CHARACTERISING THE FUNCTIONAL ECOLOGY OF SLOW SAND FILTERS THROUGH ENVIRONMENTAL GENOMICS

SARAH-JANE HAIG

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

SCHOOL OF ENGINEERING

COLLEGE OF SCIENCE AND ENGINEERING

UNIVERSITY OF GLASGOW

SEPTEMBER 2014

Abstract

Today the water industry faces a huge challenge in supplying a sustainable, energy efficient and safe supply of drinking water to an increasing world population. Slow sand filters (SSFs) have been used for hundreds of years to provide a safe and reliable source of potable drinking water, with minimal energy requirements. However, a lack of knowledge pertaining to the treatment mechanisms, particularly the biological processes, underpinning SSF operation, has meant SSFs are still operated as "black boxes". This lack of knowledge pertaining to the underlying ecology and ecophysiology limits the design and optimisation of SSFs.

This thesis represents the most comprehensive microbial community survey of full-scale SSFs to-date. Using traditional microbiological methods alongside up-to-date molecular techniques and extensive water quality analyses, specific taxa and community metrics are linked to changes in water quality production. Furthermore, it has been verified that laboratory-scale SSFs can mimic the microbial community and water quality production of full-scale filters. This allowed rigorous experiments pertaining to operational differences, pathogen and novel contaminant removal to be performed. This has revealed, for the first time, that multiple trophic interactions within SSFs are integral to optimal performance.

This thesis has shown that SSFs are phylogenetically and metabolically diverse systems capable of producing high quality water, with the ability to adapt to remove novel contaminants. Using the information gathered, improvements to filter maintenance and operation can be achieved. Future work will apply the microbial and macrobial community dynamics and impact of novel contaminants on filter performance discovered in this thesis into predictive models for water quality.

Acknowledgements

Firstly, I would like to thank my supervisors, in particular Gavin Collins for his guidance and support throughout this project. Thanks are also due to Caroline Gauchotte-Lindsay, for all her support, guidance, help and useful feedback throughout this PhD, especially with chapters 7 and 8. Special thanks also goes to the great technical staff (Bobby Boyd, Ian Scouller, and Tim Montgomery, all the secretaries and administration staff and operators at Scottish Water) without whom the work undertaken in this PhD would not have been possible. An additional thanks goes to Julie Russell and Anne McGarrity (my lab mums), who not only helped me immensely throughout the PhD, but also listened to my moaning, counselled me when I was having a bad day and advised me where possible. Additionally, I gratefully acknowledge the Lord Kelvin and Adam Smith Scholarship for the financial support and allowed me to travel and present my research at many international meetings.

Key to surviving, remaining sane and successfully finishing a PhD is lots of coffee and a good support system, therefore I owe a special thanks to my fellow students, for their constant encouragement and all the good times we had over the years. In alphabetical order: Melina Bautista De Los Santos, Stephanie Connelly, Graeme Edwards (and Kirsten), Marnie Feder (Jay and Gunther), Kazi Hassan, Mathieu Larronde-Larretche, Siding Luo, James Minto, Doug Pender (and Hayley), Ross MacKenzie (and Jay), Ben Nichols, Asha Ram, Andrea Sanchini, Melanie Schirmer, Maria Sevillano Rivera, Eri Tsagkari, and Elisa Vignaga (Seb and Leo).

In particular, my time at the University of Glasgow was made enjoyable in large part due to my great friends; Elisa, Doug, Graeme, Marnie and Melanie, that have became a part of my life and really helped support me through the PhD. I am grateful for the time spent and the memories created during many occasions including: drinking coffee in Artisan Roast

and CottonRake, sampling many whiskies in the Lios Mor, surviving the subcrawl, many Thanksgivings, Burns and meat fondue nights. Thank you all.

Furthermore, for useful advice and discussions at various points in this work including during the Viva, I would like to thank Jillian Couto, Linda D'Amore, Keith Harris, Casey Hubert, Umer Ijaz, John Kenny (and all the CGR team), Mara and Charles Knapp, Tim Pettitt, Ameet Pinto, Seung Gu Shin, and William Sloan.

Finally, I thank my family in particular my parents for instilling in me confidence and a drive for pursuing my PhD and who have encouraged and supported me throughout my life. Last, but certainly not least, I must acknowledge with tremendous and deep thanks, Martin Ellis. Thank you for your proof reading skills, patience, coding and statistical help and generally helping and supporting me through this process - it truly would not have been possible without you and words cannot convey my gratitude. You truly deserve a medal!

To conclude, I would like to dedicate this work to my Grandma, Eileen Sawyer, who always told me to never stop questioning things and that I could do anything if I put my mind to it - something summed up beautifully by Albert Einstein. I hope that this work makes you proud.

"Learn from yesterday, live for today, hope for tomorrow.

The important thing is to not stop questioning."

Albert Einstein: Relativity: The Special and the General Theory

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Simplicity is the ultimate sophistication."

Leonardo da Vinci

The requirement for access to safe drinking water is a basic human right [United Nations General Assembly, 2010] and an important factor contributing to a decrease in morbidity and mortality in developing countries [Van Leeuwen, 2000]. This, alongside the dissipation of fossil fuels and the harsh economic times currently faced by the world motivate the search for energy-efficient water treatment technologies which meet stringent drinking water standards. Therefore, there is a great necessity to adopt a water treatment scheme that meets these requirements.

1.1 Drinking Water Purification

Water purification is the process of removing undesirable chemicals, biological contaminants, suspended solids and gases from contaminated water. The goal is to produce water fit for a specific purpose. Most water is purified for human consumption (drinking water), but water purification may also be designed for a variety of other purposes, including medical, pharmacological, chemical, horticultural and industrial applications. It is also important to emphasise that access to adequate sanitation and water are inextricable, with each exacerbating the other, with water scarcity often being a problem of water quality as well as quantity [Bauer, 2004]. Water quality is, in essence, an issue of sanitation that occurs from the widespread presence of contaminants in our waterways. There are many sources of such contaminants, however, most are caused by human activities, such as:

1. Discharge of untreated sewage containing chemical wastes, nutrients, and suspended matter. Discharge includes direct input from animals or open sewage sources as well as leakage or poor management of sewage systems.

2. Industrial discharge of chemical wastes and byproducts.

3. Surface runoff from agriculture, construction sites, and mines, which result in the release of pesticides, herbicides, fertilisers, petroleum products, and heavy metals.

All or a combination of such pollution events lead to the following contaminants, which have significant issues for human health, wildlife or the environment:

- Enteric bacteria: e.g. *Escherichia coli - E.coli*;
- Heavy metals;
- Organic and synthetic compounds;
- Pesticides;
- Pharmaceutical compounds;
- Protozoan parasites: *Cryptosporidium and Giardia*;
- Viruses.

1.1.1 Regulating Drinking Water Quality in the UK

All drinking water in the UK, whether from public supplies or other sources, has to meet strict quality standards laid down in UK regulations derived from the EU Drinking Water Directive (98/83/EC). This directive sets out standards for a wide range of chemical, physical and microbiological parameters and a system for how best to monitor these parameters (Standard Methods regulated by the International Organisation for Standardisation). The directive is reviewed at least every five years by the European Commission in order to take account of changes in the World Health Organisation (WHO) guidelines. Briefly the directive states that drinking water must be *"wholesome and clean: free from any micro-organisms and parasites and from any substances which, in numbers or concentrations, constitute a potential*

danger to human health." [European Union Council Directive, 1998]. In the UK these requirements are strictly followed, with some parameters being more stringent than defined by the directive, reflecting the high standard of water supplies in the UK.

1.1.2 Drinking Water Purification Methods

Generally, the treatment of drinking water takes place in several steps to remove dissolved and suspended solids, often involving a combination of physical (filtration, sedimentation, ion-exchange and distillation), biological (slow sand filtration, biologically active carbon) and chemical (flocculation, chlorination, ozonation and UV treatment) methods. The combination of purification methods used depends upon the source of the water to be purified, economic constraints and demand, with ground water (aquifers and water locked away in polar caps and glaciers) requiring less purification than surface water (lakes, rivers, reservoirs and impoundments).

Typically in the UK the source used for drinking water comes from surface or ground water aquifers. In order to make it fit for human consumption it is impounded in large reservoirs, with residence times of 3-4 weeks, where there is some self-purification from sunlight, and from settling of particulate matter and attached bacteria. This is then normally followed by storage in an additional sedimentation basin after adding a flocculent or coagulant, and then rapid filtration through sand (depth ranging from 0.4 to 1.2 m) to remove micro-organisms and turbidity. The pH of the water is then adjusted and disinfected with chlorine before being sent to the consumer via the water distribution network (Figure 1.1). It should be stressed that these processes are all extremely energy intensive. For example 4% of the energy consumption in the United States in 2009 was used for drinking water purification [U.S. Environmental Protection Agency, 2009a], a process which releases 52 million metric tonnes of CO_2 a year [McMahon and Price, 2011]. There are less energy intensive and carbon rich solutions which could be adopted and must be explored as the above mentioned energy consumption is predicted to rise by 50% to 6% by 2020 if less energy intensive purification methods are not implemented [Spellman, 2013].

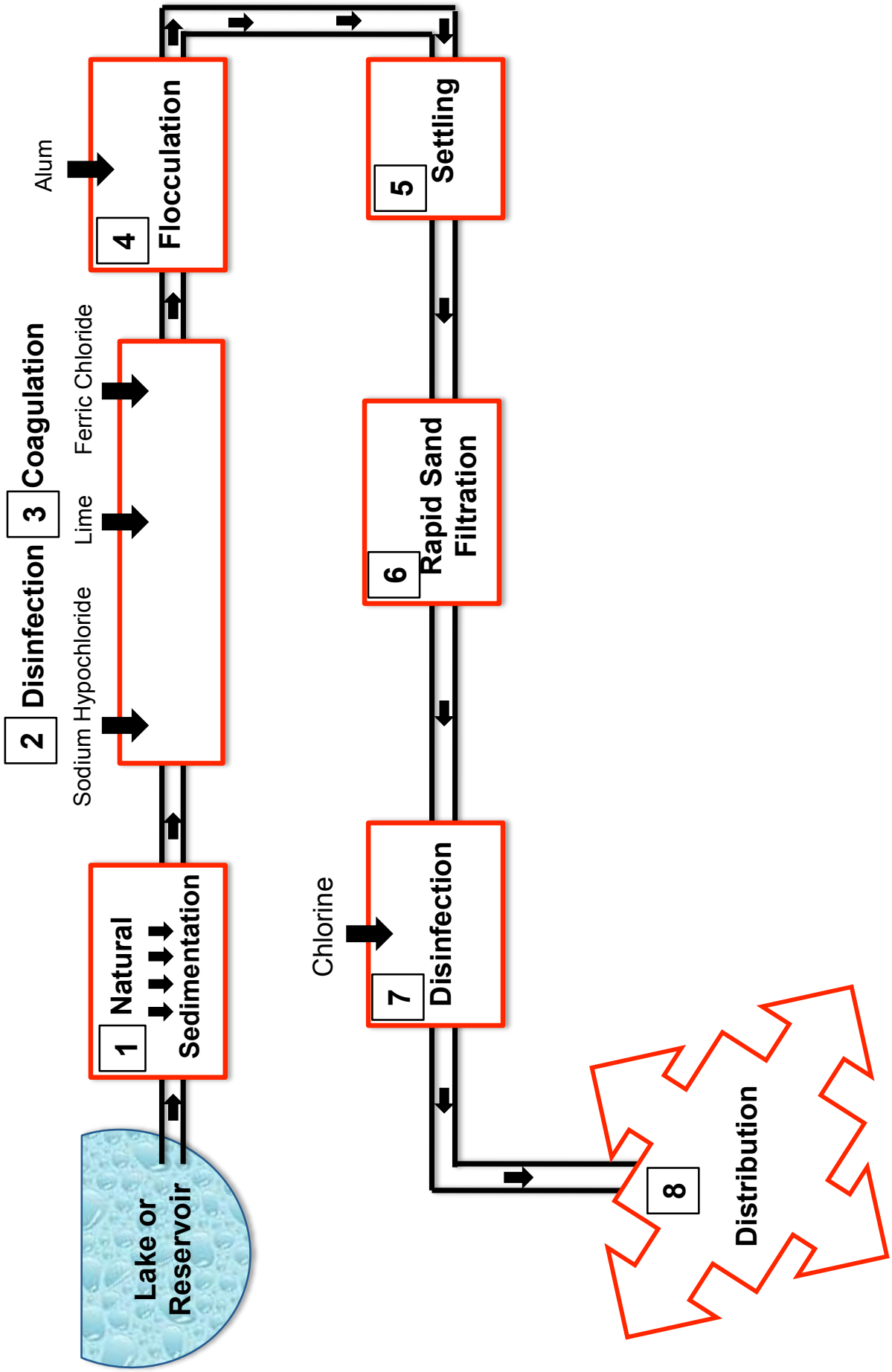


Figure 1.1: Diagram illustrating the drinking water treatment process

1.2 An Inexpensive and Less Energy Intensive Solution

For over 200 years, slow sand filtration has been used as an effective means of treating water for the control of microbiological contaminants, particularly for small community water supplies. However, such systems lost popularity to rapid sand filters which have smaller land requirements and are less sensitive to temperature and water quality variations [Huisman et al., 1974]. In recent years, there has been renewed interest in slow sand filter application, particularly because these systems do not require chemicals or electricity to operate and yet can achieve a high level of treatment. Additionally, unlike other purification methods, slow sand filters (SSFs) are relatively simple and easy to operate. It should however be stressed that this does not mean the processes involved are simple or less complex, just that they are not fully understood.

