Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Characterisation of Microbial Biomat Development in Soil Treatment Units Receiving Domestic Effluent

Submitted to Trinity College Dublin, the University of Dublin

for the Degree of Doctor of Philosophy

By Alejandro Javier Criado Monleon

Supervision: Prof Laurence W Gill

May 2024

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent / do not consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Dublin, October 31, 2024

Acknowledgments

There are so many names that require thanks it is hard to count them even, but I will attempt to give thanks regardless. I would like to thank Professor Laurence William Gill my supervisor, who without his guidance and friendship this project would not have been feasible, few men would climb into a septic tank for you. Professor Muhammad Ali who arrived midway through the PhD, but provided crucial guidance and expertise which helped me advance in my research. I would like to send a warm thanks to Professor Kim Roberts of the Moyne Institute for her collaboration throughout my research, her cooperation was essential for critical covid research and was given generously at such a challenging time for us all.

I would like to thank all the technical staff at the department of Civil, Structural and Environmental engineering. Pat Veale for getting down into the trenches with me (literally), for his humour and your energy with doing what can be described as truly disgusting work. Joseph O' Connell for all his assistance in processing samples and professional guidance in the geotechnical lab. Robert Fitzpatrick for his assistance in developing soil column pumping systems and ensuring I did not fatally electrocute myself when working on circuitry. I would like to personally thank Mark Gilligan, Michael Grimes, and David McAuley for all their work behind the scenes keeping the labs going through some very challenging times.

Quiero dar mis gracias a mis padres Javier y Rosa, la verdad es que no pensaba que podría llegar a terminar algo tan difícil. Pero cuando no quería seguir vuestro amor y comprensión me ayudo calibrar me para terminar. To my sisters Olga and Veronica, I thank you both so much for your belief in your little brother and for being there for me, with lifts to site and all the driving lessons. To Aidan and Eoin for being brothers to me and the occasional pint. To my nieces and nephews Amaia, Ruben, Lola, and Anouk, you motivate me more than you can ever imagine, if you read this, please know that your Tio will try his best.

I want to thank Julianne, that we worked through such challenges being there for each other as best we could, but aside from the sheer bulk of emotional, physical, and financial heavy lifting, for buying me a car, I am a well-kept man. I want to thank you most of all for your kindness and support that you shared so generously with me; a PhD can feel so often like a tunnel vision. A focus on a singular goal with only one light positioned at the end, but you were always my reminder that the light that guides our way through life are all these little moments on our journey together, I love you, Julianne.

Summary

Soil treatment units are an essential part on on-site wastewater treatment system. In Ireland there are approximately 500,000 on-site wastewater treatment systems with 87% configured as a septic tank followed by a soil treatment unit. The treatment of domestic wastewater which is a high in organic and nutrients involves a complex series of biogeochemical reactions. These chemical transformations are mediated by microbial activity who utilise organic and nutrients as an energy source. These organisms can often coexist in soil environments and produce extracellular polysaccharide to form microbial mats 'biomats', which have been shown to enhance the functionality of soil treatment units. With accelerated development of molecular techniques microbial ecology has become an essential component in environmental engineering. The identification of carbon and nitrogen cycling species has allowed for the optimisation of microbial activity in engineering processes. However, much of the application of microbial ecology in wastewater treatment is focused primarily on centralised wastewater treatment plant processes, particularly activated sludge treatment systems.

The aim of this research is to develop a better understanding of the microbial ecologies present within soil treatment units and the ecological profile of the microbial mats. Three on-site wastewater treatment system were surveyed for this study. The sites were built to EPA standards with the sites utilising both a standard septic system and packaged secondary treatment unit with the standard 4 percolation trenches assigned influent from either source to assess the effect of domestic wastewater pre-treatment. These sites were purpose built and instrumented with lysimeters and a three-dimensional soil sensor network to enable characterisation of the clogging of the biomat over time and chemical composition of the porewater below the systems. Additionally, four subsoil mesocosms were constructed in the laboratory, designed to mimic domestic wastewater dosing on a soil treatment unit and to analyse the effects of pre-treatment and porosity on the development of the biomat. Chemical and physical monitoring was combined with regular biomat and subsoil genomic sampling campaigns for both the field studies and the mesocosm experiments. Environmental DNA was extracted from the subsoil and was genetically sequenced to profile the biomat communities.

The results show that the pre-treatment of effluent has a significant effect on the microbial community structure of organisms present within the on-site wastewater treatment units and the laboratory mesocosms. The application of high organics and nutrients provides a selective pressure on communities with the community structure being determined by species sorting. Over time communities exposed to high organic loading exhibit increases in their diversity as the availability of substrate aided by the clogging within the soil, this promotes the transformation of nutrients and the creation of microbial hotspots. However, over time many bacteria within these communities will be out competed by copiotrophic bacteria capable of dominating the high nutrient environment. The

functionality of the bacteria is wholly linked to the reduced hydraulic conductivity delivered by a robust soil biomat. In comparison, in soils dosed with secondary treated effluent the presence of a muted biomat, due to a low concentration of organics in the effluent feed appears to delay these shifts in community structure in a pattern that is distinct to biomats dosed with primary treated effluent.

Contents

1. Introduction 1

1.1 Background and Motivations **1**

1.2 Research Aims and Objectives **2**

1.3 Thesis Organisation **3**

2 Literature Review 4

2.1 On-site wastewater treatment 4

2.1.1 Access to sanitation: global context **4**

2.1.3 On-site wastewater treatment case study: Republic of Ireland **6**

2.1.4 On-site Wastewater Treatment Technologies **10**

Septic Tank (primary treatment) 10

Packaged Treatment Systems (secondary treatment) 11

Rotating Biodisc Contactors 12

Media Filter 13

Sand Filters 14

Soil Treatment Unit (STU) 14

2.1.5 OWTS Management **16**

Planning 16

Inspection 19

2.2 Domestic Wastewater Contamination Pathways 21

2.2.1 Domestic Wastewater Characteristics **21**

Domestic Wastewater Composition 22

Organic Matter 22

Nitrogen 23

Phosphorous 23

2.2.2 Faecal Indicator Bacteria and Microbial Source Tracking **24**

2.2.3 Contaminant flow through the unsaturated zone **26**

Biomat development 26

Nutrient Removal through the STU 27

Microbial Contaminant Removal through the STU 29

2.3 Microbial Ecology functionality and community development in STUs 30

2.3.1 Dominant Phyla in Soils **31**

2.3.2 Dominant Taxa in wastewater treatment system **33**

2.3.3 Functional species in wastewater treatment **35**

Methanogenesis 36
Ammonification 37
Nitrification 38
Denitrification 38
Anaerobic Ammonia Oxidation (Anammox) 39
Biotic Phosphate Accumulation 40
2.4 Community ecology development in on-site wastewater treatment 40
2.4.1 Factors Affecting Microbiome Development **42**
2.4.2 Microbe-Environment Interactions **44**
2.4.3 Critical Points in Transition of Microbial Ecologies **45**
3. Field Studies and Laboratory Analysis 48
3.1 Field sites 48
3.11 Killmallock (Site A) **50**
3.12 Crecora (Site B) **53**
3.13 Wicklow (Site C) **55**
3.2 On-site Instrumentation 58
3.2.1 Automated data collection **58**
3.2.2 Soil moisture sensors **61**
EC5-type sensor 61
GS3-type sensor 61
Soil sensor calibration 62
Soil sensor installation 64
3.2.3 Suction cup porewater lysimeters **65**
Porewater Sampling 66
Sites 67
3.2.4 Bioports and Installation **68**
3.3 Column Construction 70
3.4 Sampling protocol 74
3.4.1 Sample collection **74**
3.4.2 Biomat Subsoil Field Sampling **74**
3.4.3 Biomat soil column sampling **75**
3.4.4 Lysimeter Porewater Sampling **75**
3.5 Pepper Mottle Virus Faecal Indicator Monitoring Sampling Campaigns 76
3.5.1 Sampling rivers impacted by on-site wastewater treatment systems. 77

3.6 Laboratory analysis 78
3.6.1 Chemical analysis 78
3.6.2 Culturing methods 79
3.6.3 Molecular analysis 79
Sample Pre-processing 79
DNA and RNA Extraction 80
16S RNA amplicon sequencing. 80
Bioinformatics 81
3.6.4 Reverse Transcription Quantitative Polymerase Chain reaction 82
Sequence Literature Study, Target and Primer Selection 82
Primer Validation, Inhibition Tests and PMMoV Quantification 83
Translating PMMoV genes copies into concentration 84
PMMoV loads 84
Limit of Detection 84
3.6.5 Bacterial growth rates 85
Calculation of NxCZ 86
Calculation of NxCsZ 86
Calculation of NxINF 87
Calculation of NxEFF 87
Solid Retention Time 87
Net growth rates 87
4. Spatial variation of the microbial community structure of on-site soil treatment units in a temperate climate, and the role of pre-treatment of domestic effluent in the development of the biomat community. 88
4.1 Abstract 90
4.2. Introduction 90
4.3. Methods 92
4.3.1 Site Description 92
4.3.2 Sampling and 16S rRNA Gene Sequencing 94
4.3.3 Sequence Processing and Analysis 95
4.3.4 Site Instrumentation 95
4.3.5 Data Availability Statement 96
4.4. Results 96
4.4.1 Meteorological Conditions 96
4.4.2 Effluent Quality and Wastewater Treatment System Performance 97

4.4.3 Biomat Position 98
4.4.4 Microbial Community Composition Within The STUs 99
4.4.5 Biomat Microbial Community Structure in Response to Effluent Dispersal 105
4.4.6 Target Organisms Screened for Biogeochemical Functionality 108
4.5 Discussion 112
4.6 Conclusion 114
5. Influent Pre-treatment of Greater Significance than Porosity in Microbial Community Assemblage Patterns in Soil Treatment Units 115
5.1 Abstract 116
5.2 Introduction 117
5.3 Materials and Methods 119
5.3.1 Column Setup and Instrumentation 119
5.3.2 Sampling and 16S DNA Gene Sequencing 120
5.3.3 Sequence Processing and Analysis 121
5.3.4 Calculation of SRT and Net Growth Rates From Amplicon Data 121
5.3.5 Data availability 122
5.4 Results 122
5.4.1 Wastewater influent pre-treatment impacts pollutant removal efficiencies 122
5.4.2 Biomat development influenced by the level of pre-treatment 125
5.4.3 Spatial Variation in Microbial Community Composition Influenced by pre-treatment 126
5.4.4 Influent pre-treatment results in increased levels of immigration and growth of bacteria from influent into the soil column 131
5.5 Discussion 135
5.6 Conclusion 137
6. Pepper Mild Mottle Virus as an Effective Tool in Microbial Source Tracking for Failing Domestic On-Site Water Treatment Systems 138
6.1 Abstract 139
6.2 Introduction 140
6.3 Methods 142
6.3.1 Soil Column Set up and Spiking Trial 142
6.3.2 Field site, Instrumentation and Spiking Trial 142
6.3.3 Rivers, Drainage Network Selection and Sampling 143
6.3.4 Organic and Nutrient analysis 144

6.3.5 Enumeration of Faecal Indicator Bacteria....144
6.3.6 RNA Extraction and PMMoV Quantification....144
6.4 Results and Discussion....145
6.4.1 Pepper Mottle Virus Mild Mottle Virus Transport and Removal: Laboratory Soil Column 145
6.4.2 Pepper Mottle Virus Mild Mottle Virus Transport and Attenuation: Field-Scale On-Site Wastewater Treatment System....149
6.4.3 PMMoV Detection and Enumeration: Catchment-Scale with High Contamination Pressure from On-Site Wastewater Treatment Systems....152
6.5 Conclusions....155
7. Temporal Variations in Microbial Ecological Patterns in Soil Treatment Unit Biomats .156
7.1 Abstract....156
7.2. Introduction....157
7.3 Methods....159
7.3.1 Sites description....159
7.3.2 Sampling and 16S RNA Gene Sequencing....160
7.3.3 Sequence Processing and Analysis....161
7.3.4 Site Instrumentation....162
7.3.5 Data Availability Statement....162
7.4 Results....162
7.4.1 Meteorological conditions....162
7.4.2 Effluent Quality and Wastewater Treatment System performance....163
7.4.3 Temporal changes to soil treatment system community composition....166
7.4.4 Temporal variation in microbial community structure in STU....171
7.4.5 Temporal changes in biogeochemical functional groups....175
7.5 Discussion....182
7.6. Conclusions....185
8. Conclusion and Recommendations....186
8.1 Summary of Findings....186
8.2 Significance of this study....189
8.3 Recommendations for Expanding Research....190
Appendix A: Supplemental Spatial Variation Study....191
Appendix B: Supplemental to Soil Mesocosm experiment....204
Appendix C: Supplemental to PMMoV trials....212

Appendix D: Supplemental to Temporal Variation Study 214
Bibliography..... 226

Introduction

1.1 Background and Motivations

Access to sanitation remains a significant global challenge with an estimated 3.9 billion people currently living without safely managed sanitation and communicable diseases from excreta responsible for the annual deaths of 1.9 million people (WHO., 2019; WHO., 2020; UN., 2022). Due to the high cost of construction and maintenance, centralised wastewater conveyance and treatment systems are more often found near population centres, creating an urban-rural divide for access to sanitation. Ireland has a relatively high rural population for Europe at 31% (CSO.,2019). This has resulted in a high number of households (~500,000) employing a decentralised on-site wastewater treatment system to safely dispose of their wastewater. Domestic wastewater is high in concentrations of organics, nutrients, and pathogens. If not adequately treated and disposed domestic wastewater can pose serious risks to both public and environmental health. At present 9% of Irish water bodies are characterised as having on-site wastewater treatment systems as a significant pressure (EPA., 2022).

On-site wastewater treatment systems most commonly consist of a septic tank and a soil treatment unit. The septic tank provides the primary treatment, solids separated from the liquid fraction, and some minor anaerobic fermentation. Following this, the effluent proceeds to a soil treatment unit (STU) where most of the treatment occurs. Whilst the development of genetic sequencing technologies has advanced understanding of centralised wastewater treatment systems for community composition and structure (Wu et al., 2019) and there has also been a great deal of study on microbial soil ecology, with fields such as biogeography providing detailed surveys on microbial soil communities globally (Delgado-Baquerizo et al., 2018), there has been little targeted research to investigate soil treatment unit ecosystems receiving different wastewater effluents. Research focused on natural subsoils usually focuses primarily on the topsoil, as natural subsoil ecosystems are not well understood and therefore less predictable (Powell et al., 2015).

This leaves a large knowledge gap in the understanding of subsoil ecosystems, which if incorporated appropriately by rigorous engineering design can provide a passive nature-based approach for the treatment of domestic wastewater. This treatment is augmented by the clogging of the soil which is enhanced by the development of a microbial mat along the infiltrative surface of the soil (Knappe et al., 2020). Most of the existing literature on the soil treatment systems focuses primarily on the transport and attenuation of contaminants through within STUs, such as organics (BOD, COD, TOC), nutrients and faecal indicator bacteria as surrogates for pathogens (Gill *et al*., 2009; O'Luanaigh *et al*., 2012; Humphrey et al., 2019). Studies have linked the effect of pre-treatment on the development of the biomat by measuring permeability of the subsoil using tensiometers and soil moisture sensors. It was found that STUs dosed with secondary treated effluent from small-scale

packaged plant treatment systems resulted in reduced clogging and nutrient attenuation when compared to those STUs receiving effluent from septic tanks (Gill et al., 2007; Gill et al., 2009; Knappe et al., 2020). It is therefore crucial to understand the ecology of an STU receiving effluent under varying degrees of pre-treatment, as to better determine the overall functionality of the system. At present there are a few papers that have focused on surveying the ecology in on-site wastewater treatment systems, but these papers focus primarily on monitoring the ecology present within the septic tank itself and not the soil treatment unit (Knisz et al., 2020; Ross et al., 2020). Tomaras et al. (2009) did survey a single soil treatment unit taking four samples and sequencing them for the microbial composition, finding that no taxa present within the wastewater were discovered in the soil treatment unit. However, it appears that no studies have carried out comprehensive investigations, spatially or temporally, into the soil microbial ecology within these globally important treatment systems, nor investigated the impact of different levels of effluent pre-treatment. There remain large gaps in the understanding of this altered subsoil microbial ecology and so further understanding of this engineered environment is essential for the optimisation of this basic but crucial sanitation system.

1.2 Research Aims and Objectives

The aims of this research develop the improved understanding of the microbial ecosystem within the on-site wastewater soil treatment unit.

This was approached via the following objectives:

- Profiling the spatial variation of the biomat ecology within the soil treatment unit in a temperate climate.
- Extending the current understanding of microbial biomass growth by relating it to the microbial composition.
- Identifying microbial hotspots and discerning temporal variations within the biomat through high frequency sampling.
- Profile young and mature biomats through long-term ecosystem monitoring.
- Studying the temporal variation of microbial biomat community structure, defining an ecological timeline and stages in the development of the community.
- Determine the effect of domestic effluent pre-treatment on the biomat community structure.
- Assessing spatial and temporal variation of functional groups within soil treatment units, by referencing wastewater treatment ecosystem-specific taxonomic databases.
- Quantifying the contaminant removal efficiencies of on-site wastewater treatment systems relative microbial biomat development, at the bench and field scale.
- Using targeted column studies to identify the deterministic factors driving community assemblage patterns in soil treatment units with established biomats.

- Validate novel microbial source tracking methods using Pepper Motile Virus for improving faecal source tracking of failing on-site wastewater treatment systems in a rural environment.

1.3 Thesis Organisation

Chapter 2 presents a comprehensive literature review to outline the relevant knowledge regarding the nexus between microbial ecology and on-site wastewater treatment.

Chapter 3 details the research sites and soil columns constructed instrumented and sampled during this research. It also details analytical methods used during this study.

Chapter 4 assesses the spatial variation of the microbial community structure of on-site soil treatment units in a temperate climate, and the role of pre-treatment of domestic effluent in the development of the biomat community.

Chapter 5 investigates the effect of influent pre-treatment and porosity in microbial community assemblage patterns in soil treatment units.

Chapter 6 details the validation of pepper mild mottle virus as a tool in microbial source tracking for failing domestic on-site water treatment systems.

Chapter 7 investigates the temporal variations in microbial ecological patterns in the development of the soil treatment unit biomat.

2 Literature Review

This chapter presents a comprehensive literature review to outline the relevant knowledge regarding the nexus between microbial ecology and on-site wastewater treatment. The chapter is divided into 4 sections.

Section 2.1 provides a brief introduction to on-site wastewater treatment systems, the global status in access to sanitation, providing an Irish case study in their dissemination and management. Section 2.2 Defines domestic wastewater sources, its composition, and the fate of contaminants within the soil environment. Section 2.3 Provides an overview of microbial composition, functionality in soils and engineered sites. Section 2.4 Summarizes factors in the development microbial communities in on-site wastewater treatment systems and how these communities interact with the engineered environment.

2.1 On-site wastewater treatment

2.1.1 Access to sanitation: global context

Wastewater treatment is a necessity to ensure human stewardship over the water cycle. The development of treatment technologies is required to mitigate the threats of wastewater to human health and environmental sustainability. It forms a major component of the United Nations Sustainability Development Goals (SDGs), specifically SDG 6 which is to ensure availability and sustainable management of water and sanitation for all.

There are many tiers and classifications of sanitation and wastewater disposal globally. Limitations in accessibility across the world often prevail because of limited financial resources, a factor which is significantly compounded in rural settings (UNESCO., 2021). High income nations on average treat 70% of their municipal and industrial wastewater, but for middle- and low-income countries this drops to 28 - 38% and 8% respectively (UN., 2017). Safely managed sanitation services are defined as a proportion of the population using improved sanitation facilities which are not shared with other households, where excreta can be disposed of safely in situ or transported offsite; in 2020 this accounted for 54% percent of the global population (WHO., 2021) (**Figure 2.1**). Treatment systems that hygienically separate human excreta from human contact are defined as *"improved"* sanitation facilities. These facilities can consist of wet sanitation technologies such as flush or pour flush toilets directed to a sewer system, septic tank, or pit latrine; and dry sanitation technologies such as dry pit latrines with slabs (built from durable and easily cleanable materials), ventilated improved pit (VIP) latrines, pit latrines with a slab and compositing toilets and container-based sanitation. If these facilities are not shared, they can be classified as "*basic facilities"* which accounted for 24% of sanitation coverage in 2020 (WHO.,

2021). If improved sanitation facilities are shared, they are defined as "*limited facilities"* which covers the sanitation needs for 7% of the global population (WHO., 2021). If a community's sanitation wet or dry facilities do not safely remove excreta from human contact, instead if it is stored in open pits or variety of ad hoc containers, if it is disposed in stormwater channels or surface waters in general this is described as unimproved. In 2020 8% of the global population's excreta was disposed of in unimproved systems (WHO.,2021).

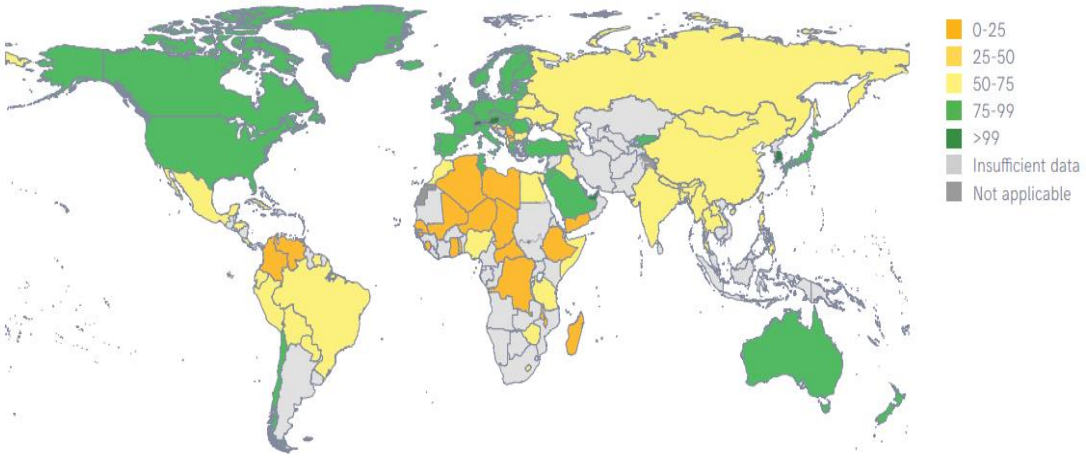


Figure 2.1 Proportions of global population employing safely managed sanitation services, 2020 (source: WHO., 2021).

Latest figures show that 3.9 billion people are without safely managed sanitation, of these 494 million people practice open defecation, representing a population total greater than all the denizens of the European Union (446.7 million) (WHO., 2020; UN., 2022). Preventable communicable diseases from excreta have a significant health burden globally accounting for 1.9 million deaths annually and 123 million disability-adjusted life years and is reckoned to be the second leading cause of death of children under the age of 5 years (WHO., 2019). Diarrhoeal diseases alone are the eighth leading cause of death worldwide, caused primarily by *Escherichia coli* and *Rotavirus.* However, *Vibrio cholera* remains endemic in 69 countries and if untreated can prove fatal in a matter of hours, with the disease killing 95, 000 people annually (WHO., 2019). Helminthiasis, i.e. macroparasitic diseases caused by helminths, affect 24% of the global population. The worms which can burrow into the gastrointestinal tract, and the urinary tract can cause abdominal pain and diarrhoea, but ultimately the long-term consequences are more severe for children who can suffer mental impairment and stunting from chronic infections (WHO., 2019). Improving access to sanitation, therefore, reduces the prevalence of these preventable diseases and the human suffering they cause and alleviates the burden on health infrastructure of nations.

The challenge of implementing SDG 6 is driven by increases in population, rates of urbanisation, and human-induced climate change. Climate change is creating significant variations in

precipitation patterns, and this is resulting larger episodes of drought or near-drought conditions in much of the natural and man-made surface water bodies across the globe (UN. 2020). It is important to note that these low flow conditions also concentrate pollutants from wastewater treatment facilities increasing the risk to human health and environmental deterioration (Van Vliet et al., 2008). Greater variation to surface water sources will put added pressure to groundwater resources which accounts for 99% of liquid global freshwater (UN., 2022). Effective protections and sustainable practices are needed to prevent over abstraction, reduce contamination and enhance recharge in a responsible manner.

2.1.3 On-site wastewater treatment case study: Republic of Ireland

Approximately 65% of the Republic of Ireland household's wastewater is treated by 1,000 centralised wastewater treatment plants, connected by 26,000 km of public sewers and water pumping stations (EPA., 2021). The wastewater is then discharged into surface waters (inland rivers or the sea), however, between 2017 – 2021 49% of this wastewater effluent discharge was deemed to be below EU standards (EPA., 2021). In terms of quantity, much of the failures are occurring at the Ringsend wastewater treatment plant which serves the capital city of Dublin and is greatly over capacity. The centralised systems failing EU standards is mainly due to insufficient removal of organic matter and nutrients. This is compounded by the fact that, as of 2021, 32 Irish towns public sewer systems were not connected to a treatment facility resulting in average release of 2,737,500 m^3 of raw sewage being released into the environment per year (EPA., 2021). The remaining third of households in Ireland treat their domestic wastewater using an On-Site Wastewater Treatment System (OWTS) (Dubber and Gill., 2014). This is due to the relatively high proportion (31.4%) of total Irish population living in a rural setting (**Figure 2.2**) (CSO.,2019). In total this accounts to nearly 489,000 households or more than 1.4 million people (**Figure 2.3; Figure 2.4**). It is estimated that 87% of these on-site treatment systems are comprised of septic tanks with some form of soil percolation system (either a soakaway or a percolation area) (Dubber et al., 2014).

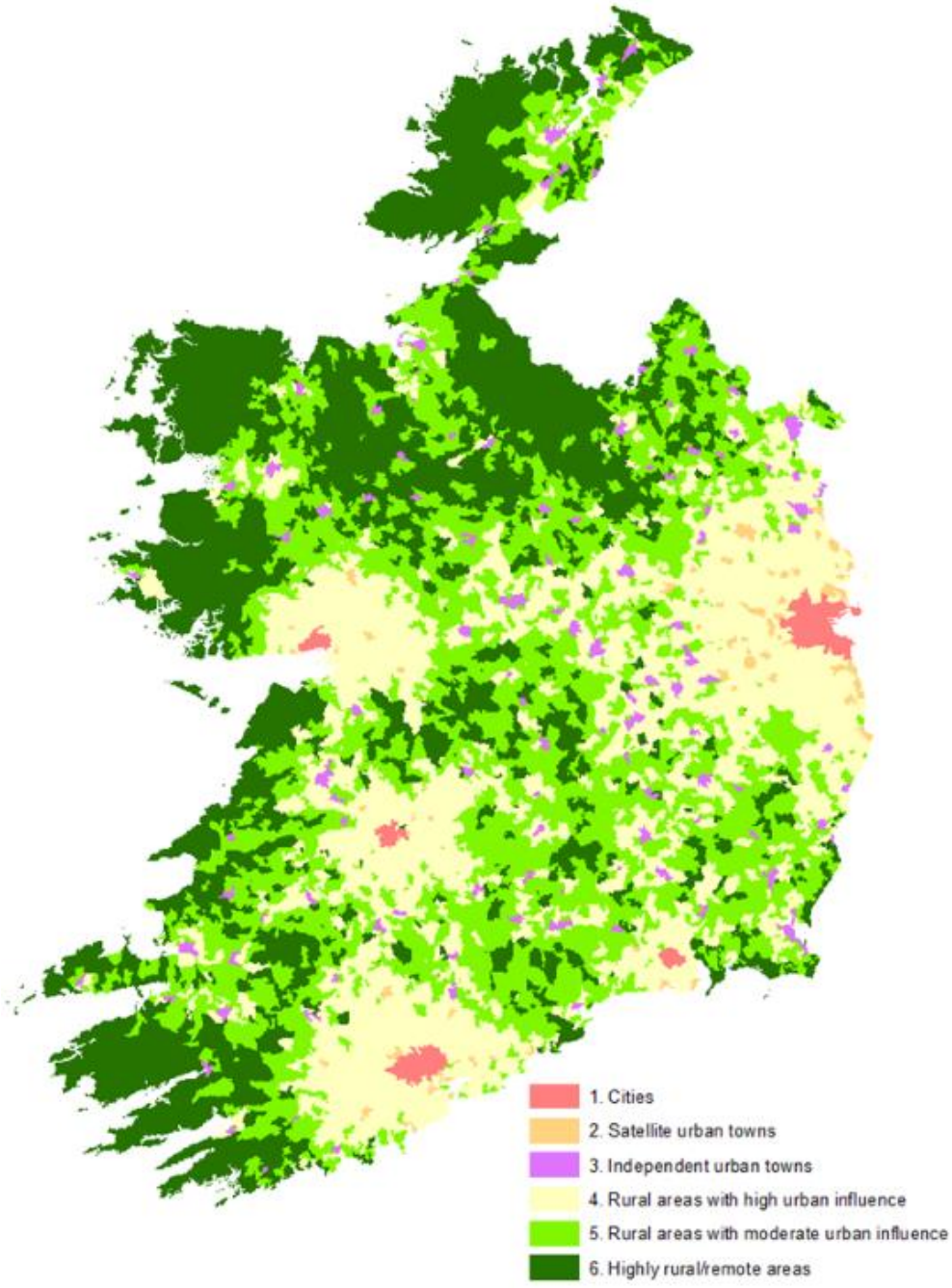


Figure 2.2 Population distributions of urban and rural locations in Ireland (Source: CSO., 2019)

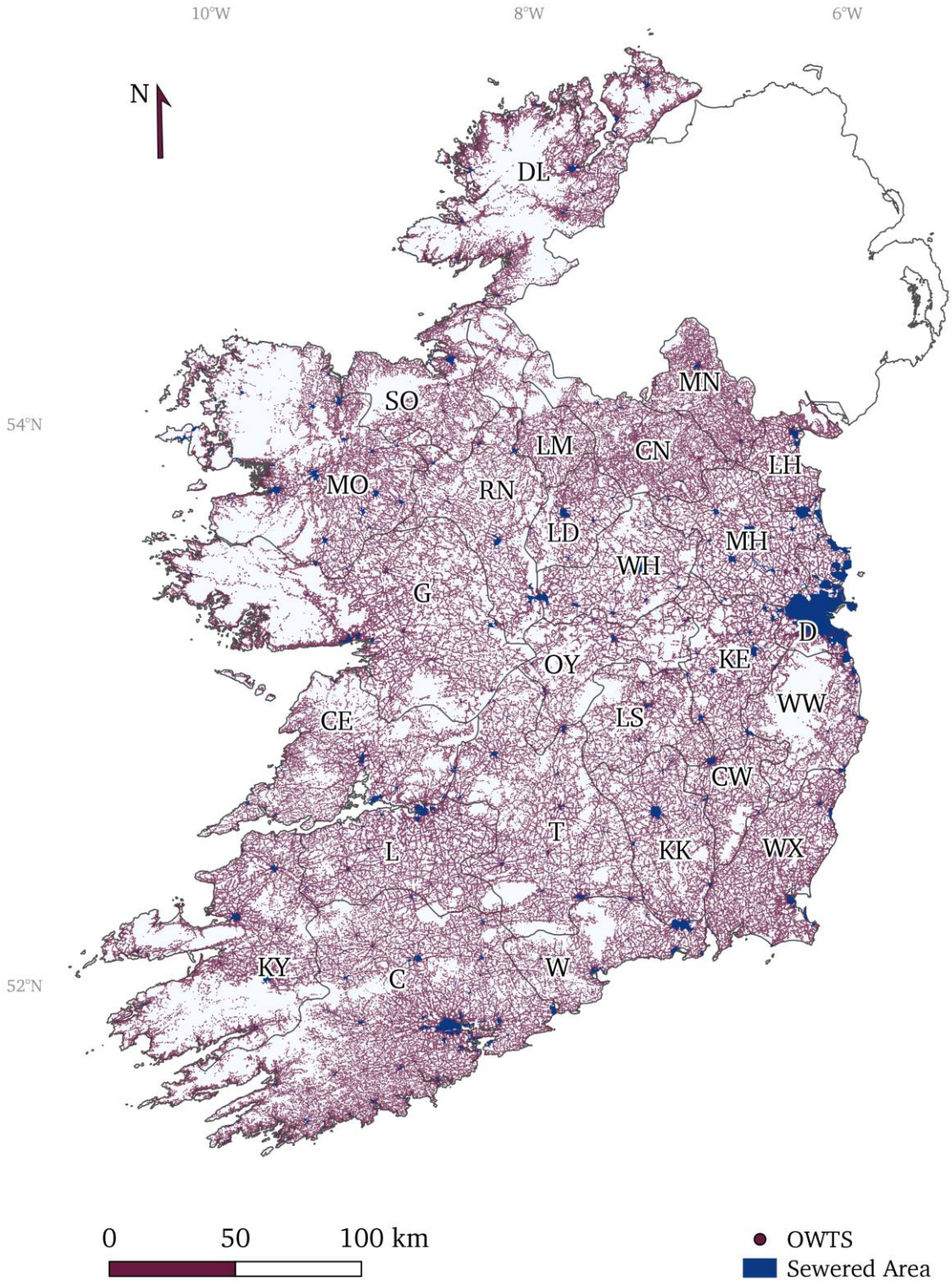


Figure 2.2 Spatial distribution of on-site wastewater treatment systems and sewered areas in Ireland. Burgundy dots mark individual OWTS, blue areas mark locations with access to a centralised sewerage network. (Source: Jan Knappe., 2020)

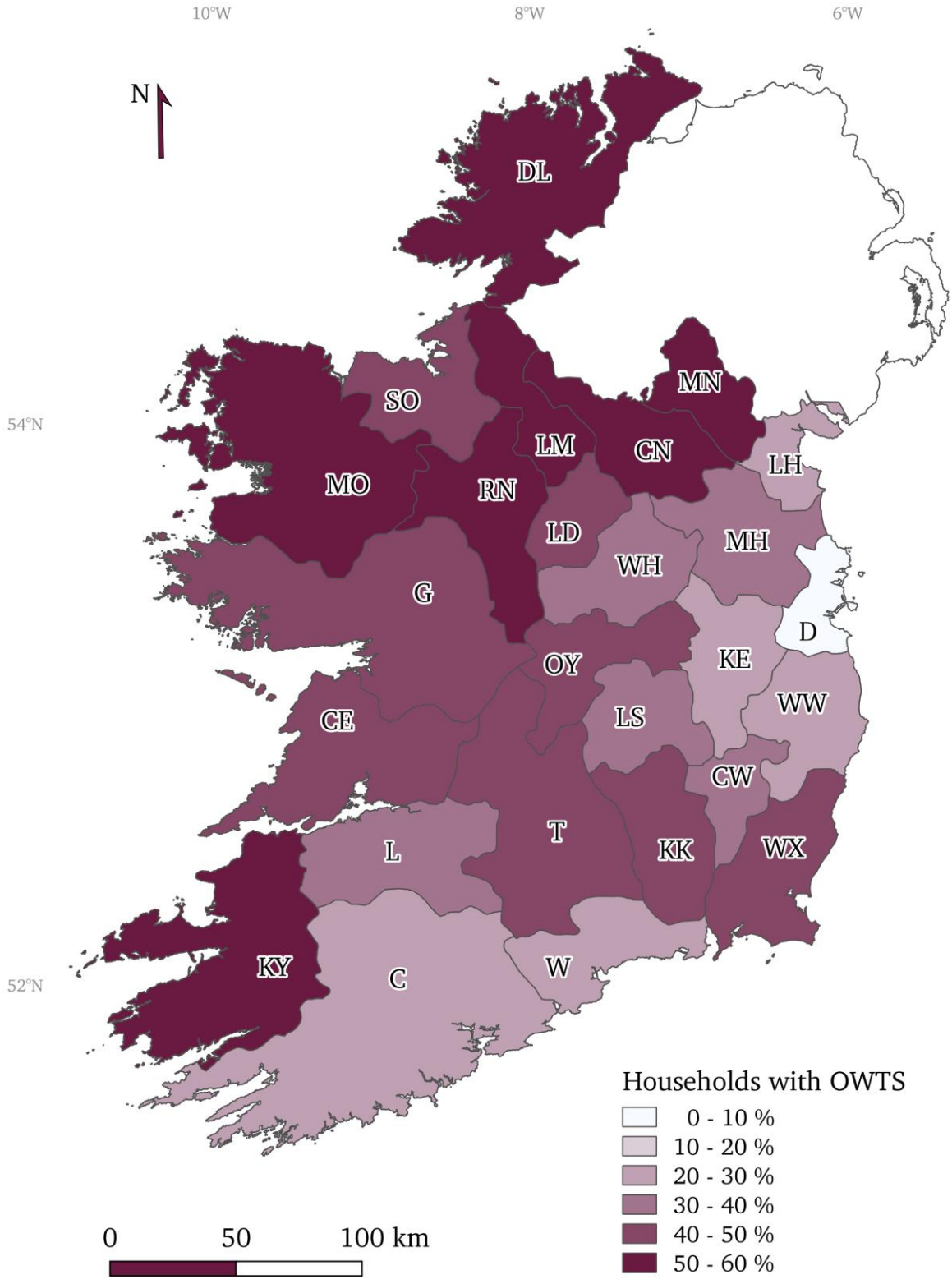


Figure 2.3 Shares of households with OWTS per county. (Source: Jan Knappe., 2020)

2.1.4 On-site Wastewater Treatment Technologies

Septic Tank (primary treatment)

A typical septic tank functions as a primary treatment system, providing a settling and skimming process, as an unheated, unmixed anaerobic digester which acts to accumulate sludge (EPA., 2009) (**Figure 2.4)**. These tanks are commonly constructed from precast concrete, or plastic such as polyethylene. As a primary treatment process, the sought outcome is the retention of suspended solids from entering the soil treatment area. The most critical component of an on-site wastewater treatment system with a septic tank is the soil treatment unit (STU), as it provides most of the treatment of the wastewater effluent (EPA., 2009) – see later. For the optimal physical removal of solids, the water must reside in the tank for at least two days' duration (i.e., hydraulic retention time (HRT)) to allow settling to occur. Particles denser than water settle creating sludge and those less dense than float forming scum. The septic tank acts as clarifier, where the clarified effluent in the tank is displaced with new flow entering the unit. The effective HRT reduces over time as the increase in the accumulated sludge volume, acts to decrease the volume available for liquid movement. If this volume becomes too small leading to too short an HRT, then the effectiveness of solids separation starts to drop. This is of significant concern as it knocks on to the effectiveness of the soil treatment unit (STU) as greater concentrations of particles can lead to an increased likelihood clogging in the STU (Amador et., 2019).

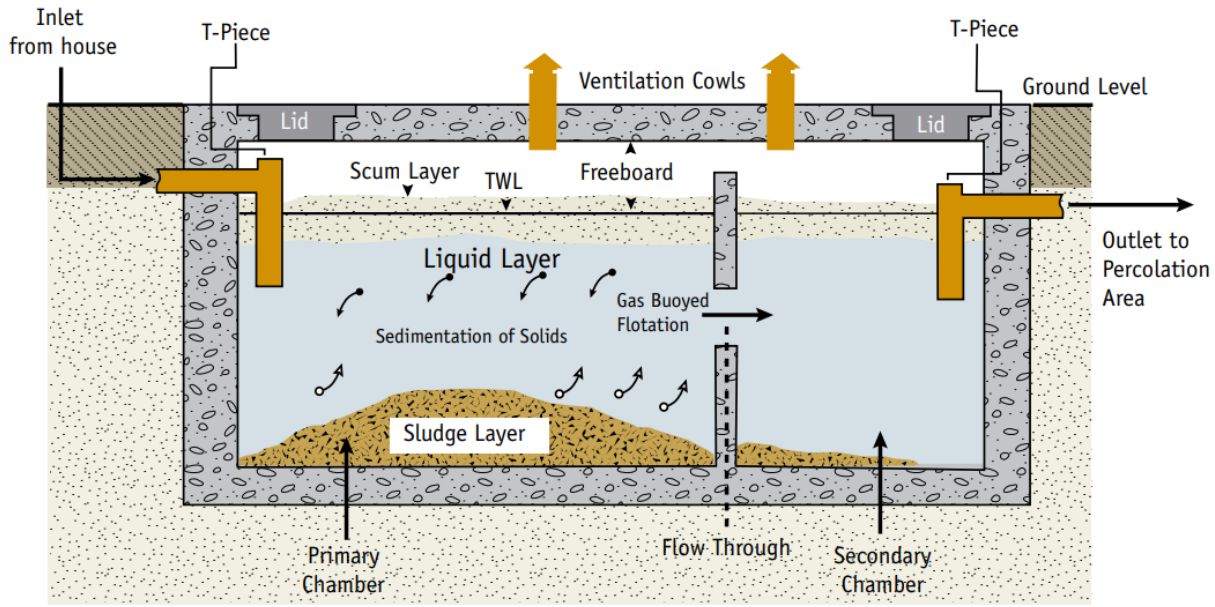


Figure 2.4. Longitudinal cross-section through a typical septic tank (EPA., 2021).

The volume necessary for sludge storage is an important factor in selecting a septic tank size. The daily design flow of wastewater from households is determined by the maximum occupancy of the household. Occupancy is assessed by the number and size of bedrooms present in the domicile. Flows range from 420 L and 570 L per bedroom per day, depending on the number of

occupants in each bedroom. Using maximum flow values allows for a conservative estimate preventing surge flows from overcoming the capacity of the system (Amador et al., 2019; EPA., 2009). In order to maintain effective hydraulic retention times, desludging should occur at a frequency of every 3 years for septic tanks over 3.5m^3 (EPA., 2016; Mac Mahon et al, 2021). For systems smaller 3.5 m^3 or older systems annual inspections must be carried out by local authorities (EPA., 2016).

Packaged Treatment Systems (secondary treatment)

Alternative or additional treatment systems can be used to replace or augment existing septic tank systems to provide secondary treatment of effluent before it is discharged to the soil treatment area. These systems utilise active mechanical processes (pumps or aeration) to facilitate an aerobic microbial interface for domestic wastewater treatment (**Figure 2.5)**. These microbial interfaces can be suspended growth reactors or a fixed surface area to permit the growth of a biofilm. Alternative treatment units (ATU) require regular monitoring and maintenance as they are more susceptible to failure in comparison to passive septic tank systems. It should be noted that most packaged treatment systems used for on-site domestic wastewater, not only promote carbonaceous (organic) removal but also are able to support at least partial nitrification of wastewater (e.g. Delaney et al., 2015).

The Biological Aerated Filter (BAF) applies a system which incorporates a primary settlement tank followed by an aerated biofilm filter and a secondary settlement tank. Such processes are installed within glass-reinforced plastic, concreted or steel cases. The biofilm media consists of a granular plastic material as to maximise the surface area of biomass material in contact with the wastewater effluent. Over time an organic microbial biofilm develops which acts to treat the wastewater.

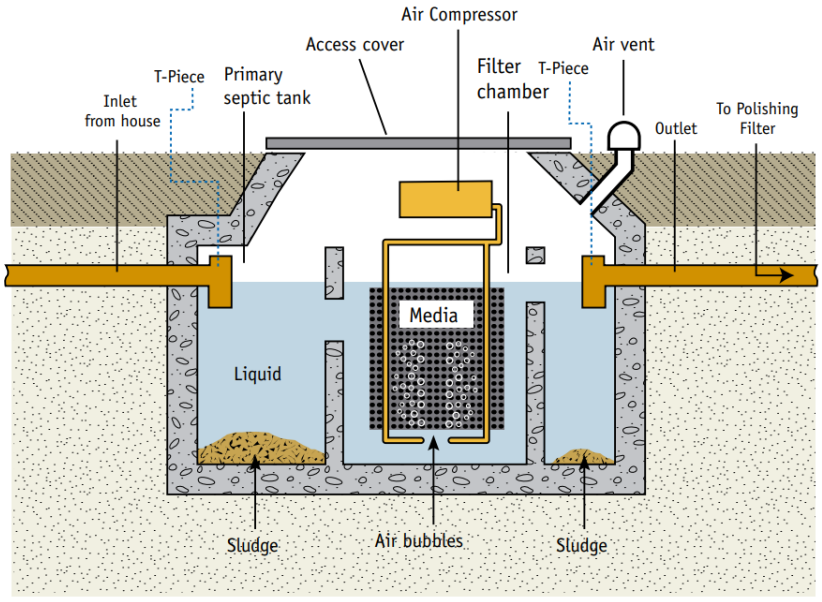


Figure 2.5. Longitudinal cross-section of a BAF system (EPA., 2021).

A Sequencing Batch Reactor (SBR) is a staged batch active sludge treatment process for domestic wastewater. The filling stage consists of mechanical mixing of the effluent to develop anoxic conditions within the tank. The effluent is then aerated by means of fixed or floating mechanical pumps. The air sparged into the effluent in fine sized bubbles created by diffusers located at the bottom of the tank. Once aerated for the required aerobic treatment time, the final process is to settle suspended solids to decant the supernatant to the soil treatment unit. SBR systems are quite technically advanced requiring high levels of maintenance and control of processes.

Membrane Bioreactor (MBR) filtration systems treat effluent by the removal of both suspended solids and dissolved molecular material from the effluent as it passes across a specific membrane material. The system involves a treatment tank with aeration into which the membrane filtration units are installed (**Figure 2.6)**. The effluent released from these systems is of a very high quality due to the very small pore size of the membranes which need to be put under a negative pressure in order to pull water through. The membranes are connected by membrane mountings on a collection manifold, in which the effluent collects and are passed through toward the outlet. These systems require regular chemical cleaning and incur high operating costs.

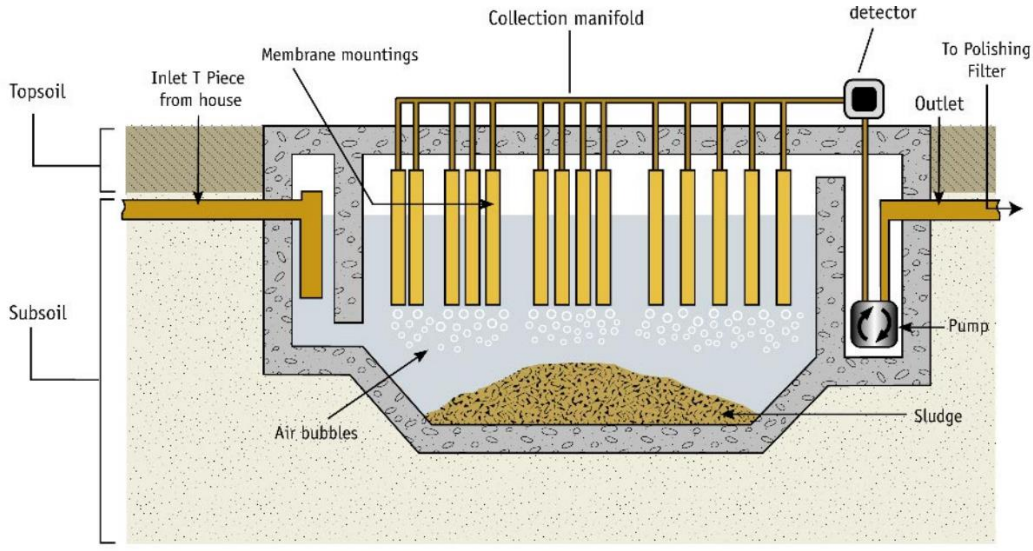


Figure 2.6. Schematic cross-section though a Membrane Bioreactor system (EPA., 2009).

Rotating Biodisc Contactors

A Rotating Biodisc Contactor (RBC) provides a material for the growth of a microbial biofilm. In the treatment process 40% of each rotating disk is immersed into anoxic wastewater, with the other 60% exposed to the atmosphere creating an aerobic environment. This rotation between oxygen states promotes the microbial degradation of wastewater constituents. The biofilm formed on the surface on the disk gradually accumulates but is then sloughed from the disk (by the hydraulic shear due caused by the speed of rotation of the disc) which accumulates as sludge settled on the bottom

of the tank (Amador et al., 2019; Langwaldt et al., 2000) (**Figure 2.7)**. These systems usually receive septic tank effluent transported to the RBC or sometimes incorporate their own primary settlement chamber up front. Depending on the rate of wastewater flow several disks may need to be installed in series to provide the required treatment (Amador et al., 2019).

Figure 2.7. Klagester Biodisc RBC. Beginning left, shows the schematic of the RBC system including the integrated primary settling chamber, two-stage biozone, and secondary clarification chamber. The middle figure displays the device upon installation prior to the development of a biofilm. The figure right displays the device following 24 months of biofilm development, the bottom disks display stage 1 biozone and the upper biodisk displays stage 2, notice the colourisation differences between both biofilms.

Media Filter

A media filter applies an intermittent dose of wastewater displaced over a filter medium. The filter medium can be formed of soil, sand or other material which provides a significant surface area for biofilm production such as coconut husk, peat or textile(**Figure 2.9)**. The effluent trickling through the media is processed through biofiltration (filtration, absorption, adsorption, ion exchange, microbial assimilation) whereby its carbonaceous substrate content is reduced.

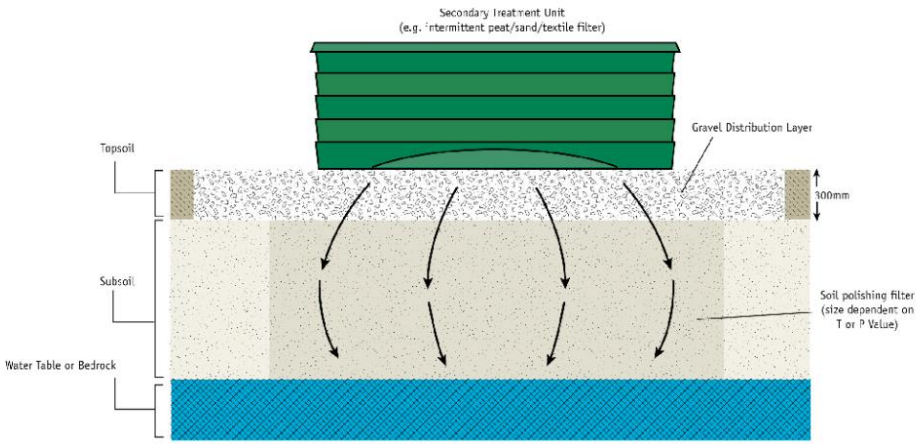


Figure 2.9. Schematic of packaged media filter (SOURCE: EPA., 2009).

Sand Filters

Intermittent sand filters are an effective form of on-site treatment. The wastewater treatment takes place under predominantly unsaturated and aerobic conditions caused by intermittent dosing of primary treated effluent through a pressurised manifold onto the bed of and media (EPA., 2009). There are two types primarily used for wastewater treatment; soil covered, and open. Soil-covered intermittent sand filters may be underground, part underground and part overground or overground (EPA., 2009). Open intermittent systems are exposed, offering access for inspection and maintenance. Intermittent sand filters are usually non–recirculating whereby biofilm develops on the surface area of the sand (EPA., 2009). The structure of the sand filter consists of several beds of graded sand, depths are commonly designed to 700-900 mm, with the base of the filter consisting of 200 mm of filter or washed pea gravel (EPA.,2009) (**Figure 2.10)**. Sand grain sizes required depend on the system design, with open systems requiring a D_{10} range from 0.4 mm to 1.0 mm, and soil covered systems a D_{10} range 0.7 from to 1.0 mm (EPA., 2009).

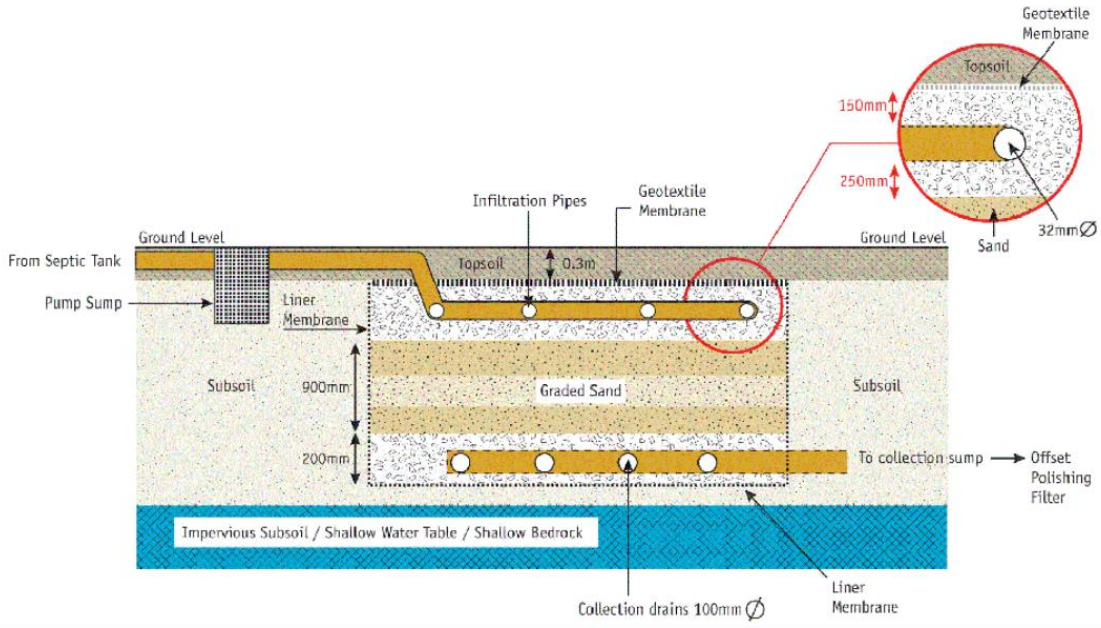


Figure 2.10. Schematic of intermittent covered Sand Filter system with underlying sand/subsoil polishing filter. If the system were to be open, the topsoil layer would be removed, exposing the gravel layer encasing the infiltration pipe (EPA., 2009).

Soil Treatment Unit (STU)

Prior to the installation of the STU a site assessment must be performed to confirm that soil conditions are suitable for the system. The final disposal of treated wastewater can be performed directly by treatment in the soil within the percolation area if there is required depth of unsaturated subsoil on-site. An unsaturated subsoil depth of 1.2 m depth is required for septic tank effluent to be disposed within the soil, if a secondary treatment system is installed this depth is reduced to 0.9 m. For a septic tank effluent the STU provides both secondary and tertiary treatment; for effluent from secondary treatment systems the STU acts as a polishing filter providing tertiary treatment.

Soakaways are a large deep pit, backfilled with stones, gravel, often construction debris where effluent is discharged into the subsoil. These systems are now considered unsuitable due to their design which results in accelerated clogging of the subsoil, which can result in failure such as ponding at the surface.

The most common STU configuration is a series of parallel percolation trenches which evenly distributes the septic tank effluent throughout the soil treatment area directly. The hydraulic distribution is usually by gravity unless contours of the land require pumping the effluent up to a higher level. Effluent is divided into of the requisite number of parallel percolation pipes using a distribution box. The number and length of percolation trenches is usually determined by the number of occupants within the domicile, but this can vary between different national legislative systems. The distribution of the wastewater is helped as wastewater seeps into a washed gravel matrix which surrounds the percolation pipe. These stones should be washed and removed of fine particles (Amador et al., 2019; EPA., 2012). These stones provide support for the perforated pipe and storage of excess wastewater. The Irish Code of Practice design for a STU percolation trench is a shallow, level excavation 850 mm in depth depending on the site, with a width of 500 mm (EPA, 2009) (**Figure 2.11**). The gravel matrix is lain onto the site to a depth of 20-30 mm and percolation pipe is lain on top at a slope of 1 in 200 with a maximum length of 18 m metres. To protect the percolation trench from surface damage 450 mm of gravel is placed on top of the percolation pipe. A minimum of 1,200 mm of unsaturated subsoil is required between the base of percolation trench and the water table or bedrock for septic tank effluent or 0.9 m of unsaturated subsoil for effluent that has already undergone secondary treatment upfront. A semipermeable geotextile is then applied over the trench to prevent the likelihood of backfill. If gravity flow is not possible, percolation trenches can also be fed by intermittent dosing of effluent from a sump via a low pressure pipe network.

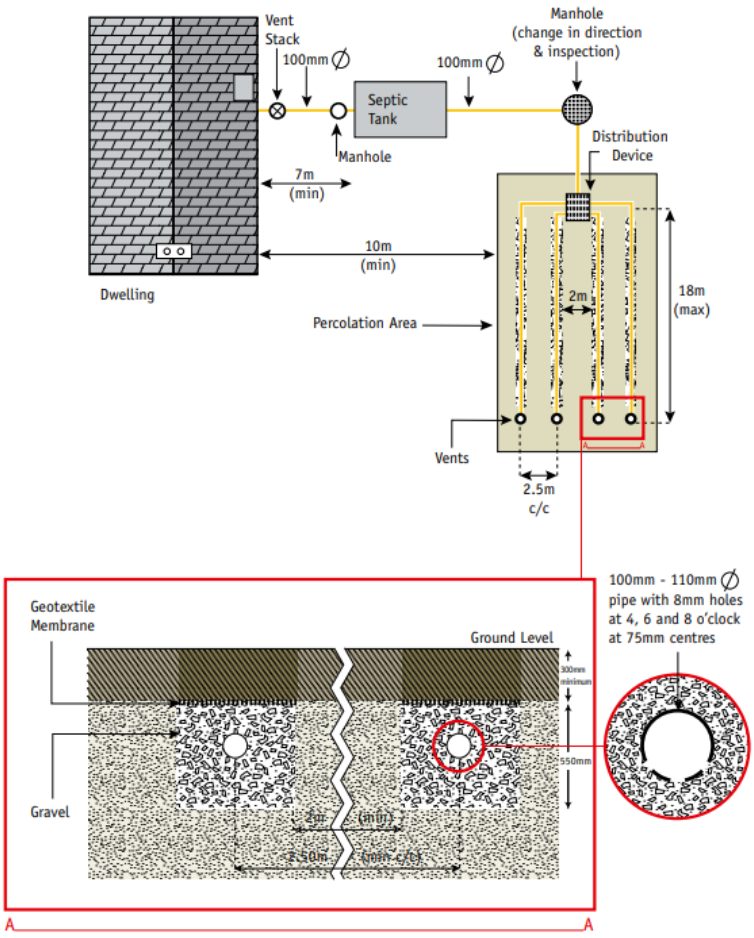


Figure 2.11. Schematic of a Soil Treatment Unit (EPA., 2021).

2.1.5 OWTS Management

Planning

The management of OTWs in Ireland are managed by Article 22(2)(c) of the Planning and Development Regulation 2006 requires all planning applications of households requiring on-site wastewater treatment systems to be designed, installed, operated, and maintained in line with the standards set in the EPA's Code of Practice (EPA, 2021). The Code of Practice requires that, as part of a planning application submitted to Local Authorities, a thorough site assessment is conducted. The assessment characterises the suitability of the site with regards to the optimum treatment option for the maximum number of residents, water supply, as well as the density of housing in the area, the presence of special areas of conservation (SACs) and special protection areas (SPAs) within 1 km of the site (EPA, 2021). A crucial component of the desk study is the assessment of the vulnerability of groundwater resources which lie below any possible system. The risk to groundwater resources is assessed by factors such as the presence and position of wells with significant supply rates (10 m^3 / day; public, group schemes or industrial abstraction), the presence of extremely vulnerable groundwater aquifers and the presence of karst features (**Table 2.1**).

Table 2.1. Response matrix for on-site treatment systems (source EPA., 2021).

Vulnerability rating	Source protection area		Resource protection area Aquifer category						
			Regionally important		Locally important			Poor aquifers	
	Inner (SI)	Outer (SO)	Rk	Rf/Rg	Lk	Lm/Lg	Ll	Pi	Pu
Extreme (X and E)	$R3^2$	$R3^1$	$R2^2$	$R2^2$	$R2^2$	$R2^1$	$R2^1$	$R2^1$	$R2^1$
High (H)	$R2^4$	$R2^3$	$R2^1$	R1	$R2^1$	R1	R1	R1	R1
Moderate (M)	$R2^4$	$R2^3$	R1	R1	R1	R1	R1	R1	R1
Low(L)	$R2^4$	R1	R1	R1	R1	R1	R1	R1	R1

[a]For public, group scheme or industrial water supply sources where protection zones have not been delineated, the arbitrary distances given in DoELG/EPA/GSI (1999a,b) of 300 m for the Inner Protection Area (SI) and 1,000 m for the Outer Protection Area (SO) should be used as a guide up-gradient of the source. Rk, Regionally Important Karstified Aquifers; Rf, Regionally Important Fissured Bedrock Aquifers; Rg, Regionally Important Extensive Sand and Gravel Aquifers; Lk, Locally Important Karstified Aquifers; Lg, Locally Important Sand/Gravel Aquifers; Lm, Locally Important – Bedrock Aquifer which is generally moderately productive; Ll, Locally Important – Bedrock Aquifer which is moderately productive in local zones; Pl, Poor – Bedrock Aquifer which is generally unproductive except for local zones; Pu, Poor – Bedrock Aquifer which is generally unproductive

Depending on the vulnerability and aquifer category, the installation of the OWTS may require additional conditions be met, as set out in Table 2.2. However, if there are already significant levels of groundwater contamination such as pathogenic organisms or high nitrate levels, greater restrictions or outright rejection of planning may be necessary. If nitrates are of concern a denitrifying secondary treatment system should be considered.

Following the desk study, physical visual inspection of the site is needed to assess the suitability of the site, identify, and qualify receptors at risk, and to provide Local Authorities with the required information as to effectively adjudicate their decision. Topography is an important factor where flatland or gently sloping convex hills are preferred. Sites may be unsuitable (or difficult to engineer) if they are in a depression or on concave slopes. Slopes must be considered in the design of the percolation area such as the position of the pipes as to ensure wastewater will remain in the soil and travel vertically for effective treatment. The gradient of slopes should be particularly assessed if the system is to be constructed adjacent to surface waters, which require minimum set-back distances regardless of gradient. The concentration of OWTS in an area must be considered, especially in areas where aquifers have an Extreme vulnerability rating and where there are pre-existing issues of surface water or groundwater contamination. This can be determined by the presence of contamination of wells in the area for example. If the site is free of contamination, minimum distances from the wells must be adhered to in the system design to reduce the risk to the receptor to an acceptable level. Although karst features are preliminary surveyed in the desk study, physical inspections of the subsoil should observe for bedrock or karst features. Subsoil characterisation then needs to be determined through a trial hole, the excavation of a large cavity in the soil 6 m x 1 m with a depth of either 2.1 m in sites over poor aquifers or 3 m in areas with an $R2^2$response of worse in terms of the response matrix (Table 2.1). The trial hole allows for the assessment of the soil structure and its likely effects on percolation capacity through the depth profile (EPA., 2021). The colour of the soil can be indicative of saturated conditions with darker colours indicating permanent

saturation due to a high water table during the winter. Additional onsite falling head percolation tests are conducted to determine the effective permeability of the soil. These tests require excavating a 300 mm x 300 mm x 400 mm deep hole from the surface at the proposed depth of effluent infiltration. The drop in head is measured over time and a PT-value (previously known as a *T value*) for the subsoil is calculated. Sites that have low permeability subsoils may need mounded soil treatment systems to provide additional depth, however, if subsoil permeability is very low (T>90) they are considered unsuitable for an OWTS (**Table 2.2**) according to the EPA Code of Practice in Ireland. Septic tank systems can be considered both diffuse sources and small point sources. In models investigating clusters of septic tank systems found in high densities (> 19/km^2) in low permeable sites were shown to constitute 22% P and 13% of N annual emissions (Gill et Mockler., 2016).

Table 2.2. The acceptability and conditionality of the installation of OWTS required for different risk response category as described in Table 2.1.

Response category	**OWTS acceptability**	**Required conditions**
R3[2]	Not generally acceptable, unless:	A secondary treatment system is installed, with a minimum thickness of 0.9 m unsaturated soil/subsoil with P/T-values from 3 to 75 (in addition to the polishing filter which should be a minimum depth of 0.9 m), beneath the invert of the polishing filter (i.e. 1.8 m in total for a soil polishing filter). and subject to the following conditions: • The authority should be satisfied that, on the evidence of the groundwater quality of the source and the number of existing houses, the accumulation of significant nitrate and/or microbiological contamination is unlikely. • No on-site treatment system should be Code of Practice: Wastewater Treatment and Disposal Systems Serving Single Houses (p.e. ≤ 10) Environmental Protection Agency 61 located within 60 m of a public, group scheme or industrial water supply source. • A management and maintenance agreement are completed with the systems supplier.
R3[1]	Not generally acceptable, unless:	A septic tank system is installed with a minimum thickness of 2 m unsaturated soil/subsoil beneath the invert of the percolation trench (i.e. an increase of 0.8 m from the requirements in Section 6) **or** A secondary treatment system, is installed, with a minimum thickness of 0.3 m unsaturated soil/subsoil with P/T-values from 3 to 75 (in addition to the polishing filter which should be a minimum depth of 0.9 m), beneath the invert of the polishing filter (i.e. 1.2 m in total for a soil polishing filter). and subject to the following conditions: • The authority should be satisfied that, on the evidence of the groundwater quality of the source and the number of

		existing houses, the accumulation of significant nitrate and/or microbiological contamination is unlikely. • No on-site treatment system should be located within 60 m of a public, group scheme or industrial water supply source. • A management and maintenance agreement is completed with the systems supplier.
R2[4]	Yes, on condition;	**1.** There is a minimum thickness of 2 m unsaturated soil/subsoil beneath the invert of the percolation trench of a septic tank system. **Or** **1.** a secondary treatment system is installed, with a minimum thickness of 0.3 m unsaturated soil/subsoil with P/T-values from 3 to 75 (in addition to the polishing filter which should be a minimum depth of 0.9 m), beneath the invert of the polishing filter (i.e. 1.2 m in total for a soil polishing filter). **2.** The authority should be satisfied that, on the evidence of the groundwater quality of the source and the number of existing houses, the accumulation of significant nitrate and/or microbiological contamination is unlikely. **3.** No on-site treatment system should be located within 60 m of a public, group scheme or industrial water supply source.
R2[3]	Yes, on condition;	Conditions: **1** & **2**
R2[2]	Yes, on condition;	Conditions: **1**
R2[1]	Yes, on condition;	Acceptable subject to normal good practice. Where domestic water supplies are located nearby, particular attention should be given to the depth of subsoil over bedrock such that the minimum depths required are met and that the likelihood of microbial pollution is minimised.
R1	Yes	subject to normal good practice

Inspection

In 2009 Ireland received a firm rebuke by the European Court of Justice (ECJ) for inadequate regulation of rural domestic wastewater infrastructure required by the 1975 EU Waste Framework Directive (75/442/EEC). In response Ireland introduced the Waste Services Act in 2012, which tasked local authorities with inspecting OTWS systems within their jurisdiction and the Irish EPA developed the National Inspection Plan to ensure compliance (Hynds et al., 2018). Since the commencement of inspections in 2012 up until 2022 there have only been 10,000 inspections of the approximately 489,000 systems within Ireland (Hynds et al., 2018; EPA, 2023). Currently, about 9% of water bodies in Ireland have been identified as experiencing substantial stress due to failing OWTS (EPA, 2022). The Water Services Act (2007) required the Environmental Protection Agency to develop a National Inspection Plan (NIP) to coordinate inspections. This plan is now its fourth iteration (2022-2026) with inspections set to increase from 1,000 to 1,200 per year. The EPA applied its Source Load Apportionment Model (SLAM) which incorporates the SANICOSE model for

pollution from OWTS (Gill and Mockler, 2016) to produce Pollution Impact Potential maps for drainage characteristics, to devise a risk profile for water bodies categorised into zones, as shown in Figure 2.23 (EPA, 2023). Zone 1 is characterised as areas which do not have public sewage network, and which river managers have identified OWTS as a significant pressure on water bodies in their river basin management plans. These zones have been allocated 2,400 inspections over this current cycle of the NIP. Zone 2, which will receive equal number of inspections over this cycle of the NIP within this zone focused on household wells particularly those with poor subsoil conditions and karst features (**Figure 2.12**). Donegal and Galway have the greatest number of water bodies at risk from OWTS, whereas household wells are deemed to be most at risk in counties Cork and Clare. Further sub prioritisation within the zones is applied with greater priority for systems at greatest proximity to surface bodies that are clustered or have older soakaway designs.

The inspection of a system has several parameters that must be met for the system to be given a passing grade. All systems must be registered with "*Protect Our Waters"*, a government-run registration portal that informs local authorities. Inspectors will determine if the system is leaking - this is mostly done at the septic tank by observing for cracks or groundwater intrusion. The system is checked for ponding primarily at percolation areas, which can occur at impermeable soils or due to clogging in older systems (>15 years). The age of the system is of importance (>20 years), as soakaway systems designs are considered significant sources of pollution with moderate to high permeable soakaway sites being critical risks to groundwater supplies (Keegan et al., 2014). These systems pose two primary risks depending on the permeability of the soils: Low permeability sites may expose the public to effluent as the preferential flow of the upper subsoil produces surface run-off (Keegan et al., 2014). In addition, rainwater from the roof of the house is assessed to confirm it is not directed toward treatment system which would reduce the hydraulic retention time of the septic tank and increase the hydraulic loading rate for percolation area. Unless licenced (which is currently not done by any local authority in Ireland) the treatment system cannot discharge directly into surface waters. Desludging is monitored as to ensure adequate desludging rates (~ every 3 years) to ensure that hydraulic retention time of septic tanks and reduce suspended solids entering the percolation area which could increase the clogging rate (Richards et al., 2015). If the system fails the homeowner will be notified within 21 days and informed on the reason for failure, the remedial works required and the deadline for their competition. Inspections have increased in the latest cycle of the plan and effective prioritisation has focused authorities on critical areas which are needed due to the limited resources of these institutions. It must be noted that the capacity to inspect existing systems remains drastically low when compared to the number of existing systems; it is therefore crucial that the planning component is significantly robust as to effectively manage new builds and ensure the best practice in site assessment and the construction of systems providing effective treatment to minimise the risks to both public health and the environment.

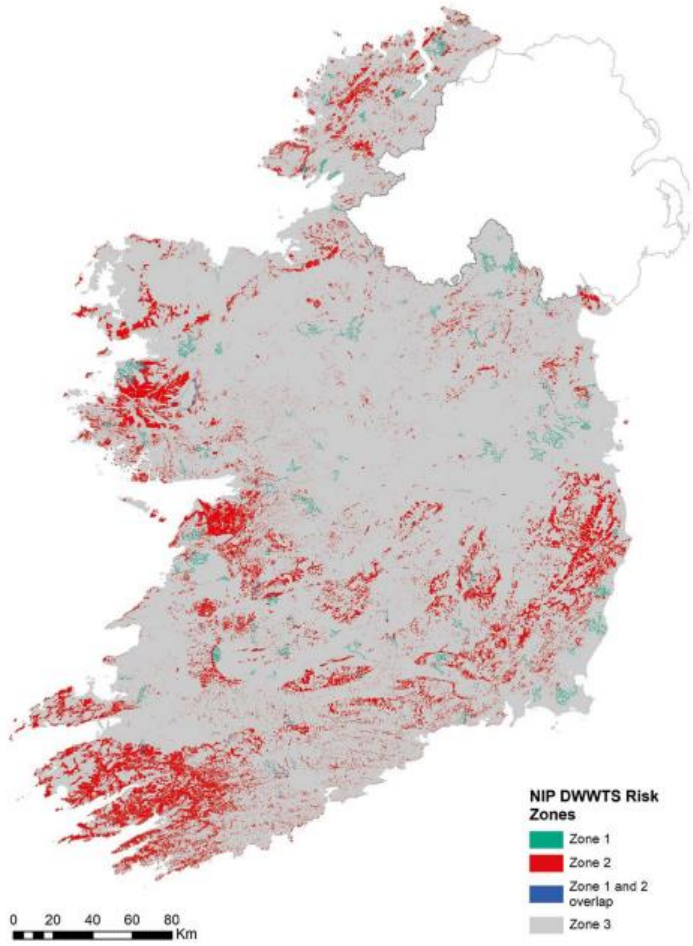


Figure 2.12. National distribution of risk zone (source: EPA., 2023)

2.2 Domestic Wastewater Contamination Pathways

2.2.1 Domestic Wastewater Characteristics

Domestic wastewater is the combined effluent from water using activities within a household. Unlike industrial effluents which can vary widely with respect to pH, temperature, organics, and pollutants (and may therefore require specialised treatment prior to being released into the environment or a central sewer system), domestic wastewater from households is relatively simple in its composition. Domestic wastewater can be subdivided based on its source within the house. Blackwater consisting of faeces and urine originates from the toilet and is most concentrated in solids, organics, and nutrients. Greywater originates from washing facilities (washing machines, dishwashers, baths, showers, washbasins, kitchen sinks) is normally a less concentrated and complex solution. The EPA code of practice states that an average effluent production of 150 litres per capita per day (Lcd) is required for tailoring the design on an on-site system. This volume of effluent production is estimated based on household water consumption statistics. However, long-term monitoring studies of effluent production in on-site systems show an average production of 101.3 Lcd, with values ranging from 60.3 – 123 Lcd (Dubber and Gill.,2014). The variation in the

production of effluent appears to be attributable to human activity, with many individuals spending only a portion of their time within their domicile.

Domestic Wastewater Composition

Domestic wastewater contains a variety of chemical and microbiological contaminants. It is crucial that prior to it reaching the aquatic environment that contaminants are removed to a sufficient standard as not to pose a threat to public health or local ecosystems. Key concerns are nutrients such as nitrogen and phosphorous which can enter surface and transitional waters and during summer periods can result in blooms of algae and Cyanobacteria. These blooms can be toxic and detrimental to ecosystems as the over production of organic natter sinks to the bed of lakes, promoting oxygen depletion as it is decomposed by heterotrophic microorganisms. Equally, any pathogenic organisms in domestic wastewater can cause outbreaks of gastrointestinal disease in recreational waters as well as contamination of private wells in rural settings. The average composition of these primary contaminants is surmised (**Table 2.3**), which lists values for raw domestic wastewater as described in the EPA Code of Practice (2021). Effluent concentrations are noted for different levels of treatment such as following primary sedimentation in a septic tank and secondary treatment effluent (Dubber et Gill., 2014).

Table 2.3. Typical range of raw domestic influent, primary and secondary effluent values as reported by EPA 2021; Dubber et Gill., 2014.

	Raw domestic wastewater	**Septic tank effluent (n = 16)**	**Secondary treated effluent (n = 12)**
COD [mg/L]	956	580 ± 297	160 ± 72
BOD [mg/L]	318	365 ± 198	40 ± 22
Total N [mg/L]	NA	128.3 ± 66.2	69.1 ± 42.8
NH4-N [mg/L]	70	79.1 ± 47.8	15.7 ± 12.3
NO3-N [mg/L]	NA	1.9 ± 1.7	26.5 ± 20.8
Org-N [mg/L]	NA	64.3 ± 41.7	18.7 ± 14.6
Ortho-P [mg/L]	18	18.2 ± 7.7	12.6 ± 10.4
Chloride [mg/L]	NA	134 ± 82.9	99 ± 63
Total coliforms [MPN/100mL] *	4.1×10^7	$1.78 \times 10^7 \pm 2.7 \times 10^6$	$3.44 \times 105 \pm 1.9 \times 105$
E. coli [MPN/100mL] *	7.1×10^5	$2.22 \times 10^6 \pm 2.4 \times 10^5$	$8.43 \times 10^4 \pm 4.7 \times 10^4$

Organic Matter

Organic matter in domestic wastewater was classically characterised by chemical oxygen demand (COD), total organic carbon (TOC) and biological oxygen demand (BOD). BOD_5 is a means of assessing the biodegradable organic fraction concentration measuring the consumption of dissolved oxygen when effluent is incubated at 20 ^{O}C degrees for a period of 5 days. COD is a quicker alterative that reduces the process to 2-3 hours, although provides a measure of all the organics in a sample (biodegradable or not) through the addition of oxidation agent potassium dichromate and addition of sulphuric acid. A photometer is used to provide the concentration of the final reaction. Alternatively, TOC is performed by thermo-catalytic oxidation and again provides a value of all the

carbon oxidised. It is considered a suitable replacement for BOD_5 and COD with analysis showing linear relationships with previous methods (Dubber and Gray., 2014). The application of TOC has added benefit that automation of the process is possible, and it eliminates the production of harmful chemical by-products produced in COD analysis. Organics in domestic wastewater are mainly composed of proteins, sugars, and lipids. Organic matter can be divided by particulate and dissolved fractions (Huang et al., 2010). Organics have been characterised by particle size distribution of substances settleable (e.g. >100 µm), supra colloidal (e.g., 1-100 µm), colloidal (e.g., 0.08-1 µm) and soluble (e.g. < 0.08) (Huang et al., 2010).

Nitrogen

Nitrogen is a critical component of proteins and nucleic acids abundant in all living cells. Nitrogen gas, although abundant in the atmosphere (N_2) accounting for ~78% of its composition, is generally unavailable for organisms due to its stable triple bond (Holmes et al., 2019). Most of the nitrogen fixing from the atmosphere is performed by groups within the phyla prokaryotes which convert N_2 to ammonia (NH_3). Lighting can also form a source of nitrogen fixation with approximately 3 – 10 Tg nitrogen fixed each year by converting N_2 to NO_2 which then combines with moisture to form nitric acid (HNO_3) (Fields., 2004). Reactive HNO_3 is transported through snow, rain or hail landing on the ground and is utilised by organisms. Nitrogen is transformed into a variety of forms: ammonia-oxidising bacteria (AOB) convert to NH_3 to nitrite (NO_2^-), followed by oxidation of NO_2^- to nitrate (NO_3^-) carried out by nitrite-oxidising bacteria (NOB), This nitrogen in nitrate form can then be transformed back to molecular (N_2) in the atmosphere by denitrifying bacteria or NO_2^- and NH_3 can be converted by anaerobic ammonium oxidation (annamox) to N_2 (Holmes et al., 2019). Excess levels of reactive nitrogen can be of major consequence to environmental health. Human activity has significantly affected the nitrogen cycle with estimates of ~210 Tg of reactive nitrogen being produced per year at the current rate compared to 7 Tg N per year 2000 years ago (Galloway et al., 2013). It is estimated that humans excrete about 23 Tg of reactive nitrogen per year, with more than half of this waste released directly into the environment (Smil, 1999). The primary treatment of wastewater in typical on-site systems which removes little reactive nitrogen (Homes et al., 2019; Knappe et al., 2020), whereas secondary treatment tends to convert the form of nitrogen into soluble nitrate. The composition of nitrogen in domestic raw wastewater is typically made up of 40% organic -N and 60% as ammonium-N (NH_4-N) which is preferred by waterborne, and soil organisms as its assimilation does not require a redox step (Lusk et al., 2017).

Phosphorous

Phosphorous (P) is an essential chemical element for life on earth with no existing substitutes and is considered to be a non-renewable resource at scale (Tarayre et al., 2016). Phosphate is primarily found and harvested from mineral deposits containing phosphate rock or phosphorite (Carrillo et al., 2020). Access to P-based fertilisers remains a crucial geopolitical challenge. However, concerns regarding phosphorous supplies run ironically in parallel to the proliferation of phosphorous in

surface and coastal waters. The largest P flows to aquatic ecosystems are agricultural run-off, soil erosion and point sources. In Europe there have been large reductions of phosphorous from centralised wastewater treatment plants with these decreases reflecting the targeted improvement in wastewater treatment under the Urban Wastewater Directive (91/271/EEC) and the reduction of phosphorous in detergents (EEA., 2012). Organic phosphorous is represented by organic phosphoric acid esters, sugars phosphates, phospholipids, nucleus acids and inositol (Tarayre et al., 2016). Phosphorous in the environment is immobilised by its assimilation by plants and microorganisms. Inorganic P is presented by polyphosphates which form tetrahedral structure, which are usually produced by industrial processes (for example, in detergents). pH is critical to the P-cycle with a low pH resulting in weathering which results in the solubilisation of phosphate as orthophosphoric acid which is readily available for biological absorption (Vymazal et al., 2008). Within wastewater, P is found in two predominant forms as soluble orthophosphate (PO_4^{3-}) or as organic phosphorous, dissolved organic phosphorous from wastewater is of concern as a source of limiting nutrients for eutrophication (Tarayre et al., 2016).

2.2.2 Faecal Indicator Bacteria and Microbial Source Tracking

In untreated wastewater total coliforms, *Escherichia coli* and *Enterococci* have been selected as good indicators of potential risk to human health from faecal sourced pathogens because their abundance in mammalian faeces (Devane et al., 2020). Coliforms are defined as the lactose fermenting, gram-negative, *Enterobacteriaceae,* which includes *Escherichia coli (E. coli), Enterobacter, Klebsiella,* and *Citrobacter* (Ishii et al., 2008). *E. coli* is a rod-shaped, gram-negative, gammaproetobacterium in the family *Enterobacteriaceae*. There are over one million *E. coli* cells present within one gram of faeces, and its dissemination into the environment is considered through deposition of mammalian or avian faeces (Savageau., 1983). However, a growing major concern of these FIBs is their lack of specificity regarding host sources, as they are abundant in mammalian (and in some cases avian) reservoirs, which is particularly challenging if trying to track diffuse sources of faecal pollution in rural settings. Previously it was considered that *E. coli* was believed to exhibit a low capacity for survival in the environment and secondary habitats such as water, sediment, and soil (Winfield and Groisman., 2003). The stresses *E. coli* faces in the environment include low and high temperatures, limited moisture, variation in soil texture, low availability of organics, salinity, solar radiation and predation by other microorganisms (Ishii et al., 2008). These FIB were first reported to survive in the environments in tropical climates where temperatures allowed for the growth of organisms (Jimenez et al., 1989). However, further studies showed that this phenomenon to occur in both subtropical and temperate climates (Solo-Gabriele et al., 2000; Desmarais et al., 2002; Whitman et al., 2003; Ishii et al., 2006). These findings indicated that these organisms had evolved to survive under different temperatures and nutrient regimes and were termed as “naturalised” *E. coli* and enterococci. Advances in molecular techniques have allowed for the whole genome analysis of *E. coli* which has shown that these naturalised FIB may have diverged

from enteric bacteria millions of years ago. These "true" environmental strains of enterococcus and *E. coli* were only recognised due to novel sequencing technology (Devane et al., 2020). The concern for public health officials is that these environmental strains have also been known to possess the enzyme activity of ß-D-glucuronidase which is used in the identification and quantification of faecal sourced *E. coli.* Therefore, the presence of environmental strains may confound water monitoring results, overestimating the abundance of FIB (Deng et al., 2014). The abundance of *E. coli* in raw wastewater system has been well defined with concentrations of approximately 1 x 10^7 CFU / 100 ml in raw influent. Wastewater treatment has shown $\log_{10}$ removal rates of 2.72 ±0.31, 0.42 ± 0.27, 1.60 ± 0.36, 4.74 ± 0.25 for activated sludge treatment, chlorination, sand filtration and the totality of treatment, respectively (Hata et al., 2013; Dubber and Gill., 2014).

Pepper Mottle Mild Virus

Pepper Mottle Mild Virus (PMMoV) is an RNA virus belonging to the genus *Tobamovirus* in the family *Virgoviridae* which was first described when identified from peppers in Italy in 1984 (Katijima et al., 2018). The virion is non-enveloped rod-shaped which encapsidates a single copy of the RNA single stranded RNA genome. The virus is pathogenic to a wide variety of peppers and is considered a major pathogen to pepper species globally (Katijima et al., 2018). PMMoV was first identified in human stools in 2006 when samples were taken from two hosts from different continents, which were enumerated using quantitative polymerase chain reaction (QPCR) to reveal viral loads of $10^6 – 10^9$ per gram of dry faeces (Zhang et al., 2006). PMMoV has been found to persist in food products even if they have been processed: examples include chilli sauce, chilli powder and Tabasco® sauce which was found to have the highest concentrations of 10^7 copies per ml (Zhang et al., 2006; Katijima et al., 2018). PMMoV was first reported in wastewater at concentrations ranging from 10^8 to 10^9 gene copies per 100 ml in raw sewage and 10^6 to 10^8 gene copies per 100 ml in treated effluent (Rosario et al., 2009). Environmental sampling of coastal waters adjacent to outflows from WWTP found detectable concentration of PMMoV in four out of six samples with concentrations of 10^4 to 10^6 gene copies per ml, with the virus being detectable for up to 7 days in temperatures ranging from 31 to 33 ^{O}C (Rosario et al., 2009). Overall, the specificity of PMMoV in anthropogenically impacted environments for domestic wastewater samples has remained 100% in wastewater based epidemiological studies (Rosario et al., 2009; Hamza et al., 2011; Symonds et al., 2017; Gyawali et al., 2019). At present PMMoV shows greater human specificity than the more conventional FIBs discussed above, but its potential as a suitable fingerprinting tracer for human effluent is dependent on diet. A variety of animals have been sampled including horses, sheep, ducks, geese, cows, coyotes, raccoons, and pigs which showed no detectable concentrations of PMMoV (Rosario et al., 2009; Hamza et al., 2011). PMMoV was found to be most prevalent in dogs, chicken and seagull samples with concentrations ranging from 1 x $10^{4.}$ to 1 x 10^5 gene copies / g, 5.24 x 10^2 to 2.2 x 10^4 copies / mg and 5.84 x 10^2 5 to 9.55 x 10^2, respectively (Rosario et al., 2009; Hamza et al., 2011; Gyawali et al., 2019). However, the source of the virus in these animals is unclear, but authors hypothesised that the chicken faeces were sourced

from a household and being kept as pets more likely exposed them to the virus through their diet (Hamze et al., 2011).

Removal rates from conventional wastewater treatment have noted a maximum removal of 3.7 $\log_{10}$ gene copies / 100 ml in an activated sludge plant in Germany (Hamza et al., 2011) or 2.7 $\log_{10}$ gene copies / 100 ml in by four stages continuous flow suspended growth with alternating anoxic/aerobic/anoxic/aerobic stages (Bardenpo) (Schmitz et al., 2016). No other treatment technology reached or exceeded more than 1 $\log_{10}$ reduction (Katijima et al., 2014; Koruda et al., 2015; Symonds et al., 2014). Hence, PMMoV appears to be a relatively conservative tracer for human wastewater with greater persistence in the environment perhaps due to capsid structure being more robust (Katijima et al., 2018).

2.2.3 Contaminant flow through the unsaturated zone

Biomat development

As discussed previously, on-site wastewater systems are often a combination of a septic tank with the effluent flowing into a STU. Pollutant transport can take several pathways prior to reaching receptors of concern, such as ground water or surface waters. These pathways have been conceptualised into three main possibilities as described in **Figure 2.13**: (1) inadequate percolation results in pollutants ponding and moving directly by surface flow overland to surface waters (2) near-surface transport through unsaturated subsoil toward surface waters (3) and pollutants travelling directly to groundwater through the unsaturated subsoil. These pathways can be modelled using the SANICOSE model which has been incorporated into the EPA's Source Load Apportionment Model (SLAM) for catchment in Ireland (Gill et al., 2016).

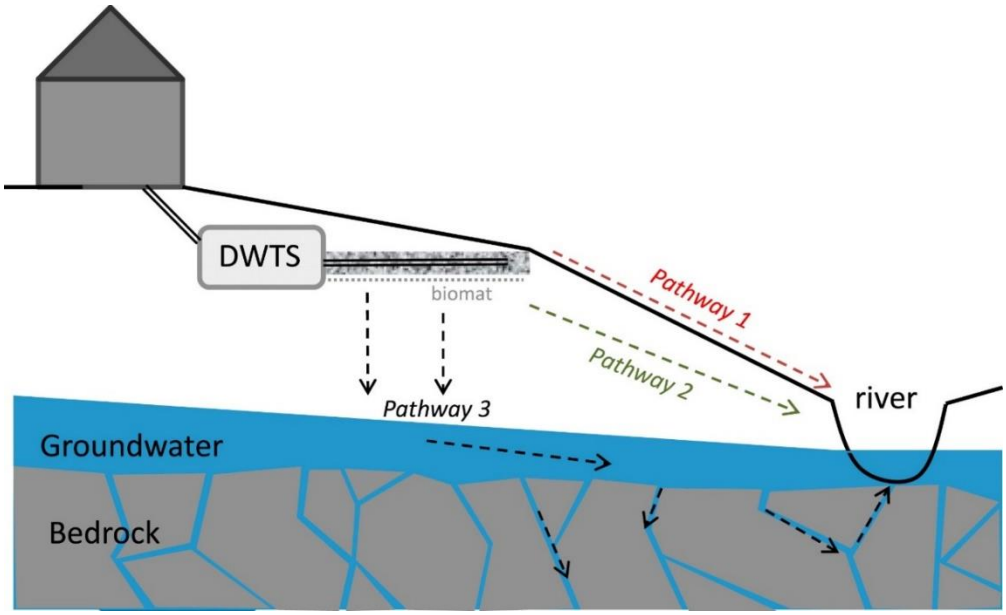


Figure 2.13. Schematic cross-section of OWTS indicating theoretical pollutant pathways (source: Gill et al., 2016).

The STU for an OWTS promotes percolation of effluent through soil to achieve purification prior to recharge into the underlying ground water (Van Cuyk et al., 2001). If these systems are constructed in accordance with the Codes of Practice of national authorities, these systems should achieve high purification performance through pathway 3. This is achieved if there is a sufficient hydraulic retention time for the percolating effluent within the soil matrix through the unsaturated subsoil (Van Cuyk et al., 2001). The effluent first passes through the infiltrative zone where the gravel of the percolation trench bases meets the subsoil base of the STU wherein a complex series of biogeochemical processes occur which are crucial in the removal of contaminants (Gill et al., 2009). Over time a clogging zone develops due to accumulation of suspended solids and organic matter contained in the effluent. This physical clogging process is described as phase 1 of the biomat development (Beal et al., 2006). This phase accounts for the physical filtration of undissolved organics found within the effluent. Over time the abundance of organics produces a microbial biomat over the surface area of the infiltrative surface which can include the base and side walls of the percolation trench (Van Cuyk et al., 2001). It has been estimated Ireland that a biomat receiving primary effluent (i.e. septic tank effluent) would grow at a rate between 3.13 - 4.6 cm / day, which was determined using soil moisture sensors monitoring the changes of the volumetric water content of the soil and the rate of growth of the biomat within the STU (Knappe et al., 2020). Using these biomat growth rates, it was estimated that the 18 m long trench could be entirely covered by the biomat within 12 – 21 months (Knappe et al., 2020). The biomat is composed of suspended solids, organics, and extracellular polysaccharides, which further reduce permeability of the trench (Van Cuyk et al., 2001; Beal et al., 2006). As the biomat develops laterally across the surface area down the trench, it promotes better distribution of the effluent by blocking macropores within the subsoil (Gill et al., 2007). Over time the transport of effluent through the system will reduce further, as effluent goes from transiently to continuously ponding on the infiltrative surface (Beal et al., 2006). The level of pre-treatment has significant effect on the development of the biomat as greater removal of organics results in reduced growth of the biomat (Gill et al., 2009; Knappe et al., 2020). Hence, STUs receiving secondary treated wastewater result in shorter and thinner biomats. This has been shown to result in reduced pollutant removal capacity of nutrients such as nitrates and pathogens down through the underlying subsoils die to the higher hydraulic loading leading to higher levels of soil moisture (saturation) and therefore greater transport velocities on through the STU (Gill et al., 2007; Gill et al., 2009; O'Luanaigh et al., 2012; Knappe et al., 2020).

Nutrient Removal through the STU

As pollutants infiltrate through the biomat zone effluent flows with advection and diffusive transport with purification taking place through processes of straining, filtration, sorption, chemical reactions, biotransformation, die–off and predation (Van Cuyk et al., 2001). In the case of nutrient removal, the primary removal mechanism of particulate nutrients contained within suspended solids is via filtration through the biomat. For the dissolved fractions of nitrogen and phosphorous subsequent

abiotic processes such as adsorption removes nutrients in the soil (Patel et al., 2018). Phase one of adsorption is advection transport which is the movement of solute from the bulk solutions onto an immobile film layer by advection flow or axial dispersion or diffusion. Phase 2 is the film transfer through the penetration and attachment of solute particles in the immobile water film layer. Phase 3 is mass transfer which is the attachment of solute particle onto the surface of the absorbent (**Figure 2.14**). The last phase is the intraparticle diffusion which is the movement of solute into the pores of adsorbent and intraparticle diffusion (Patel et al., 2018).

