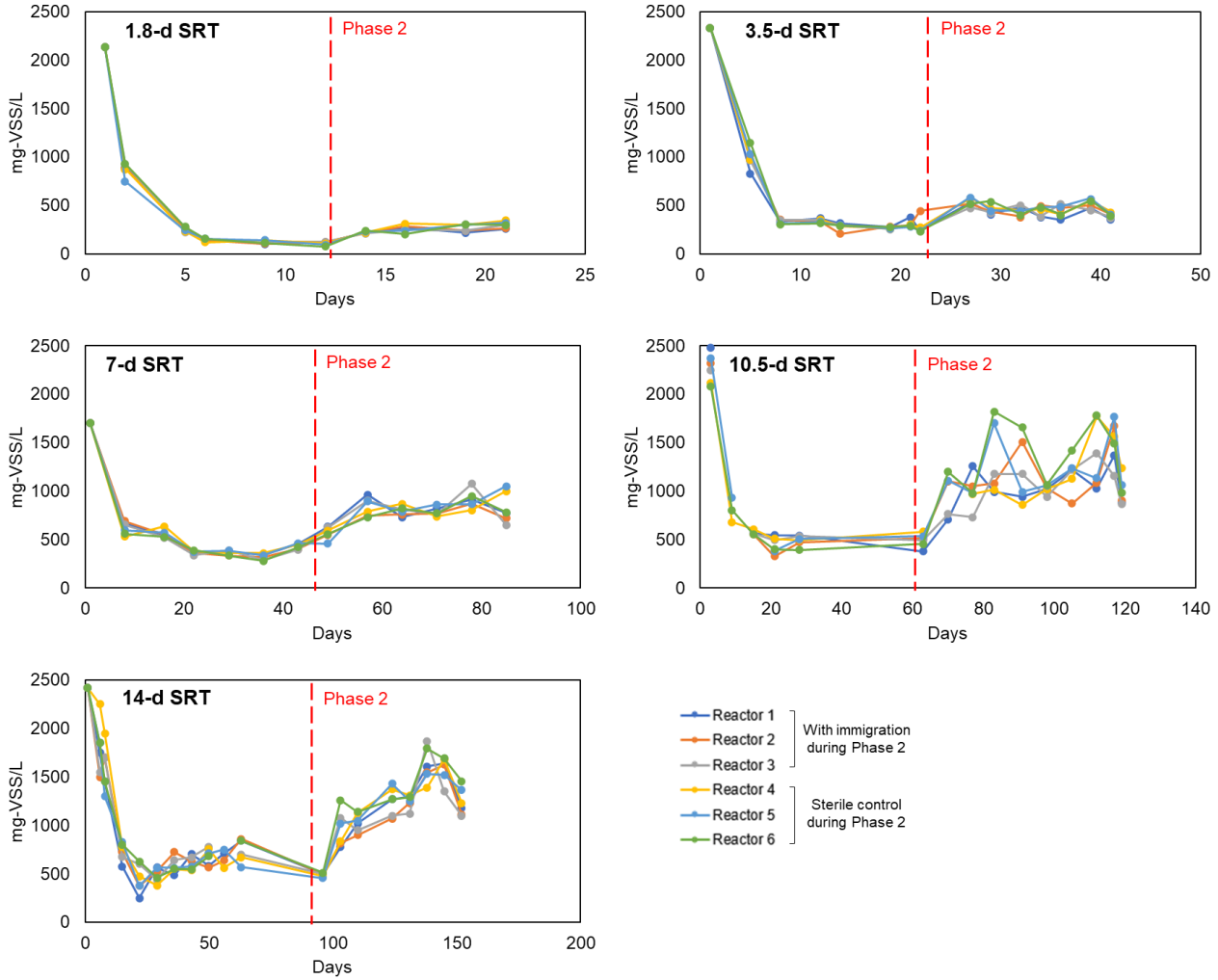


Figure S6.3- Volatile suspended solid data from reactor operation. Samples were taken from reactors and analysed in triplicate. Periods with missing data were due to limited lab access during COVID-19 pandemic.

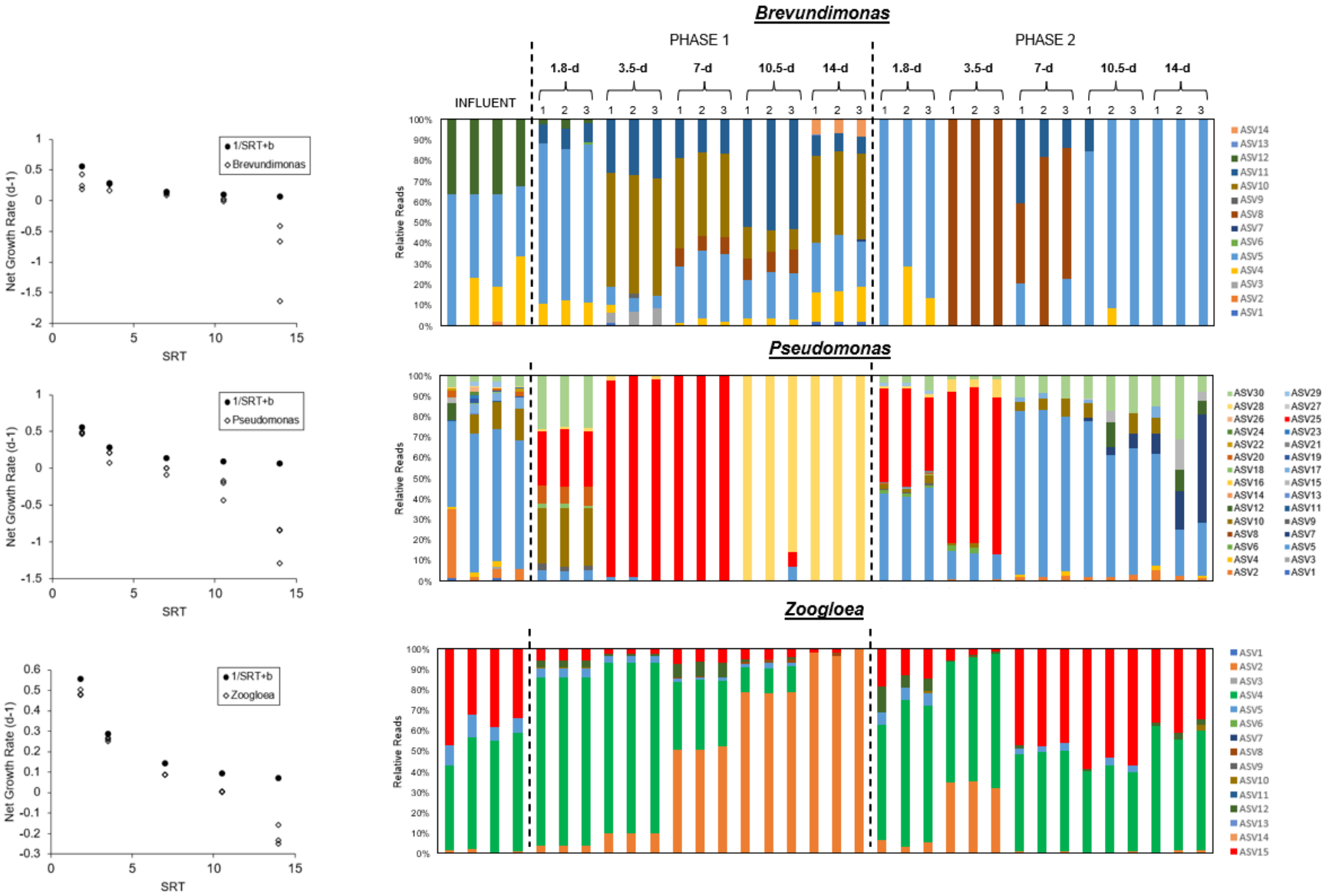


Figure S6.4 - ASV observed among SRT immigrants displaying positive net growth rates at lower SRTs