Several microbiologically mediated purification mechanisms (e.g., predation, scavenging, adsorption and bio-oxidation) have been hypothesised or assumed to occur within biofilms that form in the filter, but have never been comprehensively verified [Huisman et al., 1974, Ellis and Wood, 1985, Haarhoff and Cleasby, 1991, Fogel et al., 1993, Lloyd, 1996, Bahgat et al., 1999]. Such a gap in knowledge pertaining to the ecology and potential of SSFs to remove various pollutants has and will continue to hamper advances in the design and optimisation of slow sand filters.

Initially the role of biological purification within slow sand filters was hypothesised and was largely based on empirical observations [Huisman et al., 1974, Baker and Taras, 1948]. Since then, most SSF research and development has always assumed that "biological purification" would occur and focussed on: (a) pre-treatment methodology (particularly for application in developing countries) [Bellamy et al., 1985, Weber-Shirk and Dick, 1997b, Dorea, 2013]; (b) process development, performance and modelling [Ojha and Graham, 1994, Campos et al., 2002, Sadiq et al., 2004, Campos et al., 2006]. Some work has also been carried out on the ecological aspects of biological treatment; however much of this has been based on hypothesising about the biological treatment offered by SSFs, treating them as engineered "black boxes". As reviewed in Haig et al. [2011] there have been a number of studies which have attempted to characterise the purification mechanisms and the microbes responsible for them; however such studies have suffered from limitations in the approach or of the

techniques availability. Even recently, many of these investigations have been limited by a focus on specific elements of the filter, such as the schmutzdecke (from the German "dirt layer" a complex biological layer formed on the top of the SSF bed) [Campos et al., 2002, Rooklidge et al., 2005, Unger and Collins, 2008], or on specific biological processes, such as denitrification [Aslan and Cakici, 2007] and predation [Weber-Shirk and Dick, 1999]. Furthermore, most research so far has been limited to the microbes (and their associated processes) that could be cultured using traditional microbiological techniques; the role of uncultivable microorganisms has yet to be determined. Apart from one study [Calvo-Bado et al., 2003], the microflora of these filters has never been studied and their individual roles in purification never determined. One of the main limitations of these studies is that they have been performed in laboratory-scale microcosms with carefully controlled parameters and hence are not necessarily representative of the complex and diverse microbial community that full-scale biological systems are believed to support. New techniques to understand microbial ecology could address many of the limitations identified thus far.

1.3 Understanding Microbial Ecology

The term ecology comes from the Greek *oekologie* meaning "the study of the household of nature" and was first coined in 1866 by the German scientist Ernst Haeckel to explain the interactions between microbes and their environment [Konopka, 2009]. Therefore the primary goal of ecology is to measure, understand, and predict biodiversity and functional diversity of an ecosystem. Historically, ecological studies were performed in laboratory-scale microcosms to answer questions like: how are ecosystems assembled and how do species that make up a community arrive, survive, interact and succeed in a community? [Purdy et al., 2010]. However, understanding and answering these questions was extremely difficult and only really made possible in the 1950s when advances in molecular microbial ecology (Figure 1.2), such as the development of the polymerase chain reaction (PCR) were made. PCR made it possible for the first time to directly interrogate the genetic information of individual microorganisms and entire communities. This led to developments in obtaining and working with mRNA which have revolutionised the ways in which functional genes are determined. Further, microautoradiography coupled with fluorescence *in situ* hybridisation (FISH) and stable-isotope probing (SIP), makes it possible to detect the function of particular genes in

the community.

From the perspective of SSFs, such advancements place scientists and engineers at a juncture, which will allow them to answer both the traditional microbial ecology questions regarding community composition and assembly but also more complex questions pertaining to how to manage the SSF microbial community to improve performance and pollutant removal capabilities. Such understanding will allow slow sand filters to be designed and operated in a tailored manner, specific to water quality needs and requirements.

1.4 Thesis Statement

In order to improve the operation and design of slow sand filters, a greater understanding of the microbial community and the processes they perform is required, alongside determining the capabilities of these filters to remove new pollutants. This thesis will address the following questions:

1. Which microorganisms are present in full-scale industrially operated slow sand filters and what roles do they perform?
2. Does the microbial community structure change temporally and spatially in SSFs?
3. Can a laboratory-scale slow sand filter be constructed to mimic the performance and microbial community of full-scale industrially operated slow sand filters?
4. What is the impact of light on the microbial community and filter performance?
5. Which mechanisms are responsible for the removal of the human pathogen *E.coli* in slow sand filters?
6. How effective are slow sand filters at removing estrogen and can their performance be improved by bioaugmentation?

1.5 Publications

Journal Papers

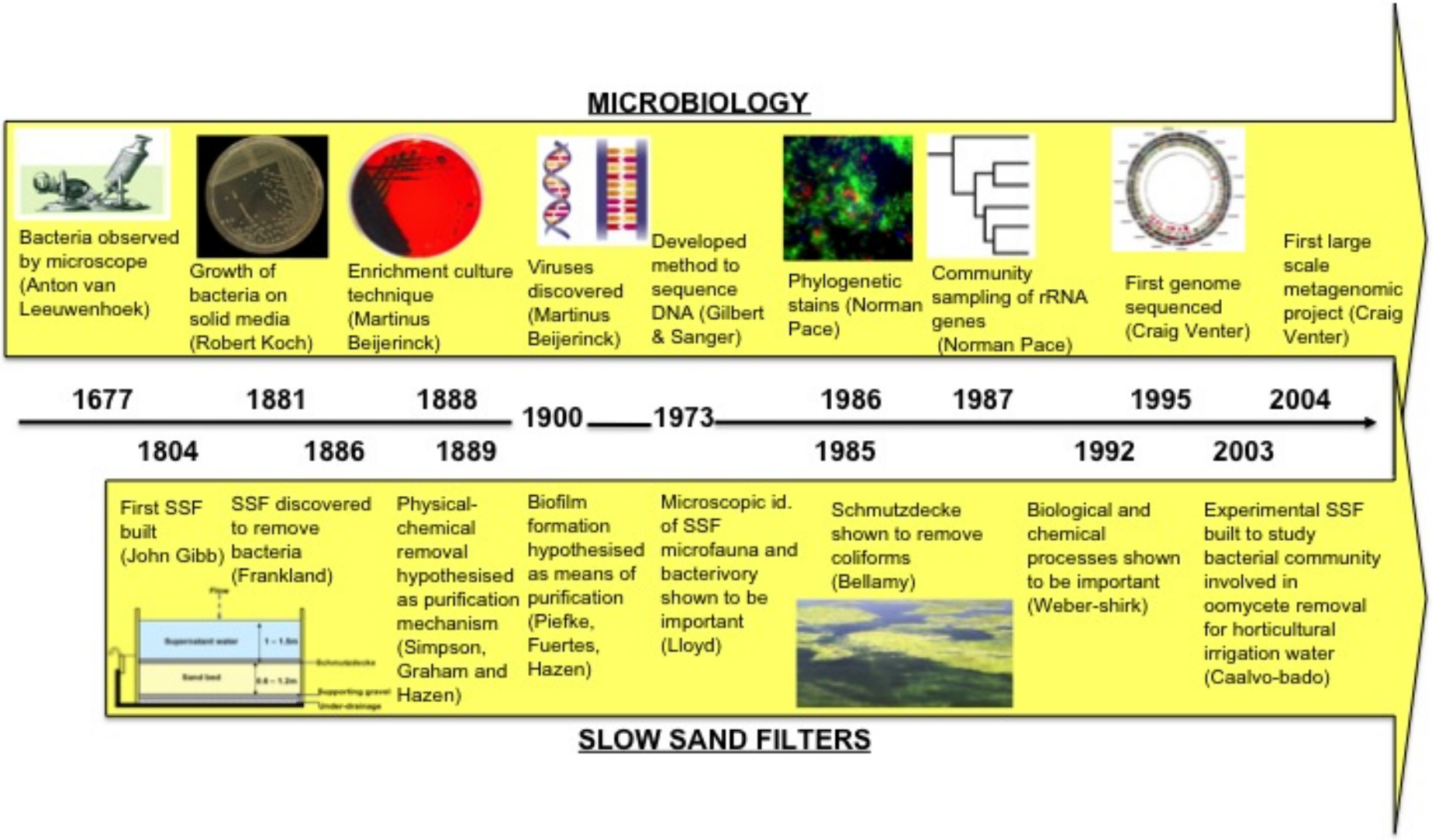


Figure 1.2: Timeline of developments made in molecular microbiology and the biological understanding of slow sand filters

- S. Haig, G. Collins, R. Davies, C. Dorea, and C. Quince. (2011). Biological Aspects of Slow Sand Filtration: Past, Present and Future. *Water Science & Technology: Water Supply*, 11 (4):468-472

- S. Haig, C. Quince, R. Davies, C. C. Dorea, and G. Collins. (2014). Validating the Performance and Microbial Community of Laboratory-Scale Slow Sand Filters with respect to Full-Scale Industrial Filters. *Water Research*, 61: 141-151

- S. Haig, C. Gauchotte-Lindsay, G. Collins, R. Davies, and C. Quince. (2014). *Progress in Slow Sand and Alternative Biofiltration Processes: Further Developments and Applications.*, Chapter 28: Bioaugmentation Reduces the Negative Effect of Estrogens on Coliform Removal in SSFs. IWA Publishing

- S. Haig, M. Schirmer, R. D'Amore, J. Gibbs, R.L. Davies, G. Collins, and C. Quince. (2014). Stable-Isotope Probing and Metagenomics Reveal Predation by Protozoa Drives *E.coli* Removal in Slow Sand Filters. *Accepted by ISME Journal*

- S. Haig, C. Quince, R. Davies, C. Dorea and G. Collins. Spatial and Temporal Microbial Community Analysis Identifies Functionally Relevant Microbes for Slow Sand Filter Performance. *Under review in mBio*

- S. Haig, A. Sewell, U.Ijaz, R. Marquez, G. Collins, C. Quince and C. Gauchotte-Lindsay. BODIPY Fluorescent Tagging of Emerging Contaminants for Rapid Isolation of Degrading Microbes. *In preparation*

- S. Haig, C. Gauchotte-Lindsay, G. Collins and C. Quince. Bioaugmentation Mitigates the Impact of Estrogen on Coliform-Grazing Protozoa. *In preparation*

Conference Publications

- S. Haig, G. Collins, R. Davies, C. Quince and C. Gauchotte-Lindsay. Bioaugmentation Reduces the Negative Effects that Estrogen Exposure has on the Pathogen Removal Capacity of Slow Sand Filters. Presented at the 15th International Society for Microbial Ecology (ISME) Conference, Seoul, Korea, August 2014. [Poster]

- S. Haig, G. Collins, R. Davies, C. Quince and C. Gauchotte-Lindsay. Bioaugmentation Reduces Negative Effect of Estrogens on Coliform Removal in SSFs. Presented at the

International Slow and Alternative Biological Filtration Conference, Nagoya, Japan, June 2014. [Talk]

- S. Haig, G. Collins, R. Davies, and C. Quince. Exploring Coliform Removal in Slow Sand Filters using DNA-SIP Coupled with Metagenomics. Presented at Microbial Ecology in Water Engineering (MEWE), Ann Arbor, USA, July 2013. [Talk]

- S. Haig, G. Collins, R. Davies, and C. Quince. Pathogen Removal in Slow Sand Filters as Revealed by Stable Isotope Probing Coupled with Next Generation Sequencing. Presented at the 14th International Society for Microbial Ecology (ISME) Conference, Copenhagen, Denmark, August 2012. [Talk]

- S. Haig, C. Dorea, G. Collins, R. Davies, and C. Quince. Validating Laboratory Slow Sand Filtration Studies Through Water Quality and Molecular Analysis. Presented at Particle Separation, Berlin, Germany, June 2012. [Talk]

- S. Haig, G. Collins, R. Davies, C. Dorea, and C. Quince. Biological Aspects of Slow Sand Filtration: Past, Present and Future. Presented at UK National Young Water Professionals, Edinburgh, Scotland, April 2011. [Best Poster Prize]

1.6 Outline

This dissertation is structured as follows:

Chapter 2 presents a detailed literature review, summarising the various aspects of SSFs including the fundamental theory, design, operation, maintenance and previous studies.

Chapter 3 presents a detailed review of the molecular techniques that are deployed to understand microbial communities.

Chapter 4 presents molecular (qPCR and next-generation sequencing) and water quality analysis of two full-scale industrially operated slow sand filters which were sampled periodically for eight months. This analysis links various water quality parameters to specific organisms and demonstrates both temporal and spatial changes in the microbial community,

providing unprecedented insight into the organisms that reside in real filters.

Chapter 5 presents the design, construction and operation of laboratory-scale slow sand filters. This chapter describes a proof-of-concept laboratory-scale unit which accurately mimic full-scale industrially operated filters in terms of both water quality and microbial consortia. The work in this chapter is a prerequisite for the subsequent work, demonstrating that findings in the following chapters are relevant and applicable to industrially operated filters.

Chapter 6 uses the laboratory filters described in Chapter 5 to examine the effect of light on slow sand filter performance and its microbial community; from an engineering perspective, this is to determine if there are differences between covered (used in the Netherlands) and uncovered filters (used in the UK and USA). This chapter further examines how the pathogen *E.coli* is removed by deploying stable-isotope probing (SIP) in combination with metagenomics. Information obtained from such work could help improve the operation of SSFs in the future.

Chapter 7 uses the laboratory filters described in Chapter 5 to explore the ability of slow sand filters to remove natural estrogens (estrone, estradiol and estriol), which have been newly designated by WHO to be harmful to wildlife and human health and recently added to the EU Drinking Water Directive (98/83/EC). Further, this chapter explores the possibility of improving estrogen removal by bioaugmentation with three estrogen metabolising bacteria (obtained via enrichment culture from the industrial SSFs discussed in Chapters 4 and 5).