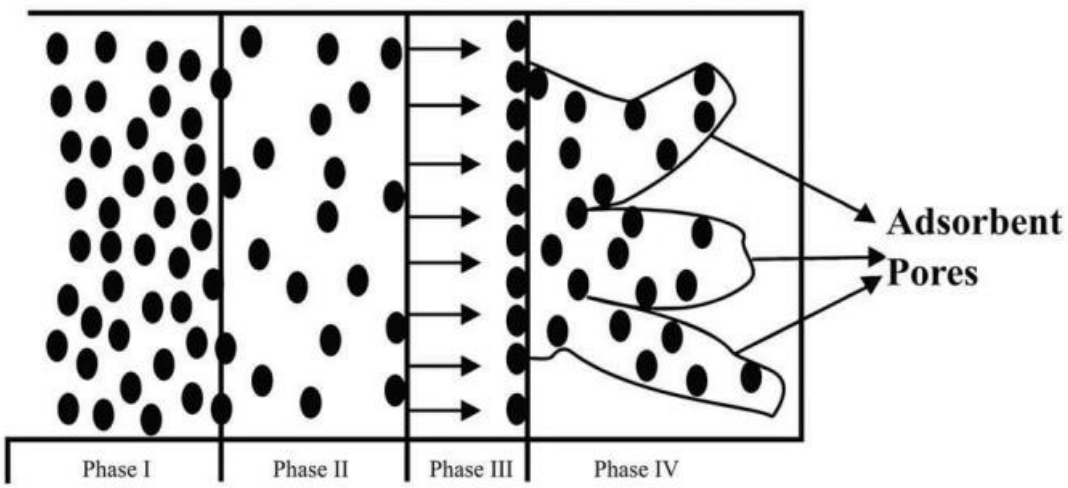


Figure 2.14. Different phases of adsorption (Patel et al., 2018).

The main attenuation mechanisms for organic chemicals in soil treatment systems are biotransformation (which would be mineralisation in the case of organic nitrogen and phosphorous) and sorption processes within the first few mm of depth of the system (Bünemann et al., 2007; Gharoon et al., 2021). This process is facilitated by fungi, aerobic and anaerobic bacteria which are affected by climate, soil properties and soil organic content (Hamidi et al., 2021). For nitrogen species mineralisation occurs when NH_4^+ is released by microbial enzymatic action from enzymes such as hydrolases, oxidases, deaminases, and lyases (Murphy et al., 2003). The process of N mineralisation produces energy but also provides the macronutrient building blocks of microbial cells through producing bioavailable C and N (Murphy et al., 2003). Phosphorous mineralisation can be separated into, (1) biochemical processes that are facilitated by phosphatases enzymes and driven by the biological demand for P, and (2) biological mineralisation that results from the oxidation of carbon by soil organisms to gain energy and secondly releasing nutrients contained in the oxidised compound (Bünemann et al., 2007). NH_4^+ is either assimilated to microbial cells or converted to NO_3- which is also assimilated by these organisms. This process is considered the immobilisation of the nitrogen species which is now considered incorporated in the microbial cell (Murphy et al., 2003). The assimilation of nitrogen species is closely linked to the availability of organic carbon and a strong inverse relationship between DOC and nitrate concentrations suggests that nitrate is depleted where DOC supplies are high, providing evidence that some portion of the

DOC losses in groundwater are due to microbial transformations, including denitrification (Murphy et al., 2003; Pabich and Hemond., 2001). Primary productivity and growth rates in soil ecosystems are very reliant on P which is why bacteria have evolved many mechanisms in the biotic immobilisation of P (Tappia-Torres et al., 2016). In its oxidised form P is found in the form of oxidised esters (H_2PO_4-) which are then metabolised by exoenzymes called phosphatases. However, genomic studies have shown soil and sediment bacteria have the capacity to acquire phosphorous in both an oxidised and reduced state (Tappia-Torres et al., 2016).

Microbial Contaminant Removal through the STU

The two mechanisms responsible for the immobilisation of pathogens in wastewater moving through soil or other porous media is straining and adsorption. Straining involves the physical mechanism of blocking movement through soil pores which are of smaller diameter than the bacteria (Stevik et al., 2004). The particle distribution of the soil is therefore an important factor for its filtering capacity (i.e. silt, clay and fine sand etc.) as the extent of bacteria retained is proportional to the size of the filter media particles (Stevik et al., 2004). It has been stated that soil straining occurs when suspended particles which are larger 0.2 times the diameter of the constituents of the porous media (Bouwer et al., 1984). In unsaturated conditions much of the transport takes place in the smaller pores (with the larger pores draining first) thereby promoting better effective straining under lower soil moisture conditions. However, it is important to note that macropores can induce significant movement of pathogens through the subsoil (Stewart et al., 1981; Stevik et al., 2003). The shape and size of the pathogen may influence its transport through the soil, with organisms larger than 1 µm in size, and rod-shaped were more likely to be filtered out by soil pores (Bouma et al., 1974; Gannon et al., 1991). In STU subsoils it was noted that the greatest removal of *E. coli* occurred at a depth of 0.35 m underneath the infiltrative surface of the trench, while bacteriophages saw the greatest reduction 0.95 m underneath the infiltrative surface, with greater depth of unsaturated subsoil required for their complete removal (O'Luanaigh et al., 2012). The hydraulic loading rate of effluent over the soil must be considered and an even distribution of the effluent is necessary otherwise high point flow rates can increase soil moisture levels causing greater transport of organisms through large pores (Stevik et al., 2003). This is why the development of a low permeability microbial biomat is essential for the proper functioning of an STU. Reduced biomat growth in STUs receiving secondary treated effluent resulted in a greater level of faecal contamination even with a lower bacterial influent load when compared to the primary effluent but due to the higher hydraulic loading rate leading to deeper depths of penetration of *E. coli* possibly contaminating the water table (O' Luanaigh et al., 2012). Microbial adsorption occurs in two processes within the soil: (1) reversible adsorption which is a weak attraction of organisms to the soil matrix surface by electrostatic forces {note, this weak attraction can allow for an organism to dislodge and return to the water phase (Escher et al., 1990; Mozes et al ., 1987)} and (2) adhesion which is considered the permanent connection between the organism and the soil which requires sufficient contact time, large amounts

of energy and for the organism to produce polymers creating bridges between the soil and cell surface (Stevik et al., 2003). Adsorption is also affected by the ionic strength of effluent, as the charged surface of the bacteria and soil will be neutralised by the accumulation of oppositely charged ions. The combined effect of ionic strength and pH of the effluent were most significant at low ionic strength, and adherence to the soil matrix increased at a low pH of 3.9 (Goldschmith et al., 1973). Clay minerals and metal oxides in soils have been shown to be effective adsorbents of virus particles, particularly at lower pH levels where the positive charges of the clay minerals and metal oxides increase enhancing absorption (Yoshimoto et al., 2012). Absorbent capacity can fluctuate temporally depending on the pH of the effluent and result in reversible absorption releasing viruses and bacteria into solution (Yoshimoto et al., 2012). Biotic factors such as natural die-off and predation are enhanced by the adhesion of viruses and bacteria to the soil matrix. Conditions which should advance bacterial die off are high temperature, low pH , and low organic content (Wang et al., 2021). Salmonella and Yersinia species have a higher survival rate than *E. coli;* however, *E. coli* dies-off slower than Shigella and *Camplylobacter jejuni*. Often microorganisms which can produce spores (such as protozoa) or are resistant to antibiotics may exhibit a competitive edge under the stress of a new ecosystem (Stevik et al., 2003). Adenonovirus are one of the most numerous virions in sewage and are considered the primary cause of gastroenteritis in children (Davies et al., 2006). The inactivation rate of adenovirus in soils varies from 0.01 – 0.047 $\log_{10}$ removal rate per day with soil temperatures ranging from 12 ^{O}C to 22 ^{O}C, with reduced removal rates at lower temperatures (Regnery et al., 2017). Ultimately, the zone receiving primary effluent develops a longer and deeper biomat which enhance the inactivation of pathogens. Enhanced microbial activity linked to greater antagonism and competition for nutrients, will further reduce the survival rates of enteric bacteria, the reduced hydraulic loading rate and pore space may enhance the adhesion of viruses (Stevik et al., 2003; Regnery et al., 2017).

2.3 Microbial Ecology functionality and community development in STUs

. There is an enormous diversity of soil bacterial communities: a single gram of soil may contain 1 × 10^3 to 1 × 10^6 unique "species" of bacteria (Fierer et al., 2007). Considerable work has been conducted to determine the role of such microbial ecosystems in altering or maintaining soil function and equally what impacts the microbiome infers on the properties of the soil. In the soil environment, these organisms are often divided into generalist and specialist classes, with generalists capable of tolerating broader environmental tolerances as compared to specialists which are often confined to niches (Xu et al., 2022). Another subdivision of soil bacteria is the evolutionary consideration of what has termed as "feast or famine". In environments characterised by low levels of nutrient and organic matter, organisms which have adapted to this regime are labelled "oligotrophs". Oligotrophs are unable to grow in high nutrient substrate conditions in comparison to "copiotrophs" which are organisms that do possess this capacity to grow and dominant in high nutrient environments (Koch 2001). Bacterial fitness and adaptation in soil are often predicated on land use. Niches in subsoil can

remain undisturbed for millennia, however, human activities increasingly alter these environments resulting to significant changes to these biomes.

2.3.1 Dominant Phyla in Soils

Several key taxa have been identified to dominate soil at a taxa level, with varying levels of abundance depending on the geographical location of the ecosystem (Fierer et al., 2007). In one global survey of soil bacteria, it was found that 2% of the global operational taxanomic units (OTUs) identified accounted for 41 % of relative read abundance (Delgado-Baquerizo et al., 2018) (**Figure 2.15**). Operational taxonomic units are closely related individuals, all global soil biomes are dominated by *Acidobacteria, Actinobacteria, Proteobacteria* and *Bacteroidetes*. *Proteobacteria* appears makes up between 10 % and 77% of soil libraries averaging at 39% (Janssen et al., 2006; Spain et al., 2009; Fierer et al., 2009; Nemergut et al., 2011). The phylogenetic resolution of OTUs can reduce when analysing lower ranks, especially in earlier years of research. Now with more developed sequenced technologies and aggregated genomic databases researchers initially analysed the ecological composition based on abundances at the phylum and class level, but now are able better determine ecological functionality based on abundances found at the genus and species level.

Proteobacteria have been attributed to be more highly distributed in grassland soils than other environments (Spain et al., 2009). Of the classes of *Proteobacteria* found in grasslands, *Alphaproteobacteria* consists of 37 %, *Beta* – (16%) and *Gamma* – *Proteobacteria* 7.6% (Spain et al., 2009). However, it has been noted that in long-term agricultural sites *Beta-Proteobacteria* and *Gamma-Proteobacteria* are dominant as they are copiotroph bacteria with high growth rates (Montecchia et al., 2015; Kim et al., 2021). *Acidobacteria* represents 20% of bacterial communities in forest soils versus 11% in pasture soils. Within the phyla there are 26 subgroups of which 1,3,5, and 6 account for 87% of the total *Acidobacteria* community in forests and 75% of pasture soils (Naverette et al., 2015). The subgroups 1,3,4,5,6 suggest shifts in land use from the conversion of natural forest to agriculture land (Naverette et al., 2015). *Actinobacteria* make up 13% of the soil bacterial communities, which contains three subclasses that are common in soil: *Actinobacteridae, Acidimicrobidae and Rubrobacteridae* (Janssen et al., 2006). Most of the *Actinobacteria* are saprophytic breaking down dead organic material and spend most of their lifecycle as semidormant spores, especially in nutrient-limited conditions. These organisms are of particular interest in the field of biotechnology as they produce the majority of naturally occurring bacterial antibiotics (Barka et al., 2016). The phylum can be divided into six subclasses, *Actinobacteria, Acidimicrobiia, Coriobacteriia, Niteiiruptoria*, *Rubrobacteria* and *Thermoleophilia* (Barka et al., 2016). Actinobacteria grows optimally at a neutral pH with optimal growth temperatures of between 25 and 30 ^{O}C (Barka et al., 2016). *Bacteroidetes* phyla compose 5% of the soil ecology, a copiotrophic bacteria, with the *Spingobacteria* class being common in the soil (Janssen et al., 2006). The *Sphingobacteria* class are diverse as some are aerobic, some anaerobic and some facultative

anaerobes, resulting in shifts in community composition depending on the oxygen levels present within the soils (Janssen et al., 2006). These phyla also compose a significant number of the organisms that reside in the human gastrointestinal tract (GTI) which is mainly composed of the *Bacteroidia* class, while classes found in the environmental classes consist of *Flavobacteria, Cyophagia* and *Spingobacteria* (Thomas et al., 2011). *Planctomycetes* are commonly found in soils across the planet. They are significantly affected by soil management with calcium content, application of nitrates and pH having been associated with variations in the *Planctomycetes* community composition (Buckley et al., 2006). *Planctomycetes* form a unique phylum of bacteria with the absence of peptidoglycan proteinaceous walls, which creates a resistance to classes of antibiotics which inhibit the synthesis of cell walls such as beta-lactams (Fuerst et al., 2011). *Planctomycetes* provide an excellent example of the divergence of classification from morphology based on the physical attributes of cells (such as the presence of fimbrils (stalks) from the cell) to the use of 16 rRNA sequencing which has allowed for the classification of three distinct classes Planctomycetia, Phycisphaerae and *candidatus Brocadia* (Jenkins and Stanley., 2013). *Candidatus Brocadia* has been of heightened interest to environmental engineers due to the presence of anammox groups (Jenkins and Staley., 2013). *Chloroflexi* phyla compose an average of 3% soil bacterial communities, with cultivatable species belonging to the *Anaeroliae* and *Caldilinea* class (Janssen et al., 2006; Yamada et al., 2009). *Chloroflexi* have been found in a variety of natural and anthropogenic environments; they dominate organic-rich deep-sea seafloor biospheres but are also found in anaerobic sludge digesters where they act as anaerobic heterotrophs with filamentous characteristics creating a fluffy sludge (Yamada et al., 2009). The phylum *Verrucomicrobia* is considered a dominant phylum in soil ecosystems. It is an oligotrophic bacterium, composing 0 - 21% of soil communities and its abundance has been shown to be impacted by soil moisture and pH (Sangwan et al., 2005; Janssen et al., 2006). *Verrocomicrobia* has 5 main lineages with the class *Spartobacteria* being the most dominant in soil (Sangwan et al., 2005). *Firmicutes* are commonly found in the rhizosphere zones of soils. They are metabolically diverse and are known to be common in culture collections (Zhang and Xu., 2008; Hashami et al., 2020). In agricultural settings they have been identified as providing a possible bioaugmentation function to aid plant growth due to their ability to cope with temperature and salt stress (Hashami et al., 2020). *Desulfobacterota* phylum corresponds to sulphate-reducing, fermentative and synthrophobic bacterial lineages, with a preference to anoxic conditions (Murphy et al., 2021). Cultured from a variety of natural, aquatic, terrestrial and engineered environments they have been shown to grow under a variety of salinity, temperature, and pH gradients (Murphy et al., 2021). Three new classes have been identified for the phyla, *Zymogenia, Anaeroferrophillalia, Anaeropigmentia*. These classes suggest evolution to oxygen exposure with differing capacity to detoxify oxygen (Murphy et al., 2021). *Synergistota* are not commonly found in the terrestrial environment but are one of the main phyla found in high-salinity environments and are involved with hydrolysis and acidification (Lu et al., 2023). *Gemmatimonadota* is cosmopolitan as it inhabits a range of terrestrial and aquatic ecosystems,

accounting for an abundance of 0.3 – 1.8% in soil (Mujakić et al., 2022). There are five class levels through the phylum with *Gemmatimonadetes* being the most represented class in soils, activated sludge in wastewater treatment plants and freshwater (Mujakić et al., 2022).

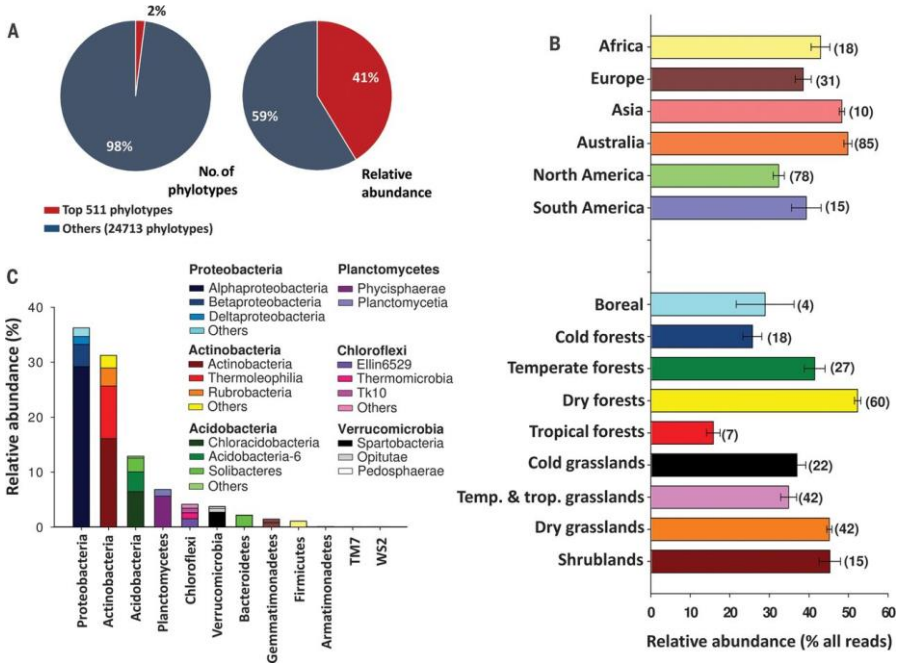


Figure 2.15. Abundance and composition of dominant soil bacterial phylotypes across the globe. (A) Percentage of phylotypes and relative abundance of 16S rRNA genes representing dominant and rare phylotype sequences. (B) Relative abundance (mean ± SE) of dominant phylotype sequences in respect of geographical position and ecology. Ecosystem type classification followed the Köppen climate classification and the major vegetation types found in the database. Grasslands include both tropical and temperate grasslands. Shrublands include polar, temperate, and tropical shrublands. The number of samples in each category is indicated in parentheses. (C) The taxonomic composition of the dominant phylotype sequences. The phylotypes assigned to the least abundant phyla are not shown (including *Armatimonadetes* = 0.08%, TM7 = 0.05%, and WS2 = 0.03%). (source: Delgado-Baquerizo et al., 2018)

2.3.2 Dominant Taxa in wastewater treatment system

It is argued that the community composition found in sewage influent accurately aggregates the human stool bacterial community. A study by Newton et al. (2015) found that, of the oligotypes found in human stools, they only accounted for 15% of the sewage sequence reads. However, 97% of the oligotypes found in human stools were captured in sewage sequence reads. The shift in community compositions has been noted in other papers where 80-90% of bacterial sequences origin from non-human sources (Vanderwalle et al., 2012; Shanks et al., 2013). *Proteobacteria* were seen to dominate influent composition with 60% of the abundance of sequence reads followed by *Firmicutes* at 20% and 15% Bacteroidetes (Shank et al., 2013). The possibility that this shift in the community from the stool biome to the influent may be due deposition of the established communities residing within the sewerage network (Vanderwalle et al., 2012). Wastewater

treatment appears to cause significant shifts to the influent community composition as it was observed that following activated sludge treatment and UV disinfection there was a significant reduction of Firmicutes from approximately 50% relative abundance of influent sequences to approximately 15% of effluent (Numberger et al., 2019). These reductions in *Firmicutes* were surprising, as spore formers Firmicutes can often survive extreme external environments. Wastewater treatment appeared result in the significant increase in the relative abundance of *Bacteroidetes, Planctomycetes, Proteobactera* and *Verrucomicrobia* of approximately 10 %, 2.5 %, 15% and 15% respectively (Numberger et al., 2019). Globally, there is a large diversity in the bacteria found in wastewater treatment plants. In one study 269 wastewater treatment plants were sampled in 23 countries across 6 continents: from the total of 61,448 OTUs found, there was only a core 28 which were present in all systems, which accounted for 12.4±0.2% (mean ± s.e.m.) of the sequences in activated sludge samples (Wu et al., 2019) (see **Figure 2.16**). The most dominant taxa in the core OTUs were the *Beta-Proteobacteria* which occurred 85% of all samples accounting for 15 of the 28-core OTUs (Wu et al., 2019). When potential sources of the communities were investigated, it was found that freshwater accounted for the greatest levels of variance in communities at 46%, followed by soil at 17% (Wu et al., 2019). An in-depth field study conducted by Li et al (2012) provided analysis of the community composition across depths in managed aquifer recharge sites from the infiltrative surface downwards. *Proteobacteria* dominated with abundances of 50.9 % (1-2 cm) and 40.8 (10-50 cm), with *Beta-Proteobacteria* being the dominant subgroup. Interestingly in the unsaturated zone *Proteobacteria* and *Bacteroidetes* remained dominant, but their abundance dropped significantly, from 50 % (1 – 2 cm) to 22 % for *Proteobacteria* and 18.8 % (1 – 2 cm) to 9.6 % for *Bacteroidetes*. Inversely, *Actinobacteria* and *Firmicutes* saw large increases in abundances from 2.5 % and 2.2 % in the infiltrative zone (1 – 2 cm) to 17.4 % and 27.4 %. The variation in abundances suggests the position of biogeochemical processes due to availability of nutrients and metabolites with phyla such *Nitrospira, Planctomycetes* and *Verrucomicrobria* which exhibited the greatest abundance between 10 cm and 50 cm depth. *Proteobacteria* appears to be the favoured in managed aquifer recharge conditions (Li et al., 2012; Barbra et al., 2019; Gorski et al., 2020; Xia et al., 2020; Schrad et al., 2022). In on-site wastewater treatment systems *Proteobactera* has been identified as the most dominant phylum of the core bacterial community within septic tanks at 82% (Tomaras et al., 2009; Knisz et al., 2021). Tomaras et al. (2009) sampled septic tank effluent and a soil treatment unit at different depths, *Bacteroidetes* and *Proteobacteria* dominated the effluent and STU samples, however *Proteobacteria* and *Acidobacteria* were the dominant taxa in the control subsoil samples.

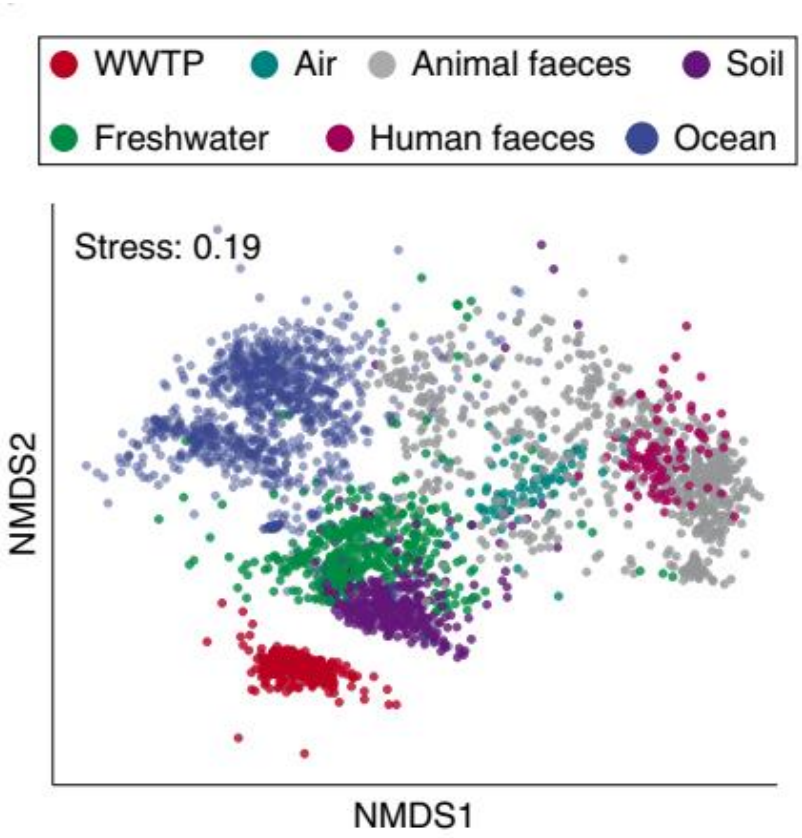


Figure 2.16. NMDS analysis showing that activated sludge of WWTPs harbours a unique microbiome compared with other habitats. For comparison, we merged our I table (n= 269 WWTPs) with that released by the EMP5, which contained thousands of bacterial communities from various habitats such as soil (n= 338 samples), ocean (n= 969 samples), freshwater (n= 447 samples), air (n= 81 samples), human faeces (n= 99 samples) and animal faeces (n= 622 samples), but not activated sludge from WWTPs (see Methods for details). Bray–Curtis distance was calculated to represent the dissimilarity in bacterial community compositions. (Source: Wu et al., 2019).

2.3.3 Functional species in wastewater treatment

Wastewater treatment is often defined as an engineering problem, with process flow charts focussing on measuring chemical and physical characteristics of influent and effluent, and bacterial ecology understood only by their required growth conditions. With advances in molecular techniques, it is now more possible than ever to apply a theoretical ecological framework on anthropogenic environments in the interest of optimising such treatment processes. Using 16S rRNA gene sequencing and the application sequence databases such as the Microbial Database for Activated Sludge, it is now possible to attribute important functional guilds to microbial ecology (McIllroy et al., 2015). In situ evidence for traits allows for the identification ammonia and nitrite oxidising bacteria (AOB and NOB), anaerobic ammonium oxidisers (Anammox), denitrifiers (nitrite reducing), polyphosphate accumulating organism (PAO) and fermenters such as methanogens. In identifying and quantifying functional groups in the system, we can profile wastewater treatment technologies fitness for the process to inhibit or enhance the microbial metabolism at play as to reduce the levels of greenhouse gases (GHG) released such as methane or to enhance immobilisation of nutrients.

Methanogenesis

Archaea produces approximately one billion tons of methane (CH_4) every year, a GHG with a greater warming capacity to CO_2 and therefore aims to mitigate such emissions are needed to start to dampen the impacts of climate change for life on the planet (Nobu et al., 2016). The process of methanogenesis is carried out by strictly anaerobic bacteria, all of which belong to the archaea phylum *Euryarchaeota* in five orders that include mesophiles to thermophiles: *Methanobateriales, Methanococcales, Methanomicrobiales, Methanopyrales* and *Methanosarcinales* (Balch et al., 1979;. Liu., 2010).

There are three major pathways for methanogenesis. The first is performed by Acetoclastic methanogens which ferment by acetate dismutation, with adenosine triphosphate (ATP) synthesis occurring by electron transport-linked phosphorylation, not substrate-level phosphorylation. Acetoclastic methanogenesis accounts for approximately two thirds of total methane formation. Acetoclastic methanogenesis is most common within freshwater sediments and anaerobic digesters. The second pathway is methylotrophic methanogenesis, where methanol or methylamines act as a substrate. Methylotrophic methanogenesis is essential in some marine sediments and anoxic environments, The third pathway for methane formation is hydrogenotrophic methanogenesis (Amador et al., 2019). Hydrogentotrophic methanogens use H_2 for the reduction of CO_2 (or CO or formate) according to:

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$$

This process is formally a type of respiration, and the organisms that use it can grow autotrophically. However, an ordinary electron transport chain is not found. The energetics of hydrogenotropic methanogenesis seem to be relatively favourable theoretically, but in practice cell growth rates and yields are lower than predicted from thermodynamic considerations. This is partly in response of the autotrophs requiring some substrate for C-assimilation, which includes a significant amount of ATP. Methanogens in comparison to sulphate reducers are relatively less competitive for substrates that 36ffect36uired by both groups of organisms: H_2 and acetate. In environments with a presence of sulphate, methanogenesis is low, but in environments where sulphate levels are depleted such as anaerobic digesters methanogens play a key role in the carbon cycle (Amador et al., 2019).

Methyl coenzyme -M reductase (MCR) catalyses the final reduction step of the methanogenesis pathway. The gene encoding for the α- subunit of MCR (mcrA) is highly conserved and is unique to anaerobic methanogens including aceticlastic and hydrogenotrophic (Fernández-Baca et al., 2018). mcrA has been used as a biomarker as to link CH_4 emissions to methanogen activity in different soil environments (Fernández-Baca et al., 2018).

Methane Oxidation

Microbial consumption of methane in anoxic environments is critical for the regulation of methane release globally. Methane oxidation in anoxic environments is estimated to consume the equivalent

of 5-20% of the net modern atmospheric methane flux (20-100 x 10^{12} g year-1) (Valentine et al., 2000). Aerobic methane oxidising bacteria (MOB) contains methane monooxygenase (MMO), which catalyse the transformation methane to methanol in the first step of methane oxidation to carbon dioxide (Fernández – Baca et al., 2018). There are two versions of the MMO in aerobic environments, particulate membrane bound, pMMO, and the soluble version, sMMO which is in the cytoplasm (Fernández – Baca et al., 2018). The pMMO version of the enzyme has been recently discovered in nitrite -dependent anaerobic methanotrophs of the *NC10* phylum (Fernández – Baca et al., 2018). sMMO is contained in a minority of MOB, and for this reason the pmoA gene, encoding the α- subunit of the pMMO genes is used as a biomarker for MOB activity (Fernández Baca et al., 2018).

Nitrogen Cycle

Nitrogen (N) concentrations commonly found in soils in the top 15 cm range from 0.1% to 0.6% (Cameron et al., 2013). This is constituted into four major forms (a) organic matter plant material, fungi and humus; (b) soil organisms and microorganisms; (c) ammonium ions (NH_4^+) held by clay minerals and organic matter and (d) mineral N forms in soil solution, this includes $NH4^+$, nitrate (NO_3^-) and low concentrations of nitrite (NO_2^-) (Cameron et al., 2013). N undergoes several transformations in a soil environment through a series of different processes as described in **Table 2.4**.

Table 2.4. Describes the processes within the Nitrogen cycle.

Process	Description	Equation
Ammonification	Microbial decomposition of organic N which results in the release of NH_4	$NH_2CONH_2 + H_2O \rightarrow 2NH_3 + CO_2$ $RCH(NH_2)COOH + H_2O \rightarrow NH_3 + CO_2$
Volatilisation	Abiotic transformation of NH_4 to NH_3	$NH_4^+ + OH^- \leftrightarrow NH_3 + H_2O$
anammox	Oxidation of NH_4 in anaerobic conditions, utilising NO_2 and NO_3 as electron acceptors instead of organic C	$NH_4^+ + 1.3NO_{2-} + 0.1HCO_3^- + 0.1H^+ \quad N_2 + 0.3NO_3^- + 0.1CH_2O_{0.5}N_{0.2} + 2H_2O$
Nitrification	Conversion of NH_4 to NO_2 and then NO_3 by aerobes	(1) $NH_3 + º.5O_2 \rightarrow NO_2^- + H^+ + H_2O$ (2) $NO_2^- + H_2O \rightarrow NO3^- + 2H^+ + 2e^-$
Denitrification	Conversion of NO_3 to N gases by anaerobes	$2NO_3^- + 5H_2 + 2H^+ \rightarrow N_2 + 6H2O$

Ammonification

Ammonification (mineralisation) is the transformation of organic nitrogen into ammonia. As urea and uric acid each lose an amine group (NH_2) this results in the production of NH_3, and this then dissolves in water of most soils as to produce NH_4 (Lusk et al., 2017). Ammonification occurs by bacteria species from *Bacillus, Clostridium, Pseudomonas, Serratia* and *Micrococcus* (Hui et al., 2019). These bacterial species are supported by extracellular enzymes like urease. Ammonification can occur in both aerobic and anaerobic conditions (Lusk et al., 2017; Saeed and Sun., 2012;

Stefanakis et al., 2014). The process is impacted by temperature, pH, C/N ratio content, and soil conditions. The optimum pH range is between 6.5 and 8.5 and temperature between 30-35 ^{o}C degrees, and it has been reported that the ammonification rate doubles with a 10^{o}C temperature increase (Lusk et al., 2017; Saeed and Sun., 2012; Stefanakis et al., 2014). The rate of ammonification tends to reduce with depth in soil as the decomposition that was performed by heterotrophic bacteria is taken over by a less efficient more restrictive anaerobic microbial community (Reddy et al., 1984; Lusk et al., 2017; Stefanakis et al., 2014). Mineralisation is likely to occur when the C:N ratio is below 20:1 (Henry et al., 1999). Ammonification is enhanced in soils with coarse texture and low clay contents (Lusk et al., 2017)

Nitrification

Nitrification is the biotic process of NH_4 conversion into NO_3 and is considered the second step in the N transformation stage. (Stefanakis et al., 2014; Hazen and Sawyer., 2009). Chemoautotrophic bacteria perform a two-stage conversion of ammonium. In the first stage ammonia NH_3 is oxidised under aerobic conditions by chemolithoptrophic *Nitrosomonas, Nitrosococcus*, *Nitrosolobus* and *Nitrosopira* bacteria. Although ammonia-oxidising archaea are present in large numbers, they are inhibited by NH_3 and are not as active as their bacterial counterparts (Amador and Loomis., 2018; Cameron et al., 2013). Stage 2 is the transformation of NO_2 to NO_3 by members of the genus *Nitrospina, Nitrospira, Nitrococcus* and *Nitrobacter* (Amador et al., 2019). The process depends on several factors such as an optimal pH (7.5 – 7.8), temperature (25 to 35 ^{O}C, although is near inhibited at temperatures less than 5^{o}C and greater than 50^{o}C), and dissolved oxygen (~2 mg/L) (Stefanakis et al., 2014; Lusk et al., 2017). The rate of nitrification corresponds to O_2 availability and NH_4 concentrations, with high levels of NO_2, and NO_3 inhibiting the process (Amador and Loomis., 2018). Methane in sufficient concentration may interfere with nitrification by inhibiting the activity of the enzyme ammonia oxygenase. The maximum nitrification rate occurs at the field capacity moisture levels, however when soils either exceed field capacity moisture or in dry soils there is a significant drop in the rate of nitrification(Cameron et al., 2013).

Denitrification

Denitrification is the reduction of NO_3 in series, converting NO_3 to NO, N_2O and N_2. The first reduction is performed by the enzyme nitrate reductase ($NO_3 \rightarrow NO$). NO is reduced to N2O by nitric acid reductase and N2O reduced by nitrous oxide reductase into N_2 (Lusk et al., 2017). The process is performed by facultative heterotrophic or autotrophic bacteria under anoxic conditions (DO< 0.3 mg/L). Organic carbon is used as an electron donor by these heterotrophs. Denitrification takes place in an anoxic environment or anoxic microsites by heterotrophs. The denitrifiers consist largely from the genus *Pseudomonas, Bacillus, Micrococcus,* and *Spirillium*. They use oxygen from the nitrate molecules as an electron donor and the breakdown compounds as an electron acceptor, producing the energy required for growth (Amador and Loomis., 2019; Stefankis et al., 2014 Lusk et al., 2017; Hazen et Sawyer., 2009). Denitrification does occur in anoxic conditions via bacteria

such as *Thiophaera pantotropha* (Amador and Loomis., 2019). The process can occur through autotrophic bacteria, using H_2S, S_0 or CH_4 as electron donors and CO_2 as a carbon source (Cooper et al., 2016; Amador and Loomis., 2018). Denitrification rates have been shown to increase with an increase in temperature. The optimum temperature for denitrification was shown to be 60-75^{O}C, with a 1.5-2.0-fold increase with increases of 10 ^{O}C (Reddy et al., 1984). More recent analysis on granular sludge bed reactor have found that the optimum temperature range for denitrification was 15-35 ^{O}C, however dentrification was accomplished at temperatures as high as 52 ^{O}C (Liao et al., 2018). The environmental conditions of the soil can impact the extent in which nitrogen is processed. If the concentration of O_2 increases with a decreasing pH, the activity of nitrous oxide reductase is impacted and results in a greater production of NO_2 (Amador and Loomis., 2019).

Anaerobic Ammonia Oxidation (Anammox)

Anammox is the oxidation of ammonia with either nitrite or nitrate in anoxic conditions, producing gaseous N_2 and H_2O (Amador and Loomis., 2018; Stefanakis et al., 2014). The process requires low energy and oxygen and does not require an external carbon source (Stefankis et al., 2014). The process is mediated by bacteria from the phylum *Planctomycetes* with six genera confirmed including *Candidatus Kuenenia, Ca. Brocadia, Ca. Anammoxoglobus, Ca. Anammoximicrobium, Ca. Jettenia,* and *Ca. Scalindua* (Nie et al., 2018; Wu et al., 2020). Of these genera mentioned five have been commonly found in wastewater treatment and freshwater, and the last on the list was found in saline environment such as seawater and sediments (Wu et al., 2020). The level of nitrogen in the environment determines which of these genera become active. In low nitrogen loading rate conditions *Ca. Brocadia anammoxidans, Ca. Jettenia, Ca. Anammoxoglobus,* and *Ca. Kuenenia* are dominant and in high loading rate results Ca. Brocadia sinica and Ca. Kuenenia stuttgartiensis being dominant (Wu et al., 2020). Under oxic conditions it is believed that anammox can use nitrite oxidoreductase enabling them to immediately use NO_2 as an energy source when available (Wu et al., 2020) Anammox are autotrophic, i.e., the presence of organic carbon can inhibit their growth. However, some species have been able to metabolise fatty acids (Wu et al., 2020). Annamox are slow growing, and it is understood that their doubling time can be from one to several weeks (Ali and Okabe., 2015). This requires that anammox have adequate time for processing nitrogen, either through an adequate sludge retention time in activated sludge treatment or reduced hydraulic conductivity within the soil. It is important to note that the group is susceptible to nutrient shock with diminished rates of anammox taking place in reactors where ammonia was increased from 13.65 ± 2.69 mg / L to 29.65 mg / L , with an optimum reaction concentration of 20-25 mg / L (Wu et al., 2020). Temperature also has an impact on the anammox rate, with instability observed at low temperatures of 15 ^{O}C due to the accumulation of nitrite. However, at lower concentrations of nitrogen (0.4 g N/L·d) it was possible to achieve the bio-reaction (Lotti et al., 2014).

Biotic Phosphate Accumulation

The storage of P compounds (mineral or organic) depends on the type of microorganism present and their ability to store P in intracellular polymers, mainly polyhydroxyalkanoates (PHAs) (Dorofeev et al., 2020). Certain archaea such as *Halobacterium salinarium* and *Halorunrum distributum* can precipitate $MgPO_4OH$ $4H_2O$ from an aqueous solution during the growth phase (Tarayre et al., 2016). *Brevibacteria* and *cyanobacteria* can accumulate P as either NH_4MgPO_4 $6H_2O$ or within their sheaths when combined with calcium in the case of cyanobacteria (Tarayre et al., 2016). Organic P accumulation occurs in some bacteria through the production of teichoic acid (Tarayre et al., 2016). Overall inorganic poly-P is the primary P reserve but attempts to isolate cultures responsible for wastewater treatment has proved challenging. However, molecular techniques identified *Accumulibacter phosphatis* of the class *Beta-Proteobacteria* which is believed to be an important subclass (Oehmen et al 2007; Tarayre et al., 2016). *Accumulibacter phosphatis* has been identified in activated sludge treatment sites and sediments of freshwater bodies (Peterson et al., 2008). Microbial phosphorus removal in wastewater is determined by the amount of consumed organic substrate as P/C ratio, which depends on the acidity of the medium. If the pH increases from 5.5 to 8.5 this will cause an increase in the P/C ratio from 0.2 to 0.75 when acetate is digested anaerobically. PAOs are defined to function effectively at an optimal temperature of 20^{o}C, although they have been shown to be highly active at temperatures as low as 5^{o}C (Dorofeev et al., 2020). The P/C ratio is an extremely variable indicator, but in technological practice it is assumed that 7-10 mg of bioavailable acetate is required to accumulate 1 mg of P. Reduction in phosphate removal in wastewater treatment systems are often attributed to competition between PAOs and glycogen accumulating organisms (GAOs) for acetate, with GAOs under anaerobic conditions accumulating organic substate to generate biomass and energy under aerobic conditions (Dorofeev et al., 2020).

2.4 Community ecology development in on-site wastewater treatment

Biogeography is the study of the spatial distribution of biological diversity with the aims of understanding and identifying the underlying mechanisms causing differing differences in community composition. Community composition is as a result of contemporary interactions among organisms, but also physical and biotic environments (Lindström et al 2012). A metacommunity framework has been devised to determine whether local factors (because of environmental conditions) or regional factors (because of dispersal) regulate the resultant biological diversity (Martiny et al., 2006). The framework was then adapted to analyse bacterial community assembly by employing four separate hypotheses for the development of the ecology: (i) patch dynamics, which assumes that diversity is determined by dispersal or species interaction among environmentally identical patches; (ii) species sorting, which assumes that dispersal is high enough to allow immigration to all patches but that local habitat conditions determine composition; (iii) mass effects, which assumes that massive immigration can rescue species from competition

exclusion, thus relaxing the connection between community composition and local environmental conditions; and (iv) neutral, which assumes that all species are similar in their competitive ability and in dispersal, and community composition will drift over time (Lindström et al 2012).

These methods have been used to understand the development of microbial biofilms in river sediments which drive many ecosystem services. A study by Besemer et al. (2012) found the community in a young river sediment biofilm to be less diverse than that of surface water. The communities present in the stream were distributed from a variety of sources within the catchment, with the presence of soil dominant proteobacteria being present within the community sampled. It was noted that there was homogeneity in the community composition of the river biofilms when randomly sampled, however, surface water communities differed to an extent. It was then hypothesised that species sorting through selective pressure of the niche environment was the factor affecting community assembly (Besemer et al., 2012). Within soil ecosystems the composition of soil communities is affected by different factors depending on depth. In topsoil, identified as either the first 10 cm or 30 cm of the soil, communiti41ffectfected by land use, large dispersals of species via air or water, root density, and variation in soil moisture and temperature. In deeper subsoils immigration of organisms is attributed to burrowing soil organisms and water movement (Powell et al., 2015). However, if dispersal of organisms is possible to deeper positions within the subsoil, they may be ill suited for survival at such levels displaying low fitness levels and biogeochemical function (Luan et al., 2020). It has been observed that with greater depth, levels of environmental filtering of species are higher, resulting in less predictability in the changes of the biome composition, suggesting that natural subsoil ecology is influenced by neutral processes (Powell et al., 2015).

In wastewater treatment plants, it was found that dispersal of species (characterised as the immigration of species from the influent into the activated sludge treatment system) did not impose a significant effect on community composition when the system was functioning regularly (Saunders et al., 2016; Vuono et al., 2016; Ali et al., 2019). Species sorting appeared to be the primary factor in community composition until there was a disturbance in the activated sludge system's running. At such s times of flux it was noted that a greater number of immigrants were colonising the system (Vuono et al., 2016). It appears in wastewater treatment plants that species sorting is the factor affecting species in developing ecological niches, with one study of 14 wastewater treatment systems finding that the geographic location of the wastewater treatment plants represented only 14% of community variance (Wang et al., 2012; Saunders et al., 2016; Vuono et al., 2016; Matar et al., 2017; Ali et al., 2019). In topsoil irrigated with reclaimed water, it was noted that species diversity increased with the addition of nutrients; whether this was due to immigration was not specified. However, functional diversity also reduces which may be due to the alteration of the environment with addition of reclaimed water resulting in the selection of generalist bacteria (Chen et al., 2017). Generalists are considered to have higher diversification, are more readily capable of transforming

from one ecological state to another and can be characterised by more stochastic processes (Xu et al., 2022).

2.4.1 Factors Affecting Microbiome Development

There is no common or uniformity in soil biomes, each ecology is different depending on its micro environmental conditions. Due to the levels of variation within each biome, it was initially difficult to assess which abiotic or biotic factors are most impactful on the community (Fierer et al., 2017). Over time researchers came to a consensus that the most important factors influencing the structure of soil microbial communities were likely pH, nitrogen availability, soil organic carbon content, temperature, and redox status (Pett-Ridge et al., 2005; Lauber et al., 2009; Sul et al., 2013; Ceberlund et al., 2014; Oliverio et al., 2017). Changes to pH particularly affected abundances of *Acidobacteria* which favour acidic soil, and *Actinobacteria* and *Bacteroidetes* both favouring alkaline soils. It must be noted that pH is not a universal predictor of community composition, and that bacterial diversity was at its highest at a neutral pH (Lauber et al., 2009). Long-term field trials found that the N-fertilisation of soils increased the availability of easily degraded carbon and because of this favoured copiotroph organisms within the soil (Cederlund et al., 2014). Temperature fluctuations have been to show not to directly correlate with taxonomy as many phyla have members with varying preferences to temperature conditions, which denotes the challenge of determining the effect of environmental factors by broad taxonomical ranks. The only phyla that have been shown to be broadly significantly temperature sensitive were *Chloroflexi* and Actinobacteria, with both exhibiting increases in abundance due to a warm response, however the phyla resolution of the study being limited to the family phylogenetic rank level (Oliviero et al., 2017). Organism specific analysis could discern the microbial physiology response to environmental pressures. Due to rainfall and other hydrological events as well as human irrigation soils can undergo differing states of saturation. Waterlogging in soils and sediments can result in the rapid depletion of oxygen from both the liquid compartments and gas-filled pores within the soil. Oxygen is the primary terminal electron acceptor for respiratory processes and can determine the soil redox status and the potential pathways for the cycling of carbon and nitrogen. However, it has been hypothesised that many bacteria are highly adaptive to fluctuating redox conditions and have evolved to exist in a constantly alternating redox state from primarily aerobic to primarily anaerobic (Pett-Ridge et al, 2005). The application of soluble organic carbon has been shown to significantly affect community structure, with *Actinobacteria* and *Acidobacteria* phyla dominating in low organic soils as oligotrophic. The deposition of carbon has been shown to reduce diversity due to nutrient availability driving a less metabolically diverse community (Sul et al., 2013).

In natural terrestrial ecosystems soil carbon content indirectly determines microbial diversity in soils, and to a lesser degree pH and soil texture also contribute (Bastida et al., 2021). There are significant differences in microbial communities across the vertical layer where key soil properties

and drivers of microbial community variation such as pH, organic carbon content, salinity, texture, moisture content and nutrient availability shif significantly with depth (Fierer et al., 2017) (**Figure 2.17**). Soil was sampled across a depth profile has shown that microbial communities exhibited significant variance even when separated vertical distances as 10-20 cm. These variations in community structure were as distinct as those of communities that were thousands of kilometres apart. It was observed that deeper soils harboured communities which were like surface soil biomes found in semi-arid and more humid climates which do not have particularly low carbon, but the similarity may be because of the type of carbon present in both types of soil (Eilers et al., 2012). When communities were sampled at intermediate depths (20-60 cm) in natural soils, a relative uniformity in community structure was found, regardless of landscape position or covering vegetation (Eilers et al., 2012). It is hard to assess in deep soils (> 60 cm) which factors result in changes to the community structure due variability in the infiltration of organics, nutrients, surface weathering, hydrological conditions, and diversity of microbes dispersed due to the filtering capacities of topsoil (Eilers et al., 2012; Powell et al., 2015).

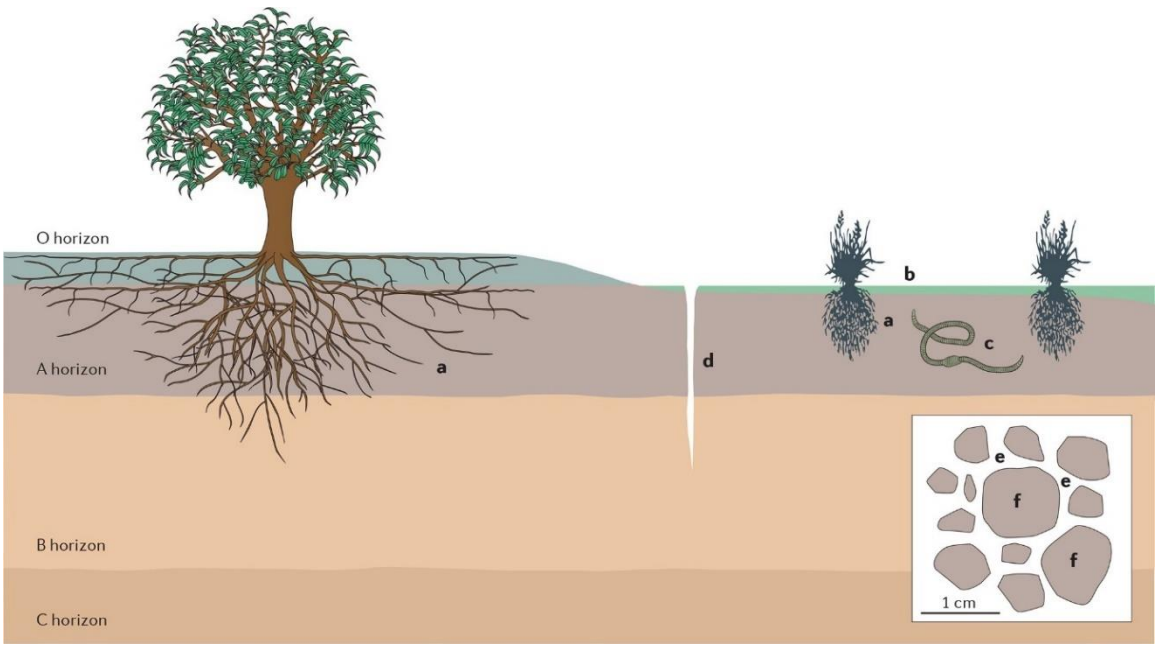


Figure 2.17. The soil environment is not monogamous it is a diverse aggregation of microbial habitats. (A) The rhizosphere is specialised habitat where bacteria work symbiotically with plant roots. (B) Photic zone, surface layers that are exposed to light. (C) Drilosphere environment described as earthworm burrows. (D) Soil found in preferential water flow paths, including cracks in the soil. (E) Communities established on aggregate surfaces or within water films between aggregates. (F) Communities found within soil aggregate structure. Majority of studies have observed the communities within the surface soil horizons (layers). However, communities found within the litter layer (O-horizon) are often distinct from those found in underlying mineral soil horizons (A and B horizons) and deeper layers (C horizons) (Source: Fierer et al., 2017).

Studies at WWTPs found that community variance correlated strongly with water temperature, conductivity, pH and dissolved oxygen content (Wang et al., 2012). A meta community analysis conducted on environmental factors affecting WWTP biomes which included 50 WWTPs across 13

countries in 6 continents found that variance in the community strongly correlated with dissolved oxygen, temperature, and pollutant concentration (Tian and Wang., 2020). When 269 WWTP activated sludge communities were assessed it was found that temperature, solid retention time and organic inputs were deterministic factors in the sludge community assembly (Wu et al., 2019). It is evident that the organic loading puts a deterministic pressure by creating an environment more suited for copiotrophs. Bioreactors run at low temperatures reduces bacterial growth rates, which if combined with a low SRT runs the risk washing out slow growing bacteria from the system (Ali et al., 2019).

2.4.2 Microbe-Environment Interactions

Bacteria soil biomes do not simply adapt or die-off due to pressures of their environment - bacteria also have the capacity to alter the environment due to biogeochemical processes they mediate. Soil pH is affected by the cycling organics and nutrients. In the carbon cycle only, it has been suggested by research that one third of CO_2 is emitted into the atmosphere, leaving the remaining CO_2 in soil to be converted to carbonic acid resulting in soil acidification. Cyanobacteria, which are photosynthetic autotrophs, fix carbon through biofilm growth which creates alkalisation events within their biome (Mergelov et al., 2018). In the nitrogen cycle, the oxidation of ammonium to nitrate by bacteria and archaea during nitrification which can result in the acidification of soil, particularly seen when N fertilisers are applied to agricultural soils (Norton et al., 2019). Inversely, ammonification and denitrification due to proton consuming processes results in greater alkalinity in soils (Philippot et al; 2023). In these geochemical processes bacteria can alter the pH in their environment which in turn can be a primary factor which affects the structure of their community (Fierer et al., 2017; Philippot et al; 2023). Bacteria have evolved to use metal ions by processing them through enzymatic action in reduction pathways which generate energy. Compounds such as iron, manganese, uranium and chromate, provide great opportunities for bioremediation of contaminated soils and as a result can alter the soil redox and sorption processes (Philippot et al; 2023).

One major effect bacterium has on the soil environment are the changes made to pore spaces between soil particles. Bacteria are important for the formation of microbial aggregates called biofilms. This is often achieved by the self-production of highly hydrated extracellular polymeric substances (EPSs) which aggregate soils as a binding agent. They are composed of polysaccharides, proteins, and DNA (Costa et al., 2018). EPS production is an energy-intensive process for bacteria, however, it provides protection from biotic and abiotic stressors, such as antimicrobial agents, temperature, pH, drought, and salinity (Costa et al., 2018; Philippot et al; 2023). The production of EPS increases the water retention within the soil, by blocking pore spaces and reducing the hydraulic conductivity. This affect has been observed in soil treatment units, as discussed earlier. Volumetric water content increases induced by EPS production have been used as an effective proxy for biomat (soil biofilm)

development. Pre-treatment of the wastewater in on-site treatment systems was shown to induce a reduction in biomat development (i.e. stunting) within the STU, which resulted in higher permeability due to the slower growth rates (Knappe et al., 2020). The development of EPS involves the interactions between polymer macromolecules, including dispersion forces, electrostatic interactions, and hydrogen bonds creating a gel-like tridimensionality structure around the cells which allows for the establishment of a stable niche (Flemming et al., 2010). This consortium of niches allows bacteria to function synergistically producing extracellular digestive enzymes which remain within the biofilm. The biofilm increases the metabolic capacity of the community in the transformation of colloidal material passing through the matrix, which is ultimately utilised as carbon and energy. Biofilms enhance collaboration between species through the exchange of metabolites and quorum sensing coordinates metabolisms which aids N cycling species (Fleming et al., 2016) (see **Figure 2.18**). Biofilms are considered as a possible location of high microbial activity or microbial hotspots. Hotspots are defined as small soil volumes with much faster process rates and much more intensive community interactions and greater diversity. Hence, it is thought that the relevant biogeochemical processes occur in a small volume of the soil (Kuzyakov et al., 2015).

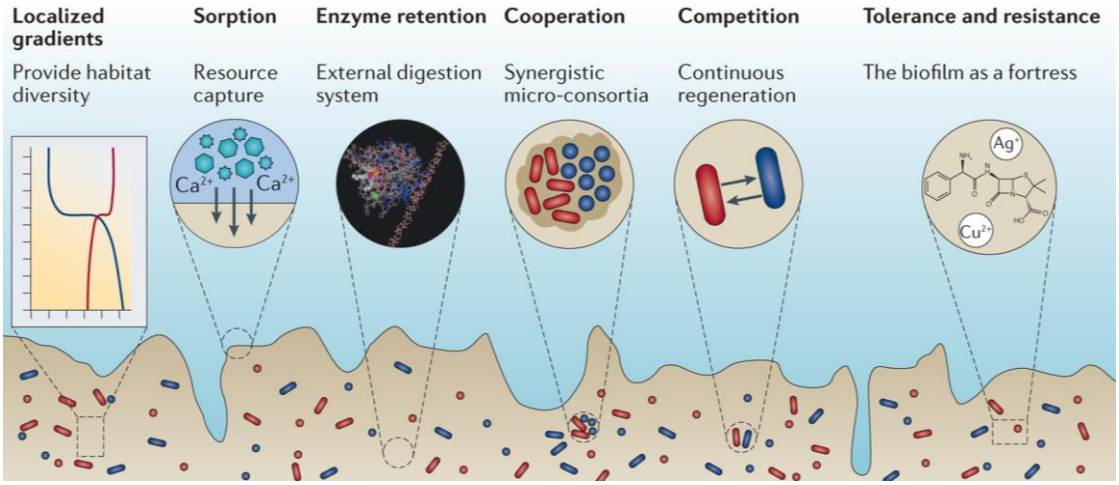


Figure 2.18. Formation of microbial biofilm, extracellular polymeric substances (EPS) composed matrix providing stability and architecture to the community. Nutrients and other compounds are enclosed within the matrix by sorption to EPS molecules and to the pores and channels of the matrix, the skin of the matrix provides an effective barrier as to prevent the community experiencing moisture shocks through desiccation. The biofilm provides abilities to the community that would otherwise be unavailable to them in their free floating. Bacteria have the capacity within the biofilm to function within localised gradients with diverse niches, resource capture by sorption, enzyme retention allowing for digestive capacities within the biofilm, social interaction, tolerances to abiotic (pH, drought, and salinity) biotic (antibiotics) stressors (Fleming et al., 2016).

2.4.3 Critical Points in Transition of Microbial Ecologies

Large changes in communities that punctuate the usual trends in ecosystems may be because of unpredictable shocks. There is the hypothesis that such a shift represents what is termed a *critical*

transition. The possibility of such a tipping point may be initially small, however, as time goes on and a community approaches a "tipping point", a relatively minor effect can result in a shift to a contrasting state. What is considered the basic mechanic for a tipping point is a feedback loop, in which when a critical point is passed, it propels toward an alternative state (Scheffer et al., 2012) (see **Figure 2.19**). Microbial communities often evolve toward a stable composition but can be altered by a transient perturbation shifting them from stable condition (Faust et al., 2015). An example of this was an experiment of a biogas reactor which was overloaded with 4 times the usual substrate load causing the cell counts to increase drastically thereby reducing the diversity and the functionality of the system (Koch et al., 2013). In structurally complex systems responses to change appear to be determined by their heterogeneity and interconnectivity of species. In heterogenous sites with low connectivity shifts in state may occur more gradually, while more homogenous connected systems may inhibit a resistive response to change and result in a more pronounced, synchronised shift. Prior to non-stochastic shifts or tipping point, a phenomenon is observed termed the "critical slowdown", which is described as an increasingly sluggish response to perturbations (Scheffer et al., 2012). Perturbations in the environment are classified in two categories: (1) increased nutrient input, especially complex carbon substrates such as nutrient amendment, eutrophication, or oil spills (2) disturbances, with heightened mortality and decrease in biomass such as drought, tillage, extreme, temperature, saltwater intrusion, pH or predation. The type of perturbation is important if it is a stochastic or deterministic process driving community assembly (Zhou et al., 2014). Nutrient drift is considered a stochastic process as it weakens niche selection by providing more resources. Selective factors such as droughts or temperature are more deterministic, by removing a large portion of regional species pool. Perturbations can also be classified temporally with pulses being discrete and short-term which can rapidly decrease in severity over a short period of time. Press perturbation may arise sharply but reaches a steady state that can be maintained over a long duration (Shade et al., 2012). If a press duration permanently alters the environmental conditions, this will have implications for community stability likely resulting in a shift to alternative stable states (Shade et al., 2012). For example, the pulse infiltration of nutrients into subsoil and groundwater and its effect on soil community assembly appear to be driven mainly by stochastic processes. It should be noted that the predictability of the community structure is more predictable if the deterministic processes play a more key role at the end of the succession (Zhou et al., 2014).

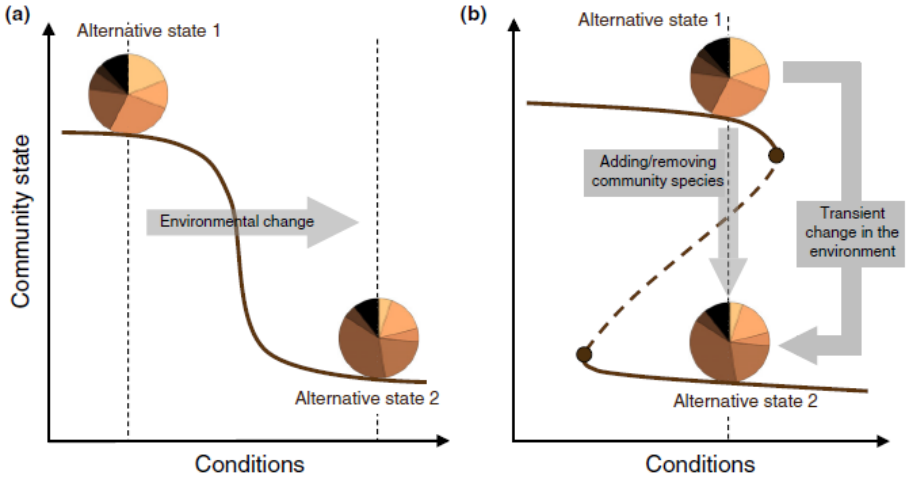


Figure 2.19. Shifts in environmental conditions (dashed vertical line) can affect a microbial community (pie chart) which pressures the composition into an alternative state (a). An ecosystem can also exhibit two (or more) stable states under the same conditions (b). A switch of the steady state composition can be induced either by a change in the composition (removal of species via antibiotics or introduction of new species) or by a transient change of the environmental conditions that pushes the system beyond the tipping points (small brown circles). In both cases, a small change in external conditions in the vicinity of a tipping point can trigger an abrupt shift in community structure. Such a state switch may be preceded by early warning signs (Source: Faust et al., 20215).

3. Field Studies and Laboratory Analysis

This chapter details the research sites and soil columns constructed instrumented and sampled during this research. In Section 3.1, the research sites and their on-site treatment systems are introduced. In Section 3.2, the instruments and monitoring techniques applied at each of these sites are described. Section 3.3 is dedicated to the construction and instrumentation of subsoil columns. The sampling protocol for porewater and subsoil is detailed in Section 3.4, covering both the research sites and the constructed columns. Section 3.5 is on the monitoring of innovative molecular techniques used for studying faecal indicator organisms. Finally, in Section 3.6, an overview of the laboratory analysis procedures is provided.

3.1 Field sites

Three research sites were monitored over the period of this research project (August 2018 to August 2022) to improve our understanding of the microbial ecology within on-site wastewater treatment systems. These systems are located within the Republic of Ireland, with the households based in rural settings. The households are homeowner occupied and continued regular domestic activities in the process of this research. The three sites (Killmallock; Site A, Crecora; Site B, Redcross; Site C) were constructed by the Department of Civil Structural and Environmental Engineering, with descriptions of the sites characterised in **Table 3.1**. Sites A and B were designed and constructed as a part of the doctoral research project "Modelling of soil biomass for on-site wastewater treatment" (Knappe et al., 2020) -which focused on determining the impact of the biomat at the infiltrative layer at the base of STU on contaminant attenuation. Site C was designed and constructed for the purpose of this doctoral research project (**Figure 3.1**).

The climate conditions of these sites are classified as maritime CFb (warm temperate, fully humid, warm summer) (De Carli et al., 2018). The mean annual temperature for Republic of Ireland is 9.8°C with mean air temperature ranges from approximately 8.5°C to 10.8°C. The mean summer temperature is 14.6 °C compared to mean winters lows of 5.4 °C. The national annual rainfall is 1288 mm per year. The highest rainfall is found in the south-west of the country at 2,044 mm per year with the lowest in the south-east at 878 mm per year (Curley et al., 2023).

Table 3.1 Characteristics of the selected study sites and wastewater treatment systems.

	Site A	Site B	Site C
County	Limerick	Limerick	Wicklow
Aquifer Category	Li	RKD	Li
Groundwater vulnerability	Moderate	High	Extreme
Subsoil	Typical Luvisol soils over a Limestone till (Carboniferous)	Typical Luvisol soils over a Limestone till (Carboniferous)	Clayey Sandstone and shale till
Land use	Garden	Pastureland	Pastureland
No. Occupants	4	4	1-2
Primary treatment	Two compartment ST	Two compartment ST	Two compartment ST
Capacity (m^3)	4.8	4.8	4.8
Hydraulic retention time (h)	10.3	10.3	10.3
Secondary treatment	Coconut filter	RBC	Open Intermittent sand-filter
Soil treatment unit	2-trench Primary effluent 2-trench Secondary effluent	2-trench Primary effluent 2-trench Secondary effluent	2-trench Primary effluent 2-trench Secondary effluent
Construction	2015	2016	2022
Area m^2	180	180	180

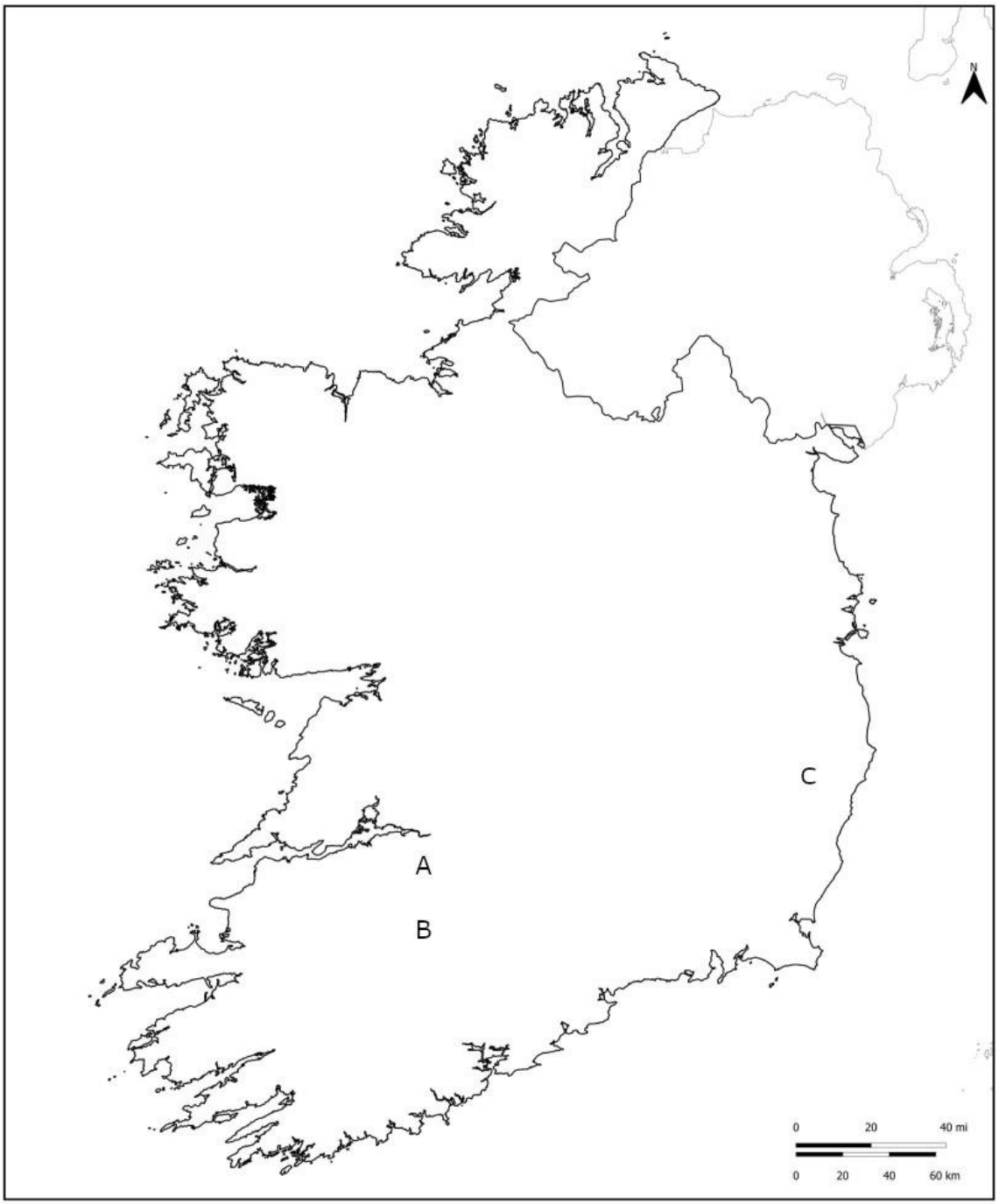


Figure 3.1. Locations of fields sites Killmallock (A), Crecora (B) and Redcross (C).

3.11 Killmallock (Site A)

Site A is in Killmallock, County Limerick. The site consists of a homogenous vegetated grassland with the system constructed in the back garden of a residential dwelling. The subsoil is classified using the British Standard classification BS 5930 (BSI, 1999) consisting of clay loam and a till derived mainly of limestone. The dwelling consists of a single household with four inhabitants. The construction of the on-site wastewater treatment system took place in September 2015. The domestic wastewater treatment system is a two-chamber septic tank, which flows out to a pump sump in which there are two separate discharge pipes. One of these pipes' pumps flow to a baffled attenuation tank where primary effluent subsequently is distributed to two percolation trenches. The second pipe

pumps effluent into a media filter consisting of coconut husk and then into two corresponding trenches (**Figure 3.2**).

Figure 3.2. Construction of the on-site system in Site A, with the installation of the septic tank (left) , coconut husk filter (centre) and the distribution buckets and STU trenches (right)

Primary Treatment is a two-chamber prefabricated concentrate septic tank (Aswasep Septic Tank NS 4 S Tank, Molloy Precast Products Ltd, the Republic of Ireland) the tank was designed under the guidelines set out by the EPA Code of Practice (**Figure 3.3**; for the layout of the site) (EPA., 2021). Effluent passes from the second compartment of the septic tank to a sump with two parallel pumps feeding the rising mains. The pumps (DOC3 Lowara , Xylem Water Solution Ltd., Republic of Ireland) were of a specification of 145 L / min , head of less than 7 m, capable of carrying suspended solids of a diameter of 10 mm, 220 V 50 Hz 2 pole, 2850 rpm. The pumps were controlled by a float switch which triggered one pump at a time once a certain level was reached in the chamber, which generated intermittent dosing of either PE or SE trenches.

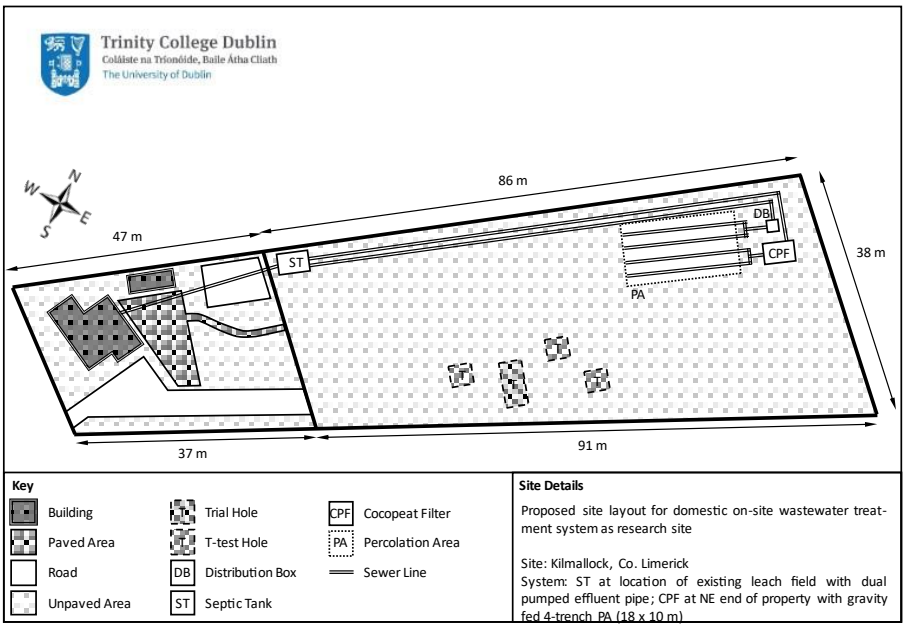


Figure 3.3. Layout of the Kilmallock site (Site A) and treatment system.

Secondary treatment at Site A consisted of media filter (Ecoflo, Premier Tech Aqua, Premier Tech). The media filter applies an intermittent trickle of wastewater displaced over a filter medium, in this case a shredded cocohusk fibre (**Figure 3.4**). The coconut husk material provides a significant surface area for biofilm production. Effluent is dosed over the media using a tipping bucket dispersion mechanism. As the effluent is processed through biofiltration (filtration, absorption, adsorption, ion exchange, microbial assimilation) its carbonaceous substrate content is reduced. In research conducted by Delaney et al. (2015) cocopeat was able to support nitrification and denitrification of wastewater. The typical lifespan of the media in these systems is approximately 15 years. The media, originally a by-product of other industries, can be replaced and the expended material is used in the anaerobic digestion to produce biogas.

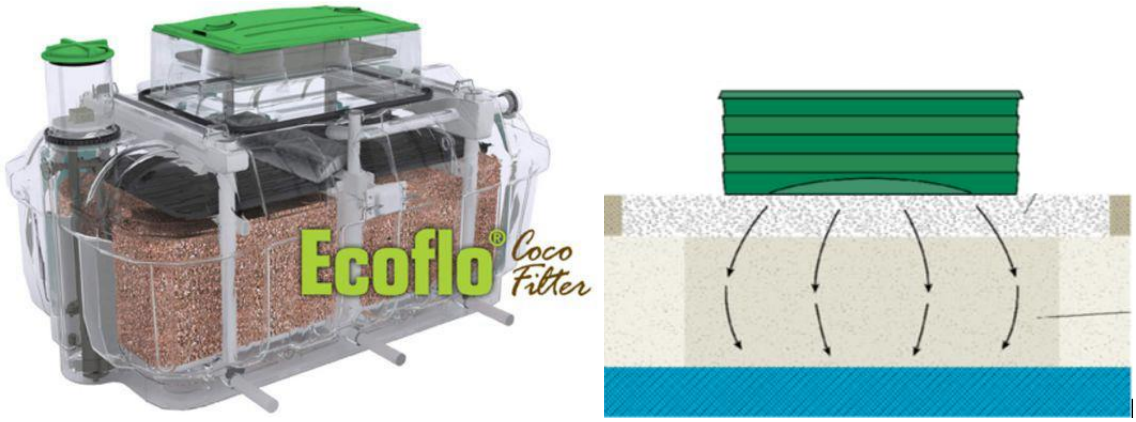


Figure 3.4. Schematic of packaged media filter (Ecoflo).

Soil Treatment Unit Effluent is divided into a four percolation pipes using a distribution box, (**Figure 3.2**). Treatment of wastewater is achieved as wastewater seeps into a washed gravel matrix which surrounds the percolation pipe. These stones should be washed and removed of fine particles

(Amador et Loomis., 2020; EPA., 2021). These stones provide support for the perforated pipe and storage of excess wastewater. The code of practice in the design of the STU percolation trench is a shallow, level excavation 850 mm in depth depending on the site, with a width of 500 mm (EPA., 2021). The gravel matrix is dispersed into the STU at a depth of 20-30mm and a percolation pipe (MFP HITEC length 18 m, diameter110mm) is laid on top of the gravel at a slope of 1 in 200 with a maximum length of 18m metres. As to protect the percolation trench from surface damage 450 mm of gravel is unloaded on top of the percolation pipe. A minimum of 1,200 mm of unsaturated subsoil is required between the base of percolation trench and the water table or bedrock. A semipermeable geotextile (Terram 1000) is then applied over the trench to prevent the likelihood of backfill.3.12 Crecora (Site B).

3.12 Crecora (Site B)

Research Site B is in Crecora, County Limerick. The site functions as a pastureland for horse grazing made up of homogenous vegetation. The subsoil is characterised as a highly permeable, gravel, sand subsoil according to British standards classification (BSI, 1999) consisting of glacial till of limestone origin. The domestic wastewater treatment system consists of a two chambered septic tank with a half of the flow then discharging into a Rotating Biodisc Contactor (RBC) treatment system. Both separate streams are dispersed into separate paired STU trenches (**Figure 3.5**).

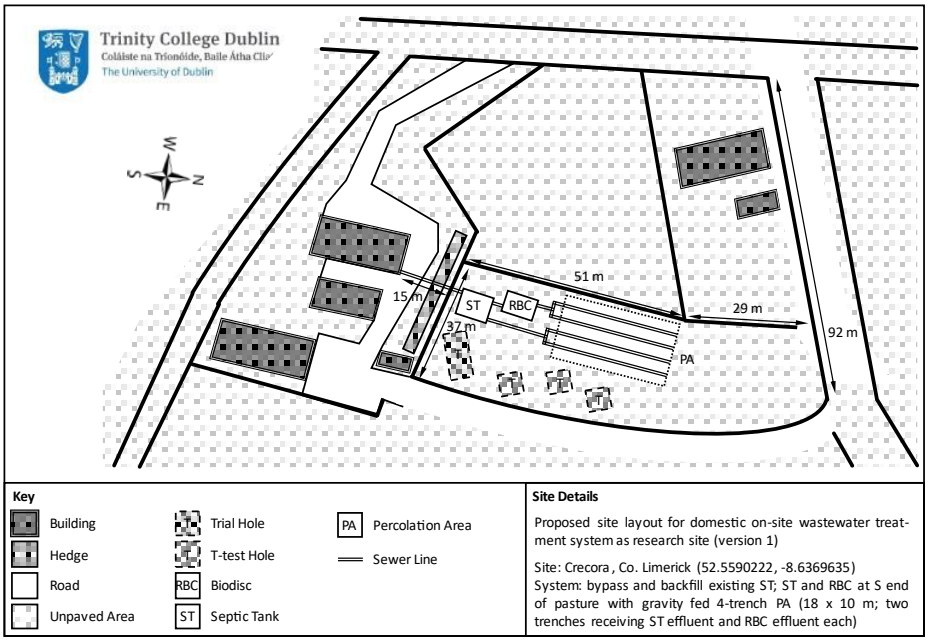


Figure 3.5. Layout of the Crecora site (Site B) and treatment system.

Primary Treatment is a two-chamber prefabricated concentrate septic tank (Aswasep Septic Tank NS 4 S Tank, Molloy Precast Products Ltd, the Republic of Ireland), the tank was designed under the guidelines set out by the EPA Code of Practice (EPA., 2021), as shown on Figure 3.6. Effluent is gravity fed into a distribution box (Molloy Precast Products Ltd., the Republic of Ireland) where it is split into separate flows using a tipping bucket, half flowing towards secondary treatment and

the other half is transferred to another distribution box where it is distributed evenly across both PE STU trenches.

Figure 3.6. Installation of septic tanks (left), RBC (centre) and STU percolation trenches (right)

Secondary Treatment Rotating biodisk contactors (RBC) function as a material for the growth and aggregation of microorganisms. The system consists of a primary settling chamber, the liquid fraction proceeds to the first stage of the biozone. The biozone first consists of small tank with a rotating disk of corrugated polypropylene disks with a total surface area of 72.5 m^2. A rotating shaft (~2 rpm) moves the disks and a cup which transfers effluent from the first biozone to the second where it undergoes treatment from a second set of rotating disks. In the treatment process 40% of the rotating disks is immersed into anoxic wastewater, as a portion is exposed to the atmosphere creating an aerobic environment. This rotation between oxygen states promotes the microbial degradation of wastewater constituents. Over time biofilm is formed as microorganisms aggregate on the surface on the disk, the biofilm is then sloughed from the disk and creates a sludge as it settles on the bottom of the tank (**Figure 3.7**). These systems have a maximum flow rate of 50 L/hr, which if a peak surpasses this limit excess flow would proceed to the series of percolation trenches.

Figure 3.7. Klagester Biodisk RBC. Beginning left, shows the schematic of the RBC system including the integrated primary settling chamber, two-stage biozone, and secondary clarification chamber. The middle figure displays the device upon installation prior to the development of a biofilm. The figure right displays the device following 24 months of biofilm development, the bottom disks display stage 1 biozone and the upper biodisk displays stage 2, with notable differences in colouration between both biofilms.

STU The effluent is divided into four percolation pipes using a distribution box (**Figure 3.6**). The STU was constructed to the same specification as the STU in Site A, i.e. according to the EPA Code of Practice.

3.13 Wicklow (Site C)

Research Site C is in Redcross, Wicklow. The site is located in a field used as pasture for dairy cattle grazing made up of homogenous vegetation. The subsoil at the site is characterised as an extremely permeable fine loamy drift to British standards classification (BSI, 1999) with siliceous stones with a clayey sandstone and shale till subsoil. The domestic wastewater treatment system consists of a two chambered septic tank with a split flow discharging either into a pump sump or directly into 2 STU trenches which are receiving PE effluent. The sump upstream of the sand filter delivering pulses of effluent to the passive treatment system. The pump (DOC3 Lowara, Xylem Water Solution Ltd., Republic of Ireland) were of a specification of 145 L / min, (for heads of less than 7 m), capable of carrying suspended solids of a diameter of 10 mm, 220 V 50 Hz 2 pole, 2850 rpm. The pump was controlled by a float switch which triggered once a certain level was reached in the chamber, this allowed for intermittent dosing of sand filter. Following treatment in the sand filter effluent is distributed into two SE STU trenches by means of a distribution box.

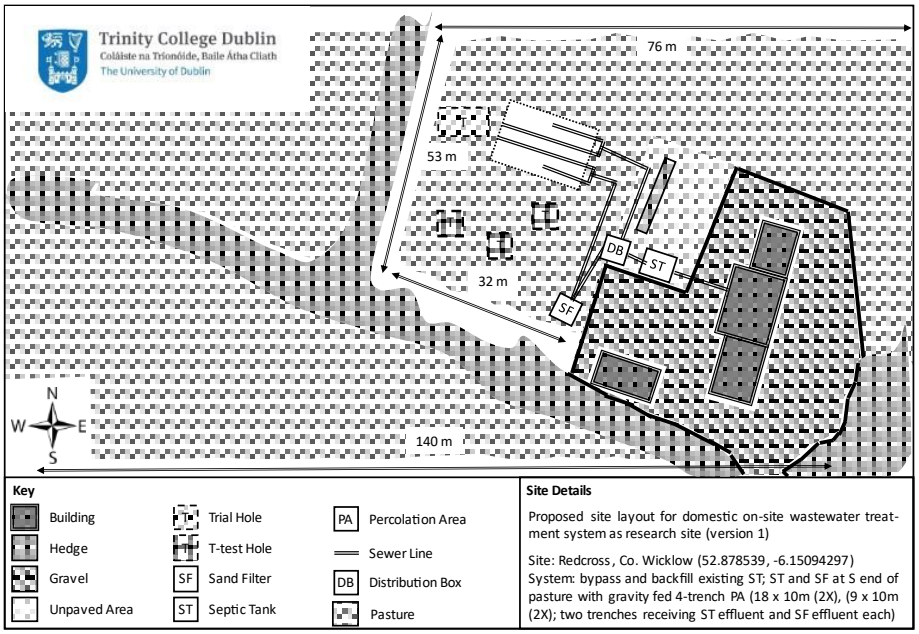


Figure 3.9. Layout of the Wicklow site (Site C) and treatment system.

Primary Treatment is a two-chamber prefabricated concentrate septic tank (Aswasep Septic Tank NS 4 S Tank, Molloy Precast Products Ltd, the Republic of Ireland), the tank was designed under the guidelines set out by the EPA Code of Practice (EPA., 2021). Effluent is gravity fed into a distribution box (Molloy Precast Products Ltd., the Republic of Ireland) where it is split into separate flows using a tipping bucket, half flowing towards sump to be pulsed into the intermittent open sand filter and the other half is transferred to another distribution box where it distributed evenly across both PE STU trenches (**Figure 3.10**).

Figure 3.11. Installation of septic tanks (left), sand filter sump relative the septic tank and household (centre) and STU percolation trenches relative to the sump position (right).

Secondary Treatment is an open intermittent sand filter - i.e. with no cover soil. The intermittent sand filters is a non–recirculating designed for the development of biofilms (EPA., 2009). The structure of the sand filter consists of several beds of graded sand, depths designed to 700 - 900 mm, with the base of the filter consisting of 200 mm of filter or washed pea gravel (EPA.,2009). The sand grain sizes used was D_{10} range from 0.4 mm to 1.0 mm, as per EPA (2021) design criteria. The manifold is a 10 mm inner diameter, 150 mm width 200 mm length (**Figure 3.11**).

Figure 3.11 Construction of the open intermittent sand filter at Site C from its basic frame and sloped base and collection pipe to the STU (left), controlled addition of the percolation sand (centre), addition of percolation pipe manifold over gravel surface (right).

STU effluent is divided into four percolation pipes using a distribution box (**Figure 3.12**). Again. the STU was constructed to the same specification as the STU in Site A, i.e. according to the EPA Code of Practice.

Figure 3.12. Construction of STU trenches at Site C (left) placement of percolation pipe (centre-left), application of geotextile membrane (centre right) and backfill with pea gravel and topsoil (right).

3.2 On-site Instrumentation

3.2.1 Automated data collection

All sites were installed with an automated weather station containing a CR1000 data logger, (Campbell Scientific USA) as shown on Figure 3.13. The system collected data on an hourly basis for mean air temperature *T*, barometric pressure *p*, net radiation *Rnet*, rainfall *r*, relative humidity *U*, wind direction *wd*, windspeed at a height of 2 m u_2. The weather stations were installed in the centre of the STU system, within 5 m distance of the inlet into the trenches. A combined temperature – humidity probe was encased in a radiation shield to prevent any misreading from thermal heating. The installation of the net radiometer was achieved by the ensuring it was south facing to reduce the effect of shadowing. Instrumentation in the weather stations was summated in **Table 3.2**.

Flow was quantified using reed switches mounted to the tipping buckets within the effluent distribution boxes. Switches were triggered when the magnet connected to the bucket passes another mounted magnet which would close the open switch and each tip is counted (**Figure 3.14**). Counts were converted to specific volumes following calibration of the site. Calibrations were performed every year, using a graduated cylinder and a syringe. The tipping buckets are filled and subsequently tipped three times in each direction as to calculate the mean volume of each basin. Tipping buckets were inspected for biofilm growth which could deviate flow volumes. Soil moisture sensors were connected to two-channel relay multiplexers (AM16/32b, Campbell Scientific USA). Stations were powered by mains supply and a back-up 7 A h lead battery as a fail-safe (NP7- 6 Yuasa Corporation, Japan). Data retrieval was performed using the LoggerNet v4.5 software, data was stored as text files for further processing.

Figure 3.13. Weather station during site build with multiplexer configuration barometer and power supply (left) soil moisture cables attached to multiplexers, backup battery supply and barometer (centre), complete installation on-site (right).

Figure 3.14. Installation of distribution boxes during site C build inlet effluent pipe is visible and leading towards one of the boxes (left) distribution box prior to installation with tipping bucket present (centre) blue reed switch magnets and communication cable installed to tipping bucket and distribution box (right).

Table 3.2 Instrumentation used, data and metrics recorded with the automated weather stations on each site

Parameter	Symbol	Unit	Instrument	Aggregation[a]	Log interval[b]
case temperature[c]	*T case*	^{O}C	Type T Thermocouple	sample	H
air temperature	*T air*	^{O}C	Campbell CS215	mean, min, max	H-T
relative humidity	*U*	-	Campbell CS215	mean, min, max	H
atmospheric pressure	*P*	mbar	Vaisala PTB110	sample	H
rain fall	*R*	mm	EML ARG100	total	H-T-M
Wind speed	*u2*	m/s	Young 05103-5	mean	H-T
Wind gust speed[d]	u_{gust}	m/s	Young 05103-5	max	H
Wind direction	*ud*	^{o}	Young 05103-5	mean	H-T
Net radiation	*Rnet*	W/m^2	Kipp & Zonen NR-Lite 2	mean	H
Wastewater flow [e]	*Q*	L	reed switch tipping bucket	total	H-T-M
Soil volumetric water content	Θ	-	Decagon EC5/GS3	sample	H(-T-M)[f]
Soil temperature	*Tsoil*	^{O}C	Decagon GS3	sample	H-T
Soil bulk electric conductivity	*Kbulk*	µS/m	Decagon GS3	sample	H-T-M

[a] sample: sample at time of logging; mean: mean of one-minute interval samples (one-second intervals for wind speed, wind direction, wind gust); min: minimum of on-minute interval samples; maximum of one-minute interval samples; total :sum of additive readings per logging interval.

[b]H: hourly interval; T: ten-minute interval; M: one-minute interval; used as approximation *Tair* in case of combined temperature-humidity probe failure; measured as maximum average wind speed over three-second intervals; separated into PE and SE flow; hourly logging for EC5 sensors only

3.2.2 Soil moisture sensors

Two different types of soil moisture sensors, typed EC5 and type GS3 (METER Group Inc., USA, former Decagon Devices, Inc., USA), were installed below the infiltrative surface of the percolation trenches within the STU, but also exterior to STU system as a control (**Figure 3.16**). These sensors allowed for the direct measurement of volumetric water content, soil temperature, and electrical conductivity over time.

Soil moisture sensors were installed to create an effective monitoring network and have been shown to be able to determine the growth of microbial biomat within the STU, by measuring the relative changes to the STU volumetric water content. Clogging has bene assumed when there is an increase of VWC of 2.5% over 30 days (Knappe et al., 2020).

EC5-type sensor

The EC5 is a two-prong capacity sensor that measures the relative dielectric permittivity ϵ_r in the soil by measuring the charge time of a hypothetical capacitor. The sensor is composed of two plastic prongs which act as electrodes and turn the surrounding soil as the dialectic of the capacitor. The variation of ϵ_r in soils in differing states of saturation which can then be calibrated soil moisture content employing a site sourced soil soil-specific standard curve (Topp et al., 2007; Decagon, 2015a). The output of the sensor is a voltage signal in units of V which is proportional to the measured charge time (**Table 3.3**). Each sensor has an estimated zone of influence in which measurement can be inferred: the zones for the EC5 is a cylindrical volume of ~0.85 L near the sensor's radial centre (**Figure 3.17**). The signal is strongest to the physical sensor, which may induce bias if the sensor has large voids induced in the soil matrix during installation.

Table 3.3 Technical specification of EC5-type soil sensor

General Specifications	
Dimensions	2 cm L x 2.5 cm W 9 cm H
Dielectric measurement frequency	70 MHz
Measurement time	150 ms
Power requirement	3.6 to 15 VDC
Operating temperature	-40 TO 60º C
Volumetric Water Content	
Accuracy	±0.02 m^3/ m^3
Resolution	0.001 m^3/ m^3
Range	0.0 to 1.0 m^3/ m^3
Volume of influence	240 mL

GS3-type sensor

Due to the cost and complexity, one of these sensors were installed into Site C, outside of the percolation field to measure soil temperature, and volumetric water content of the subsoil porewater. Its technical specifications are listed in **Table 3.4**. Temperature was measured using a thermistor connected to the prong surface with output value ^{O}C. Volumetric water content is measured in the

same manner as the EC5 sensor only with a smaller zone of influence of 0.55 L which is centred around the outermost prong (**Figure 3.18**).

Table 3.4 Technical specification of GS3-type soil sensor

General Specifications	
Dimensions	2 cm L x 2.5 cm W 9 cm H
Dielectric measurement frequency	70 MHz
Measurement time	150 ms
Power requirement	3.6 to 15 VDC
Operating temperature	-40 TO 60^{o} C
Volumetric Water Content	
Accuracy	±0.03 m^3/ m^3
Resolution	0.002 m^3/ m^3
Range	0.0 to 1.0 m^3/ m^3
Volume of influence	160 mL
Temperature	
Accuracy	± 1^{O}C
Resolution	0.1^{O} C
Range	-40 to 60 ^{O}C

Soil sensor calibration

The EC5-type sensor determines the average relative permittivity ϵ_r of the surrounding soil to transmit electromagnetic waves within a certain volume of influence, while the GS3-tpe sensor (Decagon Devices) also have the ability to measure both soil temperature and soil temperature. Both sensors were calibrated using homogenous subsoil from each of the research sites (**Figure 3.15**), which was passed through an 0.8 mm filter and air dried.

Figure 3.15. Subsoil retrieval for calibrations ensuring that sufficient depth is reached for a comparable soil matrix (left) calibration of EC5 sensors were used in triplicate (centre) and moisture was confirmed using wet dry weighing (right).

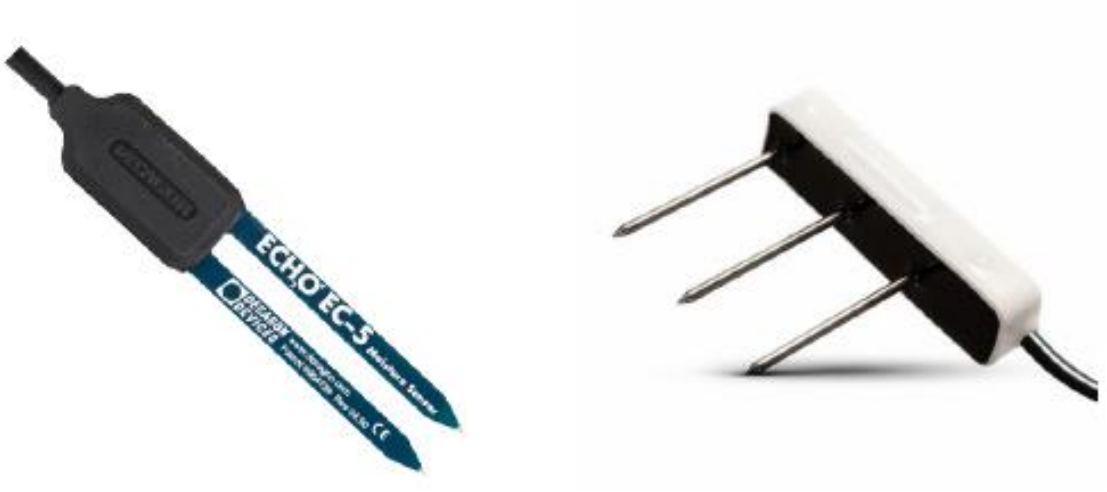


Figure 3.16. Soil sensors EC5 (left) and GS3 (right) both used in this study, EC5 measures soil water content and the GS3 measure soil water content and temperature. Images from (Decagon, 2015a,b).