Table S6.1- Resistomap quantitative PCR targets and primers

Gene	Target	Forward Primer	Reverse Primer
16S rRNA	16S rRNA	GGGTTGCGCTCGTTGC	ATGGYTGTCGTCAGCTCGTG
aac(3)-ib	Aminoglycoside	CAGCGAGACGTTCATCGC	CACGCTTCAGGTGGCTAATC
aac(6')-Ib_1	Aminoglycoside	CGTCGCCGAGCAACTTG	CGGTACCTTGCCTCTCAAACC
aac6-aph2	Aminoglycoside	CCAAGAGCAATAAGGGCATACCAA	GCCACACTATCATAACCACTACCG
aacA/aphD	Aminoglycoside	AGAGCCTTGGGAAGATGAAGTTT	TTGATCCATACCATAGACTATCTCATCA
aacC2	Aminoglycoside	ACGGCATTCTCGATTGCTTT	CCGAGCTTCACGTAAGCATTT
aadA_1	Aminoglycoside	GTTGTGCACGACGACATCATT	GGCTCGAAGATACCTGCAAGAA
aadE	Aminoglycoside	TACCTTATTGCCCTTGGAAGAGTTA	GGAACTATGTCCCTTTTAATTCTACAATCT
aph(3")-ia	Aminoglycoside	TAACAGCGATCGCGTATTTCG	TCCGACTCGTCCAACATCAATA
aph3-ib	Aminoglycoside	AACAGGTTTGGGAGGCGATG	CGCAACAAGCCTCTCCTGAA
aphA1/7	Aminoglycoside	TGAACAAGTCTGGAAAGAAATGCA	CCTATTAATTTCCCCTCGTCAAAAA
aphA3_1	Aminoglycoside	AAAAGCCCGAAGAGGAACTTG	CATCTTTCACAAAGATGTTGCTGTCT
strB	Aminoglycoside	GCTCGGTCGTGAGAACAATCT	CAATTTCGGTCGCCTGGTAGT
ampC_cefa	Beta Lactam	CAGGATCTGATGTGGGAGAACTA	TCGGGAACCATTTGTTGGC
ampC/blaDHA	Beta Lactam	TGGCCGCAGCAGAAAGA	CCGTTTTATGCACCCAGGAA
blaCMY_2	Beta Lactam	AAAGCCTCAT GGGTGCATAAA	ATAGCTTTTGTTTGCCAGCATCA
blaCTX-M_5	Beta Lactam	GCGATAACGTGGCGATGAAT	GTCGAGACGGAACGTTTCGT
blaCTX-M_8	Beta Lactam	CGTCACGCTGTTGTTAGGAA	CGCTCATCAGCACGATAAAG
blaGES	Beta Lactam	GCAATGTGCTCAACGTTCAAG	GTGCCTGAGTCAATTCTTTCAAAG
blaKPC	Beta Lactam	GCCGCCAATTTGTTGCTGAA	GCCGGTCGTGTTTCCCTTT
blaNDM	Beta Lactam	GGCCACACCAGTGACAATATCA	CAGGCAGCCACCAAAAGC
blaOXA48	Beta Lactam	TGTTTTTGGTGGCATCGAT	GTAAMRATGCTTGGTTCGC
blaOXA51	Beta Lactam	CGACCGAGTATGTACCTGCTTC	TCAAGTCCAATACGACGAGCTA
blaOXY	Beta Lactam	CGTTCAGGCGGCAGGTT	GCCGCGATATAAGATTTGAGAATT
blaSFO	Beta Lactam	CCGCCGCCATCCAGTA	GGGCCGCCAAGATGCT
blaSHV11	Beta Lactam	TTGACCGCTGGGAAACGG	TCCGGTCTTATCGGCGATAAAC
blaTEM	Beta Lactam	CGCCGCATACACTATTCTCAG	GCTTCATTCAGCTCCGGTTC
blaVIM	Beta Lactam	GCACTTCTCGCGGAGATTG	CGACGGTGATGCGTACGTT
cfxA	Beta Lactam	TCATTCCTCGTTCAAGTTTTCAGA	TGCAGCACCAAGAGGAGATGT
penA	Beta Lactam	AGACGGTAACGTATAACTTTTTGAAAGA	GCGTGTAGCCGGCAATG
intI1_1	Integrons	CGAACGAGTGGCGGAGGGTG	TACCCGAGAGCTTGGCACCCA

Table S6.1- Continued

Gene	Target antibiotics	Forward Primer	Reverse Primer
intI1_2	Integrons	CGAAGTCGAGGCATTTCTGTC	GCCTTCCAGAAAACCGAGGA
intI3	Integrons	CAGGTGCTGGGCATGGA	CCTGGGCAGCATCACCA
acrB_1	MDR	AGTCGGTGTTCGCCGTTAAC	CAAGGAAACGAACGCAATACC
cefa_qacelta	MDR	TAGTTGGCGAAGTAATCGCAAC	TGCGATGCCATAACCGATTATG
mexE	MDR	GGTCAGCACCGACAAGGTCTAC	AGCTCGACGTACTTGAGGAACAC
pcoA	MDR	TGGCGTATGGAGTTTCAATGC	GAATAATGCCGTGCCAGTGAA
qacF/H	MDR	CTGAAGTCTAGCCATGGATTCACTAG	CAAGCAATAGCTGCCACAAGC
sugE	MDR	CTTAGTTATTGCTGGTCTGCTGGA	GCATCGGGTTAGCGGACTC
tcrB	MDR	GTGCCGGAACTCAAGTAGCA	GCACCGACTGCTGGACTTAA
tolC_2	MDR	CAGGCAGAGAACCTGATGCA	CGCAATTCCGGGTTGCT
IS1247_1	MGE	CGGCCGTCACTGACCAA	TCGGCAGGTTGGTGACG
IS21-ISAs29	MGE	GGTCCGTCAGGCACAAGTC	GGGATCGTATCGGCAAGCC
IS26_1	MGE	ATGGATGAAACCTACGTGAAGGTC	CGGTACTTAATCTGTCGGTGTTCA
IS613	MGE	AGGTTCGGACTCAATGCAACA	TTCAGCACATACCGCCTTGAT
ISAba3	MGE	TCAGAGGCAGCGGTATACGA	GGTTGATTCAGTTAAAGTACGTAAAACTTT
ISPps	MGE	CACACTGCAAAAACGCATCCT	TGTCTTTGGCGTCACAGTTCTC
Tn3	MGE	GCTGAGGTGTTCAGCTACATCC	GCTGAGGTAGTCACAGGCATTC
Tn5	MGE	TCAGAGGCAGCGGTATACGA	GGTTGATTCAGTTAAAGTACGTAAAACTTT
Tn5403	MGE	AAGCGAATGGCGCGAAC	CGCGCAGGGTAAACTGC
tnpA_1	MGE	GCCGCACTGTCGATTTTTATC	GCGGGATCTGCCACTTCTT
tnpA_2	MGE	CCGATCACGGAAAGCTCAAG	GGCTCGCATGACTTCGAATC
tnpA_7	MGE	AATTGATGCGGACGGCTTAA	TCACCAAACTGTTTATGGAGTCGTT
ereA	MLSB	GATAATTCTGCTGGCGCACA	GCAGGCGTGGTCACAAC
ermA	MLSB	TCGTTGAGAAGGGATTTGCGA	TTGCATGCTTCAAAGCCTGTC
ermB_1	MLSB	TAAAGGGCATTTAACGACGAAACT	TTTATACCTCTGTTTGTTAGGGAATTGAA
ermB_2	MLSB	GAACACTAGGGTTGTTCTTGCA	CTGGAACATCTGTGGTATGGC
ermF	MLSB	TCTGATGCCCGAAATGTTCAAG	TGAAGGACAATTGAACCTCCCA
lnuB	MLSB	GGATCGTTTACCAAAGGAGAAGG	AGCATAGCCTTCGTATCAGGAA
lnuC	MLSB	GGGTGTAGATGCTCTTCTTGGA	CTTTACCCGAAAGAGTTTCTACCG
mefA_1	MLSB	CCGTAGCATTGGAACAGCTTTT	AAACGGAGTATAAGAGTGCTGCAA

Table S6.1- Continued

Gene	Target antibiotics	Forward Primer	Reverse Primer
msrE	MLSB	CGGCAGATGGTCTGAGCTTAAA	CGCACTCTTCCTGCATAAAGGA
arr3	Other	GATCGTCTTCGAACGGTCCTG	TTTGGCGATTGGTGACTTGCT
crAss56	Other	CAGAAGTACAAACTCCTAAAAAACG TAGAG	GATGACCAATAAACAAGCCATTA GC
mcr1	Other	CACATCGACGGCGTATTCTG	CAACGAGCATACCGACATCG
merA	Other	GTGCCGTCCAAGATCATG	GGTGGAAGTCCAGTAGGGTGA
qacEΔ1_1	Other	TCGCAACATCCGCATTAAAA	ATGGATTTCAGAACCAGAGAAAG AAA
sat4	Other	GAATGGGCAAAGCATAAAAACTTG	CCGATTTTGAAACCACAATTATG ATA
uidA	Other	AACCACGCGTCTGTTGACTG	CCCGGTTGCCAGAGGTG
catB3	Phenicol	GCACTCGATGCCTTCCAAAA	AGAGCCGATCCAAACGTCAT
cmlA_2	Phenicol	TAGGAAGCATCGGAACGTTGAT	CAGACCGAGCACGACTGTTG
floR_1	Phenicol	ATTGTCTTCACGGTGTCCGTTA	CCGCGATGTCGTCGAACT
qepA	Quinolone	GGGCATCGCGCTGTTC	GCGCATCGGTGAAGCC
qnrA	Quinolone	AGGATTTCTCACGCCAGGATT	CCGCTTTCAATGAAACTGCAA
qnrB_2	Quinolone	CGACGTTCAGTGGTTCAGATCTC	GCCAAGCCGCTCCATGAG
qnrB4	Quinolone	TCACCACCCGCACCTG	GGATATCTAAATCGCCCAGTTCC
qnrD	Quinolone	CGCTGGAATGGCACTGTGA	GCTCTCCATCCAACTTCACTCC
qnrS_1	Quinolone	CCACTTTGATGTCGCAGATCTTC	CCCTCTCCATATTGGCATAGGAA A
qnrS2	Quinolone	TCCCGAGCAAACTTTGCCAA	GGTGAGTCCCTATCCAGCGA
sul1_2	Sulfonamide	GCCGATGAGATCAGACGTATTG	CGCATAGCGCTGGGTTTC
sul2_2	Sulfonamide	TCATCTGCCAAACTCGTCGTTA	GTCAAAGAACGCCGCAATGT
A. baumannii	Taxonomic	TCTTGGTGGTCACTTGAAGC	ACTCTTGTGGTTGTGGAGCA
Campylobacter	Taxonomic	CTGCTTAACACAAGTTGAGTAGG	TTCCTTAGGTACCGTCAGAA
Enterococci	Taxonomic	AGAAATTCCAAACGAACTTG	CAGTGCTCTACCTCCATCATT
K. pneumoniae	Taxonomic	ACGGCCGAATATGACGAATTC	AGAGTGATCTGCTCATGAA
P. aeruginosa	Taxonomic	AGCGTTCGTCCTGCACAAGT	TCCACCATGCTCAGGGAGAT
Staphylococci	Taxonomic	CGCAACGTTCAATTTAATTTTGTTAA	TGGTCTTTCTGCATTCCTGGA
tet39	Tetracycline	TATAGCGGGTCCGGTAATAGGTG	CCATAACGATCCTGCCCATAGAT AAC
tetE	Tetracycline	TTGGCGCTGTATGCAATGAT	CGACGACCTATGCGATCTGA
tetW	Tetracycline	ATGAACATTCCCACCGTTATCTTT	ATATCGGCGGAGAGCTTATCC
dfrA1_1	Trimethoprim	GGAATGGCCCTGATATTCCA	AGTCTTGCGTCCAACCAACAG