Chapter 8 explores the deleterious effects of natural estrogens on different protozoa species, providing a potential reason for the reduced coliform removal ability observed in Chapter 7.

Chapter 9 provides a summary of the contributions and findings of this dissertation and explores avenues for future work.

Chapter 2

A Review of Slow Sand Filtration

"In every glass of water we drink, some of the water has already passed through fishes, trees, bacteria, worms in the soil, and many other organisms, including people... Living systems cleanse water and make it fit for human consumption."

Elliot A Norse, (Animal Extinctions)

For over 200 years slow sand filtration has been an effective means of treating water for the control of microbiological and chemical contaminants in both small and large community water supplies [Huisman et al., 1974, Haig et al., 2011]. However, due to advancements in engineering, various other methods, which require less land area, such as rapid sand filtration [Huisman et al., 1974] have become the technology of choice. In recent years there has been renewed interest in slow sand filter application, particularly because of its independence of fossil fuels and its efficiency at removing bacteria, viruses, cysts, amoeba, zoospores and various chemical contaminants [Rooklidge et al., 2005, Hijnen et al., 2007, Elliott et al., 2008].

Slow sand filters (SSFs) are typically composed of a 1-2m deep porous medium (sand) filter bed, through which the water to be purified percolates. The operational flow rate of these systems range from 0.1-0.2$m^3m^{-2}h^{-1}$ which is a function of the dimensions of the filters which can be rectangular or cylindrical in cross section. Although they are often the preferred technology in many developing countries, they are also used to treat water in developed countries (e.g., the UK where they are used to treat water supplied to London and Edinburgh). Further-

A condensed version of this chapter is published: Haig, S. Collins, G. Davies, R.L. Dorea, C.C. and Quince, C. (2011). Biological aspects of slow sand filtration: past, present and future. ***Water Science & Technology: Water Supply*****, 11 (4):468-472**

more, their capability to efficiently remove various contaminants has seen SSF deployment in various areas out with drinking water purification including: aquaculture [Arndt and Wagner, 2004], horticulture [Calvo-Bado et al., 2003], storm-water purification [Urbonas, 1999] and food and drink waste management [Ramond et al., 2013]. Irrespective of the adoption and utilisation of SSFs in producing energy efficient and high quality water [Lloyd, 1974], little is still understood about the functional ecology, i.e., biological mechanisms and organisms responsible for producing the diverse and efficient functional capacity of SSFs [Haig et al., 2011]. This lack of knowledge has and will continue to halt optimisation in design, management and operation of these systems.

Recently, there have been a number of studies that have attempted to characterise the purification mechanisms in SSFs and the microbes responsible for them [Weber-Shirk and Dick, 1997a, Bahgat et al., 1999, Calvo-Bado et al., 2003, Aslan, 2008, Wakelin et al., 2011, Ramond et al., 2013]. However, such studies have focused on specific aspects of SSFs, for example the schmutzdecke [Wakelin et al., 2011] or specific purification mechanisms e.g. nitrate removal [Aslan, 2008] and, with the exception of Haig et al. [2014], have been performed in non-verified laboratory-scale SSF microcosms [Burman, 1978, Weber-Shirk and Dick, 1997a,b], which may not accurately reflect the true microbial community found in real SSFs. Furthermore, these experiments have relied upon conventional plating and isolation techniques which do not allow the study of non-culturable and fastidious species generally thought to dominate environmental samples [Roszak and Colwell, 1987]. Direct methods such as pyrosequencing, denaturing gradient gel electrophoresis (DGGE) [Calvo-Bado et al., 2003], fluorescent in-situ hybridisation (FISH) and quantitative PCR (qPCR) overcome this limitation and will hopefully allow the complex ecological processes and interactions which take place in SSFs to be understood [Haig et al., 2011].

Although all of these studies have provided insight into the biological processes occurring within SSFs, a deeper analysis of the structure and dynamics of the microbial community underpinning slow sand filters as a function of performance and operational conditions is needed. Such a study has the potential to reveal important and under-appreciated structure-function relationships, which could greatly improve operation, management and design of these systems.

2.1 History of Slow Sand Filtration

Slow sand filtration or biological filtration is one of the earliest forms of potable water treatment, with its origins being traced back 4000 years to the Sanskrit text, "Sus'ruta Samhita" which documented the filtration of water through sand [Thomas, 1883]. This procedure was adopted and further developed by many civilisations including the Egyptian and Romans, where sand filter-cisterns have been documented [Lloyd, 1974]. However, slow sand filtration as recognised today dates from 1804 when John Gibb designed and built an experimental SSF for his bleachery in Paisley and sold the surplus treated water to the public [Baker and Taras, 1948]. This filter was designed based on adaptions of the Egyptian, Roman and French systems. Following the success of Gibb, slow sand filtration was further developed by Robert Thorn and then later by James Simpson who implemented the first public supply at the Chelsea Water Company, London, in 1829. Furthermore, following the cholera epidemic which devastated London in the mid 1800s it became a legal requirement to use SSFs to filter all water extracted from the River Thames within five miles of St. Paul's Cathedral [Ellis and Wood, 1985].

After the pioneering work of Gibb, Thorn and Simpson numerous improvements were made to SSFs, specifically pertaining to their construction with the first mechanised filter being installed in 1885. Today, SSFs are generally the third stage of water purification after reservoir storage and rapid filtration, and prior to disinfection [Ellis and Wood, 1985]. However, slow sand filters can also provide a single-stage treatment for raw waters within certain water quality limits of turbidity and algal content [Campos et al., 2002] and can be found in numerous cities around the world, including Amsterdam, Antwerp, London, Paris, Nagoya and Stockholm. Unlike conventional and more sophisticated water treatment methods SSFs are inexpensive, highly efficient, easy to operate and eliminate virtually all turbidity from the water together with much of the organic matter originally present. More importantly, SSFs can remove a high proportion of coliforms, pathogenic bacteria, viruses and distinct from rapid sand filters, various parasites including *Cercariae* and *Schistosomes* (Table 2.1).

However, despite its importance in providing safe, efficient and cheap water purification the fundamental biological mechanisms responsible for treatment in SSFs are poorly understood. This lack of knowledge may be partially due to the disadvantages of SSFs, such as

Table 2.1: Performance Summary of SSFs (adapted from Gimbel and Collins [2006])

Water Quality Parameter	Abbreviation	Removal Capacity	Reference
Assimilated Organic Carbon	AOC	14-40%	Lambert and Graham (1995)
Biological Dissolved Organic Carbon	BDOC	46-75%	Lambert and Graham (1995)
Cercaria		100%	Ellis (1985)
Chemical Oxygen Demand	COD	15-25%	Haaroff and Cleasby (1991)
Cryptosporidium		99.90%	Hijnen *et al.*, (2007)
Dissolved Organic Carbon	DOC	5-40%	Lambert and Graham (1995)
Dissolved Organic Matter	DOM	25-75%	Graham (1999)
Enteric bacteria		90-99.9%	Hijnen *et al.*, (2007)
Enteric viruses		99-99.9%	Poynter and Slade (1977)
Giardia cysts		99-99.9%	Bellamy *et al.,* (1985)
Iron		30-90%	Ellis (1985)
Manganese		30-90%	Ellis (1985)
Nitrate		95%	Aslan (2008)
Pesticides		0-100%	Lambert and Graham (1995)
Total Organic Carbon	TOC	15-25%	Haaroff and Cleasby (1991)
Colour		25-40%	Ellis (1985) and Smet and Vissher (1989)
Turbidity		90-98%	Smet and Vissher (1989)
Zoospores		99-100%	Calvo-Bado *et al.,* (2003)

the large land area required, the reduced run length with increased turbidity in raw water and the high cost involved in cleaning the filters [Ellis and Wood, 1985]. In recent years there has been a resurgence of interest in SSFs, mostly because SSFs are not heavily reliant on fossil fuel supply and provide excellent removal of cysts of *Giardia* and *Cryptosporidium* and dissolved organic matter (DOM) after preoxidation [Graham, 1999] (Table 2.1). Regardless of the renewed interest in slow sand filtration, the lack of knowledge pertaining to the removal mechanisms, specifically the ecological processes involved, has and continues to inhibit development and expanded application of these systems.

2.2 Elements of a Slow Sand Filter

In order to construct and operate a successful slow sand filter there are four basic components (Figure 2.1) which are required:

1. A supernatant (raw) water layer. Principle role of which is to maintain a constant level of water above the filter medium providing the pressure needed to carry the water through the filter. This water supply also provides a source of micro- and macro-organisms which form the biological components of these filters, which aids in majority of the systems purification mechanisms.

2. A sand bed which is the location of majority of the purification processes. The sand is

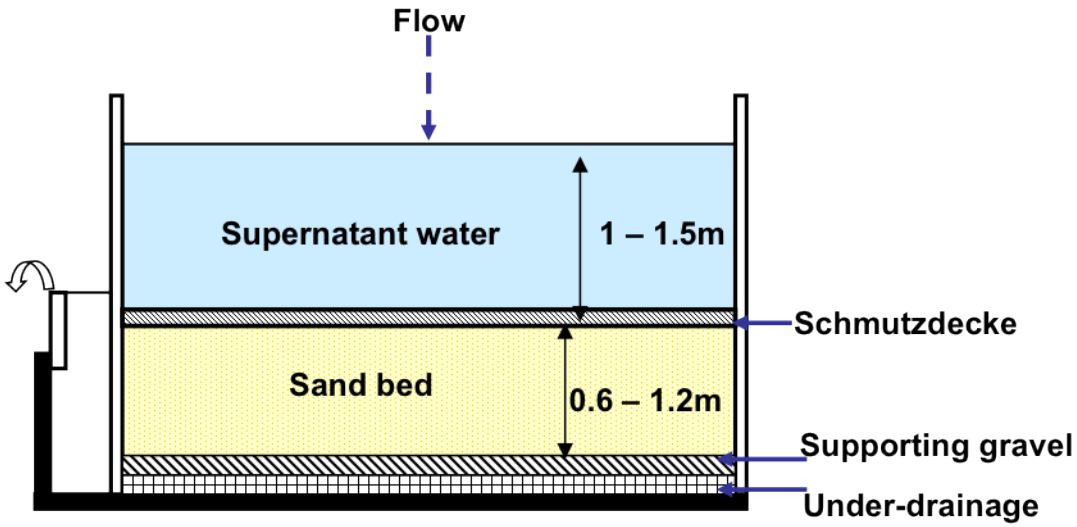


Figure 2.1: Schematic representation of a slow sand filter, adapted from Huisman et al. [1974]

usually of fine grain (0.15-0.3mm) size.

3. An under-drainage system which functions in conjunction with the sand bed. This system may consist of a false floor of porous concrete or an array of porous or unjointed pipes surrounded and covered with graded gravel to support the sand bed and prevent fine grain entering the drainage system.

4. A flow control system which regulates the velocity of flow through the sand bed in order to prevent the raw water level dropping below a predetermined level during operation.

The first three of these features are contained within a single open-topped filter box, the flow control valves being normally in adjacent structures. The box is typically rectangular and ranges in size from 2.5-4m in depth and is typically built entirely underground. The general appearance of a slow sand water filter plant can be seen in Figure 2.2

2.3 The Modes of Action in Slow Sand Filters

Several physicochemical and biological mechanisms have been proposed as responsible for the removal of particles, microorganisms and other substances (e.g., organic matter) during filtration. Biological mechanisms are those requiring (or which are enhanced by) the biological activity of the microorganisms in suspension or colonising the filter media [Weber-Shirk and Dick, 1997a]; these include predation, scavenging, decomposition and the bactericidal

Figure 2.2: Slow sand filter plant at Fairmilehead Water Treatment Plant in Edinburgh.

effects of sunlight. Physico-chemical purification mechanisms are defined as those which do not require biological activity to take place within the filter [Weber-Shirk and Dick, 1997b]. The physicochemical mechanisms taking place in SSFs have been extrapolated from rapid sand filtration theory [Cranston and Amirtharajah, 1987]. These are better understood than the biological processes within the filter bed.

The first purification mechanisms are thought to take place in the supernatant (Figure 2.1), where the levels of sunlight and nutrients allow algae to proliferate, absorbing carbon dioxide, nitrates, phosphates, and releasing oxygen. The latter reacts with organic impurities forming inorganic salts (e.g., sulphates, nitrates, and phosphates). In addition, nitrogenated compounds are oxidised forming nitrates that are easily assimilated by algae [Huisman et al., 1974, Wotton, 2002]. Wotton [2002] pointed out that that exopolymers secreted from organisms may promote the flocculation and aggregation of particles within the supernatant.

On top of and within the sand bed of the slow sand filter a diverse ecology of micro- and macroorganisms have been hypothesised to contribute to the overall biological treatment. The biological purification phenomena in SSFs have been reviewed by [Haarhoff and Cleasby, 1991] and form the basis of the mechanisms subsequently described.

In order to explain the various processes involved in slow sand filtration, the passage of the raw water through the biological filter and the different purifying methods that it undergoes

will be discussed. Firstly, the sample enters the supernatant water (Figure 2.1) and moves due to gravitational drainage through the sand bed, a process which takes between 3-12 hours depending upon the filtration velocity. As the water percolates through the sand, organic material and microorganisms are removed [Ellis and Wood, 1985, Fogel et al., 1993]. This removal is due to both mechanical and biological processes, namely the slow filtration rate of the water, the small granular size of the sand used and also biological processes such as predation, natural death and metabolic breakdown [Haarhoff and Cleasby, 1991, Bahgat et al., 1999].