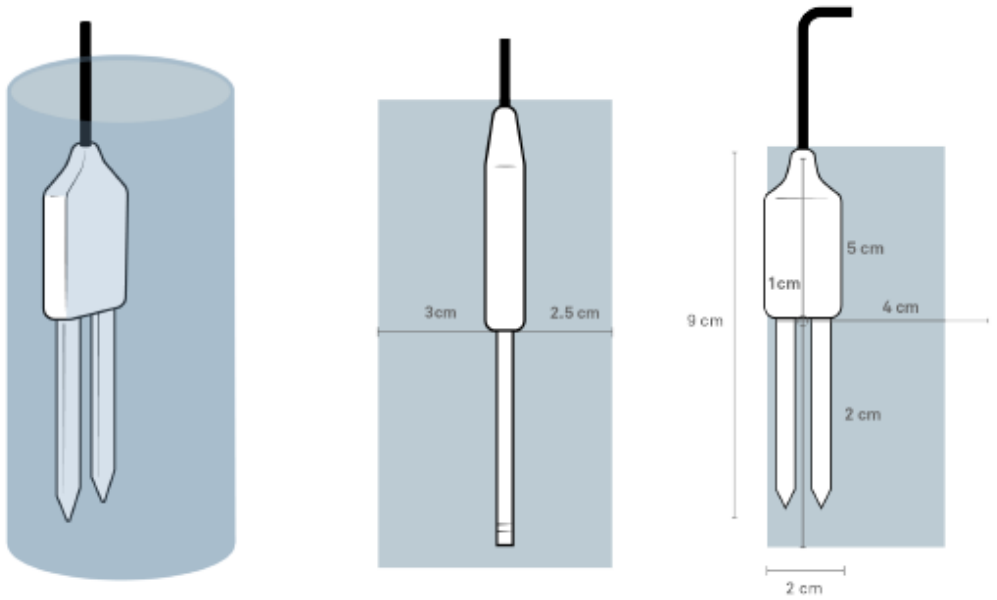


Figure 3.17. Zone of influence of EC5 soil moisture sensor (Cobos,2014)**.**

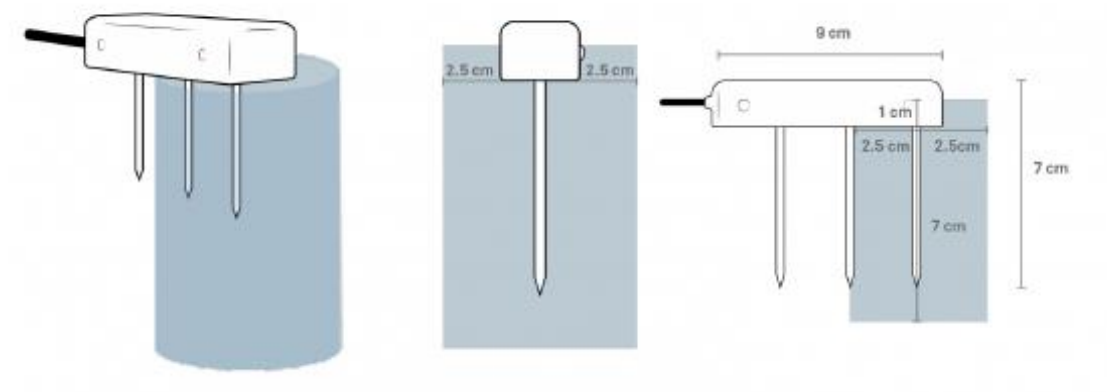


Figure 3.18. Zone of influence GS3 (Cobos,2014).

Soil sensor installation

Soil moisture sensors were installed at a range and depths and across the length of the STU percolation trenches (**Figures 3.20, 3.21** and **3.22** for soil sensor positions at Sites A, B and C, respectively). At each sensor position a 10 cm diameter hand auger was utilised to allow for desired installation depth of the sensors, positioned in the wall of the hole at depths of 5, 10 and 15 cm, as shown on Figure 3.23. It was crucial that sensors be installed laterally in such a manner as to not allow water to gather on the surface of the prongs resulting inaccurate readings. The sensor cavities were filled using a bentonite clay slurry to prevent preferential flow down through the cables.

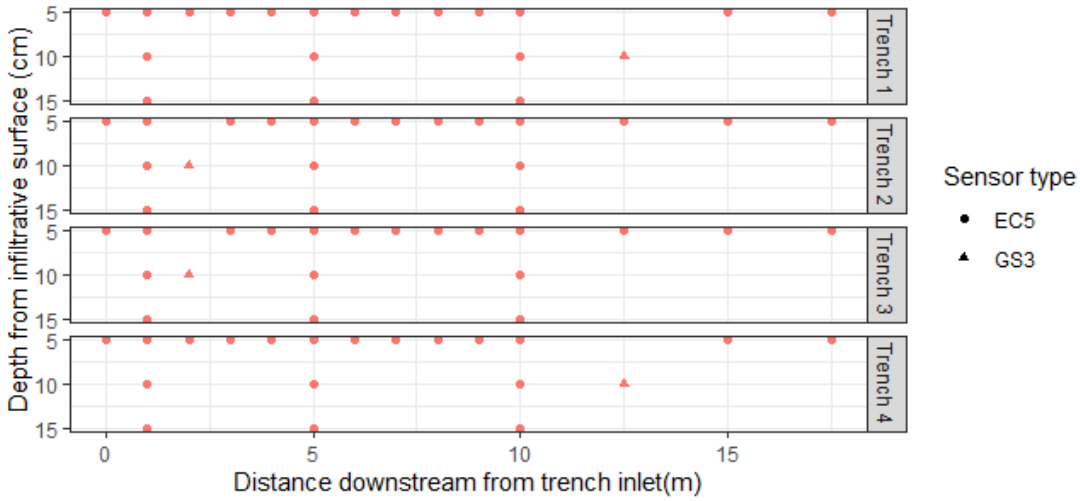


Figure 3.20. Positions of soil sensors installed within the STU in Site A. Circles represent EC5-type soil sensors capable of measuring soil VWC, triangles represent GS3-type sensors capable of simultaneous measurements of soil VWC, soil temperature, and pore water EC. Additionally, control EC5-type soil sensors are installed in natural soil adjacent to the STUs control at depths of 10, 25, 50, 75, 100, and 125 cm.

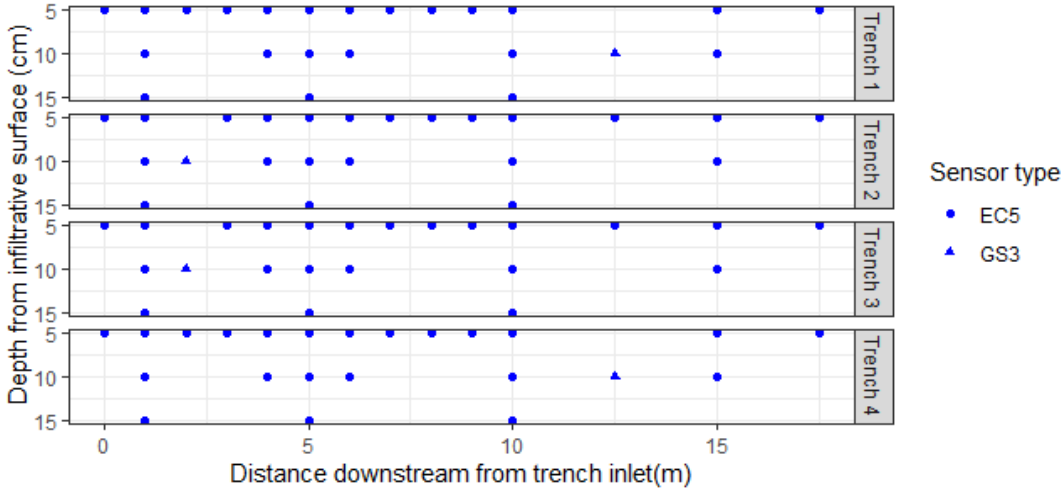


Figure 3.21. Positions of soil sensors installed within the STU in Site B. Circles represent EC5-type soil sensors capable of measuring soil VWC, triangles represent GS3-type sensors capable of simultaneous measurements of soil VWC, soil temperature, and pore water EC. Additionally, control EC5-type soil sensors are installed in natural soil adjacent to the STUs control at depths of 10, 25, 50, 75, 100, and 125 cm.

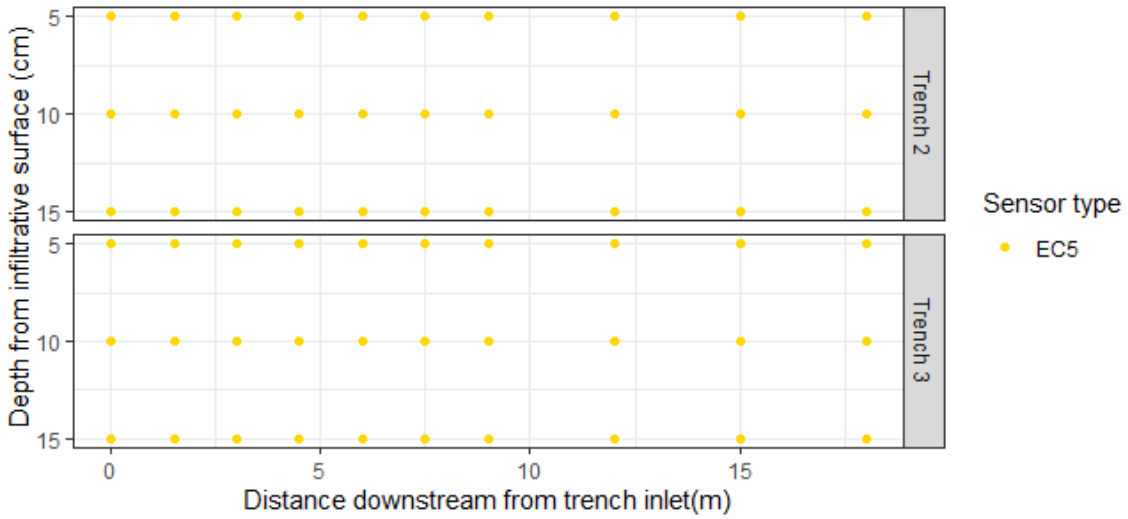


Figure 3.22. Positions of soil sensors installed within the STU in Site C. Circles represent EC5-type soil sensors capable of measuring soil VWC, triangles represent GS3-type sensors capable of simultaneous measurements of soil VWC, soil temperature, and pore water EC. Additionally, control EC5-type soil sensors are installed in natural soil adjacent to the STUs control at depths of 10, 50, and 100 cm.

Figure 3.23. The auguring of 10 cm diameter cavity for the soil sensors (left) placement of vertically aligned trenches within the trench (centre) and addition of sealing bentonite clay within the cavity (right).

3.2.3 Suction cup porewater lysimeters

For the sampling of pore water percolating from the unsaturated zone beneath the infiltrative surface of the soil treatment unit, suction cup lysimeters (Model 1900 Soil Water Sampler with z1900-200 Stopper Assembly, Soilmoisture Equipment Corp, USA) were installed at a varying range of lateral and horizontal series, to give a full porewater profile of the trenches. The lysimeters consist of a 4.8 cm diameter PVC tubes of varying lengths with a porous ceramic cup at the bottom and TPV stopper with a neoprene tubing at the top. The ceramic cup is designed with a pore radius of 1.3 µm and an air entry value of 200 kPa (Soilmoisture, 2007). Installation of the lysimeters took place

during the construction of the STU whereby lysimeters were placed in pre-augured holes at various depths (10, 30, 50 cm) below the infiltrative surface in the percolation trench.

When the lysimeter is inserted into the soil, it is crucial to ensure that porewater can travel to the ceramic cup and into the pores, this is achieved by ensuring a tight fit with between the cup and the soil by dipping the cup in a soil slurry (5 mm sieved soil and tap water). This slurry is used to fill any space in the augured hole once the lysimeter is installed, finally capped with bentonite clay to ensure there is no preferential flow from precipitation the exposed tube at the surface.

Porewater Sampling

The method for sampling of porewater from the lysimeter within the STUs and lysimeters installed exterior to the STU system required the application of a vacuum at 50 kPa of pressure. The vacuum was created through a vacuum-pressure foot pump (Model 2005G2, Soilmoisture Equipment Corp., USA). The lysimeter assembly includes a stopper and clamping ring to maintain the suction through the 1-2 days of sampling. Samples are retrieved by removing the tubing seal and inserting a sampling tube and extracting the liquid into a conical flask again by producing vacuum with a foot pump (Model 1900K3 Extraction Kit, Soilmoisture Equipment Corp., USA), as shown in Figure 3.24. Residual suction and extracted volumes were noted during sampling campaigns, to determine issues in sampling, whereby no residual pressure (i.e. 0 kPa) would indicate a broken ceramic cup, damaged tube, or loose seal.

Figure 3.24. Priming lysimeters for sampling (left) porewater collection flask and collection tubing (centre) porewater collection diagram illustrating the flow of the effluent from the percolation pipe through the gravel, the direction marked with the red arrows towards the yellow area around the cup of the lysimeter shows the zone of influence is to be extracted using the Erlenmeyer flask and foot pump at the surface level (right).

The positions and depths selected for the installation of the lysimeter were chosen to confer a better understanding of wastewater infiltration and biogeochemical processes along the entirety of

the trench up to a depth of 55 cm. However, the subsoil matrix also determined the positioning as the presence of stones hindered the installation of some lysimeters within the trenches.

Sites

Site A: 53 lysimeters were installed in the STU in Site A (Kilmallock). These were distributed across the 4 trenches with 27 in trenches receiving primary effluent, and 26 in trenches receiving secondary effluent (**Figure 3.25**).

Site B: 52 lysimeters were installed in the STU in Site B (Crecora). These were distributed across the 4 trenches with 25 in trenches receiving primary effluent, and 27 in trenches receiving secondary effluent (**Figure 3.26**).

Site C: 35 lysimeters were installed in the STU in Site C (Redcross). These were distributed across the 4 trenches with 17 in trenches receiving primary effluent, and 18 in trenches receiving secondary effluent (**Figure 3.27**).

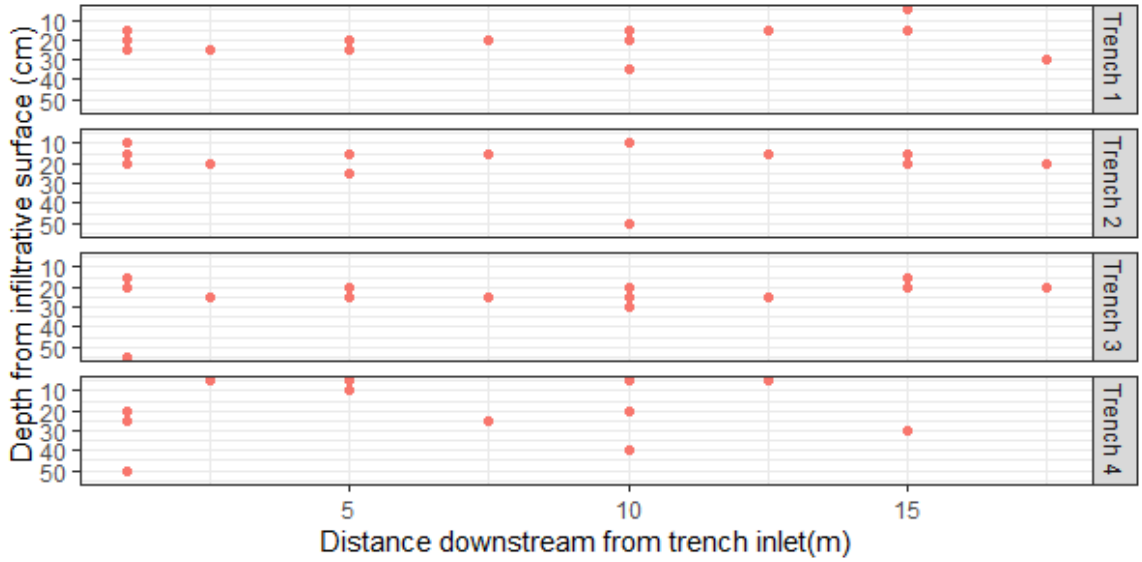


Figure 3.25. Positions of lysimeters installed in site A (Kilmallock). Trenches 1 and 2 receive primary effluent (PE); trenches 3 and 4 receive secondary effluent (SE).

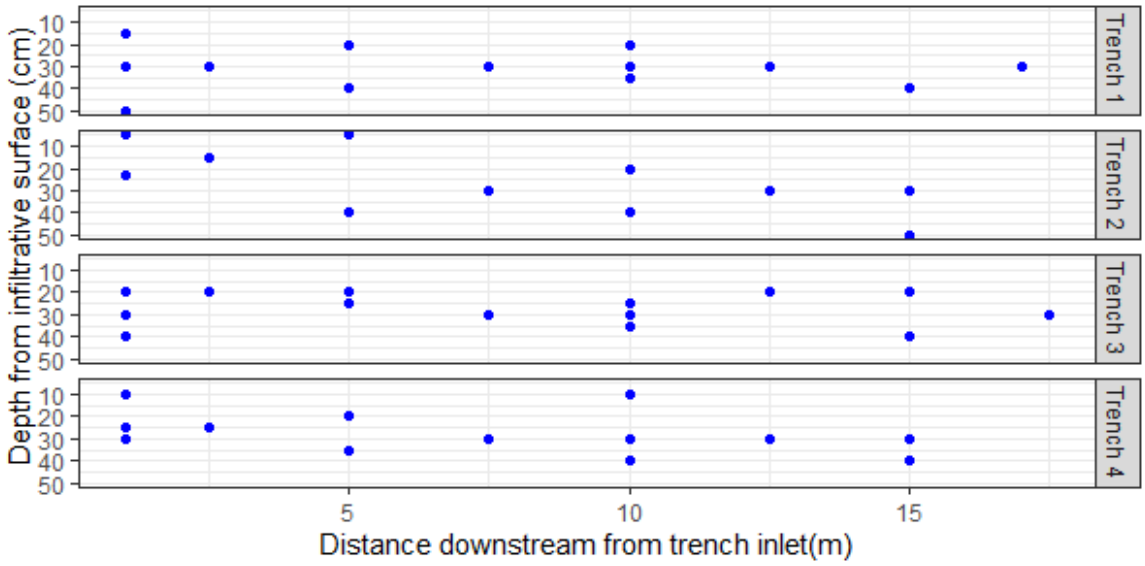


Figure 3.26. Positions of lysimeters installed in site B (Crecora). Trenches 1 and 2 receive primary effluent (PE); trenches 3 and 4 receive secondary effluent (SE).

Figure 3.27. Positions of lysimeters installed in site C (Redcross). Trench 2 receives primary effluent (PE); trench 3 secondary **effluent (SE).**

3.2.4 Bioports and Installation

Sampling the biomat requires accessing the STU infiltrative surface post construction and while the system is active. Gaining access to Sites A and B (which had already been constructed prior to this research) presented significant challenges due to multiple reasons: the exposure to biohazardous effluent from the system; the difficulty in excavating gravel with trowels without gravel collapsing back into the excavated holes which meant that 30 mm pipes were needed to reduce such infilling; and obstructions such as stones or boulders at the STU infiltrative surface preventing a sample being taken. This procedure was carried out at Sites A and B which were only sampled twice each over a 4-year period, but for Site C which was to be sampled 8 times over in less than three years, this

procedure was not deemed feasible and could have compromised the STU entirely. Therefore, access ports to the STU biomat or "Bioports" in the context of this study were installed.

Bioports consist of PVC 200 mm diameter and 900 mm length pipe with a cap to seal it at the surface. The function of the installed bioports is to allow for on-line microbial survey of the biomat without significantly impacting the structural integrity of the system. This was done by the creating a tripod flanged female end which would be installed into the trench at a depth of 2 cm of the infiltrative surface. The tripod would be designed with 10 cm wide sections exposed (**Figure 3.28** and **Figure 3.29**) to permit the percolation of water from the trench to allow the effluent to effect on the soil biome.

Figure 3.28. Construction of bioports with 100 mm wide inlets to permit effluent transport into the bioports area (left) profile of the bioports including port access lid (centre-left) installation of the port within the STU percolation trench (centre right), positioning of 1 m and 5 m bioports for trenches 2 and 3 within the Site C post construction.

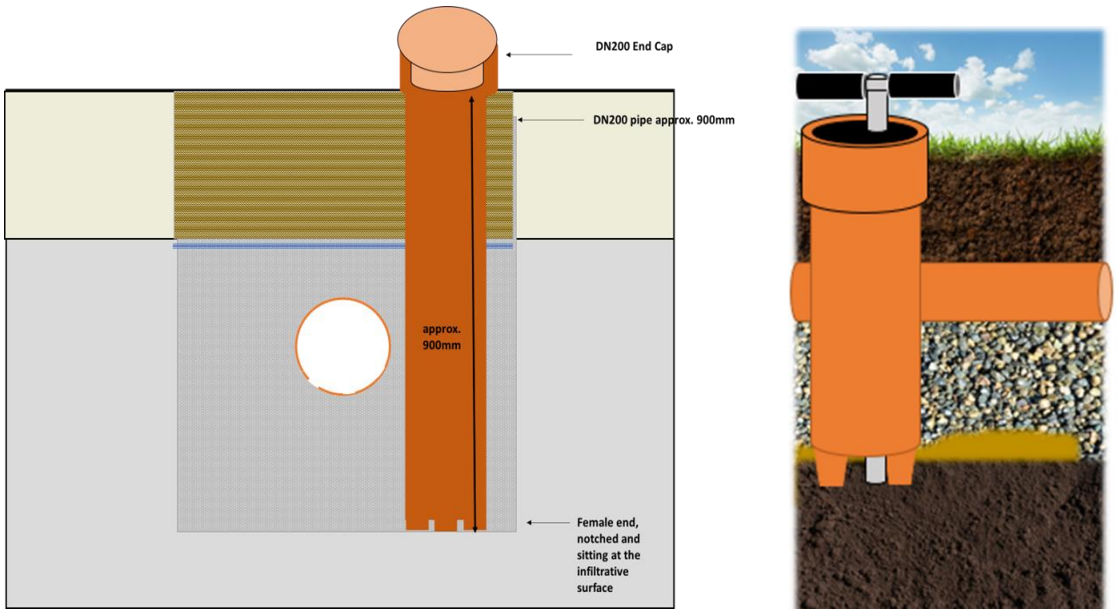


Figure 3.29. horizontal front-facing profile of the STU trench with bioports dimensions and labelled components (left) working principle of bioport diagram displaying position of horizontal positioning of bioports within the STU structure, the application of an corer within the bioport and the location of the biomat with the presence of effluent above the subsoil (right).

A total of six bioports were installed in the soil treatment area. Three bioports were added to one trench of each effluent type, at distances from effluent inlets of 0 m, 5 m, 17 m which allowed for a lateral comparison of the system biome in respect of the other measurements being taken in parallel. Samples were taken every six months, by auguring within the sites at different depths. Augured sections were backfilled with a slurry of soil taken from the site at a comparative depth, as shown on Figure 3.29.

3.3 Column Construction

The columns in the laboratory were designed from a pre-existing glass superstructure with 10 mm thickness. The superstructure was partitioned into four columns, with dimensions of 30 cm height x 30 cm width x 15 cm breath. The partition walls were created using 6 mm "MarCryl" Acrylic plastic sheet (Access plastics Ltd., Ireland). The columns were filled with a sandy loam subsoil excavated at 1.2 m deep at Site C in County Wicklow. For two of the columns, the Wicklow soils were blended to a ratio of 4:1 with percolation sand (1 mm: 2%, 90%: 600 µm, 250 µm: 8%) to create a higher permeability soil. The high permeability columns had a porosity of 26% and low permeability columns (just soil) had a porosity of 21%. Prior to packing the columns, the two types of soils were air dried for 24 hours and sieved at 6.30 mm. When packing the columns, batches of 5 kg of soil were poured into the column and subsequently compacted by a wooden tampering block, in total each column comprised of 23 kg of soil (**Figure 3.30**. The soil was filled to a height of 43 cm in the column structure, this was followed by 10-15 mm of pea gravel on top of the column to improve distribution of effluent. 10 L of distilled water was passed through each of the columns with a dosing

frequency of 1 L per hour prior to the initiating the column to ensure that the soil matrix was well settled.

Figure 3.30. The construction of the columns with the uniform tamping of soil medium at set intervals within the system (top-left) saturation of the chambers with distilled water prior to commencement of the experiment (top-right) effluent collection vessels at the base of the chambers (bottom-left) sample collection bottles and vacuum syringe for the retrieval of effluent from the collection vessels (bottom-right).

A network of four separate collecting vessels for the effluent coming through the soil in each column was created which were attached with flexible nylon tubing with a 2.5 mm inner diameter (Flomax., Ireland). The effluent was collected into a bottom drainage tank which is a part of the glass superstructure and is separated by a mesh to avoid soil particles from entering. The effluent is drained into plastic containers of 2000 ml quantity. These effluents were stored at 4 ^{O} C and analysed on every alternate day. The soil columns were intermittently dosed with 1 L per day of influent of differing degrees of treatment. Dosing was performed at 3-hour increments of 250 ml at 09:00, 12:00, 15:00, 18:00 to mimic household flow pulses for a total flow of 1 L d-1 and hydraulic loading rate of 22.22 L/m²/d. The pumps were controlled by an Arduino Uno system (Arduino LTD,Italty), which was wired to A4988 four stepper motors through a CNC shield V3 (AZ-Delivery, Germany) which was programmed using Arduino software. The pump system consists of four separate manifolds consisting of figure-of-8 shape PVC pipe (Chadwicks., Ireland) distribution manifolds of length of 20 cm and width of 8 cm and an inner diameter 10 mm. The manifolds have multiple perforations on top to allow for even distribution of effluent across the surface of the column. Each

of the four column manifolds were fed with either PE or SE influent with Vikyehxn812w6ae model stepper motor peristaltic pumps (Vikye., Italy) through Tygon S3 E-3603 non-contaminating 2 mm inner diameter peristaltic tubing (Saint-Gobain., France) (**Figure 3.33**). All pumps and manifolds were individually calibrated for their specific dimensions.

Column influents were collected from the wastewater treatment systems of Site B which employs a 4,760 litre Aswasep two-chambered septic tank (Molloy Precast Products Ltd., Ireland) as primary treatment (PE) and a packaged secondary treatment system an RBC (Klargester BioDisc, Kingspan., Ltd, United Kingdom), consisting of an integrated primary settling chamber, a two-stage biozone, and a secondary clarifying chamber as secondary treatment (SE). Each column had a soil moisture monitoring network using sensors (EC5, METER Group Inc., USA) installed laterally and horizontally at 10 cm intervals, connected to a datalogger (CR800, Campbell Scientific, U.K) (**Figure 3.32**). Each column had a designated collection bucket for the storage of its filtrate for further analysis. Each soil column was subdivided into zones relative to depth defined as follows: 0 – 1 cm is the infiltrative zone, 1 – 5 cm proximal zone, 6 -10 cm intermediary zone and 11 – 30 cm is the deep zone of the subsoil columns (**Figure 3.31**).

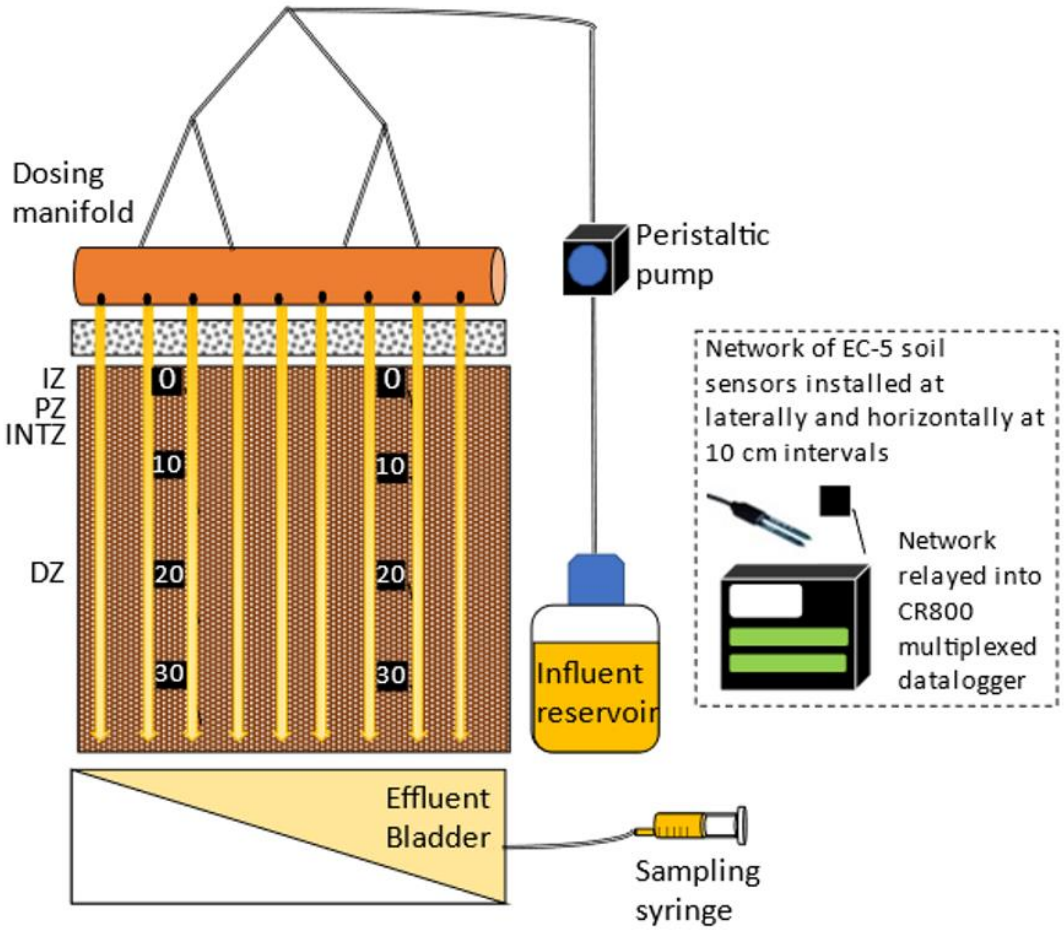


Figure 3.31. Schematic of a single soil column with positions within the column;0 – 1 cm is the infiltrative zone (IZ), 1 – 5 cm proximal zone (PZ), 6 -10 cm intermediary zone (INTZ) and 11 – 30 cm is the deep zone (DZ) of the subsoil columns. Also labelled is the peristaltic pump and influent reservoir and effluent collection vessel or "bladder". Datalogger and installed sensors are marked with black squares.

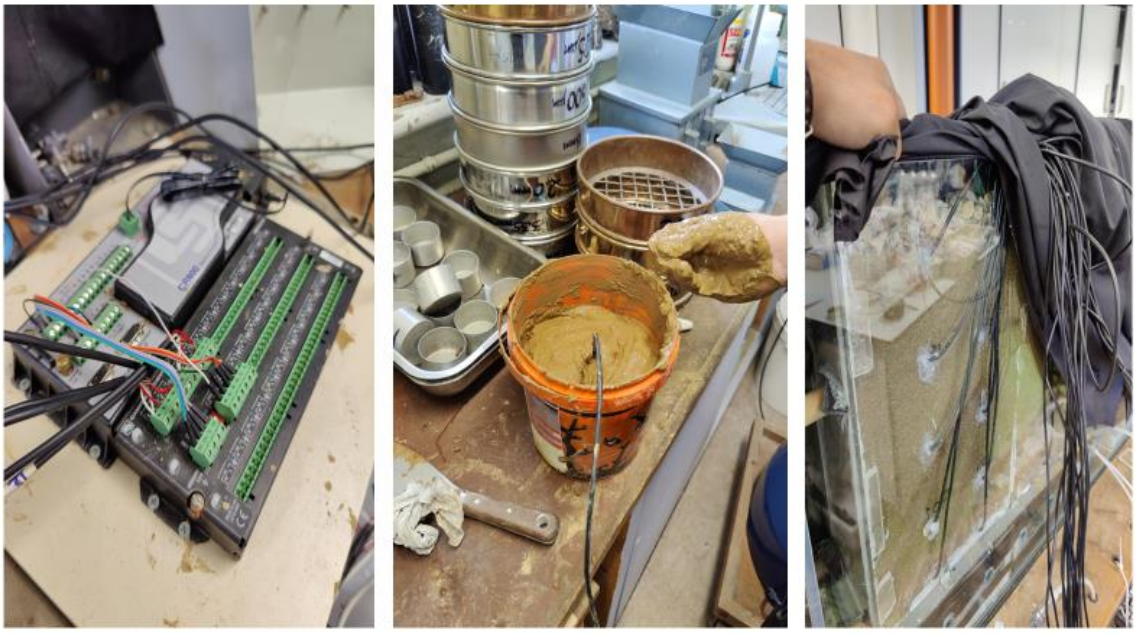


Figure 3.32. CR800 datalogger and multiplexer for soil sensor installation (left) soil sensor calibration with soil in a known volume at a certain stage of saturation with sampling cups for weighing after measuring voltage reading (centre), sensor network viewed from the rear of the column with silicon seals applied at each position.

Figure 3.33. Soil column 8 shape manifold with peristaltic tubing (left) manifolds being calibrated with weighing scale and pump manifold (centre-left) manifold positioned within the column (centre right) influent reservoirs for PE and SE influent attached with peristaltic tubing (right).

3.4 Sampling protocol

3.4.1 Sample collection

A total of 2, 2 and 36 sampling campaigns took place in site A, B and C respectively. Sampling campaigns in Sites A and B focused primarily on the subsoil biomat, whilst in Site C the subsoil biomat sampling was conducted alongside lysimeter porewater sampling. The sampling campaign for Site A and B took place in August 2018 and 2022, and for Site C the sampling campaign commenced in September 2020, however regular sampling was challenging due to the COVID pandemic and anti-infection measures.

3.4.2 Biomat Subsoil Field Sampling

In Sites A and B in August 2018 and 2022, a total of 136 and 138 samples were taken in Site A and B, respectively. Site C was sampled 8 times in this research study with a total of 146 samples taken. Biomat samples were then taken with a 25.4 mm stainless steel corer, as shown in Figure 3.34. In Sites A and B in 2018 samples were taken from infiltrative surface at depths of 0, 2.5, 5, and 7.5 cm samples at distances 1 m, 5 m, and 12 m. This was expanded in 2022 with additional depths reaching up to 12.5 cm below the infiltrative surface and further downstream of the inlet at 17 m. In Site C samples were taken from the permanently exposed subsoil within the bioports at positions 1 m, 5 m, and 17 m. Control subsoil samples were taken at locations adjacent but exterior to percolation trenches at depths ranging from 80 – 130 cm. Effluent samples were taken from all sites and refrigerated at $4^{O}C$ prior to concentrating. Sample handling was performed with a sterile metal spatula, with sterilising being performed between each sample with 70% ethanol. Efforts were made to prevent cross contamination when sampling in the field, with the use of small portable working enclosures and use of portable Bunsen burners. For DNA extraction of samples for sequencing, approximately 3 g of soil were collected from the corer and placed into a sterile 2 ml Eppendorf tube and stored at -20 for samples taken in Site A and B during the 2018 sampling campaign and 80 ^{O}C for the remaining sampling short and long-term campaigns.

Figure 3.34. Subsoil excavation, with gravel and topsoil separation for effective refilling of cavities (left) subsoil core sample with dark black organic indicative of biomat growth (centre) temporary shelter for sample processing including a refrigerator for sample storage.

3.4.3 Biomat soil column sampling

Soil column samples were taken from each column with a 25.4-mm stainless steel corer. Samples were taken after 510 days of the system performance was stable and running at a steady state. At each column, a single sample was taken in each of the depths from the infiltrative interface (0-7 cm, 10 cm, 20 cm, and 30 cm). Subsoil control samples were taken from both subsoil source site and the percolation sand mixed with the subsoil for two of the columns. Following sample retrieval, a bentonite slurry was poured into the cavity produced by the corer to prevent preferential flow and to permit the columns to continue at their steady state.

3.4.4 Lysimeter Porewater Sampling

Porewater samples were labelled on both the sample vessel and cap to avoid miss-handling and contamination. Field blanks were taken using distilled water taken from Eerlenmeyer flask to determine trace contamination even after rinsing between samplings. Quality control samples were included in sample analysis in lab workflows. 50 ml centrifuge tubes were used (Lennox Laboratory Supplies, Ireland) to transport samples. After each sampling campaign the sample vessels were washed with bactericidal, phosphorus free detergent (decon90, Decon Laboratories Limited, UK) and rinsed with distilled water. Samples were stored in an insulated storage container with reusable ice packs to assure $4^{O}C$ in transportation.

3.5 Pepper Mottle Virus Faecal Indicator Monitoring Sampling Campaigns

A component of this research was to determine the effectiveness of PMMoV as a faecal indicator for on-site wastewater treatment systems. This was achieved by means of spiking trials carried out on the low permeability soil column receiving PE influent and the PE STU percolation trench on Site C. A spiking trial in one of the soil columns was conducted following 710 days of operation, when the column had been in a steady state which would best mimic system functionality. The spiking solution was prepared by means of mixing commercial Tabasco® pepper product (McIlhenny Company, USA) with PE effluent from Site C at a concentration of 1 part tabasco for 10 parts effluent for days 1 – 3, but which reduced to 1 part tabasco 100 parts effluent from days 4-5. of the spiking trial. The effluent loading rates of the column continued remained as initially designed. Effluent samples were collected at the reservoir beneath the column with samples retrieved once per day prior to the first dose to allow for 24-hour intervals in sampling.

The field scale spiking trial took place in the Site C PE fed STU percolation trench, which was monitored for 5 days to determine the level of PMMoV transport from the treatment system. The spiking trial took place in the third year of the system's operation. The spiking solutions consisted of the household's primary effluent (septic) mixed with Tabasco pepper product (McIlhenny Company, USA) at 1:100 dilution. Porewater samples were taken at three different positions at 1 m, 8 m, and 17.5 m from the trench inlet, at depths of 10, 30 and 50 cm for each of these positions, as shown on **Figure 3.35**. An additional spiking solution sample was taken each day to determine persistence within the influent. The effluent and porewater samples extracted from lysimeters were stored on ice for <2 h of transport to be analysed in the environmental engineering laboratory at Trinity College Dublin.

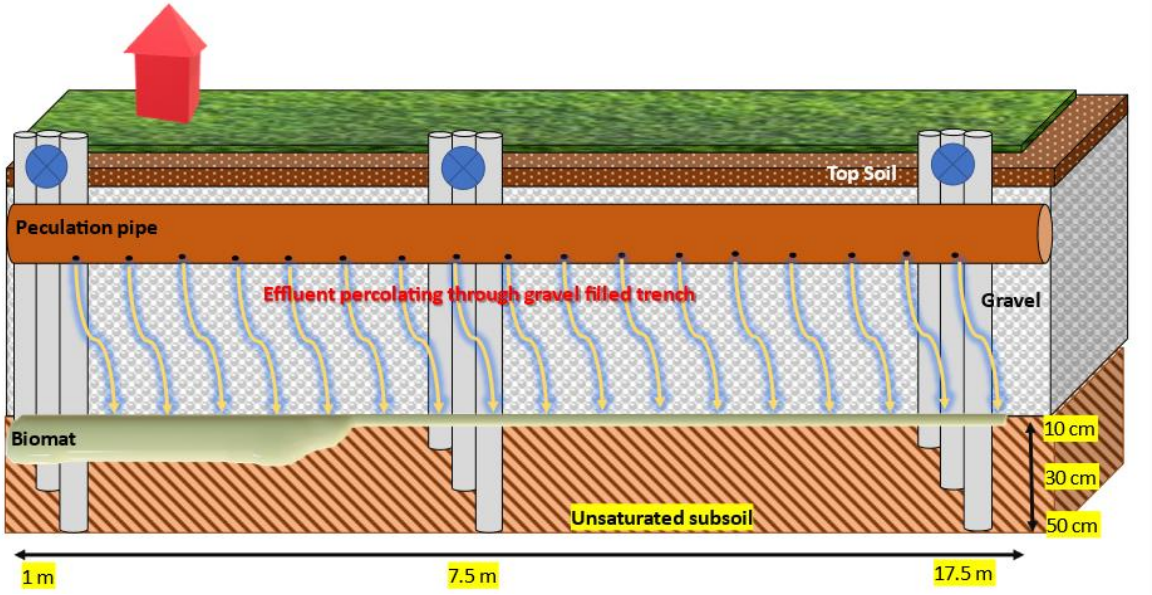


Figure 3.35. Percolation field schematic describing the microbial biomat, septic effluent transport, the position and depth of porewater lysimeter (marked with blue circled X).

3.5.1 Sampling rivers impacted by on-site wastewater treatment systems.

Several catchments which had been previously characterised as having on-site wastewater treatment systems as a significant contaminant pressure on their rivers were selected for 24-hour monitoring (**Table 3.5** and **Figure 3.36**). The high concentration of these systems and the low permeability of the subsoil in which they were constructed meant that insufficiently treated effluent was discharging into the rivers (Gill et al., 2018). Samples were taken downstream of the failing systems in all catchments using both an ISCO 3710 Portable Composite Sampler (Teledyne ISCO, USA) and a MAXX TP5 Portable Composite Sampler (MAXX-GMPH., Germany) (**Figure 3.37**). Additional grab samples were taken at all catchments on the day of sampling. Spot depth and velocity recordings were converted into discharge data by developing stage/discharge relationships at each monitoring point using the midsection area/velocity method (ISO 748:2021).

Table 3.5. Characteristics of the selected study catchments

	Area (km^2)	**No. of OWTSs**	**Density ($no./km^2$)**	**Likelihood of inadequate percolation to groundwater**	**Land use**
Catchment 1 (C1) - Duran, Co. Wicklow	3.07	78	25.4	Very high (1.08 km^2)	Pasture (46%) Arable (10%) Forestry (33%) Other (11%)
Catchment 2 (C2) - Prospect, Co. Wexford	2.95	97	32.9	Very high (2.80 km^2)	Pasture (83%) Arable (0%) Forestry (2%) Other (15%)
Catchment 3 (C3) - Kilnacrott, Co. Cavan	3.85	60	15.6	Very high (3.85 km^2)	Pasture (85%) Forestry (15%) Other (<0.5%

Figure 3.36. Catchments C1: Wicklow, C2: Wexford and C3 Cavan.

Figure 3.37. Catchment C1 (left) C2 (centre) and C3 (right) 24-hour sampling.

3.6 Laboratory analysis

3.6.1 Chemical analysis

Septic tank effluent and lysimeter samples were periodically sampled and analysed for chloride (Cl), total organic nitrogen (TON), ammonium (NH4 +N), nitrite (NO2-N), Nitrate (NO3-N), and orthophosphate (PO43-P) using the automated Konelab 20i Chemistry Analyser (Thermo Fisher Scientific, USA) (**Table 3.6**). Total organic carbon and total carbon was measured using the

Shimadzu TOC-V CSN analyser (Shimadzu Scientific Instruments Inc, USA) using a high temperature catalytic oxidation tube with subsequent detection of the remaining CO2 in a non-dispersive infrared gas analyser (ND-IRGA). Calibrations for each test was performed on an annual basis.

Table 3.6. Displays the parameters, methods and detection limits used for monitoring water quality.

Parameter	Method	Method detection limit
TON	Spectrophotometric hydrazine method (APHA, 2012, method 4500-NO_3^-G)	0.1mgN L^{-1} – 100 mg N L^{-1}
$NH4^+$N	Spectrophotometric phenate method (APHA, 2012, method 4500-NH_3-G)	0.2mgN L^{-1} – 15 mg N L^{-1} / (300 mg N L^{-1}) *
NO_2^-N	Spectrophotometric phenate method (APHA, 2012, method 4500-NH_2-G)	0.1mgN L^{-1} – 40 mg N L^{-1}
PO_4^{3-}P	Spectrophotometric phenate method (APHA, 2012, method 4500-P E)	0.4mgP L^{-1} – 10 mg P L^{-1} / (50 mg P L^{-1}) *
Cl	Spectrophotometric phenate method (APHA, 2012, method 4500-Cl^- E)	0.349mg L^{-1} – 10000 mg L^{-1}

* Method detection upper test limit extended through dilution

3.6.2 Culturing methods

Total coliform and Escherichia coli counts were analysed using the Colilert – 18 test kit (IDEXX Laboratories Inc., USA). The quantification of coliforms is employed as a faecal indicator to determine the presence of wastewater contamination within the environment. Quantification was achieved using statistical aggregation of standardised volumetric vessels which was used to determine the most probable number (MPN) of faecal indicator within the sample. The method is the addition and mixing of the nutrient substrate with 100 ml of samples in a sterile vessel. Serial dilutions were often required for values of 10^{-2}, 10^{-4}, 10^{-6} in expectation of samples with bacterial counts ranging to 10^9 MPN / 100 ml. Samples were transferred to a standardised volumetric tray where 97 compartments allow for accurate quantification. These were then incubated for at least 18 hr at 35 degrees centigrade. During the incubation the bacteria metabolise a differential media producing colorimetric reactions which allow for the quantification of either *Escherichia coli* or *Total coliforms*.

3.6.3 Molecular analysis

Samples were taken back to the lab for genetic analysis, to provide understanding on the ecological profile of communities and the quantification of faecal indicator organisms.

Sample Pre-processing

Sample preparation of water or effluent samples required the concentration of samples for downstream processing. Sample volumes were dependent on the source due to availability, as effluent and river samples were 200 ml per sample, but pore water samples from the lysimeters were

more variable with volumes as low as 1 ml retrieved. Samples were stored at (maximum 6 months) -80^{O}C, before being defrosted, transferred to 50 ml centrifuge tubes, and centrifuged at 4^{0} C for 15 minutes at 3,500 g. The supernatant was then transferred to a sterilised Centricon 100kda centrifugation filtration units (Thermofisher Scientific, the US). The remaining pellet was weighed, weight recorded and stored as a separate sample. Centrifugation was crucial as suspended particles particularly common in primary effluent would collapse the filters and render concentration infeasible. Soil samples were subdivided into 250 mg -500 mg samples, before long-term storage. All samples, once weighed, had RNA later cryoprotectant solution (Thermofisher Scientific, USA) mixed at a 50:50 ratio, to prevent damage to samples during the freezing and thawing process.

DNA and RNA Extraction

The method of extraction of genomic material depended on the sample's downstream analysis. Samples which were to be sent for genomic sequencing were to be extracted for DNA, while samples that were to be analysed using quantitative reverse transcriptions polymerase chain reaction (RT-QPCR) were to be extracted for their RNA. All samples were first centrifuged for 15 minutes at 16 g to create a pellet and that RNA Later could be aspirated from the tube. The samples were then lysed using BeadBlaster™ 24 Microtube homogeniser (Benchmark Scientific D2400-E, USA). All samples were then homogenised at a speed of 6 m/s for at least 2 cycles of 30 seconds, followed by a pause of 10 minutes refrigeration to prevent damage to molecular material. For soil substrates an additional cycle was increased to three to ensure that sample was effectively lysed. Standardised commercial kits were employed for the extractions. For DNA FastDNA™ SPIN Kit for Soil (MP bio, USA) was used, and RNA was extracted using RNeasy Power Microbiome Kit (Qiagen, Netherlands) which includes a DNA digestion step. Although the kits purchased were commercial, methods were optimised by increasing the number of homogenisation and protein elution steps using common topsoil found near the laboratory. All DNA was quantified, and DNA quality was assessed using the Nano-drop microvolume spectrometer (Thermofisher Scientific).

16S RNA gene amplicon sequencing.

Next-generation amplicon sequencing was performed ex situ of the campus with samples sent at -80^{O}C in dry ice (Polarice, Ireland) to two commercial labs. Field samples were sent to NU-OMICS institute of Northumbria University UK. At NU-OMICS samples were analysed by next-generation amplicon sequencing of the 16S rRNA in a paired end mode. DNA extracts were sequenced with an Illumina MiSeq platform (Illumina, USA) using the primer set F515/R806 targeting 294 bp of the V4 region of the bacterial 16S rRNA gene, as described by Kozich et al. (2013). Column subsoil samples were sequenced by BMKgene using next-generation amplicon sequencing of the 16S rRNA in a paired end mode. DNA extracts were sequenced with an Illumina Novaseq sequencer (Illumina, USA) using the primer set 515F /R806 targeting 294 bp of the V4 region of the bacterial 16S rRNA gene (Walthers et al., 2016). In both labs Polymerase Chain Reaction is employed to amplify the 16 S genes. Firstly, purification removes the original 16S DNA from the sample leaving only the

copies created by the reaction. Libraries are prepared by ensuring the amplified DNA is of defined lengths with defined oligomer sequences at both ends which are compatible with the sequence technique (Hess et al., 2020). Each sample is given a label or barcode using short thread of synthetic sequences unique to each sample.

These samples are sequenced applying chemical-photo-coloration of the nucleotides A, T, C, G which can be captured by digital cameras which can then read the sequence of millions of DNA molecules simultaneously. This is achieved by a process titled "*sequencing by synthesis*". As the DNA is being copied one base at a time a new colour tagged nucleotide is added allowing for the reading of the sequence by the camera continuously capturing images of specialised silicone chip called a micrograph where the DNA synthesis takes place (Hess et al., 2020). Observations of the pigments on the micrograph can indicate homogeneity of a sample, as sequencing of a monoculture will display similar pigments rather in a heterogenous environment where the coloration will be less pronounced. These micrographs have been designed into flow cells which increase the throughput of sequencing exponentially with the use of barcoding. These frames of images are digitised to sequence reads producing sequences of bases in hundreds of millions of DNA molecules.

Bioinformatics

A bioinformatic data pipeline was developed for this research project. Sequence reads are processed before community analysis. This was achieved using DADA2 pipeline package with the R programme (v1.18.0; Callahan et al., 2016). Within each read a nucleotide has a quality score which determines how confident the sequencer was in qualifying the base present. A low call is often found at the end of a sequence and are often trimmed. The trimmed forward and reverse reads were merged with settings −25 M to 230 M. Spurious sequences are filtered using core de-noising algorithm in the DADA2 R package which is built on a model of the errors in Illumina-sequenced amplicon reads (v1.18.0; Callahan et al., 2016). Chimeras which are a combination of two or more unique sequences, often caused during errors in sequencing and were removed with the "removeBimeraDenovo" function in DADA2 under default settings. Taxonomic rank was derived using the "assignTaxonomy" function linked to the MIDAS database v 4.8.1 (Dueholm et al., 2021) assigned to the Usearch software package v11 (Edgar, 2013).

Sequences taxonomy was assigned by determining the similarity of operational taxonomic units (OTUs) to sequences. OTUs were assigned by confirming a 97% similarity at the species level from a taxonomic database. Taxonomic databases are reference databases of microbial sequence catalogues assigned a taxonomic position. In this research the Microbial Database for Activated Sludge (MiDAS) was applied for amplicon libraries, the database was developed from the sampling of 20 Danish wastewater treatment plants of 13 years (Nierychlo et al., 2020). The advantage of the MiDAS database is the referencing of functional groups such as Phosphate Accumulating Organism, Denitrifying, Nitrifying, Ammonia Oxidising Bacteria Methanogens within the sequence reads. The only exception was methanotrophs which required referencing literature (**Table 3.7**).

Table 3.7. Methanotroph targets for sequencing analysis and source literature

Methanotroph (Phlyum; Species)	Source
Proteobacteria; Methylocystis	(Takeda ., 1988)
Proteobacteria; Methylobacter	(Donrina et al., 2004)
Proteobacteria; Methylocella	(Dunfield and Dedysh,. 2014)
Proteobacteria; Methylocaldum	(Eshinimaev et al., 2004)
Proteobacteria; Methylobacillus	(Kumar et Maitre., 2016)
Proteobacteria; Methylomonas	(Cheng et al., 2022)

Alpha diversity (richness and evenness within the samples) was assessed by computing the number of Operational Taxonomic Units (OTUs), abundance-based coverage estimator (Chao1) (Chao and Lee, 1992), and Shannon diversity (Shannon, 1948) for all 420 samples. Principal coordinate analysis (PCoA) plots were estimated using Bray-Curtis and weighted Unifrac metrics, which were derived from the rarefied OTU table using the ampvis2 v 2.7.11 and phyloseq v 3.6 packages, respectively. Further analysis was performed to determine the categorical variables of statistical significance to determine cause of variation within the microbial communities. Permutational analysis of variance was performed on datasets applying distance matrices (ADONIS) using the vegan package v 2.4.2 (Oksanen et al., 2021) performing 2000 permutations per variable. All analyses and plots were performed on R version 4.1.1 through the Rstudio IDE (R Core Team, 2014).

3.6.4 Reverse Transcription Quantitative Polymerase Chain reaction

Quantitative Polymerase Chain reaction has become a molecular standard practice over the past 20 years. It was employed in this study in the measurement of plant pathogen and the faecal indicator Pepper Mild Mottle Virus, known to be non-pathogenic in humans but is carried at high values within our faeces at concentrations of 9.8 $x10^4$ gene copies / mg (Zhang et al., 2006). The focus of this study was to determine if it was present within the leachate of soil treatment systems and catchments that were impacted by on-site treatment systems. As an RNA virus we applied reverse transcription qPCR protocol for the enumeration of the virus. This required several steps prior to processes samples directly.

Sequence Literature Study, Target and Primer Selection

Literature search was applied on Scopus, PubMed and Google Scholar on *Pepper Mild Mottle Virus*. Representative gene sequences of these genes were retrieved using National Centre for Biotechnology Database. Ultimately a sequence (Ascension number: AB069853) of the virus was selected (Hagiwara et al., 2001). QPCR primers developed by the highly cited Zhang et al. (**Table 3.8**) were chosen from the literature. For the assay to function, primers and probe needed to align with the chosen sequence and an optimal amplicon or replicate sequence size is selected for the assay. This was achieved using MEGA X which employed the MUSCLE algorithm in aligning the primers and probe with the target sequence (v. 10 Tamura, Stecher, and Kumar 2020). Ultimately,

120 bp long portion of the sequence (**Figure 3.38**) was selected to be designed and synthesised as a synthetic genetic sequence *"gblock"* at Integrated DNA Technologies (Integrated DNA Technologies, USA).

Table 3.8. Target faecal indicator marker, primer sequences, base pair length and reference.

Target	Primer sequences	bp	Reference
PMMoV	PMMV-Fw1-rev 5'-GAGTGGTTTGACCTTAACGTTTGA-3' PMMV-Rw1 5'-TTGTCGGTTGCAATGCAAGT-3' PMMV-Probe1 FAM-CCTACCGAAGCAAATG-BHQ1	1861-1981	(Zhang et al., 2006)

```
1861 gctgtggttt caaatgagag tggtttgacc ttaacgtttg agaggcctac cgaagcaaat
1921 gtcgcacttg cattgcaacc gacaattaca tcaaaggagg aaggttcgtt gaagattgtg
1981 tcgtcagacg taggtgagtc ctcaatcaag gaagtggttc gaaaatcaga gatttctatg
2041 cttggtctaa caggcaacac agtgtccgat gagttccaaa gaagtacaga aatcgagtcg
```

Figure 3.38. PMMoV target amplicon sequence for qPCR assays, sequence was selected for block synthesis, please not the colour configuration highlights the positioning of the primers in the table for the replication process.

Primer Validation, Inhibition Tests and PMMoV Quantification

For validation of reaction conditions set by Zhang et al. (2006), the set of primers were first tested with the standard curve generated from the serial dilution [10^1–10^8 gene copies gc / µL]of block using StepOnePlus System (Applied Biosystems, USA). The assay was performed in a 20-µl reaction mixture containing 6 µl of sample RNA, 4ul TaqPath™ 1-Step RT-qPCR Master Mix , CG (Applied Biosystems, USA), 2ul 4 mg / ml BSA (Bovine Serum Albumin, Thermofisher Scientific) 0.9 µM primers and 0.3 µM probe (Zhang et al., 2006).

PMMoV MST marker concentrations were expressed as gene copies per 100 ml (gc / 100 ml). The RT reaction was performed at 50 °C for 15 min, followed by 2 min at 95 °C for activation of DNA polymerase, then 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Negative controls (RT-less and RNA less). After confirmation an adequate standard curve with an $R^2 = 0.98$ was achieved (**Figure 3.39**). Following Validation with the standard curve, samples were tested for possible inhibitors which may reduce the effectiveness of reactions and yields. This was achieved by mixing 1 ml tabasco with Primary effluent from a septic system, effluent which was collected from the leachate of one of the soil columns, and with 100 ml of distilled water. These mixtures were concentrated with centrifugation as described, following concentration it was then mixed with RNA at a 50:50 ratio. This concentrated water tabasco solution was then compared to a sample which underwent the same process except it was not mixed with the RNA Later. The outcome showed that the centrifugation step showed no loss in the sample (**Table 3.9**). Soil column and septic tank effluent were diluted at

a ratio 1:2, 1:10, 1:100 of RNAase free ultra-clean water (Qiagen , Netherlands) to determine the effective dilution rate in which , improving detection or quantification. A dilution 1:10 dilution was chosen with samples being assayed in triplicate, however, due to some sample's low concentrations (i.e. rivers) duplicates of samples with no dilutions were run in parallel with those diluted and assayed in duplicate in the event of a repetition of any assay which initially yields undetermined results. The calculations to quantify the values are listed below.

Translating PMMoV genes copies into concentration

PMMoV concentration was calculated as follows:

$$\frac{no.of\ gene\ copies}{Volume\ of\ sample\ in\ reaction\ (\mu l)} = (A) \qquad (1)$$

$$(Volume\ of\ DNA\ extracted\ (\mu l) = (B)$$

$$\frac{Dilution\ factor\ of\ sample}{Volume\ filtered\ for\ DNA\ extraction\ (ml)} = (C)$$

PMMoV gene copies per 100 ml = (A)x(B)x(C)x10²

PMMoV loads

PMMoV gene copies per day was calculated as follows:

$$Volumetric\ flow\ rate\ (Q) = \frac{Autosampler\ reservoir\ volume}{Time\ (\text{hrs})} \qquad (2)$$

$$Daily\ load\ (24hr) = [PMMoV]x(Q)x(T) \qquad (3)$$

Limit of Detection

The limit of detection was calculated as follows:

$$\frac{No.of\ gene\ copies\ at\ the\ the\ Limit\ of\ Detection\ (copies)}{Volume\ of\ sample\ in\ reaction\ (\mu l)} = (A) \qquad (4)$$

$$\left(Volume\ of\ DNA\ extracted\ (\mu l)\right) = (B) \qquad (5)$$

$$\frac{Dilution\ factor\ of\ sample}{Volume\ filtered\ for\ DNA\ extraction\ (ml)} = (C) \qquad (6)$$

The limit of detection (No. of copies/ml) $= (\boldsymbol{A}) \times (\boldsymbol{B}) \times (\boldsymbol{C})$

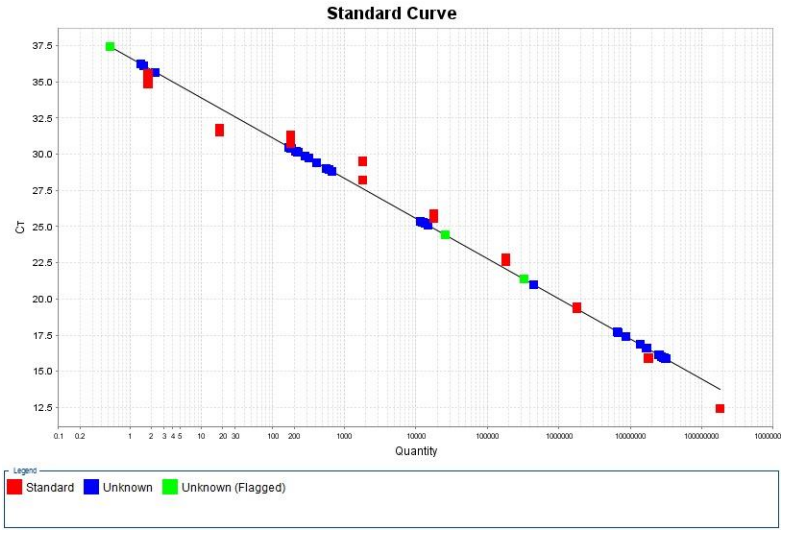


Figure 3.39. Standard curve marked in red with a R^2=0.98, with the optimised samples marked in blue

Table 3.9. Concentrations, samples, and their preparation methods in optimisation step

sample and optimisation method	Concentration (gc / 100 ml)
Tabasco	2.03E+07
Tabasco and RNA Later (50:50)	1.48E+09
Tabasco and distilled water (1 ml in 100 ml) concentrated	1.38E+09
Tabasco and Soil column effluent (1:100)	7.92E+08
Tabasco and Primary Effluent (1:100)	1.42E+09
Primary effluent	1.29E+06
Primary effluent 50:50 dilution	6.82E+05
Primary effluent 1:10 dilution	6.04E+05
Primary effluent 1:100 dilution	3.08E+04
Soil column effluent	1.68E+04
Soil column effluent 50:50 dilution	9.39E+03
Soil column effluent 1:10 dilution	1.09E+04
Soil column effluent 1:100 dilution	8.72E+01

3.6.5 Bacterial growth rates

A key objective the column study research was to determine the relative importance of species sorting and immigration in the development of biomat communities within the subsoil. This was achieved by characterising the microbial community within the influent, effluent and at different depth zones within each of the columns using 16S genomic sequencing, to estimate the solid retention time (SRT) and net growth rates of operational taxonomic units present within the soil columns. Soil masses were calculated from the bulk density and the volume of each of the zone of the column. Cells were quantified in the soil, influent and effluent by the Bradford method, a colorimetric assay (Biorad, USA) measuring total protein concentration of 250 mg of soil samples OR 250 µl of influent or effluent. These cell counts were used to quantify the SRT and determine the growth rate which provided in the calculations below.

Defining the relative abundance (%) of organisms (x) in the influent wastewater, soil column zone, control subsoil and effluent as $p_{x,WW}$, $p_{x,AGS}$, $p_{x,ES}$ and $p_{x,EF}$, respectively, equation (7) can be rearranged as:

$$\theta_x = \frac{N_{xCZ} - N_{xCsZ}}{n_{xEFF-} n_{xINF}} \qquad (7)$$

where

N_{xCZ} is the number of cells of organism x in the column depth zone

N_{xCsZ} is the number of cells of organism x in the control subsoil equivalent depth zone

n_{xEFF} is the number of cells of organism x leaving the column through the pore water effluent

n_{xINF} is the number of cells of organism x entering with influent wastewater

Primary data on the soil columns were calculated by colorimetric BSA analysis and

(cell count / g) x (bulk density) x (zone volume))

for each of the depth zones for each column (**Table 3.10**)

Table 3.10. Soil masses (g) and cell counts for each of the soil column zones, influent and effluent

	PE-High		**SE-High**		**PE-Low**		**SE-Low**	
	Soil Mass (g)	**Cell count**	**Soil Mass (g)**	**Cell count**	**Soil Mass (g)**	**Cell count**	**Soil Mass (g)**	**Cell count**
Influent		4.70E+11		1.84E+10		4.70E+11		1.84E+10
Infiltrative zone	634.5	3.05E+14	634.5	1.52E+13	603	1.48E+13	603	1.45E+13
Proximal Zone	2538	3.80E+14	2538	2.98E+14	2412	2.87E+14	2412	2.97E+14
Intermediary Zone	3172.5	5.96E+15	3172.5	3.77E+14	3015	3.55E+14	3015	3.61E+14
Deep Zone	12690	1.52E+13	12690	6.01E+15	12060	5.64E+15	12060	5.57E+15
Effluent		4.76E+11		4.64E+11		4.62E+11		4.68E+11

The parameters in equation (2) were calculated as follows:

Calculation of N_{xCZ}

Using PE-High column infiltrative zone as an example: the abundance of organisms in the zone was reported as 3.05×10^{14} cells per 634.5g using the mass quantities described in **Table 3.10.**

Hence,

$$N_{xCZ} = 2.41x10^{10}/g \; x \; 634.5g = 3.05 \times 10^{14} \text{cells} \quad (8)$$

Calculation of N_{xCsZ}

Control subsoil samples were measured, and average cell count of 2.24×10^{10} and total cells were calculated exactly as was done in equation (8).

Calculation of N_{xINF}

The total number of cells in the influent wastewater was calculated assuming cell concentration in average of about 1.52 x 10^{13} cells m^{-3} for PE influent, 1.84 x 10^{10} cells m^{-3} for SE influent. Each column received 1 m^3 d PE influent is used in this example:

Hence,

$$n_{xINF} = 4.70x10^{11}x\ 1\ m^3\ d^{-1} = 4.70x10^{11}\text{cells}$$

Calculation of N_{xEFF}

The concentration of cells in the effluent was measured for each column in 24-hour cycles, each cycle resulted in ~ 1 m^3 d^{-1} was reported**.** Using the PE-High column as an example, the total number of cells leaving PE-High column effluent is calculated as:

$$n_{xEFF} = 4.76\ x10^{11}x\ 1\ m^3\ d^{-1} = 4.76x10^{11}\text{cells}$$

Solid Retention Time

Replacing these values in equation (7), we get:

$$\theta_x = \frac{p_{xCZ}.N_{xCZ} - p_{xCsZ}.N_{xCsZ}}{P_{xEFF}.n_{xEFF} - p_{xINF}n_{xINF}} \qquad (9)$$

Values of $p_{xCZ,}$,p_{xCsZ}, P_{xEFF}. and p_{xINF} For the different organisms were determined from the rarefied OTU table. OTUs with relative read abundance ≥ 0.1% were selected for calculating equation (9).

Net growth rates

The net growth rate of organisms (*x*) was determined using equation (10):

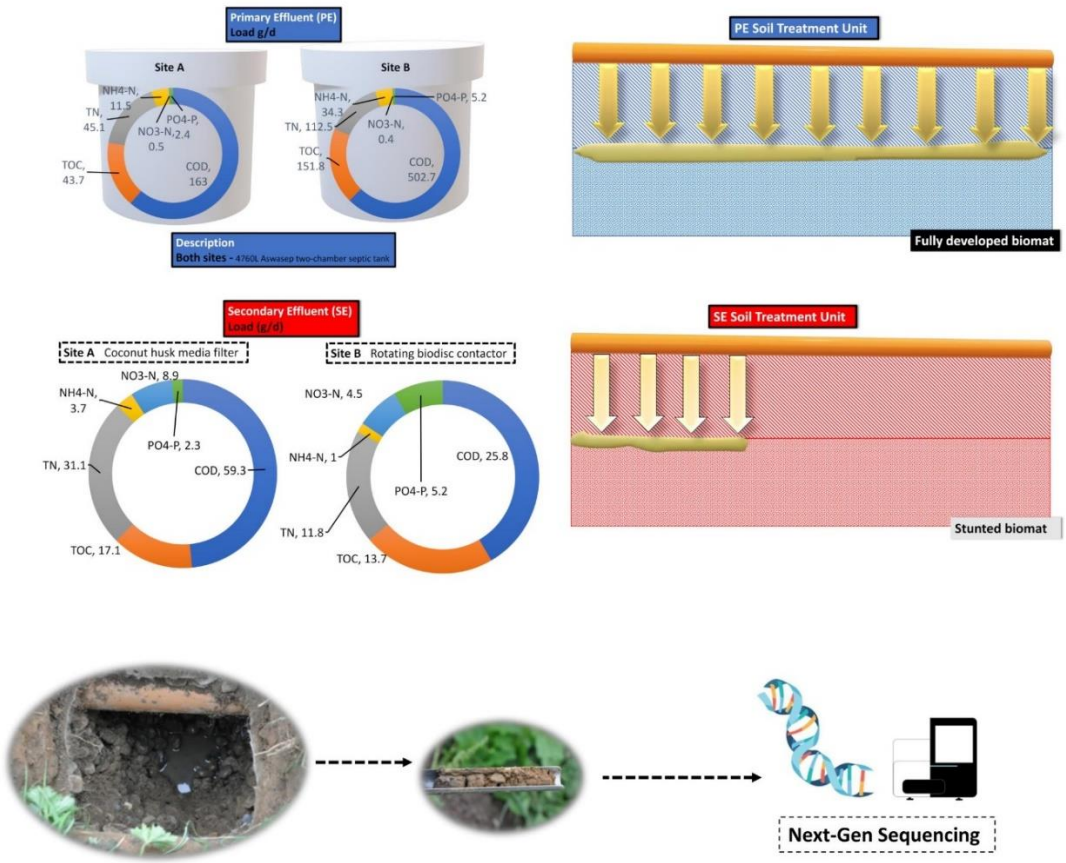
$$m_x = \frac{1}{q_x}$$

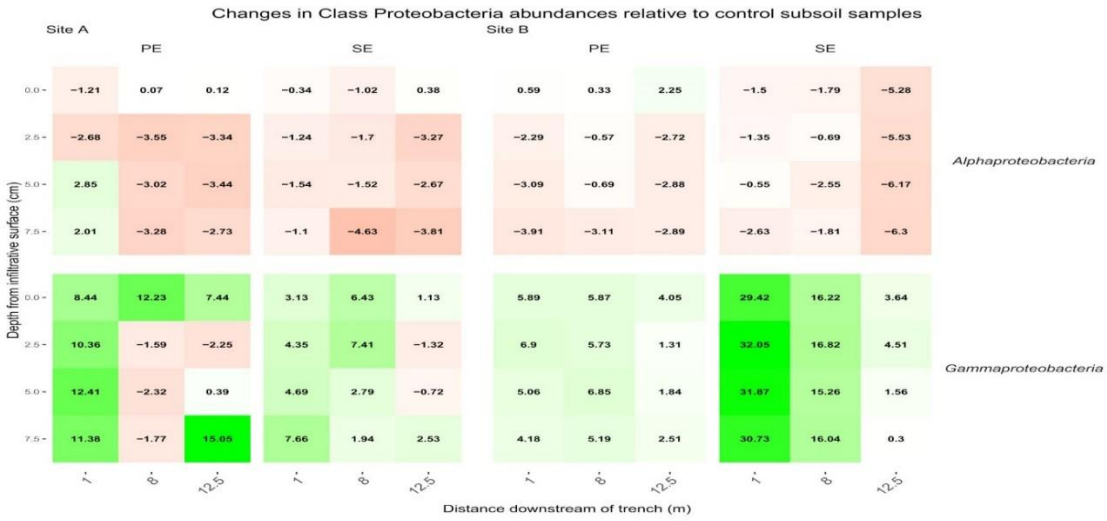
4. Spatial variation of the microbial community structure of on-site soil treatment units in a temperate climate, and the role of pre-treatment of domestic effluent in the development of the biomat community.

Criado Monleon Alejandro Javier; Knappe Jan; Somlai Celia; Ospina Carolina; Ali Muhammad ; Curtis Tom P.; Gill Laurence W

This paper was published on 24 June 2022 Volume 13 – 2022 of Frontier in Microbiology

Keywords: Bioclogging; On-site wastewater treatment; Soil Treatment Unit; Microbial diversity; Microbial community Structure; Microbial community Composition





Key Findings

- Spatial variation and pre-treatment significantly contributed to biomat community composition
- Application of pre-treatment increased abundances of key metabolic bacterial groups, including annamox

4.1 Abstract

The growth of microbial mats or '*biomats*' has been identified as an essential component in the attenuation of pollutants within the soil treatment unit (STU) of conventional on-site wastewater treatment systems (OWTSs). This study aimed to characterise the microbial community which colonises these niches and to determine the influence of the pre-treatment of raw-domestic wastewater on these communities. This was done through a detailed sampling campaign of two OWTSs. At each site, the STU areas were split whereby half received effluent directly from septic tanks, and half received more highly treated effluent from packaged aerobic treatment systems (a coconut husk media filter for site A, and a rotating biodisc contactor – RBC – at site B). Effluent from the RBC had a higher level of pre-treatment (~90% Total Organic Carbon – TOC – removal), compared to the median filter (~60% TOC removal). Ninety-two samples were taken from both STU locations and characterised by 16S rRNA gene sequencing analysis. The fully treated effluent from the RBC resulted in lower microbial community richness and diversity within the STUs compared to the STUs receiving partially treated effluent. The microbial community structure found within the STU receiving fully treated effluent was significantly different to communities receiving fully treated effluent from the RBC. Moreover, the distance along each STU appears to have a greater impact on community structure than the depth in each STU. Our findings highlight spatial variability of diversity, phylum- and genus-level taxa, and functional groups within the STUs, which supports the assumption that specialised biomes develop around the application of effluent under different degrees of treatment and distance from the source. This research indicates that the application of pre-treated effluent infers significant changes to microbial community structure, which in turn has important implications for the functionality of the STU and consequently the potential risks to public health and the environment.

4.2. Introduction

A typical on-site wastewater treatment system (OWTS) uses a septic tank to provide primary treatment and a limited amount of anaerobic digestion. Further treatment occurs as effluent percolates through the soil-stone matrix of the soil treatment unit (STU) which can vary in their configuration according to site-specific design requirements (Gill, 2011). The underlying soil or subsoil into which the wastewater effluent percolates provide a critical buffer zone for the protection of water resources. The key to effective on-site treatment is to maintain an unsaturated subsoil through which the effluent can percolate freely and wherein chemical and microbiological contaminants will be attenuated to an acceptable level before they reach the groundwater (Siegrist, 2017). In Ireland, for example, the Environmental Protection Agency's Code of Practice dictates that septic tank effluent requires at least 1.2 m unsaturated subsoil depth below the invert of the percolation trenches to the water table or bedrock with one 18 m long percolation trench required

per household occupant (EPA, 2021). Similar design criteria are used in other countries, such as the US, Canada, Australia, U.K. etc.

Recently, there has been a proliferation of packaged treatment systems that provide additional (secondary) treatment to the effluent before being discharged to the STU. Regardless of using a secondary treatment unit, the STU remains a crucial component of domestic wastewater treatment, particularly, the development of the biomat or 'clogging layer', which grows over the base of the STUs correlated to the level of organic and nutrient loading (Bouma 1975; Siegrist and Boyle 1987). This layer causes a sharp drop in infiltration rates, initially due to physical clogging processes, followed by more gradual clogging over several months resulting from the development of the microbial biomat formed by the production Extracellular Polymeric Substances (EPS) (Beal *et al.*, 2005). Studies have linked the effects of the organic loading rates on the biomat development, comparing trenches dosed with primary (septic tank) effluent compared to secondary treated effluent (Gill *et al.*, 2007; Knappe et al., 2020). These studies also showed that nitrogen loading at the base of the STUs receiving secondary effluent was 2–3 times higher than in percolation areas receiving septic tank effluent (Gill *et al*., 2009).

Most on-site wastewater treatment studies have focused on the system performance for the attenuation of hazardous contaminants from an environmental and public health perspective. These studies have focused on chemical parameters such as nutrients (different forms of nitrogen and phosphorus), bulk organics (BOD, COD, TOC) and faecal indicator bacteria such as *E. coli*, enterococci, and bacteriophages as surrogates for human enteric viruses (Gill *et al*., 2009; Van Cuyk *et* Siegrist., 2007; O'Luanaigh *et al*., 2012; Humphrey et al., 2019). Increasingly, since the advance of microbiological culture-independent techniques in the early 1990s (Wagner *et al*., 2006), the performance of wastewater treatment systems has been coupled with the dynamics and stochastic modelling of the composition of microbial communities (Curtis and Sloan, 2005; Sanz and Kochlin., 2007; Siezen and Galardini., 2008; Matar *et al*. 2021). Much of the focus to date has been about nitrogen removal processes (ammonia oxidation, nitrite oxidation, denitrification, anammox), phosphorous removal, floc, and biofilm formation (Daims *et al*., 2006). With advances in sequencing technology and expansions in genomic databases, more detailed microbial community profiles have been developed for suspended growth flocs (in activated sludge) and attached growth fixed-film treatment systems (trickling filters etc.) (Sanz and Kochlin., 2007) and suspended growth biofilm treatment systems (Ali et al., 2019).

Several factors impact the microbial community structure within a wastewater treatment system, including influent composition, environmental conditions, system processes, plant configuration and operational parameters (Hu *et al.*, 2012; Lee, Kang and Park, 2015; Chen *et al.*, 2017; Zhang *et al.*, 2020). The propagation of cheaper and more accurate next-generation sequencing has allowed for a greater resolution of community structure, deeper sequencing and thus a better determination of the

processes behind community assembly, such as dispersal (immigration from influent), selection (deterministic, driven by taxa fitness and environmental factors), and ecological drift (temporal changes in abundances caused by stochastic events) (Ali *et al.*, 2019; Frigon and Wells, 2019; Dottorini *et al.*, 2021).

Much of the research on microbiological analysis associated with wastewater treatment processes has been confined to centralised wastewater systems, as highlighted above. Within on-site soil filtration systems, Tomaras *et al.* (2009) presented one of the first sequencing profiles of soil microorganisms from the biomat of STU, the study stated that microbes found within septic tank effluent were not detected in the STU biomat sampled. Depth has also been noted as a contributor to the microbial community structure (Truu *et al.*, 2009). Effluent storage has been shown to cause reductions in microbial diversity as high nutrient contents are gradually degraded (Knisz *et al*., 2020) with comparisons of the effluent from multiple on-site systems showing the effect of treatment technologies and seasonality on the structure of nitrifying and denitrifying communities (Ross *et al.*, 2020). Crucially studies have noted that, in contrast to centralised wastewater treatment plants (WWTP), OWTS appear to be highly influenced by the inoculation of soil organisms during installation and maintenance (Wigginton *et al.*, 2020). However, the microbial community structure of STUs is understudied. This lacuna is likely due to the laborious and time-consuming nature of sampling within the soil system. Here we present a high-resolution profile of the community structure across two separate OWTs with relatively similar subsoil and land use that enables a more direct comparison of the effect of treated effluent dispersal on the STU biome. This work expands the knowledge on STU biomat growth by providing valuable insights about the community structure across the length and depth of several systems, each receiving an effluent with varying levels of pre-treatment. The findings of this study also have implications for other related fields such as reclaimed water irrigation and groundwater recharge.

4.3. Methods

4.3.1 Site Description

This study investigates the spatial distribution of microorganisms in two separate OWTSs in two owner-occupied homes in Co. Limerick, Ireland. The region is classified as a temperate oceanic climate (or *Cfb* classification within the Köppen climate classification system) (De Carli et al., 2018). The subsoil at both sites are classified as Typical Luvisol soils averaging at pH 8 with little variance (Fearly, 2009). The microbial biomats have developed at both research sites, which forms at the interface where the effluent percolates into the soil (i.e. the STU) at the base of the gravel percolation trenches. These biomats will be the principal focus of the study each percolation trench was 18 m long and 0.5 m in width, with a gradient of 1:200 filled with 300 mm pea gravel in which

a perforated rigid plastic pipe was set (as per EPA, 2021 design guidelines (**Figure 4.2**). The pipes were then covered with 150 mm gravel and a geotextile fabric to prevent backfilled topsoil from washing into the gravel layer beneath. Both sites employed a 4760L Aswasep two-chamber septic tank (Molloy Precast Products Ltd., Ireland), and four percolation trenches, two of which were fed directly from the septic tank system as primary effluent (PE), the other two percolation trenches were being fed with secondary treated effluent (SE) from packaged treatment systems (**Figure 4.2**). At Site A, secondary treatment was achieved using an intermittently dosed coconut husk filter system (Ecoflo Coco Filter, Premier Tech Aqua Ltd., Ireland). At Site B, the packaged secondary treatment system was an RBC (Klargester BioDisc, Kingspan., Ltd., UK), consisting of an integrated primary settling chamber, a two-stage biozone, and a secondary clarifying chamber. At both sites, primary and secondary effluent were distributed equally onto t own separate half of the STU. The distribution of effluent evenly between these parallel trenches was ensured using calibrated tipping bucket distribution devices (Patel *et al*., 2008), which were designed with reed switches to calculate the daily flows to each STU. The soil and hydrogeological factors from each site (as determined from a parallel research study; Knappe *et al*., 2020) are summarised in **Table 4.1**.

Table 4.1. Soil and hydrological parameters of both sites (Knappe *et al*., 2020).

Parameter	**Site A**	**Site B**
% Sand – Silt – Clay	59 - 30 - 11	49 – 34 - 17
Bulk density (cm^3 g^{-1}	1.44	1.2 0
porosity	0.386	0.448
groundwater (m)	1.6	>2.5
k_{sat} (cm d^{-1}) [1.]	30.9	13.9
k_{sat}_sd	3.5	5.4
flow mean (Ld^{-1})	269.8	500.1
flow_sd	329.1	200.8
Construction	September 2015	April 2016
Primary Treatment	Septic tank	Septic tank
Secondary Treatment	Cocopeat media filter	Rotating biological contactor
Flow Regime	Pumped – flow	Gravity flow
Number of occupants	5	4

[1.] k_{sat}= saturated hydraulic conductivity

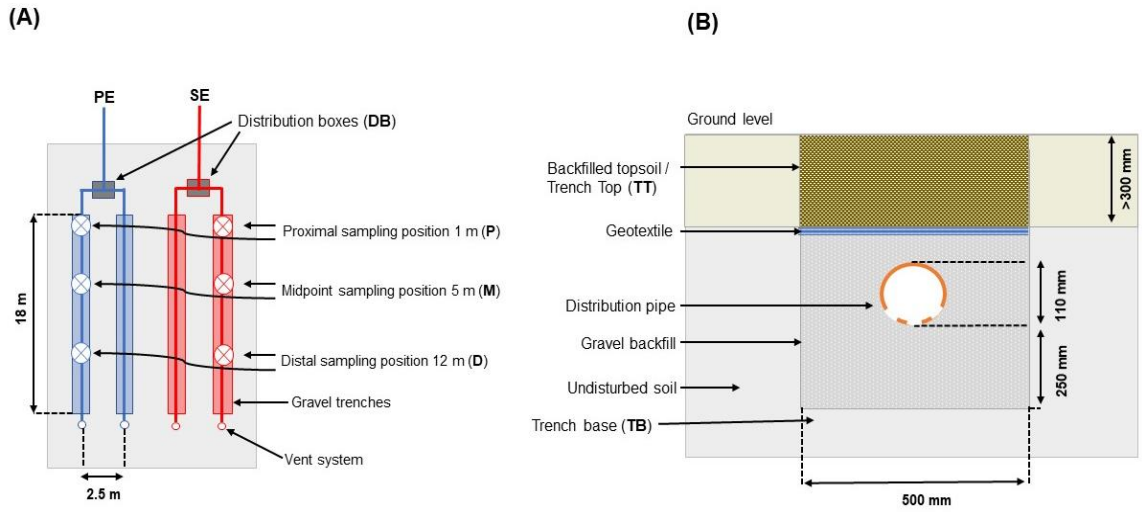


Figure 4.2. General schematic of the soil treatment system, for both Sites A and B of this study (Knappe et al., 2020).

4.3.2 Sampling and 16S rRNA Gene Sequencing

At selected locations, the soil was excavated from the surface down into the percolation trenches to the gravel subsoil interface at the base of the trenches. Core samples were then taken with a 25.4 mm stainless steel corer. At each site and for each system, a single sample was taken at different depths from the infiltrative interface (2.5 cm, 5.0 cm and 7.5 cm) (**Figure SA2**), and for each sampling position in the horizontal direction away from the inlet of the STU. A sample was also taken from the surface interface (i.e., 0 cm depth). For DNA extraction of samples for sequencing, approximately 3 g of soil were collected from the corer and placed into a sterile 2 ml Eppendorf tube and stored at -20^{+}C. Sample handling was performed with a sterile metal spatula, with sterilising being performed between each sample with 70% ethanol. 250 mg samples were then extracted from the 92 soil samples taken from the field using a DNAeasy power soil kit (Qiagen, NL) (**Table 4.2**). DNA concentration was checked using the NanoDrop spectrometer (Nanodrop ND-1000, ThermoScientific, Waltham, MA).

The microbial community within reactors was assessed by next-generation amplicon sequencing of the 16S rRNA in paired end mode. DNA extracts sequenced with an Illumina MiSeq platform (NU-OMICS, Northumbria University, U.K.) using the primer set F515/R806 targeting 294 bp of the V4 region of the bacterial 16S rRNA gene, as described by Kozich et al. 2013.

Table 4.2. The sample counts for each of the treatment sites and within reach of the on-site wastewater treatment systems. Total sample counts (n=92).

Location	Abbreviation	Site A	Site B
Primary effluent	PE	1	1
Secondary effluent	SE	1	0
Primary effluent trench topsoil	PE-TT	6	6
Secondary effluent trench topsoil	SE-TT	6	6
Primary effluent trench subsoil	PE-TB	15	12
Secondary effluent trench subsoil	SE-TB	12	12
Primary effluent distribution box	PE-DB	0	1
Secondary effluent distribution box	SE-DB	0	1
Rotating biodisc contactor	RBC	0	2
Control topsoil	CT	3	3
Control subsoil	CB	2	2

4.3.3 Sequence Processing and Analysis

Pair end reads were converted into amplicon sequence variant libraries produced using the DADA2 pipeline package with the R program (v1.18.0; Callahan *et al*., 2016). The trimmed forward and reverse reads were merged with settings -25 M – 230M. Chimeras were removed with 'removeBimeraDenovo' function in DADA2 under default settings. Taxonomic rank was derived using 'assignTaxonomy' function linked to the Silva database v 138.1. Putative functional groups were identified using the MIDAS database v 4.8.1 (Dueholm *et al*., 2021) assigned with the Usearch software package v11 (Edgar *et al*., 2013).

Alpha diversity (richness and evenness within samples) was assessed by computing the number of Operational Taxonomic Unit (OTUs), abundance-based coverage estimator (CHAO1) (Chao and Lee., 1992) and Shannon diversity (Shannon., 1948) for all 92 samples. Principal Coordinate analysis (PCoA) plots were estimated using Bray-Curtis and weighted Unifrac metrics, which were derived from rarefied OUT table using ampvis2 v 2.7.11 and phyloseq v 3.6 packages respectively. Further analysis was performed to determine the categorical variables of statistical significance to determine the variation within the microbial communities. Permutational analysis of variance applying distance matrices (ADONIS) using vegan package v 2.4.2 (Oksanen et al., 2020) was used to analyse several variables based on 2000 permutations. All analyses and plots were performed on R version 4.1.1 through the Rstudio IDE (R Core Team, 2021).

4.3.4 Site Instrumentation

Both research sites were fitted with automated weather stations (Campbell Scientific, UK) measuring air temperature, relative humidity, atmospheric pressure, net radiation, wind speed and direction, and rainfall. Hydraulic loadings were determined using calibrated tipping buckets (described previously). A network of suction lysimeters (Model 1900, Soilmoisture Equipment Corp., USA) had been installed at each site spaced longitudinally along each trench and at three

depths beneath the infiltrative surface down to 50 cm depth in the soil. For sample collection, a suction of 50 kPa was applied using a vacuum-pressure hand pump and samples collected 24 hrs later. Effluent and porewater samples extracted from lysimeters were stored on ice for <6 hr transport to be analysed in the environmental engineering laboratory at Trinity College Dublin. The organic load of the PE and SE fed into the STUs were determined as chemical oxygen demand (COD) using dichromate digestion test kits (Merck, Germany) and total organic carbon (TOC) using a Shimadzu TOC-V analyser (Shimadzu Scientific Instrument, USA). Nitrogen species were analysed as nitrate-nitrogen (NO_3-N), nitrite-N (NO_2-N), ammonium – N (NH_4-N) and phosphorus as orthophosphate (PO_4-P) using Konelab 20i chemistry analyser (Thermo Scientific, Finland). Assessing the spatial distribution of the volumetric water content (VWC) and long-term changes in water retention within the STUs was achieved by a network of 80 and 92 soil moisture sensors (EC5, Decagon Devices, USA) which had been installed during the construction of Sites A and B respectively. Sensors were installed by auguring a 10 cm diameter hole to a desired depth, with sensors positioned into undisturbed subsoil at required depth below the STU, control sensors were installed outside the STU area at corresponding depths of sensors installed within the STU. Calibrations were performed to the manufacturer's methods using site-specific subsoils retrieved from test holes excavated before the construction of the sites. All sensor data was collected hourly and stored on a CR1000 data logger with two AM15/32 multiplexers (Campbell Scientific, UK).

4.3.5 Data Availability Statement

Raw sequencing data were deposited at the National Center for Biotechnology Information (NCBI) under accession number PRJNA794316.

4.4. Results

4.4.1 Meteorological Conditions

Sites A and B received a mean annual rainfall of 928.6 mm and 972.4 mm, at mean temperatures of 10.0 ^{O}C and 10.4 ^{O}C, record lows of -7.2 ^{O}C and -7.4 ^{O}C and highs of 30.8 ^{O}C and 30.3^{O}C. respectively over the study period. During sampling in August 2018 total monthly precipitation was 41.9 mm and 17.6 mm, and mean temperatures of 15.7 ^{O}C and 15.6 ^{O}C for site A and B respectively. Detailed trend graphs of total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites can be found in the supplementary material (**Figure SA7)**. Between May 19th and August 12, 2018, Ireland was affected by a drought period resulting in no effective rainfall and severe soil drying (Met Éireann, 2018).

4.4.2 Effluent Quality and Wastewater Treatment System Performance

The average quality of effluent from the septic tanks and packaged treatment plants that were feeding the percolation trenches (**Table 4.3**). This shows the difference in effluent quality, particularly between the two packaged treatment systems where the coco-media filter on Site A was only partially nitrifying and removing just 60% of organics, compared to the RBC on Site B which was fully nitrifying and removing>90% of the organics. The mean effluent hydraulic loading on Site A was 269.8 L/d and on Site B was 500.1 L/d.

The two systems at site A and B had been in operation for approximately 35 and 29 months respectively, when the soil samples from the percolation area were taken. An intensive research study had been characterising the performance of the OWTSs in terms of the three-dimensional attenuation of contaminants as they passed down through the soil, and the effluent quality of the upstream treatment units (septic tanks and secondary treatment units) – as detailed in Knappe *et al.*, (2020). The in-line three-dimensional soil water content sensor network and the chemical quality of soil moisture percolating beneath the STUs, provided detailed surveillance of the biomat development within and below the infiltrative soil surface, using water retention as a proxy for the presence of the biomat due to bioclogging caused by the extracellular polymeric substance matrix present within the biomat. There were significant differences in mean water retention between STUs receiving septic tank primary effluent (PE) and packaged treatment system secondary effluent (SE) across all positions. The greatest amount of water retention was found at a depth of 5 cm below the biomat, followed by the upper layer of the infiltrative surface, decreasing with depth.

Table 4.3. Effluent characteristics for primary effluent (PE) from the septic tanks and secondary effluent (SE) from the packaged treatment units on Site A and Site B over a 3-year period (fr.om Knappe *et al*. (2020).

Site	Parameter	PE		SE		Mean removal efficiency
		concentration (mean ± SD)	load (mean ± SD)	concentration (mean ± SD)	load (mean ± SD)	
Site A	COD	605.8 ± 240.6 mg L-1	163 ± 64.7 g/d	220.5 ± 116.4 mg L-1	59.3 ± 31.3 g/d	0.636
	TOC	162.5 ± 82.7 mg L-1	43.7 ± 22.2 g/d	63.6 ± 42.3 mg L-1	17.1 ± 11.4 g/d	0.609
	TN	167.8 ± 69.0 mg L-1	45.1 ± 18.6 g/d	115.7 ± 44.9 mg L-1	31.1 ± 12.1 g/d	0.31
	NH4-N	42.7 ± 54.5 mg L-1	11.5 ± 14.7 g/d	13.6 ± 18.2 mg L-1	3.7 ± 4.9 g/d	0.681
	NO3-N	1.8 ± 2.5 mg L-1	0.5 ± 0.7 g/d	32.9 ± 18.0 mg L-1	8.9 ± 4.8 g/d	–
	PO4-P	8.8 ± 6.4 mg L -	2.4 ± 1.7 g/d	8.6 ± 5.8 mg L-1	2.3 ± 1.6 g/d	0.023
	Total coliforms	3.45 x 10E6 MPN/100mL		1.11 × 106 MPN/100mL		0.49 log10
	E. coli	1.35 x 10E5 MPN/100mL		8.60 × 104 MPN/100mL		0.20 log10
Site B	COD	1005.4 ± 192.7 mg L-1	502.7 ± 96.4 g/d	51.6 ± 43.5 mg L-1	25.8 ± 21.8 g/d	0.949
	TOC	303.6 ± 55.2 mg L-1	151.8 ± 27.6 g/d	27.4 ± 16.7 mg L-1	13.7 ± 8.4 g/d	0.91
	TN	245.0 ± 23.2 mg L-1	122.5 ± 11.6 g/d	23.6 ± 28.9 mg L-1	11.8 ± 14.5 g/d	0.904
	NH4-N	68.6 ± 55.7 mg L-1	34.3 ± 27.9 g/d	1.9 ± 1.8 mg L-1	1 ± 0.9 g/d	0.972
	NO3-N	0.8 ± 0.5 mg L-1	0.4 ± 0.3 g/d	8.9 ± 3.8 mg L-1	4.5 ± 1.9 g/d	–
	PO4-P	10.3 ± 4.9 mg L-1	5.2 ± 2.5 g/d	10.1 ± 5.5 mg L-1	5.1 ± 2.8 g/d	0.019
	Total coliforms	7.24 × 106 MPN/100mL		1.34 × 105 MPN/100mL		1.73 log10
	E. coli	3.09 × 105 MPN/100mL		4.09 × 102 MPN/100mL		2.88 log10

4.4.3 Biomat Position

Research performed by Knappe et al 2020 analysing the spatial variation of water retention using networks of soil moisture sensors. These sensors employed at the same research sites as this study were effective in determining the position, growth rate and the effective hydraulic conductivity of the biomats. This was achieved by classifying any region of the STU as an established biomat if it can maintain a mean increase of VWC 0.025 $cm^3\,cm^{-3}$ above the baseline value over a period of 30 days. The biomat position had extended horizontally for 15 m from the inlet for PE-STUs at site A and B. Site A exhibited a faster growth rate, with the biomat reaching 15 m in 10 months, and site B reaching that length in 13 months. Growth was more muted at SE-STU with the biomat horizontally extending for 7.5 m and 10 m for site A and B after 3 years in operation, respectively. Vertical growth of the biofilm appears to be limited to approximately 5 cm from the infiltrative surface.