Table S6.1- Continued

Gene	Target antibiotics	Forward Primer	Reverse Primer
dfrA12	Trimethoprim	CCTCTACCGAACCGTCACACA	GCGACAGCGTTGAAACAACTAC
vanA	Vancomycin	GGGCTGTGAGGTCGGTTG	TTCAGTACAATGCGGCCGTTA
vanB_1	Vancomycin	TTGTCGGCGAAGTGGATCA	AGCCTTTTTCCGGCTCGTT
vanRA_1	Vancomycin	CCCTTACTCCCACCGAGTTTT	TTCGTCGCCCCATATCTCAT
vanSA	Vancomycin	CGCGTCATGCTTTCAAAATTC	TCCGCAGAAAGCTCAATTTGTT

Table S6.2: Linear regression of relationship between SRT and reads/ASVs shared with influent microbial

	Before Immigration[a]		After Immigration				Total Shared	
			Shared with P2S[b]		Newly introduced			
Factor	ASVs	Reads	ASVs	Reads	ASV	Reads	ASVs	Reads
R2	0.88	0.97	0.66	0.46	0.05	0.68	-0.031	0.50
P value	$1.20e^{-7}$	$1.65e^{-11}$	$1.50e^{-4}$	$3.33e^{-3}$	$2.11e^{-1}$	$8.76e^{-5}$	$4.58e^{-1}$	$1.90e^{-3}$
Intercept	30.56	59.67	19.63	50.33	7.49	30.13	49.75	57.82
Slope	-1.25	-1.90	-0.92	-2.32	0.50	1.17	0.25	-1.82

[a]Before Immigration indicates samples taken at the start of Phase 2

[b]ASV/reads present at the end of Phase 2 (after immigration), which were also present at the beginning of Phase 2 (before immigration). Thus, they are considered resident (core or non-core) ASV.

Table S6.3: Richness of ARGs detected using qPCR array

	Number of ARGs			Number of MGEs		
SRT (days)	Phase 2 Start	With Immigration	Influent	Phase 2 Start	With Immigration	Influent
1.8	65	68	72	15	15	15
3.5	66	69	69	15	15	15
7	60	66	72	15	15	15
10.5	60	68		13	15	
14	58	66		15	15	

Table S6.4- Raw data from quantitative PCR array normalised to 16S rRNA gene

Gene	14d Imm	14d SC	14d Start	10.5 Imm	10.5 SC	10.5 Start	7d Imm	7d SC	7d Start	3.5d Imm	3.5d SC	3.5d Start	1.8d Imm	1.8d SC	1.8d Start	Inf 1	inf 2	Inf 3
intI3	2.2E+00	1.3E+00	1.5E+00	1.4E+00	1.9E+00		1.5E+00	1.4E+00	1.4E+00	1.4E+00	1.7E+00	1.7E+00	1.2E+00	1.4E+00	2.0E+00	1.1E+00	9.2E-01	9.9E-01
intI1_1	1.2E+00	1.0E+00	9.3E-01	1.1E+00	1.1E+00	7.3E-01	9.7E-01	9.6E-01	9.3E-01	7.5E-01	8.3E-01	8.2E-01	4.4E-01	4.8E-01	7.0E-01	4.0E-01	6.9E-01	3.7E-01
intI1_2	3.2E-02	1.4E-02	1.6E-02	1.2E-02	7.7E-03	2.7E-02	3.5E-02	2.8E-02	3.8E-02	2.1E-02	1.5E-02	4.5E-02	5.6E-02	5.2E-02	8.8E-02	3.6E-02	1.9E-02	3.5E-02
qepA	6.8E-01	4.5E-01	5.0E-01	4.7E-01	5.8E-01	5.4E-01		4.6E-01	5.0E-01	6.3E-01	5.3E-01	4.7E-01	3.5E-01	4.0E-01	4.9E-01	2.3E-01	3.3E-01	2.3E-01
qnrB4	1.0E-01	8.0E-02		7.2E-02	1.1E-01	7.4E-02	7.7E-02	5.8E-02	6.2E-02	5.3E-02	6.8E-02	7.8E-02	7.1E-02	9.9E-02	6.9E-02	4.2E-02	6.6E-02	4.4E-02
qnrB_2	4.0E-03	3.5E-03	2.7E-03	3.0E-03	4.0E-03	3.1E-03	3.8E-03	3.1E-03	2.1E-03	2.9E-03	1.8E-03	1.3E-03	1.4E-03	1.3E-03	2.5E-03	3.9E-03	2.2E-03	2.5E-03
qnrS2	2.3E-03	1.7E-03	1.3E-03	2.5E-03	1.6E-03	1.1E-03	1.6E-03	1.3E-03	1.6E-03	1.5E-03	1.8E-03	2.2E-03	3.4E-03	2.7E-03	1.8E-03	7.1E-03	3.1E-03	6.2E-03
qnrD	3.8E-04	7.0E-04	4.8E-04	5.9E-04	3.9E-04	8.1E-04	6.2E-04	8.6E-04	5.1E-04	6.8E-04	3.2E-03	6.0E-04	1.9E-03	2.2E-03	5.1E-04	3.3E-04	6.5E-04	3.4E-04
qnrS_1	2.9E-04	1.9E-04	1.8E-04	2.2E-04	2.3E-04	1.9E-04	1.7E-04	1.5E-04	2.1E-04	1.5E-04	9.3E-05	1.4E-04	1.3E-04	1.1E-04	1.0E-04	7.1E-04	1.3E-04	4.7E-04
qnrA	1.5E-04	1.9E-04	1.3E-04	1.3E-04	3.6E-04	7.2E-05	8.8E-05	7.9E-05	6.8E-05	2.1E-04	9.3E-04	1.2E-04	1.7E-04	1.4E-04	6.7E-05	1.1E-04	1.4E-04	1.3E-04
IS1247_1	3.0E-01	3.1E-01	2.9E-01	2.6E-01	3.3E-01	2.2E-01	2.2E-01	2.0E-01	1.8E-01	2.0E-01	1.9E-01	2.3E-01	2.0E-01	1.8E-01	1.9E-01	1.6E-01	2.1E-01	1.5E-01
Tn5403	3.1E-01	2.4E-01	2.3E-01	2.2E-01	2.1E-01		1.8E-01	1.6E-01	8.9E-02	1.3E-01	1.4E-01	1.6E-01	9.8E-02	9.7E-02	1.5E-01	9.6E-02	2.0E-01	9.1E-02
IS21-ISAs29	1.3E-01	9.0E-02	9.0E-02	9.6E-02	8.7E-02	1.1E-01	9.8E-02	6.5E-02	9.0E-02	6.9E-02	6.5E-02	9.1E-02	7.4E-02	5.9E-02	9.5E-02	9.0E-02	7.5E-02	8.5E-02
ISPps	2.7E-02	2.1E-02	3.9E-02	3.7E-02	2.2E-02	5.0E-02	2.6E-02	4.8E-02	4.9E-02	4.0E-02	2.6E-02	7.8E-02	4.6E-02	3.5E-02	1.3E-01	1.4E-01	2.4E-01	1.2E-01
tnpA_2	1.4E-02	2.6E-02	7.5E-02	5.2E-02	3.2E-02	2.0E-01	5.7E-02	5.8E-02	1.3E-01	4.9E-02	3.1E-02	1.6E-01	5.7E-02	4.8E-02	4.6E-02	6.8E-03	1.1E-02	7.6E-03