2.3.1 Physical Processes

Various particles such as minerals, microorganisms and amorphous debris are removed via filtration, with particle removal efficiency being documented as reaching 99.99% for mature SSFs [Bellamy et al., 1985] especially in waters of turbidity lower than 10 NTU and colour less than 5 CU [Sharpe et al., 1994]. In general physical filtration can be divided into three categories: straining, sedimentation and absorption. Straining takes place at the sand surface on particles which are too large to enter into the sandbed. Sedimentation occurs within the pore space (spaces between grains) of the SSFs and removes particles which are smaller than the pore space by settling on the sand grains. Absorption is a physicochemical removal process which favours dissolved substances and colloidal (a solution that has particles ranging between 1 and 1000 nanometers in diameter, yet are still able to remain evenly distributed throughout the solution) suspensions. The success of absorption is determined by surface forces (e.g., Van der Waals forces and electrostatic interactions) between the substance to be removed and the sand grains. For example, metals in solution (which are positively charged) are readily absorbed by quartz sand due to their negative charge. These physical processes are important, however biological processes are also integral to purification.

2.3.2 Biological Processes and the Schmutzdecke

As previously mentioned pathogenic microorganisms such as bacteria, cysts, viruses and parasites can be efficiently removed by SSFs [Poynter and Slade, 1978, Graham, 1999]. This can be explained by the long hydraulic retention time of the water above the sand bed, which allows organic matter and particles to be deposited on top of the sand, allowing

the development of a substantial biological community [Huisman et al., 1974] to form, in particular an algal mat known as the schmutzdecke. The schmutzdecke consists of threadlike algae, diatoms, plankton, protozoa, rotifers and bacteria, as shown in Figure 2.3. This layer is intensively active with the various organisms entrapping, digesting and breaking down organic matter contained within the water. For example, Bellamy et al. [1985] showed that the schmutzdecke was important for the removal of coliforms. Once the raw water has passed through the schmutzdecke it enters the top layer of sand in which a biofilm develops. Within these layers a number of biological processes occur which aid in the removal of organic matter, pathogens and chemicals, these include:

1. **Predation** - which was shown by Haarhoff and Cleasby [1991] to occur due to the algae and diatoms that were found in the guts of benthic invertebrates. Further, Lloyd [1996] and Weber-Shirk and Dick [1999] presented strong evidence of bactivory (ingestion of bacteria) by protozoa. Such predation likely occurs on the surface of the sand grains or by suspension feeding predators removing suspended particles and bacteria.

2. **Scavenging** - A considerable amount of detritus is scavenged mostly by aquatic worms in the lower strata of slow sand filters [Haarhoff and Cleasby, 1991]. In the schmutzdecke macro-invertebrates, e.g., oligochaetes and larval midges, feed on microorganisms, exopolymers, and a range of detritus particles [Wotton, 2002].

3. **Metabolic breakdown** - which occurs within the schmutzdecke and lower sand layers and accounts for the partial reduction in organic carbon levels. The bacterial population retrieves energy for growth and metabolic functions (assimilation) through microbiological oxidation of available organic matter. Die-off also occurs, liberating organic matter that is utilised by other organisms at lower depths [Huisman et al., 1974].

4. **Adsorption** - which is *per se* a physicochemical process. Nevertheless, Lloyd [1996] suggested that protozoan grazing of attached bacteria was probably playing an important role in maintaining sand surface area available for further adsorption. Hence, it cannot be seen as an exclusively physicochemical process, as it can be influenced by biological activity.

5. **Bactericidal effect of sunlight and algae** - Radiation could affect bacteria in the su-

Figure 2.3: Schmutzdecke layer of a drained industrially operated slow sand filter in Edinburgh. ***A*** *Shows an overview of the schmutzdecke over the full filter bed,* ***B-D*** *show magnified images of the schmutzdecke.*

pernatant and extracellular algal products which can increase bacterial mortality over long periods, although these (speculated) mechanisms are not proven to occur or contribute significantly in filtration [Haarhoff and Cleasby, 1991]. In addition to the bactericidal effect of sunlight, Wotton [2002] suggested that UV light can also add to the breakdown of dissolved organics into by-products that are more susceptible to bacterial assimilation.

2.3.3 Biofilms in Slow Sand Filters

Throughout history microorganisms have commonly been classified in the planktonic form, freely floating and suspended in an aqueous medium. It was not until 1664 when Van Leeuwenhoek observed that microbial cells aggregate on tooth surfaces [Madigan et al., 2011] that microbial biofilms were discovered. Later, other scientists determined that microbial attachment to a surface enhances growth and that bacteria tend to congregate on surfaces instead of freely moving in the surrounding environment. Finally, the developments in electron microscopy have enabled scientists to ascertain the composition of biofilms.

A biofilm is an aggregation of microorganisms irreversibly attached to a solid surface and enclosed by a matrix of extracellular polymeric substance (EPS). Biofilms can consist of many different types of microorganisms, such as bacteria, diatoms, fungi, algae, protozoa, and noncellular materials. Biofilms are located on solid materials in an aqueous medium and acquire organic and inorganic material floating in surrounding water. Organic compounds, such as nitrogen and phosphorous and reduced inorganic compounds provide energy for the metabolism of the biofilm [Wesley and Satheesh, 2009].

It is believed that the development of a biofilm community on a submerged surface occurs through a sequence of specific, but poorly understood processes [Cooksey and Wigglesworth-Cooksey, 1995] (see Figure 2.4). It begins with the formation of a conditioning film (organic matter) on the substratum, which facilitates the attachment of bacteria to the surface via electrochemical interactions e.g., Van der Waal. It is thought that surface colonisation by bacteria proceeds through an ordered series of recruitment processes; first, pioneer species of bacteria (primary colonisers) interact with the conditioning film and form the initial assemblage of surface biota and biopolymers [Marshall, 1992]. These primary organisms also modify

the surface characteristics of the substratum, rendering it suitable / unsuitable for subsequent colonisation by secondary microorganisms. Specific and/or non-specific interactions (e.g., quorum sensing) between the primary colonists and subsequent recruits permits new organisms to efficiently colonise, these organisms include bacteria, insect larvae and invertebrates [Wolfaardt et al., 1994]. Finally, through synergistic and competitive interactions, as well as the loss and recruitment of new species [Dang and Lovell, 2000], the mature biofilm community is formed.

The structure of biofilms varies enormously, due to the environmental conditions they inhabit [Stolz, 2000]. However all biofilms share certain structural characteristics; they are composed of microcolonies of bacterial cells embedded in a matrix of EPS; hydrodynamic channels separate the microcolonies and provide a means of communication between the bacterial cells and permit the diffusion of nutrients, oxygen, waste material and horizontal gene transfer [Laskin et al., 2004].

The biofilm matrix encloses the bacteria and determines the architecture and shape of the biofilm. EPS is the major component of the biofilm matrix and comprises on average 85% of the total organic carbon of the biofilm. Although the physical and chemical properties of the EPS of different biofilms may vary, the principal component of all EPS is polysaccharides. The polysaccharides of the EPS acquire great quantities of water through hydrogen bonding resulting in a highly hydrated matrix composed of 97% water [Romeo, 2008]. EPS production is promoted by inhibited bacterial growth and an excess of carbon and an inadequacy of other nutrients, such as nitrogen [Laskin et al., 2004].

As previously mentioned the composition of the exopolysaccharides varies depending upon the bacteria comprising the biofilm community, for example the EPS matrix of Gram negative bacteria are polyanionic (attracted to cations) whereas Gram positive bacteria produce polycationic EPS matrices. Irrespective of the composition, the matrix components cross-link the polymer strands and strengthen the biofilm and help to create a three dimensional shape which is extremely stable and resistant to toxins, antimicrobials and predators [Romeo, 2008, Wesley and Satheesh, 2009].

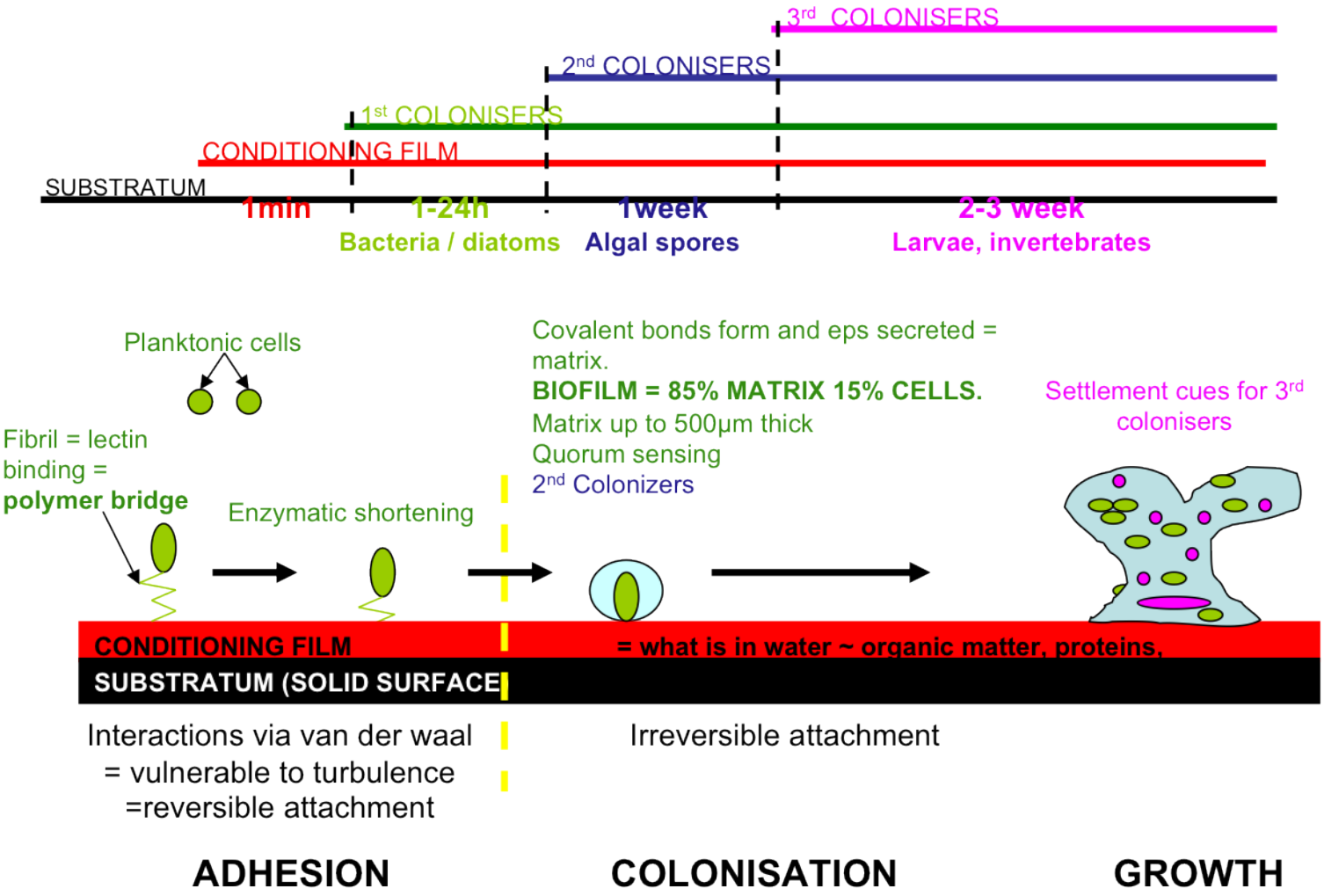


Figure 2.4: Graphical representation of biofilm development

2.4 Operating Slow Sand Filters

From an engineer's perspective, the primary consideration when operating a slow sand filter is the quantity and efficiency of water produced per unit area per day [Ellis and Wood, 1985]. This depends upon a number of factors including the quality of the raw water, the environmental conditions, the microbial community dynamics (both at the surface and within the sand bed), and also the design, construction and operation of the filter [Lloyd, 1974].

Burman [1978] suggested that there are eleven principles for good SSF operation, these include the removal of excess turbidity using effective pre-treatment, steady-state operation i.e., not leaving the beds idle when full of water, cleaning the filters as quickly as possible and resanding only during the coldest time of the year. Additionally, as suggested by Huisman et al. [1974] and implemented in the Netherlands and Japan, SSFs can be covered from the elements to prevent:

1. the deterioration in water quality during periods of low temperature (below 6°C for several months);
2. the expense and operational difficulties of ice removal during periods of cold weather;
3. sunlight exposure which has been shown to promote algae growth (particularly in warm countries) which can reduce water quality;
4. the deterioration in water quality through wind-borne contamination and wildlife droppings.

It is important to point out that as SSFs are biological in composition, if they are subjected to continuous exposure of suspended solids this will eventually lead to filter clogging after several months. The deposits of inert particles from the suspended solids, together with the growth of microorganisms, create increased hydraulic resistance to flow, resulting in headloss (when the maximum level of water above the sand and the outlet valve can no longer achieve the designated flow rate). Once headloss has been reached the filter must be drained and cleaned by scraping (removing) the top 2-3 cm of the sand bed (schmutzdecke).

2.4.1 Maturation

When putting a new filter into operation the filtration rate must be started at a quarter of the required final filtration rate and gradually increased over several weeks. This allows the schmutzdecke and microbial community on the sand grains [Huisman et al., 1974]) to develop, a process which can take more than 40 days [Ellis and Wood, 1985, Duncan, 1988] but can be speeded up by increasing the temperature [Huisman et al., 1974]. It has been widely acknowledged that the development of the microbial community (bacteria, viruses and eukaryotes) is integral to the elimination of pathogens [McConnell et al., 1984, Fogel et al., 1993, Weber-Shirk and Dick, 1997a, 1999, Hijnen et al., 2007, Bauer et al., 2011] the breakdown of organic matter [Eighmy et al., 1992] and the oxidation of ammonia to nitrate [Aslan, 2008]. It is also important to stress that throughout the maturation process and at all stages of the water distribution process water quality tests must be performed. In Europe such tests must comply with the guidelines outlined by the EU Drinking Water Directive (98/83/EC) [European Union Council Directive, 1998], which as previously mentioned in section 1.1.1 contains chemical, physical and microbiological tests to ensure excellent water quality.