4.4.4 Microbial Community Composition Within The STUs

For the 92 samples sequenced, this resulted in a total of 73,332,182 reads with a read depth per sample ranging from 29,706 (min) – 220,139 (max). Read depth was rarefied to the approximate average value of 35,000 reads per sample (**Figure SA1**). The most abundant phyla in control subsoil for Site A were *Acidobacteriota, Actinobacteriota*, and Proteobacteria (17.08 ± 0.17 %, 14.62 ± 4.47 %, 14.24 ± 6.46 %, respectively), and for Site B, *Acidobacteriota, Proteobacteria* and *Chloroflexi* (27.35 ± 4.1 %, 9.77 ± 0.41 %, 11.16 ± 7.77 %, respectively) (**Figure 4.3A; Table SA3**). Changes in abundance of key phyla relative to the control subsoil (**Figure 4.3B**). Within STUs there was a pattern of top phyla being composed of *Proteobacteria, Acidobacteria*, and *Chloroflexi* for PE site A (25.13 ± 4.02 %, 14.62 ± 2.37 %, 12.82 ± 4.4 %), PE site B (14.69 ± 1.95 %, 14.11 ± 2.27%, 15.84 ± 3.95%) and SE site A (18.32 ± 1.71%, 14.82 ± 2.25%, 14.06 ± 2.81%), respectively. At site B SE-STU where the *Proteobacteria, Firmicutes* and *Bacteroidetes* where the dominant phyla with abundances of 39.29 ± 1.75%, 19.89 ± 4.49%, 9.84 ± 1.56% respectively. *Gammabacteria* consisted of the majority abundances of Proteobacteria across all STU (**Figure SA6**).

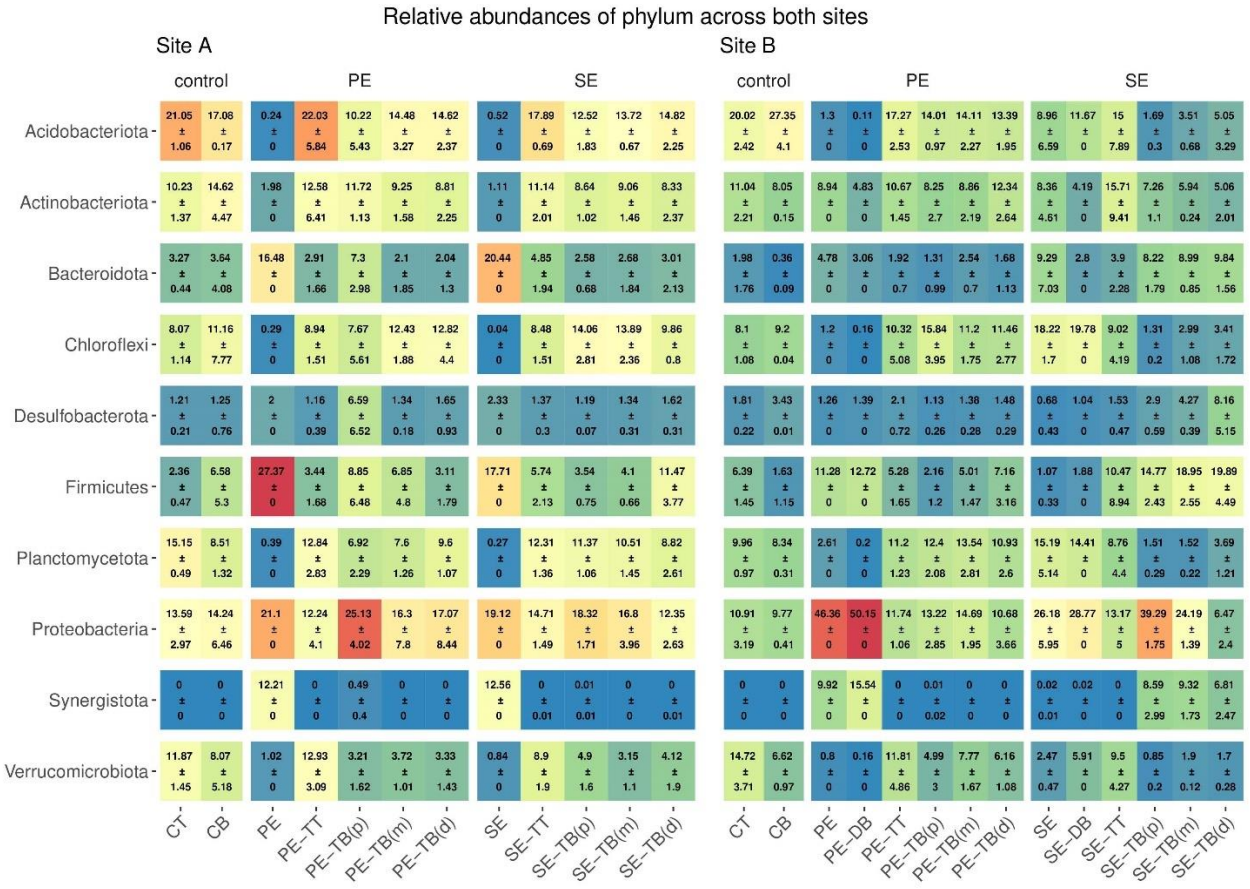


Figure 4.3A Mean ± SD of relative read abundance of phylum-level analysis for site A and site B for both the Primary (PE), secondary (SE), and control samples for each 'system': control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 12 m

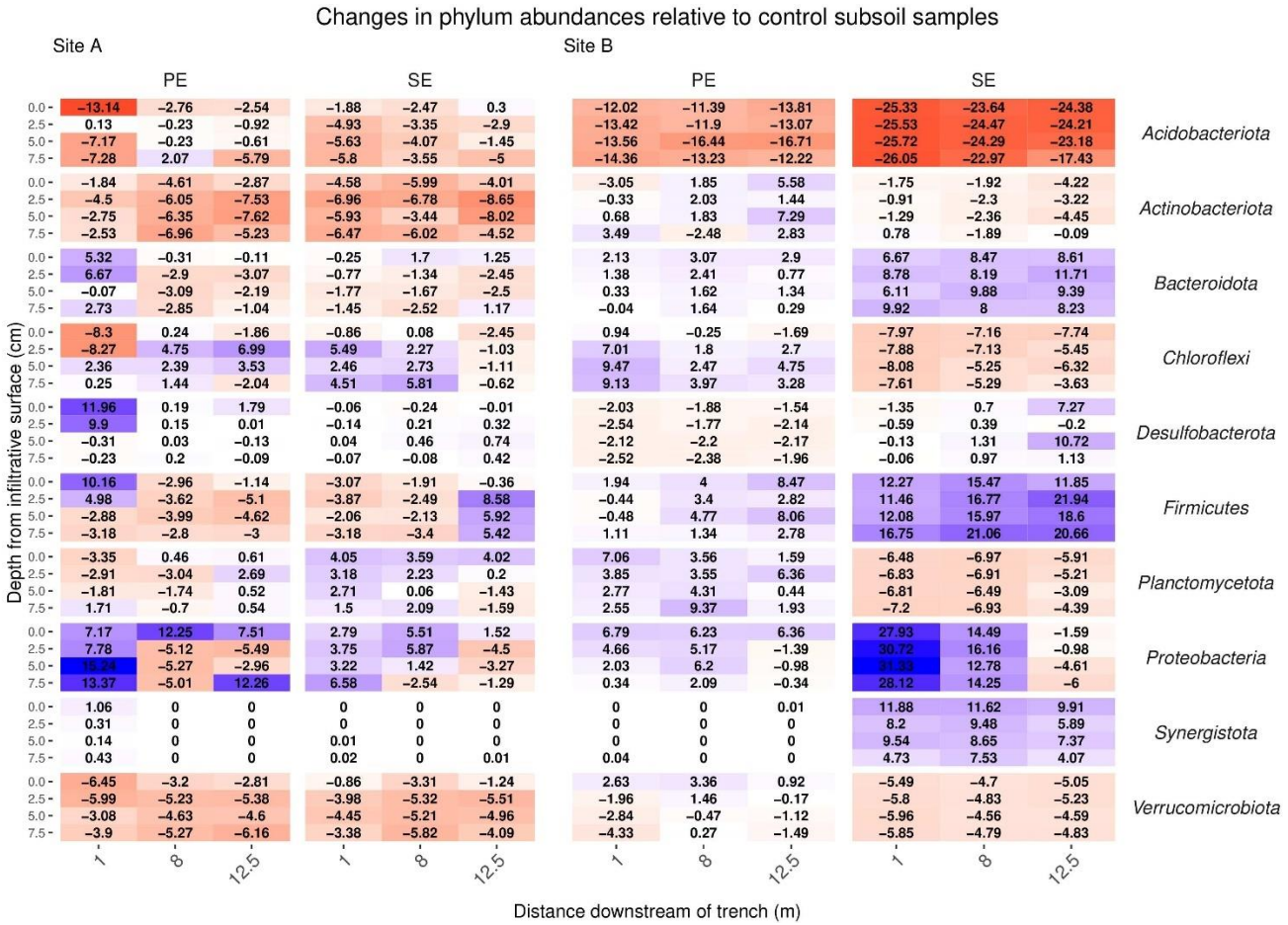


Figure 4.3B. The changes of key phyla relative abundance relative to control subsoil communities, increases are highlighted in blue and reductions in red.

The top 10 most abundant species-level taxa were selected for and compared across compartments within the systems at both sites (see supplementary table S5). The control subsoil samples at both sites contained high abundances of species belonging to the phylum *Acidobacteriota* (site A; 3.6 ± 0.54%, site B; 2.7 ± 0.42%), Actinobacteria (site A; 1.15 ± 0.77%, site B; 0.51 ± 0.16%) and *Chloroflex*i (site A; 2.02 ± 2.08%, site B; 1.29 ± 0.09%) (**Figure 4.4A**). In PE effluent samples Firmicutes was the most abundant genus, with abundances of 0.55 and 3.33 for site A and B respectively. In Site B SE samples *Metanosarcina sp* and *Thauera sp* genus were also high with abundances of 3.1% and 5.98%, respectively. In the PE fed STUs, Mycobacterium sp (site A; 0.78 ± 0.12%, site B; 0.39 ± 0.06%) was the most abundant species within both sites, followed by *Pirellula sp* (1.08 ± 0.3%) in Site A, and Bacillus (0.8 ± 0.21%) in Site B. The community profiles for topsoil samples did not vary largely between samples at both sites and effluent types. At site A with the phylum *Acidobacteriota* and Actinobacterium having abundances of 22.03 ± 5.84 %, 12.58 ± 6.41 % in the PE-TT respectively, abundances in the SE-TT were 17.89 ± 0.69% and 11.14 ± 2.01%, respectively. At site B *Acidobacteriota* and Actinobacterium abundances were 17.27 ± 2.53%; 10.67 ± 1.45% in the PE-TT respectively, abundances in the SE-TT were 15 ± 7.89% and 15.71 ± 9.41%, respectively. The greatest variance within the microbial community structure was noted from the STU base samples: changes in genus abundances were noted in respect of distance from the inlet to the trenches and the relative read abundance within the subsoil control sample.

At Site A, *Actinobacteriota* species sequences at 1 m along the PE fed trench was higher by an average of 2% in relative read abundance when compared to the control subsoil sample see **Figure 4.4B**. For the SE fed STU *Nitrospira sp* increased by an average of 1% across all distances sampled (1 m, 5 m and 12 m) along the STU compared to relative sequence read abundances found within the control subsoil. In the PE-STU at Site A, there was an average drop of 3% of the sequence read abundance relative of species belonging to the Acidobacteriota phylum at 1 m along the trench and an average of 1% at 12 m sampling point. The parallel SE fed STU saw a 1% reduction across the sampled length of the trench. There was an average drop of 1–2% of Firmicute *Bacillales* and Verrumicrobiota *Chtoniobacterale*s species sequences in both PE and SE-STUs. In Site B, there was an increase of 2% for Firmicute species sampled at 5–12 m from the head of the trench. The greatest increase in species in Site B was Firmicute species at 2% and 7% for PE and SE fed STUs respectively. There were 3–10% increases in *Methanosarcina sp*, 3–4% *Synergistota* species and 9% in Proteobacteria *Thauera sp* within the SE-STU relative to abundances found within the control subsoil, as shown in figure 4. In general, changes to microbial community composition were more evident at the phylum level with regards to the application of primary, partially, and fully treated effluent, which resulted in a negative response of Acidobacteria and a weak to very strong positive response of Proteobacteria. At a more granular species level some of the greatest variation in composition was noted at the SE fed STU, where there were large increases of the relative abundance of *Methanosarcina sp, Romboutsia sp* and *Thauer*a sp, *Romboutsia* sp and *Thauera sp*.

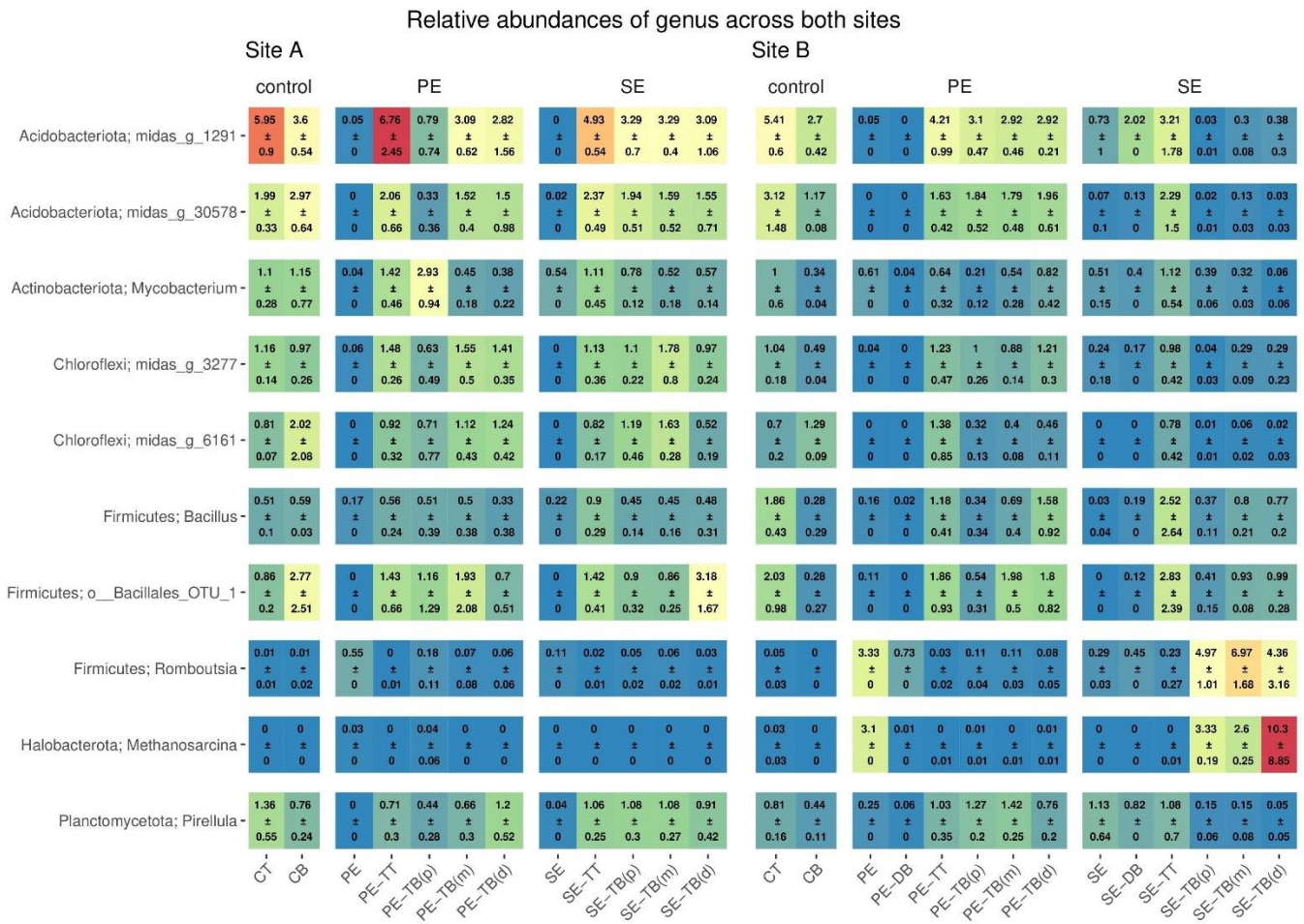


Figure 4.4A. Mean ± SD % of relative read abundance of genus-level analysis for Site A and B for both the Primary (PE), secondary (SE), and control samples for each 'system': control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 12 m

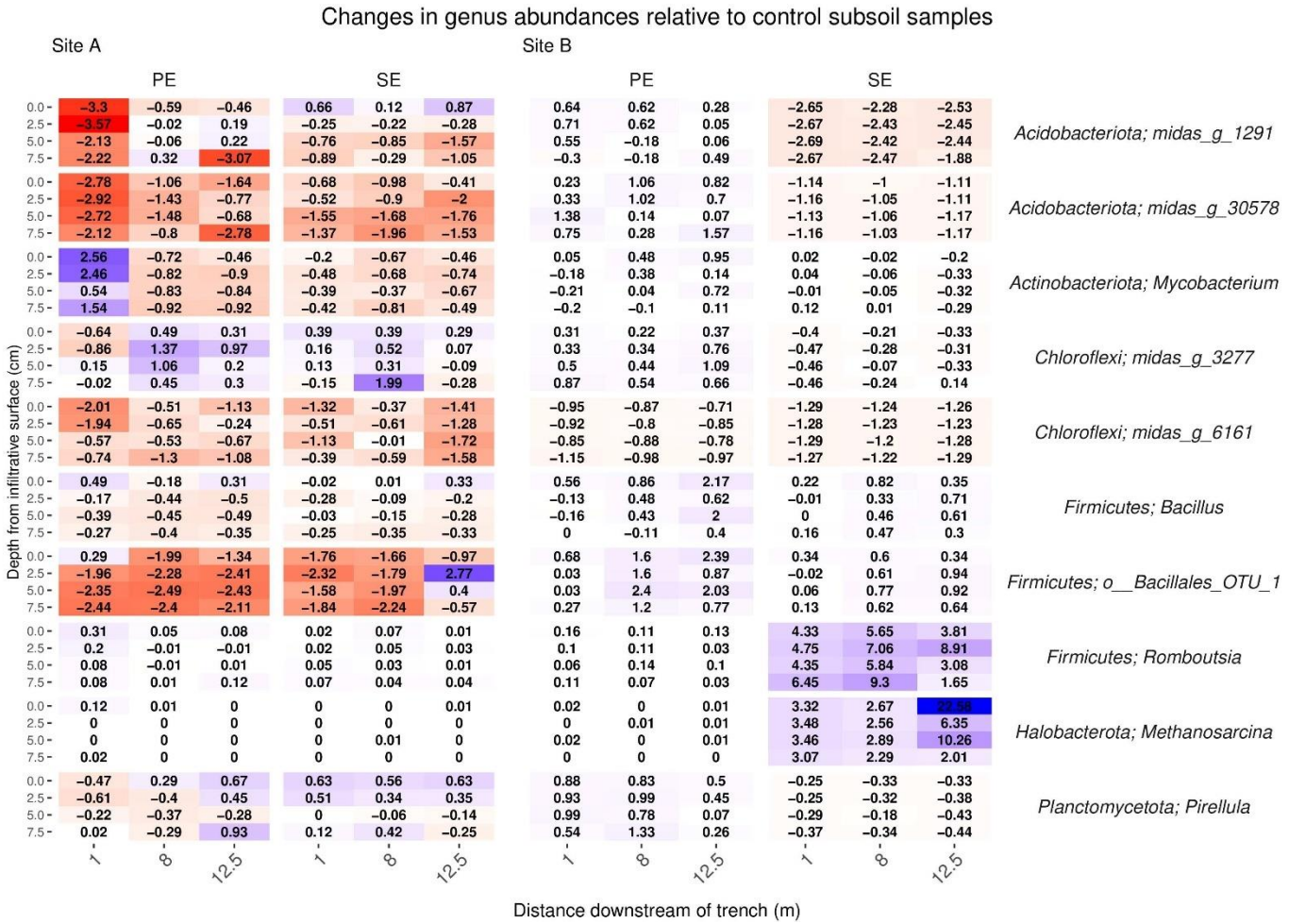


Figure 4.4B. 2D spatial profile of changes of key phyla relative abundance in respect of the control subsoil communities, increases are highlighted in blue and reductions in red

4.4.5 Biomat Microbial Community Structure in Response to Effluent Dispersal

In comparing the alpha diversity (Shannon) and species richness (Chao1) between all compartments across both sites, Site A was significantly richer in species (p=0.03, Wilcoxon test), but there was no significant difference in diversity (p>0.05, Wilcoxon test). Although site A control subsoil samples were on average more diverse and rich in species, there was no significant difference in diversity or richness between the control soil samples (**Figure SA3; Table SA2**). Mean diversity values within the PE fed STUs of Site B was significantly greater than in Site A (p=0.0008; Wilcoxon test).Site A SE-STU was significantly more rich in species ($P \leq 0.0001$; Wilcoxon test) and diverse ($P \leq 0.0001$; Wilcoxon test) than that of site B. Statistical analysis of alpha diversity and species richness (**Figure 4.5**) confirmed that the Site A PE fed STU diverged from the control. In Site B, there was a greater significant difference between the SE-STU samples and the other sample groups (**Table 4.5**).

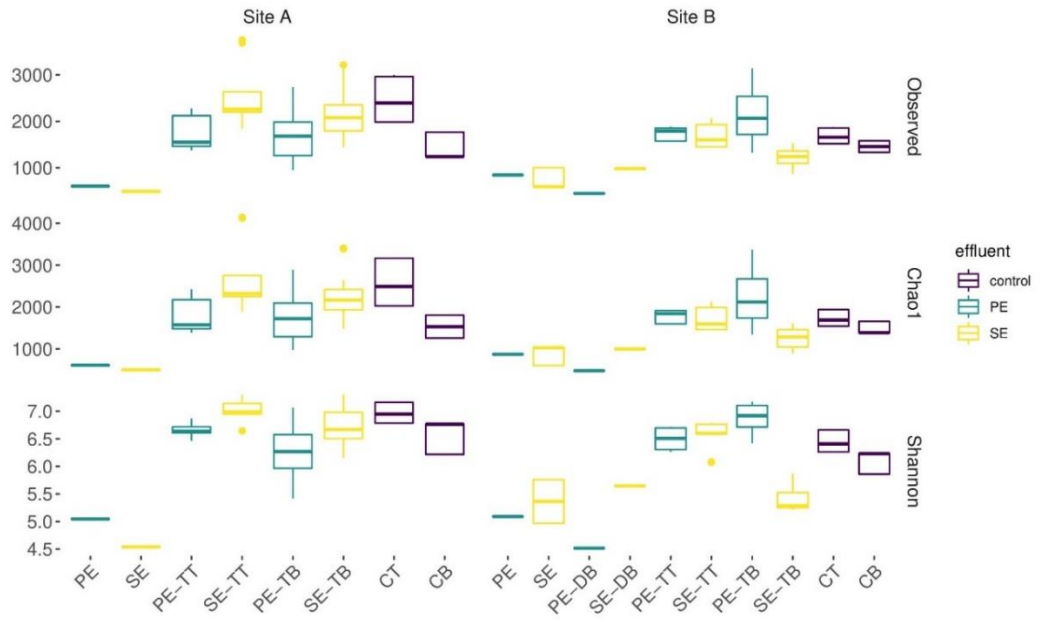


Figure 4.5. Boxplots displaying rarefied data for observed OTUs, species richness calculated using an abundance-based coverage estimates (Chao1) and alpha diversity (Shannon). Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), distribution box biofilms (DB), STU top (TT), STU base (TB), control top (CT) and base (CB).

Table 4.4. displays the Wilcoxon test values for comparative intra-site analysis.; ns P> 0.05, * $P \leq 0.05$,** $P \leq 0.01$, *** $P \leq 0.001$ and **** $P \leq 0.0001$. Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), distribution box biofilms (DB), STU top (TT), STU base (TB), control top (CT) and base (CB).

	PE-TB vs SE-TB		PE-TB vs CB		SE-TB vs CB	
	Shannon	**Chao1**	**Shannon**	**Chao1**	**Shannon**	**Chao1**
Site A	*	**	ns	ns	ns	ns
Site B	****	***	*	ns	*	ns

A cross-sectional analysis of horizontal and vertical dimensions of the different STUs across both sites indicated the presence or absence of species richness and diversity ‘’*hotspots*’ within both systems (**Figure 4.6**).

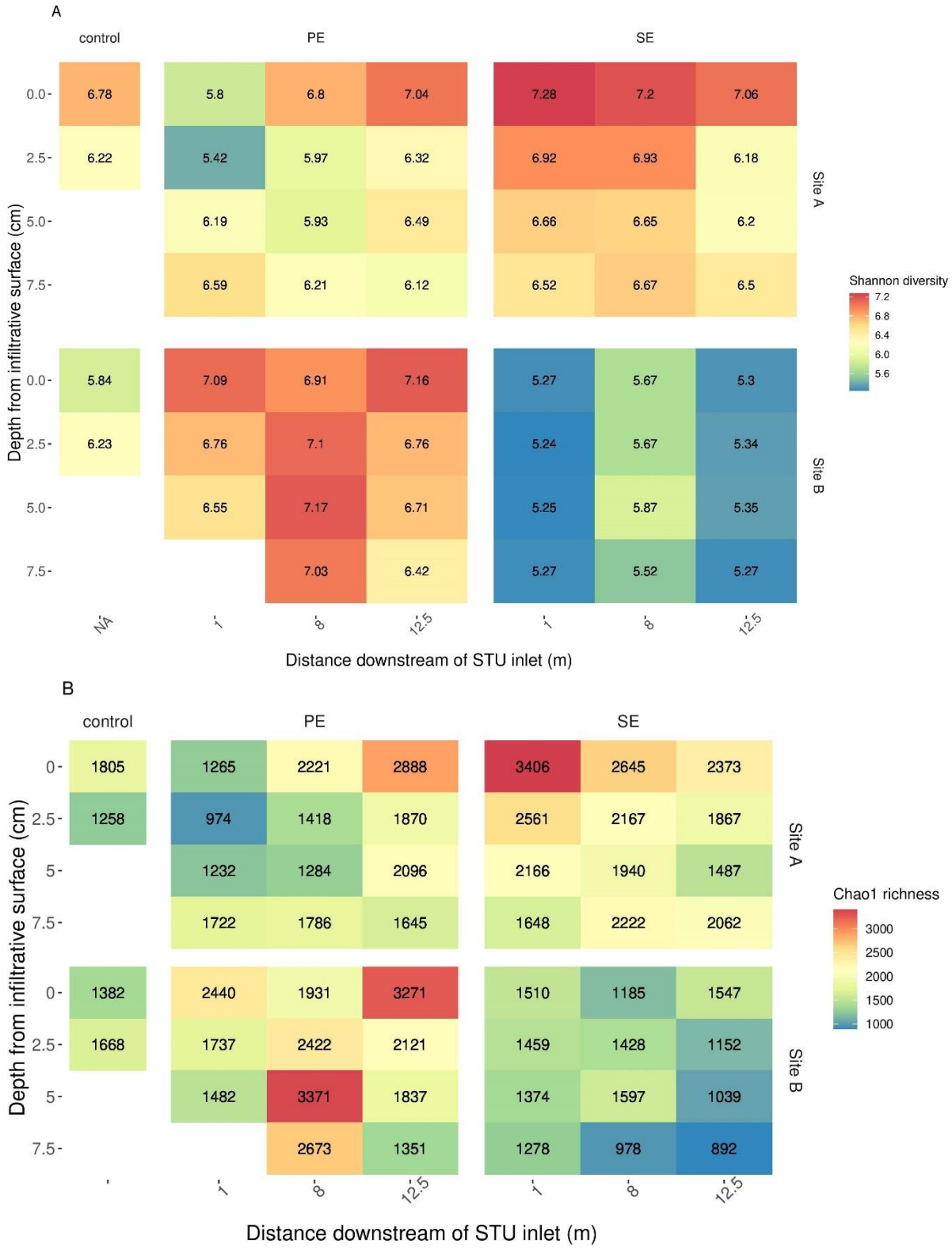


Figure 4.6. The Shannon diversity (A) and Chao1 richness (B) across distance and depth of the sites’ primary (PE) and secondary effluent (SE) STUs.

PCoA plots were used to investigate the beta diversity using weighted Unifrac, accounting for the relative read abundance within samples. One large cluster consisting mainly of STU samples are

positioned close to the subsoil controls, with the only exception being PE samples within 1 m of the STU inlet. This suggesting little dissimilarity between STUs and control subsoil samples for Site A. In contrast Site B presented much more distinct clusters (**Figure 4.7**), with a clustering of shallow subsoil samples with the control soil samples with PE-STU samples creating a minor cluster. A second separate cluster of SE-STU samples and the PE effluent sample was present, distinct from the STU, control subsoil and RBC samples.

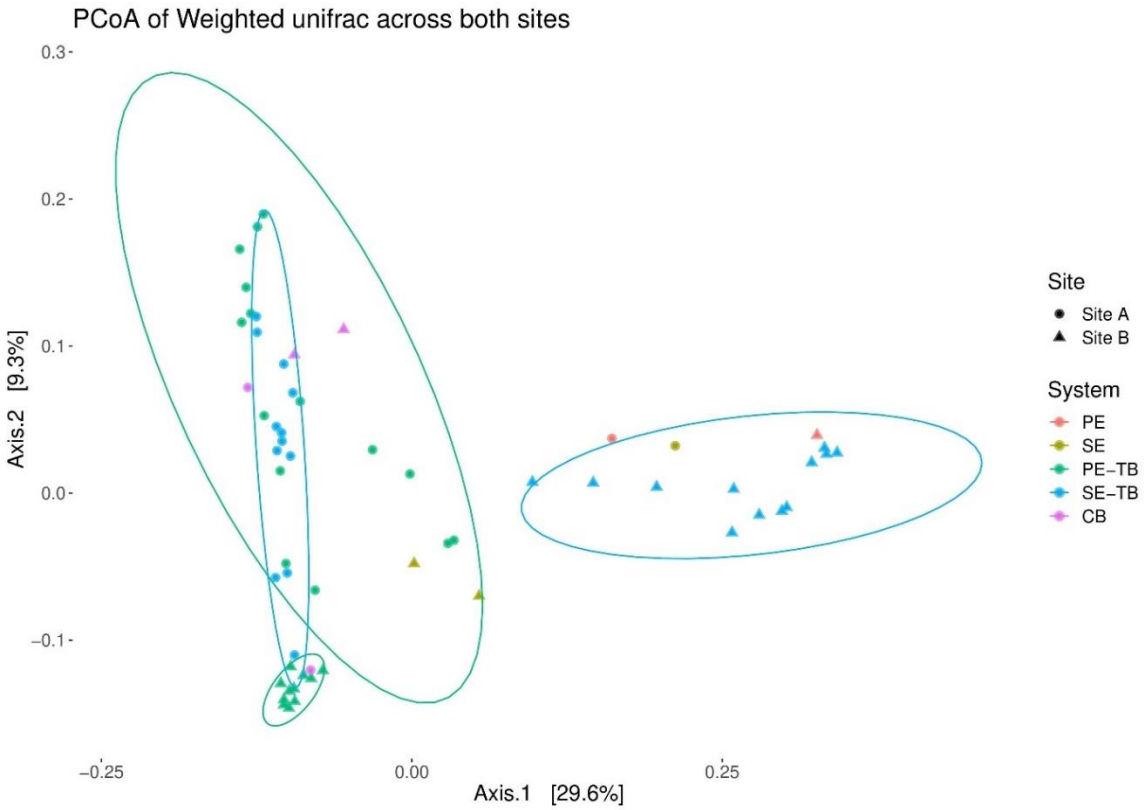


Figure 4.7. Principal coordinates of beta diversity based on weighted Unifrac distances within STUs at both sites. Each data point represents a sample taken from either PE effluent stream, SE effluent stream or control soils. Samples are further subdivided based on the position within the system, i.e. at the STU; base, top, control; base, top and pure effluent samples.

Permutational analysis of variance applying distance matrices results indicated that there was no significant difference between the control subsoil communities across sites (Pr=0.33, R^2 = 0.49). There were significant differences in the microbial community composition between the topsoil control samples and the subsoil samples (Pr=0.018; R^2 = 0.26), and statistically significant differences between STUs of the same effluent type when compared between sites (PR=0.0004; R^2 = 0.22 and PR=0.0004; R^2 = 0.44, for PE and SE respectively). Key factors impacting variance within sites showed effluent and distance as significant, whereas depth did not have any statistically significant effect on community structure (**Table 4.5**).

Table 4.5. displays the Adonis test (ns Pr> 0.05, * Pr ≤ 0.05,**Pr ≤ 0.01, ***Pr ≤ 0.001 and **** Pr ≤ 0.0001) values for comparative intra-site permutational analysis of variance applying distance matrices. Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), STU base (TB), and base (CB).

	PE-TB vs CB		SE-TB vs CB		PE-TB vs SE-TB		PE-TB vs Distance (m)		SE-TB vs Distance (m)	
	Pr	**R^2**	**Pr**	**R^2**	**Pr**	**R^2**	**Pr**	**R^2**	**Pr**	**R^2**
Site A	0.39 ns	0.06	0.23 ns	0.13	0.004 **	0.16	0.0004 ***	0.34	0.001 ***	0.34
Site B	0.009 **	0.32	0.01 **	0.38	0.004 **	0.49	0.003 **	0.36	0.0009 ***	0.65

4.4.6 Target Organisms Screened for Biogeochemical Functionality

Target putative functional groups were classified by means of the MIDAS database, with the only exception being the Anaerobic Methane Oxidisers (AMO) group which was determined from literature. Relative read abundance was measured and compared across sites, STUs, and environmental compartments (topsoil, subsoil), as shown on **Figure 4.8**. Differences in relative abundances for putative functional sequences were compared between control subsoil samples and STUs for both sites. Key changes in functional groups relative abundance is summarised in Table 6 below.

For STUs dosed with Primary effluent, denitrifying bacteria, polyphosphate-accumulating organisms (PAO), nitrite oxidising bacteria (NOB), methanogens, glycogen accumulating organisms (GAO), ammonia-oxidising Bacteria (AOB), AMO and Acetogen functional groups showed higher relative abundances when compared to the control subsoil sample (**see Table 4.6**). The only decrease in abundance was for the NOB functional group at site A which saw a reduction was proximal sampling (1 m) within the PE STU biomat (**see Table 4.6**). In the Site B PE fed STU samples also showed the presence of Anammox species sequences at relative abundances ranging from; 0.01%, 0.02%, and 0.06% in sampling positions of 1 m distance at 0 cm depth of the infiltrative surface; 5 m distance at 5 cm depth of the infiltrative surface; and 5 m distance at 7.5 cm depth of the infiltrative surface (**Table 4.6**; **Figure 4.8A**), respectively.

For the SE fed STU samples, all functional groups (Denitrifying bacteria, PAO, methanogen, GAO, AOB, AMO and Acetogen) showed an increase in relative read abundance relative to the control subsoil sample. The only notable reduction was the NOB functional group in site B, 0.6 – 0.7% lower than the abundances found within the subsoil control samples, see table 6. At the SE-STU at site B there were large increases in relative abundance of Denitrifiers, Methanogens and Acetogens (see table 6). For denitrifiers and GAOs, much of the increases were around sampling points proximal to the inlet of the SE fed STU 1 m position (see figure 4.8B). Methanogen increases in relative abundances were at a maximum at the rear of the STU at 12.5 m of the inlets whilst maximum Acetogen relative abundances occurred at the midpoint of the STU at 5 m (**Figure 4.8B**).

Table 4.6. Key changes (+/– , NC; 'No Change') in relative abundance (%) of functional groups; denitrifying bacteria, Nitrite Oxidising Bacteria (NOB), Polyphosphate-accumulating organisms (PAO), Methanogens, Glycogen Accumulating Organisms (GAO), Ammonia Oxidising Bacteria (AOB), Anaerobic Methane Oxidizers, Acetogens, Anammox in respect of control subsoil samples. Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), and Soil Treatment Unit (STU).

	Denitrifiers	**NOB**	**PAO**	**Methanogens**	**GAO**	**AOB**	**AMO**	**Acetogen**	**Anammox**
Site A									
PE STU	(+) 2 - 3.3	(+) 0.2 - 0.7 (5–8 m) (-) 0.4 (1 m)	(+) 0.1 – 0.5	(+) 0.006 – 0.3	(+) 0.1–1.2	(+) 0.006 – 0.3	(+) 0.075 –0.245	(+) 0.053 - 0.185	NC
SE STU	(+) 0.7 – 2.3	(+) 0.8 – 0.9	(+) 0.7 - 2.3	(+) 0.01 – 0.07	(+) 0.003 – 0.02	(+) 0.02 - 0.1	(+) 0.01 – 0.1	(+) 0.007 - 0.142	NC
Site B									
PE STU	(+) 0.3 – 0.55	(+) 0.1 – 0.429	(+) 0.01 – 0.04	(+) 0.08 – 0.1	(+)0.05 – 0.1	(+) 0.02 – 0.03	(+), 0.03 - 0.04	(+) 0.08 – 0.1	(+) 0.001 - 0.01
SE STU	(+) 3–13 (1–5 m) (-) 0.03 (8 m)	(-) 0.6 – 0.7	(+) 0.14 – 0.5	(+) 9–15	(+) 0.05–8	(+) 0.02 - 0.05	(+) 0.3 – 1.8	(+) 4.5 – 7.3	NC

'

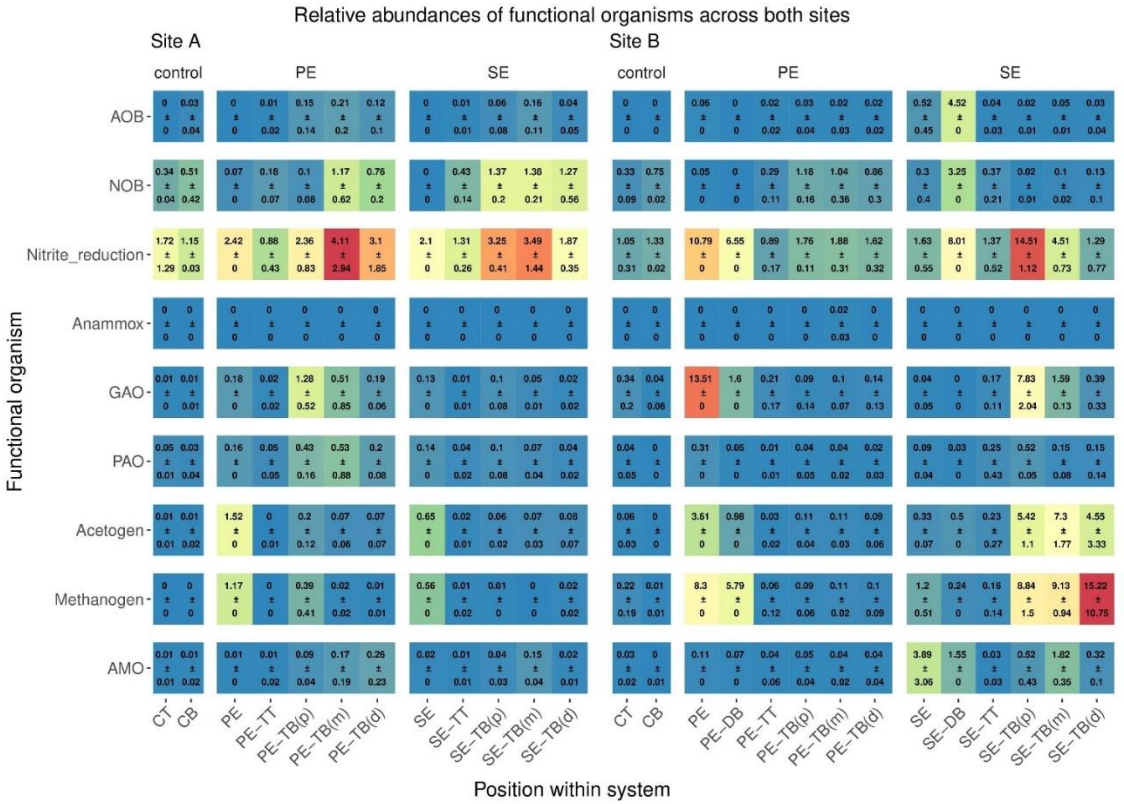


Figure 4.8A. Mean ± SD of relative read abundance of Anaerobic Methane Oxidisers (AMO), denitrifying bacteria, Polyphosphate-accumulating organisms (PAO), Nitrite Oxidising Bacteria (NOB), Methanogens, Glycogen Accumulating Organisms (GAO), Ammonia Oxidising Bacteria (AOB), AMO and Acetogen functional groups for Site A for both the Primary (PE), secondary (SE), and control samples for each 'system': control; base (CB), topsoil (CT), STU top (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 12 m.

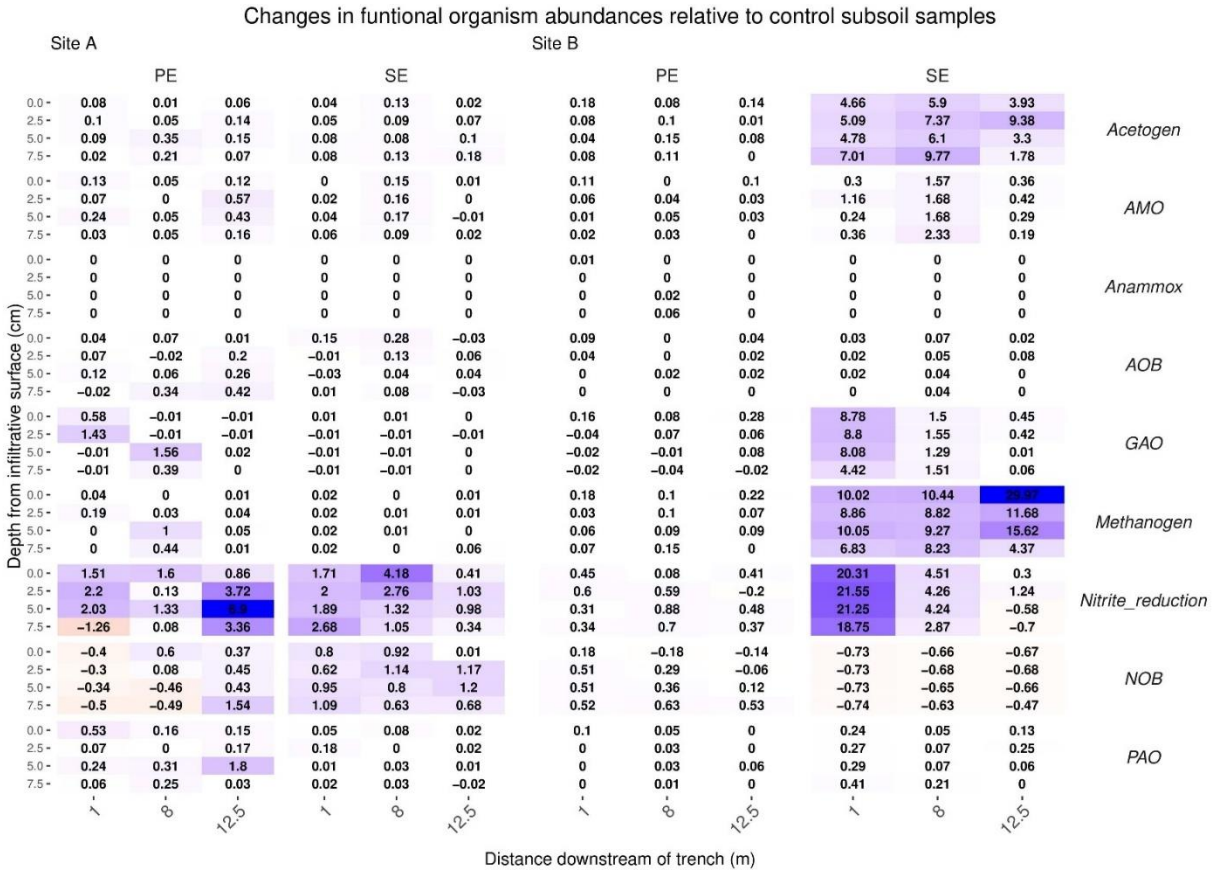


Figure 4.8B. 2D spatial profile of changes of relative abundances of Anaerobic Methane Oxidisers (AMO), denitrifying bacteria, Polyphosphate-accumulating organisms (PAO), Nitrite Oxidising Bacteria (NOB), Methanogens, Glycogen Accumulating Organisms (GAO), Ammonia Oxidising Bacteria (AOB), AMO and Acetogen functional groups sequences in respect of the control subsoil communities, increases are highlighted in green and reductions.

4.5 Discussion

The main aim of this study was to determine the influence of different levels of treated domestic wastewater on the microbial community structure of soil treatment units (STUs). Two research sites were selected due to their proximity of each other, being in the same climate region and soil type. Both sites are classified as typical Luvisol soils, averaging at pH 8 with little known variance (although not measured in this study) and both sites also employ similar land management practices (rural, unproductive domestic households) (Karimi et al., 2020; Fierer and Jackson, 2006; Fearly, 2009). The lack of significant difference in the species richness, alpha diversity, and the community composition of subsoil control samples between both sites proved the proximity effective to allow a direct comparison of the STU systems. Pre-treatment had a significant effect on the alpha diversity and the richness of species between STUs across sites. Between SE-STUs site A was significantly richer and more diverse in species although its biomat was 2.5 m shorter than site B. This shorter biomat is likely because of greater pore size noted by the higher ksat values measured at site A. Variations in richness and species diversity in PE-STUs were less significant and can be attributed to soil pore size and average loading rates, site B PE organic loading rates were more than double that of Site A. (Bastida et al., 2017; Dang et al., 2019).

The 2D spatial analysis of areas of elevated alpha diversity and species richness, which offer important insights into microbial hotspots within the STU. Areas of low richness within the trench may suggest areas of high activity as has been seen in previous studies of microbial activity within the rhizosphere which have low species richness (Reinhold-Hurek et al., 2015). Recent work on structural equation modelling (Bastida et al., 2021) has indicated that soil C content has a role in regulating soil microbial richness by a positive association with microbial biomass (Geyer et Barret., 2019). High levels of richness and alpha diversity relative to control subsoils were found in locations across the STUs with exception of the site B SE-STU. High level of species richness appears to be because of the addition of high nutrient and organic conditions which may suggest locations in which organic carbon may have been incorporated into EPS for the development of the biomat, with its high energy requirements for its production but results in less biomass (Wu et al.,2019). The low richness suggests the end of SE-STUs where effluent is likely to incur less resistance due to the absence of bioclogging may be areas of high activity (Knappe et al., 2020).

The sampling campaign allowed us to assess spatial effect of pre-treatment on alpha diversity and the taxa in the STU. The productivity diversity relationship hypothesis is that once diversity has increased beyond a certain threshold due to resource availability that the diversity outcome becomes negative (Geyer et al., 2019). At the proximal position of the PE-STU of site A, there was an observed reduced diversity relative to the rest of the STU, and an increase of abundance of copiotrophic phyla specifically proteobacteria indicating that this portion of the STU may have tipped the threshold. In Site B PE-STU diversity appears higher than in site A, and the increases in

copiotrophic organisms less pronounced, this may suggest that the younger biomat with a lower growth rate has not yet tipped into being a homogenous community (Knappe et al., 2020).The spatial profile of the taxa Proteobacteria, specifically the class *Gammaproteobacteria* accurately mirrors the position of the biomats across all sites due to its positive response to organic loading in subsoils, however spatial accuracy is lost further down the taxonomical levels (Dang et al 2019; Wu et al., 2019). Interestingly, SE-STU at site B the large increase of *Gammaproteobacteria* taxa relative to the control subsoil at the proximal sampling point (1 m) indicates the degree of pre-treatment may enhance competition. The selection for copiotrophic bacteria in SE-STU at site B is likely due to the steady flow of limited nutrient and organic inputs and lack of pressure for niche complementarity (Naeem, 2009).

Permutational multivariate analysis determined that the level of pre-treatment of the effluent had a significant impact on the community structure of the STUs. The clear divergence of SE-STU from site B from the remaining systems illustrates the variance caused by the level of pre-treatment. (Knappe et al., 2020; Guo et al., 2018). Permutational multivariate analysis also confirmed that spatial factors such as the horizontal distance accounted for a great degree of the variance in community structures, with SE-STUs particularly affected. The large variance within communities of SE-STUs highlights the presence of 'infiltrative dead zones' at the distal location of the underused trench, resulting in a heterogenous community composition within the distal portion of the STU (Knappe et al., 2020). The effect of subsoil depth was insignificant on the community structure, this is contrary to several biogeographical surveys of the natural subsoil (He et al.,2017; Uksa et al., 2015; Eiler et al.,2012). The lack of any significant effect of vertical distance from the infiltrative surface, maybe that depth of soil cores at 7.5 cm was insufficient to assess the diversity. That all depths sampled only encompassed areas of the infiltrative surface impacted by the biomat. Future studies should incorporate deeper cored samples to confirm the true effect of depth in the STU.

This study identified ammonium oxidation (annamox) bacteria Candidatus Brocadia anammoxidans in the PE fed STU of site B. Candidatus Anammoximicrobium was also detected in the effluent of the RBC. The presence of annamox in PE-STU corresponds to site descriptions noted by Knappe et al. (2020) who confirmed ponded anaerobic conditions within the PE-STU at site B. Annamox reactions have been of interest as a low-energy alternative for nitrogen removal within wastewater treatment plants but are also known to be naturally active within soils and wetlands (Kartal et al., 2010; Bagnoud et al., 2020). The presence of annamox may be due to high concentrations of organic carbon in the STU resulting in concurrent denitrification with heterotrophic Denitrifiers (Chamchoi et al., 2008). These results match previous studies that suggested that anammox was active within low flow sites and at 5 cm -7.5 cm depths within the infiltrative surface (Humphreys Jr et al., 2019 Cooper et al., 2016). Increases in functional richness may help locate metabolic activity such as the presence of Denitrifiers at the proximal position to the inlet at the SE-STU (Louca et al., 2018). However, whether an increase in functional richness resulted effective attenuation was not evident

in this study. Pre-treatment often results high nitrate and low organic carbon effluent which increased the relative read abundance of denitrifiers, but in a stunted biomat with low hydraulic retention time results significantly reduced attenuation of TN (Knappe et al., 2020; Gill et al., 2009). Our study has shown that PE-STUs has shown less pronounced increases in Denitrifiers compares to SE-STUS, but STUs receiving primary effluent has been noted as capable of removing six times the total nitrogen (Gill et al, 2009). This suggests that the increases in functional richness within the STU are secondary to bioclogging, as metabolic rates could be limited by hydraulic conductivity.

4.6 Conclusion

- This study presents the first direct microbial comparative analysis between STUs for on-site wastewater treatment systems receiving domestic effluent with different levels of pre-treatment. This analysis has been conducted in STUs which has already been successfully characterised for soil clogging, under the same environmental, hydrological and subsoil conditions.

- The microbial community richness and diversity within the STU system were significantly affected by the level of pre-treatment of the wastewater. This outcome appears to follow the productivity diversity relationship theory. The effect is not linear and the addition of a steady flow of lower concentrations of organics and nutrients may be initially more selective for copiotrophic bacteria than the raw effluent. This selection is due to known fitness such as motility exhibited by copiotrophs.

- The STU receiving the fully pre-treated effluent contained the highest relative abundance of functional communities. Functional richness may not indicate the attenuating capacity of the system. Attenuation appears to be linked to the extent of the coverage of the STU by the biomat. This suggests that there is functional redundancy within the STU community. However, the presence of functional groups is secondary to bioclogging, as metabolic rates could be limited by hydraulic conductivity.

- The community structure or beta diversity of the STUs was significantly impacted by the level of pre-treatment and the horizontal distance downstream of the inlet. Depth appeared not to impact along each STU appears to have a greater impact on community structure than depth in each STU. The microbial community structure found within the STU receiving fully nitrified effluent was significantly different from that of its raw effluent counterpart.

- This study has effectively profiled two effective and defective STUs from two separate research site. More temporal data is required to assess the development of these communities in the field. In doing so it will be possible to establish critical points of transition for the STUs, profiling communities at different stages of the growth of the biomat will provide researchers within depth understanding of biological clogging process and how to manage it.

5. Influent Pre-treatment of Greater Significance than Porosity in Microbial Community Assemblage Patterns in Soil Treatment Units

Alejandro Javier Criado Monleon, Katherine Hardgrave, Muhammad Ali , Laurence Gill

This manuscript is being prepared for submission.

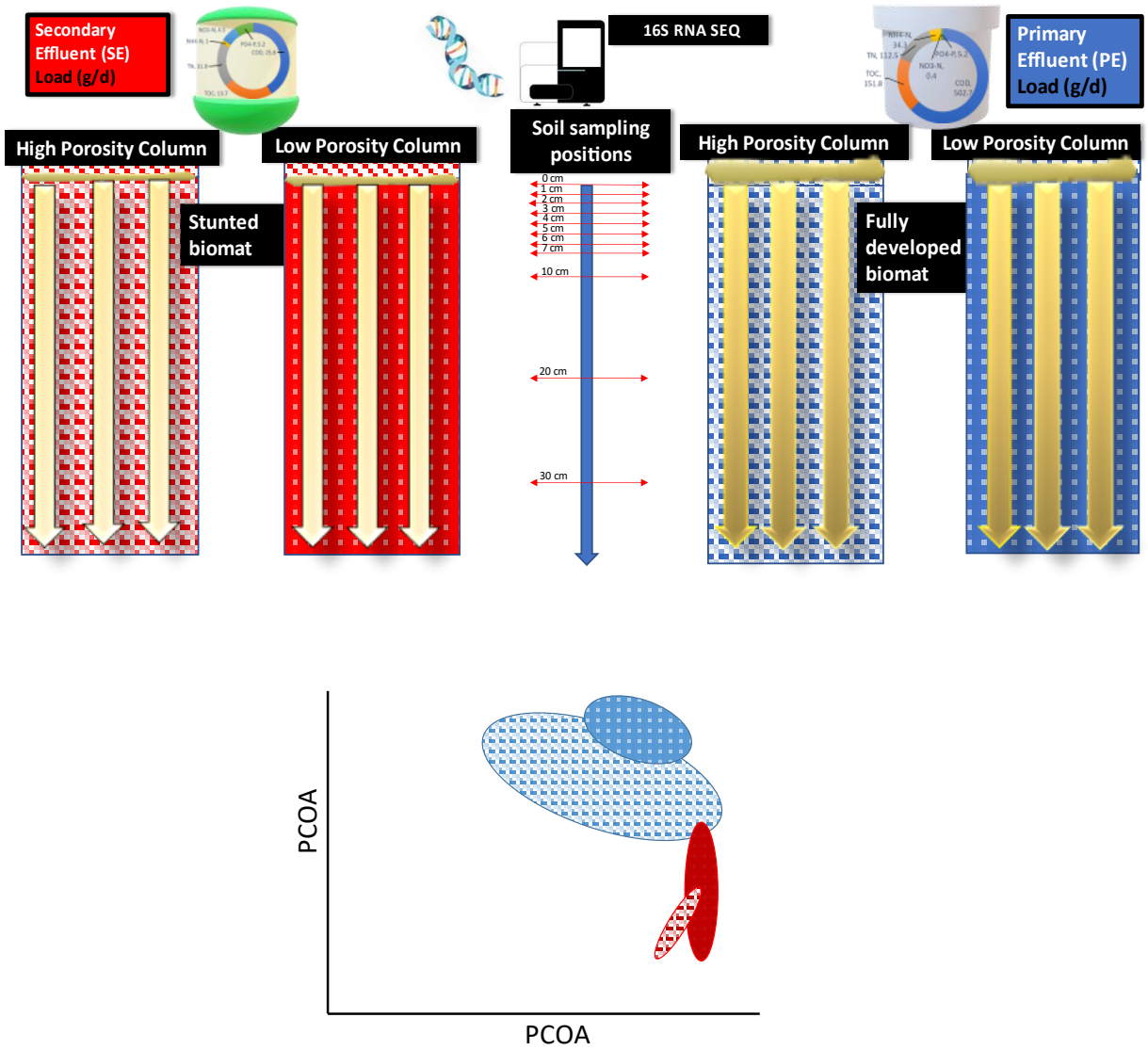


Figure 5.1. An overview of the experiment design, and the effect of effluent pre-treatment and porosity on the microbial community structure, analysed through 16S RNA sequencing.

5.1 Abstract

Domestic on-site wastewater treatment systems (OWTSs) are configured with a variety of pre-treatment technologies, resulting in varying degrees of influent treatment before being distributed to the soil treatment unit (STU). The subsoil's texture and the application of pre-treatment directly influence the development of microbial mats, or 'biomats,' within the STU. In this study, we monitored four separate soil columns for 510 days. The columns were packed with subsoil of two levels of porosity (High: 26%; Low: 21%), and these columns were dosed with either primary treated effluent (PE) from a conventional septic system or secondary treated effluent (SE) from a rotating biodisc contactor (RBC) at the same household, resulting in approximately 90% Total Organic Carbon (TOC) removal. The columns were monitored for effluent pollutants and faecal indicator bacteria removal efficiencies, and changes in volumetric water content (VWC) using an automated 2-dimensional network of sensors to determine the establishment of the biomat. The samples from varying depths of four columns, influents, and effluents (total: 53 samples) were characterised by 16S amplicon gene sequencing analysis. Our findings demonstrate that pre-treatment has a stronger effect on biomat community assemblage patterns than subsoil texture. We observed that greater pre-treatment led to more immigration at 3.8%, while columns dosed with PE exhibited greater levels of species sorting at 88.1% of sequence relative read abundance. These findings will significantly impact research and development efforts focused on rapid soil treatment systems and managed aquifer recharge projects.

Keywords:

Wastewater, Nature-based solutions, Microbial ecology, Community assemblage patterns, On-site wastewater treatment

5.2 Introduction

According to the World Health Organization (WHO), 41% of the global population employs on-site sanitation connected to a septic system (WHO, 2020). Such on-site systems are particularly predominant in middle- and low-income nations and involve the application of different nature-based solutions (NBS) treatment technologies for sanitation, from ponds, constructed wetlands, willow systems and soil infiltration systems or Soil Treatment Units (STUs). The latter can be subdivided into two categories: slow or rapid rate infiltration (IWA, 2021). Slow infiltration requires dispersal at the surface level; however, rapid infiltration can be directly applied deeper into the subsoil, offering a form of managed aquifer recharge capacity (IWA, 2021). Such soil infiltration is a widely used, successful method in on-site wastewater treatment across the world, for example, in the Republic of Ireland, 30% of the population treats its wastewater using such on-site treatment systems.

The principle of contaminant attenuation in such soil-based on-site systems is the passively controlled hydraulic loading of primarily treated domestic wastewater upon a gravel-filled trench (often by means of a gravity-fed percolation pipe), evenly distributing effluent upon the 'infiltrative surface' of the subsoil interface below which is the STU (Winstanley et Fowler, 2013). The STU serves as a passive contaminant attenuation system for the percolating effluent, utilising a combination of physical, chemical, and microbiological processes. Biological clogging is induced by the development of microbial mats or 'biomats' as the result of the product of Extracellular Polymeric Substances (EPS) aggregated within the soil pore structure (Beal et al., 2005; Or et al., 2007).

The relationship between biomat development and reduced hydraulic conductivity has been observed in various studies. Chang et al., (1974) noted that organic particles and microbial growth trapped in columns below dairy wastewater ponds led to a drop in infiltration rates and effective pond sealing within two months of operation. The presence of EPS has been shown to cause a significant reduction in porosity, resulting in 2-3 orders of magnitude drop in hydraulic conductivity within several weeks to four months (Okubo and Matsumoto, 1979; Roberson and Firestone, 1992; Vandevivere and Baveye, 1992; Seifert and Engesgaard, 2007; Or et al., 2007; Hommel et al., 2018; Knappe et al., 2020). Beal et al. (2005) described three stages of biomat development: (1) clogging due to physical processes blocking soil pores with organic matter and suspended solids, (2) enhanced microbial growth from effluent dosing, and (3) biomat development due to accumulated organic matter and EPS blocking soil pores. Studies have reported varying biomat depths of 5-50 mm, characterised by discoloured brown-black soil (Laak., 1970; Chang et al., 1974; Vandevivere et Baveye, 1992; Beal et al., 2005). Depth also appears to play a crucial role in differentiating community composition, as the abundance of organic substrates at the infiltrative surface favours microbial metabolism (Truu et al., 2009).

Improvements in molecular techniques have allowed for effective identification of soil microbial communities. One study that used genetic sequencing on the septic system including the STU biomat determined that the microorganisms found within the septic tank were absent in the biomat (Tomaras et al., 2009). Further studies with the use of next-generation sequencing have shown that the application of pre-treated wastewater effluent to a soil promotes different microbial ecology from that of natural soils (Fernandez-Baca et al., 2019; Dang et al., 2019; Criado-Monleon et al., 2022), confirming the findings in fieldwork studies which have linked the development of the biomat with the level organic and nutrient loading (Bouma, 1975; Siegrist et Boyle, 1987; Gill et al., 2007; Knappe et al., 2020). There appears to be a relationship between natural soil carbon content in regulating microbial diversity and biomass, meaning that environments rich in soil carbon will see increased biomass, but reduced diversity (Bastida et al., 2021). The turnover and residence time of subsoil bacterial communities and organic matter are significantly slower compared to those in the topsoil due to differences in depth and oxygen availability (Truu et al., 2009; Liu et al., 2019). Field studies have revealed that the level of pre-treatment applied to a system influences the diversity of subsoil communities, with systems receiving higher levels of treatment showing reduced diversity. This reduction in diversity may be attributed to a decrease in niche complementarity (Naeem, 2009; Criado-Monleon et al., 2022). Interestingly, certain functional groups within the subsoil show more significant increases relative to control subsoils when subjected to the greatest level of pre-treatment. However, it is essential to note that this increase in functional richness does not always indicate improved contaminant attenuation, as the consequential reduction in bioclogging seems to play a more significant role in the performance of the STU (Knappe et al., 2020; Criado-Monleon et al., 2022). Profiling community structure at critical points in biomat formation and subsequent bioclogging would allow for a better understanding of the biomat development and functionality.

The relative importance of various community assembly mechanisms shaping bacterial community assembly during biomat formation in STU remains unexplored, including the interplay between local processes (species sorting) and regional factors (immigration). How much the assembly of local communities in engineered systems is influenced by local factors (species sorting) versus regional factors associated with immigration, which govern the influx of immigrants from source to local (sink) habitats, remains a persistent question (Bell, 2000; Leibold et al., 2004; Mei and Liu, 2019). Metacommunity theory is an effective ecological framework for understanding the impact of immigration and species sorting in community assembly formation on engineered ecosystems (Ali et al., 2020). A key component presented in this study is to determine if species sorting, or immigration establishes the community structure within the biomat. Variation in the bacterial community composition has largely been explained by environmental factors, with robust species sorting occurring in locations with extreme environmental variation. Generalists appear to be abundant due to their capacity to survive in different environments, and specialists are rarer within a community due to the narrower range of accommodating niches (Székely et al., 2014; Zhang et

al., 2014). This differentiation in communities caused by species sorting can be a determinant in many aquatic ecosystems, specifically river sediment biofilms which are not inoculated by stream communities regardless of effective mixing (Ezzat et al., 2022). Community studies on tidal and river sediments have shown that the composition of exogenous pollutants (TN, TOC, P) and the vertical variation in species sorting considerably impact microbial community assemblage patterns at the interface of water and sediment within these ecosystems (Boer et al., 2009; Gao et al., 2017; Zhang et al., 2021; Li et al., 2022). Immigration is a neutral process, common within engineered water systems (*i.e.* WWTPs), but the application of eco-genomics-based mass balances suggests immigration may not be a significant driver in these systems (Saunder et al., 2015; Mei et al., 2019; Ali et al., 2019; Dottorini et al., 2021). These methods also have implications on the functionality of these systems as they can provide an indication of active and inactive species by determining the net growth rate of target species.

This study aims to determine whether the pre-treatment of wastewater before being dispersed within the soil-treatment unit (STU) biomat will result in conformational changes in the microbial community assembly pattern. The study has been carried out within controlled laboratory conditions to maintain soil temperature for optimum soil column functionality and preclude meteorological variables such as precipitation. This work contributes additional important considerations to the design of decentralised wastewater treatment systems, subsoil irrigation and managed aquifer recharge projects. This research provides high resolution ecological insights of augmented subsoils and trends in functionality in respect to influent pre-treatment levels and soil porosity. These columns can assist designers in determining appropriate design practices in soil augmentation prior to managed aquifer recharge.

5.3 Materials and Methods

5.3.1 Column Setup and Instrumentation

A controlled study was carried out in the laboratory using soil columns intermittently dosed with 1 L d^{-1} of influent of differing degrees of treatment. Dosing was performed at 3-hour increments of 250 ml at 09:00 , 12:00, 15:00, 18:00 to mimic household flow pulses for a total flow of 1 L d^{-1} and hydraulic lauding rate of 22.22 L/m^2/d. Column influents were collected from an on-site wastewater treatment system for a four-person household which employs a 4,760-litre Aswasep two-chambered septic tank (Molloy Precast Products Ltd., Ireland) as primary treatment (PE) and a packaged secondary treatment system an RBC (Klargester BioDisc, Kingspan., Ltd., United Kingdom), consisting of an integrated primary settling chamber, a two-stage biozone, and a secondary clarifying chamber as secondary treatment (SE). Four soil column reactors have been constructed using 6 mm 'MarCryl' Acrylic plastic sheet (Access Plastics Ltd., Ireland) to partition a large glass tank into parallel columns, each at a depth of 30 cm, width of 30 cm and breadth of 8 cm **(Figure SB1)**. Each

column had a soil moisture monitoring network using sensors (EC5, METER Group Inc., USA) installed laterally and horizontally at 10 cm intervals, connected to a datalogger, readings were automatically taken every hour and averaged by day (CR800, Campbell Scientific, U.K). Each column had a designated collection bucket for the storage of its filtrate for further analysis. Each soil column was subdivided into zones relative to depth: 0 – 1 cm is the infiltrative zone, 1 cm – 5 cm proximal zone, 6 cm -10 cm intermediary zone and 11 cm – 30 cm is the deep zone of the subsoil columns.

The columns were filled with a sandy loam subsoil excavated at 1.2 m deep at Redcross in County Wicklow. For two of the columns, the Wicklow soils were blended to a ratio of 4:1 with percolation sand (1 mm: 2%, 90%: 600 µm, 250 µm: 8%) to create a higher permeability soil. The high permeability columns had a porosity of 26% and low permeability columns (just soil) had a porosity of 21%. Total organic carbon (TOC) was determined using a Shimadzu TOC-V analyser (Shimadzu Scientific Instrument, Japan). Nitrogen species were analysed as nitrate-nitrogen (NO_3-N), nitrite-N (NO_2-N), and ammonium–N (NH_4-N) and phosphorus as orthophosphate (PO_4-P) using a Konelab 20i chemistry analyser (Thermo Scientific, Finland). Total coliforms and *Escherichia coli* most probable number (MPN) were quantified using kits (IDEXX collilert-18, Westbrook, Me). Influent and effluent samples were monitored for chemical and biological parameters for 204 days (sampled n = 35) from September 2022 to April 2023.

5.3.2 Sampling and 16S RNA Gene Sequencing

Soil column samples were taken from each column with a 25.4-mm stainless steel corer. Samples were taken after 510 days of the system performance was stable and running at a steady state. At each column, a single sample was taken in each of the depths from the infiltrative interface (0-7 cm, 10 cm, 20 cm and 30 cm) (total column samples; n = 44) **(Figure SB1)**. A sample was taken from the subsoil source (n= 1) to be used as a comparative control and from biofilms (n=2) growing within the inlet tubes to the column's influent (n=2) and column filtrate (n =4) samples were concentrated using 100 kDa a Centricon units (Millipore, Damstad, DE). For DNA extraction of the samples for sequencing, ~3 g of soil was collected from the core portioned in 500 mg and placed into a sterile 2 ml Eppendorf tube and stored at −80 °C with RNALater (ThermoScientific, Waltham, MA). Sample handling was performed with a sterile metal spatula, with sterilising performed between each sample using 70% ethanol. Then, DNA was extracted from a total of 53 samples using a FastDNA™ SPIN Kit for Soil (MP Bio, Santa Clara, CA). DNA concentration was checked using a NanoDrop spectrometer (Nanodrop ND-1000, ThermoScientific, Waltham, MA).

The microbial community was assessed by next-generation amplicon sequencing of the 16S rRNA in a paired-end mode. After extracting the total DNA of samples, the 16S rRNA genes were amplified with forward primer 515f Modified (GTGYCAGCMGCCGCGGTAA) and reverse primer 806R Modified (GGACTACNVGGGTWTCTAAT) (Walthers et al., 2016). Library QC was

performed on these libraries. Qualified libraries were paired-end sequenced on Illumina Novaseq 6000 (Illumina, Inc. USA).

5.3.3 Sequence Processing and Analysis

Paired-end reads were converted into amplicon sequence variant libraries produced using the DADA2 pipeline package with the R program (v1.18.0; Callahan et al., 2016). The trimmed forward and reverse reads were merged with settings −25 M to 230 M. Chimeras were removed with the 'removeBimeraDenovo' function in DADA2 under default settings. Taxonomic rank was derived using the 'assignTaxonomy' function linked to the MIDAS database v 4.8.1 (Dueholm et al., 2021) assigned to the Usearch software package v11 (Edgar, 2013).

Alpha diversity (richness and evenness within the samples) was assessed by computing the number of Operational Taxonomic Units (OTUs), abundance-based coverage estimator (ACE) (Chao and Lee, 1992), and Shannon diversity (Shannon, 1948) for all 53 samples. Principal Coordinate analysis (PCoA) plots were estimated using Bray-Curtis and weighted Unifrac metrics, which were derived from the rarefied OTU table using the ampvis2 v 2.7.11 and phyloseq v 3.6 packages, respectively. Taxanomic rank and putative functional groups were identified using the MIDAS database v 4.8.1 database (Dueholm et al., 2021) assigned with the Usearch software package v11 (Edgar et al., 2013). Further analysis was performed to determine the categorical variables of statistical significance to determine the variation within the microbial communities. Permutational analysis of variance applying distance matrices (ADONIS) using the vegan package v 2.4.2 (Oksanen et al., 2021) was performed to examine several variables based on 2,000 permutations. All analyses and plots were performed on R version 4.1.1 through the Rstudio IDE (R Core Team, 2019).

5.3.4 Calculation of SRT and Net Growth Rates From Amplicon Data

The solid retention time (θ_x) and the net growth rate (μ) of an organism or species level OTU were calculated for each of the depth zones using a mass balance approach already applied by Ali et al., (2019). It was assumed that each of the four column zones was under stable operation (i.e. no change in the number of cells in each of the column depth zones). Cell counts were estimated with colorimetric Bovine Serum Albumin assays (Biorad, CA: USA). Equation 1 was used for the calculation of θx. (1)

$$\theta_x = \frac{p_{xCZ}.N_{xCZ} - p_{xCsZ}.N_{xCsZ}}{P_{xEFF}.n_{xEFF-} p_{xINF}n_{xINF}}$$

where: θx is the solid retention time of organism x in the column depth zone; N_{xcz} is the number of cells of organism x in the column depth zone; N_{xCsZ} is the number of cells of organism x in the control subsoil equivalent depth zone; nxINF is the number of cells of organism x entering with influent wastewater; nxEFF is the number of cells of organism x leaving the column through the porewater effluent. Relative read abundance values of $p_{xCZ,}$,p_{xCsZ}, P_{xEFF}. and p_{xINF} For the

different organisms were determined from the rarefied OTU table. OTUs with relative read abundance ≥ 0.1% were selected for calculating equation.

Here, for simplicity we assumed that the observed relative read abundance of each species-level OTU is equal to the abundance of that organism in the different samples. This assumption has some limitations due to: (i) variation in the 16S rRNA gene copy number per genome for the different species; (ii) differences in the specificity of the primers used in this study; and (iii) biases due to PCR and DNA extraction. However, these limitations are universal for amplicon sequencing. The correction for the different 16S rRNA gene copy number per genome is impossible due to a lack of reference genomes. However, this issue could be resolved as more reference genomes become available in the future.

5.3.5 Data availability

Raw sequencing data were deposited at the National Center for Biotechnology Information (NCBI) under accession number PRJNA994025.

5.4 Results

5.4.1 Wastewater influent pre-treatment impacts pollutant removal efficiencies

Hydraulic and organic loading onto the system, influent and effluent was continuously monitored for more than 200 days. It is important to note that there were large variations in concentration of parameters due to the length of storage of influent prior to resupply. Due to the distance to influent source, influent was refrigerated at $4^{O}C$ degrees Celsius, all removal efficiencies were calculated between influent and effluent within 24-hour periods. By ensuring that removal efficiencies were determined within these 24-hour cycles it permitted statistical analysis of removal efficiencies across the timeframe of the experiment. The samples were tested for biological and chemical characteristics and removal efficiencies of all four columns were calculated **(Table 5.1)**.

Table 5.1: Soil column Primary treated (PE)* and Secondary treated (SE)* influent and column effluent chemical characteristics

	Total Coliform (MPN/100 ml)	**E. coli** (MPN/100 ml)	**NH_4^+-N (mg/l)**	**NO_2-N (mg/l)**	**NO_3-N (mg/l)**	**TOC (mg/l)**	**PO_4-P (mg/l)**	**TC (mg/l)**	**TN (mg/l)**
PE Influent	$8.06 \times 10^6 \pm 9.87 \times 10^6$	$4.73 \times 10^5 \pm 6.89 \times 10^5$	116.12 ± 37.56	0.09 ± 0.05	0.82 ± 3.90	297.61 ± 99.77	11.45 ± 4.39	624.50 ± 239.31	116.14 ± 43.30
PE-High effluent	$1.51 \times 10^3 \pm 2.16 \times 10^3$	$2.93 \times 10^2 \pm 9.40 \times 10^2$	3.32 ± 8.31	0.23 ± 0.48	37.41 ± 19.85	8.45 ± 10.20	0.59 ± 0.93	83.38 ± 30.90	68.86 ± 32.99
PE-Low effluent	$6.83 \times 10^4 \pm 2.71 \times 10^5$	$8.86 \times 10^2 \pm 3.51 \times 10^3$	0.017 ± 0.04	0.01 ± 0.01	45.86 ± 47.03	10.38 ± 10.52	0.04 ± 0.06	67.45 ± 29.85	72.64 ± 42.46
SE Influent	$1.44 \times 10^4 \pm 1.44 \times 10^4$	$6.23 \times 10^1 \pm 8.7 \times 10^1$	0.734± 1.11	0.14 ± 0.16	10.59 ± 2.64	33.02 ± 24.18	11.41 ± 3.74786	166.81 ± 62.5	18.23 ± 8.03
SE-High effluent	$4.85 \times 10^3 \pm 1.42 \times 10^4$	$3.57 \times 10^1 \pm 1.34 \times 10^2$	0.02 ± 0.03	0.30 ± 0.65	14.39 ± 13.40	21.07 ± 14.52	0.43 ± 0.51	136.66 ± 30.33	23.02 ± 15.05
SE-Low effluent	$1.84 \times 10^2 \pm 3.34 \times 10^2$	$2.02 \times 10^1 \pm 5.24 \times 10^1$	0.06 ± 0.11	0.02 ± 0.02	13.27 ± 2.86	38.55 ± 22.77	0.02 ± 0.03	129.51 ± 44.84	26.87 ± 14.83

*PE referred to primary treated effluent generated by a two-chambered septic tank and SE referred to secondary treated effluent generated by the RBC system.

The mean log10 removal efficiencies **(Table SB2)** for average total coliforms are 1.53 ± 1.35, 4.13 ± 1.09, 2.30 ± 1.14, and 4.39 ± 1.26 for SE-High, PE-High, SE-Low, and PE-Low respectively. For Escherichia coli, all columns displayed high mean log10 removal of 1.23 ± 0.81, 3.81 ± 1.39, 1.39 ± 0.76, 4.39 ± 1.26, with columns dosed with PE influent exhibiting the highest removal efficiencies. Regarding TOC, the removal efficiencies exhibit a similar trend with columns dosed with PE influent exhibiting high degrees of removal at (-) 19.9 ± 37.5 %, (-) 97.14 ± 3.3 %, (+) 42.1 ± 70.5 %, and (-) 96.5 ± 3.4 % for SE-High, PE-High, SE-Low, and PE-Low respectively. TC exhibits mean removal efficiencies of (-) 14.7 ± 24.8 %, (-) 85.3 ± 6.2 %, (-) 19.5 ± 31.5 %, and (-) 88.3 ± 6.1 %. For the nutrient components, PO_4-P shows mean removal efficiencies of (-) 97.0 ± 4.6 %, (-) 31.9 ± 312.1 %, (-) 99.9 ± 1.2 %, and (-) 92.3 ± 36.1 %. for SE-High, PE-High, SE-Low, and PE-Low respectively.

For NH_4-N, the mean removal efficiencies are (-) 90.9 ± 18.4 %, (-) 97.4 ± 6.46 %, (-) 75 ± 48 %, and (-) 100.0 ± 0.39 %, for SE-High, PE-High, SE-Low, and PE-Low respectively. Similarly, NO_2-N displays mean removal efficiencies of (-) 74.9 ± 407.36 %, (+) 342 ± 11106 %, (-) 57.7 ± 61.1 %, (-) 57.7 ± 60.9 %, and (-) 90.6 ± 13.3 %, for SE-High, PE-High, SE-Low, and PE-Low respectively. Regarding NO_3-N, the mean removal efficiencies differ with (+) 32.7 ± 74.6 %, (+) 20001 ± 15821 %, (+) 33.21 ± 46.0 %, and (+) 19042 ± 26982 % for SE-High, PE-High, SE-Low, and PE-Low respectively. Lastly, TN exhibits mean removal efficiencies of (+) 7.1 ± 19.2 %, (-) 47.8 ± 20.55 %, (-) 3.5 ± 2.8 %, and (-) 41.0 ± 29.6 % for SE-High, PE-High, SE-Low, and PE-Low respectively.

The effect of pre-treatment of influent appeared to be more significant factor in the removal efficiency of contaminants (**Table 5.2**), the only pollutants not affected being PO_4-P, and nitrate in low porosity soils. Porosity had a far less significant effect on contaminant removal efficiency, with only TOC, PO_4-P, ammonium in soils dosed with PE and TOC, PO_4-P, total nitrogen soils dosed in SE.

Table 5.2. Wilcoxon t-test comparative analysis of removal efficiencies in respect to porosity and effluent pre-treatment level

	PE High vs Low		**SE High vs Low**		**High PE VS SE**		**Low PE VS SE**	
	***P*-value**	**W- value**	***P*-Value**	**W- value**	***P*-Value**	**W- value**	***P*-Value**	**W- value**
AvgTC	NS	5	NS	383	**	113.5	***	77.5
Avgecoli	NS	117	NS	5	**	77.5	**	784.5
TOC	***	1206.5	*******	1601.5	***	1715.5	***	228.5
TC	NS	769.5	NS	1011.5	***	1976.5	***	1715.5
PO_4-P	**	1206.5	***	1601.5	NS	228.5	NS	141.5
NH_4-N	*	569.5	NS	584.5	***	1979.5	**	1927. 5
NO_3-N	NS	584.5	NS	119.5	***	724.5	NS	1979. 5
NO_2-N	NS	119.5	NS	117	***	784.5	*	724.5
TN	NS	1011.5	*	569.5	***	1927.5	***	1976. 5

(NS: Pr > 0.05, *Pr ≤ 0.05, **Pr ≤ 0.01, ***Pr ≤ 0.001, and ****Pr ≤ 0.0001)

5.4.2 Biomat development influenced by the level of pre-treatment

Over the course of the monitoring period soil moisture sensors were collecting volumetric water content (VWC) data across the depth profile of the systems. The data from the 0 cm to 10 cm depth were analysed to determine shifts in water retention at the infiltrative and proximal zone as these are depths in which main biomat development is likely to occur. The establishment of the biomat was defined by the daily mean increase of VWC of 0.025 cm^3 cm^{-3} over the initial baseline over a period of 30 days (Knappe et al., 2020). Using these qualifiers, it was possible to determine that clogging occurred first in the soils dosed with PE high porosity soil (50 days) at 0 cm depth, followed by low porosity soils dosed with PE which clogged 22 days later at 0 cm depth (72 days) and at 10 cm depth (211 days), as shown on **Figure 5.2**. The soils dosed with more treated SE took much longer to develop a biomat. The high porosity soils dosed with SE clogged 85 days prior (481 days at 0 cm depth) to the lower porosity also receiving SE (563 days at 0 cm depth; 570 days at 10 cm depth). The depth of the clogging zone occurrence appeared to extend over 10 cm depth in the low porosity soil, while in the high porosity soils had more narrow locations of clogging at 0 cm and 10 cm at the PE-High and SE-High systems respectively. It appeared the PE-Low porosity system underwent compaction due to the constant wetting and drying of the experiment, which resulted in the IZ VWC sensors being exposed; hence, values following 2022-03-17 (261 days) were ignored and expected VWC values were modelled using the prophet model **(Figure 5.2)**.

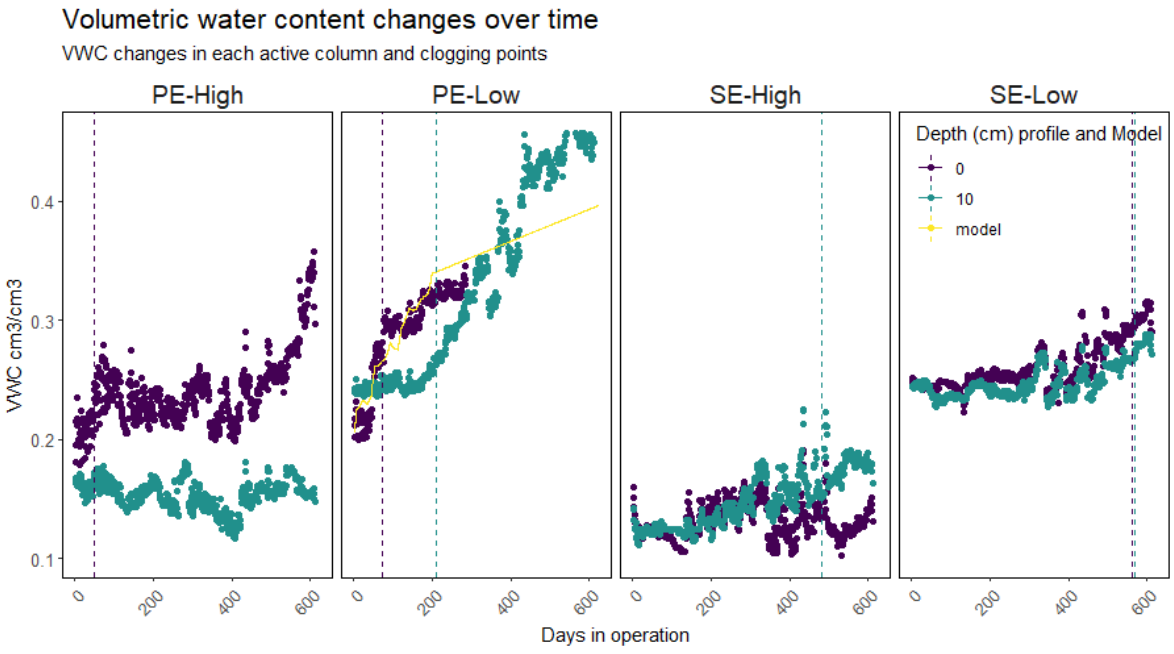


Figure 5.2. Increases in the volumetric water content over time for each system. Dashed lines represent the clogging points for each of the systems at different depths in their respective systems, assessed by increases of 5% VWC over the period of 30 consecutive days. The yellow represents the model estimating the PE-Low column changes.

5.4.3 Spatial Variation in Microbial Community Composition Influenced by pre-treatment

The 53 samples sequenced resulted in a total of 34,509 OTUs, 15,788,859 reads with a read depth per sample ranging from 132,050 (min) to 308,793 (max). Read depth was rarefied to the minimum number of reads (132,050) per sample. The control subsoil sample was dominated by phyla *Proteobacteria, Bacteroita, Actinobacteroita and Acidobacteria* with abundances of 29.3 %, 10.5 %, 7.5 % and 5.12 % respectively. PE influent was dominated by phyla *Proteobacteria, Campylobacteriota* and *Bacteroita* with abundances of 54.5 %, 22.6 %, and 14.4 % respectively. SE influent was dominated by phyla *Proteobacteria, Campylobacteriota* and *Bacteroita* with abundances of 34 %, 25.5 %, and 20.5 % respectively. The high porosity system dosed with PE influent resulted in dominant phyla *Proteobacteria, Actinobacteriota, Bacteriota* and *Acidobacteriota* with abundances of 37.9 ± 9.8 %, 8.7 ± 2.2 %, 5.2 ± 1.7 % and 5 ± 2.3 % respectively. The low porosity system dosed with PE influent was dominated by *Proteobacteria, Actinobacteriota, Firmicutes* and *Patescibacteria*, 29.5 ± 6.4 %, 12.1 ± 3.8 %, 6.9 ± 2.5 %and 5.1 ± 12.6 %respectively. The high porosity column soil dosed with SE influent resulted in dominant phyla *Proteobacteria, Actinobacteriota, Firmicutes* and *Chloroflexi* with abundances of 28.9 ± 1.58 %, 13.18 ± 4.5 %, 8.7 ± 2.8 % and 6.2 ± 1.9 % respectively. The low porosity soil dosed with SE influent was dominated by *Proteobacteria, Actinobacteriota, Chloroflexi, Acidobacteria* with abundances 29.1 ± 6.8, 9.76 ± 2.3 %, 6.9 ± 2.5 %, and 6.54 ± 2.5 % respectively. *Arcobacter sp* was the most dominant species for PE and SE influents, while two *Firmicutes* species dominated the control subsoil sample. In the columns the most dominant known species were *Arcobacter sp.* (abundance of 6.7 ± 3.1 %) and *Zoogloea sp* (abundance of 4.5 ± 0.8 %) at the infiltrative zone in the PE-High column. In the SE-Low column infiltrative zone *Gemmatimonadota OTU 6757* (abundance of 3.3 %) and *Terrimonas sp* (abundance of 3.6 %) at SE–Low infiltrative zone was most dominant (**Figure 5.3**). *Zooglea* were most dominant in the PE-Low column deep zone with an abundance of (3.1 ± 0.6 %) and *Terimonas sp* (2.9 ± 0.5 %) were most abundant in SE-High at the Intermediary zone.

Columns were analysed for alpha diversity and richness **(Figure 5.4)**. The SE-Low and PE-High soil columns alpha-diversity values were lower than that of the control subsoil sample, with only the PE-Low column having a higher alpha diversity value than that of the control. Spatial variation in alpha diversity values was observed within the depth profile of all columns, with the PE-High and SE-Low columns showing reduced diversity at the infiltrative zone and proximal zone when compared to the control sample **(Figure 5.4A)**. However, PE-High column alpha diversity exceeded that of the control sample at depths of 7 cm and deeper. The PE-Low column showed low alpha diversity at the IZ followed by an increase relative to the control samples at depths ranging from 1 – 20 cm. The SE-High column, throughout its depth profile, had a reduced lower diversity compared to the control subsoil. None of the columns exceeded the control sample values in richness (Chao1)

or OTUs. The spatial distribution of communities showed that the areas with the least species richness were at the Infiltrative Zone and the Proximal Zone apart from the PE-Low porosity column where the area of least richness was solely in the infiltrative zone **(Figure 5.4B)**. All samples exceeded effluent richness values, but the only location to exceed the control subsoil value was the PE-High porosity soil at the 20 cm mark. Low porosity soils treated with PE and SE influent showed a significant difference in the Shannon diversity and Chao1 richness ($p = 0.02^*$, $W = 111$; $p = 0.0002^{***}$, $W = 131$, respectively). High porosity soils showed no statistical difference between soils dosed with PE and SE effluent (**Figure SB3)**. Equally, the difference between the low and high porosity soils revealed no statistical difference in the alpha diversity or richness of the soil column communities.

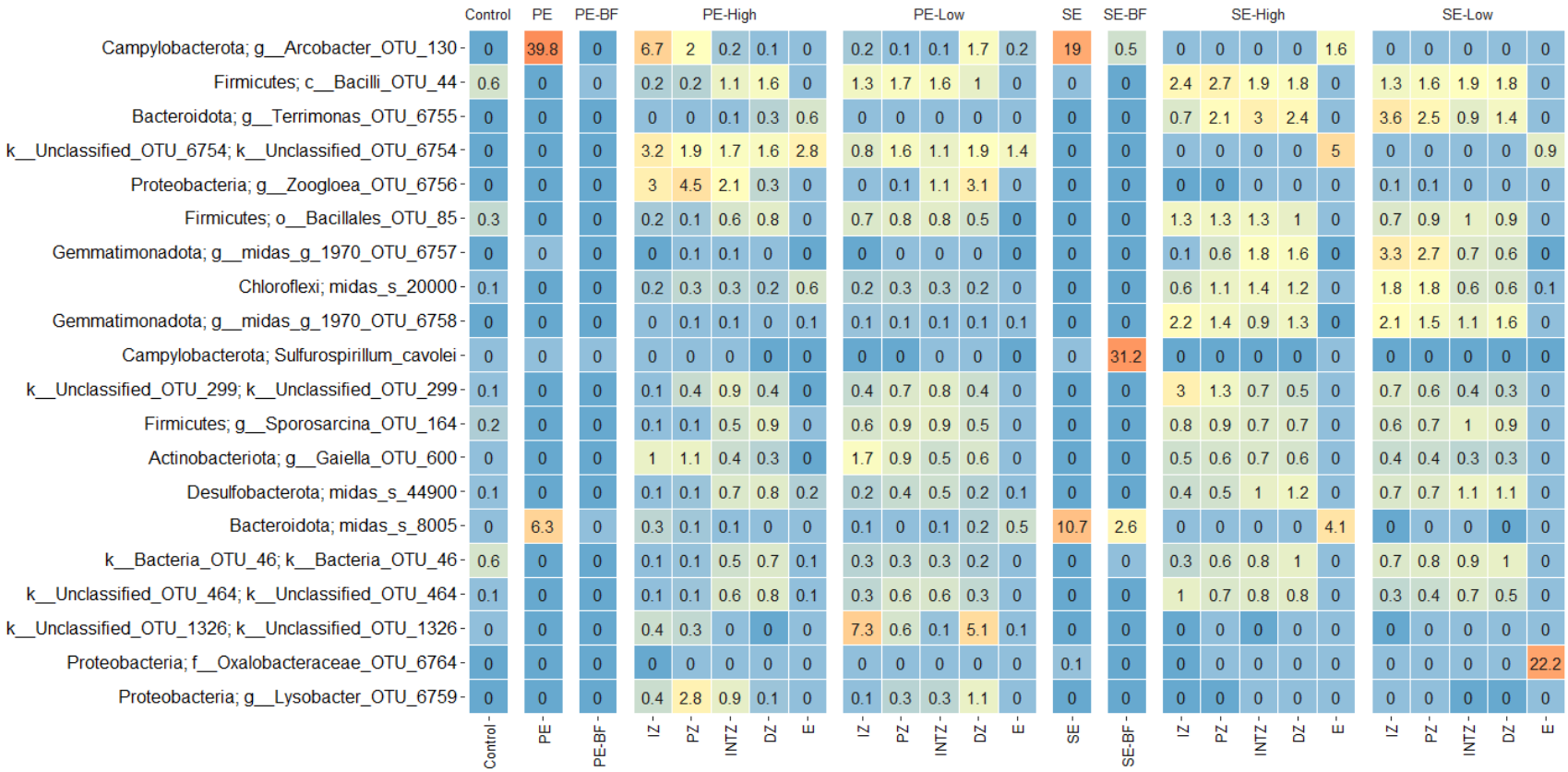


Figure 5.3. Relative read abundance (%) of top 20 species-level OTUs for soil columns of high and low porosity dosed with primary and secondary effluent (PE-High; SE-High; PE-Low; SE- High). Abundances are visualised across all columns depth profiles were 0 cm – 1 cm is the Infiltrative Zone (IZ), 1 cm – 5 cm Proximal Zone, 6 cm-10 cm Intermediary Zone and 11 cm – 30 cm in the Deep Zone of the subsoil columns, influent (PE, SE), tubing biofilms (PE-BF, SE-BF) and column effluent (E)

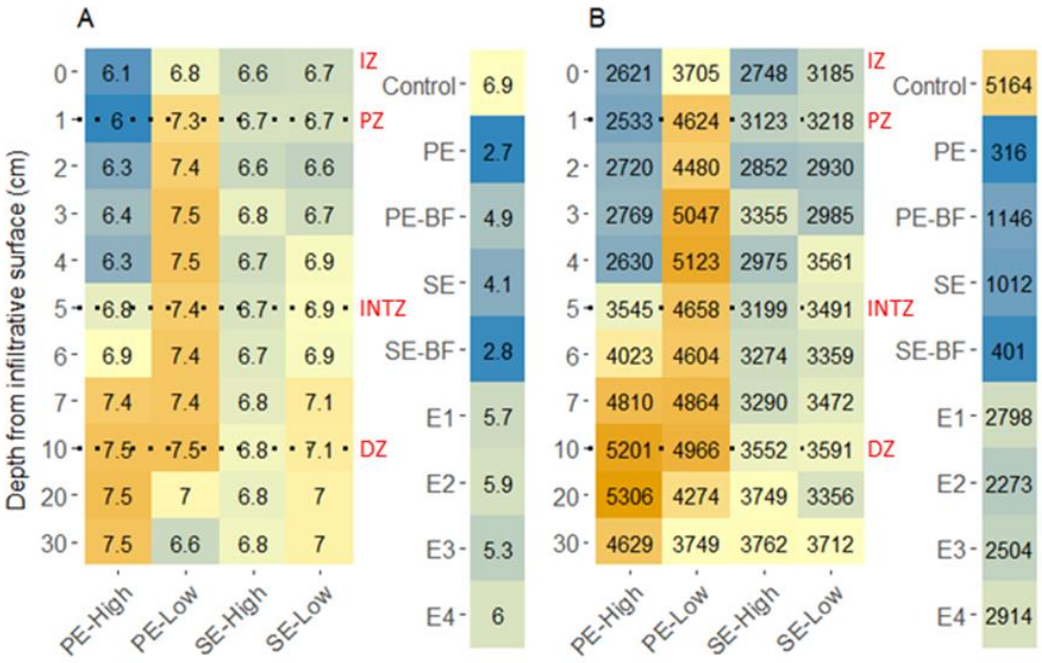


Figure 5.4. The Shannon diversity (A) and Chao1 richness (B) are visualised for soil columns of high and low porosity dosed with primary and secondary effluent (PE-High; SE-High; PE-Low; SE- High). Abundances are visualised across all columns depth profiles where 0 cm – 1 cm is the Infiltrative Zone (IZ), 1 – 5 cm Proximal Zone, 6 cm -10 cm Intermediary Zone and 11 cm – 30 cm in the Deep Zone of the subsoil columns, influent (PE, SE), tubing biofilms (PE-BF, SE-BF) and column effluent (E1-E4).