ISAba3	1.6E-02	6.3E-03	1.2E-02	2.0E-02	3.0E-03	6.0E-03	2.0E-02	1.5E-02	9.1E-02	3.4E-03	1.4E-04	6.6E-03	1.0E-02	6.8E-04	6.1E-03	1.5E-01	9.9E-03	1.3E-01
Tn5	9.2E-03	4.2E-03	1.1E-02	2.0E-02	1.7E-03	5.0E-03	2.4E-02	1.4E-02	7.5E-02	3.4E-03	3.9E-05	6.3E-03	9.1E-03	7.0E-04	7.5E-03	1.6E-01	8.9E-03	1.3E-01
tnpA_7	1.1E-02	1.9E-03	3.8E-04	8.8E-03	1.3E-03	2.0E-04	2.1E-02	2.1E-03	3.3E-04	2.6E-02	1.4E-03	2.5E-04	1.1E-02	1.6E-03	1.4E-04	1.5E-01	2.7E-03	2.2E-01
IS26_1	4.1E-03	9.1E-04	1.4E-03	3.5E-03	1.5E-03	2.2E-03	4.4E-03	2.4E-03	3.9E-03	6.8E-03	6.9E-03	2.4E-03	1.9E-02	1.4E-02	7.9E-02	3.3E-02	2.2E-03	2.2E-02
tnpA_1	1.4E-03	7.0E-04	5.3E-04	1.6E-03	1.3E-03	1.4E-03	2.0E-03	1.4E-03	1.6E-03	4.3E-03	4.8E-03	1.6E-03	1.0E-02	7.9E-03	4.7E-02	2.2E-02	1.4E-03	1.4E-02
IS613	1.4E-03	1.7E-03	1.1E-03	1.6E-03	1.1E-03	1.5E-03	9.9E-04	9.7E-04	1.1E-03	6.0E-04	7.7E-04	1.4E-03	8.2E-04	9.2E-04	1.1E-03	7.4E-03	3.4E-03	4.1E-03
Tn3	7.2E-04	3.8E-04	3.4E-04	1.0E-03	2.6E-04	1.1E-04	9.2E-04	4.6E-04	3.2E-04	7.7E-04	5.1E-04	2.1E-04	1.2E-03	8.2E-05	2.8E-04	4.9E-03	6.4E-04	7.0E-03
vanA	2.7E-01	2.2E-01	1.5E-01	2.0E-01	2.2E-01	1.3E-01	1.8E-01	1.9E-01	1.1E-01	1.3E-01	1.2E-01	8.0E-02	1.4E-01	1.3E-01	8.4E-02	7.9E-02	1.2E-01	8.3E-02
vanB_1	2.5E-02	3.9E-02	4.3E-02	3.1E-02	4.0E-02	4.2E-02	2.0E-02	1.8E-02	2.3E-02	1.3E-02	1.4E-02	1.6E-02	6.5E-03	7.5E-03	8.6E-03	6.4E-03	1.4E-02	7.9E-03
blaOXY	2.0E-01	1.3E-01	1.1E-01	1.9E-01		2.1E-01	1.3E-01	1.2E-01	1.7E-01	1.3E-01	1.5E-01	1.3E-01	1.2E-01	1.4E-01	1.1E-01	8.8E-02	1.1E-01	7.6E-02
penA	1.3E-01	1.1E-01		1.3E-01	1.3E-01	1.4E-01	8.7E-02	5.7E-02	1.0E-01	6.2E-02	7.7E-02	1.3E-01	1.7E-01	1.9E-01	1.0E-01	5.2E-02	6.8E-02	6.4E-02
blaSFO	8.6E-02	5.7E-02	7.2E-02	7.4E-02	1.0E-01	9.5E-02	9.6E-02		1.1E-01	7.5E-02	9.0E-02	1.1E-01		1.6E-01	1.1E-01	7.7E-02	5.7E-02	6.8E-02
blaOXA48	3.2E-02	2.1E-02	1.4E-02	1.9E-02	2.6E-02	2.2E-02	2.6E-02	1.8E-02	1.5E-02	1.6E-02	1.2E-02	1.1E-02	1.1E-02	9.8E-03	1.4E-02	1.5E-02	2.0E-02	1.7E-02
blaNDM	7.6E-03	3.3E-03	6.5E-03	7.1E-03	4.9E-03	1.6E-02	5.0E-03	7.2E-03	8.0E-03	3.3E-03	1.1E-02	1.2E-02	5.6E-03	5.8E-03	8.5E-03	6.8E-03	1.3E-02	7.6E-03
blaSHV11	6.8E-03	4.6E-03	3.5E-03	6.6E-03	4.7E-03	2.3E-02	9.5E-03	5.2E-03	1.1E-02	6.3E-03	6.1E-03	5.2E-03	1.0E-02	8.4E-03	5.4E-03	5.3E-03	6.6E-03	6.2E-03
blaCMY_2	3.4E-03	5.0E-03	3.5E-03	3.4E-03	5.2E-03	4.1E-03	2.1E-03	3.2E-03	2.4E-03	3.4E-03	5.0E-03	2.5E-03	4.6E-03	4.3E-03	2.6E-03	3.0E-03	2.2E-03	3.3E-03
ampC/blaDHA	1.9E-03	1.8E-03	1.2E-03	1.9E-03	2.7E-03	3.4E-03	8.0E-04	1.7E-03	1.8E-03	1.8E-03	2.1E-03	1.4E-03	2.7E-03	3.8E-03	1.3E-03	1.3E-03	1.0E-03	1.4E-03
blaKPC	1.1E-03	1.8E-03	1.8E-04	1.3E-03	6.0E-04	3.2E-04	9.4E-04	1.8E-03	2.1E-04	5.4E-04	6.3E-04	2.9E-03	9.0E-04	7.5E-04	4.4E-04	1.1E-03	8.7E-04	1.3E-03
blaVIM	1.1E-03	1.0E-03	1.4E-03	1.5E-03	8.6E-04	1.1E-03	1.2E-03	8.7E-04	1.0E-03	5.9E-04	9.1E-04	4.6E-04	6.0E-04	6.0E-04	4.1E-04	5.6E-04	1.8E-03	7.5E-04
blaCTX-M_5	1.4E-03	1.0E-03	8.5E-04	9.2E-04	1.4E-03	1.1E-03	8.3E-04	9.2E-04	1.2E-03	4.5E-04	1.3E-03	1.0E-03	4.0E-04	4.3E-04	4.2E-04	9.1E-04	1.4E-03	6.8E-04
blaOXA51	6.3E-04	7.8E-04	5.9E-04	6.6E-04	5.8E-04	4.6E-04	7.4E-04	5.4E-04	4.3E-04	5.1E-04	5.0E-04	3.3E-04	2.0E-04	1.9E-04	2.8E-04	2.3E-04	1.2E-03	3.1E-04
blaTEM	2.2E-03	3.7E-04	4.3E-04	6.6E-04	5.3E-04	3.0E-04	3.0E-04	3.3E-04		2.4E-04	2.9E-04	1.1E-04	2.1E-04	3.9E-04	1.3E-04	1.2E-03	5.9E-04	6.4E-04
cfxA	4.1E-04	5.4E-04	6.0E-04	5.1E-04	5.1E-04	3.3E-04	3.3E-04	3.0E-04	3.5E-04	1.7E-04	2.8E-04	8.5E-05	4.9E-05	3.5E-05	5.0E-05	9.6E-04	5.5E-04	8.1E-04
blaGES	1.8E-04	7.3E-05		2.6E-04			1.7E-04			1.3E-04		9.3E-05	3.0E-04	5.9E-05		1.5E-03	6.2E-04	2.1E-03
ampC_cefa	2.1E-04	4.7E-04	8.5E-05	1.8E-04	9.6E-05	9.5E-05	9.5E-05	1.8E-04	2.7E-04	1.1E-04	1.2E-04	8.8E-05	2.4E-04	1.6E-04	9.5E-05	1.6E-04	2.7E-04	1.5E-04
blaCTX-M_8																7.4E-04	1.6E-04	4.0E-04
ereA		1.6E-01		1.3E-01		1.2E-01		1.2E-01	8.0E-02	1.3E-01	1.2E-01	1.2E-01	1.5E-01	1.8E-01	1.1E-01	1.1E-01	1.4E-01	1.1E-01
msrE	7.6E-03	1.0E-03	6.8E-03	2.2E-02	7.3E-03	4.1E-02	2.2E-02	1.3E-03	5.9E-03	6.4E-03	3.3E-03	3.2E-03	8.8E-03	1.8E-03	2.2E-02	1.3E-01	6.1E-03	1.3E-01