2.4.2 Cleaning and Re-sanding

Depending upon the turbidity and suspended solid content in the water to be purified the length of time a slow sand filter can be operated for before water quality diminishes varies greatly (60 days to more than fifteen years). In order to maintain good filter performance and to reduce the period of time a filter is non-operational, SSFs need to be cleaned regularly to prevent the filter bed from becoming clogged. This usually involves removing the schmutzdecke and the top 2-3 cm of sand. These layers have been reported to contain the highest levels of bacterial colonisation in the sand filter matrix [Haarhoff and Cleasby, 1991, Eighmy et al., 1992] and the greatest biological activity [Calvo-Bado et al., 2003]. In order to determine when a filter needs to be cleaned, the headloss of the filter is constantly measured piezometrically. Briefly, referring to Figure 2.5 when the sum of the height of the filtered water (H2) subtracted from the height of the column of water supported by the raw water (H1) exceeds 1m the filters are cleaned.

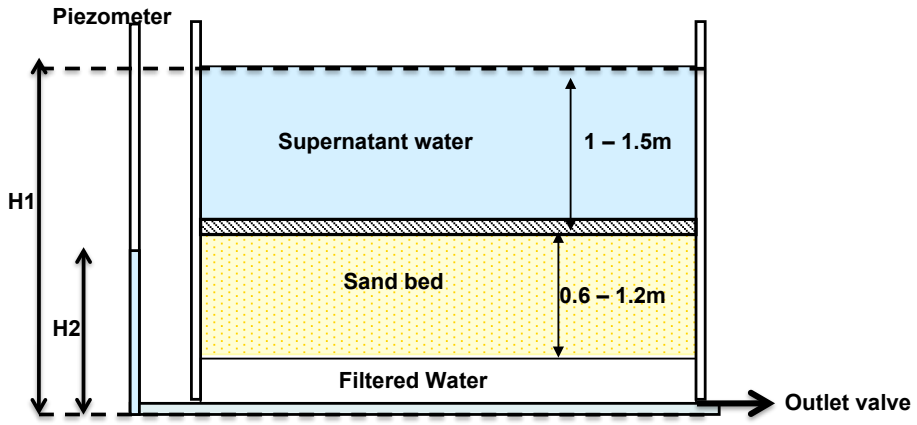


Figure 2.5: Schematic filter showing how headloss is measured, adapted from Lloyd [1974]

In order to clean a slow sand filter, the water in the filter is first drained to 20 cm below the sand surface, a process called "wet harrowing" (Figure 2.6 A). After being drained, the top 2 - 3 cm of the schmutzdecke is removed (scraped) either manually or by using a tractor with a special device attached. The scraped sand is then cleaned through a hydraulic ejector which forces it through a hose to a washer (Figure 2.6 B) the washed sand is then left to dry in the sunlight and is ready for reuse (Figure 2.6 C and D). After the filter bed is scraped, it is refilled until it is completely covered and run at a third of the normal filtration rate for at least 4 days to allow the microbial community to reform. Before being put back into service the filter must pass water quality checks. Resanding of the filters occurs after approximately 16 scraping sessions, once the level of the sand bed has reached 60 cm in height [Burman and Lewin, 1961].

2.5 Advantages and Disadvantages of Slow Sand Filters

Today in developed countries SSFs tend to be overlooked when designing new drinking water treatment plants mainly due to their disadvantages, which include:

1. Their large land footprint, which increases the initial startup costs. For example the average surface area required for one industrial SSF in Europe is 1620m^2 compared to 196m^2 for rapid sand filters [Huisman et al., 1974]. Further, based on current average land prices in the UK (one acre for £4000 [UK Land Directory, 2014]), one SSF would cost approximately £1600, however, for the same cost and space eight rapid sand filters could be installed.

Figure 2.6: Photographs showing (a) drained filter, (b) hydraulic ejector sand cleaning machine, (c) cleaned sand dune and (d) cleaned sand.

2. Poor removal of colour.

3. Poor operational performance if the water to be purified has a high algal and or turbidity content as this increases the rate of clogging and decreases the concentration of dissolved oxygen available for respiration and hence the development of aerobic microbial community members.

4. Increased expense in countries where the winter is very cold as special structures may need to be installed to prevent filter freezing.

However, these disadvantages are greatly outweighed by the advantages SSFs offer such as:

1. Simple design and construction.

2. Ease of operation which requires limited supervision.

3. Low operational and maintenance costs, mainly due to the renewable nature of the filter-bed material and limited mechanical equipment. For example, Visscher et al. [1987] calculated that the construction of one rapid sand filter would cost $70,000, whereas a SSF of the same capacity costs less than half this value.

4. Limited power requirements as SSFs are typically gravity fed.

5. Excellent overall water quality, including pathogen removal.
6. Small amount of water wastage (2–3% of the total treated).
7. Usually not necessary to chemically pre-treat the raw water.
8. Sand required for the filter bed is readily available and reusable making SSFs a very environmentally friendly technology.
9. Filtered water is less corrosive and more uniform in quality than that found from chemically treated water.
10. Complete ammonia removal.

These advantages in particular their carbon neutral footprint and the lack of chemical pre and post treatment makes them a very attractive technology for the future [Johnson et al., 2009].

2.6 Previous Slow Sand Filter Studies

Extensive research exists describing the biomass development which occurs in the sand-bed and in the schmutzdecke, with Huisman et al. [1974] and Ellis and Wood [1985] showing that the level of bacterial activity decreases with depth, but normally continues to a depth of 700 mm. Although significant details are known about biomass development under different conditions and in different positions in SSFs [Eighmy et al., 1992, Campos et al., 2002], relatively little is known about the microbial species and their ecological interactions which are responsible for purification. This lack of knowledge inhibits further optimisation and functional advancement of SSFs.

There have been a few studies which have begun to address these questions. However, these have been performed in controlled laboratory conditions [Burman, 1978, Weber-Shirk and Dick, 1997a,b, Calvo-Bado et al., 2003, Bourne et al., 2006, Aslan, 2008] which have not been shown to represent real industrially operated full-scale SSFs. Further they have used conventional culture-dependent or carbon utilisation methods [Eighmy et al., 1992, Weber-Shirk and Dick, 1999] which do not allow the study of the non-culturable and fastidious species generally thought to dominate environmental samples [Roszak and Colwell, 1987].

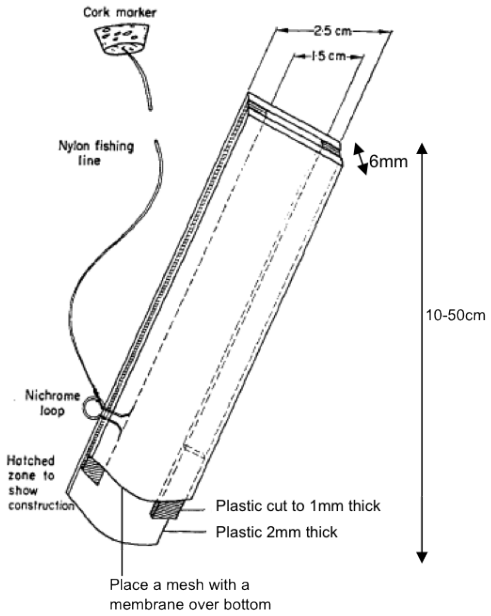


Figure 2.7: Schematic of the Sand profile sampler created by Lloyd [1973]

From earlier work it is known that the biological (grazing) activity of protozoa is responsible for the high reduction of coliforms [Brook, 1955, Lloyd, 1973, Weber-Shirk and Dick, 1999] and other pathogenic bacteria; however the specific species responsible remain unknown. This breakthrough in understanding was made in 1973, when Lloyd created a simple and inexpensive method for sampling and microscopically visualising the protozoa and Rotifera in the sand of pilot-scale and full-scale SSFs at the London Metropolitan Water Board's Walton and Ashford Common Treatment Works (Figure 2.7). Following Lloyd's discovery, a pioneering study by Weber-Shirk and Dick [1997a] involving the construction of a laboratory-scale SSF revealed that biological activity was responsible for the marked reduction in *E.coli* and particles smaller than 2 μm. However, within this study it was also concluded that protozoan grazing of the bacteria was the only removal mechanism, a conclusion which cannot be substantiated as the authors failed to try other methods. Since the creation of the first laboratory-scale SSFs by Weber-Shirk and Dick, several other laboratory-scale filters have been created, with the most notable being that of Calvo-Bado et al. [2003] which possessed sample ports to allow access to several depths down the sand bed. This provided for the first time some insight into the spatial variability and diversity in the microbial community of SSFs.

Slow sand filters have also been shown to be extremely effective in virus removal [Windle-Taylor, 1969, Poynter and Slade, 1978, McConnell et al., 1984, Wotton, 2002, Elliott et al.,

2008, Bauer et al., 2011]. Windle-Taylor first provided strong evidence that the microbial community in SSFs is integral to viral removal in 1969 when he reported that viruses were not removed by sterile sand. However, when sand from a mature filter was used to filter water containing 100 plaque-forming units (PFU) per millilitre of attenuated poliovirus, 99.9% of the virus was removed. He further highlighted the importance of the depth of the sand bed, showing that a sand bed of 600 mm removed 75% more viral particles than a filter of 300 mm in depth.

More recently, some studies have tried to characterise the whole microbial community and determine the functional boundaries of slow sand filters [Wakelin et al., 2011, Hunter et al., 2012, Bai et al., 2013, Ramond et al., 2013] using modern molecular techniques, instead of focussing on specific aspects such as viral or coliform removal. In spite of all of the work which has been done, the complex foodwebs and metabolic processes which make up the SSF community have not been studied in detail. Although Campos et al. [2006] produced models describing schmutzdecke development and Devadhanam and Pillay [2008] visualised the microbial biomass growth within the sand using environmental scanning electron microscopy (ESEM), very little is still known about the ecosystem of the SSFs. Additionally, the interaction between the schmutzdecke layer and the underlying biofilm community in the sand adds another level of complexity to the interaction and hence the processes of water purification. Understanding such interactions and mechanisms would enable the development of ecological mechanistic models of SSF systems, which may improve the operation and efficiency of slow sand filters [Campos et al., 2002].

To date, slow sand filter studies have focussed on characterising and validating the biologically mediated purification mechanisms, using carefully controlled laboratory conditions and conventional plating and isolation techniques, thus ignoring the uncultivable majority. Today, molecular microbiological techniques are available, which, when used alongside conventional microbiological tools, allow scientists to understand the ecology of SSF systems. This will, in turn, promote optimisation of SSF design and operation, creating more efficient filters with the ability to tailor water quality performance to the specific needs of a site.

Chapter 3

Microbial Community Analysis Reviewed

"The true method of knowledge is experiment."

William Blake, (All Religions are One)

Traditionally, the identification and characterisation of microbial communities has been limited to those microorganisms that are culturable, therefore these studies have overlooked the non-culturable, fastidious and less adaptable species generally thought to dominate environmental samples [Roszak and Colwell, 1987]. Molecular techniques overcome many of the disadvantages associated with traditional culture-based techniques and provide an exciting opportunity to greatly increase our understanding of microbial diversity and functionality in various environments. These methods rely on the identification of cellular components such as nucleic acids, proteins, fatty acids and other taxa-specific compounds [Amann et al., 1995] and unlike culture dependent methods can be extracted directly from the environment of interest and hence *in situ* metabolic function and the microbial community composition can be preserved and analysed easily [Malik et al., 2008].

Application of these molecular techniques have led to more rapid and accurate strategies to study microbial diversity, including the discovery and identification of novel organisms and their ecophysiology. Individual bacterial identification and community diversity characterisation has been enhanced by using the highly conserved 16S rRNA gene which is ubiquitous in all prokaryotes [Pace, 1997]. The 16S rRNA gene sequence is conserved enough to enable the design of PCR primers which target different taxonomic groups from kingdom to genus, but have enough variability to provide phylogenetic comparisons of microbial communities [Woese, 1987]. There are however a plethora of molecular techniques in the scientific arse-

nal used to assess microbial community composition. In order to discuss them, they will be divided into six categories:

1. Biochemical Methods
2. Nucleic Acids Based Methods
3. Techniques Linking Microbial Identity To Function
4. Next Generation Sequencing
5. Metagenomics
6. Other "Omic" Methods

It is however important to stress that in order to successfully identify and quantify microbial communities, a combination of both molecular and traditional culture-based techniques are often required.

3.1 Biochemical Methods

Phospholipid Fatty Acids (PLFAs)

Phospholipids are integral components of living organism's membranes and constitute a significant proportion of an organism's biomass under natural conditions [Kozdroj and van Elsas, 2001]. Unlike higher organisms, microbes have the ability to change their membrane lipid composition in response to environmental conditions such as chemical stress and temperature fluctuations [Malik et al., 2008]. Further, PLFA rapidly degrade upon cell death; such regulated fluctuation in PLFA composition makes it an ideal indicator of organism's status and viability [Drenovsky et al., 2004].

Additionally, changes in phospholipid content are generally related to changes in the abundance of different microbial phyla. Although a useful tool PLFA as a microbial community profiling method produces profiles of limited complexity and has a bias towards species with a larger PLFA content. Today, PLFA is rarely used alone; instead it is deployed alongside other molecular profiling methods to assess microbial diversity [Malik et al., 2008].