PCoA clustering was applied to determine beta diversity using weighted Unifrac accounting for the relative read abundance within the samples and Bray Curtis for dissimilarity between samples **(Figure 5.5)**. Two main clusters are present: the first consists of subsoil columns dosed with PE influent, which is spread loosely and contains a tight subcluster consisting primarily of low porosity subsoil samples and the intermittent zone, deep zone subsoil samples of the PE-High subsoil column; the second cluster which is more tightly confined and consists entirely of subsoil column samples dosed with SE influent. None of the clusters or sub-clusters overlapped with the control leachate, dosing tube biofilms and effluent samples, the former being the most distant from subsoil column communities. Permutational analysis of variance applying distance matrices results showed porosity appears to have had a significant impact on community structure in both subsoil columns treated with PE and SE influent (R2 = 0.23, Pr = 0.002**; R2 = 0.21, Pr = 0.0009). Effluent pre-treatment impact on community structure was more significant and accounted for greater variance within the community than porosity (R2= 0.43, Pr = 0.0005 ***; R2 = 0.5, Pr = 0.0005 ***, respectively

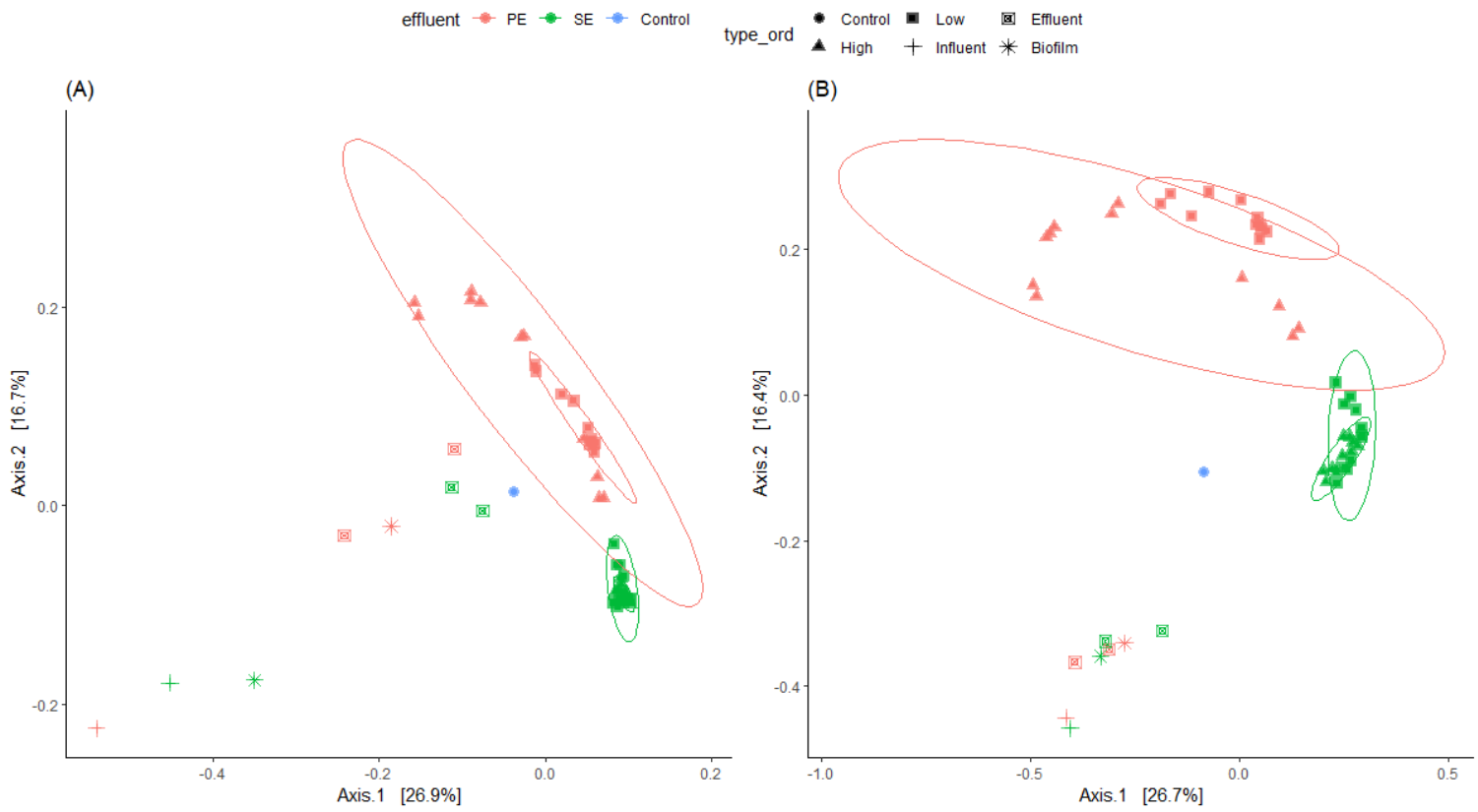


Figure 5.5. Principal coordinates of beta diversity based on weighted Unifrac (A) and Bray-Curtis (B) distances across all subsoil columns and control. Samples were aggregated based on porosity (High, Low), influent (PE, SE), and tubing biofilms (PE-BF, SE-BF), confidence interval was set to 0.75. Ellipses are set to a multivariate t-distribution, with confidence level set to 95%.

5.4.4 Influent pre-treatment results in increased levels of immigration and growth of bacteria from influent into the soil column

The influence of immigration from the effluent or source habitat into the subsoil ecosystems was investigated by calculating the net growth rate of 7962 dominant species (relative read abundance > 0.01) level OTUs within the different layers of the subsoil column by pairing amplicon sequence abundance data with mass balancing. PE and SE influent had 199 and 650 OTUs, respectively. Of these present only 106 (30 growing and 76 not growing) were also found within PE-dosed STUs and 121 were found within the SE-dosed STUs (95 growing and 197 not growing). The greater number of OTUs present with the SE STU may be because biomat dispersal from the RBC disks into the column influent. The relative abundances of the dominant growing species were impacted by the level of pre-treatment of the effluent **(Table 5.3)**. Dominant OTUs consisted of the phyla *Firmicutes*, *Actinobacteriota*, *Proteobacteria*, *Chloroflexi*, *Planctomycetota*, *Bdellovibrionota*, *Myxococcota* and *Gemmatimonadota*, the greatest abundances were only found within the STUs and those that were growing. SE columns harboured a relatively high abundance of growing OTUs sourced from influent **(Figure 5.6)**.

The greatest level of growth was found at the infiltrative zone of all columns: with *Campylobacteria Arcobacter sp* and (*Gamma*) *Proteobacteria midas species 18794* in the PE-Low porosity STU, (*Beta*) *Proteobacteria order Oxalobacteraceae* in SE-Low STU; (*Beta*) *Proteobacteria order Burkholderiales* in the PE-High porosity STU; and *Campylobacteria Arcobacter sp* in the SE-High porosity STU. Except for *Burkholderiales* in the PE-High column the remaining prominent growing species were present within the influent. The species whose growth was most negatively effective were: *Campylobacterota midas_s_2255*, (*Gamma*) *Proteobacteria; Pseudomonas caeni*, (*Beta*) *Proteobacteria; Comamonas sp , Bacteroidota midas_s_8005* and *Bacteroides graminisolvens* in the PE-Low column; *Proteobacteria; midas_s_9838* in the SE-High column; (*Gamma*) *Proteobacteria midas_s_18794* in the PE-High porosity column; and (*Gamma*) *Proteobacteria; midas_s_8839* in the SE low column which was the only species that was found only in the STU. The most variable location for species growth was the infiltrative zone of all the columns, PE – dosed columns had immigrants account for 3.4 % or relative read abundance, and 3.8% in SE – dosed columns (**Figure SB5; Table SB3**).

Table 5.3. Growing and not growing species relative read abundances (%), and (total OTUs) across both influent streams (PE STUs: n=22, SE STUs: n=22).

	PE		**SE**	
	Growing	Not Growing	Growing	Not Growing
Influent	0.2 (30)	3.4 (76)	3.77 (95)	3.8 (197)
STU	88.1 (2134)	1.7 (650)	86.7 (1872)	1.28 (656)

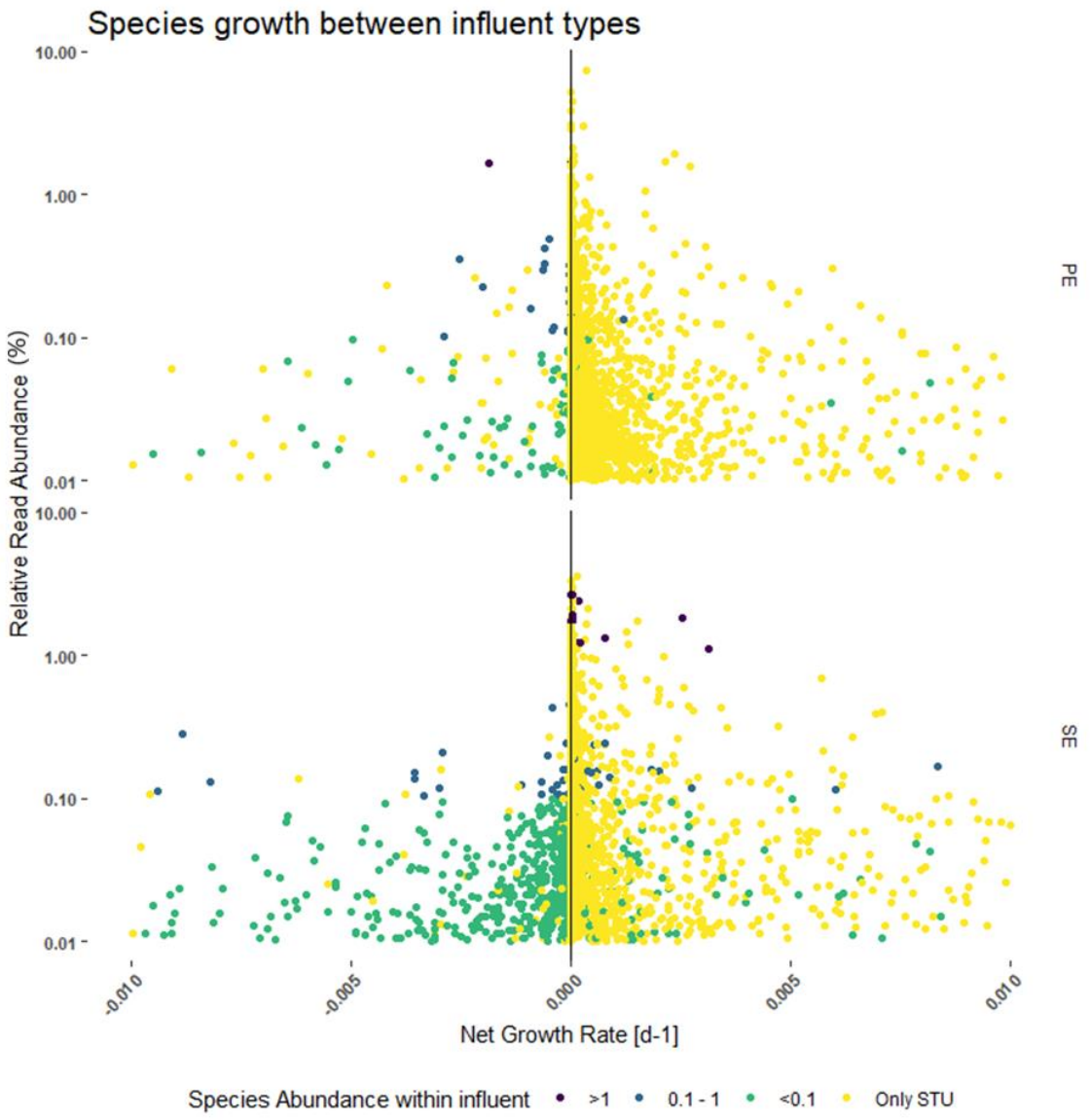


Figure 5.6. Net Growth rate for dominant and rare species aggregated across columns for both influent types

The study also investigated the role of functional groups connected through the MiDAS reference database with respect to nutrient removal across different zones of the four columns. Among these groups, ammonia oxidizing bacteria (AOB) were represented by Nitrosospira sp. and Nitrosomonas sp. OTUs in all four systems, including wastewater and control soils (Figure 7). Nitrosomonas shared dominance with Nitrosospira sp. in soil columns dosed with PE, compared to the soil columns dosed with SE, the control subsoil and the SE influent where Nitrosospira sp. was dominant. In the PE-dosed columns, the highest abundances of AOB were found at the proximal and deep zone for the high and low porosity columns, respectively. In SE-dosed columns, the greatest abundances of AOB were observed at the proximal zone and the proximal zone. The relative abundance of AOB in the control subsoil and columns treated with PE and SE were 0.23 %, 0.48 ± 0.29 %, and 0.27 ± 0.1 %, respectively. Nitrite oxidizing bacteria (NOB) were represented by Nitrospira sp. OTUs in all four systems, including wastewater and in control soils Nitrotoga sp. accounted as a rare out, Nitrobacter as an NOB was not present within any of the systems. The abundances of NOB were

highest at the intermediary and deep zone in the PE-high and low porosity columns, respectively. SE-dosed columns exhibited the highest abundances of NOB at the infiltrative zone, and the abundances across all columns increased significantly compared to the control. The relative abundance of NOB in control subsoil and columns dosed with PE and SE were 0.92%, 2 ± 0.8%, and 2.1 ± 0.38%, respectively.

The highest abundances of denitrifiers were found at the proximal zone and deep zone for the PE-high and low porosity soils, respectively. In the SE-dosed soils, the highest abundances of denitrifiers were observed at the DZ and IZ for the high and low porosity soils, respectively. The most common dominant species found were Nitrospira sp. and Ca_Accumulibacter sp, with relative read abundance of denitrifiers in the control subsoil and columns treated with PE and SE of 1.66 ± 3.06%, 0.85 ± 1.29%, and 0.66 ± 0.79%, respectively. The greatest abundances of polyphosphate-accumulating organisms (PAOs) were observed at the proximal zone and infiltrative zone for both PE-dosed high and low porosity columns. In the SE-dosed columns, the highest abundances of PAOs were found at the deep zone in the high porosity soil and at the intermediary zone and deep zone in the low porosity soil. Five unique PAO species were found across all samples (Tetrasphaera sp., Ca_Accumulibacter sp., Microlunatus sp.Ca_Obscuribacter sp.Corynebacterium sp.) The relative abundance of PAO in control subsoil and columns dosed with PE and SE were 0.03 ± 0.03%, 017 ± 0.06%, and 0.01 ± 0.01%, respectively. The greatest abundances of methanogens were found at the proximal zone and intermediary zone for the PE-High and PE-Low porosity columns respectively. In the SE dosed columns, the locations of greatest abundances were intermediary zone and proximal zone for high and low porosity columns respectively. The only dominant species of methanogens was Methanomassiliicoccus sp in the control subsoil. Of the rare species of Methanogens, Methanobrevibacter sp., Methanospirillum sp. Methanosarcina sp were most frequently observed. The mean relative abundances for the control, PE columns, and SE columns were 0.98 ± 0%, 0.02 ± 0.01%, and 0.005 ± 0.01%, respectively. In the PE-High column, methanotrophs were dominant at the intermediary zone and deep zone locations. The PE-Low column had an average total abundance of 0.01 across all depth zones. In both SE-High and Low porosity columns, the greatest total abundance of methanotrophs was observed at the proximal and intermediary zones. Methanotrophs consisted of four species; Methylocaldum sp.Methylobacter sp. Methylocella sp. and Methylocystis sp. The mean relative abundances for the control, PE columns and SE columns were 0.01 ± 0 %, 0.14 ± 0%, and 0.01 ± 0.01%, respectively.

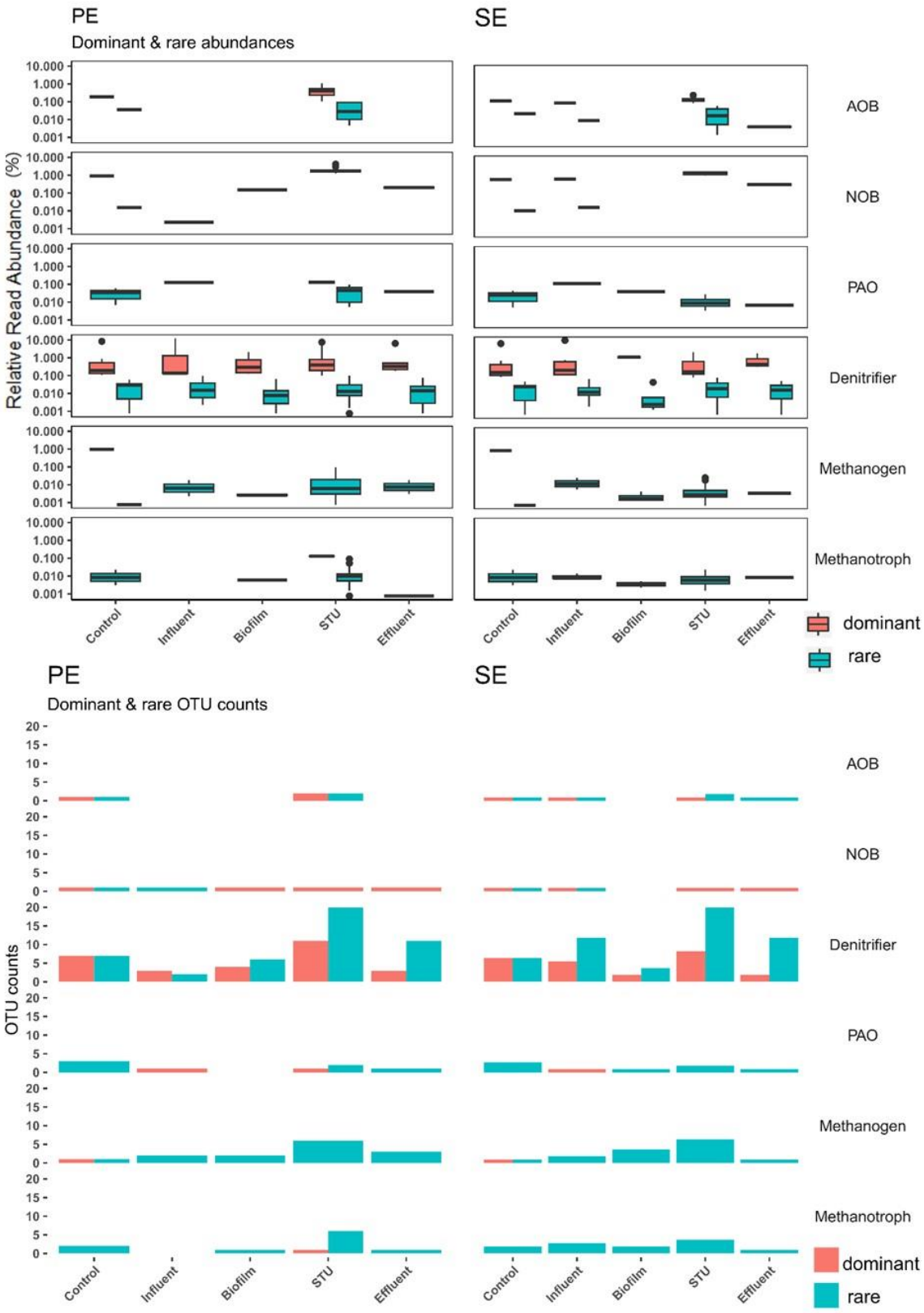


Figure 5.7. Dominant and rare functional organism abundances, and (OTU counts) averaged across both influent streams, within Control, influent, inlet tube biofilm, soil treatment unit (STU) and effluent.

5.5 Discussion

This study was illustrated the significance of pre-treatment of influent, porosity, in in the structuring of the community assemblages of soil treatment units. The findings of this study mirror the ecological surveys of field scale observations of STUs, where effluent pre-treatment significantly altered community composition. The use of VWC sensors confirmed the accelerated bioclogging action (indicative of biomat development) from the application of high organic primary treated influent which was achieved over a year in advance in comparison to the SE dosed columns. These findings match field observations of STU trenches which observed biomats fed with PE reaching 15 m (along a total 18 m of trench) after 10 to 13 months, compared to those fed with SE where the biomat only reached 7.5 to 10 m after 3 years of operation (Knappe et al.,2020). Somewhat counter-intuitively, the column with high porosity dosed with primary treated influent exhibited the formation of a biomat 22 days earlier than the low porosity soil columns. Literature states that the biomat is assumed to have a thickness of only 1-2 cm, however, its development was found at greater depths of 10 cm for both the PE and SE-dosed low porosity columns. This greater biomat thickness could be attributed to the soil pore network due to greater pore connectiveness and reduced pore diameter enhancing the distribution of the biomat. (Beal et al 2006; Mckinley et Siegrist., 2010, Mckinley et Siegrist., 2011). The level of *E. coli* removal efficiencies was significantly less for columns receiving secondary effluent, regardless of the lower loads of *E. coli* being loaded onto the columns. This demonstrates the physical effect of filtration and removal provided by the established biomats in the PE columns. PE effluent fed systems also demonstrated better removal rates of total nitrogen which has been seen at the field level (Gill et al., 2009). However, in soils with high clay content pre-treatment would be required due to the low soil permeability and greater risk of ponding. The advanced establishment of the biomat within PE STUs enhances the role of the consortium of synergistic microbial activity, and results in greater removal of total nitrogen, but also the utilisation of organic material in metabolic processes.

Pre-treatment had a significant effect on the alpha species richness and diversity only for the low porosity soil columns whereas porosity or soil texture had no direct significant effect on the species richness or alpha diversity. Bioclogging, organic and nutrient transformations within the PE columns is shown by the possible positions of their biomat and the presence of microbial activity hotspots. There appears to be a pattern of ecological tipping points in the development of microbial biomats dosed with primary effluent, the first is the increase of diversity due to high microbial activity due to the abundance of substrate, second is that clogging continues and the infiltrative surface enters a steady state where community diversity drops due to the selection of copiotrophs (usually *Gamma/BetaProteobacteria)* driving a less metabolically diverse community (Sul et al., 2013; Kuzyakov et al., 2015; Knappe et al., 2020; Criado Monleon et al., 2022; Cui et al., 2023). In field studies of on-site systems, the lateral spatial variation of niches in the biomat development displayed high diversity initially in locations next to the effluent ingress of the trench due to

clogging. However, in older systems as they develop the diversity drops and a large increase in abundance of copiotroph appear to dominate the STU. (Reinhold-Hurek et al., 2015; Geyer and Barrett, 2019; Criado Monleon et al., 2022). It was observed that in the PE dosed columns areas of low diversity at the infiltrative zone corresponded to the locations with the highest abundance of AOB groups, which shows the selective pressure for these specialists in ammonium rich effluent. The significantly higher concentration of ammonium in the effluent in the PE-High column suggests that greater permeability of the soil resulted in incomplete nitrification.

Microbial hotspots can be identified within the PE dosed columns by areas of high species diversity located at the intermediary zone (5 – 10 cm) of the column which also contain the highest relative abundance of NOB, and deep zone (10 – 30 cm) for denitrifiers. The positions of these hotspots and their functional groups within the column suggest where there is more diverse biogeochemical processes of nitrogen species and organic matter as it broken down into new metabolites as it leaches through the soil. Both the SE columns showed similar diversity and richness increase with distance away from the infiltrative surfaces. All nitrogen cycling organisms were most abundant between a depth of 0-10 cm, suggesting the effects pre-treatment which provides an effluent low in diverse organics and nutrients substrates that is loaded at a steady state was then selective for specialist bacteria but only at shallower depths when compared to soils dosed with PE (Sul et al., 2013).

Our findings using sandy loam of varying degrees of porosity have shown that the level of pre-treatment or influent quality is of greater significance than porosity or soil texture to community structure. Our findings are contrary to several other studies that have stated that soil texture of differing soil types is more significant than influent characteristics (Krause et al., 2020; Oboyumi et al., 2021). Note, however, that those studies used treated wastewater and groundwater, potable or freshwater as a control, with relatively low organic concentrations of 12.7 mg/ l (TOC) and 55.50 ± 4.65 mg/l (COD). Due to the absence of a control column dosed with distilled water or a duplicate column, it is not possible conclude on the dominance of pre-treatment as a factor solely from the lab study on community assembly. However, there is a clear link in the lab findings with the survey attained from STUs from sites where the effluent was sourced for this lab experiment, as these STUs also showed a significant divergence in communities based on pre-treatment (Criado Monleon et al., 2022). Paddy fields have also demonstrated similar findings, where the distribution of dissolved organic matter (DOM) correlates with the microbial community structure (Li et al., 2020). The growth levels of species within the columns where those indigenous to the soil, and not dispersed through the influent. These growth patterns indicates that the nutrient and organic loading result in species sorting being the more important driver of the mesocosm community assembly. Species sorting has been suggested in previous studies as a contributor to sediment bacterial community composition in rivers impacted by urban discharge, with different contaminants having selective forces on community composition (Roberto et al., 2018). Immigration did occur, but accounted for a small fraction of the community, immigrants were more abundant in SE dosed mesocosms with a

thinner biomat with less filtering capacity (Gill et al., 2009; Knappe et al., 2020). Organisms dispersed within the soil matrix are exposed to highly competitive and diverse environment, where mesophiles are often outcompeted for organisms, more adapt to the colder temperatures. The soil environment in these systems are found in two different conditions, one loaded with a very high concentration of organics resulting in large and quick paced changes to ecosystems which initially selects for a diverse community until such time clogging ensures only the most competitive copiotrophs dominate. The second condition is also selective, but for specialists ready to for a steady flow of low concentrations of organics, but spatially limited to their niche at or near the infiltrative surface of the soil.

The findings from this research have shown that the levels of pre-treatment of wastewater influent then added to a soil matrix had a more significant role in the development of the soil ecology than the initial soil texture. Porosity did result in less attenuation of nutrients in columns. However, the less developed biomat due to pre-treatment resulted in less effective removal of organics and nutrients. It was determined that species sorting due to the alteration of the environment by the level of organics or nutrients within the influent selected for indigenous species within the sandy loam STU communities rather than the successful colonisation of wastewater borne species. This is likely to be the case in most STUs with appropriate levels of permeability.It Spatial profiling shows a linear demarcation in niches identified by diversity, functional group abundances which correspond to positions of microbial activity, nitrogen transformation, but only in high organic environments resulting in enhanced contaminant removal efficiency. These findings will benefit research and development projects in rapid soil treatment systems and managed aquifer recharge. In the application of managed aquifer recharge in municipal sites were pretreatment due to flow will be mandated, the incorporation of organics such as biochar into the gravel matrix of the percolation area could accelerate the development of the biomat. This organic augmentation could enhance performance in systems dispersing highly treated effluent.

5.6 Conclusion

- Clogging and biomat development occurred faster in PE dosed soil. Contaminant removal efficiency was greater in columns dosed with primary influent.
- Species diversity and functionality highlighted microbial hotspots corresponding to the vertical flow of nitrogen compounds in columns dosed in PE influent. Columns dosed in secondary treated effluent showed no microbial hotspots and nitrogen speciality bacteria were concentrated in shallow positions of the soil columns.
- Pre-treatment of influent was more significant than porosity for the community structure of soil columns' biomes.
- Species sorting from the nutrient and organic composition of the influent was the main driver in community assembly and not immigration of species present in the influent.

6. Pepper Mild Mottle Virus as an Effective Tool in Microbial Source Tracking for Failing Domestic On-Site Water Treatment Systems

Alejandro Javier Criado Monleon, Laurence Gill

This manuscript is under review in the journal of Science of the Total Environment.

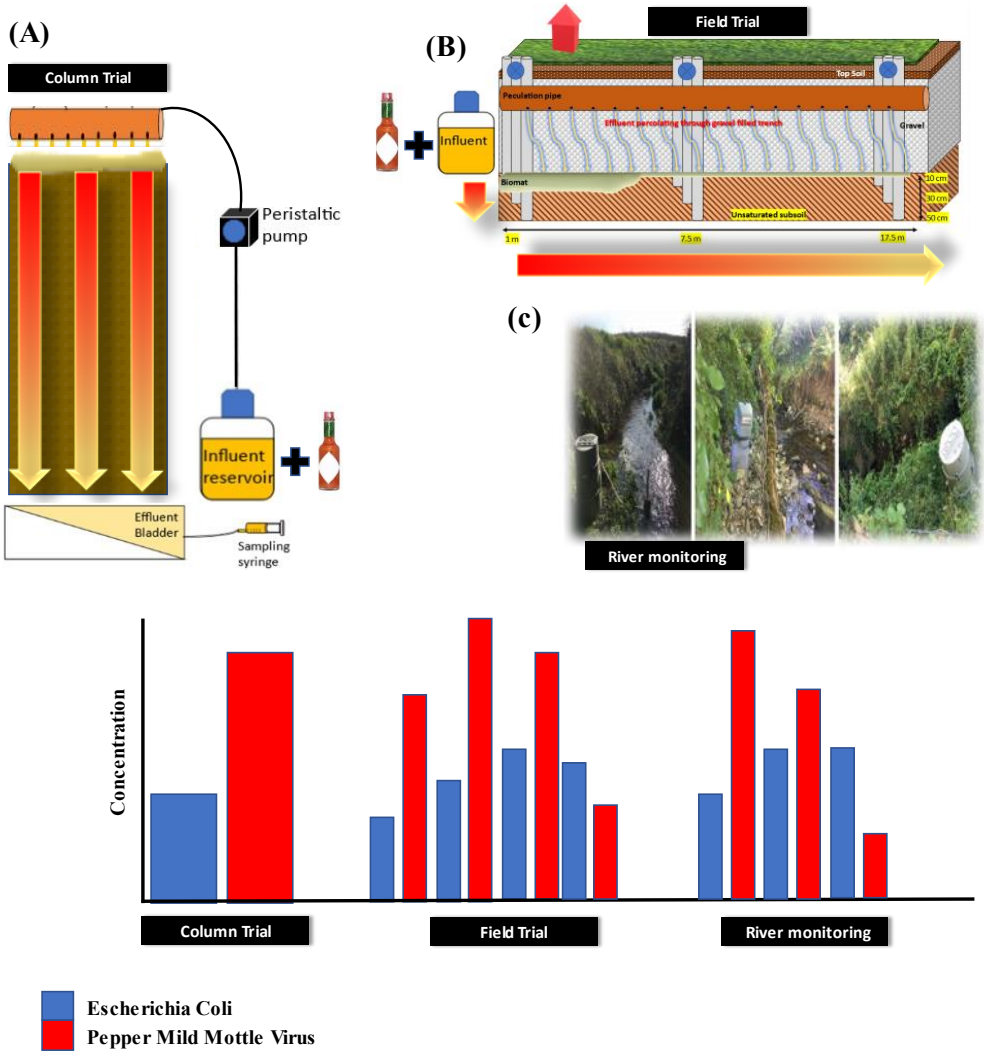


Figure 6.1 Experiment design for (A) soil column pepper mild mottle spiking trial (B) soil treatment unit spiking trial (C) catchment monitoring of rivers known to be impacted by failing on-site wastewater treatment systems, please note concentrations are not based on values from the study and are only illustrative.

6.1 Abstract

Pollution from domestic on-site wastewater treatment systems (OWTS) is a significant contaminant pressure in many rural catchments. However, due to their design, and dispersed proliferation, it is difficult to assess their impact. Water testing methodologies employ bacterial culturing methods and chemical analysis which may lose resolution and/or specificity being confounded by diffuse agricultural sources within a rural environment. In this study, we successfully assess the applicability of Pepper Mild Mottle Virus (PMMoV) as a faecal source tracker for failing on site wastewater treatment systems. The transport of PMMoV was first studied in the effluent of a 30 cm deep soil column which was dosed for 510 days with primary influent from a conventional septic system. The removal of PMMoV through the 30 cm deep soil column was quantified with a 5-day spiking trial employing primary influent mixed with Tabasco® pepper product. The trial was then carried out at field scale with the spiking solution dosed into an operational percolation trench receiving septic tank effluent which had been instrumented for porewater sampling. Samples were taken at depths of 10 cm, 30 cm and 50 cm across the length of the trench at distances of 1 m, 7.5 m and 17.5 m from the inlet of the trench. At the field-level PMMoV was detected on all days of the trial, with a peak concentration of 1×10^6 found at the rear of the trench on day 2 of the spiking trial. Three rural catchments with high densities of OWTSs were also sampled and analysed for hourly variations in biological parameters which include *E. coli*, total conforms, PMMoV, and chemical parameters total organic carbon, total nitrogen, and total carbon. PMMoV was detected in all river samples over a 24-hour period, thereby indicating its suitability as a tracer of human wastewater effluent in such environments with multiple diffuse sources.

Keywords:
Microbial source tracking, pollution, Faecal contamination fingerprinting, Water, On-site wastewater treatment

6.2 Introduction

The UN Sustainable Development Goal 6 requires the effective management and availability of water resources, to ensure the stewardship of water, and prevention of anthropogenic contamination. Global failings in achieving this goal are complex, especially the challenge of faecal contamination and its repercussions on human health and the environment. In lower to middle income countries over 2 billion people are impacted by contaminated drinking water supplies (Bain et al., 2020). At present 41% percent of the global population employ on site sanitation, often households or communities connected to a septic system with varying degrees of functionality (WHO., 2020). In Europe the effects of climate change are severely impacting water infrastructure, with high rainfall events overloading sewage wastewater treatment and sewerage systems resulting in the faecal contamination of surface waters (Nnane et al., 2011). The Republic of Ireland, for example, has a high frequency of persistent, rainfall events (Andrade et al., 2018). The country also abstracts 80% of its drinking water from surface waters bodies, almost half of which are achieving less than 'Good' status according to the EU Water Framework Directive and there has been an ongoing declining trend in quality (Rolston et al., 2020; EPA IE.,2022). A key stressor to Irish water bodies is excess nutrients (nitrogen and phosphorous), primarily from agriculture and incomplete wastewater treatment (EPA IE; 2022). It is difficult to separate the relative pollution impact in river catchments between agricultural non-point sources and human wastewater effluent discharge sources, particularly with such a large degree of one-off rural housing in Ireland with on-site wastewater treatment systems accounting for 30% of the total national domestic wastewater treatment (CSO, 2023a). It has been estimated that about 9% of water bodies in Ireland are experiencing substantial stress due to failing On-site Wastewater Treatment Systems (OWTS). However, the dominant and most impactful pressure on water quality continues to be agriculture (Mockler et al., 2017). This sector's influence extends to more than 1000 water bodies (EPA IE 2022; 2023).

Faecal Source Tracking (FST) uses microbial and chemical methods to ascertain the source of faecal contamination within a catchment area. A variety of chemical markers have been identified in the detection of faecal contamination such as caffeine, fluorescent whitening compounds used in detergents, fragrances, pesticides, faecal sterols and polycyclic aromatic compounds (Field and Samadpour., 2008). Ionic ratios using K/Na and Cl/Br ratios which can be unique to a contamination source have been used to screen aquifers for septic tank leachate (Brown et al., 2009). Pharmaceutical and antibiotic resistance to compounds used in the treatment of enteric infections (amoxicillin, ampicillin, cephalexin, cefuroxime and ciprofloxacin) prevalence of resistance in total coliforms and *Escherichia coli (E. coli)* can be indicative of a source of faecal contamination. However, the selection of antibiotics must be specific to either veterinary or medical application to determine the host (Gheethi et al., 2013; Grenni et al., 2018). Faecal sterol ratios have been successful for identifying sources at the catchment scale (Derrien et al., 2012; Jauković et al., 2022).

However, monitoring potentially contaminated groundwater using sterol analysis for host identification may be ineffective due to sterols readily adsorbing to sediments (Fennel et al., 2021).

At present *E. coli* and faecal enterococci are commonly used as faecal indicator bacteria applied in FST profiling of catchments. However, these techniques can be susceptible to confounding as both types of bacteria are endemic to a variety of animals, and they therefore do not identify the source (Field and Sandpour, 2007). Advancements in molecular techniques have been well developed in the monitoring of SARs-CoV2 within wastewater samples (Bivins et al., 2020). These techniques have provided novel and more specific targets which can better discern sources of contamination. For example, there has been significant correlation of the human-specific biomarker HF-183 and ammonium concentrations in known impacted sites in small urban streams in Ireland (Reynolds et al., 2023). Biomarkers have also sourced faecal contamination to dog fouling within Dublin's coastline (Martin et al., 2021). These markers demonstrate an enormous versatility in the selection of targets, although they are limited by the inability to discern viable from dead cells; however, the selection of RNA biomarkers may increase sensitivity due to the differing persistence times of RNA (13 hr) and DNA (94 hr) in a maritime environment (Wood et al., 2020).

Pepper Mild Mottle Virus (PMMoV) is an RNA virus plant pathogen that has been identified in pepper food products and subsequently quantified in human faeces at quantities up to 9.8×10^6 GC/mg (Zhang et al 2006; Hamze et al., 2011). PMMoV has already successfully been applied as a faecal source tracker and has been found in concentrations of 1.50×10^8 to $1 \times 10^{9.7}$ GC/100 ml in raw sewage and 1.10×10^6 to 7.00×10^7 GC/100 ml in treated wastewater (Rosario et al., 2009; Hamze et al., 2011; Gyawali et al., 2019). PMMoV has a faster inactivation rate (−0.09/d) than human faecal bacteria biomarkers HF183/BacR287 (−3.5/d) (Greaves et al., 2020). PMMoV has also been successfully identified within surface water samples at ranges 3.0×10^3–1.1×10^5 GC/100 ml for rivers, up to $8{\cdot}73 \times 10^4$ GC/100 ml in coastal waters and $1 \times 10^{3.9}$ GC/100 ml in shellfish fisheries (Rosario et al., 2009; Hamza et al., 2011; Kuroda et al., 2015; Symonds et al., 2016; Gyawali et al., 2019). Due to its intact capsid PMMoV can survive within sediments for up to 121 days (Calderon et al., 2022) and has been detected in groundwater wells next to soil aquifer treatment systems, which act as a large soil treatment unit for a centralised wastewater treatment plant (Morison et al., 2020). However, there are few studies which have analysed the presence, pathways, and persistence of PMMoV from on-site domestic wastewater treatment systems (Kuroda et al., 2015; Rosiles-González G et al., 2017). This study will analyse the presence of PMMoV at increasing scales of complexity; at controlled laboratory scale with regular dosing of spiked PMMoV solution onto a soil column, at field scale spiking PMMoV into a soil treatment (percolation) area and monitoring its prevalence at different depths, and at a catchment scale in three streams in rural catchments with high densities of OWTSs. The aim of this staggered investigation is to conclude whether PMMoV can be used as a suitable biomarker for faecal source tracking in rural catchments which have suspected failing OWTSs.

6.3 Methods

6.3.1 Soil Column Set up and Spiking Trial

A soil column reactor had been constructed in the laboratory using 6 mm 'MarCryl' Acrylic plastic sheet (Access plastics Ltd., Ireland) to partition a large glass tank into parallel columns, each at a depth of 30 cm, width of 30 cm and breadth of 8 cm. At the time the spiking trial, the column was in a steady state having been dosed continuously with septic tank effluent (see average loading rate below) for a total 710 days, thereby reflecting full-scale soil treatment system functionality. The column was filled with a sandy loam subsoil excavated at 1.2 m deep at Redcross in County Wicklow. The septic tank effluent was collected from an on-site wastewater treatment system for a four-person household which employs a 4,760-litre Aswasep two-chambered septic tank (Molloy Precast Products Ltd., Ireland) as primary treatment. To prepare the spiking solution, the septic tank effluent was combined with commercial Tabasco ® pepper product (McIlhenny Company, USA) at a dilution rate of 1:10 for days 1, 2 and a dilution of 1:100 for days 3–5. A controlled study was carried out using the soil column intermittently dosed with 1 L/d of spiking influent for five consecutive days. Dosing was performed in 4 daily cycles of 250 ml at 3-hour increments at 09:00, 12:00, 15:00, 18:00 to mimic household flow pulses for a total flow of 1 L/d and hydraulic loading rate of 22.2 L/m²/d. Effluent samples, collected at the reservoir beneath the column, were retrieved once per day before the first dose to allow for 24-hour intervals in sampling. Samples were stored at -80^{o}C until concentrated.

6.3.2 Field site, Instrumentation and Spiking Trial

The field site was a single occupancy household, 50 km south-east of Dublin. The domestic wastewater was treated by a Aswasep two-chambered septic tank (Molloy Precast Products Ltd., Ireland), followed by a series of four percolation trenches, each 18 m long and 0.5 m in width, with a gradient of 1:200 filled with 300 mm pea gravel surrounding a perforated rigid plastic pipe, as per the Irish EPA Code of Practice (see figure S3) (EPA IE 2021). The pipes were then covered with 150 mm gravel and a geotextile fabric to prevent backfilled topsoil from washing beneath the gravel layer. During the spiking trial, the site had been operational for 1043 days and monitored as an ongoing research site (see Criado Monleon, 2022). Following the spiking, one percolation trench was monitored for five days, as to determine the quantities of PMMoV transport from the treatment system. The spiking solutions consisted of the household primary effluent (septic) mixed with Tabasco pepper product (McIlhenny Company, USA) at 1:100 dilution. The spiking solution totalled a volume of 20 L. Four pulses of 5 L were injected into the study trench evenly over an hour, i.e. one pulse every 15 minutes. This field research site had an automated weather station (Campbell Scientific, United Kingdom) measuring air temperature, relative humidity, atmospheric pressure, net radiation, wind speed and direction, and rainfall. Hydraulic loadings were determined using calibrated tipping buckets. A network of suction soil porewater lysimeters (Model 1900, Soil

moisture Equipment Corp., United States) were installed spaced longitudinally along each trench and at three depths beneath the infiltrative surface down to 50 cm depth in the soil. For sample collection, a suction of 50 kPa was applied using a vacuum-pressure hand pump and samples were collected 24 hours later. Porewater samples (n=31) were taken at three positions at 1 m, 8 m and 17.5 m from the trench inlet, at depths of 10 cm, 30 cm and 50 cm for each of these positions. An additional spiking solution sample was taken each day from storage on-site to determine persistence within the influent (n=5) in field conditions. The effluent and porewater samples extracted from lysimeters were stored on ice for <2 h of transport to be analysed in the environmental engineering laboratory at Trinity College Dublin. For comparative analysis matching volumes were used in qPCR and Colilert analysis.

6.3.3 Rivers, Drainage Network Selection and Sampling

The rural catchments selected for this research had previously been selected and profiled due to their high density of OWTS sited on low permeability soils which created a significant potential pressure on the catchments (**Table 6.1**) (EPA IE., 2018). Twenty-four-hour sampling was performed downstream of the clusters of OWTS in all catchments using the ISCO 3710 Portable Composite Sampler (Teledyne ISCO, USA) and the MAXX TP5 Portable Composite Sampler (MAXX-GMPH., Germany). Samples were analysed in aggregated 2-hour blocks; additional three grab samples were taken for each location except catchment 3 in which one grab sample was taken. Spot depth and velocity recordings were converted into discharge data by developing stage/discharge relationships at each monitoring point using the midsection area/velocity method (ISO 748:2021).

Table 6.1. Characteristics of the selected study catchments.

	Area (km^2)	**No. of OWTSs**	**Density (no./km^2)**	**Likelihood of inadequate percolation to groundwater**	**Land use**
Catchment 1 (C1) – Duran, Co. Wicklow	3.07	78	25.4	Very high (1.08 km^2)	Pasture (46%) Arable (10%) Forestry (33%) Other (11%)
Catchment 2 (C2) – Prospect, Co. Wexford	2.95	97	32.9	Very high (2.80 km^2)	Pasture (83%) Arable (0%) Forestry (2%) Other (15%)

	Area (km^2)	No. of OWTSs	Density ($no./km^2$)	Likelihood of inadequate percolation to groundwater	Land use
Catchment 3 (C3) – Kilnacrott, Co. Cavan	3.85	60	15.6	Very high (3.85 km^2)	Pasture (85%) Forestry (15%) Other (<0.5%

6.3.4 Organic and Nutrient analysis

Total organic carbon (TOC) was determined using a Shimadzu TOC-V analyser (Shimadzu Scientific Instrument, United States). Nitrogen species were analysed as nitrate-nitrogen (NO_3-N), nitrite-N (NO_2-N), and ammonium – N (NH_4-N) and phosphorus as orthophosphate (PO_4-P) using a Konelab 20i chemistry analyser (Thermo Scientific, Finland).

6.3.5 Enumeration of Faecal Indicator Bacteria

E. coli were enumerated for these samples using the IDEXX Colilert-18 assay according to ISO 9308-2:2012. 1- 100 ml of samples were mixed with dehydrated Colilert medium and incubated in Quanti-Trays at 36.0 ± 2.0 °C. *E. coli* concentrations for these samples are determined as most probable number per 100 ml (MPN/100 ml).

6.3.6 RNA Extraction and PMMoV Quantification

Water samples (10–250 ml) were concentrated by filtration through 100kda centricon unit membranes (Sigma-Aldrich, Germany) and stored at −80 °C and with RNA later (Thermofisher, USA) for long-term storage (3–6 months). Concentrated samples were extracted using RNeasy Power Microbiome Kit (Qiagen) Immediately following DNA extraction, triplicate qPCR reactions from undiluted and 10-fold diluted samples for each RNA extraction were conducted. For each RNA extraction were conducted with a positive control (1 ml of tabasco) which were included in every PCR assay.

To determine the copy number of PMMoV in porewater and surface water samples, a TaqMan assay was used to quantify the viral load in these samples, according to the manufacturer's recommendation (Applied Biosystems). Briefly, the assay was performed in a 20-µl reaction mixture containing 6 µl of sample RNA, 4ul TaqPath™ 1-Step RT-qPCR Master Mix, CG (Applied Biosystems), 2 µl of 4 mg/ml BSA (Bovine Serum Albumin, Thermofisher Scientific) 0.9 µM primers (PMMV-FP1; GAG TGG TTT GAC CTT AAC GTT TGA and PMMV-RP1; TTG TCG GTT GCA ATG CAA GT), and 0.3 µM probe (PMMV-Probe1; /56-FAM/CC TAC CGA AGC AAA TG/3BHQ_1/) (Zhang et al., 2006). The real-time RT-PCR reactions were carried out in a

StepOnePlus System (Applied Biosystems). PMMoV MST marker concentrations were expressed as gene copies per 100 ml (GC/100 ml).

The RT reaction was performed at 50 °C for 15 min, followed by 2 min at 95 °C for activation of DNA polymerase, and then 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Negative controls (RT-less and RNA-less) ,serial dilutions [10^1–10^8 gene copies GC/μL] of Pepper mild mottle virus gene ssRNA positive-strand viruses gBlock fragment (GenBank accession AB069853) was used to generate a standard curve (Haiwagara et al., 2001). The threshold of detection was established as the minimum DNA concentration detected in 95% or more of the replicates, while the threshold of quantification was ascertained as the minimum DNA concentration quantified within a range of 0.5 SDs from the logarithmic concentration. Optimisation steps determined inhibition, concentration, and extraction losses for developing spiking solutions. One hundred millilitres of influent and effluent from the soil columns was concentrated; two additional 100 ml of influent and effluent were spiked with 1 ml of Tabasco, as was 100 ml of distilled water to determine losses in concentration. Losses from long-term storage were analysed by 50:50 mix of RNA Later (Thermofisher) and Tabasco®.

6.4 Results and Discussion

6.4.1 Pepper Mottle Virus Transport and Removal: Laboratory Soil Column

A 5-day trial was performed on a subsoil column (**Figure SC1**) which had been dosed with effluent and monitored extensively for 710 days and developed an established biomat. The experiment provided a lab scale reproduction of a section of the percolation trench receiving domestic wastewater from a septic tank. Initial PMMoV concentrations measured the day before the trial showed concentrations for effluent and influent measured of 8.72×10^5 and 8.3×10^7 GC/100 ml respectively giving a $\log_{10}$ reduction value (LRV) of 1.74. Total coliforms and *E. coli* LRVs values monitored before the experiment averaged at 4.9 ± 0.9 and 4.5 ± 0.5, respectively. During the trial the PMMoV concentration of the spiked influent rose to 1.3×10^8 and 2.69×10^9 GC/100 ml for days 1 and 2 respectively. The average LRV for days 1 and 2 of the trial were 2.58 (**Figure 6.2A/B)**, however, in days 3 and 4 a large drop in the LRV of 1.35 and 0.1 was noted respectively. On day 5 effluent PMMoV concentrations exceeded influent values, these were discounted. The addition of the tabasco to the influent increased PMMoV marker concentrations, concentrations increased at a range of $\log_{10}$ 0.4 and $\log_{10}$ 1.78 for days 1 and 2 relative to non-spiked influent. The increase in dilution of 1:100 resulted in a significant drop in influent concentrations of PMMoV for days 3,4, 5, with concentrations of PMMoV being an average $\log_{10}$ 1.75 less than days 1 and 2 of the experiment. The increased dilution of the spiking solution also coincided with a log 10 increase of effluent PMMoV following day 2 of the experiment. The high acidity caused by the relatively high concentration of Tabasco® spiked into the influent in days 1 and 2 seemed to diminish *E. coli* concentrations found in the influent sample resulting in far lower LRVs averaging at 1.09 ± 0.46

during the experiment, far lower than the prior long-term monitoring LRVs for the system for *E. coli* of 4.5 ± 0.5.

There have been few column studies which look at viral attenuation within the infiltrative zone. Van Cuyk and Siegrist (2007) showed LRVs of 1.02 to 2.0 for a selection of bacteriophages, but this was only through 4 cm in depth of soil compared to the 30 cm depth soil columns used in this study which may account for the higher LRV values found. The adsorption of viruses onto porous media is intricately linked to the properties of proteins within the virus's outer capsid. Additionally, the electrostatic repulsion between viruses and the media can be mitigated through factors such as lower pH levels, elevated ionic strength, and including divalent cations within the systems (Wang et al., 2021). PMMoV has been noted to adsorb at greater rates in higher acidity soils (Yoshimoto et al., 2012). Note that the initial influent dosed into columns on days 1 and 2 at dilution of 1:10 had low pH averaging at 3.85 due to the high concentration of Tabasco within the influent. In subsequent days 3 to 5 in the spiking trial, the influent pH increased to a range of 5.46 – 5.73 due to the lower concentration of the spiking solution dosed at 1:100 dilution of Tabasco product to influent. As the pH increased, this resulted in an increase in the concentration of PMMoV within the effluent, which is shown by the significant reduction in LRV through the column. The adsorption of PMMoV has been noted to be greater in soils of greater acidity, which appears to match our findings over the first days of the trial in which influent of an average pH of 3.85 was dosed onto the system showed the highest PMMoV LRV value of 2.58 (Dana et al., 2012; Yoshimoto et al., 2012). The concentration of PMMoV was then reduced in influent on days 3–5 by $\log_{10}$ 1.75 by dilution causing the influent pH to rise to an average 5.49. However, this seemed to promote a reduction in LRV values to an average of 0.3 ± 0.7 which suggests the higher pH may have either inhibited adsorption of PMMoV to the soil and/or promoted desorption of previously sorbed virus congruently. The initial pH values caused by the addition of acidic Tabasco ® product were extreme when compared to domestic household wastewater, which has a more neutral pH of 6–7. It is important to note that effluent pH across the trial ranged from 6.75 – 7.31 which may indicate adsorption capacity was greatest near the infiltrative surface and that there was neutralisation of the highly acidic spiked influent solute down through the soil column.

The addition of organics within wastewater into soil environment greatly accelerates the development of the biomat which plays a key role in the removal of viruses and other pathogens. The importance of pre-treatment and its impact on biomat development, and corresponding pollutant attenuation has been shown in field studies of soil treatment units (Gill et al., 2009; Knappe et al., 2020). For comparison, our column was dosed with 297.6 ± 99.8 mg/L per day TOC. The use of primary treated influent (i.e. septic tank effluent) on the column and the establishment of well-developed biomat may be why the column in this study with at a depth of 30 cm yielded a PMMoV LRV average of 2.58 which exceeded LRV values observed by Betancourt et al. (2019) who found an LRV of 1.3 at depths of 1.8 m in columns dosed with secondary treated effluent DOC of

approximately 4 – 6.8 mg/L and a HLR of 6.22 L/m^2/d. The development of the biomat and its effect on viral removal was noted by Van Cuyk and Siegrist (2007) who observed that phage removal efficiencies increased with time and were influenced by hydraulic loading rates which ranged from 5 to 25 L/m^2/d, dosing effluent with an organic concentration of organics of 504 mg/L (COD). The development of the biomat is crucial for effective filtration properties of the soil media as the clogging results in reduced pore sizes (Stevik et al., 2004) which appears to be a causative factor in this study for the large LRVs for *E. coli* at 4.39 within the column effluent. Finally, a multiparametric analysis was conducted on column effluent samples (**Figure 6.2C)** which shows a significant correlation between PMMoV with *E. coli* and total coliforms in effluent concentrations (R^2 =1, P value <0.05; R^2 =1, P value <0.005, Spearman), indicating its potential as an alternative fingerprinting tracer for faecal contamination by human effluent.

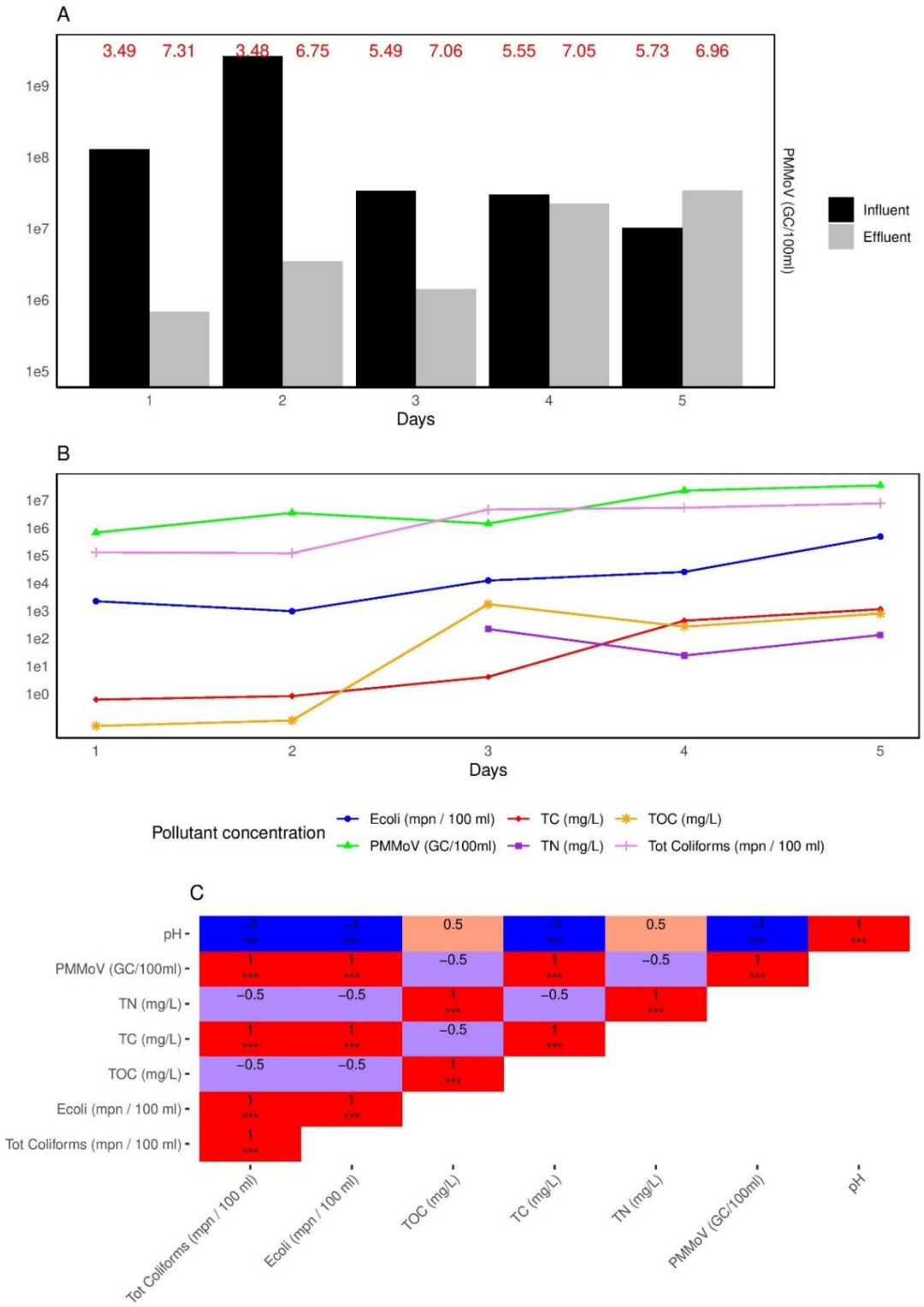


Figure 6.2 (A) Concentration of PMMoV (Pepper Mild Mottle Virus) gene copies per 100 ml of soil column influent and effluent across a span of 5 days during the trial. Influent and effluent pH values in red **(B**) Column effluent biological parameters include *E. coli* (*Escherichia coli*) and total conforms (Most Probable Number per 100 ml), PMMoV GC/100 ml. Column effluent chemical parameters encompass total carbon (TC), total nitrogen (TN), and total organic carbon (TOC). (**C)** Spearman's correlation and regression values are presented for the concentrations observed in the soil column effluent. Significance levels are denoted as follows: $p < 0.05$, $p < 0.01$, and $p < 0.001$.

6.4.2 Pepper Mottle Virus Mild Mottle Virus Transport and Attenuation: Field-Scale On-Site Wastewater Treatment System.

A spiking trial was conducted for 5 days in August 2023 on the full-scale on-site wastewater treatment system serving a private house in County Wicklow. This site had been monitored as a research site for 1043 days before the spiking trial. During the trial, concentrations of PMMoV and *E.coli* collected in the porewater samples within the soil treatment area varied diurnally (see Table 6.2). An additional 5 Primary influent sample which had been mixed with the spiking solution was stored at 4 ^{o} C and analysed to determine LRVs for each corresponding day. Detection of PMMoV was lowest on day 1 with only 25% of samples detecting the virus compared to *E. coli* which was detected in 50% of the samples, as shown in Table 6.2. The concentration of PMMoV and *E. coli* peaks on day 2 at a concentration of 1×10^6 GC/100 ml and 1×10^3 MPN/100 ml, respectively (**Figure 6.3A**). The peaks in PMMoV in day 2 may be linked to heavy rainfall (19.4 mm) which preceded it in day 1, given the relative high soil hydraulic conductivity, measured in-site by a constant head permeability test to be 0.163 m/h (**Figure SC1**). Conditions onwards across the spiking trial were dry with rainfall ranging from 0 to just 0.8 mm per day. These peaks in PMMoV on day 2 were observed at 17.5 m and 1 m downstream, respectively, from the inlet of the trench (**Figure 6.3A**). The highest concentrations PMMoV 1.03 × 106 was found on the second day of the trial found at 17.5 m from the effluent inlet at a depth of 15 cm, no E. coli was found at this position.

Table 6.2. Detection rates in porewater samples

	Day 1 (n=4)	**Day 2 (n=8)**	**Day 3 (n=8)**	**Day 4 (n=4)**	**Day 5 (n=4)**
PMMoV	25%	62.5%	75%	75%	50%
E. coli	50%	75%	50%	75%	50%

LRV values for *E. coli* and PMMoV across the 5 days of the trial varied diurnally and spatially within the system. The highest levels of LRV for PMMoV were for day 1 at 4.36, which may be due to the low dispersal of the spiking solution at this time in the trial (**Figure 6.3B**). The lowest PMMoV LRV 2.5 was recorded in the middle of the trench at 7.5 m at a depth of 10 cm with daily concentrations ranging at this position from 6×10^4 to 4×10^5 GC/100 ml. The highest PMMoV LRV across depths for an average of 3.11 was positioned proximal to the trench inlet at 1 m. The fact that the most proximal point to the inlet is where the greatest removal takes place mirrors previous observations on biomat development and the influence on the soil filtering capacity (Gill et al., 2009; Knappe et al., 2020). *E. coli* LRV values were highest at 3.74 at the rear of the trench at 17.5 m from the inlet. Lowest LRV were 2.5 at the midpoint of the trench at a depth of 10 cm with concentrations ranging from 364–419 MPN/100 ml. Across the trench LRVs appear to increase with time for *E. coli* but reduced with time for PMMoV, with a relatively steady state of removal for both indicators over days 2 to 4 (**Figure 6.3B**) The average LRV for *E.coli* within the STU (3.01)

was greater than that of PMMoV (2.88), due to pore size viral filtration is more challenging than bacteria in subsoils.

The risk of viral exposure due to transport is greater than bacteria as shown in groundwater samples taken from karst aquifers in Mexico. In the study of Mexican karst systems low *E. coli* measurements at 8 and 1 MPN/100 ml (in wet and dry conditions respectively) when compared to 1.7×10^1 to 1.0×10^3 GC/100 ml for the viral PMMoV with neither indicator significantly correlating (Rosilis et al., 2017). A significant correlation between *E. coli* and PMMoV concentrations was also not observed in this study (**Figure 6.3D**). The lack of correlation may also be due to the limitations in sample volume extraction given sometimes the lysimeters only accumulated low quantities of porewater. The small volume samples required dilution for further analysis which may have reduced resolution. Furthermore, it should be noted that there was no significant correlation between PMMoV and any of the other parameters monitored in this study. Another study has shown no correlation between *E. coli* and human Rotavirus in contaminated well samples; however, *E.coli* has been known to survive in groundwater conditions to up 470 days which may explain the lack of correlation within this study (Cook et al., 2007; Ferguson et al., 2012).

Using the LRV values taken from the spiking trial we estimate PMMoV concentrations values within trench which would likely correspond to conventional dosing of primary influent with a concentration of 5 $x10^7$ GC/100 ml. The likely values found within the PE fed STU would result in max and minimum concentration of 1×10^5, 2×10^3 GC/100 ml respectively. The legislation in Ireland (where many parts of the country have relatively high-water tables) require on-site wastewater treatment systems to have at least 1.2–2 m depth of unsaturated subsoil in the soil treatment unit (EPA IE, 2021) like many other jurisdictions around the world. Column studies by Betancourt et al. (2019), have shown a 95% reduction PMMoV at a depth of 1.8 m, starting from the estimated minimum (depth 30 cm) and maximum (depth of 50 cm). Employing the Betancourt et al. (2019) LRV values infer that between 100 GC/100 ml and 7 $x10^3$ GC/100 ml may be reaching the groundwater. PMMoV has been shown to reach groundwater at far lower depths of 40 m to 83 m at monitoring wells next to soil aquifer treatment plant in Tucson, Arizona, where 22% of samples from the wells tested positive with PMMoV with average concentrations of 6 GC/ L and 3 GC/ L Morrison et al., 2020). Hence, PMMoV has been shown to be an effective faecal indicator in field studies. It has the advantage that it is a source specific to humans (and being non-infectious) and is present in far higher concentrations than for example *E. coli* and so would likely reach water tables from on-site systems but not pose a threat to human health (Zhang et al., 2006).

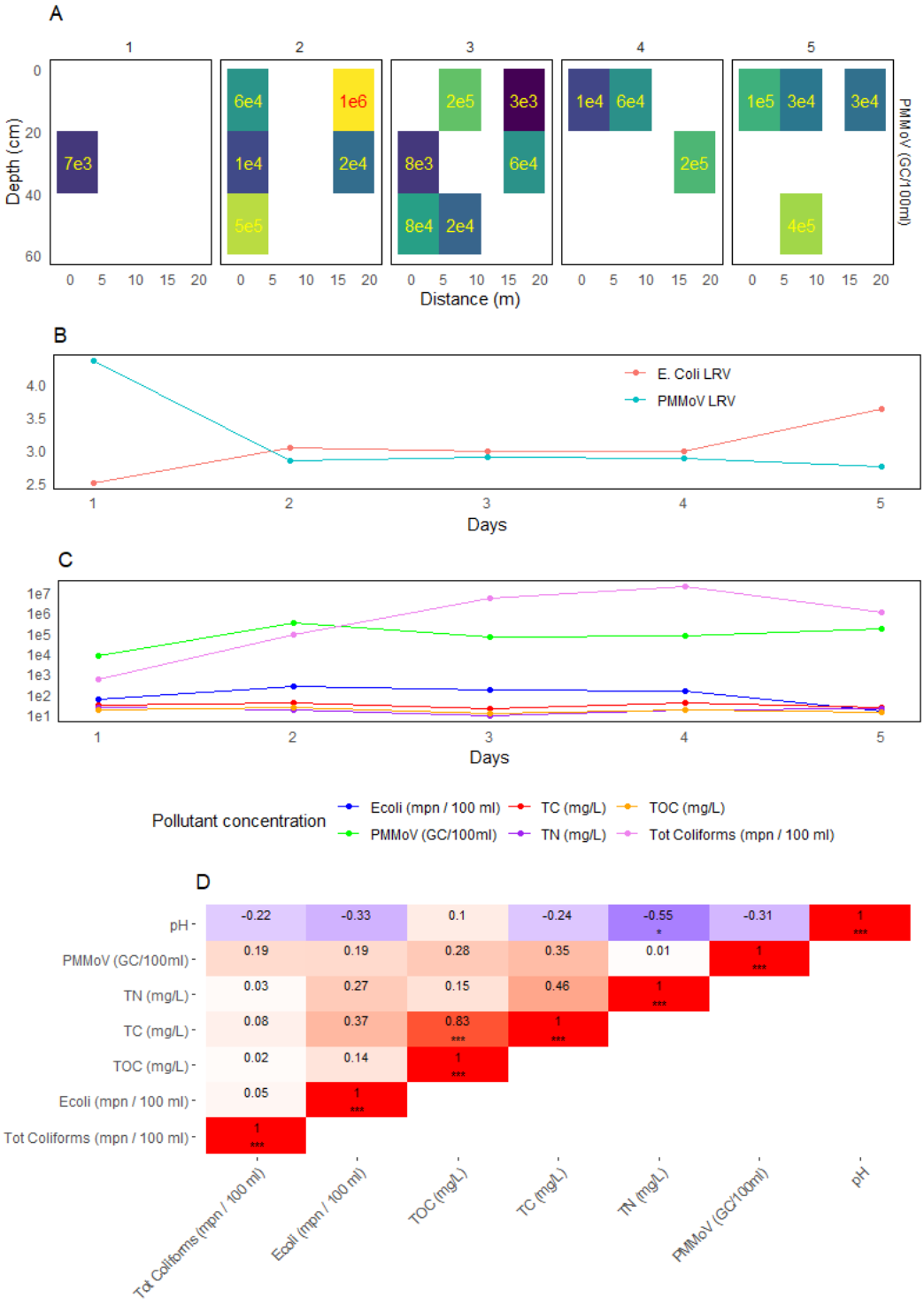


Figure 6.3. (A) Concentration of PMMoV (Pepper Mild Mottle Virus) gene copies (GC) per 100 ml across depth (cm) and distance (m) of percolation trench across a span of 5 days during the trial. **(B)** Log10 removal values for *E. coli* and PMMoV across the five days of the trial **(C)** Biological parameters averaged across depths at different distances within the percolation trench variables include *E. coli* and total conforms (Most Probable Number per 100 ml), and the PMMoV GC/100 ml. Chemical parameters encompass total carbon (TC), total nitrogen (TN), and total organic carbon (TOC). **(D)** Spearman's correlation and regression values are presented for the concentrations observed in the soil column effluent. Significance levels are denoted as follows: p <0.05, p <0.01, and p <0.001.

6.4.3 PMMoV Detection and Enumeration: Catchment-Scale with High Contamination Pressure from On-Site Wastewater Treatment Systems

Three small river catchments in rural areas were selected for monitoring having been previously selected for being of high risk of on-site effluent pollution due to high densities of improperly installed and maintained systems in close proximity of the rivers (<100 m) and in areas of low permeability subsoils (Gill et al., 2018). The 24-hour monitoring results are shown below (**Figure 6.4A**). PMMoV was detected in all samples. Note, however, that samples from 05:00 to 12:00 at catchment C3 registered concentrations lower than LOQ and therefore were conservatively assumed to be the lowest concentration (i.e. limit of detection) of 180 GC/100 ml. Concentrations in PMMoV ranged from $8 \times 10^2 – 1 \times 10^4$, $6 \times 10^3 – 5 \times 10^4$, 2×10^2 - 1×10^4, GC/100 ml for rivers C1, C2 and C3 respectively. In comparison, *E. coli* concentrations ranged from $2 \times 10^3 – 2 \times 10^6$, $4 \times 10^1 – 4 \times 10^3$, 2×10^2 - 1×10^5, for rivers C1, C2 and C3 respectively. River C2 showed the highest average concentration of PMMoV (2×10^4 GC/100 ml) which had the greatest number and highest density of OTWSs, although this was not matched by *E. coli* concentrations where C3 had the greatest concentrations. Both sites accounted for large proportions of pastureland accounting for 83% land use for C2 and 85% for C3. Previous research on these catchments (Gill et al., 2018) estimated that OWTS contributed 8.5 ± 1.3% of total phosphorous loading and 12.1 ± 1.6% total nitrogen loading to catchment C2. Catchment C3 is noted to be less impacted by OWTSs with 1.3 ± 1.2% of total phosphorous loading and 6.8 ± 3.1% total nitrogen loading estimated to originate from OTWSs. There were notable correlations in catchment C2 between *E. coli* and total phosphorus concentrations ($R^2 = 0.8$, $P < 0.05$) (**Figure 6.4B)**. In catchment C2 PMMoV concentrations showed correlations with total phosphorous, *E. coli* and chloride ($R^2 = 0.7, 0.6, 0.5$ respectively), but none were significant. In catchments C1 and C3, there were no significant correlations between any of the parameters. Monitoring the diurnal variance of measured parameters within the three rivers (see Table SC2) showed biological parameters (total coliforms, *E. coli*, PMMoV) exhibited greatest variances, followed by chemical parameters NH_4-N, PO_4-P and NO_2-N. Dry sampling conditions resulted in little variance of chemical parameters, with little fluctuation in concentrations over the 24 hours monitoring periods. NH_4-N is an effective parameter for monitoring pollution, correlating with human biomarker HF-183 in urban catchments. However, the likelihood of biomarkers correlating with nutrient concentrations will be far lower in a rural catchment due to agricultural practices. In Ireland grassland pasture accounts for 92% of agricultural land use and the application of mineral fertilisers in Ireland is extensive with 343,193 and 34, 240 tonnes of N and P fertilisers applied annually, respectively (Reynolds et al., 2021, CSO 2023). Aside from the nutrient emissions from mineral fertilisers, the large population of cattle and subsequent production of urine and manure, with 35% of the manure being stored and spread, is also a significant source and contributes to agriculture's status as the most significant pressure to Irish catchment water quality (Mockler et al., 2007; Úlen et al., 2007; EPA IE., 2022). Therefore, the correlations seen with *E.coli* and total phosphorous within this catchment may be indicative that the *E.coli* in the rivers are more linked to

agricultural pollution sources, as catchment C2 is in the south-east of Ireland which is known for intensive livestock farming where the average stocking rate (grazing cattle and sheep) is 1.5 LU ha^{-1} (Úlen et al., 2007).

The lack of strong correlation between the PMMoV concentrations and the more traditionally used marker of faecal pollution, *E. coli*, is not surprising due to (i) the lack of correlation picked up in the soil treatment area trials (Section 3.2) where the only contaminant source in the soil is the human effluent being applied and (ii) the multiple agricultural sources of *E.coli* in such rural catchments in Ireland. These findings therefore suggest that *E. coli* is not a suitable marker of on-site wastewater effluent contamination into rivers. The human host specificity of biomarkers such as PMMoV holds an advantage faecal indicator bacterium in rural catchments. *E. coli* has been used as a standardised faecal indicator due to the belief that its survival outside the host is limited and that it is unable to grow in secondary habitats such as water, sediment, or soil (Ishii et al., 2008). Currently the sole use of *E. coli* as a faecal indicator has come under greater scrutiny as it is not hosting specific and has been accepted that *E. coli* can create self-sustaining naturalised populations in high nutrient, warm environments which can include temperate climates, especially with increasing global temperatures (Gordon et al., 2001; Ishii et al.,2008). *E. coli* may be suitable to determine water quality and public health risk in urban catchments or coastlines where point sources from municipal treatment plant discharges are readily identified by large fluxes in concentration (Devane et al., 2020). However, persistent diffuse faecal contamination of low concentration typical within rural environments characterised by dispersed on-site wastewater treatment plans and agricultural run-off requires more sophisticated techniques such as the use of PMMoV markers, possibly in conjunction with other fingerprinting compounds. PMMoV is ubiquitous in human effluent, with only small concentrations detected in dogs, wild fowl and chickens (Rosario et al., 2009; Hamza et al., 2011; Gyawali et al., 2019). It is therefore likely that all the PMMoV detected within these catchments originated from domestic household systems, proving its effectiveness in human faecal source tracking in a confounding rural environment. The factors controlling the source, transport and attenuation of other contaminants and any potential infectious microorganisms from OWTS effluent sources would be the same as those influencing the dissemination of PMMoV, such as concentration of systems in a locality, how well they have been installed and maintained and the contamination pathways influenced by soil type, meteorology, geology, topography, etc.

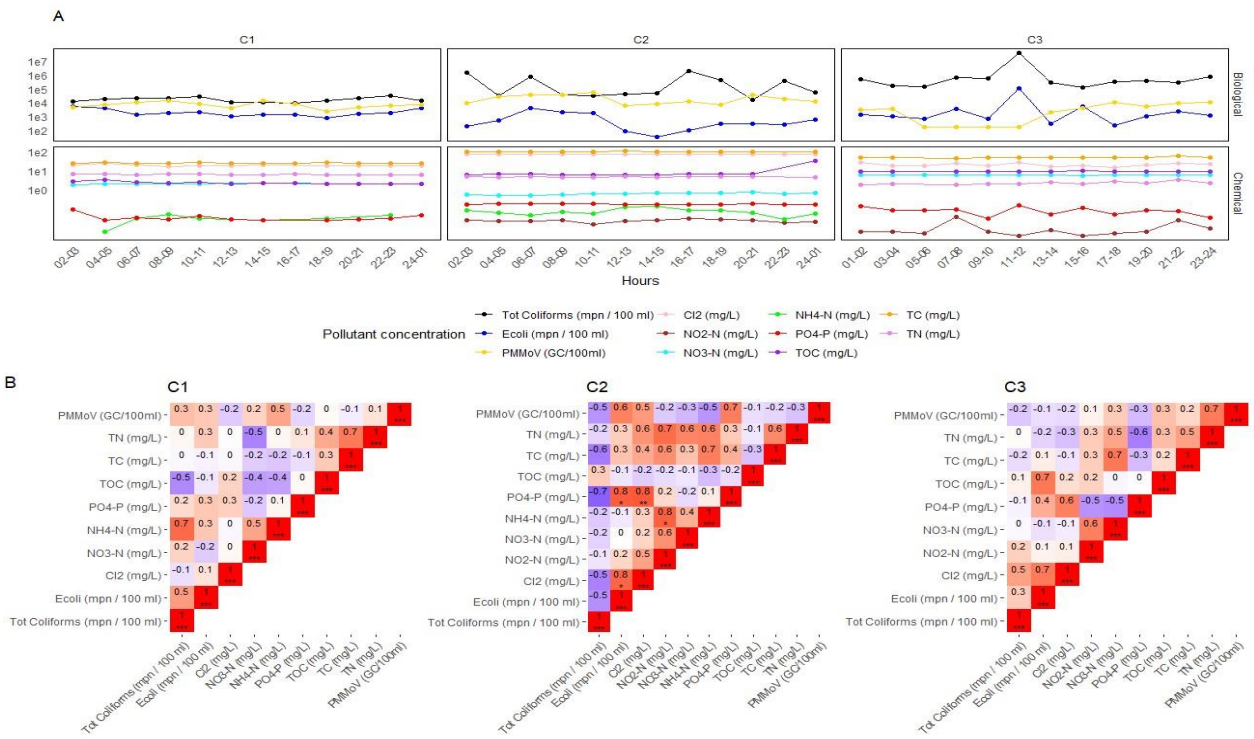


Figure 6.4. (A) (C) Biological parameters include *E. coli* and total conforms (Most Probable Number per 100 ml), and the PMMoV GC/100 ml. Chemical parameters (mg/L) encompass total carbon (TC), total nitrogen (TN), and total organic carbon (TOC), total phosphorous (PO4-P), ammonium (NH4-N), nitrite (NO2-N) nitrate (NO3-N) and chloride (CL-2) for rivers C1-C3. **(B)** Spearman's correlation and regression values are presented for the concentrations observed in rivers C1-C3. Significance levels are denoted as follows: $p < 0.05$, $p < 0.01$, and $p < 0.001$.

6.5 Conclusions

- PMMoV exhibited a $\log_{10}$ removal rate of 2.58 through a 30 cm soil column depth, demonstrating its transport through the microbial biomat and column receiving primary influent at the lab scale.
- Correlations were observed between PMMoV and existing faecal indicator bacteria, and total carbon in the soil column effluent. Additionally, pH exhibited a significant negative correlation with PMMoV concentrations.
- PMMoV was quantified in a full-scale soil treatment area with an average $\log_{10}$ removal rate of 2.88 at depths ranging from 10 cm to 50 cm.
- No significant correlation was found between PMMoV and *E. coli* concentrations in the soil treatment and PMMoV concentrations were much higher than *E. coli* across all the different sampling depths within the soil.
- *E. coli* LRVs were 3.01, much higher than that PMMoV which suggest that classical FIB approaches may underplay the risk of enteric viral transmission.
- PMMoV was found consistently in small rivers in rural areas with high densities of OWTSs during low flow conditions, indicating its potential as a reliable microbial source tracker.

7. Temporal Variations in Microbial Ecological Patterns in Soil Treatment Unit Biomats

Alejandro Javier Criado Monleon, Laurence Gill

This manuscript is being prepared for submission.

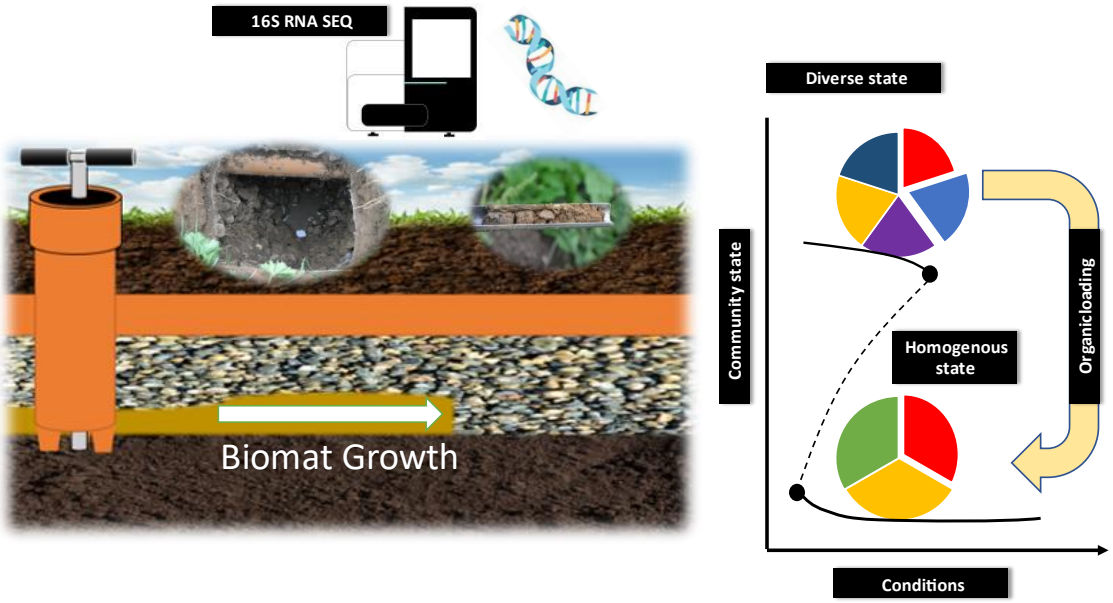


Figure 7.1, Physical and ecological changes to soil treatment unit biomat, with biomat lateral growth over time and significant shifts in biomat community structure due organic substrate loading.

7.1 Abstract

The development of microbial mats or 'biomats' within on-site wastewater treatment system percolation area is essential in protection of groundwater resources. Though essential in processing nutrients and organics from domestic households, little is known in how these microbial communities change over time as to facilitate their function in the engineered environment. This study to determine to frequently monitor subsoil community, to identify shifts in community composition and define a time of transition. The approach of the study was to monitor 3 research sites at different stages of operation and biomat maturation. At each site, the STU areas were split whereby half received effluent directly from septic tanks, and half received more highly treated effluent from packaged aerobic treatment systems (a coconut husk media filter, a rotating biodisc contactor – RBC – and an intermittent sand filter). One site was monitored intensively, to identify initial shifts in community composition as the 'young' biomat develops. The two remaining sites varying ages were monitored less frequently, but at an interval of four years to determine long-term

trends in biomat ecological developed in the more mature sites. In total 420 samples were taken over a period of four years across the three sites and characterised by 16S RNA sequencing. What was observed that within 3 years the young biomat de organic rich effluent exhibited significant changes in alpha diversity and community structure in the first 5 metres of the soil treatment unit, indicative of the development of nutrient cycling niches. The development of these niches corresponded increases in abundance of functional groups responsible for biogeochemical processes and decreases in concentration of nutrients below the soil. In the more mature sites, the effect of nutrient loading had caused tipping points with systems being dosed with high organic loads shifting at a more accelerated rate to more homogenous communities. This research shows that the combination of high nutrient loads, clogging and the resulting biomat development leads to a lateral drift of substrates creating lagged response of soil communities to a high organic environment.

Keywords: on-site wastewater treatment, microbial hotspots, tipping points, Community assemblage patterns, time-series

7.2. Introduction

On-site wastewater treatment systems (OTWSs) are extensively used globally with 41% of the world's population using these systems as their means of sanitation (WHO, 2020). These are usually comprised of a septic tank which acts primarily as a settling chamber with anaerobic conditions which aid the removal of organic and suspended solids but has little to no effect on nutrients and microorganisms (Gill et al., 2007). However, it is increasingly common that households install a packaged secondary treatment system that usually applies aerobic processes to treat waste, either in series with the septic tank or replacing it altogether. Treatment is completed through the dispersal of the effluent into a soil treatment unit (STU), often known as the percolation area. The STU comprises of a series of percolation trenches or dispersal beds in which an even distribution of effluent can be dispersed over the subsoil surface for effective attenuation (EPA., 2021).

STUs are a nature-based solution which can provide sustainable passive wastewater treatment. STUs should be constructed following an appropriate on-site assessment for percolation characteristics, unsaturated subsoil depth and the risk profile of underlying aquifers (EPA., 2017). The treatment in these systems is enhanced by the development of microbial mats at the interface where the effluent meets the soil or subsoil bound biofilms which reduce hydraulic conductivity due to clogging by the production of extracellular polymeric substances (EPS) (Beal et al., 2006; Gonçalves et al., 2007; Adessi et al., 2018). The reduction of hydraulic conductivity through biomat induced bioclogging improves the effectiveness of the system by aiding effluent dispersal across the entirety of the system surface area, but also improves pollutant removal through adsorption, biodegradation, and filtration by the biomat directly (Knappe et al., 2020). The biomat is understood to develop in three phases, the first phase is noted by a large drop in hydraulic conductivity caused by physical clogging from deposition of suspended solids (Beal et al., 2006). Stage two is characterised by a more gradual drop

in hydraulic conductivity over time and stage three marks a point in equilibrium with low infiltration rates (Beal et al., 2005).Column studies assessed that soil ecologies within an STU reach a 'steady-state' when there are no fluctuations in microbial populations, unless there are significant alterations to the environment such as changes to the soil volumetric water content (VWC) (Fernandez-Baca et al., 2019). Previous work has noted that the level of pre-treatment has shown to have significant effect on the growth rate of the biomat, with lower organic loads resulting in stunted biomat growth and altered community structure (Knappe et al., 2020; Criado Monleon et al., 2022). The application of higher organic loading and subsequently higher biomat growth showed these systems to be more resilient to drought conditions than systems dosed with secondarily treated effluent (Knappe et a., 2020). Ultimately, this subdued, less diverse biomats have areas of greatly reduced attenuation capacity described as "infiltrative dead zones" (Gill et al., 2009;. Knappe et al., 2020; Criado Monleon et al., 2022).

With advances in molecular techniques, more extensive sampling and analysis for microbial community structures have been possible. A significant advancement is the development of the Microbial Database for Activated Sludge by Dueholm et al. (2021) which was then applied to independent dataset consisting on amplicon libraries of 269 wastewater treatment plants across the globe. Temporal analysis of microbial communities using 16S amplicon sequencing has been increasing employed in wastewater treatment systems to gain increased levels of insights into the processes such as activated sludge systems (Ju and Zhang et al., 2014; Jiang et al., 2018), anaerobic digesters (Hashimoto et al., 2014) Seasonal changes can result in strong taxonomic shifts, as a combination of short sludge retention times combined with a drop in temperature results in reduced growth rates and washing out of slow growing bacteria, however it could be argued that this was more aligned to changes in operation (Griffin and Well., 2017). In comparison to natural ecosystems bacterial communities within activated sludge plants generally did not succumb to seasonal succession, due to controlled artificial nature of the ecosystems shifts in communities took place annually in the winter which coincided issues in processes of the plant, specifically severe foaming within the aeration tank within the plant (Ju and Zhang et al., 2014).

Within the human impacted environment temporal dynamics of polluted rivers sediments have shown greater levels of diversity and was significantly higher in the winter than when sampled in the summer (Wang et al. 2023). The shift in community structure may result in increases in organisms involved in nitrogen and methane cycling due to the degradation and deposition of organic matter products through the summer and early winter (Cruaude et al. 2020; Wang et al., 2023). Monitoring of the seasonal alteration of the microbial community structures in the urban estuary of the Yangtze River showed clear seasonal differentiation, with losses in alpha diversity in the summer whilst beta diversity revealed seasonal clustering (Yi et al., 2020). The authors speculated that the seasonality of precipitation may have caused the migration of certain phyla from terrestrial soil to sediment causing the shifts in the community structure. Sediment community

structures has shown to be more impacted to temporal temperature changes than communities located within the water of urban estuaries (Zhang et al., 2020). In soils irrigated with treated wastewater it was noted using T-RFLP that bacterial communities changed in response to irrigation in the first 8 weeks (Krause et al., 2020). Amplicon sequencing within the same study determined that, regardless of irrigation water quality, diversity declined with the most responsive phlya being Proteobacteria, Bacteroidetes, Actinobacteria, and Planctomycetes (Krause et al., 2020). Long term studies of wastewater irrigation systems found that seasonality impacted community structure in irrigated soils but did not affect community structure within non-irrigated soils, which may be attributed to higher moisture content found within the topsoil of the constantly irrigated soils which showed less seasonal fluctuations than the non-irrigated soils (Wafula et al., 2015). In less impacted deeper subsoils sites have shown seasonal dynamics in groundwater ecologies of river water intrusion that was correlated to the alteration of some microbial assemblages located at top of the aquifer, specifically the phyla Actinobacteria (Lin et al., 2012)

To date there appears to be no temporal analysis on the microbial community structure within soil treatment systems at field scale. This is likely due to the challenges involved in sampling subsoil located under the percolation pipe infrastructure installed to distribute the wastewater. In this study over 400 subsoil samples were taken from three different full-scale operational STUs over a period of 3 years. The aim of this study is to determine critical points in transition for bacterial ecologies found within biomats fed by two separate influent streams for each site – septic tank (primary treated) effluent versus secondary treated effluent. Two of the sites were sampled in parallel to compare the variation exhibited by the biomat ecologies over the space of 2 years. Intense monitoring was conducted for one of the sites for three years where changes in the community structure were measured in conjunction with environmental and influent chemical factors monitoring the temporal composition dynamics.

7.3 Methods

7.3.1 Sites description

This study investigates the spatial distribution of microorganisms in three separate OWTSs located in three owner-occupied homes in Ireland. The region is classified as a temperate oceanic climate (or Cfb classification within the Köppen climate classification system) (De Carli et al., 2018). Subsoils at sites A and B (in Co. Limerick) are classified as typical Luvisol soils over a Limestone till (Carboniferous) averaging at pH 8 with little observed variance; the subsoil at site C (in Co. Wicklow) is classified fine loamy drift with siliceous stones with a clayey sandstone and shale till subsoil (Lower Paleozoic) (diamictons) (Fealy, 2009). Microbial biomats - the principle focus of the study - had fully developed at research sites A and B (developing in C), which form at the interface where the effluent percolates into the soil (i.e., the STU) at the base of the gravel percolation trenches. Each percolation trench was 18 m in length and 0.5 m in width, with a gradient of 1:200

filled with 300 mm pea gravel in which a perforated rigid plastic pipe was set (as per EPA, 2021 design guidelines; **Figure 7.2**).

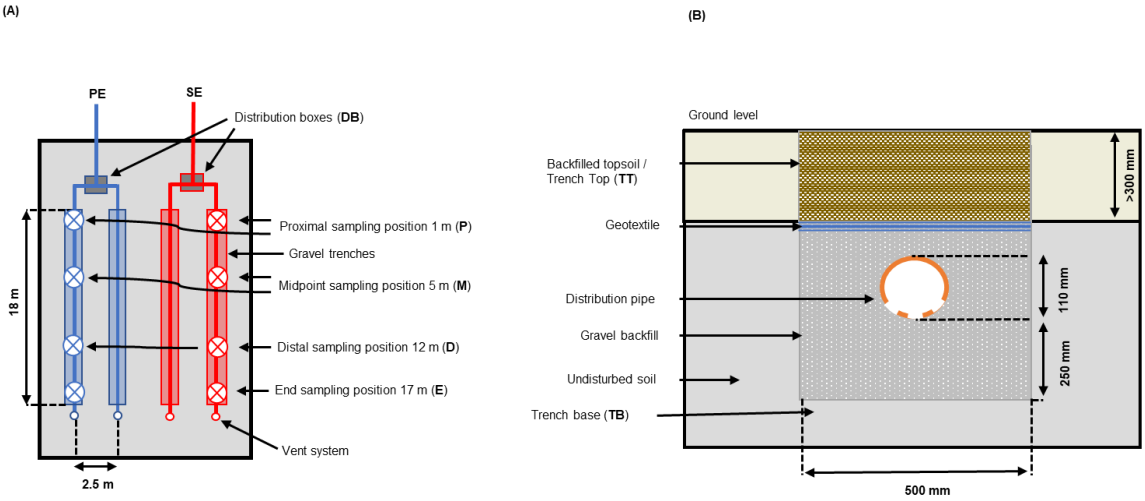


Figure 7.2. General schematic of the soil treatment unit for all sites, with sampling positions proximal 1 m, midpoint 5 m, distal sampling point 12 m and end sampling position 17 m.