'Imm' represents reactors receiving influent solids during Phase 2, 'SC' represents the sterile control, 'Start' indicates samples taken at the start of Phase 1 before immigration was introduced. Samples analysed were pooled from the triplicate reactors. Each pooled sample was analysed in triplicate by qPCR Array.

Table S6.4 Continued

Gene	14d Imm	14d SC	14d Start	10.5 Imm	10.5 SC	10.5 Start	7d Imm	7d SC	7d Start	3.5d Imm	3.5d SC	3.5d Start	1.8d Imm	1.8d SC	1.8d Start	Inf 1	inf 2	Inf 3
ermB_1	7.0E-03	5.0E-03	4.6E-05	9.8E-03	5.6E-03		4.1E-02	8.1E-03		3.3E-02	4.6E-03	5.4E-04	1.5E-02	7.3E-03	1.5E-03	8.5E-02	2.3E-03	7.2E-02
ermB_2	8.1E-03	5.3E-03	5.4E-04	1.0E-02	6.2E-03	1.1E-03	4.1E-02	8.4E-03	4.2E-04	3.3E-02	5.4E-03	6.6E-04	1.5E-02	8.3E-03	1.4E-03	8.1E-02	2.1E-03	6.7E-02
mefA_1	4.8E-03	5.2E-03	5.7E-04	7.4E-03	5.3E-03	7.9E-04	2.4E-02	7.7E-03	4.0E-04	3.0E-02	5.9E-03	1.5E-03	1.6E-02	1.2E-02	2.4E-03	3.1E-02	2.1E-03	3.7E-02
lnuB	1.2E-03	1.1E-03		2.3E-03	1.7E-03		1.1E-02	2.6E-03		1.3E-02	2.5E-03	8.2E-04	5.6E-03	9.3E-03	1.6E-03	1.7E-02	5.6E-04	2.1E-02
ermF	1.5E-03	2.0E-03	4.7E-04	1.7E-03	7.6E-04	1.8E-03	5.8E-04	4.3E-04	2.6E-04	1.2E-03	6.2E-04	1.6E-03	6.8E-04	1.3E-04	2.8E-04	5.0E-03	3.2E-03	3.6E-03
ermA	8.7E-04	7.4E-04	2.4E-04	4.9E-04	6.2E-04	4.4E-04	3.6E-04	3.2E-04	4.3E-04	3.3E-04	4.4E-04	1.9E-04	3.6E-04	3.8E-04	2.2E-04	4.8E-04	2.7E-03	5.5E-04
lnuC	2.0E-04			2.0E-04	7.1E-05		4.3E-05	9.6E-05	6.2E-05	2.0E-04	5.9E-05		6.0E-05	2.1E-04		3.5E-03	7.6E-04	2.0E-03
merA	2.3E-02	8.2E-03	3.6E-02	2.7E-02	2.5E-02	4.4E-02	5.8E-02	4.0E-02	9.1E-02	3.9E-02	2.0E-02	8.2E-02	6.7E-02	7.6E-02	1.6E-01	8.9E-02	1.5E-02	6.9E-02
mcr1	3.8E-02	3.2E-02	2.1E-02	2.9E-02	2.7E-02	2.3E-02	2.4E-02	2.1E-02	2.2E-02	1.6E-02	1.7E-02	2.0E-02	1.5E-02	1.5E-02	1.9E-02	1.1E-02	1.9E-02	1.1E-02
arr3	4.9E-03	3.8E-03	3.1E-03	4.3E-03	3.6E-03	2.7E-03	4.8E-03	6.2E-03	4.1E-03	3.1E-03	4.0E-03	6.3E-03	6.5E-03	7.8E-03	5.5E-03	6.6E-03	4.7E-03	7.2E-03
uidA	4.6E-03	4.9E-03	2.7E-03	4.1E-03	4.5E-03	3.0E-03	4.9E-03	5.1E-03	4.7E-03	4.6E-03	4.1E-03	7.4E-03	4.2E-03	3.7E-03	5.0E-03	4.4E-03	6.9E-03	4.1E-03
qacEâˆ†1_1	1.5E-03	1.1E-03	1.7E-03	1.9E-03	4.3E-03	5.1E-03	5.2E-03	3.2E-03	2.1E-03	3.9E-03	4.3E-03	3.6E-03	5.4E-03	7.7E-03	5.6E-03	4.9E-03	8.3E-03	5.5E-03
crAss56	4.1E-04	5.5E-04		4.4E-04	1.4E-04		2.2E-04	3.7E-04		2.2E-04	9.3E-05		2.6E-04	9.8E-05		2.1E-03	3.3E-03	2.9E-03
sat4																1.8E-04	8.6E-05	1.5E-04
aph3-ib	2.7E-01	1.6E-01	1.5E-01	2.0E-01		9.3E-02	1.3E-01	2.0E-01	7.8E-02	1.3E-01	1.4E-01	9.6E-02	6.1E-02	5.2E-02	6.9E-02	6.1E-02	1.3E-01	5.8E-02
aac(3)-ib	2.9E-02	2.3E-02	2.9E-02	2.9E-02	3.0E-02	2.1E-02	2.5E-02	1.8E-02	2.0E-02	2.1E-02	2.0E-02	1.9E-02	1.0E-02	8.3E-03	1.6E-02	9.5E-03	1.9E-02	8.8E-03
aac(6')-Ib_1	1.1E-02	1.2E-02	8.8E-03	1.1E-02	9.2E-03	8.7E-03	5.2E-03	4.8E-03	4.8E-03	4.1E-03	3.6E-03	7.6E-03	4.6E-03	3.4E-03	4.6E-03	5.8E-03	9.6E-03	7.1E-03
strB	2.3E-02	1.6E-02	3.0E-03	1.4E-02	7.9E-03	1.7E-03	4.6E-03	2.3E-03	2.8E-03	5.4E-03	4.0E-03	1.9E-03	3.6E-03	3.5E-03	4.4E-03	8.1E-03	3.2E-03	7.5E-03
aadA_1	2.8E-03	3.6E-04	3.8E-04	6.6E-03	5.2E-04	1.3E-03	1.0E-03	4.3E-04	1.6E-03	3.6E-03	2.5E-03	4.0E-03	3.7E-03	1.0E-02	3.2E-03	6.3E-03	5.2E-03	7.4E-03
aphA3_1	1.5E-04	2.1E-04	7.8E-05	2.0E-04	8.7E-05	2.0E-04	7.2E-05	7.9E-05	5.3E-04	1.1E-04	8.5E-05	1.0E-04	5.6E-05	5.2E-05	5.7E-05	3.1E-04	2.5E-04	3.1E-04
aadE	1.2E-04			1.2E-04			6.8E-05			1.8E-04	1.9E-04	1.1E-04	2.4E-04	4.3E-04	1.3E-04	3.3E-04	5.1E-05	8.0E-04
aph(3")-ia	1.8E-04		6.7E-05	7.9E-05	4.3E-04	6.2E-05	1.0E-04	1.1E-04		3.2E-04	1.1E-04	7.8E-05	3.8E-05	5.0E-05	5.8E-05	3.2E-04	2.0E-04	4.1E-04
aacC2				5.6E-05			1.0E-04	7.0E-05	5.8E-05	4.1E-05		7.5E-05	9.2E-04	4.2E-04	8.2E-05	2.4E-04		1.6E-04
aac6-aph2	5.6E-05	1.0E-04	5.0E-05	9.1E-05		1.1E-04	5.4E-05	1.1E-04	8.7E-05	4.9E-05	1.6E-04	8.7E-05	5.6E-05	1.0E-04	7.9E-05	3.6E-04	1.3E-04	2.4E-04
aphA1/7					2.0E-04					1.3E-04			2.2E-05			9.8E-05		1.6E-04
aacA/aphD																1.3E-04		4.6E-05
floR_1	4.2E-02	2.7E-02	2.9E-02	2.8E-02	3.3E-02	1.9E-02	3.1E-02	2.0E-02	2.0E-02	2.6E-02	2.1E-02	1.9E-02	8.3E-03	7.0E-03	1.2E-02	7.1E-03	1.1E-02	7.7E-03
cmlA_2	2.7E-03	6.4E-03	3.1E-03	2.8E-03	5.8E-03	3.2E-03	2.6E-03	3.0E-03	2.0E-03	6.6E-03	5.0E-03	1.9E-03	1.9E-03	1.6E-03	9.8E-04	2.4E-03	2.8E-03	3.2E-03
catB3	3.5E-04	1.3E-04	1.3E-04	2.1E-04	2.6E-04	2.7E-04	1.9E-04	2.2E-04		2.0E-04	1.6E-04	3.4E-04	3.2E-04	3.4E-04	2.1E-04	5.4E-04	2.5E-04	5.5E-04
sul1_2	2.0E-02	1.2E-02	2.3E-03	5.7E-03	4.7E-03	4.5E-03	8.6E-03	1.4E-02	2.6E-03	7.7E-03	6.6E-03	1.0E-02	1.1E-02	2.6E-02	2.0E-02	6.9E-03	1.5E-02	8.0E-03
sul2_2	5.1E-03	2.6E-03	4.0E-03	8.2E-03	3.4E-03	8.2E-03	4.6E-03	2.0E-03	3.4E-03	2.4E-03	4.0E-03	1.0E-02	2.2E-03	2.6E-03	2.2E-03	4.3E-03	3.9E-03	3.0E-03
tet39	5.1E-03	1.1E-03	6.0E-03	1.7E-02	6.0E-03	9.2E-02	1.5E-02	3.4E-04	4.9E-03	2.7E-03	8.0E-05	9.9E-04	5.3E-03	2.0E-04	1.0E-02	7.8E-02	2.6E-03	8.1E-02
tetW	2.2E-04	4.3E-04	1.4E-04	3.8E-04	3.7E-04	1.9E-04	3.0E-04	2.6E-04	1.9E-04	2.5E-04	1.5E-04	3.1E-04	2.3E-04	9.9E-05	8.5E-05	3.2E-03	1.1E-03	3.1E-03
tetE	5.5E-04	4.8E-04	4.6E-04	4.3E-04	6.7E-04	2.3E-04	5.1E-04	5.2E-04	4.3E-04	3.3E-04	4.3E-04	5.5E-04	1.0E-03	9.3E-04	5.4E-04	1.2E-03	9.7E-04	8.3E-04
sugE	4.2E-02	3.0E-02	3.3E-02	2.6E-02	2.6E-02	3.1E-02	1.6E-02	1.7E-02		1.7E-02	1.8E-02	1.4E-02	1.6E-02	1.7E-02	1.6E-02	1.1E-02	1.9E-02	1.1E-02
qacF/H	9.2E-03	5.1E-03	8.6E-03	9.0E-03	6.1E-03	1.0E-02	1.1E-02	7.5E-03	7.0E-03	1.8E-02	2.3E-02	9.2E-03	1.0E-02	9.9E-03	1.4E-02	3.5E-02	7.5E-03	3.3E-02
cefa_qacelta	1.6E-03	1.7E-03	3.4E-03	2.6E-03	8.7E-03	4.1E-03	6.7E-03	3.9E-03	2.4E-03	4.9E-03	5.5E-03	4.9E-03	6.9E-03	1.0E-02	6.5E-03	5.3E-03	1.2E-02	7.3E-03
tolC_2	5.7E-03	6.1E-03	4.2E-03	5.0E-03	4.4E-03	2.4E-03	4.4E-03	4.1E-03	2.6E-03	4.2E-03	7.3E-03	5.4E-03	5.2E-03	7.3E-03	1.2E-02	3.5E-03	4.6E-03	4.0E-03
pcoA	2.9E-03	3.4E-03	2.6E-03	2.8E-03	1.7E-03	3.7E-03	1.7E-03	1.3E-03	1.4E-03	1.7E-03	2.2E-03	1.9E-03	6.6E-03	6.5E-03	1.1E-02	5.4E-03	1.6E-03	3.5E-03
mexE	1.1E-03	9.2E-04	9.8E-04	1.1E-03	9.6E-04	7.5E-04	1.4E-03	1.2E-03	1.3E-03	4.2E-03	7.2E-03	1.4E-03	6.1E-03	8.5E-03	2.0E-03	1.8E-03	1.9E-03	1.5E-03
tcrB	6.0E-04	1.3E-04		6.4E-04	2.0E-04		7.2E-04	1.8E-04	8.4E-05	1.1E-03	1.6E-04	4.7E-04	1.2E-03	5.4E-04	1.0E-04	3.6E-03	3.2E-04	5.0E-03
acrB_1	7.8E-04	5.6E-04	4.9E-04	6.0E-04	6.4E-04	7.6E-04	3.6E-04	4.0E-04	5.5E-04	4.5E-04	1.2E-03	1.5E-03	1.7E-03	1.8E-03	1.1E-03	5.7E-04	7.5E-04	6.0E-04
dfrA1_1	2.4E-04	3.7E-04	4.6E-05	1.1E-04	1.0E-04	3.0E-04	1.0E-04	1.0E-04	1.2E-04	1.3E-04	3.4E-04	1.8E-03	1.6E-03	7.3E-03	3.8E-04	3.4E-04	1.8E-04	3.7E-04