3.2 Nucleic Acid Based Methods

In 1983 Kary Mullis developed the polymerase chain reaction (PCR) [Bartlett and Stirling, 2003], a process which has the ability to produce millions of copies of a desired gene with high fidelity within 3 to 4 hours. However like all molecular methods, it has its biases and is reliant on the quality and yield of the initial DNA extraction. Insufficient lysis of cells can result in the preferential extraction of DNA from Gram-negative bacteria, while excessively harsh treatments may result in the shearing of DNA [Wintzingerode et al., 1997]. In addition, PCR amplification efficiency can be severely reduced by the presence of inhibitory substances which are co-extracted with nucleic acids (e.g., humic acids [Kirk et al., 2004]). Due to these issues it is vital that methods used for sample collection, transportation and storage are effective in preventing the addition of non-native organisms into the microbial community of the sample [Schneegurt et al., 2003].

PCR-based molecular techniques have completely revolutionised the detection of DNA/RNA, especially in microbial ecology studies. However, differential amplification and primer choice of target genes such as 16S rRNA can bias PCR-based diversity studies [Wintzingerode et al., 1997]. For example, sequences with lower guanine and cytosine content are thought to separate more efficiently in the denaturing step of PCR and hence could be preferentially amplified. Also, products seen on gels or in real-time may be as a result of artefacts or chimeric PCR product formation [Wang and Wang, 1996]. Therefore, care must be taken to minimise chimera formation, something which can be achieved through the use of high fidelity polymerases [Oyola et al., 2012] and low numbers of PCR amplification cycles.

Microbial community composition can be analysed based on profiles generated from the physical separation of RNA or DNA sequences on a gel [Muyzer et al., 1993]. In this regard, several techniques based on the amplification and comparison of PCR-amplified DNA sequences have been developed and used to characterise microbial communities. These methods detect differences between DNA/RNA sequences of the target gene, with the 16S rRNA gene being the most extensively used in bacterial studies, whereas in eukaryotic studies the lack of a universal primer site has resulted in several genes being targeted (18S rRNA, and ITS) depending on the subkingdom of interest.

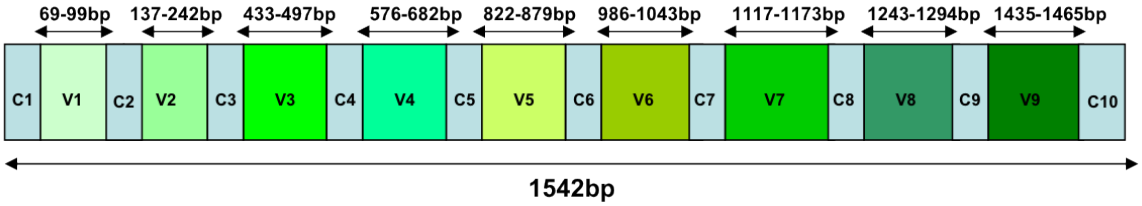


Figure 3.1: Schematic of 16S rRNA gene illustrating hypervariable (green) and conserved (blue) regions

The 16S rRNA gene is composed of nine hypervariable regions (V1-V9) (Figure 3.1) interspersed by conserved regions. By amplifying selected targets (combination of variable and conserved areas) within the 16S rRNA gene, bacterial and archaeal genera and species identification can be achieved. It is however important to note that depending upon the hypervariable regions chosen to amplify, different efficacies with respect to species identification are achieved [Schmalenberger et al., 2001, Luna et al., 2007]. The different genetic community profiling methods include amplified ribosomal DNA restriction analysis (ARDRA), denaturing gradient gel electrophoresis (DGGE)/temperature gradient gel electrophoresis (TGGE) and terminal-restriction length polymorphism (T-RFLP).

3.2.1 Reverse-transcription PCR (RT-PCR)

Reverse transcription polymerase chain reaction (RT-PCR) is a variant of PCR, which is used to qualitatively detect gene expression through the creation of complementary DNA (cDNA) transcripts from RNA [Freeman et al., 1999]. RT-PCR has been widely used in environmental studies to measure functional gene expression levels [Smith et al., 2007, Hatzenpichler et al., 2008] and to detect viable viruses in water filters [Langlet et al., 2009]. RT-PCR has many advantages over conventional methods for measuring RNA including its sensitivity in detecting low expression levels of mRNA, accurate quantification and the potential for high throughput. However, it also has a number of problems associated with its use, including the inherent variability and sensitivity of RNA, difficult extraction procedures and variability in PCR efficiencies [Huggett et al., 2005]. Therefore, it is important that a reliable normalisation method is implemented to control for these problems.

3.2.2 Quantitative Polymerase chain reaction (qPCR)

A variant of the PCR technique which can simultaneously detect and quantify the amplified product whilst the reaction is occurring is real-time PCR or qPCR. This approach enables the detection and quantification of PCR amplicons during the early exponential phase of the reaction [Invitrogen Corporation, 2008]. Real-time PCR involves the use of fluorescent markers (Sybr-Green or TaqMan) to quantify the product at the end of each amplification cycle; the intensity of fluorescence is directly related to the amount of product at the end of each cycle in the PCR reaction [Saleh-Lakha et al., 2005]. qPCR is one of the most widely used molecular tools for determining relative and absolute numbers of different members of a microbial community [O'Farrell and Janssen, 1999, Silva et al., 2006, Philippot et al., 2009, De Gregoris et al., 2011]. The advantages that real-time PCR offers include speed, sensitivity, accuracy and the possibility of robotic automation [Smith and Osborn, 2009]. Although real-time PCR can measure gene abundance, the results obtained do not link gene expression with a specific measurable microbial activity or population.

3.2.3 Amplified ribosomal DNA restriction analysis (ARDRA)

Amplified ribosomal DNA restriction analysis (ARDRA) is a technique in which PCR-amplified 16S rRNA fragments are digested at specific sites with restriction enzymes and the resulting digest separated by gel electrophoresis. Due to the differing DNA content of microbes, ARDRA results in unique banding patterns being generated for each organism. These patterns can be compared and analysed to determine community composition. It should be stressed that this technique allows analysis of similarities or differences between samples, however the identity of individual banding patterns is unknown unless extracted and sequenced. Overall, ARDRA is a simple, rapid, and cost-effective method which has been widely used to study microbial communities from various environments [Smit et al., 1997, Gich et al., 2010].

3.2.4 Clone library

The creation of a clone library for a particular gene, typically 16S rRNA is one of the most widely used tools for initial exploratory analysis of the microbial community in environmental samples. Clone libraries are created by amplifying extracted DNA with primers specific

to the gene to be used to survey diversity (i.e., 16S or 18S rRNA). Amplified products are then cloned / inserted into typically an *E.coli* vector (using commercially available kits) and screened using ARDRA. Representatives of different banding patterns are typically Sanger sequenced and characterised by comparing the sequence to databases such as Ribosomal Database Project (RDP), and Greengenes. Typically, cloned sequences are assigned to phylum, class, order, family, subfamily, or species (DeSantis et al. 2007). While clone libraries of 16S rRNA genes permit an initial survey of diversity and identify novel taxa, studies have shown that environmental samples may require >40,000 clones to document 50% of the richness (Dunbar et al. 2002). Due to the need for such a large screen to pick up on a samples diversity, next generation sequencing has today became the methodology of choice, owing to the high sample throughput and increased sequencing resolution. Despite the limitations of clone libraries (e.g., labour-intensive, time-consuming, and cost factors), they are still considered the "gold standard" for preliminary microbial diversity surveys (DeSantis et al. 2007). Furthermore unlike other methods they provide template of representative groups for use as standards and positive controls in other methods such as qPCR.

3.2.5 Terminal-restriction fragment length polymorphism (T-RFLP)

Terminal-restriction fragment length polymorphism (T-RFLP) is a modification of ARDRA; however, the primers used in T-RFLP are fluorescently labelled at their 5' end so that the resulting PCR products are also labelled [Slater et al., 2010]. The products are then digested using a restriction enzyme (usually 2 - 4 different enzymes are used) and the mixture of fragments analysed by capillary electrophoresis. The resulting electropherogram consists of only the fluorescently labelled terminal fragments each of a particular length and height, which represent a single operational taxonomic unit (OTU) [Tiedje et al., 1999]. This simplified banding pattern enables complex microbial communities to be analysed both in terms of diversity and relative abundance and has been used extensively in environmental studies of contaminated land and water bodies [Jeon et al., 2003, Winderl et al., 2008, Ramond et al., 2013]. Like ARDRA and the other molecular techniques described in this section, T-RFLP shares the same drawbacks (problems associated with PCR bias, chimera formations, incomplete digestion, noise associated with small fragments and the subjective nature of interpretation). However, it is an extremely valuable tool which provides great insight into the community dynamics of a sample in an economical and time efficient way.

3.2.6 Denaturing / Temperature gradient gel electrophoresis

Denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) separate amplified rRNA fragments of the same length but with different base pair compositions. The separation of bands is dependent on both the decreased electrophoretic mobility of partially melted double stranded DNA molecules in polyacrylamide gels containing a linear gradient of DNA denaturants (DGGE) and a linear temperature gradient (TGGE) [Muyzer and Smalla, 1998]. The amplified PCR products are separated on the basis of sequence differences, not variation in length, with the number of bands produced being proportional to the number of dominant species in the sample. Unlike ARDRA and T-RFLP, DGGE/TGGE are used when information on the phylogenetic composition and abundance of the microbial community is either unknown or only required for the dominant members, specifically when determining the effect that contaminants and pathogens have on a community [Boon et al., 2000, Calvo-Bado et al., 2003, Pereira et al., 2010, Hunter et al., 2012].

The main advantage of DGGE/TGGE is that it allows spatial and temporal changes in the microbial community to be monitored. The limitations include those mentioned for ARDRA and that the 16S fragment that can be analysed is limited to 500 bp, which may lack the specificity required for the phylogenetic identification of some organisms. Additionally, strong band intensity may not necessarily mean a more abundant member of the microbial population, meaning that diversity may be overestimated [Muyzer and Smalla, 1998]. Finally, unlike T-RFLP, DGGE and TGGE are extremely labour-intensive techniques, which provide limited information compared with faster methods.

3.3 Techniques Linking Identify to Function

Today there are a plethora of techniques available which allow the identification of microorganisms. However, the real challenge for modern-day microbial ecologists is being able to characterise them and their individual functions whilst in their natural environment [Pernthaler, 2010]. The aforementioned biochemical and nucleic acid based methods are very effective at determining the identity and structure of microbial communities and certain processes can also be correlated with these measures (e.g., nitrogen fixation, and respiration). However, these methods do not provide direct information on the ecophysiology of individ-

ual members of a community. This gap in knowledge drove the development of "single-cell technologies" which allow both identity and function of individual members to be determined in a culture-independent manner [Sen and Ashbolt, 2011].

3.3.1 Microarray and Phylochips

A microarray is a high-throughput screening tool that is used to study gene expression or in the case of phylochip used as an identification method. They work by hybridising the mRNA or DNA (in the case of phylochip) of a sample to oligonucleotides or probes of known genes or organisms (up to 240,000 probe arrays are available) which are attached to a chip. The presence of a gene is detected by fluorescence, emitted when the sample (which has been fluorescently labelled) binds to a probe on the chip. Lasers then scan the chip and capture images of which locations are fluorescing, and relate this back to identity of the probe(s). Conventionally, both a control and an experimental sample are added to a chip; one is labelled with Cy3 (red fluorophore) and the other Cy5 (green fluorophore), allowing conclusions to be drawn about the effect of the experimental procedure on gene expression.

Due to their high-throughput, microarrays have been extensively used in various environmental studies ranging from understanding methane cycling, total microbial diversity studies to pathogen detection [Roh et al., 2010, Wakelin et al., 2011]. However, they are being used less frequently today, due to several limitations [Hurd and Nelson, 2009] including:

- the prior knowledge about the organisms or genes required to make probes;
- cross hybridisation between similar sequences, limiting analysis to areas of a genome which are non-repetitive;
- reproducibility of results is questionable due to issues with sample preparation and analysis methods;
- the cost to perform compared to next-generation sequencing (in terms of per base costs) especially if custom chips need to be designed.

3.3.2 Fluorescence *in situ* hybridisation (FISH)

Fluorescence *in situ* hybridisation (FISH) is a method used to visualise (via epiflourescence microscopy) and quantify the presence and relative abundance of microbial populations [DeLong et al., 1989]. Samples are fixed on a glass slide and hybridised using phylogenetically fluorescent probes, either group or domain specific. FISH is a method predominantly used to examine whether members of a specific phylogenetic group are absent or present and allows nonculturable organisms to be visualised and quantified [Lara-Martín et al., 2007]. On its own, FISH provides limited insight into the ecophysiology of the microorganism in question, and it is therefore typically coupled with other techniques such as microautoradiography (MAR-FISH) [Gray et al., 2000, Thayanukul et al., 2010], and Raman microspectroscopy (Raman-FISH) [Huang et al., 2007] and qPCR [Schippers et al., 2005]. Further, FISH is used in numerous microbial community dynamic studies in conjunction with other genetic fingerprinting methods such as DGGE [Straub et al., 2001, Collins et al., 2006] and T-RFLP [Collins et al., 2006, Slater et al., 2010]

FISH is an extremely useful tool when trying to identify community dynamics particularly of non-culturable organisms and link identity to function. However, it has several drawbacks, namely that a limited number of probes can be used in a single hybridisation experiment hence prior knowledge of which organisms are in a sample is often needed [Lara-Martín et al., 2007], background fluorescence is difficult to correct for, and the fact that the design and optimisation of hybridisation conditions for new probes is extremely time consuming and complex [Pernthaler, 2010].