The pipes were then covered with 150 mm gravel and a geotextile fabric to prevent backfilled topsoil from washing beneath the gravel layer. All three sites employed a 4,760-L Aswasep two-chambered septic tank (Molloy Precast Products Ltd., Ireland), and four percolation trenches, two of which were fed directly from the septic tank system as the primary effluent (PE), and the other two percolation trenches were fed with the secondary treated effluent (SE) from packaged treatment systems (**Figure 7.2**). At Site A, secondary treatment was achieved using an intermittently dosed coconut husk filter system (Ecoflo Coco Filter, Premier Tech Aqua Ltd., Ireland). At Site B, the packaged secondary treatment system was an RBC (Klargester BioDisc, Kingspan., Ltd, United Kingdom), consisting of an integrated primary settling chamber, a two-stage biozone, and a secondary clarifying chamber. In site C, an open intermittent sand filter was installed to a depth of 0.9 m with sand grain sizes of D10 range from 0.4 to 1.0 mm. In all sites, the primary and secondary effluents were distributed equally onto their own separate half of the STU. The even distribution of the effluents between these parallel trenches was ensured using calibrated tipping bucket distribution devices (Patel et al., 2008), which were instrumented with reed switches to calculate the daily flows to each STU. The soil and hydrogeological factors from each site (as determined from a parallel research study; Knappe et al., 2020).

7.3.2 Sampling and 16S RNA Gene Sequencing

For sites A and B, the soil was excavated from the surface down into the percolation trenches to the gravel subsoil interface at the base of the trenches. In site C subsoil biomat access ports or" bioports" were installed within the percolation trench to add ease to sampling (see section 3.4.3). Core samples were then taken with a 25.4-mm stainless steel corer. Samples were taken after 1080 (n=46) and 2540 (n = 90) days of the system running at Site A and 860 (n = 46) and 2320 (n= 92) days on Site

B, these sites represent the longer duration observation of the development of the biomat. Site C which represents the initial observations of the biomat was sampled for intensely (n = 146) on 60, 275, 547, 641, 793, 882, 944, and 1037 days in operation.

For each system, a single sample was taken at three horizontal positions measured from the head of the gravel trenches as follows: Site C (0 m, 5 m, 17m), Site A and B (0 m. 5 m, 12.5 m and 17 m). At each position different depths from the unsaturated subsoil beneath the point of wastewater infiltration at the base of the trench were sampled from 0 to 12.5cm (**Figure S2**). A sample was also taken from the surface interface (i.e., 0 cm depth). For DNA extraction of the samples for sequencing, ~3 g of soil was collected from the core and placed into a sterile 2 ml Eppendorf tube and stored at −20°C. Sample handling was carried out using a sterile metal spatula, with sterilizing performed between each sample using 70% ethanol. Then, 250 mg of the samples were extracted from the soil samples taken from the field using a FastDNA™ SPIN Kit for Soil (MP bio, US). DNA concentration was checked using a NanoDrop spectrometer (Nanodrop ND-1000, ThermoScientific, Waltham, MA).

The microbial community was assessed by next-generation amplicon sequencing of the 16S rRNA in a paired end mode. DNA extracts were sequenced with an Illumina MiSeq platform (NU-OMICS, Northumbria University, United States) using the primer set F515/R806 targeting 294 bp of the V4 region of the bacterial 16S rRNA gene, as described by Kozich et al. (2013).

7.3.3 Sequence Processing and Analysis

Pair end reads were converted into exact amplicon sequence variant libraries produced using the DADA2 pipeline package with the R program (v1.18.0; Callahan et al., 2016). The trimmed forward and reverse reads were merged with settings −25 M to 230 M. Chimeras were removed with the "removeBimeraDenovo" function in DADA2 under default settings. Taxonomic rank was derived using the "assignTaxonomy" and functional groups were identified using the MIDAS database v 4.8.1 (Dueholm et al., 2021) assigned with the Usearch software package v11 (Edgar, 2013).

Alpha diversity (richness and evenness within the samples) was assessed by computing the total number of Operational Taxonomic Units (OTUs), abundance-based coverage estimator (Chao1) (Chao and Lee, 1992), and Shannon diversity (Shannon, 1948) for all 420 samples. Principal Coordinate analysis (PCoA) plots were estimated using Bray-Curtis and weighted Unifrac metrics, which were derived from the rarefied OTU table using the ampvis2 v 2.7.11 and phyloseq v 3.6 packages, respectively. Further analysis was performed to determine the categorical variables of statistical significance to determine the variation within the microbial communities. Permutational analysis of variance applying distance matrices (ADONIS) using the vegan package v 2.4.2 (Oksanen et al., 2021) was performed to examine several variables based on 2,000 permutations. All analyses and plots were performed on R version 4.1.1 through the Rstudio IDE (R Core Team, 2014).

7.3.4 Site Instrumentation

All three research sites had automated weather stations (Campbell Scientific, United Kingdom) installed which measured air temperature, relative humidity, atmospheric pressure, net radiation, wind speed and direction, and rainfall. Hydraulic loadings were determined using calibrated tipping buckets (described previously). A network of suction lysimeters (Model 1900, Soilmoisture Equipment Corp., United States) was installed at each site spaced longitudinally along each trench and at three different depths beneath the infiltrative surface down to 50 cm depth in the soil. For sample collection, a suction of 50 kPa was applied using a vacuum-pressure hand pump; samples were then collected 24 h later. The effluent and porewater samples extracted from lysimeters were stored on ice for <6 h of transport to be analysed in the environmental engineering laboratory at Trinity College Dublin. The organic loadings of the PE and SE fed into the STUs were determined as total organic carbon (TOC) was determined using a Shimadzu TOC-V analyser (Shimadzu Scientific Instrument, United States). Nitrogen species were analysed as nitrate-nitrogen (NO_3-N), nitrite-N (NO_2-N), and ammonium–N (NH_4-N) and phosphorus as ortho-phosphate (PO_4-P) using a Konelab 20i chemistry analyser (Thermo Scientific, Finland). The spatial distribution of the volumetric water content (VWC) and long-term changes in water retention within the STUs was monitored using a network of 80, 92 and 64 soil moisture sensors (EC5, Decagon Devices, United States), which had been installed during the construction of Sites A, B, and C respectively. Sensors were installed by auguring a 10-cm diameter hole to a desired depth, with sensors positioned into undisturbed subsoil at a required depth below the STU. Control sensors were installed outside of the STU area at the corresponding depths to those below the STU. Calibrations were performed according to the manufacturer's instructions using site-specific subsoils retrieved from test holes excavated prior to the construction of the sites. All sensor data were collected every hour and stored on a CR1000 data logger with two AM16/32 multiplexers (Campbell Scientific, United Kingdom).

7.3.5 Data Availability Statement

Raw sequencing data will be deposited at the National Center for Biotechnology Information (NCBI).

7.4 Results

7.4.1 Meteorological conditions

The mean annual temperature for Republic of Ireland is 9.8°C with mean air temperature ranges from approximately 8.5°C to 10.8°C. Summer account for highs with mean summer mean temperature 14.6 °C and Winters account for mean lows of 5.4 °C. The national annual rainfall is 1288 mm per year. The highest rainfall is found in the south-west (site A and B) of the country at 2,044 mm per year with the lowest in the south-east at 878 mm per year (site C) (Curley et al.,2023).

During the month of sampling in 2018 at site A and B monthly precipitation was 41 mm 18 mm and approximately 38 mm in 2022. In 2018 the mean temperature was (max: 30.5 ^{O}C) 10.2 ^{O}C (min: -7.2 ^{O}C), in 2022 this rose to (max: 31.5 ^{O}C) 10.7 (min: -8 ^{O}C). It is important to note that when sampling in 2018 Ireland was affected by a drought period resulting in no effective rainfall and severe soil drying (Met Éireann, 2018), (**Figure SD1 Figure SD2**) For site C the meteorological conditions for each of the sampling days are summated in **Table 7.1. (Figure SD3)**

Table 7.1 summating meteorological conditions of precipitation, evapotranspiration, temperature and subsoil temperature (depth 100 cm)

Sampling Date	**20-09-2020**	**23-04-2021**	**20-01-2022**	**24-04-2022**	**23-09-2022**	**21-12-2022**	**21-02-2023**	**25-05-2023**
Days in Operation	**60**	**275**	**547**	**641**	**793**	**882**	**944**	**1037**
Season	**Autumn**	**Spring**	**Winter**	**Spring**	**Autumn**	**Winter**	**Winter**	**Spring**
Mean Monthly precipitation (mm)	71*	26*	41*	12.8	76.6	181	25	58
Monthly Evapotranspiration (mm)	54*	70*	10*	64	50	4	27	93
Mean daily temperature (^{O}C)	13.2*	7.2*	5.5	9.6	13	6.4	9	14
Mean daily subsoil temperature (^{O}C)	14.3	8.6	7.7	8.8	14	7.6	8.2	11

*Values approximated from MET Eireann Oak Park weather monitoring system (Carlow, Ireland)
** approximated from existing site averages

7.4.2 Effluent Quality and Wastewater Treatment System performance

The average quality of effluent from the septic tanks and packaged treatment plants that were feeding the percolation trenches is shown in **Table 7.2**. The level of pre-treatment varied according to the different packaged treatment systems used at each site with the coco-media filter on Site A only partially nitrifying and removing just 60% of organics, followed by the sand filter on Site C with greater levels of organic removal at 80%, up to the Rotating Biodisc Contactor on Site B achieving >90% organic removal. The mean effluent hydraulic loadings across the sampling periods were: Site A = 269.8 L/d, Site B = 500.1 L/d and-Site C = 129 L/d.

Site C was monitored intensely to observe concentrations in the porewater concentrations of effluent contaminants within the subsoil (**Figure SD10; Figure SD11)**. Variations in contaminants were aggregated by season to analyse the level of variation across the years. In the PE fed trench changes to contaminant concentrations were most noted in proximal and midpoint of the trench at depths of up to 50 cm below the infiltrative surface, while in the SE trench variation in contaminant concentrations was high across the length of the trench. Shifts in total organic carbon (TOC) concentrations were large between the years, in 2021 the highest concentration in organics were located at the proximal position of the PE trench (**Table 7.3**), however in 2023 it dropped by ~80%. The end (distal) position of the PE trench exhibited a large variation in organic concentrations with TOC tripling between 2021 and 2023. Total nitrogen (TN) concentrations stayed relatively steady across the three of years of monitoring with the greatest concentrations found at the proximal positions. In the midpoint there was a 46% drop from 2021 to 2023. There was also a notable increase in TN at the end zone of the PE trench of 63% increase over the monitoring period.

Concentrations in ammonium dropped in both the proximal and the midpoint zone of the PE trench by 99% in 2023 when compared by 2021. There was an 83% increase NH_4-N concentration in the PE STU end zone between 2021 and 2023 campaign. For NO_3-N the proximal position peaked in concentration in 2022, but then 2023 return similar concentrations to 2021. At the midpoint of the trench NO_3-N concentrations dropped by 72% between 2021 and 2023, while in the end position of the trench saw an increase of 57%.

Table 7.2. Effluent characteristics for primary effluent (PE) from the septic tanks and secondary effluent (SE) from the packaged treatment units on Site A (2015 – 2018), Site B (2016-2023) and Site C (2020-2023).

Site	Parameter	PE		SE		Mean removal efficiency
		concentration (mean ± SD)	load (mean ± SD)	concentration (mean ± SD)	load (mean ± SD)	
Site A	COD	605.8 ± 240.6 mg L-1	163 ± 64.7 g/d	220.5 ± 116.4 mg L-1	59.3 ± 31.3 g/d	0.636
	TOC	162.5 ± 82.7 mg L-1	43.7 ± 22.2 g/d	63.6 ± 42.3 mg L-1	17.1 ± 11.4 g/d	0.609
	TN	167.8 ± 69.0 mg L-1	45.1 ± 18.6 g/d	115.7 ± 44.9 mg L-1	31.1 ± 12.1 g/d	0.31
	NH4-N	42.7 ± 54.5 mg L-1	11.5 ± 14.7 g/d	13.6 ± 18.2 mg L-1	3.7 ± 4.9 g/d	0.681
	NO3-N	1.8 ± 2.5 mg L-1	0.5 ± 0.7 g/d	32.9 ± 18.0 mg L-1	8.9 ± 4.8 g/d	–
	PO4-P	8.8 ± 6.4 mg L-	2.4 ± 1.7 g/d	8.6 ± 5.8 mg L-1	2.3 ± 1.6 g/d	0.023
	Total coliforms	3.45 x 10^6 MPN/100mL		1.11 x 10^6 MPN/100mL		0.49 log10
	E. coli	1.35 x 10^5 MPN/100mL		8.60 x 10^4 MPN/100mL		0.20 log10
Site B	COD	1005.4 ± 192.7 mg L-1	251.35 ± 48.2 g/d	51.6 ± 43.5 mg L-1	25.8 ± 21.8 g/d	0.949
	TOC	297.61 ± 99.8 mg L-1	74.4025 ± 24.9425 g/d	33.02 ± 24.18 mg L-1	8.3 ± 6.1 g/d	0.90
	TN	116.1 ± 43.3 mg L-1	29.035 ± 10.825 g/d	18.23 ± 8.03 mg L-1	4.6 ± 2.01 g/d	0.85
	NH4-N	116.1 ± 37.6 mg L-1	29.03 ± 9.39 g/d	0.74 ± 1.11 mg L-1	0.19 ± 0.28 g/d	0.99
	NO3-N	0.8 ± 3.9 mg L-1	0.20475 ± 0.975 g/d	10.59 ± 2.64 mg L-1	2.7 ± 0.7 g/d	–
	PO4-P	11.5 ± 4.4 mg L-1	2.8625 ± 1.0975 g/d	11.41 ± 3.74 mg L-1	2.9 ± 0.9 g/d	0.34
	Total coliforms	8.06 x 10^6 MPN/100mL		1.44 x 10^4 MPN/100mL		2.76 log10
	E. coli	4.73 x 10^5 MPN/100mL		6.23 x 10 MPN/100mL		3.88 log10
Site C	TOC	179 ± 107 mg L-1	14.2842 ± 8.5386 g/d	35.7 ± 46.1 mg L-1	1.75644 ± 2.26812 g/d	0.8
	TN	96.5 ± 114.1 mg L-1	7.7007 ± 9.10518 g/d	42.1 ± 87.5 mg L-1	2.07132 ± 4.305 g/d	0.6
	NH4-N	100.7 ± 42.6 mg L-1	8.03586 ± 3.39948 g/d	11.2 ± 16.3 mg L-1	0.55104 ± 0.80196 g/d	0.89
	NO3-N	0.4 ± 1.6 mg L-1	0.03192 ± 0.12768 g/d	6.4 ± 6.3 mg L-1	0.31488 ± 0.30996 g/d	

Site	Parameter	PE		SE		Mean removal efficiency
		concentration (mean ± SD)	load (mean ± SD)	concentration (mean ± SD)	load (mean ± SD)	
	PO4-P	10.5 ± 8.1 mg L-1	0.8379 ± 0.64638 g/d	2.2 ± 2.5 mg L-1	0.10824 ± 0.123 g/d	0.8
	Total coliforms	8.78 x 10^6 MPN/100mL		1.42 x 10^4 MPN/100mL		1.79 log10
	E. coli	4.27 x 10^5 MPN/100mL		6.44 x 10^4 MPN/100mL		0.82 log10

PO_4-P concentrations remained relatively similar throughout sampling at the proximal position, in the midpoint we see a 140% relative increase in PO_4-P concentration in 2022. However, PO_4-P dropped in concentration across all position in 2023 averaging at 0.4 ± 0.5 mg/L.

Table 7.3. Average porewater concentrations across the trench lateral profile (mean ± sd) in the three years of monitoring of site C primary effluent (PE) and secondary effluent (SE) at each of the positions in the STU base: (p) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end position 18 m (e).

		TOC (mg/L)	TN (mg/L)	NH4-N (mg/L)	NO3-N (mg/L)	PO4-P (mg/L)
PE-TB (p)	**2021**	50.5 ± 40.4	40.0 ± 37.9	44.0 ± 24.8	15.1 ± 26.3	1.4 ± 2.4
	2022	19.1 ± 37.2	33.9 ± 32.8	13.8 ± 15.5	22.6 ± 25.9	1.5 ± 2.1
	2023	9.7 ± 5.0	37.5 ± 26.4	0.5 ± 1.3	13.0 ± 4.6	0.6 ± 0.6
PE-TB (m)	**2021**	23.8 ± 19.4	26.5 ± 44.1	15.3 ± 19.3	26.1 ± 30.8	0.7 ± 1.7
	2022	12.9 ± 11.5	22.5 ± 30.1	9.1 ± 13.8	15.9 ± 26.3	1.7 ± 2.7
	2023	9.4 ± 9.2	13.8 ± 16.5	0.04 ± 0.1	7.3 ± 4.0	0.2 ± 0.3
PE-TB (e)	**2021**	10.0 ± 5.9	6.9 ± 6.9	0.1 ± 0.2	6.8 ± 4.2	0.7 ± 1.7
	2022	11.7 ± 11.2	6.9 ± 4.5	0.1 ± 0.1	4.2 ± 4.0	1.7 ± 2.7
	2023	28.6 ± 25.9	11.3 ± 7.4	0.6 ± 0.6	11.7 ± 5.9	0.2 ± 0.3
SE-TB (p)	**2021**	17.6 ± 17.5	5.2 ± 2.5	0.3 ± 0.7	7.4 ± 7.4	0.1 ± 0.1
	2022	12.2 ± 9.5	15.5 ± 27.7	0.5 ± 1.2	9.5 ± 15.6	0.1 ± 0.2
	2023	11.1± 10.6	35.9 ± 29.5	2.1 ± 3.6	33.6 ± 31.0	0.5 ± 0.6
SE-TB (m)	**2021**	12.9 ± 8.1	4.4 ± 3.8	0.2 ± 0.8	3.9 ± 2.5	0.2 ± 0.2
	2022	10.7 ± 11.9	10.1 ± 10.4	1.2 ± 3.5	5.9 ± 5.6	0.1 ± 0.2
	2023	7.0 ± 5.6	28.9 ± 20.4	1.1± 3.2	18.7 ± 15.6	0.3 ± 0.4
SE-TB (e)	**2021**	10.6 ± 10.5	2.0 ± 3.1	0.1 ± 0.2	3.9 ± 3.6	0.2 ± 0.3
	2022	16.7 ± 22.7	4.3 ± 3.7	3.4 ± 3.8	2.8 ± 1.2	0.6 ± 0.7
	2023	20.7 ± 6.8	8.0 ± 3.0	0.0 ± 0.0	4.8± 1.5	0.0 ± 0.0

TOC concentrations in the SE trench appeared to follow a similar pattern to those in the PE trench, with decreases in the proximal (37%) and midpoint (46%) and increases at the end position (49%). Total nitrogen increased in all positions by 4 to 6-fold between 2021 to 2023. Ammonium concentrations increased by 6 and 4.5 times for the proximal and midpoint zones respectively. There was a 330% increase of the ammonium concentrations between 2021 to 2022 but dropped to 0 mg / L by 2023. The increase in nitrates was far greater over the monitoring time in the proximal (3.75-fold) and midpoint (3.9-fold), but the end position only marginally increased by 23%. Phosphorus showed the greatest increase at the proximal position by 80%, followed by the midpoint position 33% and the end of the trench showed a large increase between 2021-2022 (66%) but like NH_4-N dropped to lower than the limit of detection in spring 2023.

7.4.3 Temporal changes to soil treatment system community composition

The soil samples were subdivided as follows. For Sites A and B, the 271 samples taken in 2018 and 2022 had a total of 10,071,026, sequence reads, ranging from min (8605), median (58289) and max (220139) per sample: these samples were the rarefied to their median values. For the 145 samples collected from Site C there was a total of 2,822,135 sequence reads, ranging from min (14), median (41149) and max (246657) per sample, with the samples rarefied to the median value of 41149.

Monitoring of site C provides insights of changes occurring from the early inception of biomat. In Site C the most abundant and variable phylas within the control subsoil throughout the 1037 days in operation were *Proterobacteria Actinobacteriota* and *Acidobacteriota* with relative read abundances of (mean ± sd) 16.79 ± 12.11%, 9.67 ± 15.17% and 11.31 ± 6.98% respectively. The proximal position of this STU revealed that *Proteobacteria, Actinobacteriota* and *Firmicutes* exhibited the greatest level of variation. Notably in day 547 sampled in the winter Proteobacteria had its lowest relative read abundance at 30.2 ± 14%, whilst both *Firmicute*s and *Actinobacteria* exhibiting their highest relative read abundance of 18 ± 1.8% and 18 ±8.7% respectively. Overall, the midpoint of the PE fed STU was the most variable location at the phyla level. At the midpoint position for the PE fed STU *Proteobacteria*, *Firmicutes* and *Desulfobacterota* showed the most variation across the days in operation of the system. *Proteobacteria* relative read abundance peaked at 46.7 ± 14.2% at 793 days from an initial abundance of 11.29 ± 4.5 %. *Firmicutes* at the midpoint position began high initially at 23.88 ± 10.19% but steadily declined over time to a relative read abundance of 5.02 ± 4.76 %. At the end point of the PE fed trench *Proteobacteria* and *Firmicutes* were again the most variable of the phyla - the *Proteobacteria* relative read abundance started at 4.25 ± 0.4% and peaked at 944 days of operations at 24 ± 17.1%. At the end position of the trench the relative read abundance of *Firmicutes* began low at 1.01 ± 0.31% but increased significantly by 275 days operation up to 10.67 ± 8.11%, then remaining at an average abundance of 8.54 ± 5.6% for the remainder of the monitoring campaign. *Gammaproteobacteria* were the most dominant phylogenetic class at the proximal (50.61 ± 11.62 %) and midpoint (34.21 ± 12.32 %) position of the PE trench, with abundances peaking after 1037 days and 793 days, respectively (**Figure SD4**). Following 275 days in operation *Alphaproteobacteria* and *Bacilli* were the most dominant classes at the end of the trench. Average *Gammaproteobacteria* abundances taken from control subsoil (10.79 ± 8.42 %) were exceeded by all positions of the PE trench within 60 (proximal), 275 (midpoint) and 944 (end) days.

The SE fed trench of Site C found that the proximal and midpoint sampling positions of the trench were the most variable within the STU. In the proximal position the most variable phyla were *Proteobacteria* and *Acidobacteria*. Here *Proteobacteria* remained relatively low at 9.4 ± 3.44% until day 641 when there was a notably rise which peaked on day 944 of operation at a relative read abundance of 47 ± 4.24. *Acidobacteria* appeared to be susceptible to seasonal patterns with peaks occurring in spring sampling campaigns and lowest abundances noted during winter sampling. At the midpoint of the SE fed trench, *Proteobacteria* and *Firmicutes* were the most variable phyla

present. *Proteobacteria* began to increase from an average relative abundance of 9.59 ± 2.38% in day 547 to 16.17 ± 0 % and finally peaking at 944 days of operation at 39.33 ± 4.68 %. The variability of *Firmicutes* can be attributed to a large peak after 547 days of system operation at a relative abundance of 24.09 % no other peak exceeded an abundance of 7.43 ± 1.93%. At the end of the SE fed STU, *Proteobacteria*, *Firmicutes* and *Bacteroidota* were the most variable phyla. *Proteobacteria* increased from a relative read abundance of 6.14 ± 0.23% to an average of 18.15 ± 12.78%, peaking at 35.95 ± 27.01% after 944 days of operation. Firmicutes and *Bacteroidota* increased from an average relative read abundance of 1.15 ± 0% and 0.45 ± 0.26% to an average of 6.47 ± 5.89% and 2.86 ± 5.83% respectively. At the end position of the SE fed trench, *Firmicutes* and *Bacteroidota* peaked at 275 (14.75 ± 7.55% and 9.51 ± 14.9%, respectively). At the class level *Nitrososphaeria*, *Alphaproteobacteria*, *Bacilli*, *Vicinibacteria* and *Verrucomicrobiae* appear to be dominant in the SE-STU, however the ecology shifts by 882 days. This shift is characterised by a large, sustained increase in abundance of *Gammaproteobacteria* which exceed average abundances found in control subsoil samples, achieved after 882 days in the proximal and midpoint position, later the threshold. Although a large increase of *Gammaproteobacteria* is achieved at the end the trench after 944 days this increase is not sustained falling under the control subsoil average (**Figure SD5**). The variable and dominant species within both trenches for *Proteobacteria* consisted of species from the *Zoogloea sp*, *Rhodanobacter sp*, and *Polarmonas sp*, for *Firmicutes* the species *Bacillales sp*, *Sporosarcina sp*, *Bacillus sp*; from the *Actinobacteriota* family *Micrococaceae sp*; and from *Acidobateriota* genus the species *Geothrix sp* (**Figure SD6**).

The most abundant phyla in 2018 at 1080 days in operation at Site A control subsoil were *Acidobacteriota*, *Proteobacteria*, *Actinobacteriota*, *Chlorflexi* and *Planctomycetota*, with *Verrucomircobiota* replacing *Actinobacteriota* in 2022 after 2540 days in operation. In Site A the relative abundances of *Acidobacteriota*, *Proteobacteria*, *Actinobacteriota*, *Chlorflexi*, *Planctomycetota* and *Verrucomircobiota* changed in relative read abundance in 2018 (1080 days in operation) to 2022 (2540 days in operation) by + 2%, -3.8%, -6.1%, +1.3%, -1% and -1.5% respectively. In Site B in 2018 (860 days in operation) and 2022 (2320 days in operation) *Proteobacteria*, *Acidobacteria*, *Actinobacteriota*, *Verrucomicrobiota* and *Chloroflexi* were the most common phyla in control subsoils, the phyla relative read abundance changed by +4.6%, -11.6%, -1.1%, +2.2% respectively. Over the space of four years there was no notable shift in phyla composition in both sites.

In Site A much of the variation was noted in the proximal position of the trench of the PE dosed STU. The most variable phyla across the system were *Actinobacteria* and *Proteobacteria*. At the class level *Gammaproteobacteria* dominated the proximal position of the trench in 2018, but the remaining positions in the trench were dominated by the *Vicinamibacteria*, in 2022, there was a shift in 2022 with *Gammaproteobacteria* dominating the midpoint of the trench (**Figure SD7**). *Proteobacteria* was dominant at the species level variation, with members of the family

Rhodocyclaceae and genus *Dechloromonas* increasing in abundance when sampled at 2540 days of operation (**Figure 7.3**). Within the SE dosed STU at Site A the most variable location was the midpoint of the trench. *Actinobacteriota* and *Firmicutes* were the most variable phyla across the length of the trench. At the class level *Gammaproteobacteria* was dominant at the proximal position of the trench in both 2018 (1080 days in operation) and 2022 (2540 days in operation) with its abundance increasing at 2540 days of operation, in the remaining positions of the trench *Vicinamibacteria* remained dominant and increased its abundance when sampled in 2022. Firmicute species *Bacillales* within the SE dosed STU at Site A after 2540 days in operation appeared to decrease in relative read abundance across the length of the trench when compared to 1080 days.

Site B showed far greater levels of variability in both PE and SE trenches when compared to Site A, but the SE trench showed greatest levels of variability at the phyla level. In the PE dosed STU in Site B, the most variable position was as the proximal zone. The most variable phyla were *Synergistota* and *Proteobacteria* across the length of trench. At the class level *Viciniamibacteria* and *Planctomycetes* were the most dominant after 860 days in operation in PE trench, however there was a large increase in *Gammaproteobacteria* and *Synergistia* both dominating the STU at 2320 days of operation (**Figure SD8**). At the species level, there were notable increases in the abundance of three species of the *Synergistaceae* family are attributable to increases at 2320 days (**Figure 7.4**). Variation of *Proteobacteria* species, increases in species from the *Competibacteraceae* family, *Rhizobiales* order were attributable to these shifts. In the SE fed trench of the STU for Site B the greatest variations were found in the distal position of the trench. The most variable of the phyla across the trench were *Firmicutes* and *Acidobacteria*. At the class level *Gammaproteobacteria* which only dominated the proximal and midpoint of the trench after 860 days of the system being in operation proceed to dominate the entirety of the trench after 2320 days. The class *Clostridia* dominated the distal position of the SE trench in 2018 (860 days in operation), however by 2022 (2320 days in operation) their abundance drops from 12.4 to 0.1 %. The variability in the *Firmicutes* phyla after 2320 days in operation was attributable to a reduction in the relative read abundance of species belonging to the genus *Romboutsia*, order *Bacillales*, and genus *Clostridium sensu stricto*. The *Proteobacteria* phyla in the SE fed STU at Site B showed a large reduction in species such as members of the order *Rhizobiales* and family *Comamonadacaeae* (**Figure 7.4**).

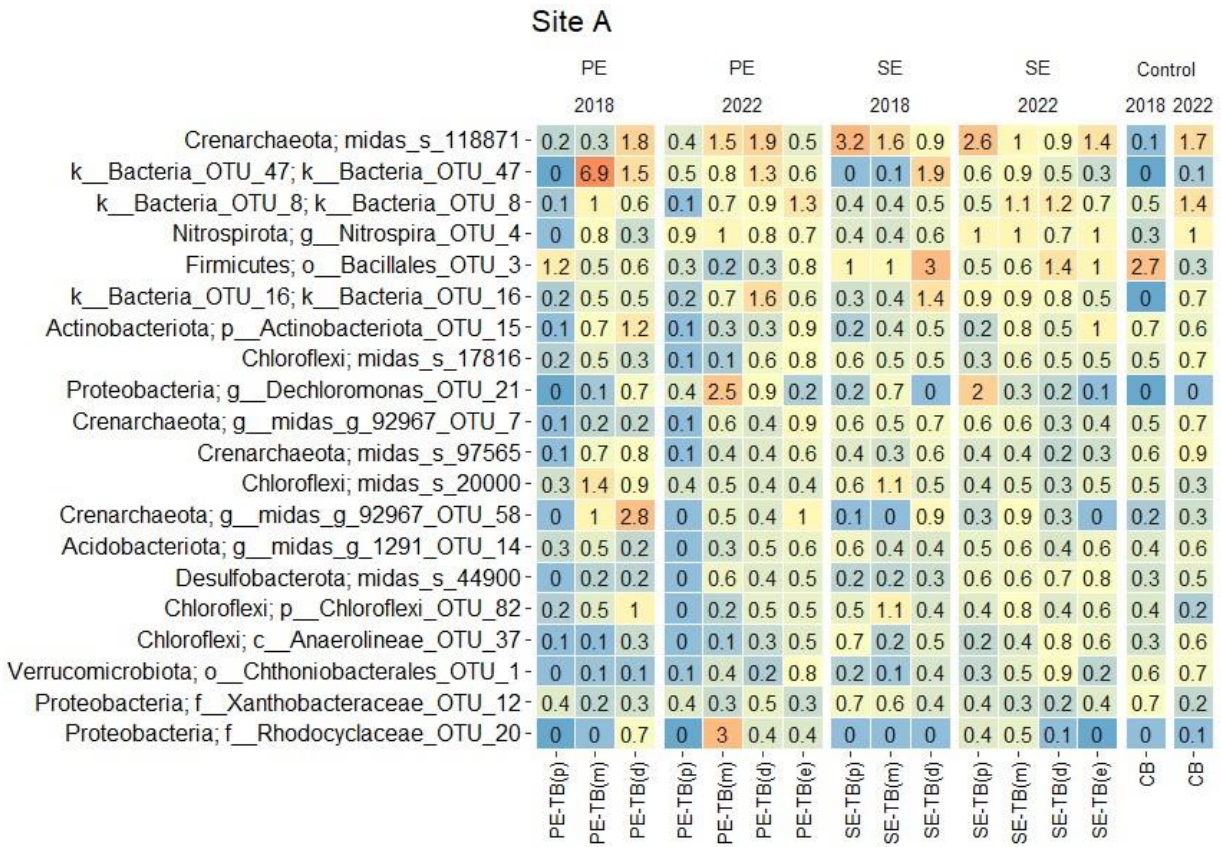


Figure 7.3 Mean of relative read abundance of top 20 genus for Site A both 2018 (1080 days in operation) and 2022 (2540 days in operation) within the Primary (PE) secondary (SE) STU, control subsoil samples; base (CB), and STU subsoil "trench base" (TB). STU base is further divided into proximal (p) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end position 18 m (e).

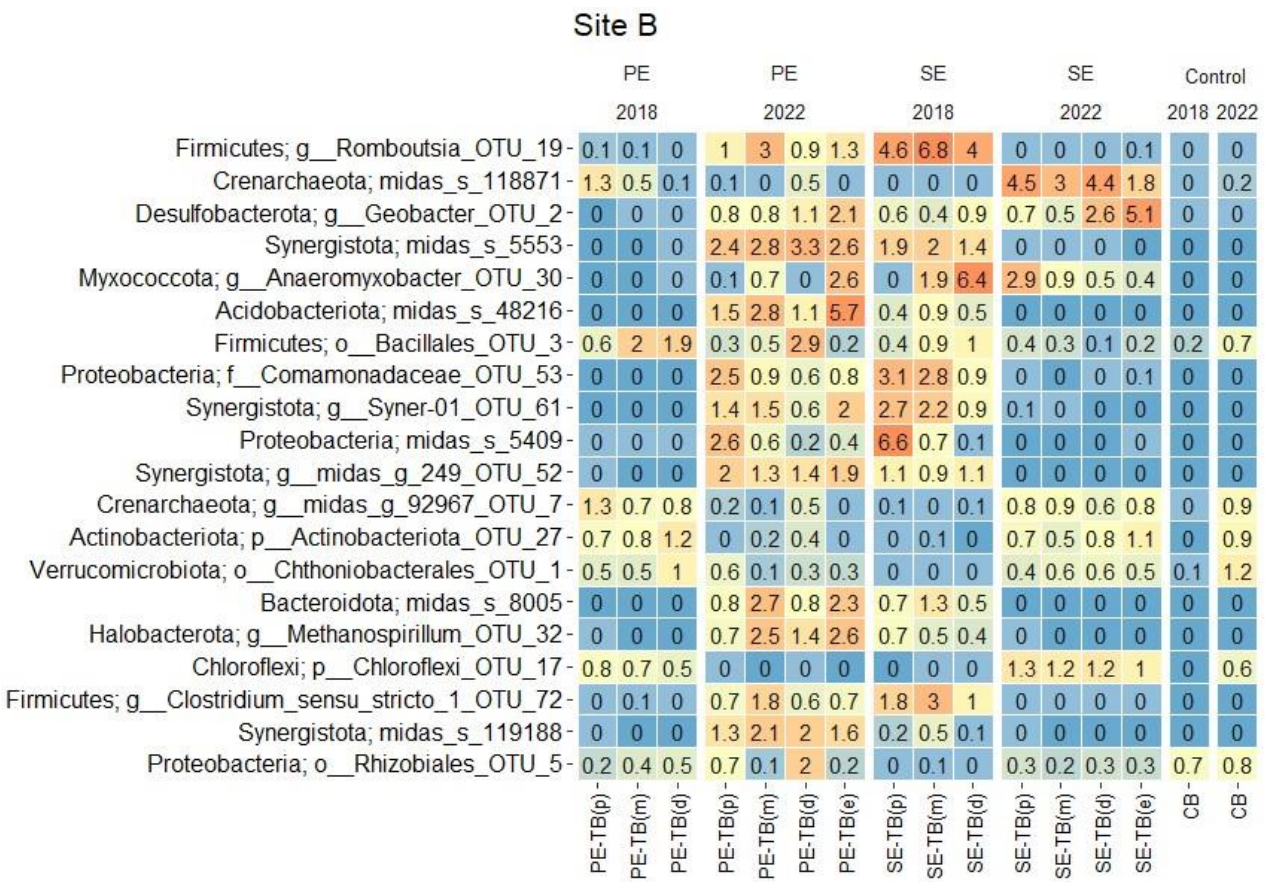


Figure 7.4 Mean of relative read abundance of top 20 analysis for Site B both 2018 (860 days in operation) and 2022 (2320 days in operation) within the Primary (PE) secondary (SE) STU, control subsoil samples; base (CB), and STU subsoil "trench base" (TB). STU base is further divided into proximal (p) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end position 18 m (e).

7.4.4 Temporal variation in microbial community structure in STU

Each site was analysed temporally for alpha diversity (Shannon) and species richness (Chao1) for each of the sampling positions. In site C which represents the initial stages of biomat development there were notable shifts in species richness and diversity over the monitored operation of the system (**Figure 7.5**). Positions within the trench were compared between PE and SE fed trenches as well as controls subsoils for each year. There was only one significant difference in alpha diversity between PE and SE trenches at the proximal sampling position in the year 2022 (P = 0.02). In the PE trench proximal zone there was an average drop in diversity per year. The was an increase in richness in the PE trench until day 882 of operation (winter 2022) with no successive increase in richness until the end of the experiment.

Diversity at the PE-TB proximal position was lowest at the infiltrative zone (0 cm), when sampled in days 547 (winter 2022), 793 (autumn 2022), and 944 (winter 2022). When comparing annual changes in diversity, PE-TB(p) samples collected in only in winter did show a significant positive correlation with diversity over time (R^2= 0.5, P = 0.05). The midpoint of the PE trench also did show a significant correlation with diversity increasing over time (R^2= 0.7, P = 0.05), with winter being the only season where a significant correlation was noted. There was a great deal of variability at the end position of the PE fed STU between the days in operation. The end position of the PE-TB showed significant positive correlation diversity (R^2 = 0.34, P = 0.01) and richness (R^2 = 0.31, P = 0.01) over the entire 1037 days of operation. In the SE fed STU, there was only one significant positive correlation with richness increasing over time (R^2 = 0.9, P = 0.01) at the midpoint of the trench.

At the long-term monitoring sites, there was no significant difference in the alpha diversity or species richness between the control subsoil samples in site A. For Site B there was a significant increase in alpha diversity (p = 0.01) for the subsoil samples taken as a control, but the increase in richness was not significant. For Site A there was no significant difference between the control subsoil, the PE or SE fed trenches in either 2018 (1080 days in operation) or 2022 (2540 days in operation). In 2018 Site A showed the SE fed STU was significantly richer and more diverse than PE fed STU (**Table 7.4**), however by 2022 this difference between the two trenches was no longer significant. In Site B the difference between PE and SE fed trenches changed significantly: after 860 days in operation the PE fed STU was significantly richer and more diverse than the SE fed STU and the control subsoil, however, after 2320 days in operation the richness and alpha diversity in the PE fed STU had dropped significantly below that of the SE fed STU and the control subsoil (**Figure SD9)**.

Within both PE and SE trenches of Site A there was no change in the richness and diversity within any of the sampling positions between 1080 and 2540 days in operation. In Site B, however, there was significant difference in the richness and diversity from the sampling positions between 860 and 2320 days in operation. In the PE fed STU there was a significant decrease in richness and diversity,

with the greatest reduction occurring at the midpoint (Shannon: p = 0.05; Chao1: p = 0.05) followed by the distal position of the trench (Shannon: p = 0.01; Chao1: p = 0.01). In the SE fed STU at site B, the species richness and alpha diversity significantly increased at the proximal sampling zone (Shannon: p = 0.05; Chao1: p = 0.05).

Table 7.4. Wilcoxon test values for comparative intra-site analysis.; ns P > 0.05, * P ≤ 0.05,**P ≤ 0.01, ***P ≤ 0.001 and **** P ≤ 0.0001. Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), STU base (TB), and control base (CB).

	PE-TB VS SE-TB				**PE-TB VS CB**				**SE-TB vs CB**			
	Shannon		**Chao1**		**Shannon**		**Chao1**		**Shannon**		**Chao1**	
	2018	**2022**	**2018**	**2022**	**2018**	**2022**	**2018**	**2022**	**2018**	**2022**	**2018**	**2022**
Site A	*	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS
Site B	****	****	***	****	*	****	NS	****	*	NS	NS	NS

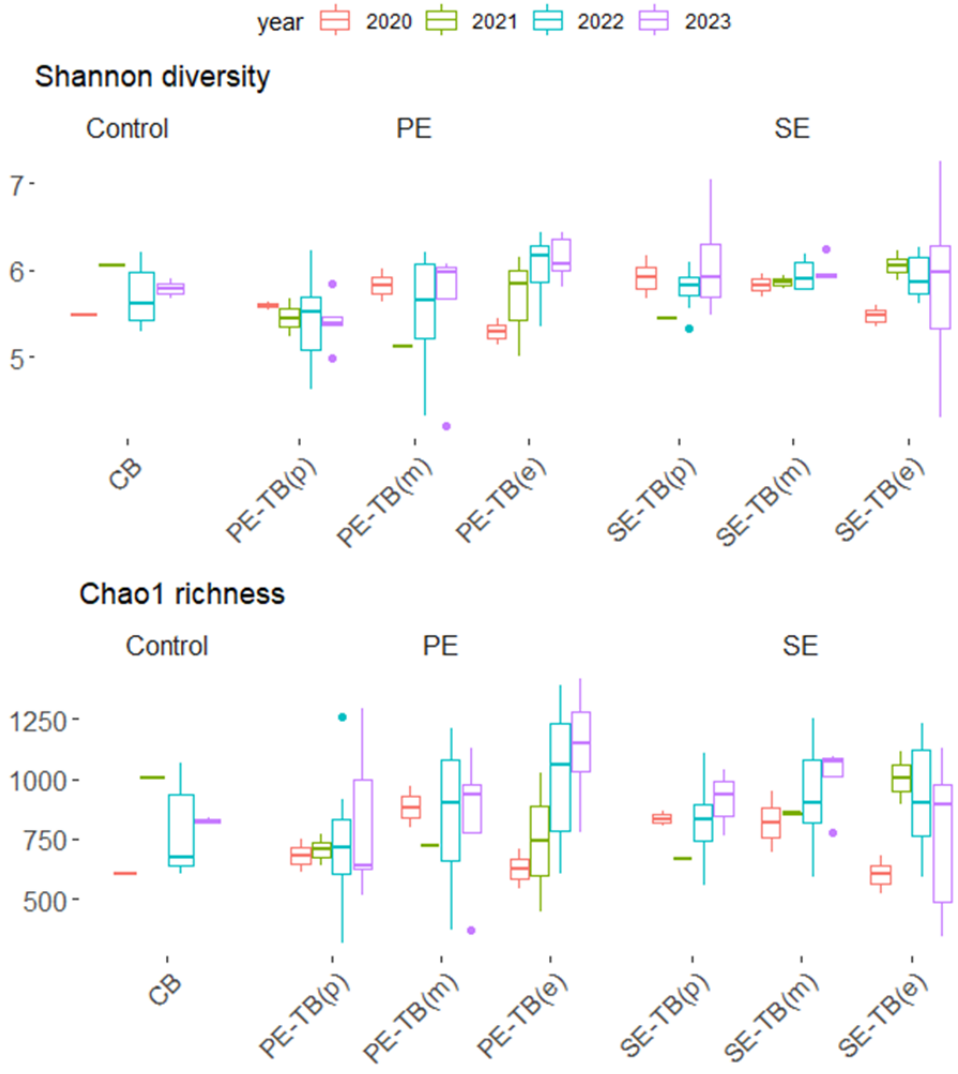


Figure 7.5. Rarefied data for observed species richness calculated using an abundance-based coverage estimates (Chao1) and alpha diversity (Shannon) for years 2020 to 2023 at site C. Samples were aggregated on the bases of positions proximal (1 m), midpoint (7.5 m) and end point (17 m) in trenches fed by primary (PE) and secondary (SE) effluent and base (CB). STU base is further divided into proximal (p) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end position 18 m (e).

Permutational analysis of variance applying distance matrices indicated that only community structures which at site C representing the initial development of the biomat. It was not until day 547 that there was a significant shift in the overall PE trench community structure due to the position

in the biomat (R^2 = 0.65, P = 0.05), which was also observed on day 793 (R^2 = 0.45, P = 0.01), 882 (R^2 = 0.41, P = 0.05), 944 (R^2 = 0.63, P = 0.003) of system operation. (**Figure 7.6**). For the SE-fed STU this shift was not observed until 882 days in operation (R^2 = 0.45, P = 0.05) and again after 944 days (R^2 = 0.37, P = 0.01). The level of effluent pre-treatment showed to have a significant effect on community structure on day 60 (R^2 = 0.18, P = 0.05), 547 (R^2 = 0.27, P = 0.001), day 793 (R^2 = 0.24, P = 0.001), day 882 (R^2 = 0.17, P = 0.05), and day 944 (R2 = 0.22, P = 0.001).

In the sites measured between 3 to 7 years in operation there was no significant difference between the control subsoil communities at either site between the years 2018 and 2022. Time had a greater overall effect on variation in communities of both PE and SE fed trenches from Site B (R^2 = 0.44, P = 0.001; R^2 = 0.39, P = 0.001, respectively), than the PE and SE trenches of Site A (R^2 = 0.10, P = 0.002; R^2 = 0.12, P = 0.001, respectively) (**Figure 7.7**)..Variation between the PE and SE fed trenches reduced in four years between sampling campaigns at both sites, but remained relatively high in Site B (**Table 7.5**). The spatial variation in communities increased in the PE fed trenches at both sites but did drop in the SE fed STUs, with Site B exhibiting a greater drop in variation. Depth continued to have no significant effect on community structure.

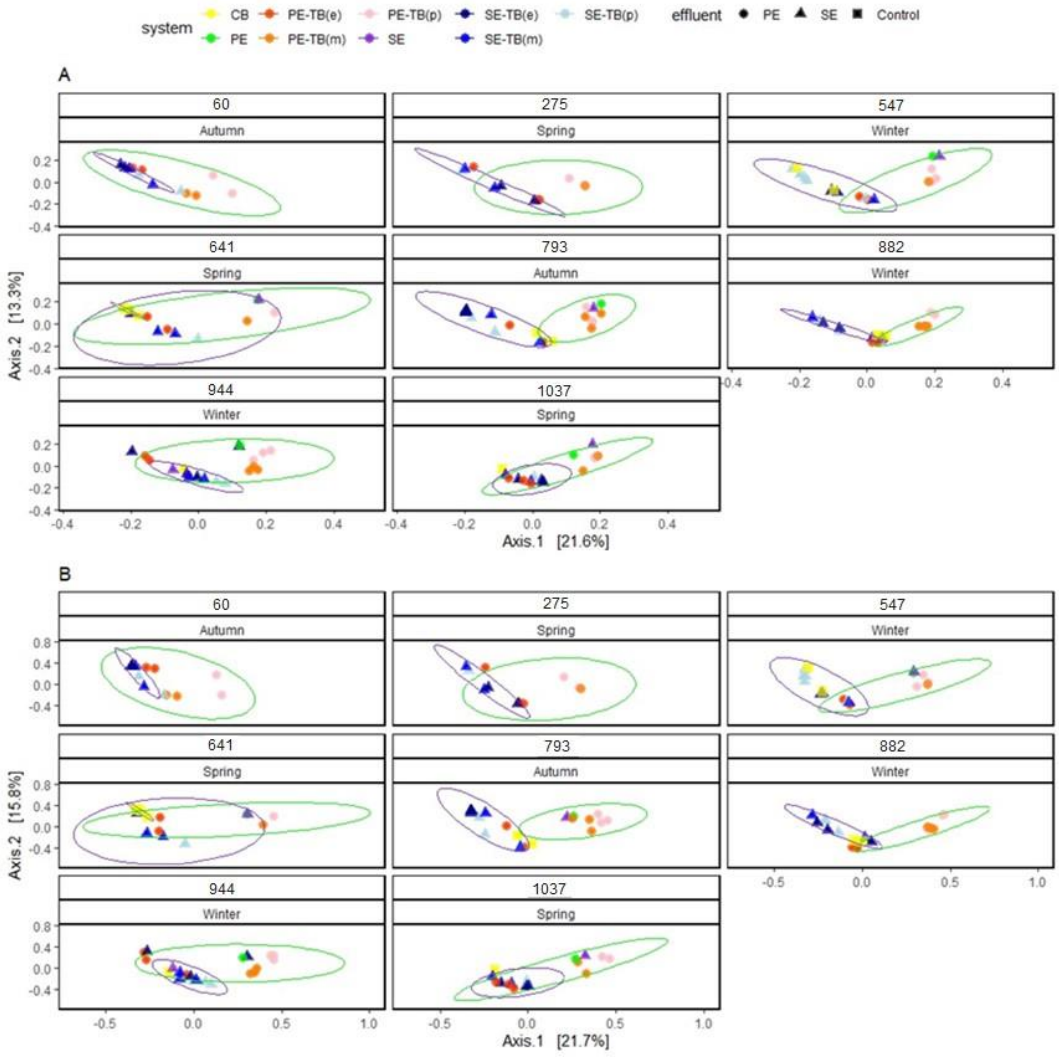


Figure 7.6, Principal coordinates of beta diversity based on weighted Unifrac (A) and Bray (B) distance matrix within STUs at Site C, over the days in operation and seasons sampled. Each data point represents a sample taken from either PE effluent stream, SE effluent stream or control soils. Samples are further subdivided based on the position within the system, i.e., at the STU; base, control; base, top and pure effluent samples.

Table 7.5. Permutational tests with weighted unifrac distance matrices (ns Pr > 0.05, * Pr ≤ 0.05,**Pr ≤ 0.01, ***Pr ≤ 0.001 and **** Pr ≤ 0.0001) values for comparative intra-site permutational analysis of variance applying distance matrices . Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE), STU base (TB), and base (CB) for each year of sampling (2018, 2022).

		PE-TB vs CB		SE-TB vs CB		PE-TB vs SE-TB		PE-TB vs Distance (m)		SE-TB vs Distance (m)	
		Pr	R2	Pr	R2	Pr	R2	Pr	R2	Pr	R2
Site A	**2018**	0.1 NS	0.23	0.1 NS	0.13	0.004 **	0.16	0.0004 ***	0.34	0.001 ***	0.34
	2022	0.0004 ***	0.12	0.0004 ***	0.13	0.003 **	0.04	0.0004 ***	0.41	0.1 NS	0.24
Site B	**2018b**	0.01 **	0.32	0.01 **	0.38	0.004 **	0.49	0.003 **	0.36	0.001 ***	0.65
	2022	0.0005 ***	0.16	0.0005 ***	0.12	0.0005 ***	0.35	0.0005	0.45	0.0005 ***	0.29

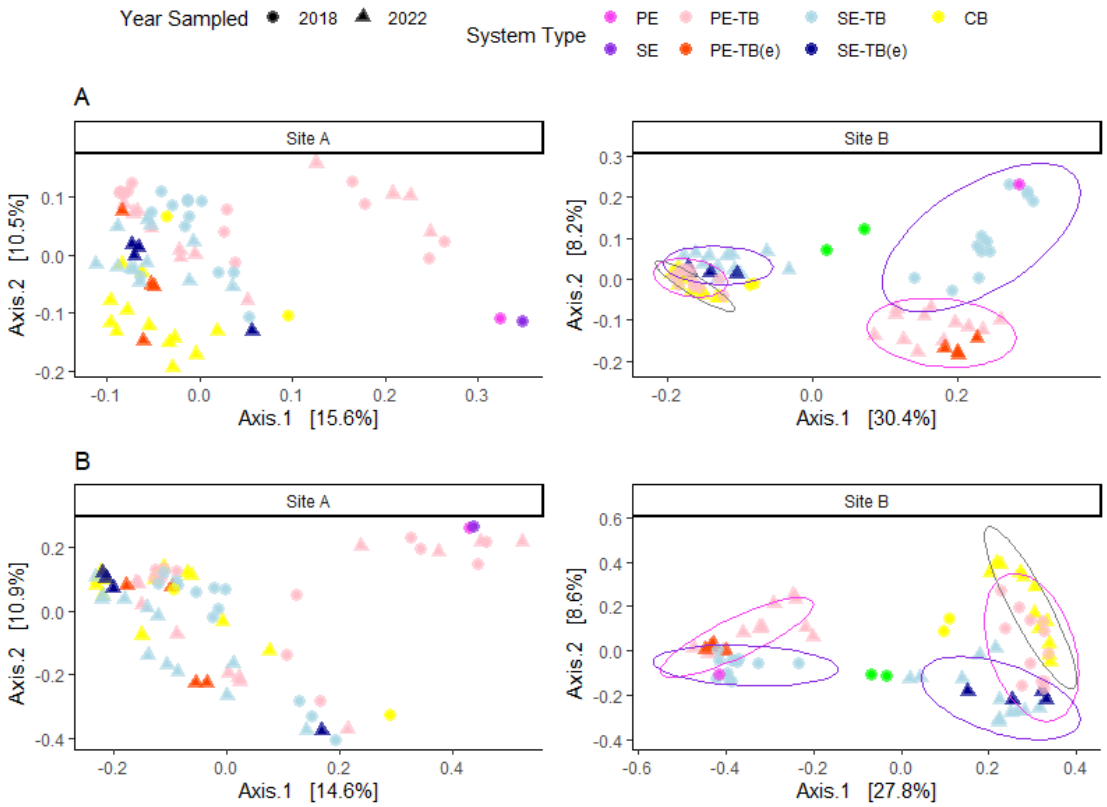


Figure 7.7. Principal coordinates of beta diversity based on weighted Unifrac (A) and Bray (B) distance matrix within STUs at both sites and years sampled. Each data point represents a sample taken from either PE effluent stream, SE effluent stream or control soils. Samples are further subdivided based on the position within the system, i.e., at the STU; base, control; base, and pure effluent samples.

7.4.5 Temporal changes in biogeochemical functional groups

Using the MiDAS database, samples were screened for biogeochemical functional bacteria. At Site C samples were collected over 3 years throughout the operation of the system, monitoring the initial formation of the biomat. Functional groups relative read abundances were monitored across both systems (**Figures 7.8** and **Figure 7.9**). Changes in functional groups were compared with respect to season to determine what changes had taken place within the biomat. Changes in relative read abundances were relativised to seasonal comparisons, to avoid seasonal variation. The greatest change in functional group was noted for denitrifiers, in the proximal zone for the SE trench and the PE trench with an average increase in relative read abundance over time of 6.92 ± 3.96%, and 4.12 ± 2.98%, respectively. Relative read abundances of denitrifiers found in both STUs exceeded average abundances found in the control subsoil (0.12 ± 0.72 %) after only 60 days in operation. The highest relative read abundance were noted at 944 days and 882 days for the proximal zone for the PE fed trench and the midpoint zone (5 m) of the SE fed trench respectively. For AOBs the SE trench saw a greater increase at 1.24 ± 1.06 % (**Figure 7.9)**, followed by the midpoint of the PE trench with an average increase of relative read abundance of 0.69 ± 0.38%. The highest relative read abundance for AOB (1.49 ± 1.22 %) and NOB (0.92 ± 1.04 %) were found at the midpoint zone for the PE trench at 793 days. AOB abundances found in the control subsoil (0.12 ± 0.28 %) were exceeded by 60 days of operation in proximal and midpoint of the SE trench, but this exceedance did not continuously occur in the end position until 793 days of operation. In the SE trench AOB (2.41 ± 0.44 %) and NOB (2.08 ± 0.57 %) were found at their highest relative read abundance at the proximal position at 944 and 1037 days respectively. NOB showed the greatest increase in relative read abundance of 1.07 ± 0.74% at the proximal position in the SE trench, followed by the end sampling position and midpoint position for PE trench with average increases in relative read abundances of 0.31 ± 0.21% and 0.26 ± 0.26 %, respectively. NOB relative read abundances in the STU surpassed average abundance of the control subsoil (0.06 ± 0.12 %) at nearly every position in both STUs by 60 days in operation, this did not occur at the end of the trench until day 793 of operation. The only exception being midpoint of the PE trench which achieved exceeded the control NOB abundance after 793 days of operation.

For methanogens the greatest increase in relative read abundance was noted at the proximal zone of the PE trench with an increase of 0.6 ± 1.82%, the highest level of abundance was found at the same position after 547 days. The PE trench exceeded the abundances of methanogens found in the control subsoil (<0.01%) in the proximal and midpoint position by 60 days of operation. In the remaining STU positions methanogen abundances only surpassed the control subsoils at 793 days of operation for the end position of PE trench, in the SE trench the threshold was reached after 882 days for all the trench positions. Changes in methanogen and methanotroph abundances were less pronounced in the SE trench, with spikes in abundance found in the end of the trench after 275 days and 1037 days for methanogens and methanotrophs respectively. Methanotrophs were found at the

highest abundance and average increase at the proximal zone within the PE trench at 0.16 ± 0.78 %, the peak in abundance was found after 641 day of operation. Methanotrophs only surpassed control subsoil (<0.01%) values at the PE trench proximal and midpoint position after 60 and 793 days of operation, respectively.

PAO saw the greatest average increase in relative read abundance at the proximal point within the PE fed trench at 0.39 ± 0.16%. PE trench PAO abundances were exceeded control abundances (0.01 ± 0.04 %) after 60 days at the proximal and midpoint positions. the remaining positions at both STUs exceeded control PAO abundances after 882 days of operation. PAO peaked in the PE trench at the proximal position after 882 days of system being in operation. In the SE trench PAO had muted increases across the entirety of the trench with average increases of 0.07 ± 0.11%. Spatial and temporal variability in functional groups appears to be affected by the level of pre-treatment of effluent dosed within the PE stream the proximal zone was the position of greatest change for most of the functional groups except for AOB and NOB which showed the greatest variation within the midpoint of the trench. Within the SE trench, the functional groups involved with the cycling of carbon and nutrients showed greatest variability at the proximal position of the trench; for methanogens it was towards the distal and for PAO and Methanotrophs variability was similar across the entirety of the trench.

Changes in relative read abundances of the functional organisms were compared for each of the sampling locations at Sites A and B between the two sampling campaigns in 2018 and 2020, which represent the long-term development of the biomat after approximately 3 and 7 years of the systems being in operation (**Table 7.6** and **Figure 7.10**). For both sites there was mostly an increase in the relative read abundance of ammonia oxidizing bacteria, all STUs continued to exceed control subsoil abundances in both years sampled. In Site A PE trench NOB proximal position only exceeded abundances found in the control subsoil after 2540 days in operation. In Site B there was a notable reduction in NOB groups in trench dosed with PE which no longer exceeded the control subsoil abundance, which exceeded across the entire trench in 2018. The proximal point of the SE trench was the only position in site B exceeded the NOB abundances found in the control subsoil, this was the first occurrence of the SE trench surpassing control abundances.

Table 7.6. Changes in functional organism abundance (%) in site A and B: ammonia oxidizing bacteria (AOB), nitrite oxidizing bacteria (NOB), denitrifiers, anammox, methanogens, methanotroph and phosphate accumulating organisms (PAO) between August 2018 and August 2022 in the proximal (1 m), midpoint (7.5 m) and distal point (12.5 m) in trenches fed by primary (PE) and secondary (SE) effluent.

	Site A						**Site B**					
	PE			**SE**			**PE**			**SE**		
	P	**M**	**D**	**P**	**M**	**D**	**P**	**M**	**D**	**P**	**M**	**D**
AOB	(+) 1.0	(+) 0.5	(+) 0.1	(+) 0.4	(-) 0.04	(+) 0.1	(+) 0.1	(+) 0.02	(+) 0.5	(+) 1.1	(+) 0.2	(+) 0.5
NOB	(+) 1.9	(+) 2	(+) 1.4	(+) 1.4	(+) 0.4	(+) 0.3	(-) 0.8	(-) 0.9	(-) 0.3	(+) 2.9	(+) 0.5	(+) 0.6
Denitrifier	(+) 3.3	(+) 4.2	(+) 1.2	(+) 3.8	(-) 0.7	(+) 1.0	(+) 2	(+) 1.1	(+) 1.6	(-) 8.3	(-) 1.5	(+) 2
Anammox	(+) 0.03	NC	(+) 0.002	(+) 0.006	NC	NC	(-) 0.002	(-) 0.02	(+) 0.1	(+) 0.6	(+) 0.02	(+) 0.004
Methanogen	(+) 1.3	(-) 0.01	(-) 0.004	(-) 0.01	(+) 0.01	(+) 1.1	(+) 6.0	(+) 9.4	(+) 8.6	(-) 7.9	(-) 8.3	(-) 13.9
Methanotroph	(+) 0.5	(-) 0.01	(-) 0.02	(+) 0.0004	(+) 0.01	(+) 0.1	(+) 0.2	(+) 0.3	(+) 0.7	(-) 0.1	(-) 0.5	(-) 0.02
PAO	(-) 0.002	(-) 1.0	(-) 0.2	(-) 0.1	(-) 0.1	(-) 0.01	(+) 0.2	(+) 0.3	(+) 0.14	(-) 0.2	(+) 0.1	(+) 0.2

For Site A there was a notable increase in denitrifier relative read abundances at the proximal and midpoint of the trench for PE fed trench which exceeded control subsoil values in both 2018 and 2022. At the Site A SE trench the greatest increases of denitrifiers were located only at the proximal sampling the only position that continued to surpass the control subsoil abundance. Denitrifiers in site B increased in all of PE trench position and the distal position of the SE trench. Although there was a drop in denitrifier abundance in the proximal and midpoint position of SE trench at site B, all positions in both trenches continued to be greater in denitrifier abundance than the control subsoil.

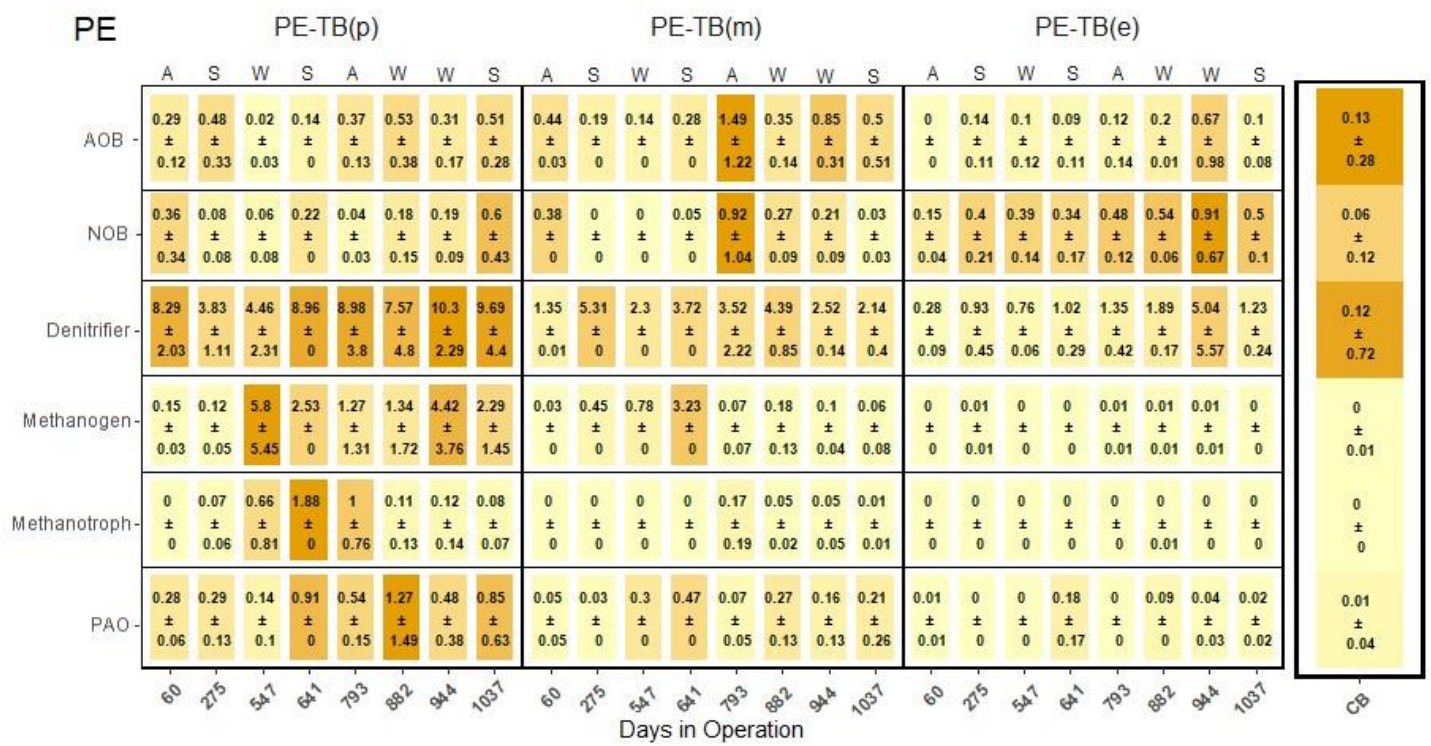


Figure 7.8. Mean ± SD of relative read abundance of Anaerobic Methane Oxidizers (AMO), Denitrifying bacteria, polyphosphate-accumulating organisms (PAO), Nitrite Oxidi zing Bacteria (NOB), Methanogens, Glycogen Accumulating Organisms (GAO), Ammonia Oxidizing Bacteria (AOB), AMO, and Acetogen functional group over time for prim ary effluent (PE) trench at the proximal (1 m), midpoint (5 m) and end point (17 m) and control subsoil at site C.

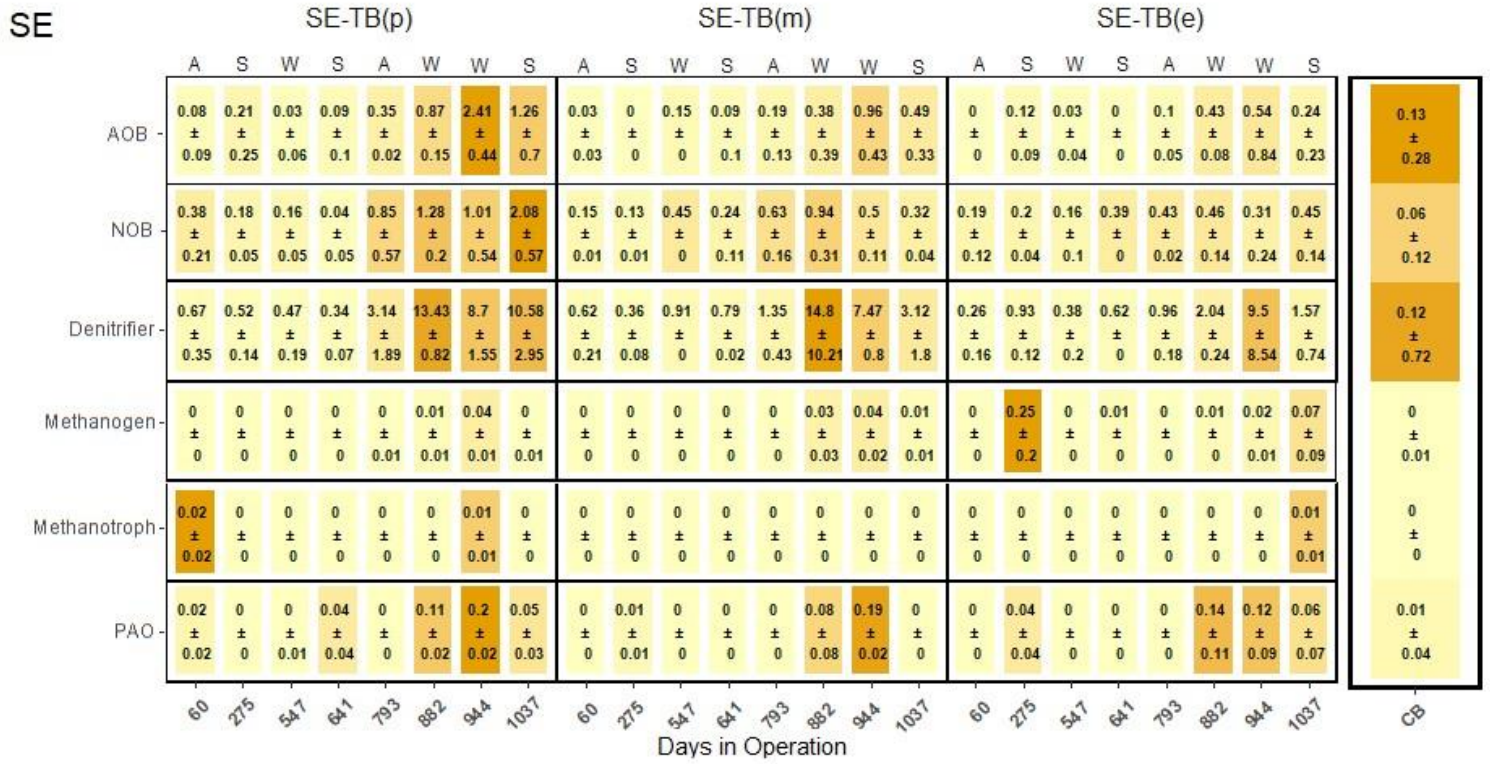


Figure 7.9. Mean ± SD of relative read abundance of Anaerobic Methane Oxidizers (AMO), Denitrifying bacteria, polyphosphate-accumulating organisms (PAO), Nitrite Oxidizing Bacteria (NOB), Methanogens, Glycogen Accumulating Organisms (GAO), Ammonia Oxidizing Bacteria (AOB), AMO, and Acetogen functional group over time for primary efflu ent (SE) trench at the proximal (1 m), midpoint (5 m) and end point (17 m) and control subsoil at site C.

Anammox were discovered in greater numbers across both sites in 2022. The largest relative abundance of anammox was found at the SE trench at Site B at the proximal zone, with abundances varying with depth resulting in 0.3%, 0.6%, 0.7%, 0.3%,0.5%, 1.2 at depths of 0, 2.5, 5, 7.5, 10, 12.5 cm respectively. In the PE fed trench at Site B the other position of high relative abundance of anammox was located at the distal position at a depth of 12.5 cm, the only increase of the anammox group within the trench. In the midpoint of the same trench there was a noted drop in anammox abundance when compared to 860 days in operation. At Site A the greatest increases of annamox were located at the proximal positions at both trenches. The greatest increases of relative read abundance of annamox were 0.1% and 0.05% found at depths of 7.5 cm and 10 cm in the PE trench and 0.02% in the proximal zone at the SE trench at a depth of 12.5 cm. No annamox was found in the control subsoils of any year sampled.

Methanogens showed the largest increase within the proximal sampling and midpoint positions for the PE fed trenches at site A and B respectively. For the SE fed trenches, site A showed the greatest increase in relative read abundance at the distal point in the trench, while Site B exhibited a reduction across the trench, however both sites continued to exhibit higher abundances of methanogens than the control subsoil samples. At site A only the proximal position continued to exceed the control subsoil in its abundance of methanotrophs, however it was only after 2540 days that a (proximal) position in SE trench exceed this abundance level. In site B, all positions in both trenches continued exceeding the control subsoil in abundances of methanotroph, with the midpoint of PE STU only surpassing the control subsoil in 2022. At Site A all the positions of the PE trench, and distal position of the SE trench continued surpass control subsoil PAO relative read abundances. Additionally, after 2540 days in operation the distal position of PE trench and the proximal position of the SE trench now exceeded the PAO abundance in the control subsoil. In Site B there were increases across all of trench dosed with PE and at the midpoint and distal sampling position in the SE fed trench, which continued to surpass control subsoil values.

The end (distal) position (16 m) of all the trenches were sampled in 2022 to compare relative read abundances of functional groups of the distal position of the corresponding trench in 2018. This allowed us to determine if the development of the biomat exhibited any functional changes with growth. In Site A within the PE trench samples taken after 2540 days in operation at the end point showed relative read abundances of methanogens denitrifiers, NOB and AOB that equalled or exceeded values from the distal position samples after 1080 days in operation. For the SE trench of Site A, denitrifiers, NOB and AOB relative read abundances at the end position in 2022 matched or surpassed values from the same position sampled in 2018. In Site B samples taken after 2320 days in operation from the end PE fed trench showed that PAO, methanogens, methanotrophs, denitrifiers and AOB were found to equal or exceed the relative read abundance from those samples taken at 860 days at the distal position. While in the SE fed trench in Site B PAO, denitrifier, NOB and AOB sampled after 2320 days exceeded those distal positioned samples taken in 860 days..

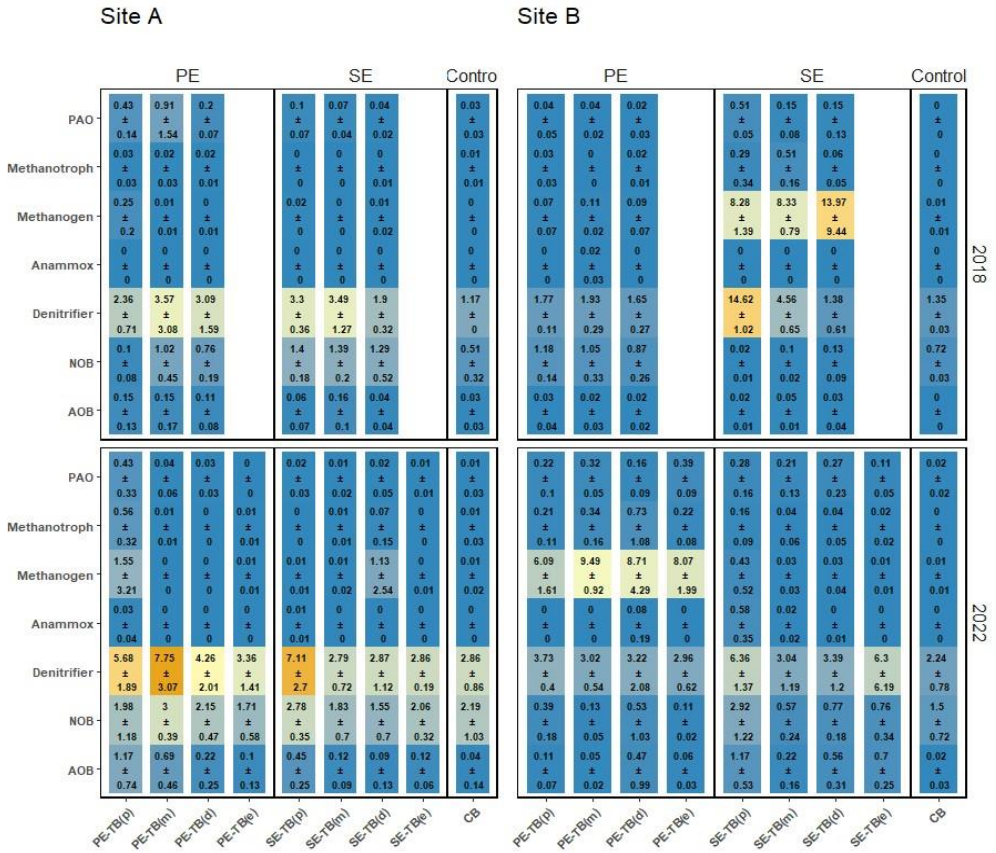


Figure 7.10. Average relative read abundances ± SD for each functional organism abundance: ammonia oxidizing bacteria (AOB), nitrite oxidizing bacteria (NOB), denitrifiers, anammox, methanogens, methanotroph and phosphate accumulating organisms (PAO) between August 2018 and August 2022 in the proximal (1 m), midpoint (7.5 m) and distal point (12.5 m) in trenches fed by primary (PE) and secondary (SE) effluent.

7.5 Discussion

The main aim of this research was to monitor the shifts community structure over time, and to be able to identify critical points in the transition of the ecologies and how this might affect the overall functionality of the system. This was achieved by two separate monitoring campaigns. The initial biomat growth campaign spanned from the early operation of the system at site C, up to 1037 days of operation. The third Site C was designed to improve accessibility to the biomat through the construction of direct sampling ports or "Bioports" to allow for greater sampling frequencies. These increased sampling frequencies would enhance resolutions to determine points of critical transitions of STU ecologies. The second campaign involved sampling Sites A and B in 2018 and 2022. Both site A and B systems having already been running for 35 and 29 months respectively by 2018 are approximately equivalent to when sampling on site C ceased and therefore represent the next stage of biomat development in well-established STUs.

Frequent monitoring of the early establishment of the biomat in site C provided important insights into the spatial temporal variability of the ecologies present. By day 60 of operation the biomat at the proximal position and day 275 for the midpoint of the PE trench at site C had significantly shifted from the composition of the biomat community within SE trench. SE trench biomats have been previously characterised by muted growth levels due to removal of organics due to pre-treatment (Criado Monleon et al., 2022). The application of organics has a crucial effect on the community structure of the biomat ecologies, the variation of *Gammaproteobacteria* which are known exhibit copiotropic strategies highlight these critical shifts in ecologies (Li et al., 2012; Cui et al.,2023). Within the STU the dominance of copiotrophs in the early development of the biomat appears affected by the level of pre-treatment of effluent. *Gammaproteobacteria* at site C dominated only 1 m of the inlet of the PE trench within 60 days of operation, however copiotrophs only dominated the 5m position after 275 days in operation and a spike in abundance of *Gammaproteobacteria* was noted at 17 m downstream of the inlet after 944 days of operation. Although, changes in relative abundance does not necessarily reflect changes in microbial community strategies, the annual reduction in average porewater TOC concentrations suggests that there is a breakdown of organics by this increasing abundance of copiotrophs.

As labile organics are required for denitrification it is unsurprising that the shifts in abundances of denitrifiers match those of copiotrophs across the trench (Murphy et al., 2003; Pabich and Hemond., 2001). The greatest reductions in porewater concentration of total nitrogen occurring at the midpoint of the trench, where there is a lower abundance of copiotrophs suggesting nitrogen removal may be inhibited in positions of the trench due to competition. It's evident that the lower levels of organics found in the SE trench at site C resulted in a significant delay in the establishment of copiotropic dominance with the trench which did not take place until 882 days in operation. Although exceeding abundances of functional groups found within control subsoils after 60 days of operation at 882

microbial hotspots seem to form in the proximal position of SE STU at site C, noted by sustained elevations of PAO and nitrogen cycling bacteria, this also corresponds with the increased average levels of nitrate and phosphorous found in porewater over time. The increase in nitrates across the SE trench may be a combination of increased clogging, an increase in the abundance of nitrifiers and possible competition for organic carbon with denitrifiers.

Whether through immigration or species sorting through organic loading both the PE and SE STUs at site C exhibited increases in diversity, however these positive correlations were only present when comparing samples taken in winter. Organic loading and temperature have been noted to have significant effects on subsoil communities in managed aquifer recharge systems (Li et al., 2012). Studies have also shown a lack of seasonal variation in other wastewater biofilm and subsoil samples, showing similar community structure in summer and winter month (Li et al., 2012; Turki et al., 2017). The effect of adding wastewater to a STU initially resulted in increases diversity with the addition of organics, and the development of microbial hotspots. The seasonal increases in site C alpha diversity observed in winter may be the favourable conditions due to the dispersal of organics allowing for greater abundances of habitat generalists and reduced competition (Bell and Bell., 2021; Xiang et al., 2023). These correlations matched with trends in diversity in previous work that as the biomat develops initially there is a period of heightened diversity, which commences from the most proximal position due to the input of organics and nutrients to the system (Criado Monleon et al., 2022). The lower diversity of the proximal position at the PE-STU at site C after 1037 days of operation compared to the midpoint position suggests that ecology closest to the inlet of the trench had likely reached a tipping point that has resulted in a dominance of copiotrophs (Sul et al., 2013). The increasing diversity of the midpoint and end of PE-STU trench suggests this tipping point was not yet to be reached, as the reduced level of organics and diverse forms of substrates can select for more diverse niches (Kuzyakov et al., 2015). It was evident that that trenches in site C dosed with primary effluent with a higher organic content resulted in quicker shifts in community structure.

In the long-term observations of shifts in copiotropic abundance showed that generally *Gammaproteobacteria* increased its abundances throughout the entire length (16 m) between 2018 and 2022, in nearly all positions exceeding abundances found in the 2022 control subsoils. With varied levels of organic loading between the sites, site B PE trench received ~170% more organic carbon than its site A counterpart and subsequently saw a greater increase in copiotrophs resulting in a significant shift in its community structure. The productivity diversity relationship hypothesis is that once diversity has increased beyond a certain threshold due to resource availability that the diversity outcome becomes negative, which evidently occurred in site B PE trench (Geyer et al., 2019). Although the RBC at site B removes a greater deal of organics than the cocohusk filter at site A it saw a greater increase in copiotrophs within the SE STU and a significant shift in the community. The greater abundance of copiotroph may be attributable to the gravity flow and volume

of effluent, which is more regularly dispersed to the trench, when compared to the intermittent pump-sump responsible for the distribution of effluent in site A. It has been seen in lab experiments of sand filters that the schmutzdecke of continuously fed filters delivers better filtering capacity of *E. coli* than intermittently fed systems (Young-Rojanschi and Madramootoo., 2014). The distribution of nitrogen cycling organisms remained spatially distinct with increases in AOB, NOB, Denitrifier and Annamox in Site A were localised to the proximal or midpoint of the PE trench. In Site B increases of nitrogen cycling organisms were distributed across the trench, or in the case annamox at the distal position. Annamox increases are indicative of areas where the biomat is well established as this has often been found in low flow sites (Humphreys Jr et al., 2019 Cooper et al., 2016). As hydraulic conductivity would have reduced in the STUs with clogging, this would result in a greater distribution of organics resulting in a higher carbon to nitrogen ratio which may be the cause of the increase in denitrifiers further downstream in the STU in Sites A and B (Gill et al., 2009; Knappe et al., 2020; Zuo et al., 2023).

In the long-term sites variation between the PE and SE trenches and the control subsoils diminished, as did the variation within the SE trenches. Variation within the PE trenches increased with time, attributable to spatial shifts in diversity due to organic loading, with copiotrophs generally being higher in abundance close to the inlet of the trench. Pre-treatment of effluent appeared to account for less of the variation in the communities in both sites over time. In Site A it appeared that the ecologies became notably more similar by the midpoint of the trenches, and as there was no significant difference between the two trenches in alpha diversity and species richness. This may suggest that the SE biomat at site A has expanded further downstream of the inlet. Nonlinear growth modelling of the SE biomat (Knappe et al., 2020) predicted that is should reach a final steady state at 8.9 – 13.9 m (95% confidence level) between 6-10 years of operation at Site A, while the PE trench was observed to reach 15 m in length after 10 months in operation.

In proximal position in the SE trench at site B was the only location to significantly increase in alpha diversity and species richness suggesting after 2320 days of operation this portion of the biomat was shifting to a high diversity and richness ecology. While in site B interestingly communities sampled in the PE STU after 860 days of operation overlapped with the more mature SE STU sampled after 2320 days in operation. There was a partial overlap also with SE STU sampled after 860 days of operation and samples taken from end of PE STU after 2320 of activity. These overlaps in community demonstrate the shift in community structure due to the level organics, the time required for these shifts to expand laterally downstream of the inlet. The modelled growth rates estimated biomat length of 10.7 m to 17.1 m (95 % confidence intervals) after approximately 5 to 7 years of operation for the SE fed trench at Site B, while the biomat of the PE STU was observed to reach 15 m after only 13 months (Knappe et al.,2020). These estimated growth rates appear to match changes in the microbial community structure of these sites, as the SE trench biomats expanded, the variance in community structure has reduced drastically across the trenches, reducing the surface area of

"infiltrative dead zones" within the trench (Knappe et al., 2020; Criado Monleon et al., 2022). It is important to note that although there appears to be a lateral expansion in the SE fed biomats, previous studies have shown that it is unlikely that SE fed biomats will continue to grow to the full length of the STU trench. (Gill et al., 2009; Amador and Loomis, 2018). It highlights the risks these slow developing ecologies may pose to groundwater sources if in the highly likely situation that the packaged treatment unit were to fail, and non-treated effluent was dosed to the STU.

7.6. Conclusions

- First successful temporal analysis of subsoils STU systems analysing the effect of effluent pre-treatment on bacterial community composition.
- Copiotrophs (*Gammaproteobacteria*) have been shown to match the lateral extension of expected biomat growth.
- Two significant states in biomat community structure defined, first being in an increase in diversity, combined with rapid biotic and abiotic clogging, biomat development, with a shift in community structure driven by copiotrophic *Gammaproteobacteria*. The second shift of the ecology appears defined by a drop in diversity again driven by the overabundance of copiotrophs.
- System design affected the rate at which change in state occurred in the biomat. The important variables being the level organics applied, and whether they were continuously (gravity) fed or intermittently (pump-sump) affected the rate of biomat development.

8. Conclusion and Recommendations

The main aim of this research was to improve the understanding of the development of microbial biomat communities present with percolation areas of on-site wastewater treatment systems, and the roles they play in the treatment of pollutants. This was accomplished in dedicated full-scale research sites designed and installed by the Department of Civil, Structural and Environmental Engineering, as well as laboratory mesocosm experiments. On the full-scale sites, microbial communities were analysed by accessing subsoil at the infiltrative surface of the percolation trench, and sampling from three engineered soil treatment units receiving different levels of wastewater pre-treatment; primary treated septic tank effluent versus secondary treated effluent. The research finding from these investigations is of relevance to scientists and engineers wishing to gain an understanding soil-community interaction in such engineered treatment systems that incorporate the natural soil, but is also of relevance to more centralised systems that might want to incorporate managed aquifer recharge for treated effluent discharge.

8.1 Summary of Findings

Effluent pre-treatment and species sorting a more significant factor in microbial community composition than soil porosity or immigration in a sandy loam soil treatment unit.

Using soil columns, it was also possible to determine the vertical spatial variation in the development of the biomat. The soil columns receiving PE showed a low species richness and diversity from depths ranging up to 1 cm in low porosity soils and 7 cm in high porosity soils, this was then followed by a microbial hotspot. This suggests that within the clogged soil (i.e., the biomat) there is a reduction in microbial diversity. The low diversity suggests specialised niches optimised for the transformation of nutrients and organic substrates. Consequently, as nutrients are transformed and organics are taken up at the shallower depths, these reduced organics and nutrients may account for the greater diversity observed at the lower depths. In columns dosed with SE, a similar outcome occurred in the laboratory as did at full-scale at Site B, whereby low diversity and species richness were observed and attributed to the steady flow of limited strength substrate.

It was found that the microbial community structure was significantly affected by the level of pre-treatment of the effluent compared to porosity which explained less of the variation between communities. Growth analysis found that the effect of pre-treatment was due to conformational changes in the column biomes, as it was observed that immigration accounted for a small (> 4%) of the relative read abundance of species present within the soil column. However, immigration skewed

towards SE columns likely due to the poor filtering capacity of the muted biomat growth. It was found that community structure was determined by species sorting due to the steady loading of organics and nutrients of differing concentrations.

Effluent pre-treatment effects on spatial ecology of the microbial biomat transposed to known biomass position

The effect of effluent pre-treatment has been shown to have significant effects on the microbial ecology of the biomat at the field scale. Microbial diversity and richness in the biomat change over time. Initially the addition of organic carbon results in a significant increase in diversity and species richness as the availability of substrate creates the conditions for a microbial hotspot of activity. The positions of these hotspots are significant as they demarcate where the biomat is growing and most active. Diversity and species richness decline with the change of environment due to enhanced clogging and the dominance of copiotrophic bacteria which have evolved to thrive in high nutrient environments. In Site A the trench receiving primary effluent exhibited low diversity and richness at the sampling point at the start of the trench, after 1080 days in operation, with microbial hotspots being located at the midpoint and near the end of the trench. In Site B trenches sampled following 860 days in operation receiving primary effluent were receiving the highest organic loads (74.4025 g/d ± 24.9425 g/d), nearly 170% greater than Site A. The organic loading rates at Site B resulted in a distribution of microbial hotspots across the entirety of the trench sampled, with greatest species richness observed at 12.5 m along the trench suggesting where current biomat formation is occurring. In the trenches receiving secondary treated effluent, differences in diversity and species niches compared to the parallel PE fed trenches were significant. In Site A the SE fed biomat showed the highest diversity and richness in species at the sampling location most proximal to the start of the trench. In Site B the SE fed trench biomat appeared to show no microbial hotspots, instead exhibiting a low level of diversity and species richness across the entirety of the trench. The biomat (site B) located at 1 m from the start of the SE trench resulted in the lowest diversity where it appeared that the steady application of low organics and nitrates was selecting for copiotrophs and denitrifiers. The SE fed biomat at Site A receives a partially treated effluent with the cocohusk filter having ~60% removal rate of organics. This incomplete nitrification and higher organic loading appeared to promote a more diverse substrate and subsequently greater diversity and richness within the biomat. These growth patterns match the biomat lengths observed by Knappe et al. (2020) on the same sites, which were determined by analysis of changes of volumetric water content within the STU observed through the EC5 soil sensor.

Critical shifts in community composition observed by continuous monitoring of subsoil biomat

The temporal effects of species sorting were monitored for three research sites at differing points in their operation and while receiving effluent of differing degrees of treatment. At Site C the design and installation of bioports allowed for the continuous monitoring of the biomat, from the point of

installation up to approximately three years in operation. With regular sampling across the length of the trenches it was feasible to identify tipping points in the structure of the biomat community. There are few available studies that have sampled the microbial biomat of STUs and this study represents is one of the first to observe the temporal variation of the community, seasonally and in its development over time. The application of organics and nutrients to the STU appears to result in two main shifts in states of the microbial ecology of a system. The first state can be described by an increase in diversity, combined with rapid biotic and abiotic clogging, biomat development, with a shift in community structure driven by copiotrophic *Gammaproteobacteria*. The second shift of the ecology appears defined by a drop in diversity again driven by the overabundance of copiotrophs. From a functionality perspective the abundance of copiotrophs could impede rates of denitrification as both copiotrophs and denitrifiers compete for liable organic carbon. Copiotrophs have been shown to match the lateral extension of the biomat and with increased bioclogging so does the system's capacity to remove nitrogen from the system due to enhanced hydraulic retention. Bioclogging is also principal cause of failure of these systems resulting in ponding. None of the systems monitored during this experiment exhibited any ponding, therefore it was not possible to profile the ecology at the point of failure. However, it was evident that that the level organics applied, and whether they were continuously (gravity) fed or intermittently (pump-sump) affected the rate of biomat development. In the PE trench at site B, although a relatively younger STU was the first STU to shift in its entirety from a high diversity state to a low diversity copiotrophic dominant state due to its high organic loading. The early development of the biomat is an advantage in domestic households when disposing of wastewater which is composed mostly of nutrient and organics. The use of primary effluent from municipal wastewater treatment plants in MAR may contain too many contaminants of emerging concern such antibiotics or Phospholipid-derived fatty acids. It would be necessary to assure that these contaminants are removed at sufficient rates as to significantly reduce the risk of groundwater resources if this cannot be achieved secondary and or tertiary would be applied to the effluent prior to groundwater recharge. However, this study has shown that to reduce the risk of a steady rate of infiltration into groundwater from pre-treated effluent, augmentation of the percolation area with an organic source such as biochar could accelerate biomat development and enhance the functionality of the MAR system, further reducing the risk to groundwater.

Pepper motile virus is an effective faecal indicator in rural catchments.

This study also validated the effectiveness of pepper mild mottle virus as faecal indicator to assess faecal contamination emanating from on-site wastewater treatment systems. This was determined by the quantification removal rates of the virus from soil mesocosm up to $\log_{10}$ 2.58. Mesocosm studies showed strong correlations in concentrations of PMMoV to classical faecal indicator bacteria such as *E. coli*. The mesocosm experiment also highlighted the importance of pH of both the soil and effluent in the immobilisation of PMMoV, suggesting that the filtration of the virus would be greater in soils of higher acidity.

When trialled at the field level, PMMoV matched the correlation profiles of *E. coli* quantified from porewater samples. The highest removal values of PMMoV at $\log_{10}$ 3.11 were found at the proximal position to the inlet of the STU, likely the position of densest biomat growth. The highest concentration of PMMoV was found at the rear of the trench suggesting the displacement of the influent across the surface. Concentrations found 15 cm deeper showed a drop in PMMoV by two orders of magnitude highlighted the filtering capacity of STU for viral material. None of the faecal indicator bacteria or PPMoV demonstrated any significant correlation with each other or any chemical parameter measured. The fate of FIB and PMMoV can be significantly different within the STU due to their size and replication or reproduction. In STUs the overall removal rate for *E. coli* was higher at $\log_{10}$ 3.01. This higher LRV may diminish the risk that poorly installed STUs may pose for viral contamination of groundwater sources.

Finally, PMMoV was successfully detected and quantified within catchments which have failing on-site domestic wastewater treatment systems as a significant pressure. However, these catchments are also located in agricultural areas where intensified practices may provide additional sources of nutrients and faecal indicator bacteria. PMMoV provides a specific marker which can successfully detect faecal contamination providing results in a fraction of the time of classic culturing methods.