Table S6.4 Continued

Gene	14d Imm	14d SC	14d Start	10.5 Imm	10.5 SC	10.5 Start	7d Imm	7d SC	7d Start	3.5d Imm	3.5d SC	3.5d Start	1.8d Imm	1.8d SC	1.8d Start	Inf 1	inf 2	Inf 3
dfrA1_1	2.4E-04	3.7E-04	4.6E-05	1.1E-04	1.0E-04	3.0E-04	1.0E-04	1.0E-04	1.2E-04	1.3E-04	3.4E-04	1.8E-03	1.6E-03	7.3E-03	3.8E-04	3.4E-04	1.8E-04	3.7E-04
16S rRNA	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00	1.0E+00
Enterococci	1.2E-02	2.0E-02	7.5E-05	2.1E-02	1.6E-02		9.4E-02	2.2E-02		9.3E-02	1.9E-02	2.6E-03	5.8E-02	5.7E-02	6.9E-03	1.2E-01	5.9E-03	1.6E-01
A. baumannii	1.3E-03	2.4E-03	1.5E-03	2.5E-03	1.5E-03	3.9E-03	3.5E-03	1.8E-03	2.1E-03	2.7E-03	1.1E-03	1.4E-03	2.5E-03	1.0E-03	2.2E-03	5.6E-03	2.2E-03	9.4E-03
P. aeruginosa	1.1E-03	1.2E-03	1.1E-03	6.2E-04	6.3E-04	1.2E-03	6.5E-04	6.0E-04	5.4E-04	4.5E-04	4.7E-04	5.1E-04	2.9E-04	3.2E-04	3.3E-04	4.5E-04	6.2E-04	4.1E-04
K. pneumoniae	1.8E-03	8.1E-04	4.6E-05	6.7E-04	4.9E-04	9.1E-05	4.4E-04	1.7E-04	6.2E-05	6.7E-05	6.6E-05		6.2E-05	4.6E-05	5.9E-05	1.5E-04	1.4E-04	1.3E-04
Staphylococci																2.8E-05		
Campylobacter																		

Table S6.5: Values used for the calculation of $f_{OHO,ML}$ when $\mu_{net} = 0\ d^{-1}$

θ_X (Days)	θ (Days)	$X_{Tot,inf}$[a] (mg-VSS/L)	$X_{Tot,ML}$[b] (mg-VSS/L)	$Y_{Tot,inf}$[c] (mg-DNA/mg-VSS)	$Y_{Tot,ML}$ (mg-DNA/mg-VSS)	$f_{OHO,capture}$
1.8 + 0.2	1.8	120	247.7 + 27.7	4.8 + 0.6	6.3 $\pm$ 1.0	1
3.5 + 0.3	1.8	120	444.8 + 198.2	4.8 + 0.6	4.9 $\pm$ 0.1	1
7.0 + 0.5	1.8	120	817.8 + 106.1	4.8 + 0.6	3.6 $\pm$ 1.2	1
10.5 + 0.8	1.8	120	1121.1 + 213.7	4.8 + 0.6	2.4 $\pm$ 0.2	1
14 + 1.1	1.8	120	1066.0 + 358.2	4.8 + 0.6	1.2 $\pm$ 0.2	1

Where $\pm$ is the standard deviation between reactors

[a] Influent VSS was controlled at 120 mg-VSS/L so no standard deviation is reported

[b] Average of reactor VSS during Phase 2, after steady state was reached

[c] As quantified using Nanodrop one

6.7 References

1. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ. 2013;447:345–60.
2. de Oliveira M, Farias Frihling BE, Velasques J, Corrêa Magalhães Filho FJ, Sabioni Cavalheri P, Migliolo L. Pharmaceuticals residues and xenobiotics contaminants: Occurrence, analytical techniques and sustainable alternatives for wastewater treatment. Sci Total Environ. 2020;705:135568.
3. Gibson C, Kraemer SA, Klimova N, Vanderweyen L, Klai N, Guo B, et al. Multiplexed amplicon sequencing reveals high allelic diversity of antibiotic resistance genes in Québec sewers. In preparation. 2023.
4. Baquero F, Martínez J-L, Cantó R. Antibiotics and antibiotic resistance in water environments. Curr Opin Biotechnol. 2008;19:260–5.
5. Petrovich M, Chu B, Wright D, Griffin J, Elfeki M, Murphy BT, et al. Antibiotic resistance genes show enhanced mobilization through suspended growth and biofilm-based wastewater treatment processes. FEMS Microbiol Ecol. 2018;94:fiy041.
6. Qin K, Wei L, Li J, Lai B, Zhu F, Yu H, et al. A review of ARGs in WWTPs: Sources, stressors and elimination. Chinese Chem Lett. 2020;31(10):2603–13.
7. Guo B, Liu C, Gibson C, Frigon D. Wastewater microbial community structure and functional traits change over short timescales. Sci Total Environ. 2019;662:779–85.
8. Zhu Y, Wang Y, Zhou S, Jiang X, Ma X, Liu C. Robust performance of a membrane bioreactor for removing antibiotic resistance genes exposed to antibiotics: Role of membrane foulants. Water Res. 2018;130:139–50.
9. Xia S, Jia R, Feng F, Xie K, Li H, Jing D, et al. Effect of solids retention time on antibiotics removal performance and microbial communities in an A/O-MBR process. Bioresour Technol. 2012;106:36–43.
10. Staley C, Kaiser T, Vaughn BP, Graiziger C, Hamilton MJ, Kabage AJ, et al. Durable long-term bacterial engraftment following encapsulated fecal microbiota transplantation to treat Clostridium difficile infection. MBio. 2019;10(4):e01586-19.
11. Zhang M, Chen S, Yu X, Vikesland P, Pruden A. Degradation of extracellular genomic, plasmid DNA and specific antibiotic resistance genes by chlorination. Front Environ Sci Eng. 2019;13(3):38.

12. Liu SS, Qu HM, Yang D, Hu H, Liu WL, Qiu ZG, et al. Chlorine disinfection increases both intracellular and extracellular antibiotic resistance genes in a full-scale wastewater treatment plant. Water Res. 2018;136:131–6.

13. Pang Y, Huang J, Xi J, Hu H, Zhu Y. Effect of ultraviolet irradiation and chlorination on ampicillin-resistant Escherichia coli and its ampicillin resistance gene. Front Environ Sci Eng. 2016;10(3):522–30.

14. Gibson C, Kraemer SA, Klimova N, Guo B, Dominic F. Antimicrobial resistance gene variant sequencing is necessary to reveal the complex dynamics of immigration from sewers to activated sludge. Submitted. 2023.

15. Gibson C, Jauffur S, Guo B, Frigon D. Activated sludge microbial community assembly: The role of influent microbial community immigration. bioRxiv. 2023;525574.

16. Frigon D, Wells G. Microbial immigration in wastewater treatment systems: analytical considerations and process implications. Curr Opin Biotechnol. 2019;57:151–9.

17. Vuono DC, Munakata-Marr J, Spear JR, Drewes JE. Disturbance opens recruitment sites for bacterial colonization in activated sludge. Environ Microbiol. 2016;18(1):87–99.

18. Kaewpipat K, Grady CPL. Microbial population dynamics in laboratory-scale activated sludge reactors. Water Sci Technol. 2002;46:19–27.

19. American Public Health Association. Standard methods for the examination of water and wastewater. 22nd ed. Rice EW, Baird RB, Eaton AD, Clesceri LS, editors. Washington DC; 2012.

20. Isazadeh S, Frigon D. Biosolids minimization by partial ozonation of return activated sludge: Model development and bacterial population dynamics. McGill University, Canada; 2014.

21. Apprill A, Mcnally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75:129–37.

22. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2015;18(5):1403–14.

23. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

24. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. Vegan: Community Ecology Package. 2022.

25. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–7.

26. McIlroy SJ, Kirkegaard RH, McIlroy B, Nierychlo M, Munk Kristensen J, Karst SM, et al. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017;1–9.

27. Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686.

28. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat Microbiol. 2019;4:1183–95.

29. Lee S-H, Kang H-J, Park H-D. Influence of influent wastewater communities on temporal variation of activated sludge communities. Water Res. 2015;73:132–44.

30. Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 2016;10:11–20.

31. Harris K, Parsons TL, Ijaz UZ, Lahti L, Holmes I, Quince C. Linking statistical and ecological theory : Hubbell's unified neutral theory of biodiversity as a hierarchical dirichlet process. Proc IEEE. 2017;105(3):516–29.

32. Dai D, Brown C, Bürgmann H, Larsson DGJ, Nambi I, Zhang T, et al. Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. Microbiome. 2022;10:20.

33. Meerburg FA, Vlaeminck SE, Roume H, Seuntjens D, Pieper DH, Jauregui R, et al. High-rate activated sludge communities have a distinctly different structure compared to low-rate sludge communities, and are less sensitive towards environmental and operational variables. Water Res. 2016;100:137–45.

34. Dueholm MKD, Nierychlo M, Andersen KS, Rudkjøbing V, Knutsson S, Arriaga S, et al. MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. Nat Commun. 2022;13:1908.

35. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 2005;3:722–32.

36. Tong J, Liu J, Zheng X, Zhang J, Ni X, Chen M, et al. Fate of antibiotic resistance bacteria and genes during enhanced anaerobic digestion of sewage sludge by microwave pretreatment. Bioresour Technol. 2016;217:37–43.

37. Wu D, Huang X-H, Sun J-Z, Graham DW, Xie B. Antibiotic resistance genes and associated microbial community conditions in aging landfill systems. Environ Sci Technol. 2017;51(21):12859–67.

38. Zhou Z-C, Zheng J, Wei Y-Y, Chen T, Dahlgren RA, Shang X, et al. Antibiotic resistance genes in an urban river as impacted by bacterial community and physicochemical parameters. Environ Sci Pollut Res. 2017;24:23753–62.

39. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nat Lett. 2014;509:612–6.