3.3.3 Stable-Isotope Probing (SIP)

Stable-isotope probing (SIP) is a relatively new method which is under constant development and is increasingly being used to link identity to function. SIP works by using stable isotopically-labelled substrates (e.g., ^{13}C and ^{15}N) to follow the fate of the substance as it is metabolised and incorporated into the biomass of the microbial community [Dumont and Murrell, 2005]. Once metabolised the labelled marker (typically DNA or RNA) can be separated from the unlabelled biomass using CsCl and density-gradient centrifugation [Neufeld et al., 2007]. Labelled (heavy) fractions will be found at the bottom of the cen-

trifuge tubes and unlabelled (light) fractions at the top. Fractions can then be excised either using a needle or a fractionator and used with traditional community analysis methods such as DGGE [Moreno et al., 2010], T-RFLP [Sul et al., 2009] and qPCR or with more modern methods such as metagenomics [Sul et al., 2009], proteomics [Jehmlich et al., 2008], FISH, NanoSIMS or transcriptomics.

SIP is a powerful technique and has provided invaluable insights into the ecophysiology of microorganisms and promises to deliver much more when used in combination with metagenomics and proteomics. However, as mentioned before this technique is in its infancy and still has a lot of issues which must be overcome [Dumont and Murrell, 2005], such as:

- possible biases caused by the incubation with the isotope and the cycling of the stable isotope within the microbial community;
- cross feeding with naturally forming variants of the labelled substrate which may dilute labelled incorporation into the organism;
- formation of multiple heavy fractions in a mixed community, which can be overcome by using a fractionator;
- uncertainty of how long it takes for the labelled substrate to be incorporated into the biomass of the community;
- the complexity of the stages involved in SIP.

3.3.4 NanoSIMS

Nanoscale secondary ion mass spectrometry (NanoSIMS) is a technique that determines the absolute isotopic composition of a sample in combination with high-resolution microscopy. The process is destructive and involves bombarding the subject with Cs^+ or O^- ion beams, resulting in the production of secondary ions. These secondary ions can be measured through mass spectrometry and give details about the isotopic ratios within the sample. Currently, a resolution of 50nm can be achieved, allowing very detailed spatial maps to be created. NanoSIMS has recently been used by Fike et al. [2008] to explain how phosphorus cycling occurs in cyanobacterial mats. Furthermore, NanoSIMS can be combined with other techniques such as SIP [Ploug et al., 2011] which allows accurate substrate rates and turnovers

to be calculated. Although NanoSIMS is a highly sensitive and accurate tool for functional identification, the currently high cost of equipment limits further use of the technique.

3.4 Next Generation Sequencing

The development of DNA sequencing technology has a rich and diverse history [Shendure and Ji, 2008], with its evolution being driven by clinical and research fraternities requiring ever more efficient, cost effective and robust methods of genomic sequencing. In the past the overwhelming majority of DNA sequencing relied on some version of the Sanger biochemistry3 methodology which allowed the full human genome to be sequenced. However, within the last nine years at least five entirely new sequencing methods, coined "next-generation sequencing" (NGS) platforms, have emerged. Roche's 454 pyrosequencing was formerly the most extensively used platform, however due to better read lengths and cost Illumina's MiSeq and HiSeq platforms have today became the platforms of choice. In addition to these platforms, Life Technologies' Ion Torrent and Helicos Biosciences' HeliScope also exist.

NGS platforms have made it possible to recover and characterise genomic material directly from environmental samples. It is this very reason which explains the vast number and range of microbial interactions, ecophysiologies and biomes that are understood today [Tyson et al., 2004, Amaral-Zettler et al., 2009, Petrosino et al., 2009, Bartram et al., 2011, Koskinen et al., 2011, Dunthorn et al., 2012]. It is important to stress that different NGS platforms have various advantages and disadvantages in terms of their read length, quantity of data produced, run time and cost (Table 3.1) and all of these must be considered when choosing a platform. However, this continual development and evolution is not without problems, namely issues with computational power, storage space and analytical methods not keeping pace. Furthermore, the data produced by these platforms is affected by the same issues mentioned in Section 3.2 (e.g. primer choice and PCR cycle number). Regardless of these problems, next generation-sequencing methods are cheaper, faster and produce copious amounts of genomic data. For this reason NGS has and will continue to revolutionise and accelerate biological and biomedical research; in particular, allowing scientists to begin to understand the complexity of bacterial metabolism, virulence mechanics, genetic exchange, phylogeny, roles and interactions within specific niches.

Table 3.1: Comparison of currently available next-generation sequencing platforms, adapted from Schirmer et al. [2012] and Shokralla et al. [2012].

Platform	Read length (bp)	Sequencing output (per run)	Run time	Cost per Mb ($)
Roche 454 FLX Titanium	400-500	≤500 Mb	10 h	12.40
Roche 454 FLX +	600-800	≤700 Mb	23 h	7.00
Illumina Hiseq 2000	50-100	540-600 Mb	11 d	0.10
Illumina GAIIx	50-150	≤ 95 Gb	14 d	0.12
Illumina Miseq	100-250	1-2 Gb	19-27 h	0.74
Ion Torrent 318 Chip	100-200	1 Gb	5.5 h	0.75
HeliScope	30-55	20-28 Gb	1 d	2.40

3.4.1 454 (GS-FLX Pyrosequencing)

Released in 2005, Roche's 454 sequencing platform was the first available next-generation sequencing system and is probably the most well known of all the new methodologies commercially available. 454, the same as the other methods begins by the creation of a library of amplified template-DNA fragments by emulsion PCR. Then using the 454 methodology the DNA-template fragments undergo *in vitro* ligation to adaptor molecules (aid in amplification and sequencing), which in turn are immobilised onto DNA capture beads. The DNA on the beads are then subjected to emulsion PCR in order to amplify the nucleic acid. Amplified template DNA are then isolated and inserted into wells of a picotitre plate. The sequencing process that then follows consists of alternating cycles of adding one species of dNTP and DNA polymerase in the presence of luciferase; thus, the incorporation of a complementary nucleotide results in the release of pyrophosphate, which is used to make ATP, which drives the chemoluminescence reaction of luciferase. The resulting light that is produced is directly proportional to the amount of ATP available and software is used to interpret which dNTP has been incorporated onto the complementary strand.

Of all next-generation platforms 454 sequencing provides the longest sequence reads (500 bases, with a single run potentially generating 500Mb) making it well suited to *de novo* genome assemblies. However, it has several disadvantages; such as inaccuracies in calling homopolymeric stretches of sequence (i.e., AAAA, CCCCC) [Hurd and Nelson, 2009] and perhaps one of the biggest problems, the generation of chimeras (sequence hybrids usually consisting of two phylogentically distinct parent sequences). Chimeras are well documented in community profiling with 16S rRNA and are believed to account for at least $\frac{1}{20}$ of the

16S sequences held in public repositories [Ashelford et al., 2005], hence their removal is of paramount importance in order to provide a true reflection of the composition of a microbial community. There are several algorithms available which check the amplicon for the presence of chimeras and remove them (Pintail [Ashelford et al., 2005], Greengenes [DeSantis et al., 2006] and Perseus [Quince et al., 2009]). Finally, unlike the other NGS platforms, 454 is very expensive in terms of megabase sequencing (Table 3.1) output and for that reason is being used less often [Schirmer et al., 2012].

3.4.2 Illumina - MiSeq and HiSeq

In 2006 Illumina (formally Solexa) introduced the Genome Analyser which, unlike 454, works in a similar way to Sanger sequencing, incorporating nucleotides in separate sequencing cycles; hence accounting for its major errors "base substitutions" as compared to 454's "indels" (insertions and deletions) [Suzuki et al., 2011]. Briefly, Illumina works by creating templates of the DNA sample using bridge amplification on individual fragments of the sample, which are attached onto the solid surface of a flow cell. Attachment to the flow cell occurs through covalent bonding between complementary oligos on the flow cell to the adaptors on the fragments. Subsequent hybridisation is achieved by a combination of heating and cooling stages, followed by incubation with the reagents needed for amplification which results in the formation of 100-200 million clusters of amplified DNA fragments [Shokralla et al., 2012]. Clusters are sequenced by supplying them with four fluorescently labelled nucleotides which have a reversible terminator at their 3' end (this terminator ensures that only a single base can be incorporated per cycle). The sequence of each cluster is then computed and quality filtered [Shendure and Ji, 2008] using phred scores.

Today, five versions of Illumina sequencer are commercially available: HiSeq 2500, HiSeq 1000, Genome Analyser, Genome Analyser IIx and MiSeq. According to Eisenstein [2012] Illumina is the biggest and most widely used NGS platform, accounting for 60% of the platforms used. Compared to 454, Illumina sequencing is extremely economical in terms of its per base pair costs [Schirmer et al., 2012], and due to paired-end sequencing is now competitive with the read lengths of 454 (Table 3.1).

3.5 Metagenomics

Metagenomics refers to the culture-independent analysis of the complex and diverse ("meta") populations of prokaryotes, eukaryotes and viruses found in environmental niches, or in animal hosts. Unlike traditional microbial genomic sequencing, metagenomics is based on the concept that the entire genetic composition of environmental communities could be sequenced and analysed in the same way as sequencing a whole genome of a pure bacterial culture, assuming sufficient sequencing depth. This in theory will allow the estimated 99% of prokaryotes which are thus far uncultivable [Amann et al., 1995, Rappé and Giovannoni, 2003] to be identified. More importantly, metagenomics allows scientists to glimpse into fully functional microbial communities, allowing observation over how they interact with each other by exchanging nutrients, metabolites and signalling molecules [Wooley et al., 2010].

Conventionally, bacterial communities in soils, sediment and water are analysed by targeting the 16S rRNA gene by PCR amplification and subsequent analysis via the creation of clone libraries, DGGE or T-RFLP. Most of these approaches provide limited insights into the structure of the bacterial communities, as the survey sizes and the number of compared sampling sites are too small with respect to the enormous bacterial diversity present [Will et al., 2010]. Metagenomics can overcome these shortcomings as it can directly examine the phylogenetic diversity of complex microbial communities [Petrosino et al., 2009]. The metagenomic approach includes both functional and sequence-based analyses of DNA extracted directly from the environment. Extensive studies have been performed on mammalian microbiomes, specifically the mucosal and epidermal surfaces [Turnbaugh et al., 2007, Grice et al., 2008, Zhu et al., 2010] however scientists are only now beginning to explore the natural world [Dinsdale et al., 2008, Petrosino et al., 2009].

Today, metagenomics is one of the fastest-developing research areas in science and has made huge advancements since its development in 1998 [Simon and Daniel, 2011]. Such advancements have been due to several keystone environmental metagenomic projects by Craig Venter and Gene Tyson. The first and probably best known extensive large-scale environmental sequencing project was carried out by the J. Craig Venter Institute in 2004 in which fragments of DNA derived from the entire microbial population of the nutrient-limited Sargasso

Sea were sequenced. Following Venter's approach, Tyson et al. [2004] chose a much simpler community to analyse: acid mine drainage in the Richmond mine, Iron Mountain, California, one of the most extreme environments on earth. In this environment the microbiota was composed of three bacterial and three archaeal species existing as a pink biofilm that forms on the surface of the mine water. Both of these projects have shown that metagenomics is a powerful technique for exploring the ecology of complex and simple microbial communities.

The main analytical challenge for metagenomics pertains to the need to obtain accurate identification of all isolates in the sample in a cost and time efficient manner. This need has led to the development of various bioinformatic tools which allow sequences to be trimmed, vetted for their quality and assembled. However, by far the biggest issue for metagenomics and to a lesser extent NGS is the inability of computational advancements (data storage and processing) to keep pace and until this is addressed any further advancements will be hindered [Röling et al., 2010]. Regardless of these issues metagenomics is an extremely powerful tool for analysing microbial communities. However, its true power will only ultimately be harnessed when it is integrated with classical ecological approaches.

3.6 Other "Omic" Methods

Metagenomics provides information on the metabolic and functional capacity of a microbial community. However, it cannot differentiate between expressed and nonexpressed genes and hence fails to provide information about metabolic activity [Cardenas and Tiedje, 2008]. This deficiency has led to the creation of metatranscriptomics and metaproteomics (the studies of gene expression and translation to proteins, respectively). Today, there has been an explosion in the number of studies using either or both techniques mainly due to methods overcoming the previous associated difficulties with RNA (collecting enough high quality RNA, instability of mRNA and its low abundance in the total RNA of a sample) [Simon and Daniel, 2011]. Metatranscriptomics and metaproteomics are approaches that have the potential to allow us to understand the functional dynamics of microbial communities. When used in combination with metagenomics they promise to significantly impact the measurement and prediction of *in situ* microbial responses, activities, and productivity, which will dramatically improve our understanding of ecosystems and aid in biotechnological advancements.

3.7 Systems Biology for Microbial Ecology

All organisms, whether bacteria, protozoa, plants or humans, interact with the physical and biological world that surrounds them, and understanding these interactions and their functional potential is the role of ecology [Purdy et al., 2010]. Therefore, the major challenge facing microbial ecology is the need to determine how a whole ecosystem operates and therefore be able to:

- Predict responses to perturbations, whether from environmental conditions or changes in operational conditions (in engineered systems);
- Predict performance capabilities of biological engineered systems;
- Manage the community members of ecosystems in order to sustain and improve functionality.