8.2 Significance of this study

This study presents the first systematic and high-resolution monitoring of microbial ecology present within the infiltrative zone of on-site soil treatment units. High frequency sampling of an engineered subsoil is a laborious and time-consuming task. The installation of access ports for direct sampling of the infiltrative surface provides a novel approach for observing the ecology without risk of compromising the integrity of the treatment system. This study was the first to spatially assess the microbial ecology coordinating with volumetric water sensors for an effective transpose of microbial biomat position with the genomic profile of the soil treatment unit and system performance. Sampling soil treatment units of any configuration is a time consuming and hazardous process, the data produced within this study provide high spatial and temporal resolution of the microbial ecology who would otherwise be out of arms reach. This study sampled three sites over 7 years of their operation, making it possible to assess each stage of the biomat, from its inception to its maturation. Findings from this study can assist in the design of future on-site wastewater treatment systems. It is evident from the findings that the loading of organics and nutrients has a significant response on the desired ecological and performance outcome.

Soil treatment units are by their nature a form of localised groundwater recharge, and so have the capability of recycling water as a resource as long as the effluent is treated adequately. On-site wastewater treatment has often been overlooked in genomic surveys of wastewater treatment systems, with much of the research focused on centralised systems. This research is one of the first

to employ tools devised for centralised systems, such as Microbial Database for Activated Sludge in characterising the engineered subsoil. Although soil treatment units are an engineered environment, due to the lack of studies on subsoil ecologies, microbial community assembly is considered a random process at these depths. This study was able to assess the temporal variation of a subsoil to a quasi-constant pressure of contamination, which gives valuable insights into the ecological impact of subsoil contamination. Finally, Pepper Mild Mottle Virus was successfully quantified and its effectiveness as a faecal source tracker verified through a scaled experimental process, from laboratory column study to field scale to full catchment assessment. Its utility in rural environments has been proven, as faecal indicator bacteria such as *E. coli* may not have specificity required to source failing on-site wastewater treatment systems in rural environments where confounding is likely due to the diffuse agricultural sources.

8.3 Recommendations for Expanding Research

- The data accrued within this research was significant, a combination of physical, chemical, biological, and genomic. The data could be applied to machine learning algorithms to better determine the pre-treatment level required for optimum performance.
- Further studies to assess biomat capacity to attenuate pollutants of particular concern such as PFAS, pharmaceuticals, antibiotics, antibiotic resistant genes, and antimicrobial resistant bacteria.
- Integration of this research with isotope probes to gain more insight into microbial nutrient transformations (e.g. calculate rates of nitrification, denitrification, anammox etc.). This would be combined with transcriptomics for a temporal understanding of gene expression within the biomat as it develops, and the community attenuates contaminants and nutrients.
- Test soil cores with X-Ray microscopy to assess the biomat surface area within soils of differing porosities. Visualisation and quantification of the biomat surface area and comparing these results against functionality of the system.
- Unfortunately, due to their age none of the soil treatment had matured to the point where they would be at risk of failing. Sampling and profiling a failing system would identify community compositions at the point of failure.
- The application of microsensors to directly measure pH, O_2, CO_2 online sensors to be installed at the infiltrative surface to directly measure changes in soil parameters and microbial respiration.
- Wastewater irrigation has been applied at increasing levels globally. However, it comes with significant risks to human and environmental health. A column study assessing the effects of subsoil biomats on the yield of arable crops. Columns of comparable depths could assess whether it is feasible, safe and beneficial to grow crops above percolation trenches.

Appendix A: Supplemental Spatial Variation Study

Contents

Figures:

- **Figure SA1** Histogram of sequence reads across all samples.
- **Figure SA2** Accessing STU biomat, gravel matrix and percolation pipe visible (right) and sampling using corer.
- **Figure SA3** Boxplots displaying rarefied data for observed OTUs, species richness calculated using an abundance-based coverage estimates (Chao1) and alpha diversity (Shannon). Samples were aggregated on the bases of sites.
- **Figure SA4** Means in concentration of ammonium, nitrate, ortho-phosphate, total nitrogen, and total organic carbon within the porewater samples at site A, throughout the year of sampling.
- **Figure SA5** Means in concentration of ammonium, nitrate, ortho-phosphate, total nitrogen, and total organic carbon within the porewater samples at site B, throughout the year of sampling.
- **Figure SA6** changes in relative abundance of *Gammaproteobacteria* and Alphaproteobacteria relative to the control subsoils.
- **Figure SA7** Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between January 2016 to August 2018.

Tables

- **SA1** Summary of sequence read data
- **SA2** Table S2 displays the Wilcoxon test values for inter-site analysis Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE)
- **SA3** Mean relative abundance of key phyla within site A and site B,
- **SA4** Relative abundance of key genus within site A and site B

Table S1. Summary of sequence read data

Min	Median	Mean	Max	1st Qu.:	3rd Qu.:
29706	74885	79273	220139	57310	91556

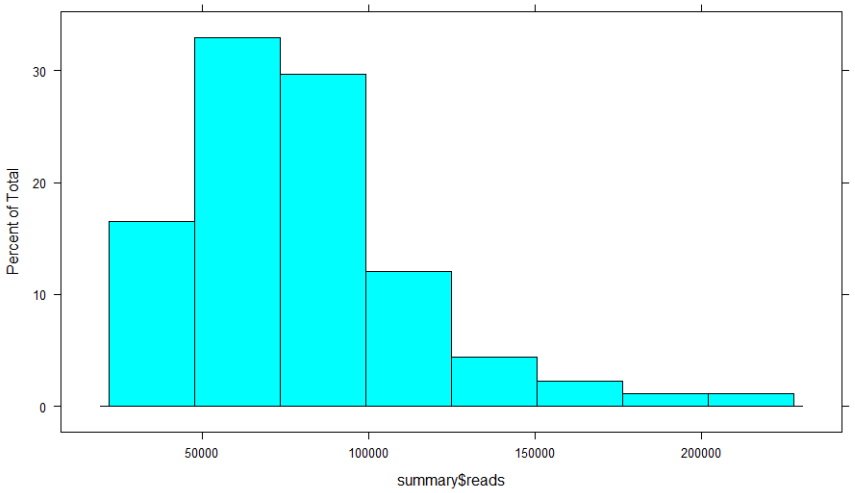


Figure SA1 Histogram of sequence reads across all samples (n=92)

Figure SA2. (left) Accessing STU biomat, gravel matrix and percolation pipe visible (right) and using a corer.

Table SA2 displays the Wilcoxon test values for intersite analysis (ns Pr > 0.05, * Pr ≤ 0.05,**Pr ≤ 0.01, ***Pr ≤ 0.001 and **** Pr ≤ 0.0001) values for comparative intra-site analysis. Samples were aggregated on the bases of systems primary, secondary effluent (PE, SE

Control		PE		SE	
Shannon	Chao1	Shannon	Chao1	Shannon	Chao1
ns	ns	**	*	****	****

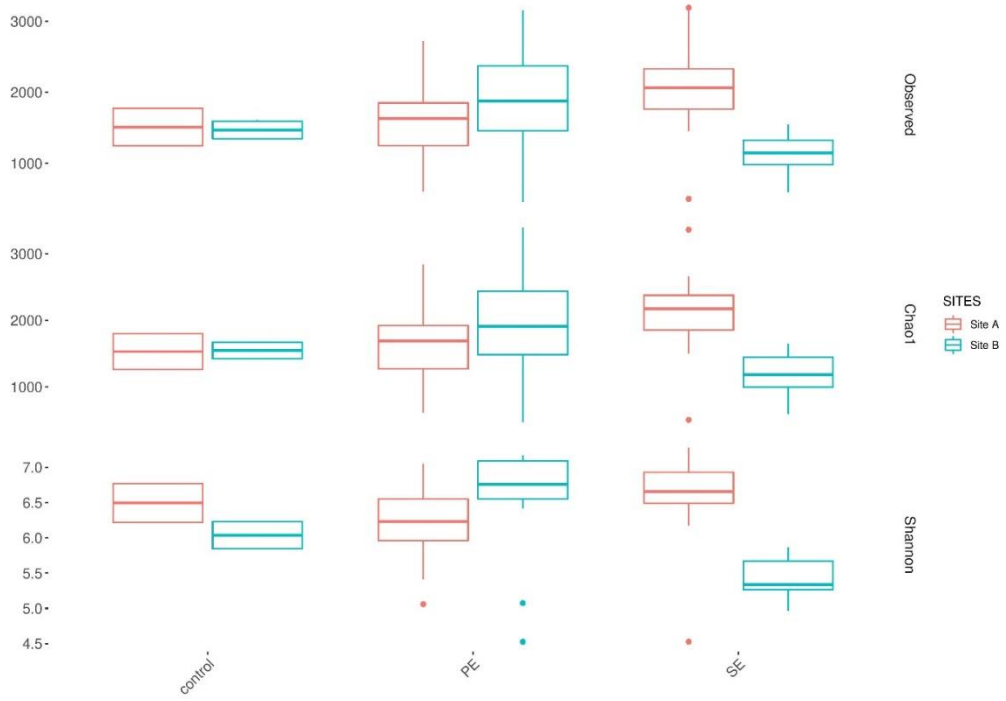


Figure SA3. Boxplots displaying rarefied data for observed OTUs, species richness calculated using an abundance-based coverage estimates (Chao1) and alpha diversity (Shannon). Samples were aggregated on the bases of sites.

)

Table SA3. Mean relative abundance of key phyla within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Acidobacteriota	Actinobacteriota	Bacteroidota	Campylobacterota	Chloroflexi
Control	**A**	**CB**	17.08 ± 0.17	14.62 ± 4.47	3.64 ± 4.08	0 ± 0.01	11.16 ± 7.77
	B	**CB**	27.35 ± 4.1	8.05 ± 0.15	0.36 ± 0.09	0 ± 0	9.2 ± 0.04
	A	**CT**	21.05 ± 1.06	10.23 ± 1.37	3.27 ± 0.44	0 ± 0	8.07 ± 1.14
	B	**CT**	20.02 ± 2.42	11.04 ± 2.21	1.98 ± 1.76	0 ± 0	8.1 ± 1.08
Primary Effluent	**A**	**PE**	0.24 ± 0	1.98 ± 0	16.48 ± 0	9.42 ± 0	0.29 ± 0
	B	**PE**	1.3 ± 0	8.94 ± 0	4.78 ± 0	0.93 ± 0	1.2 ± 0
	A	**PE-TT**	22.03 ± 5.84	12.58 ± 6.41	2.91 ± 1.66	0 ± 0	8.94 ± 1.51
	B	**PE-TT**	17.27 ± 2.53	10.67 ± 1.45	1.92 ± 0.7	0 ± 0	10.32 ± 5.08
	A	**PE-TB(p)**	10.22 ± 5.43	11.72 ± 1.13	7.3 ± 2.98	1.91 ± 2.17	7.67 ± 5.61
	B	**PE-TB(p)**	14.01 ± 0.97	8.25 ± 2.7	1.31 ± 0.99	0.01 ± 0.01	15.84 ± 3.95
	A	**PE-TB(m)**	14.48 ± 3.27	9.25 ± 1.58	2.1 ± 1.85	0.04 ± 0.07	12.43 ± 1.88
	B	**PE-TB(m)**	14.11 ± 2.27	8.86 ± 2.19	2.54 ± 0.7	0.01 ± 0.01	11.2 ± 1.75
	A	**PE-TB(d)**	14.62 ± 2.37	8.81 ± 2.25	2.04 ± 1.3	0.01 ± 0.01	12.82 ± 4.4
	B	**PE-TB(d)**	13.39 ± 1.95	12.34 ± 2.64	1.68 ± 1.13	0 ± 0	11.46 ± 2.77
Secondary Effluent	**A**	**SE**	0.52 ± 0	1.11 ± 0	20.44 ± 0	17.56 ± 0	0.04 ± 0
	B	**SE**	8.96 ± 6.59	8.36 ± 4.61	9.29 ± 7.03	0 ± 0	18.22 ± 1.7
	A	**SE-TT**	17.89 ± 0.69	11.14 ± 2.01	4.85 ± 1.94	0.02 ± 0.05	8.48 ± 1.51
	B	**SE-TT**	15 ± 7.89	15.71 ± 9.41	3.9 ± 2.28	0.01 ± 0.01	9.02 ± 4.19
	A	**SE-TB(p)**	12.52 ± 1.83	8.64 ± 1.02	2.58 ± 0.68	0.01 ± 0.01	14.06 ± 2.81
	B	**SE-TB(p)**	1.69 ± 0.3	7.26 ± 1.1	8.22 ± 1.79	2.05 ± 0.21	1.31 ± 0.2
	A	**SE-TB(m)**	13.72 ± 0.67	9.06 ± 1.46	2.68 ± 1.84	0.02 ± 0.03	13.89 ± 2.36
	B	**SE-TB(m)**	3.51 ± 0.68	5.94 ± 0.24	8.99 ± 0.85	2.81 ± 0.43	2.99 ± 1.08
	A	**SE-TB(d)**	14.82 ± 2.25	8.33 ± 2.37	3.01 ± 2.13	0 ± 0	9.86 ± 0.8
	B	**SE-TB(d)**	5.05 ± 3.29	5.06 ± 2.01	9.84 ± 1.56	0.65 ± 0.48	3.41 ± 1.72

Table SA3. Mean relative abundance of key phyla within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Desulfobacterota	Firmicutes	Gemmatimonadota	Halobacterota	Latescibacterota
Control	**A**	**CB**	1.25 ± 0.76	6.58 ± 5.3	1.4 ± 1.36	0 ± 0	0.56 ± 0.23
	B	**CB**	3.43 ± 0.01	1.63 ± 1.15	1.76 ± 0.19	0 ± 0	0.4 ± 0.09
	A	**CT**	1.21 ± 0.21	2.36 ± 0.47	0.49 ± 0.13	0 ± 0	1.51 ± 0.03
	B	**CT**	1.81 ± 0.22	6.39 ± 1.45	0.51 ± 0.17	0.04 ± 0.05	1.5 ± 0.81
PE	**A**	**PE**	2 ± 0	27.37 ± 0	0.01 ± 0	0.1 ± 0	0 ± 0
	B	**PE**	1.26 ± 0	11.28 ± 0	0 ± 0	5.01 ± 0	0 ± 0
	A	**PE-TT**	1.16 ± 0.39	3.44 ± 1.68	0.37 ± 0.16	0 ± 0	1.26 ± 0.54
	B	**PE-TT**	2.1 ± 0.72	5.28 ± 1.65	1.02 ± 0.59	0 ± 0.01	1.54 ± 0.4
	A	**PE-TB(p)**	6.59 ± 6.52	8.85 ± 6.48	1.42 ± 1.1	0.12 ± 0.17	0.14 ± 0.09
	B	**PE-TB(p)**	1.13 ± 0.26	2.16 ± 1.2	1.58 ± 0.39	0.01 ± 0.01	1.2 ± 0.21
	A	**PE-TB(m)**	1.34 ± 0.18	6.85 ± 4.8	2.19 ± 1.17	0 ± 0	0.73 ± 0.13
	B	**PE-TB(m)**	1.38 ± 0.28	5.01 ± 1.47	1.19 ± 0.19	0 ± 0.01	1.17 ± 0.11
	A	**PE-TB(d)**	1.65 ± 0.93	3.11 ± 1.79	1.38 ± 0.66	0 ± 0	0.81 ± 0.52
	B	**PE-TB(d)**	1.48 ± 0.29	7.16 ± 3.16	1.26 ± 0.17	0.01 ± 0.01	1.45 ± 0.22
SE	**A**	**SE**	2.33 ± 0	17.71 ± 0	0 ± 0	0.09 ± 0	0 ± 0
	B	**SE**	0.68 ± 0.43	1.07 ± 0.33	0.42 ± 0.45	0.02 ± 0.01	0.11 ± 0.16
	A	**SE-TT**	1.37 ± 0.3	5.74 ± 2.13	0.87 ± 0.41	0 ± 0	1.23 ± 0.22
	B	**SE-TT**	1.53 ± 0.47	10.47 ± 8.94	0.69 ± 0.29	0.01 ± 0.01	1.32 ± 0.49
	A	**SE-TB(p)**	1.19 ± 0.07	3.54 ± 0.75	1.63 ± 0.28	0 ± 0	1.07 ± 0.18
	B	**SE-TB(p)**	2.9 ± 0.59	14.77 ± 2.43	0.02 ± 0.01	5.5 ± 0.9	0.04 ± 0.03
	A	**SE-TB(m)**	1.34 ± 0.31	4.1 ± 0.66	1.87 ± 0.48	0 ± 0	1.32 ± 0.33
	B	**SE-TB(m)**	4.27 ± 0.39	18.95 ± 2.55	0.07 ± 0.05	4.87 ± 0.75	0.16 ± 0.02
	A	**SE-TB(d)**	1.62 ± 0.31	11.47 ± 3.77	2.42 ± 1.35	0 ± 0	0.75 ± 0.5
	B	**SE-TB(d)**	8.16 ± 5.15	19.89 ± 4.49	0.24 ± 0.09	12.94 ± 10.61	0.12 ± 0.01

Table SA3. Relative abundance of key phyla within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Myxococcota	Planctomycetota	Proteobacteria	Synergistota	Verrucomicrobiota
Control	**A**	**CB**	1.5 ± 0.51	8.51 ± 1.32	14.24 ± 6.46	0 ± 0	8.07 ± 5.18
	B	**CB**	1.43 ± 0.14	8.34 ± 0.31	9.77 ± 0.41	0 ± 0	6.62 ± 0.97
	A	**CT**	1.23 ± 0.11	15.15 ± 0.49	13.59 ± 2.97	0 ± 0	11.87 ± 1.45
	B	**CT**	1.17 ± 0.39	9.96 ± 0.97	10.91 ± 3.19	0 ± 0	14.72 ± 3.71
PE	**A**	**PE**	0.1 ± 0	0.39 ± 0	21.1 ± 0	12.21 ± 0	1.02 ± 0
	B	**PE**	0.13 ± 0	2.61 ± 0	46.36 ± 0	9.92 ± 0	0.8 ± 0
	A	**PE-TT**	1.26 ± 0.54	12.84 ± 2.83	12.24 ± 4.1	0 ± 0	12.93 ± 3.09
	B	**PE-TT**	1.13 ± 0.28	11.2 ± 1.23	11.74 ± 1.06	0 ± 0	11.81 ± 4.86
	A	**PE-TB(p)**	0.87 ± 0.71	6.92 ± 2.29	25.13 ± 4.02	0.49 ± 0.4	3.21 ± 1.62
	B	**PE-TB(p)**	0.6 ± 0.12	12.4 ± 2.08	13.22 ± 2.85	0.01 ± 0.02	4.99 ± 3
	A	**PE-TB(m)**	1.02 ± 0.12	7.6 ± 1.26	16.3 ± 7.8	0 ± 0	3.72 ± 1.01
	B	**PE-TB(m)**	1.03 ± 0.45	13.54 ± 2.81	14.69 ± 1.95	0 ± 0	7.77 ± 1.67
	A	**PE-TB(d)**	0.97 ± 0.26	9.6 ± 1.07	17.07 ± 8.44	0 ± 0	3.33 ± 1.43
	B	**PE-TB(d)**	1.35 ± 0.49	10.93 ± 2.6	10.68 ± 3.66	0 ± 0	6.16 ± 1.08
SE	**A**	**SE**	0.04 ± 0	0.27 ± 0	19.12 ± 0	12.56 ± 0	0.84 ± 0
	B	**SE**	0.79 ± 0.12	15.19 ± 5.14	26.18 ± 5.95	0.02 ± 0.01	2.47 ± 0.47
	A	**SE-TT**	1.39 ± 0.41	12.31 ± 1.36	14.71 ± 1.49	0 ± 0.01	8.9 ± 1.9
	B	**SE-TT**	1.47 ± 0.66	8.76 ± 4.4	13.17 ± 5	0 ± 0	9.5 ± 4.27
	A	**SE-TB(p)**	1.59 ± 0.17	11.37 ± 1.06	18.32 ± 1.71	0.01 ± 0.01	4.9 ± 1.6
	B	**SE-TB(p)**	0.1 ± 0.04	1.51 ± 0.29	39.29 ± 1.75	8.59 ± 2.99	0.85 ± 0.2
	A	**SE-TB(m)**	1.42 ± 0.66	10.51 ± 1.45	16.8 ± 3.96	0 ± 0	3.15 ± 1.1
	B	**SE-TB(m)**	1.94 ± 1.28	1.52 ± 0.22	24.19 ± 1.39	9.32 ± 1.73	1.9 ± 0.12

	Site	System	Myxococcota	Planctomycetota	Proteobacteria	Synergistota	Verrucomicrobiota
SE	**A**	**SE-TB(d)**	1.18 ± 0.25	8.82 ± 2.61	12.35 ± 2.63	0 ± 0.01	4.12 ± 1.9
	B	**SE-TB(d)**	6.67 ± 5.96	3.69 ± 1.21	6.47 ± 2.4	6.81 ± 2.47	1.7 ± 0.28

Table SA4. Mean relative abundance of key genus within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Acidobacteriota; midas_g_1291	Acidobacteriota; midas_g_30578	Actinobacteriota; Gaiella	Actinobacteriota; Mycobacterium	Chloroflexi; midas_g_3277
Control	**A**	**CB**	3.6 ± 0.54	2.97 ± 0.64	0.85 ± 0.41	1.15 ± 0.77	0.97 ± 0.26
	B	**CB**	2.7 ± 0.42	1.17 ± 0.08	0.51 ± 0.16	0.34 ± 0.04	0.49 ± 0.04
	A	**CT**	5.95 ± 0.9	1.99 ± 0.33	0.86 ± 0.23	1.1 ± 0.28	1.16 ± 0.14
	B	**CT**	5.41 ± 0.6	3.12 ± 1.48	1.28 ± 0.06	1 ± 0.6	1.04 ± 0.18
Primary	**A**	**PE**	0.05 ± 0	0 ± 0	0 ± 0	0.04 ± 0	0.06 ± 0
EFfluent	**B**	**PE**	0.05 ± 0	0 ± 0	0.03 ± 0	0.61 ± 0	0.04 ± 0
	A	**PE-TT**	6.76 ± 2.45	2.06 ± 0.66	1.11 ± 0.5	1.42 ± 0.46	1.48 ± 0.26
	B	**PE-TT**	4.21 ± 0.99	1.63 ± 0.42	1.19 ± 0.21	0.64 ± 0.32	1.23 ± 0.47
	A	**PE-TB(d)**	2.82 ± 1.56	1.5 ± 0.98	0.45 ± 0.43	0.38 ± 0.22	1.41 ± 0.35
	B	**PE-TB(d)**	2.92 ± 0.21	1.96 ± 0.61	1.14 ± 0.35	0.82 ± 0.42	1.21 ± 0.3
	A	**PE-TB(m)**	3.09 ± 0.62	1.52 ± 0.4	0.35 ± 0.24	0.45 ± 0.18	1.55 ± 0.5
	B	**PE-TB(m)**	2.92 ± 0.46	1.79 ± 0.48	0.8 ± 0.28	0.54 ± 0.28	0.88 ± 0.14
	A	**PE-TB(p)**	0.79 ± 0.74	0.33 ± 0.36	0.15 ± 0.07	2.93 ± 0.94	0.63 ± 0.49
	B	**PE-TB(p)**	3.1 ± 0.47	1.84 ± 0.52	0.72 ± 0.3	0.21 ± 0.12	1 ± 0.26
Secondary	**A**	**SE**	0 ± 0	0.02 ± 0	0 ± 0	0.54 ± 0	0 ± 0
Effluent	**B**	**SE**	0.73 ± 1	0.07 ± 0.1	0.06 ± 0.08	0.51 ± 0.15	0.24 ± 0.18
	A	**SE-TT**	4.93 ± 0.54	2.37 ± 0.49	0.94 ± 0.23	1.11 ± 0.45	1.13 ± 0.36
	B	**SE-TT**	3.21 ± 1.78	2.29 ± 1.5	1.47 ± 1.01	1.12 ± 0.54	0.98 ± 0.42
	A	**SE-TB(p)**	3.29 ± 0.7	1.94 ± 0.51	0.3 ± 0.24	0.78 ± 0.12	1.1 ± 0.22
	B	**SE-TB(p)**	0.03 ± 0.01	0.02 ± 0.01	0.06 ± 0.02	0.39 ± 0.06	0.04 ± 0.03
	A	**SE-TB(m)**	3.29 ± 0.4	1.59 ± 0.52	0.34 ± 0.2	0.52 ± 0.18	1.78 ± 0.8
	B	**SE-TB(m)**	0.3 ± 0.08	0.13 ± 0.03	0.33 ± 0.11	0.32 ± 0.03	0.29 ± 0.09
	A	**SE-TB(d)**	3.09 ± 1.06	1.55 ± 0.71	0.56 ± 0.32	0.57 ± 0.14	0.97 ± 0.24
	B	**SE-TB(d)**	0.38 ± 0.3	0.03 ± 0.03	0.21 ± 0.05	0.06 ± 0.06	0.29 ± 0.23

Table SA4. Relative abundance of key genus within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Chloroflexi; midas_g_6161	Firmicutes; Bacillus	Firmicutes; Clostridium_sensu_stricto_1	Firmicutes; o__Bacillales_OTU_1	Firmicutes; Romboutsia
Control	**A**	**CB**	2.02 ± 2.08	0.59 ± 0.03	0.02 ± 0.01	2.77 ± 2.51	0.01 ± 0.02
	B	**CB**	1.29 ± 0.09	0.28 ± 0.29	0.03 ± 0.02	0.28 ± 0.27	0 ± 0
	A	**CT**	0.81 ± 0.07	0.51 ± 0.1	0 ± 0	0.86 ± 0.2	0.01 ± 0.01
	B	**CT**	0.7 ± 0.2	1.86 ± 0.43	0.07 ± 0.06	2.03 ± 0.98	0.05 ± 0.03
Primary	**A**	**PE**	0 ± 0	0.17 ± 0	0.91 ± 0	0 ± 0	0.55 ± 0
Effluent	**B**	**PE**	0 ± 0	0.16 ± 0	2.75 ± 0	0.11 ± 0	3.33 ± 0
	A	**PE-TT**	0.92 ± 0.32	0.56 ± 0.24	0.04 ± 0.03	1.43 ± 0.66	0 ± 0.01
	B	**PE-TT**	1.38 ± 0.85	1.18 ± 0.41	0.03 ± 0.01	1.86 ± 0.93	0.03 ± 0.02
	A	**PE-TB(d)**	1.24 ± 0.42	0.33 ± 0.38	0.05 ± 0.04	0.7 ± 0.51	0.06 ± 0.06
	B	**PE-TB(d)**	0.46 ± 0.11	1.58 ± 0.92	0.14 ± 0.1	1.8 ± 0.82	0.08 ± 0.05
	A	**PE-TB(m)**	1.12 ± 0.43	0.5 ± 0.38	0.06 ± 0.06	1.93 ± 2.08	0.07 ± 0.08
	B	**PE-TB(m)**	0.4 ± 0.08	0.69 ± 0.4	0.1 ± 0.05	1.98 ± 0.5	0.11 ± 0.03
	A	**PE-TB(p)**	0.71 ± 0.77	0.51 ± 0.39	0.32 ± 0.28	1.16 ± 1.29	0.18 ± 0.11
	B	**PE-TB(p)**	0.32 ± 0.13	0.34 ± 0.34		0.54 ± 0.31	0.11 ± 0.04
Secondary	**A**	**SE**	0 ± 0	0.22 ± 0	0.2 ± 0	0 ± 0	0.11 ± 0
Effluent	**B**	**SE**	0 ± 0	0.03 ± 0.04	0.35 ± 0.13	0 ± 0	0.29 ± 0.03
	A	**SE-TT**	0.82 ± 0.17	0.9 ± 0.29	0.05 ± 0.1	1.42 ± 0.41	0.02 ± 0.01
	B	**SE-TT**	0.78 ± 0.42	2.52 ± 2.64	0.15 ± 0.09	2.83 ± 2.39	0.23 ± 0.27
	A	**SE-TB(p)**	1.19 ± 0.46	0.45 ± 0.14	0.06 ± 0.01	0.9 ± 0.32	0.05 ± 0.02
	B	**SE-TB(p)**	0.01 ± 0.01	0.37 ± 0.11	3.3 ± 0.59	0.41 ± 0.15	4.97 ± 1.01
	A	**SE-TB(m)**	1.63 ± 0.28	0.45 ± 0.16	0.06 ± 0.03	0.86 ± 0.25	0.06 ± 0.02
	B	**SE-TB(m)**	0.06 ± 0.02	0.8 ± 0.21	4.89 ± 1.01	0.93 ± 0.08	6.97 ± 1.68
	A	**SE-TB(d)**	0.52 ± 0.19	0.48 ± 0.31	0.39 ± 0.63	3.18 ± 1.67	0.03 ± 0.01
	B	**SE-TB(d)**	0.02 ± 0.03	0.77 ± 0.2	1.71 ± 1.11	0.99 ± 0.28	4.36 ± 3.16

Table SA4. Relative abundance of key genus within site A and site B, aggregated on the bases of systems primary, secondary effluent (PE, SE), STU top (TT), STU base (TB), control top (CT) and base (CB).

	Site	System	Halobacterota; Methanosarcina	Latescibacterota; midas_g_2686	Nitrospirota; Nitrospira	Planctomycetota; Pirellula	Proteobacteria; Thauera
Control	**A**	**CB**	0 ± 0	0.43 ± 0.26	0.5 ± 0.43	0.76 ± 0.24	0 ± 0
	B	**CB**	0 ± 0	0.29 ± 0.09	0.75 ± 0.02	0.44 ± 0.11	0 ± 0
	A	**CT**	0 ± 0	1.14 ± 0.08	0.34 ± 0.04	1.36 ± 0.55	0 ± 0
	B	**CT**	0.03 ± 0.03	1.21 ± 0.83	0.33 ± 0.09	0.81 ± 0.16	0 ± 0
Primary	**A**	**PE**	0.03 ± 0	0 ± 0	0.07 ± 0	0 ± 0	0 ± 0
Effluent	**B**	**PE**	3.1 ± 0	0 ± 0	0.05 ± 0	0.25 ± 0	5.98 ± 0
	A	**PE-TT**	0 ± 0	1.09 ± 0.47	0.18 ± 0.07	0.71 ± 0.3	0 ± 0
	B	**PE-TT**	0 ± 0.01	1.33 ± 0.36	0.29 ± 0.11	1.03 ± 0.35	0 ± 0
	A	**PE-TB(d)**	0 ± 0	0.67 ± 0.48	0.54 ± 0.14	1.2 ± 0.52	0.09 ± 0.1
	B	**PE-TB(d)**	0.01 ± 0.01	1.22 ± 0.23	0.86 ± 0.3	0.76 ± 0.2	0.02 ± 0.01
	A	**PE-TB(m)**	0 ± 0	0.57 ± 0.11	1.09 ± 0.5	0.66 ± 0.3	0.18 ± 0.21
	B	**PE-TB(m)**	0 ± 0.01	0.86 ± 0.09	1.02 ± 0.34	1.42 ± 0.25	0.04 ± 0.02
	A	**PE-TB(p)**	0.04 ± 0.06	0.11 ± 0.09	0.09 ± 0.08	0.44 ± 0.28	0.38 ± 0.3
	B	**PE-TB(p)**	0.01 ± 0.01	0.96 ± 0.2	1.17 ± 0.18	1.27 ± 0.2	0.05 ± 0.03
Secondary	**A**	**SE**	0 ± 0	0 ± 0	0 ± 0	0.04 ± 0	0.02 ± 0
Effluent	**B**	**SE**	0 ± 0	0.11 ± 0.16	0.3 ± 0.4	1.13 ± 0.64	0 ± 0
	A	**SE-TT**	0 ± 0	0.88 ± 0.17	0.43 ± 0.14	1.06 ± 0.25	0 ± 0
	B	**SE-TT**	0 ± 0.01	1.06 ± 0.43	0.37 ± 0.21	1.08 ± 0.7	0 ± 0
	A	**SE-TB(p)**	0 ± 0	0.8 ± 0.13	1.37 ± 0.2	1.08 ± 0.3	0 ± 0.01
	B	**SE-TB(p)**	3.33 ± 0.19	0.01 ± 0.01	0.02 ± 0.01	0.15 ± 0.06	9.21 ± 1.05

	Site	System	Halobacterota; Methanosarcina	Latescibacterota; midas_g_2686	Nitrospirota; Nitrospira	Planctomycetota; Pirellula	Proteobacteria; Thauera
Secondary Effluent	**A**	**SE-TB(m)**	0 ± 0	1 ± 0.31	1.09 ± 0.22	1.08 ± 0.27	0.17 ± 0.18
	B	**SE-TB(m)**	2.6 ± 0.25	0.09 ± 0.02	0.09 ± 0.01	0.15 ± 0.08	1.49 ± 0.23
	A	**SE-TB(d)**	0 ± 0	0.57 ± 0.34	1.18 ± 0.49	0.91 ± 0.42	0 ± 0
	B	**SE-TB(d)**	10.3 ± 8.85	0.03 ± 0.03	0.12 ± 0.11	0.05 ± 0.05	0.31 ± 0.25

.

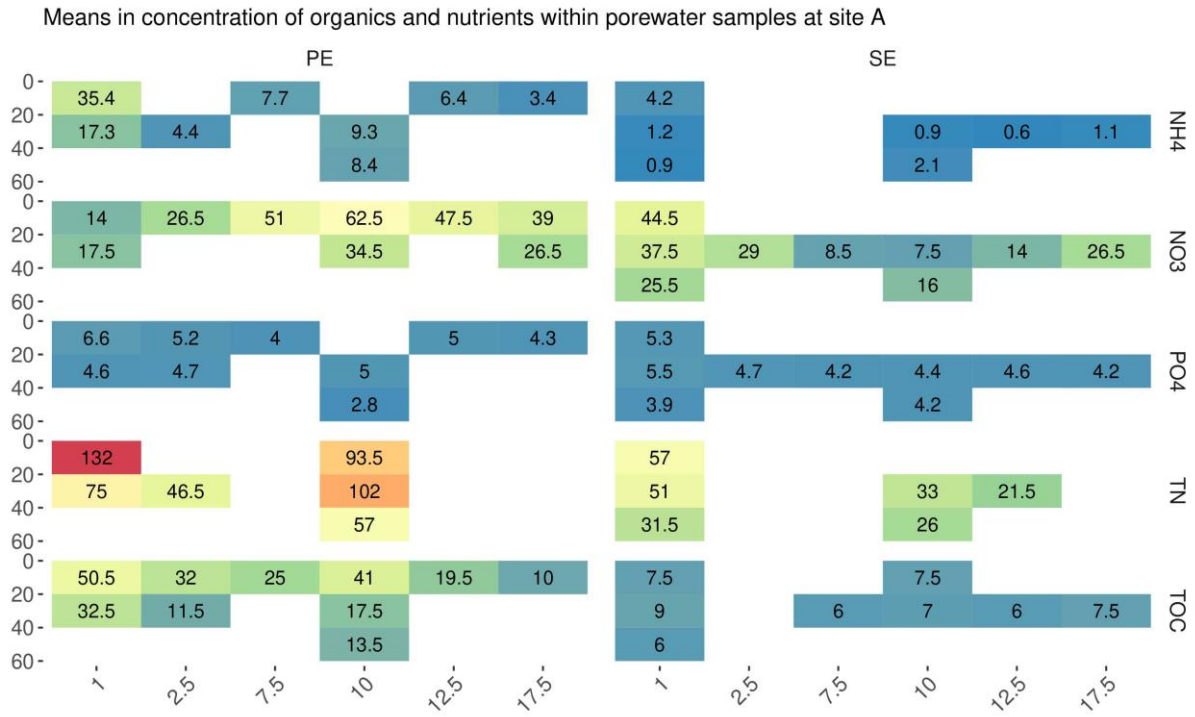


Figure SA4 Means in concentration of ammonium, nitrate, ortho-phosphate, total nitrogen and total organic carbon within the porewater samples at site A, throughout the year of sampling.

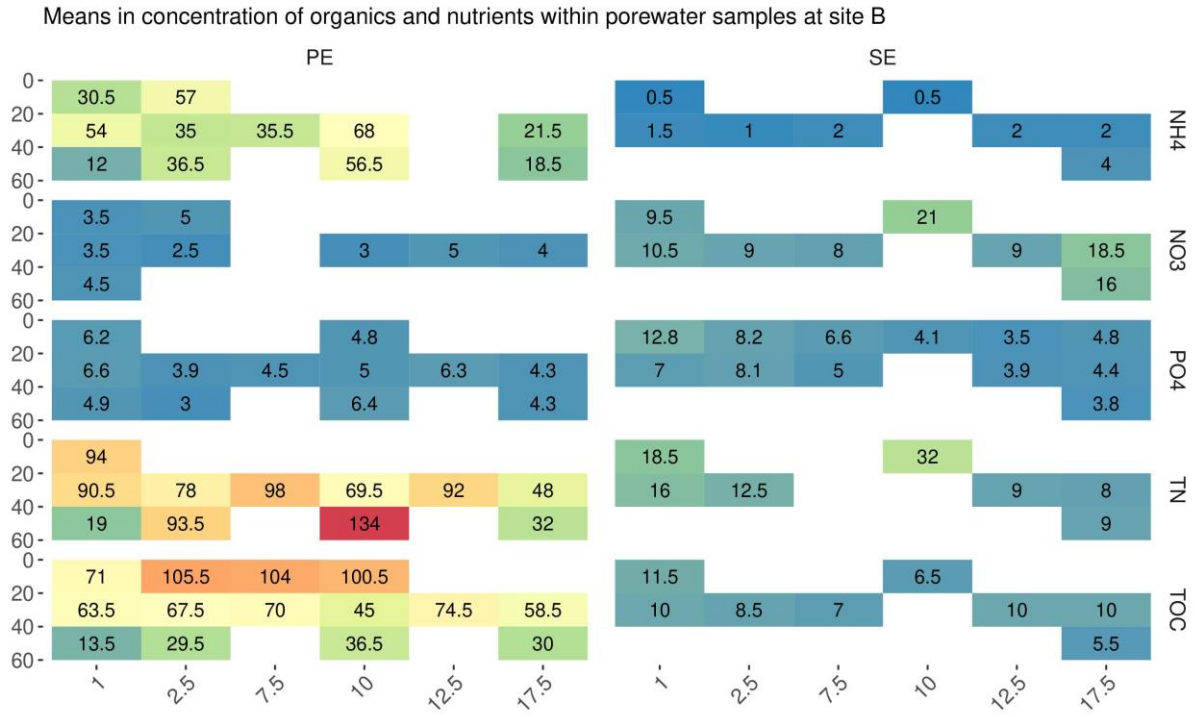


Figure SA5 Means in concentration of ammonium, nitrate, ortho-phosphate, total nitrogen and total organic carbon within the porewater samples at site B, throughout the year of sampling.

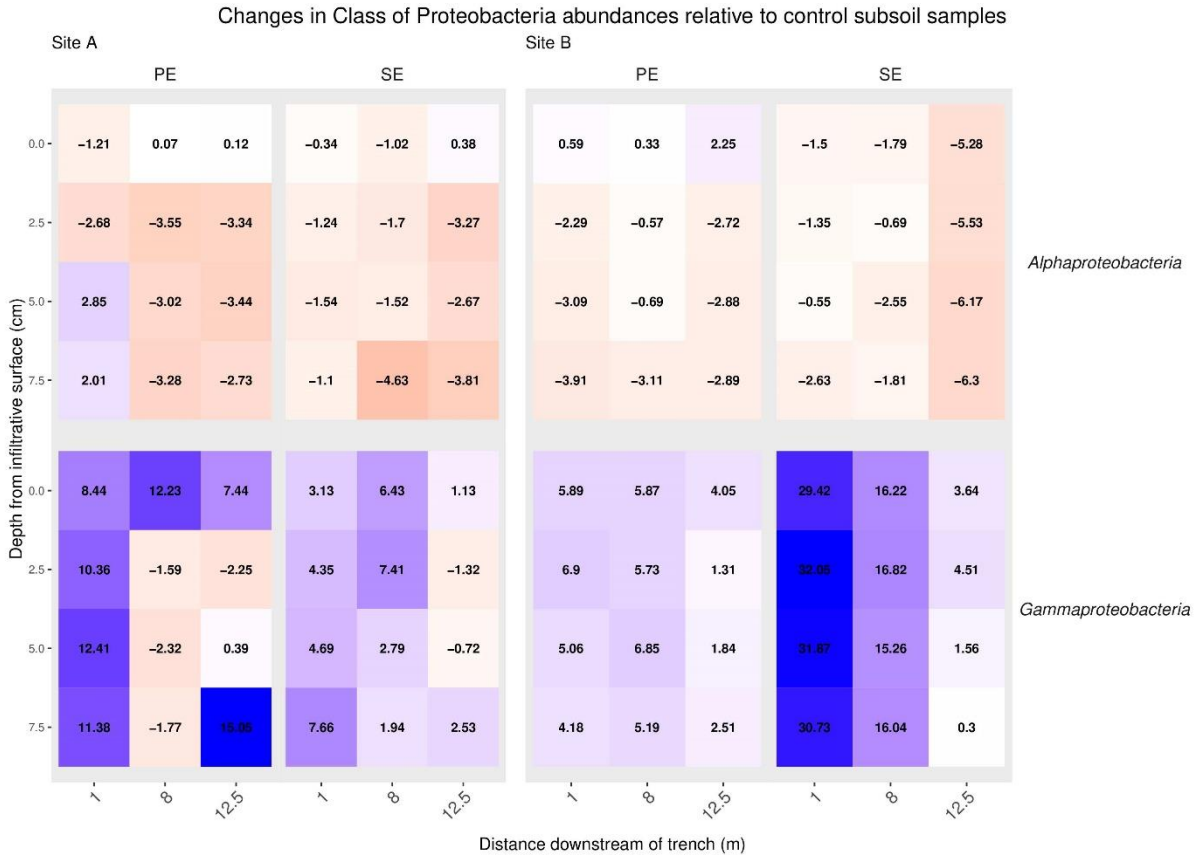


Figure SA6 changes in relative abundance of *Gammaproteobacteria* and Alphaproteobacteria relative to the control subsoils, increases are marked in blue and reductions in red.

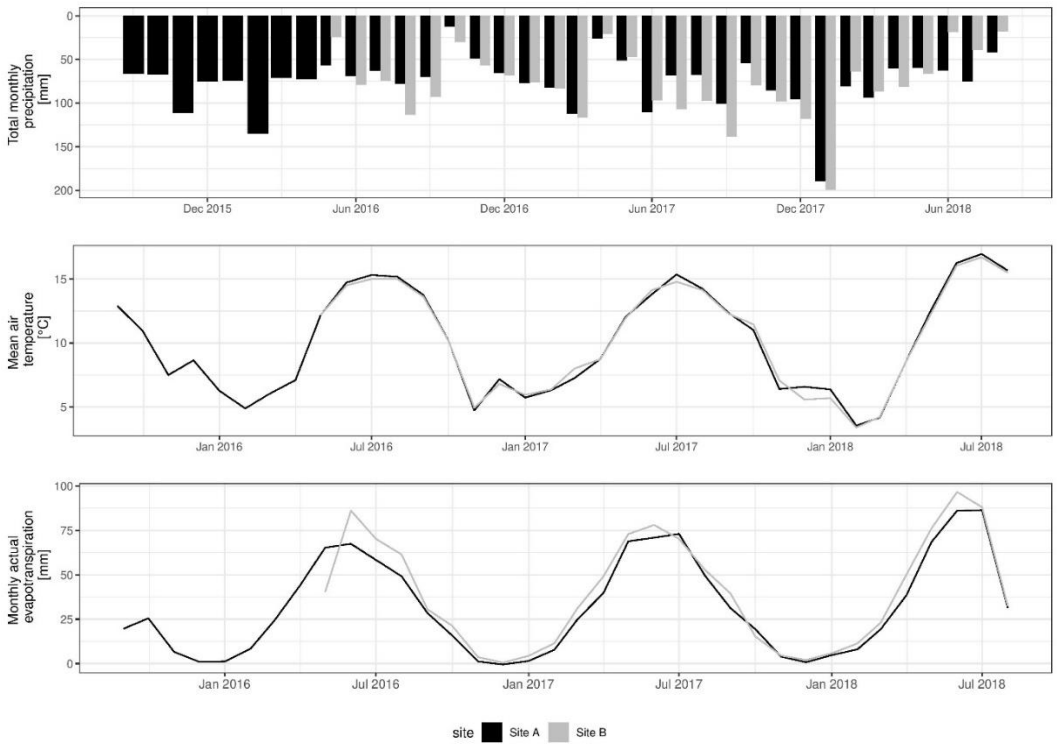


Figure SA7 Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between January 2016 to August 2018.

Appendix B: Supplemental to Soil Mesocosm experiment

Contents

Figures:

- **Figure SB1** Column design (a) and construction (B)
- **Figure SB2.** Influent and Effluent concentrations …
- **Figure SB3** Boxplots displaying species richness and diversity of soil columns ….
- **Figure SB4** Net Growth rate absent and present within wastewater across all columns PE and SE influent High and Low porosity soil …
- **Figure SB5** Net Growth rate absent and present within wastewater across all columns depth profiles …Pg (vii)

Tables:

- **Table SB1** Methanotroph targets for sequencing analysis and source literature …
- **Table SB2** Median removal efficiencies of column effluent relative to influent concentrations, and mean concentrations and standard deviations for each of effluent components
- **Table SB3** Dominant and rare growing and not growing species abundances, measured across depth profiles.

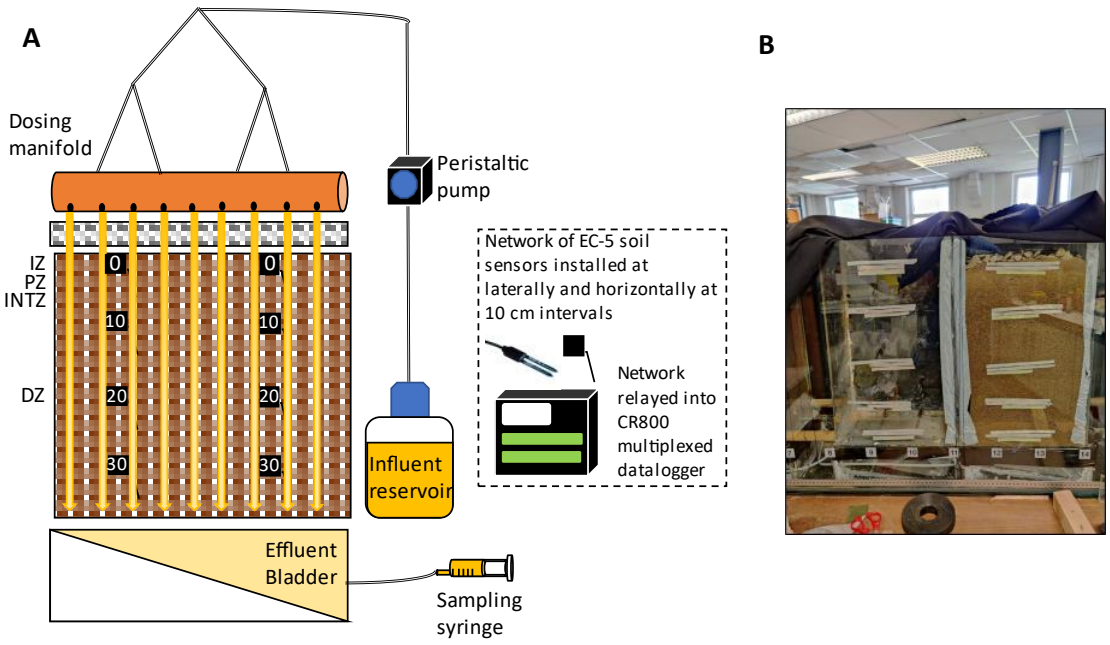


Figure SB1 Column design (a) and construction (B)

Table SB1 Methanotroph targets for sequencing analysis and source literature

Methanotroph	**Source**
Proteobacteria; Methylocystis	(Takeda ., 1988)
Proteobacteria; Methylobacter	(Donrina et al., 2004)
Proteobacteria; Methylocella	(Dunfield and Dedysh,. 2014)
Proteobacteria; Methylocaldum	(Eshinimaev et al., 2004)
Proteobacteria; Methylobacillus	(Kumar et Maitre., 2016)
Proteobacteria; Methylomonas	(Cheng et al., 2022)

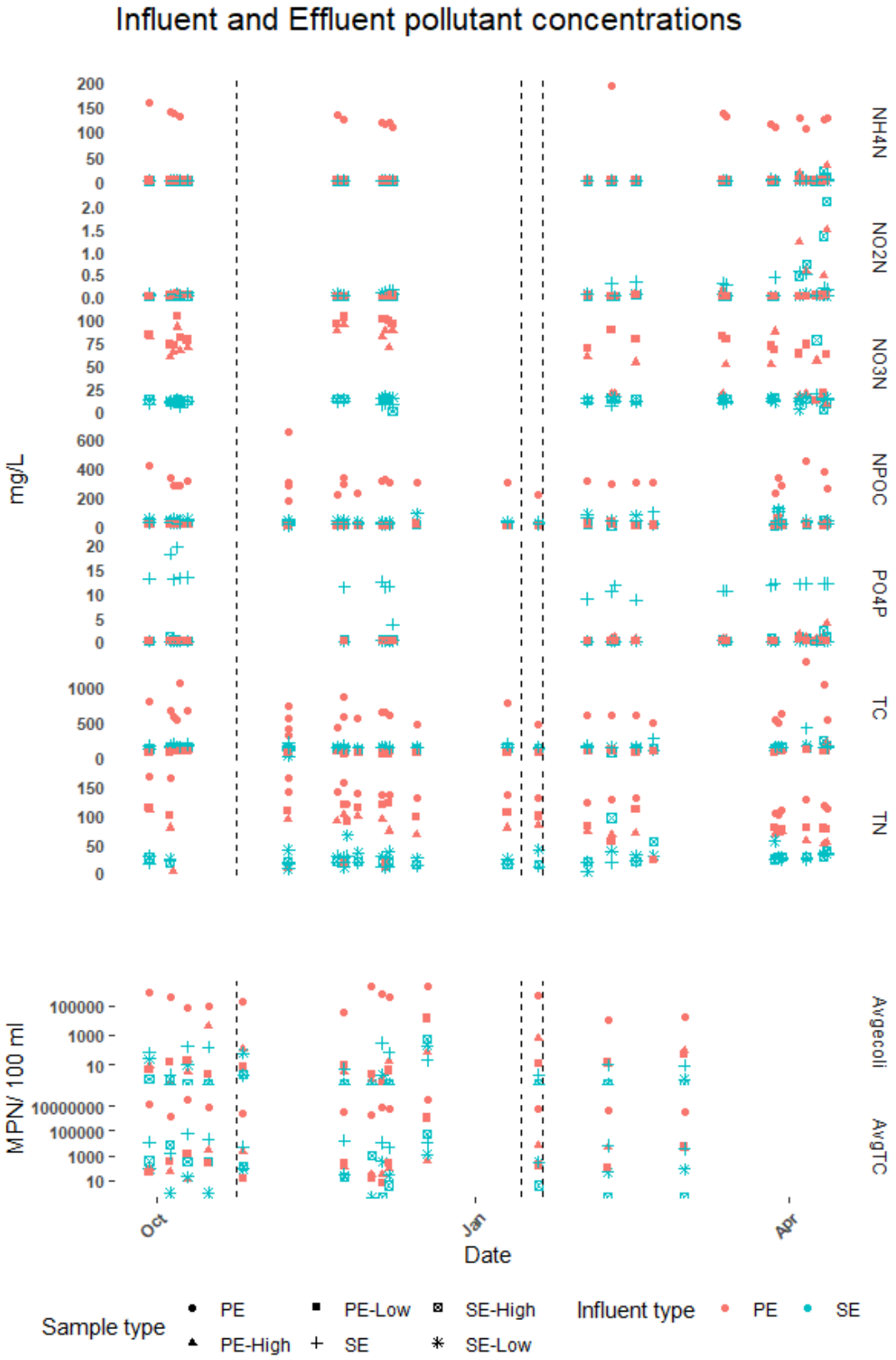


Figure SB2. Influent and Effluent concentrations for Average Total Coliforms (AvgTC) Average Escherichia Coli (Avgecoli), NH4-N, NO2-N, NO3-N, PO4-P, Non purgeable Organic Carbon (NPOC), Total Carbon (TC) and Total Nitrogen (TN).

Table SB2 Median removal efficiencies of column effluent relative to influent concentrations, and mean concentrations and standard deviations for each of effluent components

	SE-High		PE-High		SE-Low		PE-Low	
	Median removal efficiency	**Mean Concentration ± SD**	**Median removal efficiency**	**Mean Concentration ± SD**	**Median removal efficiency**	**Mean Concentration ± SD**	**Median removal efficiency**	**Mean Concentration ± SD**
AvgTC	(-) 99.43	4850.37 ± 14291.33	(-) 100.00	1509.18 ± 2155.80	(-) 99.34	184.46 ± 334.57	(-) 100.00	68304.43 ± 270556.6
Avgecoli	(-) 100.00	35.73 ± 134.01	(-) 99.99	293.93 ± 940.71	(-) 99.34	20.15 ± 52.40	(-) 100.00	886.37 ± 3506.06
NPOC	(-) 18.82	21.07 ± 14.52	(-) 97.82	8.45 ± 10.20	(+) 31.81	38.55 ± 22.77	(-) 97.29	10.38 ± 10.52
TC	(-) 13.71	136.66 ± 30.33	(-) 85.38	83.38 ± 30.90	(-) 13.01	129.51 ± 44.84	(-) 87.91	67.45 ± 29.85
PO4-P	(-) 97.00	0.43 ± 0.51	(-) 97.56	0.59 ± 0.93	(-) 99.93	0.02 ± 0.03	(-) 99.75	0.04 ± 0.06
NH4-N	(-) 95.87	2.66 ± 5.62	(-) 100.00	3.32 ± 8.31	(-) 99.14	0.06 ± 0.11	(-) 100.00	0.017 ± 0.04
NO2-N	(-) 96.32	0.30 ± 0.65	(-) 90.72	0.23 ± 0.48	(-) 84.09	0.02 ± 0.02	(-) 95.68	0.01 ± 0.01
NO3-N	(+) 20.00	14.39 ± 13.40	(+) 18639.29	37.41 ± 19.85	(+) 26.38	13.27 ± 2.86	(+) 19042.75	45.86 ± 47.03
TN	(+) 7.06	23.02 ± 15.05	(-) 40.49	68.86 ± 32.99	(+) 38.56	26.87 ± 14.83	(-) 30.97	72.64 ± 42.46

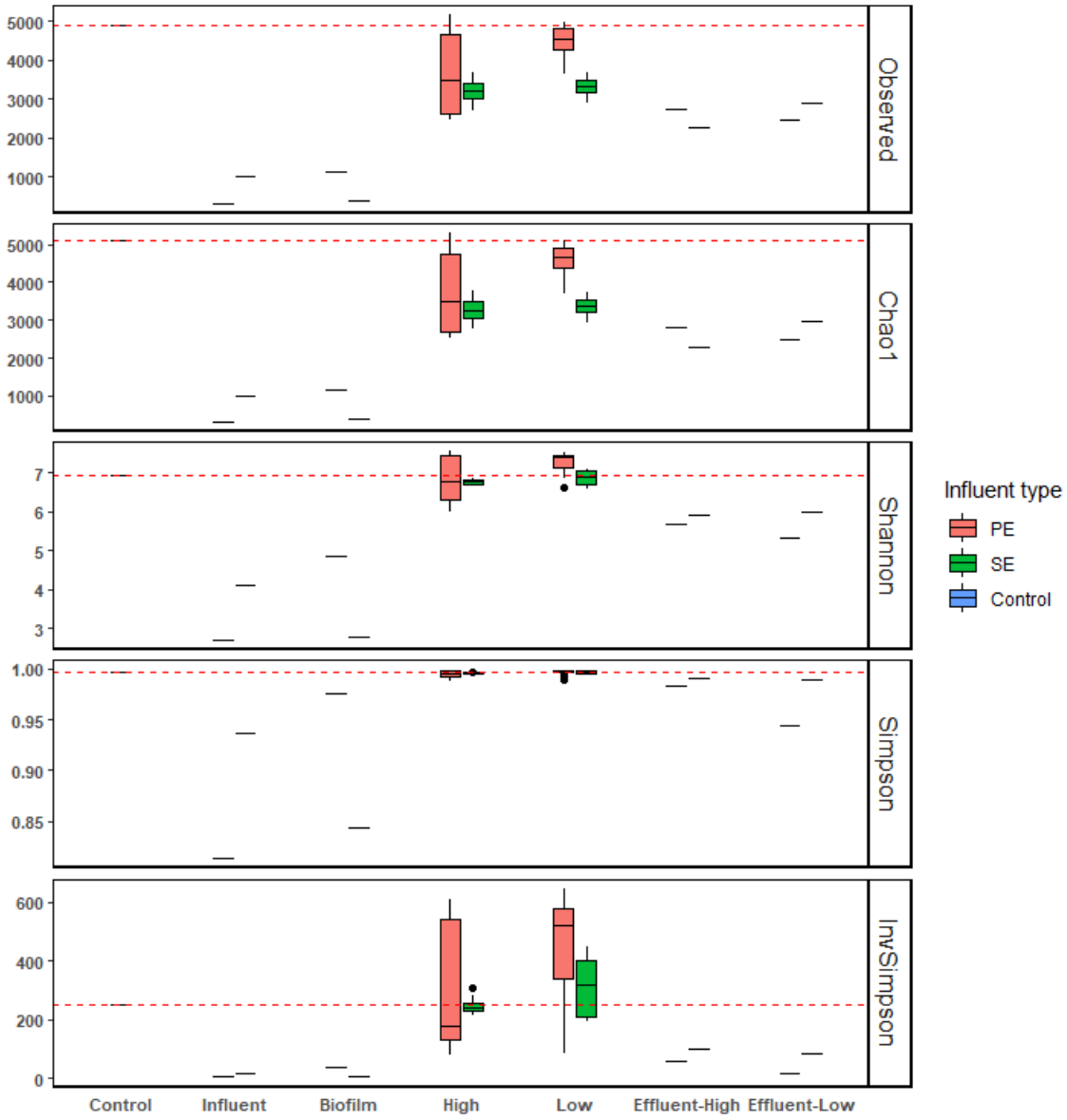


Figure SB3 Boxplots displaying rarefied data for the observed OTUs, species richness calculated using an abundance-based coverage estimates (Chao1) and alpha diversity (Shannon). Samples were aggregated on the basis of porosity (High, Low), effluent (PE, SE), and tubing biofilms (PE-BF, SE-BF))

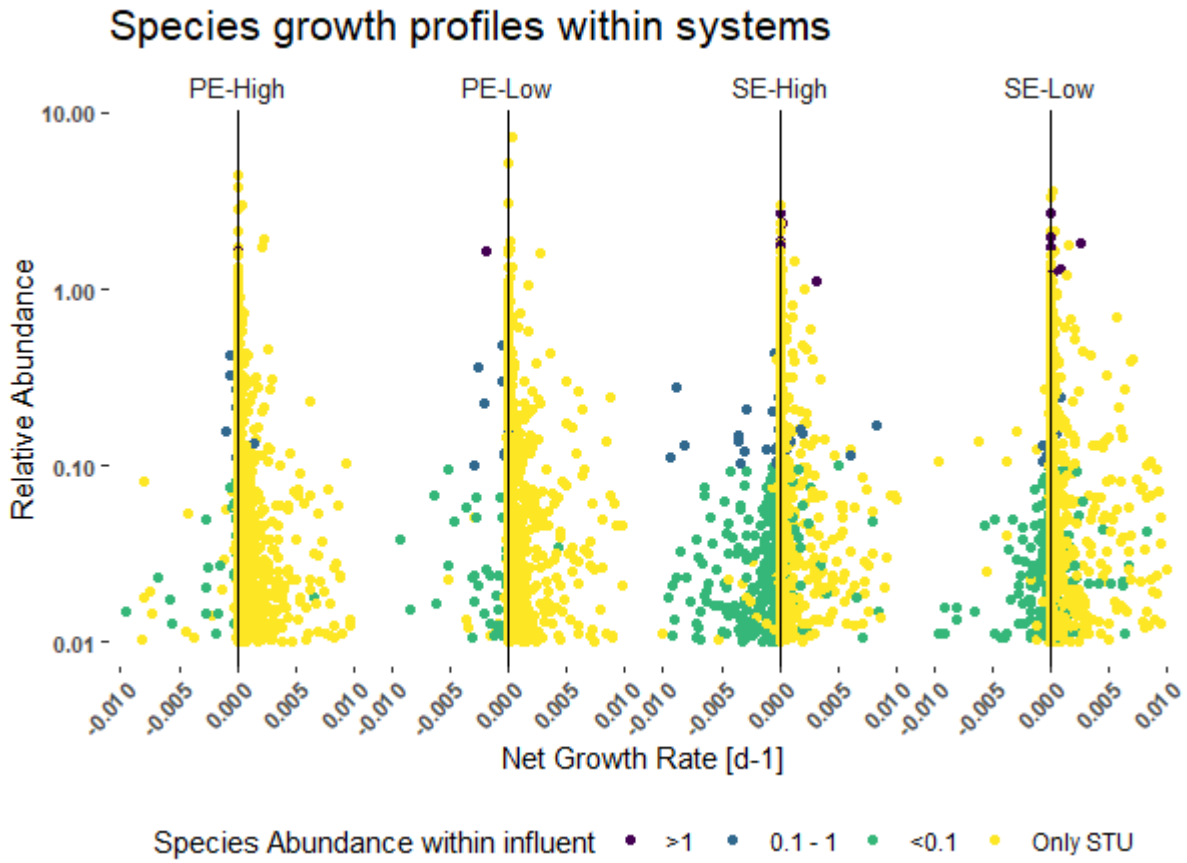


Figure SB4 Net Growth rate absent and present within wastewater across all columns PE and SE influent, High and Low porosity soil

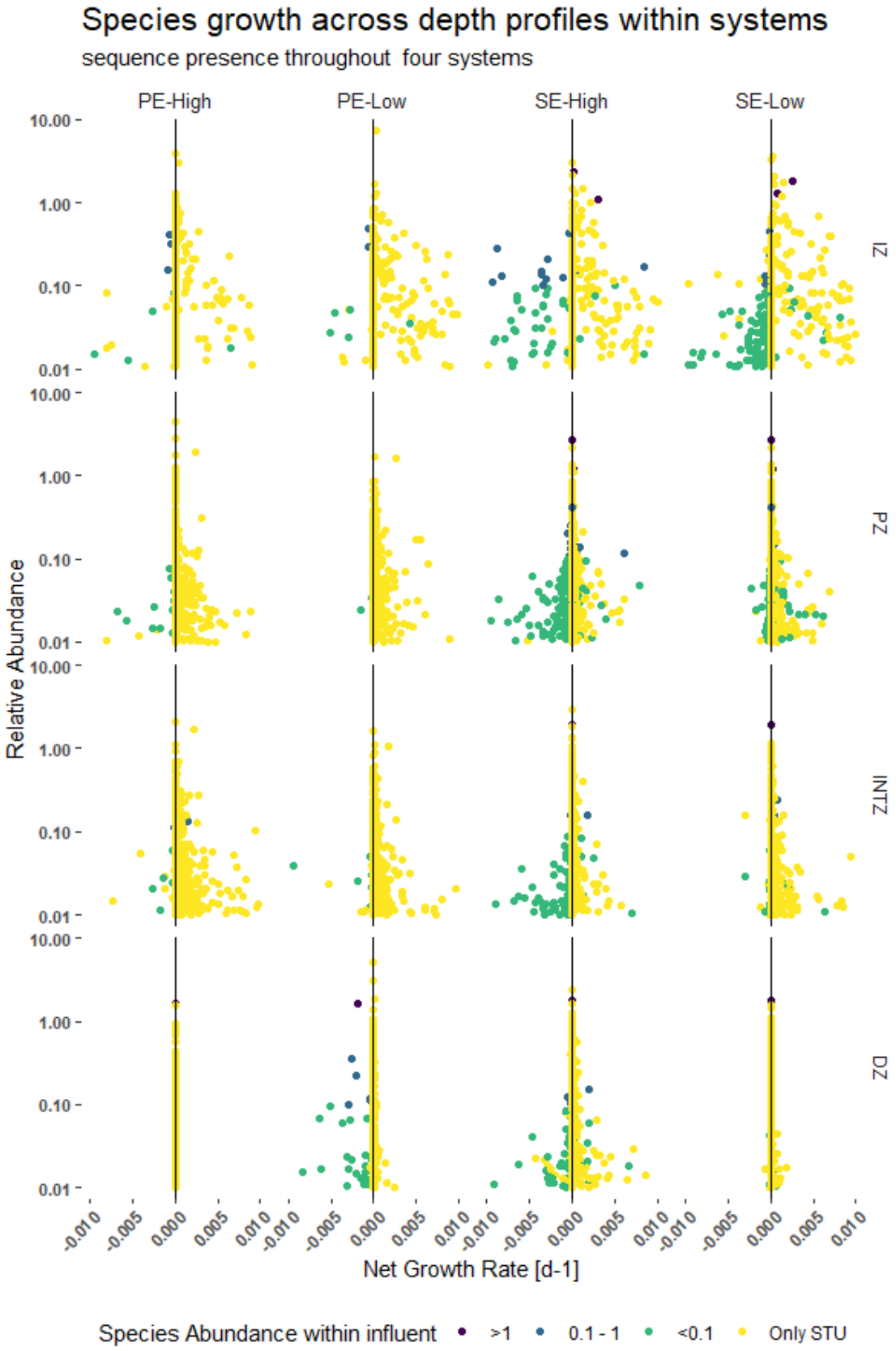


Figure SB5 Net Growth rate absent and present within wastewater across all columns depth profiles where 0 – 1 cm is the Infiltrative Zone (IZ) , 1 – 5 cm Proximal Zone, 6 -10 cm Intermediary Zone and 11 – 30 cm defined as the Deep Zone of the subsoil columns.

Table SB3 Dominant and rare growing and not growing species abundances, measured across depth profiles. Where 0 – 1 cm is the Infiltrative Zone (IZ) , 1 – 5 cm Proximal Zone, 6 -10 cm Intermediary Zone and 11 – 30 cm defined as the Deep Zone of the subsoil columns

		PE-High		PE-Low		SE-High		SE-Low	
		G	**NG**	**G**	**NG**	**G**	**NG**	**G**	**NG**
Influent	**IZ**	0.31 (13)	11.2 (35)	0.1 (3)	1.44 (19)	6.42 (83)	4.5 (73)	5.25 (76)	3.95 (67)
	PZ	0.12 (6)	4.3 (32)	0.16 (4)	0.79 (13)	6.52 (83)	4.17 (74)	6.79 (114)	0.86 (8)
	INTZ	0.16 (2)	1.11 (22)	0.22 (4)	0.99 (22)	3.08 (33)	1.79 (55)	2.62 (15)	0.75 (26)
	DZ	0.2 (1)	0.62 (15)	0.35 (11)	3.79 (31)	2.52 (24)	1.22 (35)	2.42 (13)	0.91 (31)
Only found in STU	**IZ**	80.57 (918)	0.39 (7)	88.14 (1504)	0.22 (4)	82.71 (1200)	0.25 (8)	82.71 (1288)	0.24 (14)
	PZ	85.2 (1218)	0.25 (11)	83.76 (1834)	0.03 (2)	79.11 (1242)	0.22 (14)	75.56 (1200)	6.63 (95)
	INTZ	80.89 (1874)	0.25 (10)	84.31 (1899)	0.12 (5)	83.75 (1383)	0.28 (15)	84.94 (1552)	0.27 (15)
	DZ	83.42 (1877)	0.11 (3)	83.42 (1520)	0.25 (9)	84.76 (1365)	0.18 (8)	86.07 (1555)	0.14 (7)

NG- Not growing

G - Growing

Appendix C: Supplemental to PMMoV trials

Contents

Figure SC1 Figure SC1. The location of site suitability assessment in which percolation subsoil t-tests, and surface p tests are performed on site. T tests are performed 400 mm below ground level. Each whole is carved to set dimensions, filled and time measured for the falling head of water to percolate through the soil.

Table SC1 Average Biological parameters: total coliforms, Escherichia coli, PMMoV and chemical parameters total phosphorous (PO4-P), ammonium (NH4-N), nitrite (NO2-N), nitrate (NO3-N), total nitrogen (TN), total organic carbon (TOC) and total carbon (TC)

Figure SC1. The location of site suitability assessment in which percolation subsoil t-tests, and surface p tests are performed on site. T tests are performed 400 mm below ground level. Each whole is carved to set dimensions, filled and time measured for the falling head of water to percolate through the soil.

Table SC1 Average Biological parameters: total coliforms, Escherichia coli, PMMoV and chemical parameters total phosphorous (PO4-P), ammonium (NH4-N), nitrite (NO2-N), nitrate (NO3-N), total nitrogen (TN), total organic carbon (TOC) and total carbon (TC)

Rivers		**Cl2 (mg/L)**	**PO4-P (mg/L)**	**NH4-N (mg/L)**	**NO2-N (mg/L)**	**NO3-N (mg/L)**	**TN (mg/L)**	**TC (mg/L)**	**TOC (mg/L)**	**Tot Coliforms (mpn / 100 ml)**	**Ecoli (mpn / 100 ml)**	**PPMOV (GC/100ml)**
C1	**Mean ± SD**	20.07 ± 2.16	0.04 ± 0.02	0.03 ± 0.01	0 ± 0	2.14 ± 0.07	6.61 ± 0.11	26.64 ± 0.58	2.44 ± 0.42	17719.93 ± 7231.15	2257.18 ± 1434.51	7558.33 ± 3632.04
	Variance %	10.76	56.02	45.22	0	3.27	1.69	2.19	17.18	40.81	63.55	48.05
C2	**Mean ± SD**	77.08 ± 1.21	0.18 ± 0.01	0.08 ± 0.03	0.02 ± 0	0.64 ± 0.07	4.95 ± 0.2	106.41 ± 1.47	9.24 ± 8.51	407737.08 ± 593806.19	853.77 ± 1176.65	21950 ± 15549.9
	Variance %	137.82	70.84	16.64	11.61	3.15	92.08	4.09	145.63		40.12	1.38
C3	**Mean ± SD**	21.64 ± 4.31	0.08 ± 0.04	0.01 ± 0.01	0 .01± 0.01	6.13 ± 0.12	2.34 ± 0.47	50.87 ± 3.43	9.58 ± 0.2	3449500.42 ± 10662073.1	10243.8 ± 29669	4151.67 ± 4050.63
	Variance %	19.92	46.63	101.18	101.18	2.03	19.9	6.73	2.14	309.09	289.63	97.57

Appendix D: Supplemental to Temporal Variation Study

Contents

Figures:

- **Figure SD1** Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between January 2016 to August 2018, sites A and B.
- **Figure SD2** Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between May 2020 to August to November 2022, for sites A and B.
- **Figure SD3** Total monthly precipitation, mean monthly air temperatures, mean daily subsoil temperature and monthly actual evapotranspiration for both sites between July 2020 to August to September 2023, for sites C, Met Eireann Oak park monitoring station, and an approximation of values based on monthly variations. Mean subsoil temperatures are monthly averages based on existing site data.
- **Figure SD4.** Mean ± SD of relative read abundance (%) of top 10 Class for site C for both the Primary (PE) STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 17 m over days in operation.
- **Figure SD5.** Mean ± SD of relative read abundance (%) of top 10 class for site C for secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 17 m, across days in operation.
- **Figure SD6.** Mean ± SD of relative read abundance (%) of top 20 Species for site C for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m.
- **Figure SD7** Mean of relative read abundance (%) of class-level analysis for site A in 2018 and 2022 for both the Primary (PE), secondary (SE), and control samples for each 'system': coconut husk filter, control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end (e) at 17 m.
- **Figure SD8** Mean ± SD of relative read abundance (%) of class-level analysis for site B in 2018 and 2022 for both the Primary (PE), secondary (SE), and control samples for each 'system': RBC effluent (RBC); RBC disk 1 and 2 (RBC-1; RBC-2) control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end (e) at 17 m.
- **Figure SD9**. The Shannon diversity (A) and Chao1 richness (B) across distance and depth of the sites A and B primary (PE) and secondary effluent (SE) STUs for years 2018 and 2022.
- **Figure SD10** Mean concentration (mg/l) of ammonium (NH4-N), nitrate, orthophosphate (PO4-P), total nitrogen (TN) and non-purgeable organic carbon (NPOC) , total carbon (TC), nitrate (NO3-N), nitrite (NO2-N), total oxidized nitrogen (TON) within the porewater samples at site C, for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m throughout the years of sampling monitoring.
- **Figure SD11** Mean concentration (mg/l) of ammonium (NH4-N), nitrate, orthophosphate (PO4-P), total nitrogen (TN) and non-purgeable organic carbon (NPOC) , total carbon (TC), nitrate (NO3-N), nitrite (NO2-N), total oxidized nitrogen (TON) within the porewater samples at site C, for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m, seasonal values throughout the years of sampling monitoring.

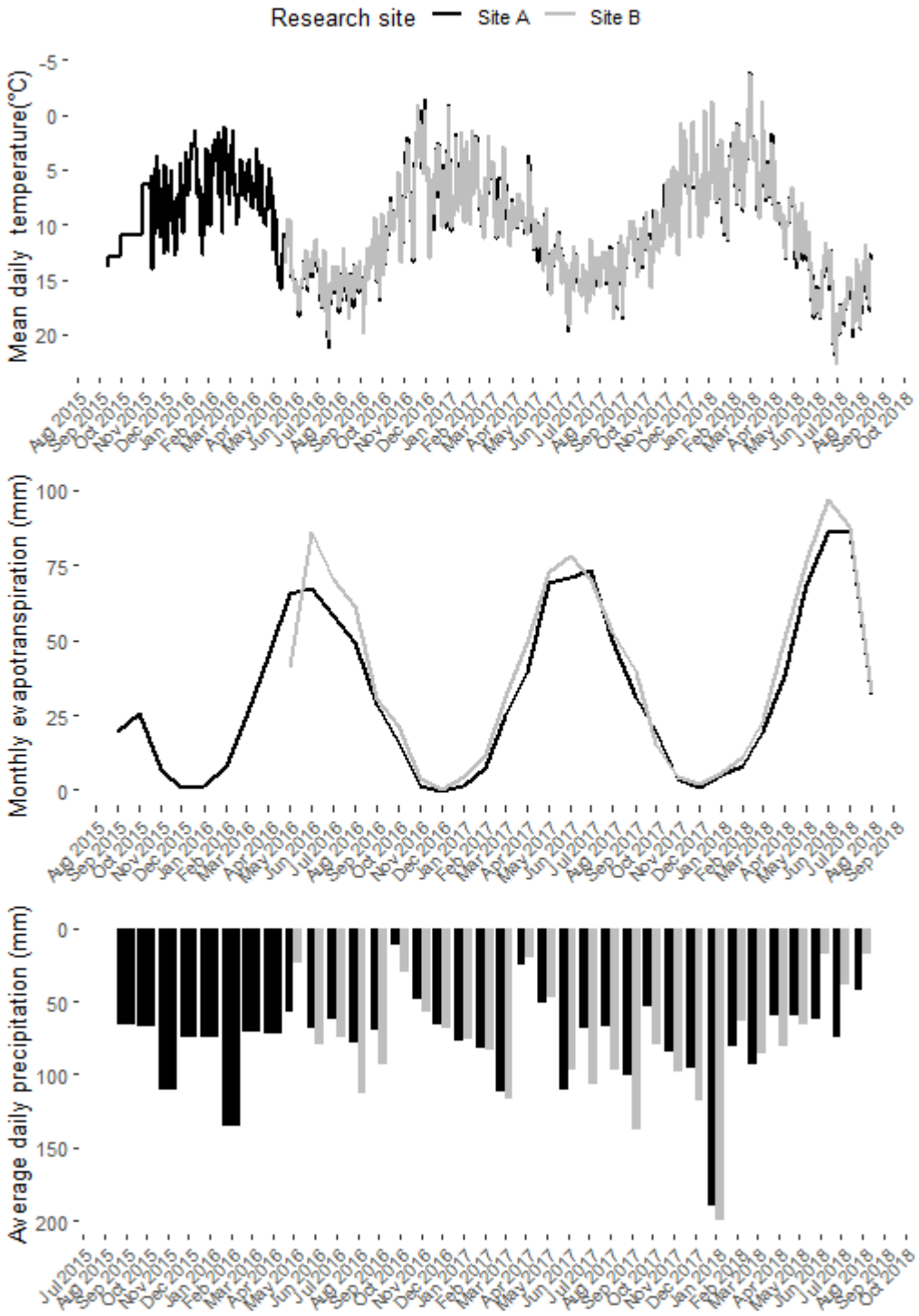


Figure SD1 Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between January 2016 to August 2018, sites A and B.

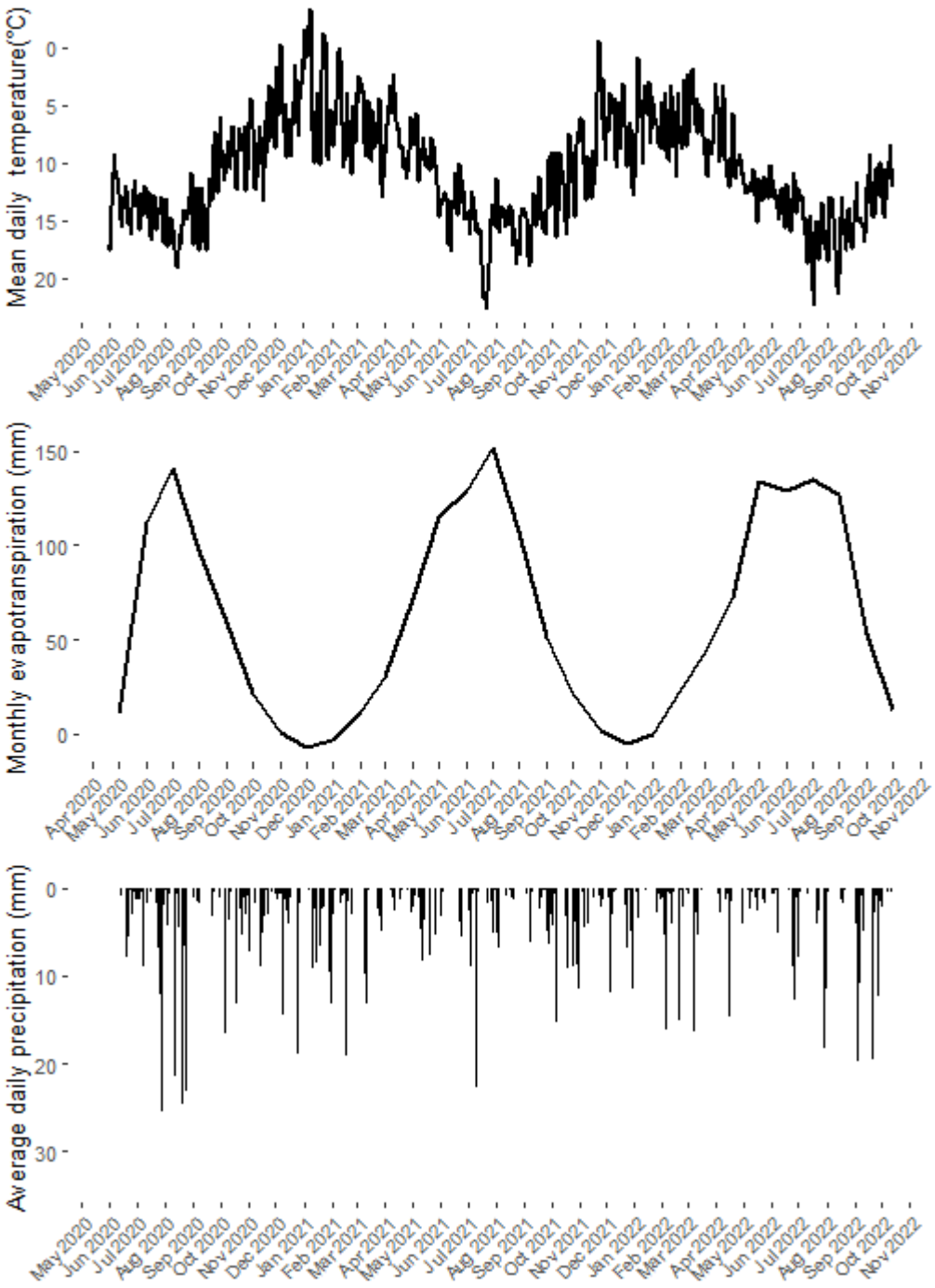


Figure SD2 Total monthly precipitation, mean monthly air temperatures and monthly actual evapotranspiration for both sites between May 2020 to August to November 2022, for sites A and B.

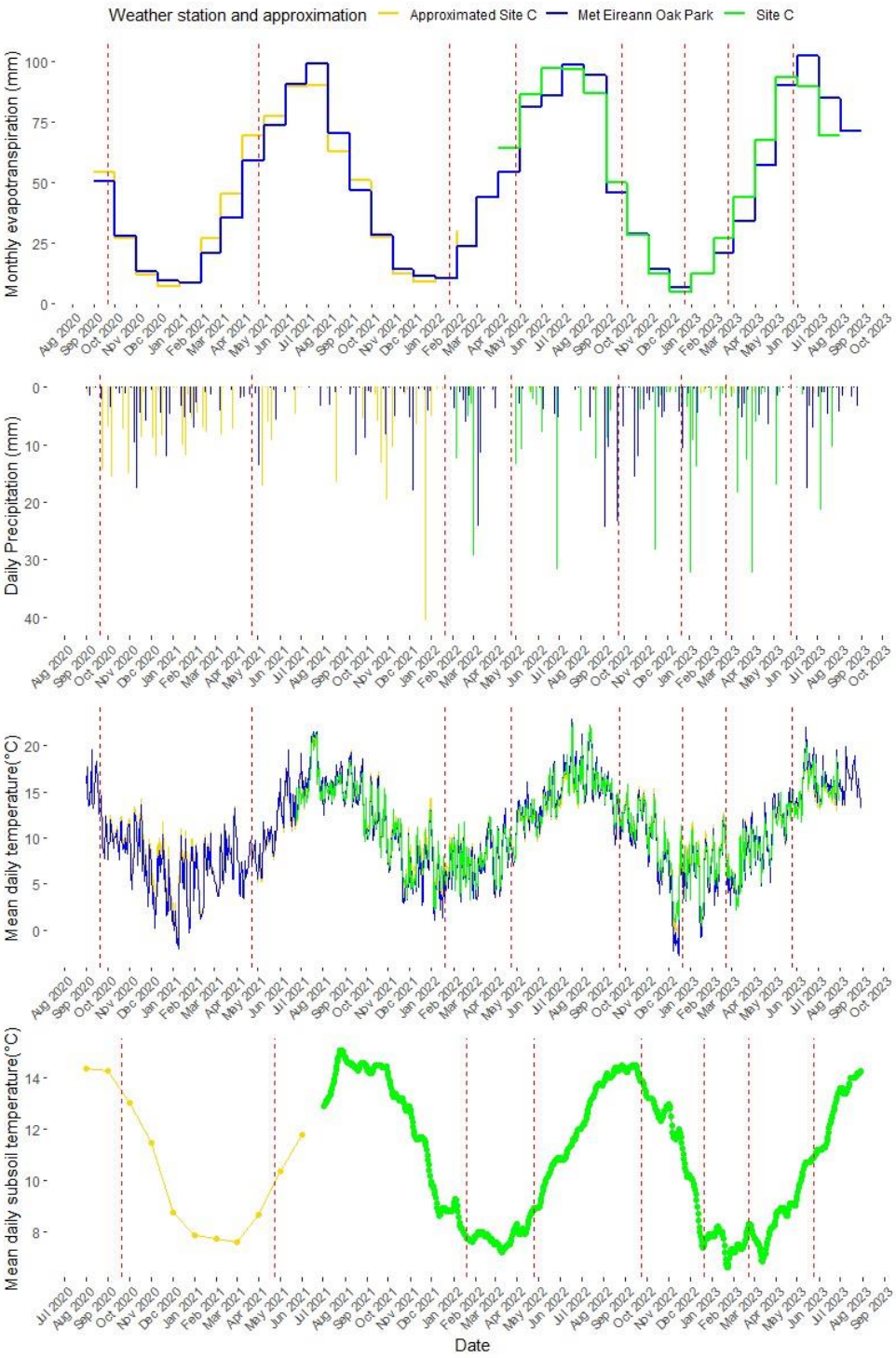


Figure SD3 Total monthly precipitation, mean monthly air temperatures, mean daily subsoil temperature and monthly actual evapotranspiration for both sites between July 2020 to August to September 2023, for sites C, Met Eireann Oak park monitoring station, and an approximation of values based on monthly variations. Mean subsoil temperatures are monthly averages based on existing site data. Red dash lines mark each of the genomic sampling days of the experiment.

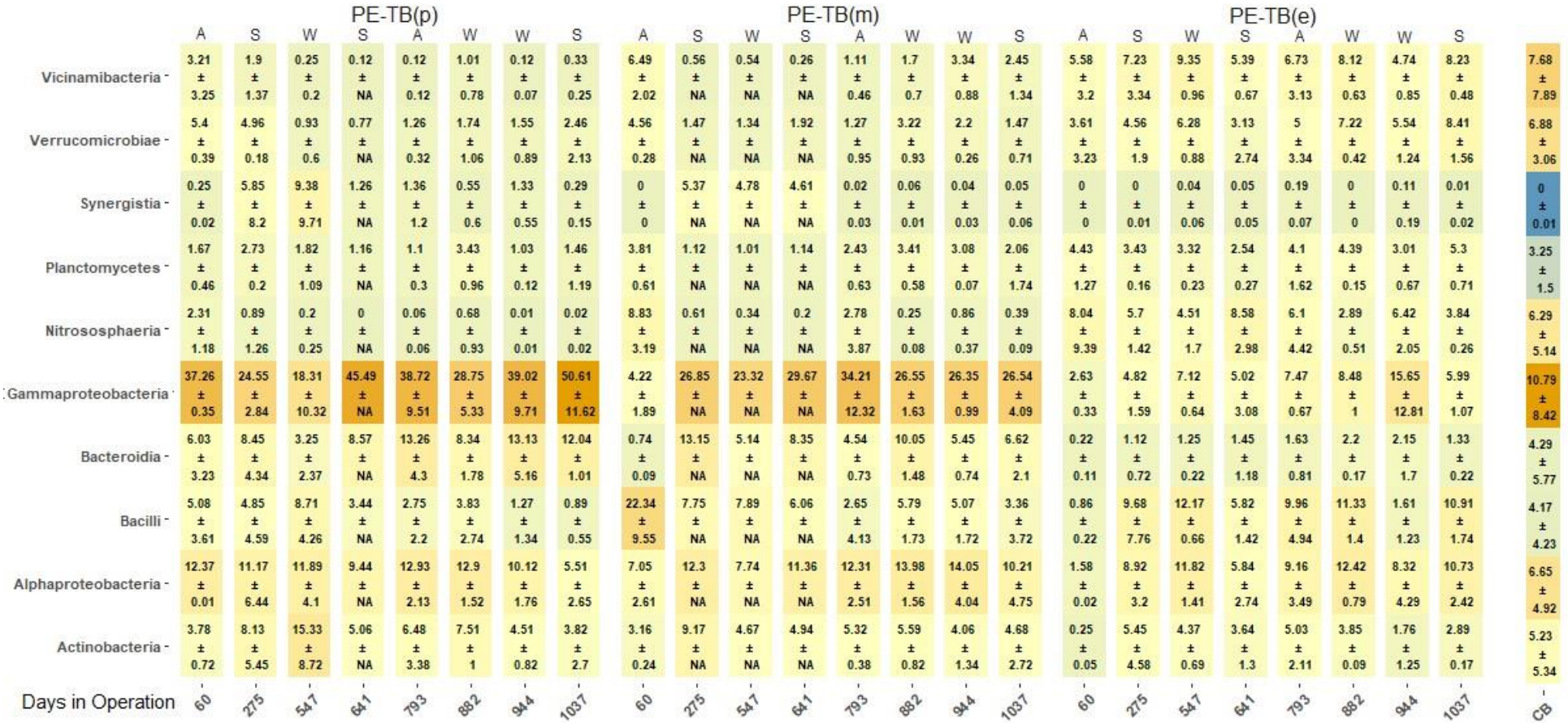


Figure SD4. Mean ± SD of relative read abundance (%) of top 10 Class for site C for both the Primary (PE) and control samples (CB) for each 'system': STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 17 m over days in operation through seasons Autumn (A), Spring (S) and Winter

(W).

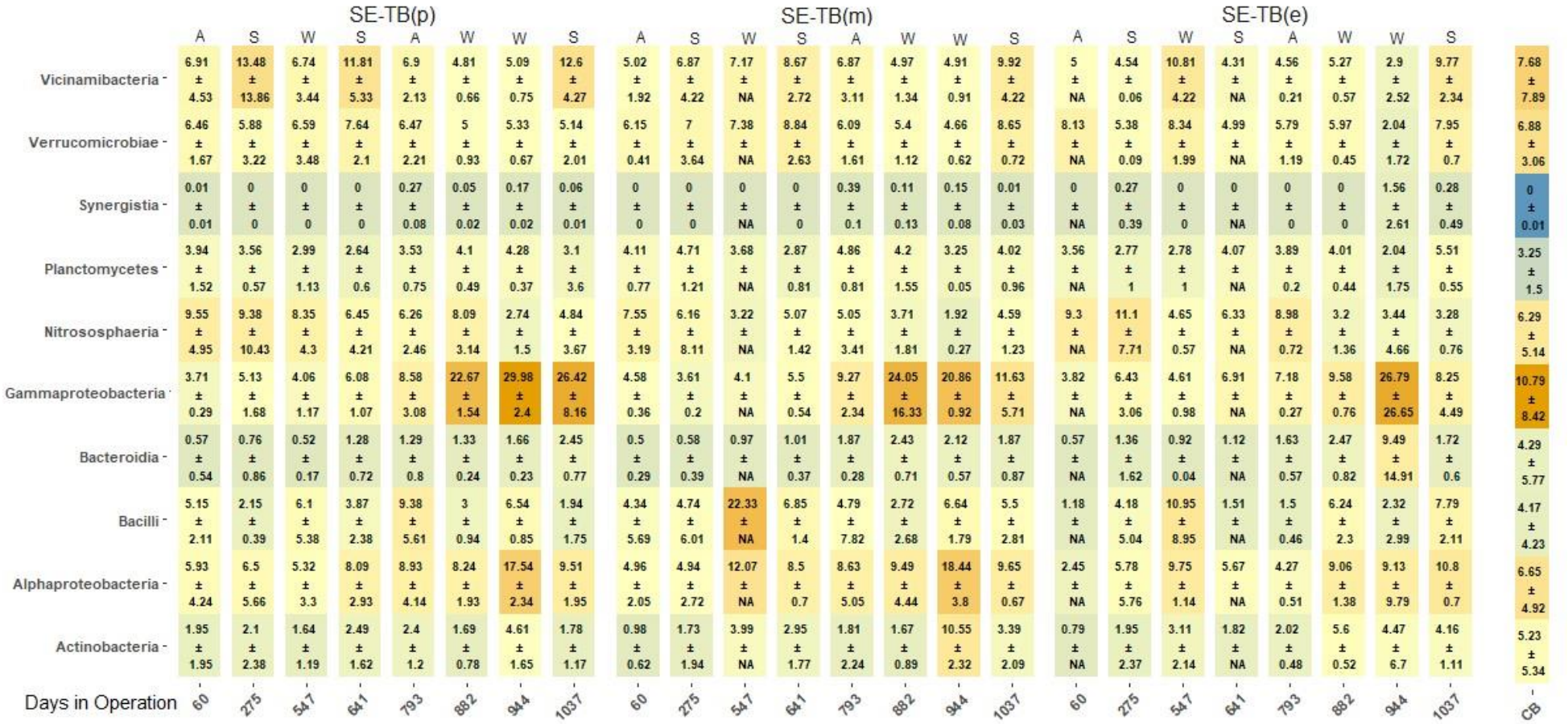


Figure SD5. Mean ± SD of relative read abundance (%) of top 10 Class for site C for both the secondary (SE), and control samples (CB) for each 'system': STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and distal (d) at 17 m over days in operation through seasons Autumn (A), Spring (S) and Winter (W).