CHAPTER 7

General Discussion and Future Considerations

Without a firm understanding of factors impacting the persistence of antimicrobial resistance (AMR) in the environment, we cannot begin to establish appropriate control measures to tackle this growing problem. In this thesis, the relationship between microbial community assembly and AMR in activated sludge wastewater treatment plants (AS-WWTPs) was explored in depth by developing new approaches to study immigration and monitor antimicrobial resistance genes (ARGs) in the environment

7.1 Immigration in Other Environments

Microbial immigration is omnipresent whether it be within AS-WWTPs, the gut or the wider environment. Within the wastewater treatment process alone, immigration occurs at numerous interfaces (Figure 1). This thesis focuses specifically on the impact of immigration between the influent and activated sludge. However, the concepts explored are likely valid in other environments. Microbial community composition is a key determinant of antimicrobial resistance. Thus, control of the microbial community composition may be critical in managing the spread of AMR in numerous settings. To fully understand how complex microbial communities are assembled, it is critical that future studies consider the role of immigration, or community coalescence in other environments.

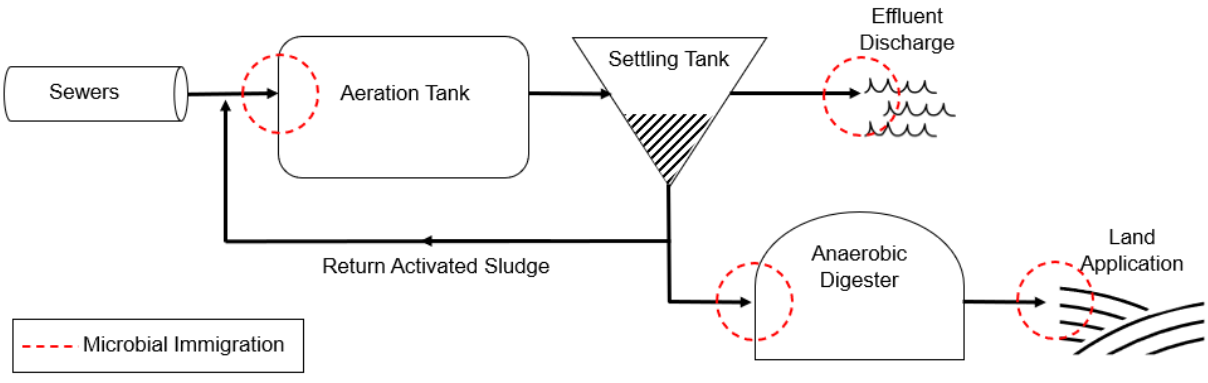


Figure 7.1- Immigration Interfaces in the Wastewater Treatment Process. Immigration interfaces are defined as points were a unique, immigrating microbial community enters an environment where an established resident community is present.

7.2 The Use of Reactors to Study Immigration

For many years, the impact of immigration on the activated sludge community has been debated, with both modelling and full-scale studies producing inconsistent results. Experimental protocols deployed in full-scale AS-WWTP suffer from various limitations related to the lack of reproducibility and controls, and the multitude of variables impacting the microbial community at any one time. In this thesis, by simplifying the activated sludge process and employing a controlled experimental protocol, specific questions could be precisely evaluated, including;

1. Does the influent microbial community alone impact the activated sludge community?
2. Are immigrants growing in the activated sludge?
3. Do immigrants persist in the absence of influent immigration?
4. Is immigration dependent upon the influent microbial community composition?
5. Is immigration dependent upon the activated sludge community composition?
6. Are immigration dynamics impacted by WWTP operation?

Through the operation of four sets of reactors (Chapter 4 and 6), with four different influent communities, immigration was shown to have a significant and reproducible effect on the activated sludge microbial community. This effect was replicated using different source activated sludge microbial communities validating the method for future studies using different inocula. By using a system where all conditions could be carefully controlled, we were able to define specific resident and immigrant populations in the activated sludge, and precisely characterise and quantify the impacts of immigration. In the future, the laboratory-scale reactor protocol developed in this thesis can be modified to reproducibly test the impact of other parameters of interest and improve our mechanistic description of the various phenomena described.

The operation of reactors both with and without immigration enabled limitations in the typical assumptions used to study immigration in the literature to be identified. These studies often identify immigrants based upon shared ASVs (or OTUs or genera) between environments. Under the reactor conditions explored herein, the microbial community was found to be inherently more similar to that of the influent solids at lower SRTs (Chapter 6). Therefore, the assumption of immigration based upon shared genera would likely over-estimate the contribution of immigration-dependent taxa at low SRTs. By extension, this assumption may also lead to misestimations of the impacts of immigration for certain taxa under different operational conditions. To enable immigration dependent populations to be identified, we first need to distinguish between resident taxa of the activated sludge which are already shared among the environments, and those which are newly introduced by immigration. In full-scale WWTPs, it is unlikely that this clarification can be made due to the lack of no-immigration controls. Whether using reactors or studying full-scale AS-WWTPs, future studies aiming to explore the role of influent immigration should carefully consider how each population can be defined.

7.3 Immigration in the Activated Sludge

Influent immigration was shown to significantly impact the microbial community of the activated sludge. Given the abundance and diversity of the immigrant population, they likely play a critical role in substrate consumption in the activated sludge and impact the outcomes of wastewater treatment (1). Among the immigrants there were specialised genera such as the genus *Candidatus Accumulibacter* and *Tetrasphaera* which have been shown to play important roles in enhanced biological phosphorous processes (2) and others such as genus *Gordonia* which have been implicated in bulking and foaming events (3). Taken together, these results suggest that influent

immigration is likely critical not only to the performance of certain processes, but also in operational problems.

Immigrating genera typically displayed low or negative net growth rates within the activated sludge (Chapter 4 and 6), which suggested that washout would occur over time without continuous influent immigration. This should be carefully considered in future WWTP design. For example, in the design of enhanced biological phosphorous removal processes it may be necessary to ensure a continuous stream of specialised polyphosphate accumulating organisms in the influent wastewater. The influent microbial community has been shown to vary over time (4,5), thus on occasions where the concentration of specialised taxa is reduced, bioaugmentation of the influent wastewater could be utilised to maintain function within the WWTP. Similarly, to prevent entry of unfavourable genera, treatment plant operators could consider the introduction of an advanced primary treatment to reduce the degree of immigration.

7.4 Complex Dynamics of ARG Immigration

In Chapter 3, the antimicrobial resistance genes detected in influent wastewater were found to display high sequence diversity which was likely influenced by the microbial community composition in the samples. The diversity among the detected ARGs provided the opportunity to study immigration dynamics between the influent and activated sludge in greater detail than previous studies. Immigration caused significant increases in the abundance of ARGs in the activated sludge. Using multiplexed amplicon sequencing, the immigration dynamics of ARGs between the influent and activated sludge were shown to be complex, with counter selection dynamics being most commonly observed. The increase in the abundance of ARGs with immigration was most frequently associated with the changes in the microbial community composition. Taken together, this suggests that control over the microbial community composition

in the activated sludge rather than the influx of ARGs themselves, may be a better approach to manage the spread of AMR. In addition, although monitoring of ARGs in influent wastewater may be a useful tool to assess the prevalence of AMR in a population within a catchment area, given the complexity of ARG dynamics at the immigration interface and selection observed, it cannot be used to accurately assess the fate and persistence of a given ARG in the wider environment.

7.5 WWTP Design to Minimise AMR Dissemination

Control over microbial community assembly in the activated sludge is likely critical in managing the persistence of AMR in AS-WWTP. To investigate how AS-WWTP design could be optimised to control the microbial community composition and AMR, the reactor protocol developed in Chapter 4 was modified to evaluate the impact of operational solid retention time (SRT). SRT impacted both the persistence of ARGs in the core resident community, and the dynamics at the immigration interface (Chapter 6). The controlled conditions within the reactor systems enabled the microbial community to be defined into specific populations, which each displaying distinct dynamics. This allowed a unique perspective into the relationship between microbial community assembly and AMR in the activated sludge. ARGs in the activated sludge could be linked to populations which were present within the reactors regardless of immigration (the core resident population), those which were immigration dependent and successful under all conditions (universal immigrants), and those which were dependent upon both SRT and immigration (SRT immigrants). Given the different behaviours of ARGs, in the future a multi-prong approach may be required to manage AMR in activated sludge or specific ARGs may need to be prioritised based upon their risk. By utilising novel methods to investigate microbial community assembly alongside the dynamics of ARG sequence variants, the role of specific populations in the persistence of AMR in WWTPs was identified for the first time. Future studies should consider applying the analysis

framework developed herein to investigate the impact of other operational parameters to inform WWTP design and minimise the persistence of AMR.

7.6 Multiplexed Amplicon Sequencing to Manage AMR in the Environment

The United Nations Environmental Protection Agency have called for increased surveillance of AMR in the environment (6). The value of tracking ARG sequence variants is becoming increasingly recognised in the literature (7). In this thesis, we present a low-cost and sensitive method for the detection of ARG sequence variants in environmental samples and demonstrate its potential in both ARG source tracking and risk assessment. In Chapter 3 and 5, unique sequence variants were identified within sewers and activated sludge. Due to the limited sample types analysed thus far, markers have not yet been identified to enable source tracking in the wider environment. However, in the future, with further sampling of known hotspots and sources of AMR, ARG sequence variant markers in each environment could be established which would allow multiplexed amplicon sequencing to be utilized in routine monitoring and source tracking of AMR in the environment.

In Chapter 3, it was discussed that the risk of a given antimicrobial resistance gene (ARG) can be associated with its mobility and occurrence within pathogens, which relates to the ability of the host to cause infection. Based upon this, publicly available databases were used to assess the occurrence of ARG sequence variants in pathogens and on plasmids. Risk assessment using multiplexed amplicon sequencing could be further improved by designing primers to target specific resistance phenotypes. For example, among the *bla*OXA resistance gene, mutations in specific regions of the ARG have been demonstrated to result in increased phenotypic resistance (8). Based upon these known mutations, the targets for multiplexed amplicon sequencing should be optimised to enable higher risk resistance phenotypes to be distinguished from those with

narrow spectrum activity. ARG are widespread in numerous environments, however by monitoring high risk sequence variants, the management of AMR can be prioritised.