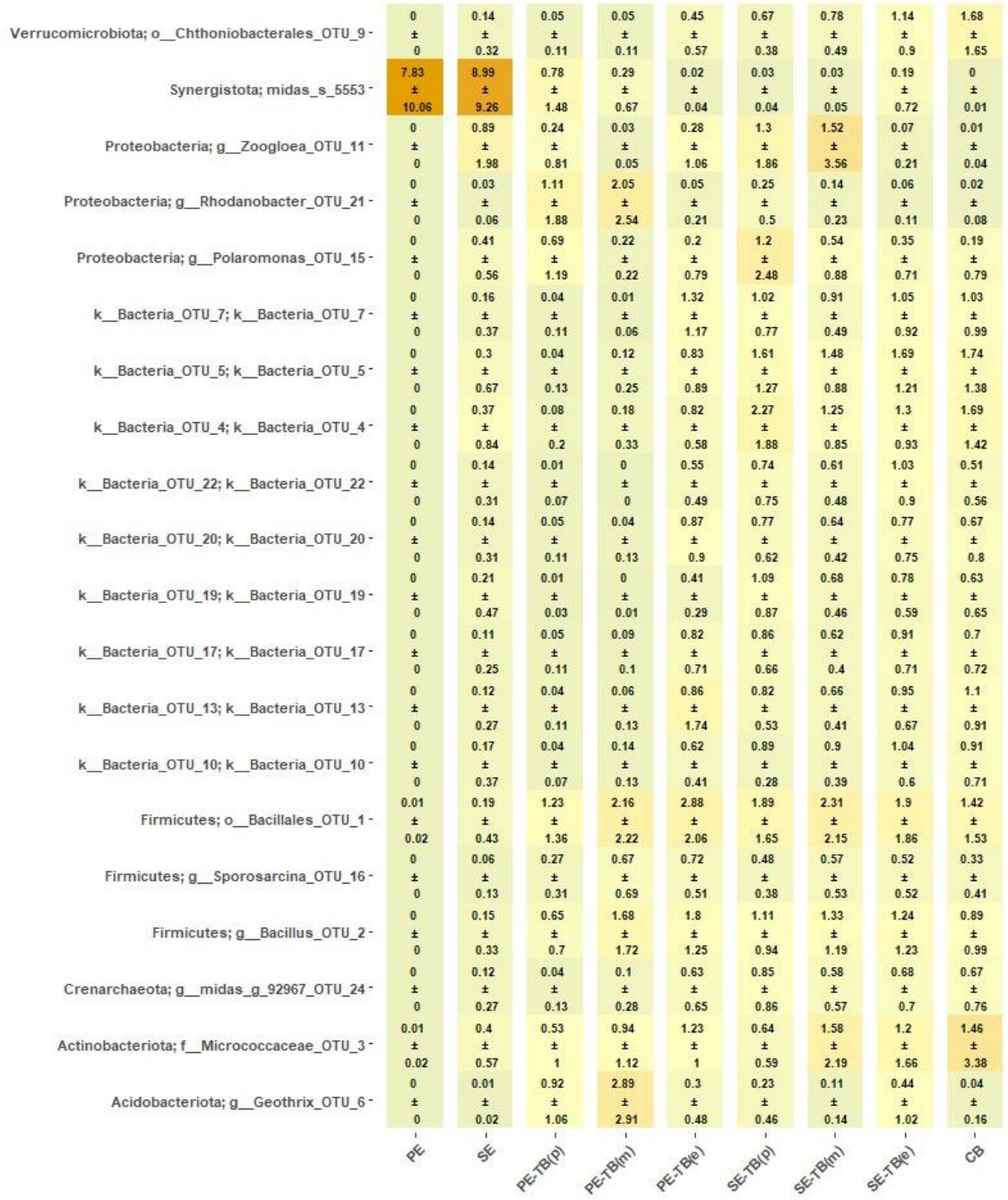


Figure SD6. Mean ± SD of relative read abundance (%) of top 20 Species for site C for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m.

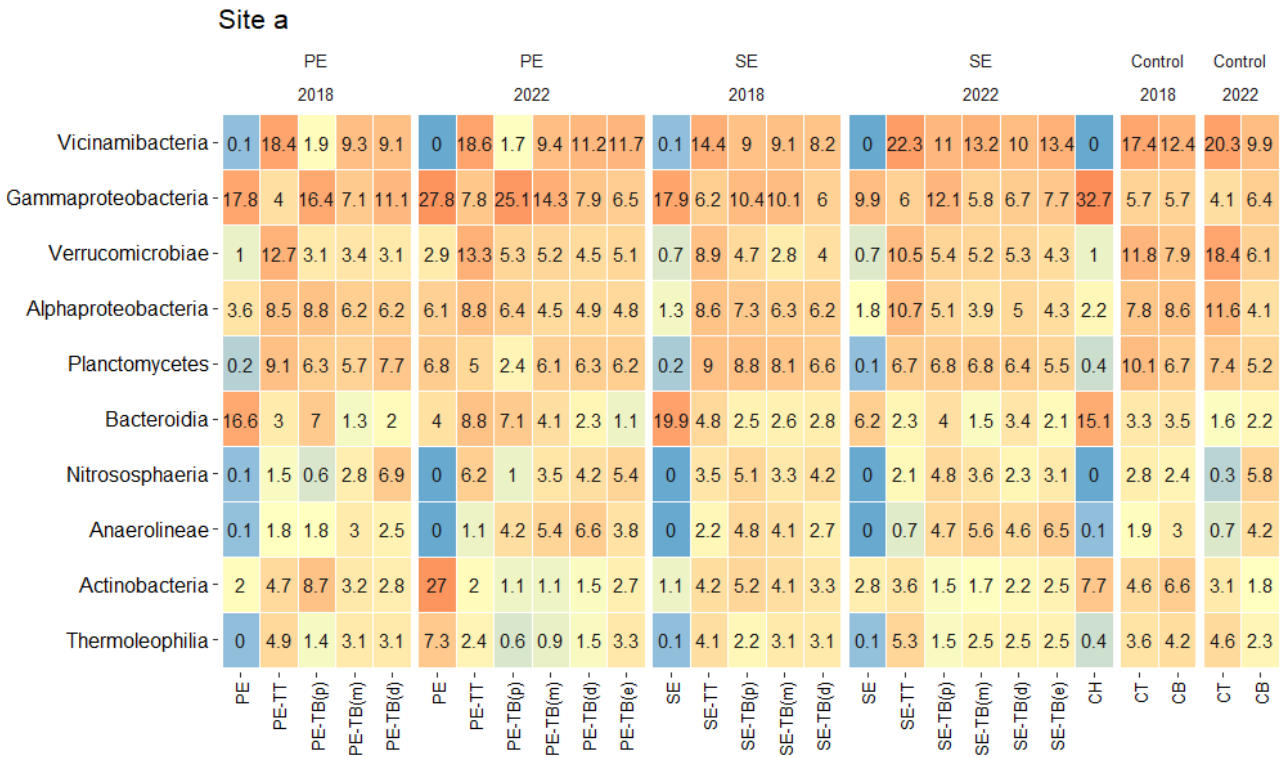


Figure SD7 Mean of relative read abundance (%) of class-level analysis for site A in 2018 and 2022 for both the Primary (PE), secondary (SE), and control samples for each 'system': coconut husk filter, control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end (e) at 17 m.

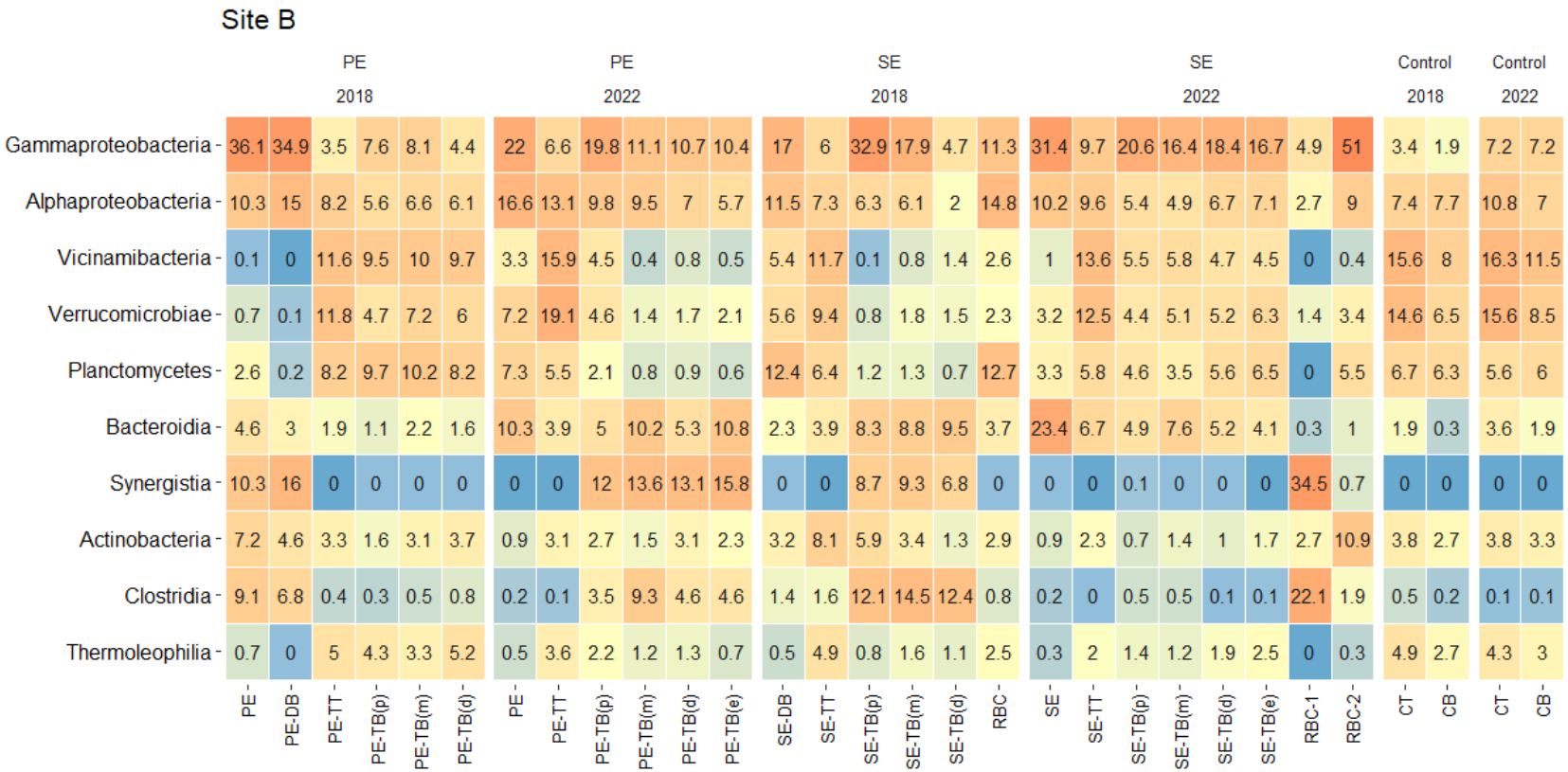


Figure SD8 Mean of relative read abundance (%) of Class-level analysis for site B in 2018 and 2022 for both the Primary (PE), secondary (SE), and control samples for each 'system': RBC effluent (RBC); RBC disk 1 and 2 (RBC-1; RBC-2) control; base (CB), top (CT), STU topsoil (TT) and STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m, distal (d) at 12 m and end (e) at 17 m.

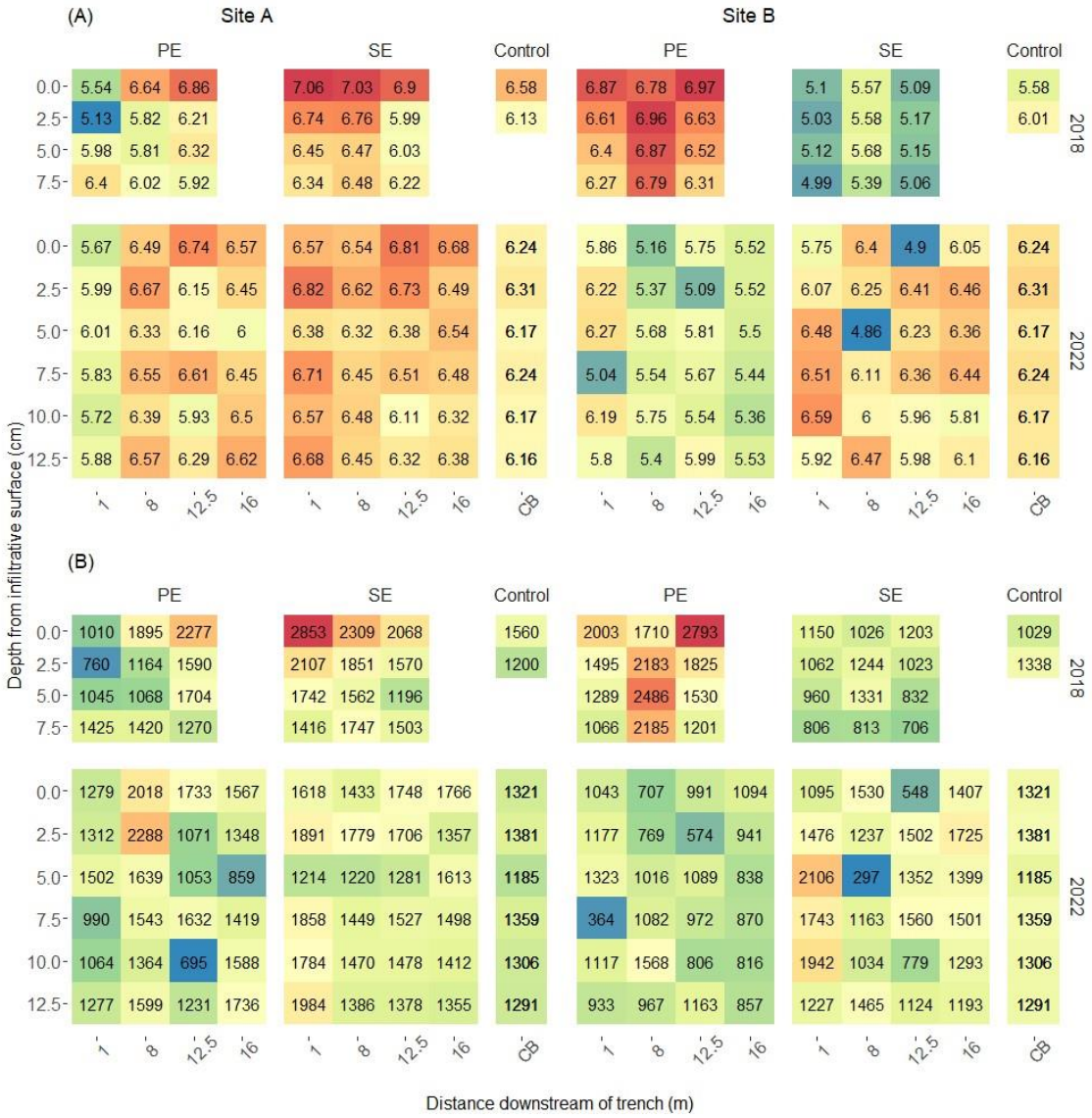


Figure SD9. The Shannon diversity (A) and Chao1 richness (B) across distance and depth of the sites A and B primary (PE) and secondary effluent (SE) STUs for years 2018 and 2022.

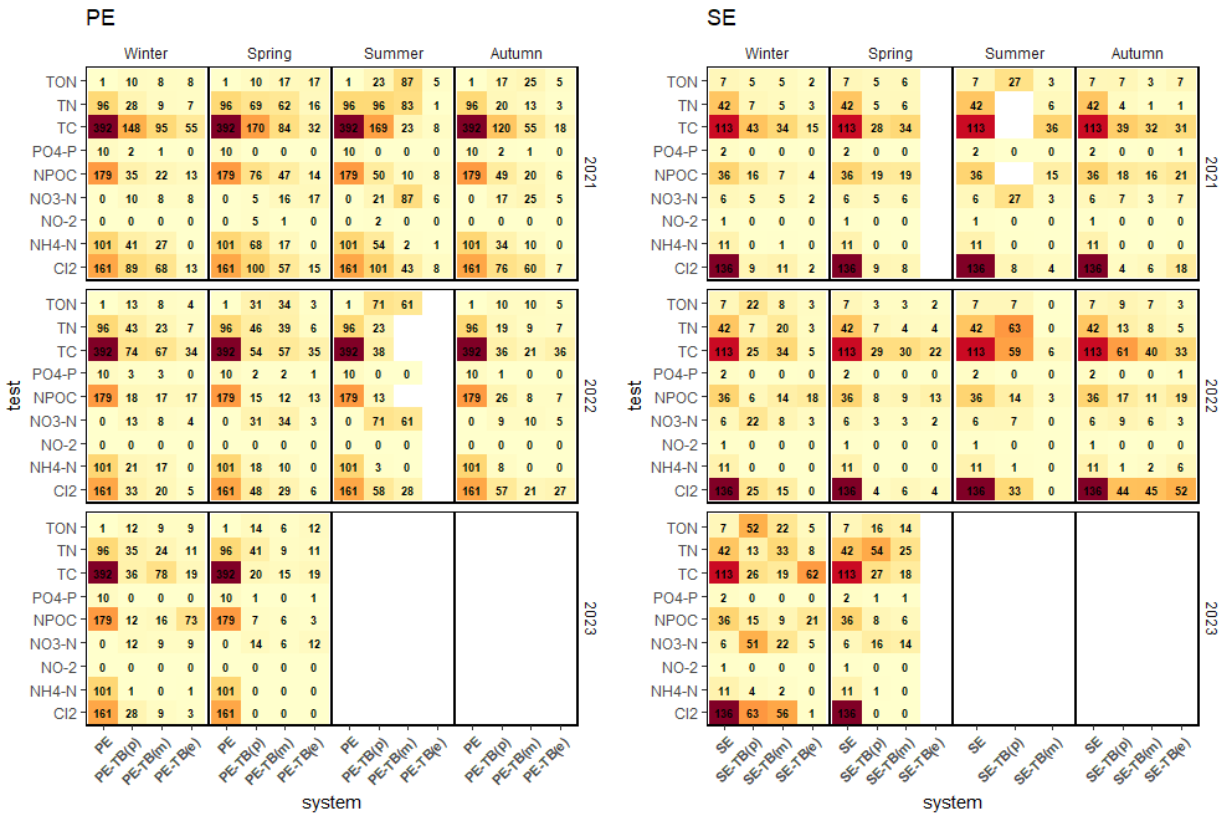


Figure SD10 Mean concentration (mg/l) of ammonium (NH4-N), nitrate, orthophosphate (PO4-P), total nitrogen (TN) and non-purgeable organic carbon (NPOC) , total carbon (TC), nitrate (NO3-N), nitrite (NO2-N), total oxidized nitrogen (TON) within the porewater samples at site C, for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m throughout the years of sampling monitoring.

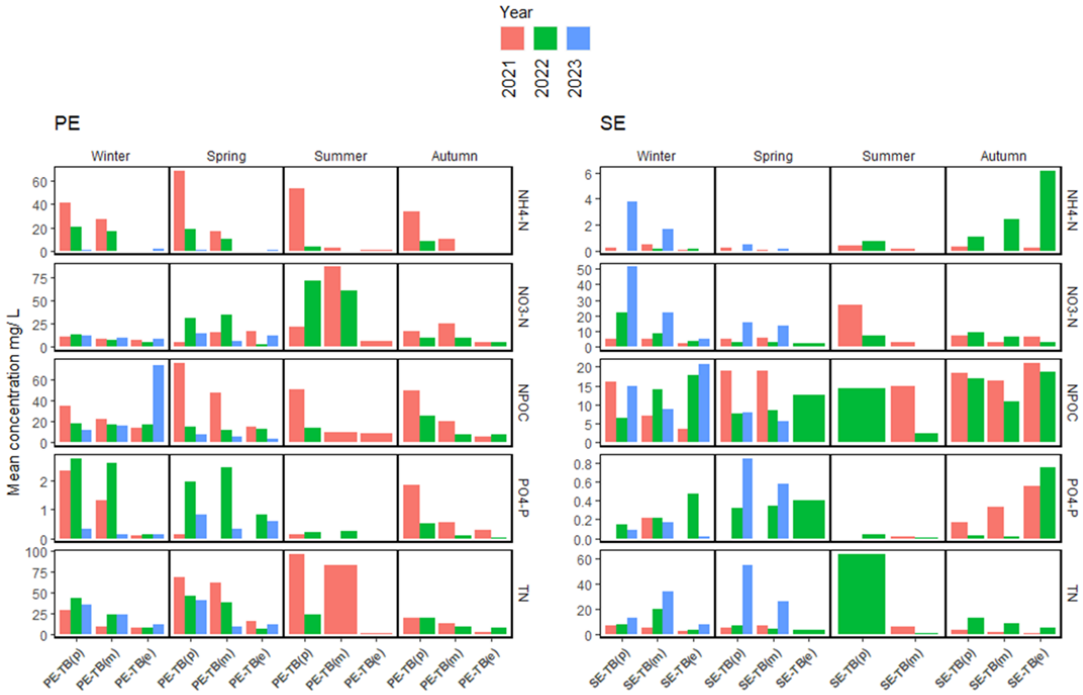


Figure SD11 Mean concentration (mg/l) of ammonium (NH4-N), nitrate, orthophosphate (PO4-P), total nitrogen (TN) and non-purgeable organic carbon (NPOC) , total carbon (TC), nitrate (NO3-N), nitrite (NO2-N), total oxidized nitrogen (TON) within the porewater samples at site C, for both the Primary (PE), secondary (SE), and control samples for each 'system': control base (CB), STU subsoil 'trench base' (TB). STU base is further divided into proximal (P) at 1 m, midpoint (m) at 5 m and end (e) at 17 m, seasonal values throughout the years of sampling monitoring.

Bibliography

Achenbach, L.A., Michaelidou, U., Bruce, R.A., Fryman, J. and Coates, J.D., 2001. Dechloromonas agitata gen. nov., sp. nov. and Dechlorosoma suillum gen. nov., sp. nov., two novel environmentally dominant (per) chlorate-reducing bacteria and their phylogenetic position. International journal of systematic and evolutionary microbiology, 51(2), pp.527-533.

Adessi, A., de Carvalho, R.C., De Philippis, R., Branquinho, C. and da Silva, J.M., 2018. Microbial extracellular polymeric substances improve water retention in dryland biological soil crusts. Soil Biology and Biochemistry, 116, pp.67-69.

AL-Gheethi, A.A., Ismail, N., Lalung, J., Talib, A., Efaq, A.N. and Kadir, M.O.A., 2013. Susceptibility for antibiotics among faecal indicators and pathogenic bacteria in sewage treated effluents. Water Practice and Technology, 8(1), pp.1-6.

Ali, M. and Okabe, S., 2015. Anammox-based technologies for nitrogen removal: advances in process start-up and remaining issues. Chemosphere, 141, pp.144-153.

Ali, M., Wang, Z., Salam, K.W., Hari, A.R., Pronk, M., van Loosdrecht, M.C. and Saikaly, P.E., 2019. Importance of species sorting and immigration on the bacterial assembly of different-sized aggregates in a full-scale aerobic granular sludge plant. Environmental science & technology, 53(14), pp.8291-8301.Amador, J.A. and Loomis, G., 2019 Soil-based wastewater treatment (Vol. 180). John Wiley & Sons.

Alonso, E., García-Luque, I., De la Cruz, A., Wicke, B., Avila-Rincon, M.J., Serra, M.T., Castresana, C. and Díaz-Ruíz, J.R., 1991. Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistance-breaking tobamovirus in pepper. Journal of General Virology, 72(12), pp.2875-2884.

Amador, J.A. and Loomis, G., 2020. Soil-based wastewater treatment (Vol. 180). John Wiley & Sons.

Andrade, L., O'Dwyer, J., O'Neill, E. and Hynds, P., 2018. Surface water flooding, groundwater contamination, and enteric disease in developed countries: A scoping review of connections and consequences. Environmental pollution, 236, pp.540-549.

Bagnoud, A., Guye-Humbert, S., Schloter-Hai, B., Schloter, M. and Zopfi, J., 2020. Environmental factors determining distribution and activity of anammox bacteria in minerotrophic fen soils. *FEMS microbiology ecology*, *96*(2), p.fiz191.

Bain, R., Johnston, R. and Slaymaker, T., 2020. Drinking water quality and the SDGs. npj Clean Water, 3(1), p.37.

Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R. and Wolfe, R., 1979. Methanogens: reevaluation of a unique biological group. Microbiological reviews, 43(2), pp.260-296.

Barba, C., Folch, A., Gaju, N., Sanchez-Vila, X., Carrasquilla, M., Grau-Martínez, A. and Martínez-Alonso, M., 2019. Microbial community changes induced by Managed Aquifer Recharge activities: linking hydrogeological and biological processes. Hydrology and Earth system sciences, 23(1), pp.139-154.

Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.P., Clément, C., Ouhdouch, Y. and van Wezel, G.P., 2016. Taxonomy, physiology, and natural products of Actinobacteria. Microbiology and molecular biology reviews, 80(1), pp.1-43.

Bastida, F., Eldridge, D.J., García, C., Kenny Png, G., Bardgett, R.D. and Delgado-Baquerizo, M., 2021. Soil microbial diversity–biomass relationships are driven by soil carbon content across global biomes. *The ISME Journal*, *15*(7), pp.2081-2091.

Bastida, F., Torres, I.F., Abadía, J., Romero-Trigueros, C., Ruiz-Navarro, A., Alarcón, J.J., García, C. and Nicolás, E., 2018. Comparing the impacts of drip irrigation by freshwater and reclaimed wastewater on the soil microbial community of two citrus species. *Agricultural Water Management*, *203*, pp.53-62.

Bastida, F., Torres, I.F., Romero-Trigueros, C., Baldrian, P., Větrovský, T., Bayona, J.M., Alarcón, J.J., Hernández, T., García, C. and Nicolás, E., 2017. Combined effects of reduced irrigation

and water quality on the soil microbial community of a citrus orchard under semi-arid conditions. *Soil Biology and Biochemistry*, *104*, pp.226-237.

Beal, C.D., Gardner, E.A. and Menzies, N.W., 2005. Process, performance, and pollution potential: A review of septic tank–soil absorption systems. *Soil Research*, *43*(7), pp.781-802.

Beal, C.D., Gardner, E.A., Vieritz, A. and Menzies, N.W., 2004. The role of the biomat in the sustainable performance of soil absorption systems in Australia: A review. In On-Site Wastewater Treatment X, 21-24 March 2004 (p. 1). American Society of Agricultural and Biological Engineers.

Becerra-Castro, C., Lopes, A.R., Vaz-Moreira, I., Silva, E.F., Manaia, C.M. and Nunes, O.C., 2015. Wastewater reuse in irrigation: A microbiological perspective on implications in soil fertility and human and environmental health. *Environment international*, *75*, pp.117-135.

Bell G. The distribution of abundance in neutral communities. Am Nat. 2000;155(5):606–17.

Bell, T.H. and Bell, T., 2021. Many roads to bacterial generalism. FEMS microbiology ecology, 97(1), p.fiaa240.

Besemer, K., Peter, H., Logue, J.B., Langenheder, S., Lindström, E.S., Tranvik, L.J. and Battin, T.J., 2012. Unraveling assembly of stream biofilm communities. The ISME journal, 6(8), pp.1459-1468.

Betancourt, W.Q., Schijven, J., Regnery, J., Wing, A., Morrison, C.M., Drewes, J.E. and Gerba, C.P., 2019. Variable non-linear removal of viruses during transport through a saturated soil column. Journal of contaminant hydrology, 223, p.103479.

Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A.B., Brown, J. and Buttiglieri, G., 2020. Wastewater-based epidemiology: global collaborative to maximize contributions in the fight against COVID-19.

Böer, S.I., Hedtkamp, S.I., Van Beusekom, J.E., Fuhrman, J.A., Boetius, A. and Ramette, A., 2009. Time-and sediment depth-related variations in bacterial diversity and community structure in subtidal sands. The ISME journal, 3(7), pp.780-791.

Bouma J, Baker FG, Veneman PLM. Measurement of water movement in soil pedons above the water table. Information circular nr. 27. Madison, WI: University of Wisconsin; 1974.

Bouma, J., 1975. Unsaturated flow during soil treatment of septic tank effluent. Journal of the Environmental Engineering Division, 101(6), pp.967-983.

Bouwer H. Elements of soil science and groundwater hydrology. In: Bitton G, Gerba CP, editors. Ground water pollution microbiology. New York, NY: Wiley; 1984. p. 9–38.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry, 72(1-2), pp.248-254.

Brown, C.J., Starn, J.J., Stollenwerk, K.G., Mondazzi, R.A. and Trombley, T.J., 2009. *Aquifer chemistry and transport processes in the zone of contribution to a public-supply well in Woodbury, Connecticut, 2002-06* (No. 2009-5051). US Geological Survey.

Buckley, D.H., Huangyutitham, V., Nelson, T.A., Rumberger, A. and Thies, J.E., 2006. Diversity of Planctomycetes in soil in relation to soil history and environmental heterogeneity. *Applied and Environmental Microbiology*, *72*(7), pp.4522-4531.

Bünemann, E.K., Marschner, P., McNeill, A.M. and McLaughlin, M.J., 2007. Measuring rates of gross and net mineralisation of organic phosphorus in soils. Soil Biology and Biochemistry, 39(4), pp.900-913.

Cai, L., Ju, F. and Zhang, T., 2014. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Applied microbiology and biotechnology*, *98*, pp.3317-3326.

Calderon, J.S., Verbyla, M.E., Gil, M., Pinongcos, F., Kinoshita, A.M. and Mladenov, N., 2022. Persistence of fecal indicators and microbial source tracking markers in water flushed from riverbank soils. Water, Air, & Soil Pollution, 233(3), p.83.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*(7), pp.581-583.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), pp.581-583.

Cameron, K.C., Di, H.J. and Moir, J.L., 2013. Nitrogen losses from the soil/plant system: a review. Annals of applied biology, 162(2), pp.145-173.

Carrillo, V., Fuentes, B., Gómez, G. and Vidal, G., 2020. Characterization and recovery of phosphorus from wastewater by combined technologies. Reviews in Environmental Science and Bio/Technology, 19, pp.389-418.

Cederlund, H., Wessén, E., Enwall, K., Jones, C.M., Juhanson, J., Pell, M., Philippot, L. and Hallin, S., 2014. Soil carbon quality and nitrogen fertilization structure bacterial communities with predictable responses of major bacterial phyla. Applied soil ecology, 84, pp.62-68.

Central Statistics Office, 2019. Urban and Rural Lie in Ireland, 2019 https://www.cso.ie/en/releasesandpublications/ep/purli/urbanandrurallifeinireland2019/introduction/ [ACCESSED: 13-10-23]

Central Statistics Office. (2023)(a). Domestic Waste Water Treatment Systems 2022 Background Notes. CSO Statistical Publication. https://www.cso.ie/en/releasesandpublications/ep/pdwwts/domesticwastewatertreatmentsystems2022/backgroundnotes/Online ISSN: 2737-758X. Published on 13 March 2023.

Central Statistics Office. (b) (2023, August 23). PFSA01 – Agricultural Output and Income. Retrieved August 23, 2023, from https://data.cso.ie/table/PFSA01

Chamchoi, N., Nitisoravut, S. and Schmidt, J.E., 2008. Inactivation of ANAMMOX communities under concurrent operation of anaerobic ammonium oxidation (ANAMMOX) and denitrification. *Bioresource Technology*, *99*(9), pp.3331-3336.

Chang, A.C., Olmstead, W.R., Johanson, J.B. and Yamashita, G., 1974. The sealing mechanism of wastewater ponds. Journal (Water Pollution Control Federation), pp.1715-1721.

Chao, A. and Lee, S.M., 1992. Estimating the number of classes via sample coverage. Journal of the American statistical Association, 87(417), pp.210-217.

Characklis, W.G. and Marshall, K.C., 1990. Biofilms. (No Title).

Chau, J.F., Bagtzoglou, A.C. and Willig, M.R., 2011. The effect of soil texture on richness and diversity of bacterial communities. Environmental Forensics, 12(4), pp.333-341.

Chen, L., Feng, Q., Li, C., Wei, Y., Zhao, Y., Feng, Y., Zheng, H., Li, F. and Li, H., 2017. Impacts of aquaculture wastewater irrigation on soil microbial functional diversity and community structure in arid regions. Scientific Reports, 7(1), p.11193.

Chen, Y., Lan, S., Wang, L., Dong, S., Zhou, H., Tan, Z. and Li, X., 2017. A review: driving factors and regulation strategies of microbial community structure and dynamics in wastewater treatment systems. *Chemosphere*, *174*, pp.173-182.

Cobos, D., 2015. Measurement volume of Decagon volumetric water content sensors. Application Note Decagon Devices; Decagon Devices Inc.: Pullman, WA, USA, pp.1-4.

Cook, K.L. and Bolster, C.H., 2007. Survival of Campylobacter jejuni and Escherichia coli in groundwater during prolonged starvation at low temperatures. Journal of applied microbiology, 103(3), pp.573-583.

Cooper, J.A., Morales, I. and Amador, J.A., 2016. Nitrogen transformations in different types of soil treatment areas receiving domestic wastewater. Ecological Engineering, 94, pp.22-29.

Cooper, R.J., Fitt, P., Hiscock, K.M., Lovett, A.A., Gumm, L., Dugdale, S.J., Rambohul, J., Williamson, A., Noble, L., Beamish, J. and Hovesen, P., 2016. Assessing the effectiveness of a three-stage on-farm biobed in treating pesticide contaminated wastewater. *Journal of Environmental Management*, *181*, pp.874-882.

Costa, O.Y., Raaijmakers, J.M. and Kuramae, E.E., 2018. Microbial extracellular polymeric substances: ecological function and impact on soil aggregation. Frontiers in microbiology, 9, p.1636.

Criado Monleon, A.J., Knappe, J., Somlai, C., Betancourth, C.O., Ali, M., Curtis, T.P. and Gill, L.W., 2022. Spatial Variation of the Microbial Community Structure of On-Site Soil Treatment

Units in a Temperate Climate, and the Role of Pre-treatment of Domestic Effluent in the Development of the Biomat Community. Frontiers in Microbiology, 13, p.915856.
Cruaud, P., Vigneron, A., Fradette, M.S., Dorea, C.C., Culley, A.I., Rodriguez, M.J. and Charette, S.J., 2020. Annual bacterial community cycle in a seasonally ice-covered river reflects environmental and climatic conditions. Limnology and Oceanography, 65, pp.S21-S37.
Curley, M., Coonan, B., Ruth, C.E. and Ryan, C. 2023. Ireland's Climate Averages 1991-2020. Climatological Note No. 22. Met Éireann, Ireland.
Curtis, T.P. and Sloan, W.T., 2005. Exploring microbial diversity--A vast below. *Science*, *309*(5739), pp.1331-1333.
Cui, J., Yang, B., Zhang, M., Song, D., Xu, X., Ai, C., Liang, G. and Zhou, W., 2023. Investigating the effects of organic amendments on soil microbial composition and its linkage to soil organic carbon: A global meta-analysis. Science of The Total Environment, p.164899.
Dai, Z., Su, W., Chen, H., Barberán, A., Zhao, H., Yu, M., Yu, L., Brookes, P.C., Schadt, C.W., Chang, S.X. and Xu, J., 2018. Long-term nitrogen fertilization decreases bacterial diversity and favors the growth of Actinobacteria and Proteobacteria in agro-ecosystems across the globe. *Global change biology*, *24*(8), pp.3452-3461.
Daims, H., Taylor, M.W. and Wagner, M., 2006. Wastewater treatment: a model system for microbial ecology. *Trends in biotechnology*, *24*(11), pp.483-489.
Dana, M.J., Bostam, S. and Nyanti, L., 2012. Domestic wastewater quality and pollutant loadings from urban housing areas. Iranian (Iranica) Journal of Energy & Environment, 3(2).
Dang, Q., Tan, W., Zhao, X., Li, D., Li, Y., Yang, T., Li, R., Zu, G. and Xi, B., 2019. Linking the response of soil microbial community structure in soils to long-term wastewater irrigation and soil depth. *Science of the total environment*, *688*, pp.26-36.
Davies, C.M., Logan, M.R., Rothwell, V.J., Krogh, M., Ferguson, C.M., Charles, K., Deere, D.A. and Ashbolt, N.J., 2006. Soil inactivation of DNA viruses in septic seepage. Journal of applied microbiology, 100(2), pp.365-374.
De Carli, M., Bernardi, A., Cultrera, M., Dalla Santa, G., Di Bella, A., Emmi, G., Galgaro, A., Graci, S., Mendrinos, D., Mezzasalma, G., Pasquali, R., 2018. A database for climatic conditions around Europe for promoting GSHP solutions. Geosciences, 8(2), p.71.
Decagon (2015a), "EC-5 Soil Moisture Sensor, Operator's Manual." Technical report, Decagon Devices,
Decagon (2015b), "GS-3Water Content, EC and Temperature Sensors, Operator's Manual." Technical report, Decagon Devices, Inc., Pullman, WA, USA.
Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. and Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. Science, 359(6373), pp.320-325.
Derrien, M., Jardé, E., Gruau, G., Pourcher, A.M., Gourmelon, M., Jadas-Hécart, A. and Wickmann, A.P., 2012. Origin of fecal contamination in waters from contrasted areas: Stanols as Microbial Source Tracking markers. Water research, 46(13), pp.4009-4016.
Desmarais, T.R., Solo-Gabriele, H.M. and Palmer, C.J., 2002. Influence of soil on faecal indicator organisms in a tidally influenced subtropical environment. Applied and environmental microbiology, 68(3), pp.1165-1172.
Devane, M.L., Moriarty, E., Weaver, L., Cookson, A. and Gilpin, B., 2020. Faecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research, 185, p.116204.
Dhakar, V. and Geetanjali, A.S., 2022. Role of pepper mild mottle virus as a tracking tool for fecal pollution in aquatic environments. Archives of Microbiology, 204(8), p.513.
Dorofeev, A.G., Nikolaev, Y.A., Mardanov, A.V. and Pimenov, N.V., 2020. Role of phosphate-accumulating bacteria in biological phosphorus removal from wastewater. Applied biochemistry and microbiology, 56, pp.1-14.
Dottorini, G., Michaelsen, T.Y., Kucheryavskiy, S., Andersen, K.S., Kristensen, J.M., Peces, M., Wagner, D.S., Nierychlo, M. and Nielsen, P.H., 2021. Mass-immigration determines the

assembly of activated sludge microbial communities. *Proceedings of the National Academy of Sciences*, *118*(27), p.e2021589118.

Dubber, D. and Gill, L., 2014. Application of on-site wastewater treatment in Ireland and perspectives on its sustainability. Sustainability, 6(3), pp.1623-1642.

Dubber, D. and Gray, N.F., 2010. Replacement of chemical oxygen demand (COD) with total organic carbon (TOC) for monitoring wastewater treatment performance to minimize disposal of toxic analytical waste. Journal of Environmental Science and Health Part A, 45(12), pp.1595-1600.

Dubber, D., Knappe, J. and Gill, L.W., 2021. Characterisation of organic matter and its transformation processes in on-site wastewater effluent percolating through soil using fluorescence spectroscopic methods and parallel factor analysis (PARAFAC). *Water*, *13*(19), p.2627.

Dueholm, M.K.D., Nierychlo, M., Andersen, K.S., Rudkjøbing, V., Knutsson, S., Albertsen, M. and Nielsen, P.H., 2022. MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. Nature communications, 13(1), p.1908.

Duffy, P., Hanley, E., Hyde, B., O'Brien, P., Ponzi, J., Cotter, E. and Black, K., 2014. Ireland National Inventory Report 2014, Greenhouse Gas Emissions 1990–2012. Reported to the United Nations Framework Convention on Climate Change. Environmental Protection Agency, Johnstown Castle Estate, Co. Wexford, Ireland.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, *10*(10), pp.996-998.

Eilers, K. G.; Debenport, S.; Anderson, S.; Fierer, N. Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. Soil Biol. Biochem. 2012, 50, 58–65.

Environmental Protection Agency (EPA)., NATIONAL INSPECTION PLAN Domestic Waste Water Treatment Systems 2022 – 2026; 2022

Environmental Protection Agency (EPA)., Urban Wastewater Treatment in 2021;2022

Environmental Protection Agency Ireland Domestic Waste Water Treatment System (DWWTS) Inspections 2022

Environmental Protection Agency Ireland Water Quality in 2022 An Indicators Report

Environmental Protection Agency Ireland. Code of Practice for Domestic Waste Water Treatment Systems (Population Equivalent ≤ 10)., Ireland: Environmental Protection Agency; 2021

Environmental Protection Agency. Water Quality in Ireland 2013–2018 EPA; 2018.

Escher A, Characklis WG. Modelingthe initial events in biofilm accumulation. In: Characklis WG, Marshall KC, editors. Biofilms. NewYork, NY: Wiley; 1990. p. 445–86.

European Environment Agency European waters — current status and future challenges.2012

EUROSTAT, 2022. Population projections in the EU Population projections in the EU - Statistics Explained (europa.eu) [ACCESSED: 13-10-23]

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D. and Tito, R.Y., 2016. Population-level analysis of gut microbiome variation. Science, 352(6285), pp.560-564.

Fealy, R., 2009. *Teagasc-EPA Soils and Subsoils mapping project: Final report V. 1*. Teagasc; Environmental Protection Agency.

Fennell, C., Misstear, B., O'Connell, D., Dubber, D., Behan, P., Danaher, M., Moloney, M. and Gill, L., 2021. An assessment of contamination fingerprinting techniques for determining the impact of domestic wastewater treatment systems on private well supplies. Environmental Pollution, 268, p.115687.

Ferguson, A.S., Layton, A.C., Mailloux, B.J., Culligan, P.J., Williams, D.E., Smartt, A.E., Sayler, G.S., Feighery, J., McKay, L.D., Knappett, P.S. and Alexandrova, E., 2012. Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. Science of the Total Environment, 431, pp.314-322

Fernández-Baca, C.P., Omar, A.E.H., Pollard, J.T. and Richardson, R.E., 2018. Microbial communities controlling methane and nutrient cycling in leach field soils. Water research, 151, pp.456-467.

Fernández-Baca, C.P., Truhlar, A.M., Omar, A.E.H., Rahm, B.G., Walter, M.T. and Richardson, R.E., 2018. Methane and nitrous oxide cycling microbial communities in soils above septic leach fields: Abundances with depth and correlations with net surface emissions. Science of the Total Environment, 640, pp.429-441.

Field, K.G. and Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water research, 41(16), pp.3517-3538.

Fields, S., 2004. Global nitrogen: cycling out of control.

Fierer, N. and Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, *103*(3), pp.626-631.Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A. and Cleveland, C.C., 2009. Global patterns in belowground communities. Ecology letters, 12(11), pp.1238-1249.

Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology, 15(10), pp.579-590.

Fierer, N., Bradford, M.A. and Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. Ecology, 88(6), pp.1354-1364.

Filippelli, G.M., 2002. The global phosphorus cycle. Reviews in mineralogy and geochemistry, 48(1), pp.391-425.

Flemming, H.C. and Wingender, J., 2010. The biofilm matrix. *Nature reviews microbiology*, *8*(9), pp.623-633.

Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A. and Kjelleberg, S., 2016. Biofilms: an emergent form of bacterial life. Nature Reviews Microbiology, 14(9), pp.563-575.

Frigon, D. and Wells, G., 2019. Microbial immigration in wastewater treatment systems: analytical considerations and process implications. *Current opinion in biotechnology*, *57*, pp.151-159.

Fuerst, J.A. and Sagulenko, E., 2011. Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nature Reviews Microbiology, 9(6), pp.403-413.

Galloway, J.N., Leach, A.M., Bleeker, A. and Erisman, J.W., 2013. A chronology of human understanding of the nitrogen cycle. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), p.20130120.

Gannon JT, Manilal VB, Alexander M. Relationship between cell surface properties and transport of bacteria through soil. Appl Environ Microbiol Am Soc Microbiol 1991;57:190–3.

Gao, Y., Wang, C., Zhang, W., Di, P., Yi, N. and Chen, C., 2017. Vertical and horizontal assemblage patterns of bacterial communities in a eutrophic river receiving domestic wastewater in southeast China. Environmental Pollution, 230, pp.469-478.

Gerba, C.P., Dickenson, E.R. and Drewes, J.E., 2017. The importance of key attenuation factors for microbial and chemical contaminants during managed aquifer recharge: a review. Critical Reviews in Environmental Science and Technology, 47(15), pp.1409-1452.

Geyer, K. M., Barrett, J. E., 2019. Unimodal productivity – diversity relationships among bacterial communities in a simple polar soil ecosystem. Environmental microbiology, 21(7), pp.2523-2532.

Gharoon, N. and Pagilla, K.R., 2021. Critical review of effluent dissolved organic nitrogen removal by soil/aquifer-based treatment systems. Chemosphere, 269, p.129406.

Gill, L., O'Flaherty, V., Misstear, B., Brophy, L., Fennell, C., Dubber, D., O'Connell, D., Kilroy, K., Barrett, M., Johnston, P., Pilla, F., and Geary, P.The Impact of On-site Domestic Wastewater Effluent on Rivers and Wells. Report No.251 EPA IRELAND; 2018

Gill, L., O'Flaherty, V., Misstear, B., Brophy, L., Fennell, C., Dubber, D., O'Connell, D.,Kilroy, K., Barrett, M., Johnston, P., Pilla, F., & Geary, P. (2018). Report No.251 EPA IRELAND: The Impact of On-site Domestic Wastewater Effluent on Rivers and Wells

Gill, L.W., 2011. The development of a Code of Practice for single house on-site wastewater treatment in Ireland. *Water Science and Technology*, *64*(3), pp.677-683.

Gill, L.W., O'luanaigh, N., Johnston, P.M., Misstear, B.D.R. and O'suilleabhain, C., 2009. Nutrient loading on subsoils from on-site wastewater effluent, comparing septic tank and secondary treatment systems. Water Research, 43(10), pp.2739-2749.Martin, N.A., Stephens, J. and Meijer, W., The Impact of Dog Fouling on Bathing Water Quality in Dublin Bay.

Gill, L.W., O'luanaigh, N., Johnston, P.M., Misstear, B.D.R. and O'suilleabhain, C., 2009. Nutrient loading on subsoils from on-site wastewater effluent, comparing septic tank and secondary treatment systems. Water Research, 43(10), pp.2739-2749.

Gill, L.W., O'Súlleabháin, C., Misstear, B.D.R. and Johnston, P.J., 2007. The treatment performance of different subsoils in Ireland receiving on-site wastewater effluent. *Journal of Environmental Quality*, *36*(6), pp.1843-1855.

Goldschmith J, Zohar D, Argamon Y, Kott Y. Effects of dissolved salts on the filtration of coliform bacteria in sand dunes. In: Jenkins SH, editor. Advances in water pollution research. New York, NY: Pergamon Press; 1973. p. 147–55.

Gonçalves, R.A., Folegatti, M.V., Gloaguen, T.V., Libardi, P.L., Montes, C.R., Lucas, Y., Dias, C.T. and Melfi, A.J., 2007. Hydraulic conductivity of a soil irrigated with treated sewage effluent. Geoderma, 139(1-2), pp.241-248.

Gordon, D.M., 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology, 147(5), pp.1079-1085.

Gorski, G., Dailey, H., Fisher, A.T., Schrad, N. and Saltikov, C., 2020. Denitrification during infiltration for managed aquifer recharge: Infiltration rate controls and microbial response. Science of The Total Environment, 727, p.138642.

Greaves, J., Stone, D., Wu, Z. and Bibby, K., 2020. Persistence of emerging viral fecal indicators in large-scale freshwater mesocosms. Water Research X, 9, p.100067.

Grenni, P., Ancona, V. and Caracciolo, A.B., 2018. Ecological effects of antibiotics on natural ecosystems: A review. Microchemical Journal, 136, pp.25-39.

Griffin, J.S. and Wells, G.F., 2017. Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly. The ISME journal, 11(2), pp.500-511.

Griffiths, B.S. and Philippot, L., 2013. Insights into the resistance and resilience of the soil microbial community. *FEMS microbiology reviews*, *37*(2), pp.112-129.

Guo, Y. S., Furrer, J. M., Kadilak, A. L., Hinestroza, H. F., Gage, D.J., Cho, Y. K. = Shor, L. M., 2018. Bacterial extracellular polymeric substances amplify water content variability at the pore scale. Frontiers in Environmental Science, p.93.

Gyawali, P., Croucher, D., Ahmed, W., Devane, M. and Hewitt, J., 2019. Evaluation of pepper mild mottle virus as an indicator of human faecal pollution in shellfish and growing waters. Water research, 154, pp.370-376.

Hagiwara, K., Ichiki, T.U., Ogawa, Y., Omura, T. and Tsuda, S., 2002. A single amino acid substitution in 126-kDa protein of Pepper mild mottle virus associates with symptom attenuation in pepper; the complete nucleotide sequence of an attenuated strain, C-1421. Archives of virology, 147, pp.833-840.

Hamidi, N.H., Ahmed, O.H., Omar, L. and Ch'ng, H.Y., 2021. Soil nitrogen sorption using charcoal and wood ash. Agronomy, 11(9), p.1801.

Hamza, I.A., Jurzik, L., Überla, K. and Wilhelm, M., 2011. Evaluation of pepper mild mottle virus, human picobirnavirus and Torque teno virus as indicators of fecal contamination in river water. Water research, 45(3), pp.1358-1368.

Hashimoto, K., Matsuda, M., Inoue, D. and Ike, M., 2014. Bacterial community dynamics in a full-scale municipal wastewater treatment plant employing conventional activated sludge process. Journal of bioscience and bioengineering, 118(1), pp.64-71.

Hashmi, I., Bindschedler, S. and Junier, P., 2020. Firmicutes. In Beneficial microbes in agro-ecology (pp. 363-396). Academic Press.

Hata, A., Kitajima, M. and Katayama, H., 2013. Occurrence and reduction of human viruses, F-specific RNA coliphage genogroups and microbial indicators at a full-scale wastewater treatment plant in Japan. Journal of applied microbiology, 114(2), pp.545-554.

He, S.; Guo, L.; Niu, M.; Miao, F.; Jiao, S.; Hu, T.; Long, M. Ecological diversity and co-occurrence patterns of bacterial community through soil profile in response to long-term switchgrass cultivation. Sci. Rep. 2017, 7, 3608

Henry, C.L., Sullivan, D., Rynk, R., Dorsey, K. and Cogger, C., 1999. Managing nitrogen from biosolids. Seattle^ eWA WA: Washington State Department of Ecology.

Hess, J.F., Kohl, T.A., Kotrová, M., Rönsch, K., Paprotka, T., Mohr, V., Hutzenlaub, T., Brüggemann, M., Zengerle, R., Niemann, S. and Paust, N., 2020. Library preparation for next generation sequencing: A review of automation strategies. Biotechnology advances, 41, p.107537.Inc., Pullman, WA, USA.

Holmes, D.E., Dang, Y. and Smith, J.A., 2019. Nitrogen cycling during wastewater treatment. Advances in applied microbiology, 106, pp.113-192.

Hu, M., Wang, X., Wen, X. and Xia, Y., 2012. Microbial community structures in different wastewater treatment plants as revealed by 454-pyrosequencing analysis. *Bioresource technology*, *117*, pp.72-79.

Huang, M.H., Li, Y.M. and Gu, G.W., 2010. Chemical composition of organic matters in domestic wastewater. Desalination, 262(1-3), pp.36-42.

Hui, C., Wei, R., Jiang, H., Zhao, Y. and Xu, L., 2019. Characterization of the ammonification, the relevant protease production and activity in a high-efficiency ammonifier Bacillus amyloliquefaciens DT. International Biodeterioration & Biodegradation, 142, pp.11-17.

Humphrey Jr, C.P., Iverson, G., Underwood, W.J., Cary, S.S., Skibiel, C. and O'Driscoll, M., 2019. Nitrogen treatment in soil beneath high-flow and low-flow onsite wastewater systems. *Journal of Sustainable Water in the Built Environment*, *5*(4), p.04019006.Humphrey Jr, C.P., Iverson, G.,

Hynds, P., Naughton, O., O'Neill, E. and Mooney, S., 2018. Efficacy of a national hydrological risk communication strategy: Domestic wastewater treatment systems in the Republic of Ireland. Journal of Hydrology, 558, pp.205-213.

Ibekwe, A.M., Gonzalez-Rubio, A. and Suarez, D.L., 2018. Impact of treated wastewater for irrigation on soil microbial communities. Science of the Total Environment, 622, pp.1603-1610.

International Water Association. 2021. Nature-Based Solutions for Wastewater: A Series of Factsheets and Case Studies.

Ishii, S. and Sadowsky, M.J., 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes and environments, 23(2), pp.101-108.

Ishii, S., Ksoll, W.B., Hicks, R.E. and Sadowsky, M.J., 2006. Presence and growth of naturalized Escherichia coli in temperate soils from Lake Superior watersheds. Applied and environmental microbiology, 72(1), pp.612-621.

Janssen, P.H., 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Applied and environmental microbiology, 72(3), pp.1719-1728.

Jauković, Z., Grujić, S., Bujagić, I.M., Petković, A. and Laušević, M., 2022. Steroid-based tracing of sewage-sourced pollution of river water and wastewater treatment efficiency: Dissolved and suspended water phase distribution. Science of The Total Environment, 846, p.157510.

Jenkins, C. and Staley, J.T., 2013. History, classification and cultivation of the Planctomycetes. In Planctomycetes: cell structure, origins and biology (pp. 1-38). Totowa, NJ: Humana Press.

Jiang, X.T., Ye, L., Ju, F., Li, B., Ma, L.P. and Zhang, T., 2018. Temporal dynamics of activated sludge bacterial communities in two diversity variant full-scale sewage treatment plants. Applied microbiology and biotechnology, 102, pp.9379-9388.

Jimenez, L., Muniz, I., Toranzos, G.A. and Hazen, T.C., 1989. Survival and activity of Salmonella typhimurium and Escherichia coli in tropical freshwater. Journal of Applied Bacteriology, 67(1), pp.61-69.

Ju, F. and Zhang, T., 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. The ISME journal, 9(3), pp.683-695.

Karimi, B., Villerd, J., Dequiedt, S., Terrat, S., Chemidlin-Prévost Bouré, N., Djemiel, C., Lelièvre, M., Tripied, J., Nowak, V., Saby, N.P. Bispo, A., 2020. Biogeography of soil microbial habitats across France. Global Ecology and Biogeography, 29(8), pp.1399-1411.

Karpouzas, D.G., Ntougias, S., Iskidou, E., Rousidou, C., Papadopoulou, K.K., Zervakis, G.I. and Ehaliotis, C., 2010. Olive mill wastewater affects the structure of soil bacterial communities. Applied soil ecology, 45(2), pp.101-111.

Kartal, B., Kuenen, J.V. and Van Loosdrecht, M.C.M., 2010. Sewage treatment with anammox. *Science*, *328*(5979), pp.702-703.

Kim, H.S., Lee, S.H., Jo, H.Y., Finneran, K.T. and Kwon, M.J., 2021. Diversity and composition of soil Acidobacteria and Proteobacteria communities as a bacterial indicator of past land-use change from forest to farmland. Science of the Total Environment, 797, p.148944.

Kitajima, M., Iker, B.C., Pepper, I.L. and Gerba, C.P., 2014. Relative abundance and treatment reduction of viruses during wastewater treatment processes—identification of potential viral indicators. *Science of the Total Environment*, *488*, pp.290-296.

Kitajima, M., Sassi, H.P. and Torrey, J.R., 2018. Pepper mild mottle virus as a water quality indicator. NPJ Clean Water, 1(1), p.19.

Knappe, J., Somlai, C. and Gill, L.W., 2022. Assessing the spatial and temporal variability of greenhouse gas emissions from different configurations of on-site wastewater treatment system using discrete and continuous gas flux measurement. Biogeosciences, 19(4), pp.1067-1085.

Knappe, J., Somlai, C., Fowler, A.C. and Gill, L.W., 2020. The influence of pre-treatment on biomat development in soil treatment units. *Journal of Contaminant Hydrology*, *232*, p.103654.

Knisz, J., Shetty, P., Wirth, R., Maróti, G., Karches, T., Dalkó, I., Bálint, M., Vadkerti, E. and Bíró, T., 2021. Genome-level insights into the operation of an on-site biological wastewater treatment unit reveal the importance of storage time. Science of the Total Environment, 766, p.144425.

Koch, A.L., 2001. Oligotrophs versus copiotrophs. Bioessays, 23(7), pp.657-661.

Koch, C., Fetzer, I., Schmidt, T., Harms, H. and Müller, S., 2013. Monitoring functions in managed microbial systems by cytometric bar coding. Environmental science & technology, 47(3), pp.1753-1760.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. , Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and environmental microbiology, 79(17), pp.5112 5120.

Krause, S.M., Dohrmann, A.B., Gillor, O., Christensen, B.T., Merbach, I. and Tebbe, C.C., 2020. Soil properties and habitats determine the response of bacterial communities to agricultural wastewater irrigation. Pedosphere, 30(1), pp.146-158.

Kuroda, K., Nakada, N., Hanamoto, S., Inaba, M., Katayama, H., Do, A.T., Nga, T.T.V., Oguma, K., Hayashi, T. and Takizawa, S., 2015. Pepper mild mottle virus as an indicator and a tracer of fecal pollution in water environments: comparative evaluation with wastewater-tracer pharmaceuticals in Hanoi, Vietnam. *Science of the Total Environment*, *506*, pp.287-298.

Kuzyakov, Y. and Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry, 83, pp.184-199.

Laak, R., 1970. Influence of domestic wastewater pretreatment on soil clogging. *Journal (Water Pollution Control Federation)*, pp.1495-1500.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N., 2009. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. Appl Environ Microbiol, 75(15), pp.5111-5120.

Lee, S.H., Kang, H.J. and Park, H.D., 2015. Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water research*, *73*, pp.132-144.

Leibold, M. A.; Holyoak, M.; Mouquet, N.; Amarasekare, P.;Chase, J. M.; Hoopes, M. F.; Holt, R. D.; Shurin, J. B.; Law, R.;Tilman, D.; Loreau, M.; Gonzalez, A. The Metacommunity Concept:A Framework for Multi-Scale Community Ecology. Ecol. Lett. 2004, 7(7), 601−613

Li, D., Sharp, J.O., Saikaly, P.E., Ali, S., Alidina, M., Alarawi, M.S., Keller, S., Hoppe-Jones, C. and Drewes, J..E., 2012. Dissolved organic carbon influences microbial community composition and diversity in managed aquifer recharge systems. Applied and environmental microbiology, 78(19), pp.6819-6828.

Li, H.Y., Wang, H., Wang, H.T., Xin, P.Y., Xu, X.H., Ma, Y., Liu, W.P., Teng, C.Y., Jiang, C.L., Lou, L.P. and Arnold, W., 2018. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. Microbiome, 6, pp.1-16.

Li, Y., Xie, X., Zhu, Z., Liu, K., Liu, W. and Wang, J., 2022. Land use driven change in soil organic carbon affects soil microbial community assembly in the riparian of Three Gorges Reservoir Region. Applied Soil Ecology, 176, p.104467.

Liao, R., Miao, Y., Li, J., Li, Y., Wang, Z., Du, J., Li, Y., Li, A. and Shen, H., 2018. Temperature dependence of denitrification microbial communities and functional genes in an expanded granular sludge bed reactor treating nitrate-rich wastewater. RSC advances, 8(73), pp.42087-42094.

Lin, X., McKinley, J., Resch, C.T., Kaluzny, R., Lauber, C.L., Fredrickson, J., Knight, R. and Konopka, A., 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. The ISME journal, 6(9), pp.1665-1676.

Lindström, E.S. and Langenheder, S., 2012. Local and regional factors influencing bacterial community assembly. Environmental Microbiology Reports, 4(1), pp.1-9.

Liu, N., Hu, H., Ma, W., Deng, Y., Liu, Y., Hao, B., Zhang, X., Dimitrov, D., Feng, X. and Wang, Z., 2019. Contrasting biogeographic patterns of bacterial and archaeal diversity in the top-and subsoils of temperate grasslands. MSystems, 4(5), pp.e00566-19.

Liu, P., Jia, S., He, X., Zhang, X. and Ye, L., 2017. Different impacts of manure and chemical fertilizers on bacterial community structure and antibiotic resistance genes in arable soils. *Chemosphere*, *188*, pp.455-464.

Liu, Y., 2010. Methanopyrales. In Handbook of hydrocarbon and lipid microbiology.

Lotti, T., Kleerebezem, R., van Erp Taalman Kip, C., Hendrickx, T.L., Kruit, J., Hoekstra, M. and Van Loosdrecht, M.C., 2014. Anammox growth on pretreated municipal wastewater. Environmental science & technology, 48(14), pp.7874-7880.

Louca, S., Polz, M.F., Mazel, F., Albright, M.B., Huber, J.A., O'Connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., Crowe, S.A. and Doebeli, M., 2018. Function and functional redundancy in microbial systems. *Nature ecology & evolution*, *2*(6), pp.936-943.

Lu, C.S., Qiu, J.M., Yang, Y., Hu, Y., Li, Y.Y., Kobayashi, T. and Zhang, Y.J., 2023. A review of anaerobic granulation under high-salinity conditions: Mechanisms, influencing factors and enhancement strategies. Journal of Water Process Engineering, 55, p.104227.

Luan, L., Liang, C., Chen, L., Wang, H., Xu, Q., Jiang, Y. and Sun, B., 2020. Coupling bacterial community assembly to microbial metabolism across soil profiles. Msystems, 5(3), pp.10-1128.

Lusk, M.G., Toor, G.S., Yang, Y.Y., Mechtensimer, S., De, M. and Obreza, T.A., 2017. A review of the fate and transport of nitrogen, phosphorus, pathogens, and trace organic chemicals in septic systems. Critical Reviews in Environmental Science and Technology, 47(7), pp.455-541.

Lyu, Z., Shao, N., Akinyemi, T. and Whitman, W.B., 2018. Methanogenesis. Current Biology, 28(13), pp.R727-R732.

MacArthur, R.H., 1965. Patterns of species diversity. Biological reviews, 40(4), pp.510-533.

Matar, G.K., Ali, M., Bagchi, S., Nunes, S., Liu, W.T. and Saikaly, P.E., 2021. Relative importance of stochastic assembly process of membrane biofilm increased as biofilm aged. *Frontiers in Microbiology*, *12*, p.708531.

Matar, G.K., Bagchi, S., Zhang, K., Oerther, D.B. and Saikaly, P.E., 2017. Membrane biofilm communities in full-scale membrane bioreactors are not randomly assembled and consist of a core microbiome. Water research, 123, pp.124-133.

McIlroy, S.J., Saunders, A.M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A.A., Karst, S.M., Nielsen, J.L. and Nielsen, P.H., 2015. MiDAS: the field guide to the microbes of activated sludge. Database, 2015, p.bav062.

McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N. and Sogin, M., 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental microbiology*, *12*(2), pp.378-392.

Mei, R. and Liu, W.T., 2019. Quantifying the contribution of microbial immigration in engineered water systems. Microbiome, 7(1), pp.1-8.

Mergelov, N., Mueller, C.W., Prater, I., Shorkunov, I., Dolgikh, A., Zazovskaya, E., Shishkov, V., Krupskaya, V., Abrosimov, K., Cherkinsky, A. and Goryachkin, S., 2018. Alteration of rocks by endolithic organisms is one of the pathways for the beginning of soils on Earth. Scientific Reports, 8(1), p.3367.

Met Éireann,, 2018. A Summer of Heat Waves and Droughts. Irish Meteorological Sevice: Dublin, Ireland.

Mockler, E.M., Deakin, J., Archbold, M., Gill, L., Daly, D. and Bruen, M., 2017. Sources of nitrogen and phosphorus emissions to Irish rivers and coastal waters: Estimates from a nutrient load apportionment framework. *Science of the Total Environment*, *601*, pp.326-339.

Montecchia, M.S., Tosi, M., Soria, M.A., Vogrig, J.A., Sydorenko, O. and Correa, O.S., 2015. Pyrosequencing reveals changes in soil bacterial communities after conversion of Yungas forests to agriculture. PloS one, 10(3), p.e0119426.

Morrison, C.M., Betancourt, W.Q., Quintanar, D.R., Lopez, G.U., Pepper, I.L. and Gerba, C.P., 2020. Potential indicators of virus transport and removal during soil aquifer treatment of treated wastewater effluent. Water research, 177, p.115812

Mozes N, Marcha F, Hermesse MP, Van Haecht JL, Reuliaux l, Leanonard AJ, Rouxhet PG. Immobilization of microorganisms by adhesion: interplay of electrostatic and non-electrostatic interactions. Biotechnol Bioeng 1987;30:439–50

Mujakić, I., Piwosz, K. and Koblížek, M., 2022. Phylum Gemmatimonadota and its role in the environment. Microorganisms, 10(1), p.151.

Murphy, C.L., Biggerstaff, J., Eichhorn, A., Ewing, E., Shahan, R., Soriano, D., Stewart, S., VanMol, K., Walker, R., Walters, P. and Elshahed, M.S., 2021. Genomic characterization of three novel Desulfobacterota classes expand the metabolic and phylogenetic diversity of the phylum. Environmental microbiology, 23(8), pp.4326-4343.

Murphy, D.V., Recous, S., Stockdale, E.A., Fillery, I.R.P., Jensen, L.S., Hatch, D.J. and Goulding, K.W.T., 2003. Gross nitrogen fluxes in soil: theory, measurement and application of^ 1^ 5n pool dilution techniques. Advances in Agronomy, 79(69), p.e118.

Naeem, S., Bunker, D.E., Hector, A., Loreau, M. and Perrings, C. eds., 2009. *Biodiversity, ecosystem functioning, and human wellbeing: an ecological and economic perspective*. OUP Oxford.

Naether, A., Foesel, B.U., Naegele, V., Wüst, P.K., Weinert, J., Bonkowski, M., Alt, F., Oelmann, Y., Polle, A., Lohaus, G. and Gockel, S., 2012. Environmental factors affect acidobacterial communities below the subgroup level in grassland and forest soils. *Applied and Environmental Microbiology*, *78*(20), pp.7398-7406.

Napieralski, S.A., Buss, H.L., Brantley, S.L., Lee, S., Xu, H. and Roden, E.E., 2019. Microbial chemolithotrophy mediates oxidative weathering of granitic bedrock. Proceedings of the National Academy of Sciences, 116(52), pp.26394-26401.

Navarrete, A.A., Venturini, A.M., Meyer, K.M., Klein, A.M., Tiedje, J.M., Bohannan, B.J., Nüsslein, K., Tsai, S.M. and Rodrigues, J.L., 2015. Differential response of Acidobacteria subgroups to forest-to-pasture conversion and their biogeographic patterns in the western Brazilian Amazon. Frontiers in microbiology, 6, p.1443.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L. and Knight, R., 2011. Global patterns in the biogeography of bacterial taxa. Environmental microbiology, 13(1), pp.135-144.

Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M. and Sogin, M.L., 2015. Sewage reflects the microbiomes of human populations. *MBio*, *6*(2), pp.10-1128.

Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Jiang, C., Albertsen, M., Dueholm, M.S. and Nielsen, P.H., 2020. MiDAS 3: an ecosystem-specific reference database, taxonomy and

knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge. Water Research, 182, p.115955.

Nnane, D.E., Ebdon, J.E. and Taylor, H.D., 2011. Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. *Water research*, *45*(6), pp.2235-2246.

Nobu, M.K., Narihiro, T., Kuroda, K., Mei, R. and Liu, W.T., 2016. Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. The ISME journal, 10(10), pp.2478-2487.

Norton, J. and Ouyang, Y., 2019. Controls and adaptive management of nitrification in agricultural soils. Frontiers in microbiology, 10, p.1931.

Numberger, D., Ganzert, L., Zoccarato, L., Mühldorfer, K., Sauer, S., Grossart, H.P. and Greenwood, A.D., 2019. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. *Scientific reports*, *9*(1), p.9673.

O'Boyle, S., Trodd, W., Bradley, C., Tierney, D., Wilkes, R., Longphuirt, S.N., Smith, J., Stephens, A., Barry, J., Maher, P. and McGinn, R., 2019. Water quality in Ireland 2013–2018. Environmental Protection Agency, Wexford.

Obayomi, O., Seyoum, M.M., Ghazaryan, L., Tebbe, C.C., Murase, J., Bernstein, N. and Gillor, O., 2021. Soil texture and properties rather than irrigation water type shape the diversity and composition of soil microbial communities. *Applied Soil Ecology*, *161*, p.103834.

Oehmen, A., Lemos, P.C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L.L. and Reis, M.A., 2007. Advances in enhanced biological phosphorus removal: from micro to macro scale. Water research, 41(11), pp.2271-2300.

Oksanen, J., 2010. Vegan: community ecology package. *http://vegan. r-forge. r-project. org/*.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., (2017). vegan: community ecology package. R package version 2.4–4.2.https://CRAN.R project.org/package=vegan

Okubo, T. and Matsumoto, J., 1979. Effect of infiltration rate on biological clogging and water quality changes during artificial recharge. *Water Resources Research*, *15*(6), pp.1536-1542.Oliverio, A.M., Bradford, M.A. and Fierer, N., 2017. Identifying the microbial taxa that consistently respond to soil warming across time and space. Global Change Biology, 23(5), pp.2117-2129.

Oliverio, A.M., Bradford, M.A. and Fierer, N., 2017. Identifying the microbial taxa that consistently respond to soil warming across time and space. *Global change biology*, *23*(5), pp.2117-2129.

O'Luanaigh, N.D., Gill, L.W., Misstear, B.D.R. and Johnston, P.M., 2012. The attenuation of microorganisms in on-site wastewater effluent discharged into highly permeable subsoils. *Journal of contaminant hydrology*, *142*, pp.126-139..

Or, D., Phutane, S. and Dechesne, A., 2007. Extracellular polymeric substances affecting pore-scale hydrologic conditions for bacterial activity in unsaturated soils. *Vadose Zone Journal*, *6*(2), pp.298-305.

Ostvar, S., Iltis, G., Davit, Y., Schlüter, S., Andersson, L., Wood, B.D. and Wildenschild, D., 2018. Investigating the influence of flow rate on biofilm growth in three dimensions using microimaging. *Advances in Water Resources*, *117*, pp.1-13.

Ottoni, J.R., dos Santos Grignet, R., Barros, M.G.A., Bernal, S.P.F., Panatta, A.A.S., Lacerda-Júnior, G.V., Centurion, V.B., Delforno, T.P., da Costa Silva Goncalves, C. and Passarini, M.R.Z., 2022. DNA Metabarcoding from Microbial Communities Recovered from Stream and Its Potential for Bioremediation Processes. Current Microbiology, 79(2), p.70.

Oved, T., Shaviv, A., Goldrath, T., Mandelbaum, R.T. and Minz, D., 2001. Influence of effluent irrigation on community composition and function of ammonia-oxidizing bacteria in soil. *Applied and environmental microbiology*, *67*(8), pp.3426-3433.

Pabich, W.J., Valiela, I. and Hemond, H.F., 2001. Relationship between DOC concentration and vadose zone thickness and depth below water table in groundwater of Cape Cod, USA. Biogeochemistry, 55, pp.247-268.

Patel, H., 2018. Charcoal as an adsorbent for textile wastewater treatment. Separation Science and Technology, 53(17), pp.2797-2812.

Patel, T., O'Luanaigh, N. and Gill, L.W., 2008. A comparison of gravity distribution devices used in on-site domestic wastewater treatment systems. *Water, air, and soil pollution*, *191*, pp.55-69.

Peterson, S.B., Warnecke, F., Madejska, J., McMahon, K.D. and Hugenholtz, P., 2008. Environmental distribution and population biology of Candidatus Accumulibacter, a primary agent of biological phosphorus removal. Environmental microbiology, 10(10), pp.2692-2703.

Pett-Ridge, J. and Firestone, M.K., 2005. Redox fluctuation structures microbial communities in a wet tropical soil. Applied and environmental microbiology, 71(11), pp.6998-7007.

Philippot, L., Chenu, C., Kappler, A., Rillig, M.C. and Fierer, N., 2023. The interplay between microbial communities and soil properties. Nature Reviews Microbiology, pp.1-14.Regnery, J.,

Powell, J.R., Karunaratne, S., Campbell, C.D., Yao, H., Robinson, L. and Singh, B.K., 2015. Deterministic processes vary during community assembly for ecologically dissimilar taxa. Nature communications, 6(1), p.8444.

Powelson, D.K. and Gerba, C.P., 1994. Virus removal from sewage effluents during saturated and unsaturated flow through soil columns. Water Research, 28(10), pp.2175-2181.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Reddy, K.R., Patrick, W.H. and Broadbent, F.E., 1984. Nitrogen transformations and loss in flooded soils and sediments. Critical Reviews in Environmental Science and Technology, 13(4), pp.273-309.

Reinhold-Hurek, B., Bünger, W., Burbano, C. S., Sabale, M. and Hurek, T., 2015. Roots shaping their microbiome: global hotspots for microbial activity. Annual review of phytopathology, 53, pp.403-424.

Reynolds, L.J., Martin, N.A., Sala-Comorera, L., Callanan, K., Doyle, P., O'Leary, C., Buggy, P., Nolan, T.M., O'Hare, G.M., O'Sullivan, J.J. and Meijer, W.G., 2021. Identifying sources of faecal contamination in a small urban stream catchment: a multiparametric approach. Frontiers in Microbiology, 12, p.661954.

Rieke, E.L., Soupir, M.L., Moorman, T.B., Yang, F. and Howe, A.C., 2018. Temporal dynamics of bacterial communities in soil and leachate water after swine manure application. *Frontiers in Microbiology*, *9*, p.3197.

Roberson, E.B. and Firestone, M.K., 1992. Relationship between desiccation and exopolysaccharide production in a soil Pseudomonas sp. Applied and environmental microbiology, 58(4), pp.1284-1291.Roberto, A.A., Van Gray, J.B. and Leff, L.G., 2018. Sediment bacteria in an urban stream: spatiotemporal patterns in community composition. Water Research, 134, pp.353-369.

Rolston, A. and Linnane, S., 2020. Drinking water source protection for surface water abstractions: an overview of the group water scheme sector in the Republic of Ireland. Water, 12(9), p.2437.

Rosario, K., Symonds, E.M., Sinigalliano, C., Stewart, J. and Breitbart, M., 2009. Pepper mild mottle virus as an indicator of fecal pollution. Applied and environmental microbiology, 75(22), pp.7261-7267.

Rosiles-González, G., Ávila-Torres, G., Moreno-Valenzuela, O.A., Acosta-González, G., Leal-Bautista, R.M., Grimaldo-Hernández, C.D., Brown, J.K., Chaidez-Quiroz, C., Betancourt, W.Q., Gerba, C.P. and Hernández-Zepeda, C., 2017. Occurrence of pepper mild mottle virus (PMMoV) in groundwater from a karst aquifer system in the Yucatan Peninsula, Mexico. *Food and Environmental Virology*, *9*, pp.487-497.

Ross, B.N., Wigginton, S.K., Cox, A.H., Loomis, G.W. and Amador, J.A., 2020. Influence of season, occupancy pattern, and technology on structure and composition of nitrifying and denitrifying bacterial communities in advanced nitrogen-removal onsite wastewater treatment systems. *Water*, *12*(9), p.2413.

Saeed, T. and Sun, G., 2012. A review on nitrogen and organics removal mechanisms in subsurface flow constructed wetlands: dependency on environmental parameters, operating conditions and supporting media. Journal of environmental management, 112, pp.429-448.

Sangwan, P., Kovac, S., Davis, K.E., Sait, M. and Janssen, P.H., 2005. Detection and cultivation of soil Verrucomicrobia. Applied and environmental microbiology, 71(12), pp.8402-8410.

Saunders, A.M., Albertsen, M., Vollertsen, J. and Nielsen, P.H., 2016. The activated sludge ecosystem contains a core community of abundant organisms. The ISME journal, 10(1), pp.11-20.

Savageau, M.A. 1983. Escherichia coli habitats, cell types, and molecular mechanisms of gene control. Am. Nat. 122:732-744.

Schmitz, B.W., Kitajima, M., Campillo, M.E., Gerba, C.P. and Pepper, I.L., 2016. Virus reduction during advanced bardenpho and conventional wastewater treatment processes. *Environmental science & technology*, *50*(17), pp.9524-9532.

Schrad, N., Pensky, J., Gorski, G., Beganskas, S., Fisher, A.T. and Saltikov, C., 2022. Soil characteristics and redox properties of infiltrating water are determinants of microbial communities at managed aquifer recharge sites. FEMS Microbiology Ecology, 98(12), p.fiac130.

Seifert, D. and Engesgaard, P., 2007. Use of tracer tests to investigate changes in flow and transport properties due to bioclogging of porous media. Journal of contaminant hydrology, 93(1-4), pp.58-71.

Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Bürgmann, H., Huber, D.H., Langenheder, S., Lennon, J.T., Martiny, J.B. and Matulich, K.L., 2012. Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology*, *3*, p.417.

Shanks OC, Newton RJ, Kelty CA, Huse SM, Sogin ML, McLellan SL.2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl Environ Microbiol 79:2906 –2913. http://dx.doi.org/10.1128/AEM.03448-12.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell system technical journal, 27(3), pp.379-423.

Shaw, E., Hill, D.R., Brittain, N., Wright, D.J., Täuber, U., Marand, H., Helm, R.F. and Potts, M., 2003. Unusual water flux in the extracellular polysaccharide of the cyanobacterium Nostoc commune. Applied and environmental microbiology, 69(9), pp.5679-5684.

Siegrist, R.L. and Boyle, W.C., 1987. Wastewater-induced soil clogging development. *Journal of environmental engineering*, *113*(3), pp.550-566.

Siegrist, R.L. and Siegrist, R.L., 2017. Treatment using subsurface soil infiltration. *Decentralized Water Reclamation Engineering: A Curriculum Workbook*, pp.547-639.

Siezen, R.J. and Galardini, M., 2008. Genomics of biological wastewater treatment. *Microbial biotechnology*, *1*(5), p.333.

Smil, V., 1999. Nitrogen in crop production: An account of global flows. *Global biogeochemical cycles*, *13*(2), pp.647-662.

Soininen, J., 2014. A quantitative analysis of species sorting across organisms and ecosystems. Ecology, 95(12), pp.3284-3292.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R. and Palmer, C.J., 2000. Sources of Escherichia coli in a coastal subtropical environment. Applied and environmental microbiology, 66(1), pp.230-237.

Spain, A.M., Krumholz, L.R. and Elshahed, M.S., 2009. Abundance, composition, diversity and novelty of soil Proteobacteria. The ISME journal, 3(8), pp.992-1000.

Stefanakis, A., Akratos, C.S. and Tsihrintzis, V.A., 2014. Treatment processes in VFCWs. Vertical flow constructed wetlands, pp.57-84.

Stevik, T.K., Aa, K., Ausland, G. and Hanssen, J.F., 2004. Retention and removal of pathogenic bacteria in wastewater percolating through porous media: a review. Water research, 38(6), pp.1355-1367.

Stewart LW, Reneau RB. Spatial and temporal variation of faecal coliform movement surroundingseptic tank-soil adsorption systems in two Atlantic coastal plain soils. J Environ Qual 1981;10:528–31.

Sul, W.J., Asuming-Brempong, S., Wang, Q., Tourlousse, D.M., Penton, C.R., Deng, Y., Rodrigues, J.L., Adiku, S.G., Jones, J.W., Zhou, J. and Cole, J.R., 2013. Tropical agricultural land management influences on soil microbial communities through its effect on soil organic carbon. Soil Biology and Biochemistry, 65, pp.33-38.

Symonds, E.M., Verbyla, M.E., Lukasik, J.O., Kafle, R.C., Breitbart, M. and Mihelcic, J.R., 2014. A case study of enteric virus removal and insights into the associated risk of water reuse for two wastewater treatment pond systems in Bolivia. *Water research*, *65*, pp.257-270.

Székely, A.J. and Langenheder, S., 2014. The importance of species sorting differs between habitat generalists and specialists in bacterial communities. FEMS microbiology ecology, 87(1), pp.102-112.

Tamura, K. Stecher, G., and Kumar, S., 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. Molecular biology and evolution, 37(4), pp.1237-1239.

Tapia-Torres, Y., Rodríguez-Torres, M.D., Elser, J.J., Islas, A., Souza, V., García-Oliva, F. and Olmedo-Álvarez, G., 2016. How to live with phosphorus scarcity in soil and sediment: lessons from bacteria. Applied and environmental microbiology, 82(15), pp.4652-4662.

Tarayre, C., De Clercq, L., Charlier, R., Michels, E., Meers, E., Camargo-Valero, M. and Delvigne, F., 2016. New perspectives for the design of sustainable bioprocesses for phosphorus recovery from waste. Bioresource Technology, 206, pp.264-274.

Tarayre, C., Nguyen, H.T., Brognaux, A., Delepierre, A., De Clercq, L., Charlier, R., Michels, E., Meers, E. and Delvigne, F., 2016. Characterisation of phosphate accumulating organisms and techniques for polyphosphate detection: a review. Sensors, 16(6), p.797.

TASK, A., 2009. Florida On-site Sewage Nitrogen Reduction Strategies Study.

Thomas, F., Hehemann, J.H., Rebuffet, E., Czjzek, M. and Michel, G., 2011. Environmental and gut bacteroidetes: the food connection. Frontiers in microbiology, 2, p.93.

Thullner, M. and Baveye, P., 2008. Computational pore network modeling of the influence of biofilm permeability on bioclogging in porous media. Biotechnology and Bioengineering, 99(6), pp.1337-1351.

Tian, L. and Wang, L., 2020. A meta-analysis of microbial community structures and associated metabolic potential of municipal wastewater treatment plants in global scope. Environmental Pollution, 263, p.114598.

Tomaras, J., Sahl, J.W., Siegrist, R.L. and Spear, J.R., 2009. Microbial diversity of septic tank effluent and a soil biomat. *Applied and Environmental Microbiology*, *75*(10), pp.3348-3351..

Truu, M., Juhanson, J. and Truu, J., 2009. Microbial biomass, activity and community composition in constructed wetlands. *Science of the total environment*, *407*(13), pp.3958-3971.

Turki, Y., Mehri, I., Lajnef, R., Rejab, A.B., Khessairi, A., Cherif, H., Ouzari, H. and Hassen, A., 2017. Biofilms in bioremediation and wastewater treatment: characterization of bacterial community structure and diversity during seasons in municipal wastewater treatment process. Environmental Science and Pollution Research, 24, pp.3519-3530.

Uksa, M.; Schloter, M.; Kautz, T.; Athmann, M.; Köpke, U.; Fischer, D. Spatial variability of hydrolytic and oxidative potential enzyme activities in different subsoil compartments. Biol. Fert. Soils 2015, 51, 517–521.

Ulén, B., Bechmann, M., Fölster, J., Jarvie, H.P. and Tunney, H., 2007. Agriculture as a phosphorus source for eutrophication in the north-west European countries, Norway, Sweden, United Kingdom and Ireland: a review. Soil use and Management, 23, pp.5-15.

UNESCO, 2021. The United Nations World Water Development Report 2021: Valuing Water. United Nations.

Valentine, D.L. and Reeburgh, W.S., 2000. New perspectives on anaerobic methane oxidation: minireview. Environmental microbiology, 2(5), pp.477-484.

Van Cuyk, S. and Siegrist, R.L., 2007. Virus removal within a soil infiltration zone as affected by effluent composition, application rate, and soil type. Water research, 41(3), pp.699-709.

Van Cuyk, S., Siegrist, R., Logan, A., Masson, S., Fischer, E. and Figueroa, L., 2001. Hydraulic and purification behaviors and their interactions during wastewater treatment in soil infiltration systems. Water Research, 35(4), pp.953-964.

Van Vliet, M.T.H. and Zwolsman, J.J.G., 2008. Impact of summer droughts on the water quality of the Meuse river. Journal of Hydrology, 353(1-2), pp.1-17.

Vandevivere, P. and Baveye, P., 1992. Saturated hydraulic conductivity reduction caused by aerobic bacteria in sand columns. Soil Science Society of America Journal, 56(1), pp.1-13.

Vandewalle JL, Goetz GW, Huse SM, Morrison HG, Sogin ML, Hoffmann RG, Yan K, McLellan SL. 2012. Acinetobacter, Aeromonas and Trichococcus populations dominate the microbial community within urban sewer infrastructure. Environ Microbiol 14:2538 –2552. http://dx.doi.org/10.1111/j.1462-2920.2012.02757.x

Vogt, S.J., Sanderlin, A.B., Seymour, J.D. and Codd, S.L., 2013. Permeability of a growing biofilm in a porous media fluid flow analyzed by magnetic resonance displacement-relaxation correlations. Biotechnology and bioengineering, 110(5), pp.1366-1375.

Volk, E., Iden, S.C., Furman, A., Durner, W. and Rosenzweig, R., 2016. Biofilm effect on soil hydraulic properties: Experimental investigation using soil-grown real biofilm. Water Resources Research, 52(8), pp.5813-5828.

Vuono, D.C., Munakata-Marr, J., Spear, J.R. and Drewes, J.E., 2016. Disturbance opens recruitment sites for bacterial colonisation in activated sludge. Environmental microbiology, 18(1), pp.87-99.

Vymazal, J. and Kröpfelová, L., 2008. *Wastewater treatment in constructed wetlands with horizontal sub-surface flow* (Vol. 14). Springer science & business media.

Wafula, D., White, J.R., Canion, A., Jagoe, C., Pathak, A. and Chauhan, A., 2015. Impacts of long-term irrigation of domestic treated wastewater on soil biogeochemistry and bacterial community structure. Applied and environmental microbiology, 81(20), pp.7143-7158.

Wagner, M., Nielsen, P.H., Loy, A., Nielsen, J.L. and Daims, H., 2006. Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays. *Current opinion in biotechnology*, *17*(1), pp.83-91.

Walsh, C.M., Gebert, M.J., Delgado-Baquerizo, M., Maestre, F.T. and Fierer, N., 2019. A global survey of mycobacterial diversity in soil. Applied and Environmental Microbiology, 85(17), pp.e01180-19.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J.A., Jansson, J.K., Caporaso, J.G., Fuhrman, J.A. and Apprill, A., 2016. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. Msystems, 1(1), pp.e00009-15.

Wang, H., Liu, X., Wang, Y., Zhang, S., Zhang, G., Han, Y., Li, M. and Liu, L., 2023. Spatial and temporal dynamics of microbial community composition and factors influencing the surface water and sediments of urban rivers. Journal of Environmental Sciences, 124, pp.187-197.

Wang, M., Zhu, J. and Mao, X., 2021. Removal of pathogens in on-site wastewater treatment systems: A review of design considerations and influencing factors. Water, 13(9), p.1190.

Wang, X., Hu, M., Xia, Y., Wen, X. and Ding, K., 2012. Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China. Applied and environmental microbiology, 78(19), pp.7042-7047.

Water Quality in Ireland 2013-2018 EPA. 23.

Water, U.N., 2017. 2017 UN World Water Development Report: Wastewater the Untapped Resource.

Water, U.N., 2020. Water and climate change. The United Nations World Water Development Report.

Water, UN., 2022. The United Nations World Water Development Report 2022: groundwater: making the invisible visible

Whitman, R.L. and Nevers, M.B., 2003. Foreshore sand as a source of Escherichia coli in nearshore water of a Lake Michigan beach. Applied and environmental microbiology, 69(9), pp.5555-5562.

Wigginton, S.K., Brannon, E.Q., Kearns, P.J., Lancellotti, B.V., Cox, A., Moseman-Valtierra, S., Loomis, G.W. and Amador, J.A., 2020. Nitrifying and denitrifying microbial communities in centralized and decentralized biological nitrogen removing wastewater treatment systems. *Water*, *12*(6), p.1688.

Winfield, M.D., and E.A. Groisman. 2003. Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. Appl. Environ. Microbiol.69:3687-3694

Wood, S.A., Biessy, L., Latchford, J.L., Zaiko, A., von Ammon, U., Audrezet, F., Cristescu, M.E. and Pochon, X., 2020. Release and degradation of environmental DNA and RNA in a marine system. Science of the Total Environment, 704, p.135314.

World Health Organization, 2019. Water, sanitation, hygiene and health: a primer for health professionals (No. WHO/CED/PHE/WSH/19.149). World Health Organization.

World Health Organization, 2020. State of the world's sanitation: an urgent call to transform sanitation for better health, environments, economies and societies.

World Health Organization, 2021. Progress on household drinking water, sanitation and hygiene 2000-2020: five years into the SDGs.

Wu, G., Zhang, T., Gu, M., Chen, Z. and Yin, Q., 2020. Review of characteristics of anammox bacteria and strategies for anammox start-up for sustainable wastewater resource management. Water Science and Technology, 82(9), pp.1742-1757.

Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M.R., Li, Z., Van Nostrand, J.D. and Ling, F., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nature microbiology, 4(7), pp.1183-1195.

Wu, Y., Cai, P., Jing, X., Niu, X., Ji, D., Ashry, N.M., Gao, C., Huang, Q., 2019. Soil biofilm formation enhances microbial community diversity and metabolic activity. Environment International, 132, p.105116.

Xia, J., Ye, L., Ren, H. and Zhang, X.X., 2018. Microbial community structure and function in aerobic granular sludge. Applied microbiology and biotechnology, 102, pp.3967-3979.

Xia, L., Gao, Z., Xu, H. and Feng, G., 2020. Variations in bacterial community during bioclogging in managed aquifer recharge (MAR): a laboratory study. International Biodeterioration & Biodegradation, 147, p.104843.

Xia, Q., Rufty, T. and Shi, W., 2020. Soil microbial diversity and composition: Links to soil texture and associated properties. Soil Biology and Biochemistry, 149, p.107953.

Xiang, Q., Zhu, D., Qiao, M., Yang, X.R., Li, G., Chen, Q.L. and Zhu, Y.G., 2023. Temporal dynamics of soil bacterial network regulate soil resistomes. Environmental Microbiology, 25(2), pp.505-514.

Xu, Q., Vandenkoornhuyse, P., Li, L., Guo, J., Zhu, C., Guo, S., Ling, N. and Shen, Q., 2022. Microbial generalists and specialists differently contribute to the community diversity in farmland soils. Journal of Advanced Research, 40, pp.17-27.

Yamada, T. and Sekiguchi, Y., 2009. Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured chloroflexi'subphylum i'with natural and biotechnological relevance. Microbes and Environments, 24(3), pp.205-216.

Yang, W.Y., Lee, Y., Lu, H., Chou, C.H. and Wang, C., 2019. Analysis of gut microbiota and the effect of lauric acid against necrotic enteritis in Clostridium perfringens and Eimeria side-by-side challenge model. *PLoS One*, *14*(5), p.e0205784.

Ye, L. and Zhang, T., 2013. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Applied microbiology and biotechnology*, *97*, pp.2681-2690.

Yi, J., Lo, L.S.H. and Cheng, J., 2020. Dynamics of microbial community structure and ecological functions in estuarine intertidal sediments. Frontiers in Marine Science, 7, p.585970.

Yoshimoto, R., Sasaki, H., Takahashi, T., Kanno, H. and Nanzyo, M., 2012. Contribution of soil components to adsorption of Pepper Mild Mottle Virus by Japanese soils. Soil Biology and Biochemistry, 46, pp.96-102.

Young-Rojanschi, C. and Madramootoo, C., 2014. Intermittent versus continuous operation of biosand filters. Water research, 49, pp.1-10.

Zhang, L. and Xu, Z., 2008. Assessing bacterial diversity in soil: a brief review. Journal of soils and sediments, 8, pp.379-388.

Zhang, L., Li, X., Fang, W., Cheng, Y., Cai, H. and Zhang, S., 2021. Impact of different types of anthropogenic pollution on bacterial community and metabolic genes in urban river sediments. Science of The Total Environment, 793, p.148475.

Zhang, L., Shen, Z., Fang, W. and Gao, G., 2019. Composition of bacterial communities in municipal wastewater treatment plant. Science of the Total Environment, 689, pp.1181-1191.

Zhang, L.Y., Zhang, L., Liu, Y.D., Shen, Y.W., Liu, H. and Xiong, Y., 2010. Effect of limited artificial aeration on constructed wetland treatment of domestic wastewater. *Desalination*, *250*(3), pp.915-920.

Zhang, Q., Lv, X., Wei, C., Lu, W., Wang, J., Zhou, Z., Chang, G., Gao, T. and Zhang, H., 2020. Microbial community structure diversity in the dewatered sludge from 4 different waste water treatment plants used for CSRB in colder season. In *E3S Web of Conferences* (Vol. 194, p. 04063). EDP Sciences.

Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F. and Ruan, Y., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS biology, 4(1), p.e3.

Zhang, W., Bougouffa, S., Wang, Y., Lee, O.O., Yang, J., Chan, C., Song, X. and Qian, P.Y., 2014. Toward understanding the dynamics of microbial communities in an estuarine system. PloS one, 9(4), p.e94449.

Zhou, J., Deng, Y.E., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J.D., Yang, Y., He, Z., Wu, L., Stahl, D.A. and Hazen, T.C., 2014. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proceedings of the National Academy of Sciences, 111(9), pp.E836-E845.

Zhou, Z., Meng, Q. and Yu, Z., 2011. Effects of methanogenic inhibitors on methane production and abundances of methanogens and cellulolytic bacteria in in vitro ruminal cultures. Applied and Environmental Microbiology, 77(8), pp.2634-2639.

Zuo, J., Xu, L., Guo, J., Xu, S., Ma, S., Jiang, C., Yang, D., Wang, D. and Zhuang, X., 2023. Microbial community structure analyses and cultivable denitrifier isolation of Myriophyllum aquaticum constructed wetland under low C/N ratio. Journal of Environmental Sciences, 127, pp.30-41.