7.7 References

1. Guo B, Liu C, Gibson C, Klai N, Lin X, Frigon D. Wastewater influent microbial immigration and contribution to resource consumption in activated sludge using taxon-specific mass-flow immigration model. bioRxiv. 2022;504022.
2. Albertsen M, Hansen LBS, Saunders AM, Nielsen PH, Nielsen KL. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. ISME J. 2012;6:1094–106.
3. Jiang X-T, Guo F, Zhang T. Population dynamics of bulking and foaming bacteria in a full-scale wastewater treatment plant over five years. Sci Rep. 2016;6:24180.
4. Guo B, Liu C, Gibson C, Frigon D. Wastewater microbial community structure and functional traits change over short timescales. Sci Total Environ. 2019;662:779–85.
5. Peces M, Dottorini G, Nierychlo M, Andersen KS, Dueholm MKD, Nielsen PH. Microbial communities across activated sludge plants show recurring species-level seasonal patterns. ISME Commun. 2022;2:18.
6. United Nations Environment Programme (UNEP). Environmental Dimensions of Antimicrobial Resistance: Summary for Policymakers. 2022. Available at: https://wedocs.unep.org/bitstream/handle/20.500.11822/38373/antimicrobial_R.pdf. [Accessed: 27.05.2022]
7. Smith SD, Choi J, Ricker N, Yang F, Hinsa-Leasure S, Soupir ML, et al. Diversity of Antibiotic Resistance Genes and Transfer Elements-Quantitative Monitoring (DARTE-QM): a method for detection of antimicrobial resistance in environmental samples. Commun Biol. 2022;5:216.
8. Danel F, Hall LM, Duke B, Gur D, Livermore DM. OXA-17, a further extended-spectrum variant of OXA-10 beta-lactamase, isolated from Pseudomonas aeruginosa. Antimicrob Agents Chemother. 1999;43(6):1362–6.

APPENDIX 1

Optimisation of Multiplex Amplicon Tool

A1.1 Introduction

To allow antimicrobial resistance gene amplicon sequence variants (ARG-ASVs) to be accurately detected in environmental samples, an appropriate multiplexed amplicon sequencing tool and bioinformatic pipeline had to be developed. This appendix details the optimisation steps used to assess the performance of the tool, and validate its use for the analysis of environmental samples.

A1.2 Methods

A1.2.1 Amplicon Sequencing Panels

In total four multiplexed amplicon sequencing panels were produced, each with different combinations of antimicrobial resistance genes (ARGs) and mobile genetic elements (MGEs). Panels A-C were developed in collaboration with Swift Biosciences and utilised the Accel-Amplicon™ NGS Panel, whilst Panel D used the commercially available Qiagen multiplex PCR kit and was developed in house. Panel A-C were prepared as described in Chapter 3, whilst samples using Panel D were prepared as described in Chapter 5. All samples were sequenced on the Illumina MiSeq sequencing platform with PE250 reads.

Table A1.1: Multiplex Amplicon Sequencing Panels Designed

Version	Total targets	ARGs Targeted	MGEs Targeted
A	26	25	1
B	444	404	40
C	110	109	1
D	16	16	0

A1.2.2 Samples

To aid in the development of a bioinformatics pipeline which could be used to distinguish between actual amplicon sequencing variants and sequencing error, controlled samples with a known number of sequence variants were analysed. These mock samples which contained a known

assembly of ARG amplicons were also used to test the sensitivity of the panel. Finally, to further test the performance of the panels on representative environmental samples, six samples originating from wastewater (influent, mixed liquor and anaerobic digesters) and manure (chicken, cattle and swine) were included in each panel run. Both the wastewater and manure samples were composites of six samples taken from different locations across Canada. These samples were also analysed using shotgun metagenomic sequencing on the NovaSeq 6000 platform with PE150 reads and a sequencing depth of 35 million reads per sample to allow comparison with existing methods for the detection of ARG sequence diversity.

A1.2.3 Bioinformatics Pipeline

Unique sequences detected for each gene may have represented real ASVs or the results of sequence errors. To ensure an appropriate bioinformatics pipeline was developed, two filtering approaches were tested. The first method aimed to remove singletons and doubletons from the sequences obtained. The second method tested was more stringent, and assumed that the distribution of ASV frequencies under an assumption of pure error would follow a Poisson distribution centred around the mean of all counts. For each sample and gene, we tested whether we could detect ASVs with frequencies that represented outliers of a Poisson distribution (after correction for multiple testing using the Bonferroni method), representing higher abundance, real ASVs after PCR. If such ASVs were detected, we removed low count “noise” ASVs from the sample by finding the minimum of the distribution function and cutting all ASVs below the minimum. Both filtering methods were then tested on a controlled sample containing a known number of actual ASVs.

A1.3 Results

A1.3.1 Development of Bioinformatics Pipeline

A bioinformatics pipeline was designed by comparing the number of ASVs obtained after filtering with the known number of actual ASVs included in the control samples. An example of this approach is documented in Table A1.2. Using a filtering method which only aimed to remove sequence singletons and doubletons, it was observed that a large number of sequencing artifacts remained. Whilst using the filtering approach which assumed a Poisson distribution, the number of ASVs remaining was more representative of the actual ASVs in the sample (Table A1.2). This filter more accurately identified actual ASVs, however caution should be taken when using more than one primer pair to target the same gene.

Table A1.2: Comparison of actual ASVs and those detected after sequence filtering using Panel C

Gene Name	Expected ASVs	ASVs Remaining After Filtering	
		Singleton and Doubleton Removal	Poisson Filter
aph9	1	1	1
CTX-M	1	76	1
IMP-1	1	6	1
blaPAO/PDC	2	139	2
blaSHV	1	53	2*
blaTEM	2	424	2

*Amplified by two primer pairs

A1.3.2 Performance on Defined Samples

Control samples containing 15 ARGs targets were used to test the performance of the four panels. Within Panel A and D all targets were successfully identified, whilst in Panels B and C only 30 % and 67 % of genes were identified, respectively (Table A1.3). Sequences which did not significantly match with any of the targeted genes were classified as noise. This noise likely occurs

due to cross amplification of primer pairs and sequencing errors. It was found that panels with a greater number of targets typically had more sequencing noise reads, with up to 75 % reported in Panel B (Table A1.3). This noise may have impacted the sensitivity of the panels. To minimise the sequencing noise, smaller panels with fewer primers were chosen for future use. In addition, due to the potential to customise targets in house, Panel D was selected for further development.

Table A1.3: Performance of Multiplex Amplicon Sequencing Panels

Version	Total targets	Detected (%)[1]	Noise Reads (%)
A	26	100	<1
B	444	30	75
C	110	67	52
D	16	100	10

A1.3.3 Performance on Environmental Samples

To test the performance of Panel C and D compared to existing methods for the detection of ARGs, a selection of samples originating from wastewater and manure were analysed. These samples were also sequenced on the Illumina NovaSeq platform. Figure A1.1 displays the performance of Panel C and D compared to metagenomic shotgun sequencing. In this analysis, only 16 ARGs were considered, as this was the maximum number of targets within Panel D. It should also be noted that the samples used for this analysis were undefined, thus may not have contained the targeted ARGs.

Panel C consistently detected the lowest number of ARGs within the environmental samples. In mixed liquor, anaerobic digester, cattle manure and swine manure samples, Panel D outperformed shotgun metagenomic sequencing with a greater number of genes detected. Whilst in the influent

and chicken manure samples shotgun metagenomic sequencing detected a greater number of the ARGs.

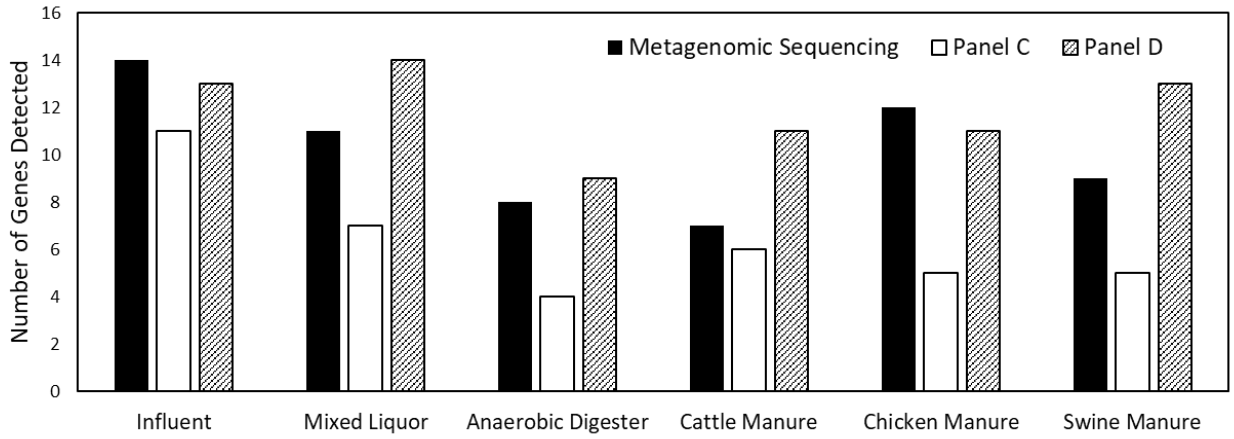


Figure A1.1- Genes detected using shotgun metagenomic sequencing and multiplexed amplicon sequencing using Panel C and D.

A1.4 Conclusions

By analysing controlled samples with known sequence diversity, a stringent Poisson filtering method was developed allowing actual ARG-ASVs to be distinguished from sequencing errors. It was found that multiplexed amplicon sequencing panels with fewer primers generated less sequencing noise and often had better sensitivity. When analysing undefined samples, the performance of amplicon sequencing Panel D appeared to be better than that of Shotgun metagenomic sequencing in terms of genes detected, in mixed liquor, anaerobic digesters, cattle manure and swine manure samples. These results highlight the potential for multiplex amplicon sequencing to be used in the future for the routine monitoring of environmental samples.