The resilience and stability of an ecosystem ultimately depends on the contributions made by its constituent individual organisms; however, understanding these complex relationships and interaction at the ecosystem level of organisation and larger spatial-temporal scales (Figure 3.2) is challenging. Achieving this understanding requires far more information than can be obtained by simply scaling up from studies of single species in isolation and measuring bulk stocks and fluxes as performed in the classic Lindeman (father of ecosystem biology) ecosystem ecology studies [Lindeman, 1942]. In essence, the diversity of species, their ecological roles, and the interactions between them are key to understanding the functionality of an ecosystem [Purdy et al., 2010]; therefore a proper systems biology approach is needed.

Today, microbial ecology is undergoing a revolution because of the recent advancements in NGS technologies, which are providing huge quantities of both taxonomic and functionally relevant information [Woodward, 2010], allowing general theories about biodiversity, biogeography and ecosystem functioning to be addressed for the first time. Furthermore by combining NGS and modern molecular techniques with traditional methodologies it is now possible to study a range of levels of organisation, from individual members of a community to whole ecosystems (Figure 3.2), allowing relationships and interconnections to be determined.

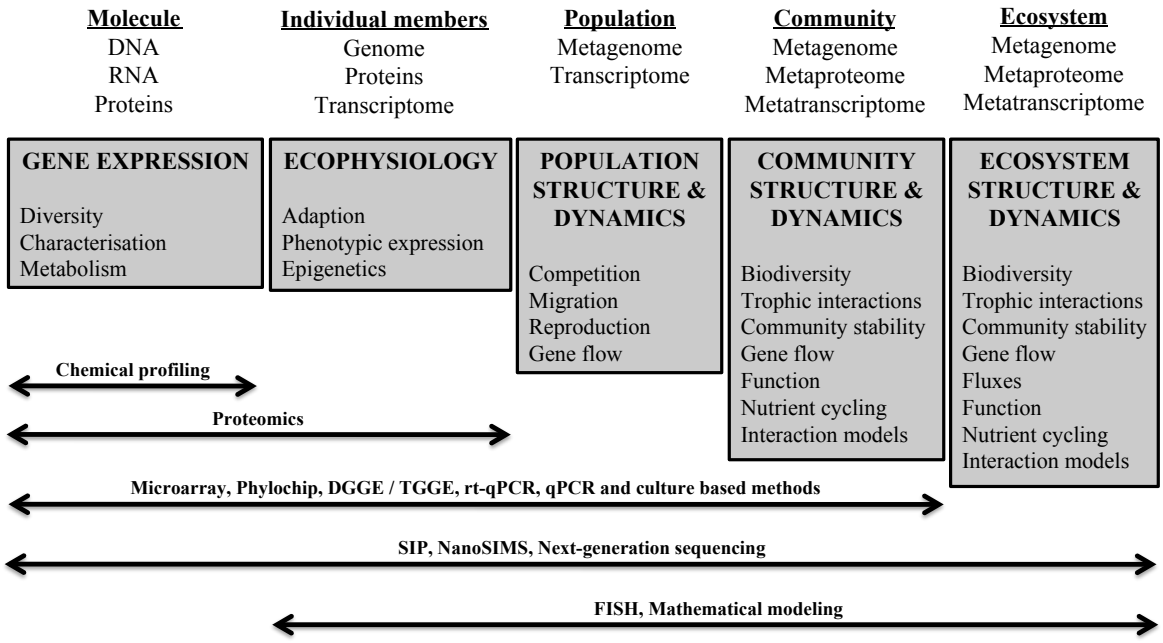


Figure 3.2: Hierarchical organisation of biology from molecules to ecosystems. Below each heading the genetic and ecological (shaded box) measures and fields studied are shown. Arrows represent some of the tools / techniques available to study these areas. Adapted from Purdy et al. [2010]

3.8 Implications for Understanding the Ecology of SSFs

The biology of most naturally-occurring microbial communities, including those found in SSFs, is complex. This complexity is due to the fact that microorganisms are rarely found alone, but thrive in diverse biofilm communities, which collectively process a range of chemicals and nutrients entering the system. Identifying the organisms, their functions and the interrelatedness and interdependency between different species and trophic groups underpins modern microbial ecology [Kuypers, 2007] and is highly relevant to SSFs.

To date, studies of SSFs have focused on characterising and validating the biologically mediated purification mechanisms, by using carefully controlled, laboratory conditions, and conventional plating and isolation techniques, thus ignoring the uncultivable majority [Haig et al., 2011]. Further, these studies have never shown that laboratory-scale SSFs accurately represent full-scale SSFs in terms of their water quality production or microbial community composition, limiting the applicability of their results to full-scale slow sand filters. In addition the true metabolic / degradation capabilities of SSFs have never been examined.

Today, as previously shown there are a wealth of molecular microbiological techniques avail-

able, which when used alongside conventional microbiological tools, will allow several deficiencies in the knowledge of the SSFs ecology and hence performance to be addressed. By building on pivotal slow sand filter studies like those carried out by Barry Lloyd, Monroe Weber-Shirk and Luiza Campos through the use of qPCR, next-generation sequencing, SIP and metagenomics, questions such as:

1. Which organisms are present in full-scale SSFs and what roles do they perform?
2. How does the microbial community structure change both spatially and temporally?
3. Do laboratory-scale slow sand filter mimic the performance and microbial community of full-scale SSFs?
4. What effect does covering SSFs have on the microbial community and filter performance?
5. What mechanisms are responsible for the removal of human pathogens, such as *E.coli*?
6. How effective are slow sand filters at removing modern-day chemical contaminants, such as pharmacological contaminants and heavy metals?
7. Is it possible to improve the performance of SSFs through bioaugmentation?

can be answered, potentially allowing for better operations, system design and tailored water purification for a range of chemical and microbiological contaminants. These four techniques unlike the others mentioned in this chapter provide, accurate quantification of dominant groups abundances (qPCR), high throughput and resolution in assessing community composition and diversity (next generation amplicon sequencing) and provides an unbiased, yet tractable method for determining how pathogens are removed without prior knowledge requirement of which organisms to study (SIP and metagenomics).

Chapter 4

Characterising the Microbiome of Full-Scale Slow Sand Filters

"All have their worth and each contributes to the worth of the others."

J.R.R. Tolkien (The Silmarillion)

Two industrially operated full-scale slow sand filters (ISSFs) were sampled periodically from April until November 2011 to study the spatial and temporal structure of the bacterial community comprising the SSFs. To monitor global changes in the microbial community, DNA from sand samples taken from different depths, locations within the ISSFs, filters ages and operational stages were used for qPCR and Illumina 16S rRNA sequencing. Additionally, fifteen water quality parameters were monitored to assess filter performance, with functionally relevant microbial members being identified using an array of statistical techniques. The bacterial diversity in SSFs was found to be much larger than previously documented, with community composition being shaped by the: characteristics of the SSF (age, and depth) and sampling characteristics (month, side and distance from the influent and effluent pipe). This study is the first to comprehensively characterise the microbial community of SSFs and link specific microbes (*Acidovorax, Halomonas, Sphingobium* and *Sphingomonas*) to water quality parameters and overall filter performance. Additionally, this study's results indicate that species evenness is critically related to SSF performance. By better understanding the SSF community structure it will be possible to improve water quality performance through

A condensed version of some of the work in this chapter is under review: Haig, S. Quince, C. Davies, R.L. Dorea, C.C. and Collins, G. Spatial and Temporal Microbial Community Analysis Identifies Functionally Relevant Microbes for Slow Sand Filter Performance. ***Under review in mBio***

optimisation of SSF operation and design in the future.

4.1 Introduction

As discussed in Chapter 2 slow sand filtration has been an effective means of treating water for the control of microbiological contaminants in both small and large community water supplies for over 200 years. However, such systems lost popularity to rapid sand filters mainly due to smaller land requirements and less sensitivity to water quality variations. Slow sand filtration is still a particularly attractive process because its operation does not require chemicals or electricity. Despite, this SSFs can still achieve a high level of treatment, which is mainly attributed to naturally-occurring, biochemical processes in the filter. Several microbiologically-mediated purification mechanisms (e.g. predation, scavenging, adsorption and bio-oxidation) have been hypothesised or assumed to occur in the biofilm that forms in the filters but these have not yet been comprehensively verified. Thus, SSFs are operated as "black boxes" and knowledge gaps pertaining to the underlying ecology and ecophysiology limit the design and optimisation of the technology.

Recently, there have been a number of studies that have attempted to characterise the purification mechanisms and the microbes responsible for them [Weber-Shirk and Dick, 1997a, Bahgat et al., 1999, Calvo-Bado et al., 2003, Aslan, 2008, Wakelin et al., 2011, Ramond et al., 2013]. However, such studies have focused on specific aspects of SSFs (e.g. Schmutzdecke [Wakelin et al., 2011]) or specific purification mechanisms (e.g. nitrate removal [Aslan, 2008]), and have been performed in non-verified laboratory-scale SSF microcosms, which may not accurately reflect the true microbial community found in real SSFs. Although all of these studies have provided great insight into the biological processes occurring within SSF, a deeper analysis of the structure and dynamics of the microbial community which underpins slow sand filters as a function of performance and operational conditions is needed. Such a study has the potential to reveal important and under-appreciated structure-function relationships, which could greatly improve operation, management and design of these systems. Previous, microbial ecology papers on engineered systems with a biological component have shown that functional stability and robustness is correlated with several components of biodiversity, such as species richness and evenness [Hashsham et al., 2000, Bell et al., 2005,

Allison and Martiny, 2008, Wittebolle et al., 2009, Werner et al., 2011] however no such study has ever been performed on SSFs.

This chapter will present and discuss the findings of periodical sampling of two industrially operated full-scale slow sand filters (ISSFs), in order to study the spatial and temporal structure of the bacterial community comprising the filters. It will also strive to determine how specific microbial groups are related to overall filter performance. By determining the organisms and community metrics comprising SSFs, along with their functional performance over their lifecycle (Figure 4.1), it will enable the development of a detailed slow sand filter microbiome blueprint which can be used as a benchmark for subsequent comparisons.

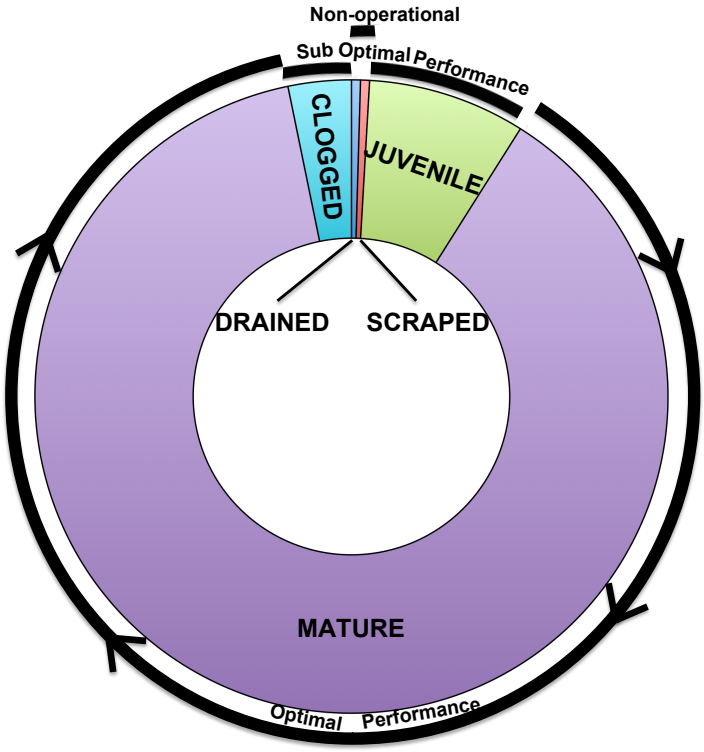


Figure 4.1: Lifecycle of slow sand filter. The size of the ring component corresponds to the proportion of time the SSFs are at that stage. The black outer lines provide performance related information.

Hypotheses

Hypothesis 4.1 *Slow sand filters become more spatially heterogenous with age.*

Hypothesis 4.2 *Decrease in species diversity from inlet to the outlet of slow sand filters is related to substrate concentration gradients.*

Hypothesis 4.3 *The age of slow sand filters will explain the largest proportion of the variance in the microbial community.*

Hypothesis 4.4 *The quality of water produced (number of parameters met within the European Union Council Directive) will positively correlate with SSF age.*

Hypothesis 4.5 *Slow sand filters with identical influent compositions and filter performance will share genus level similarity in community composition.*

4.2 Materials and Methods

Within this section a detailed description of the various methods and approaches used to sample and characterise the microbial community of ISSFs is presented.

4.2.1 Operation and Sampling of Industrial SSFs

Two dimensionally-identical ISSFs (Filter A and Filter B), at Scottish Water's Fairmilehead Water Treatment Works in Edinburgh were sampled approximately monthly from April until November 2011, with the filters being decommissioned by the addition of chlorine in November 2011. The filters differed in only their age (days since scraped). Additional to the monthly sampling, an eight-week intensive sampling strategy was adopted from May to June. The purpose of the intensive sampling programme was to monitor the SSF community closely during the time it was hypothesised to be more microbially active. In total, 16 sampling sessions were conducted, providing data from representative points in the filters lifecycle (Figure 4.1); it should be noted that drained filters were sampled 20h after draining had occurred. Further, the first sampling points taken during decommissioning were collected 20h after chlorine delivery and; both filters remained operational with the produced water entering the distribution system until November 2011.

Fairmilehead is a drinking water treatment works with seven filters, each with an approximate area of 1800m^2 and composed of a 1m filter bed and 30cm of under-drainage. The sand filters receive their influent water from several upland reservoirs (Talla: N 55.4745°, W 3.3848°and Megget: N 55.4851°, W 3.2819°) in the borders of Scotland via 45km of gravity fed tunnels, pipelines and aqueducts (Figure 4.2). The Talla reservoir was built in the late

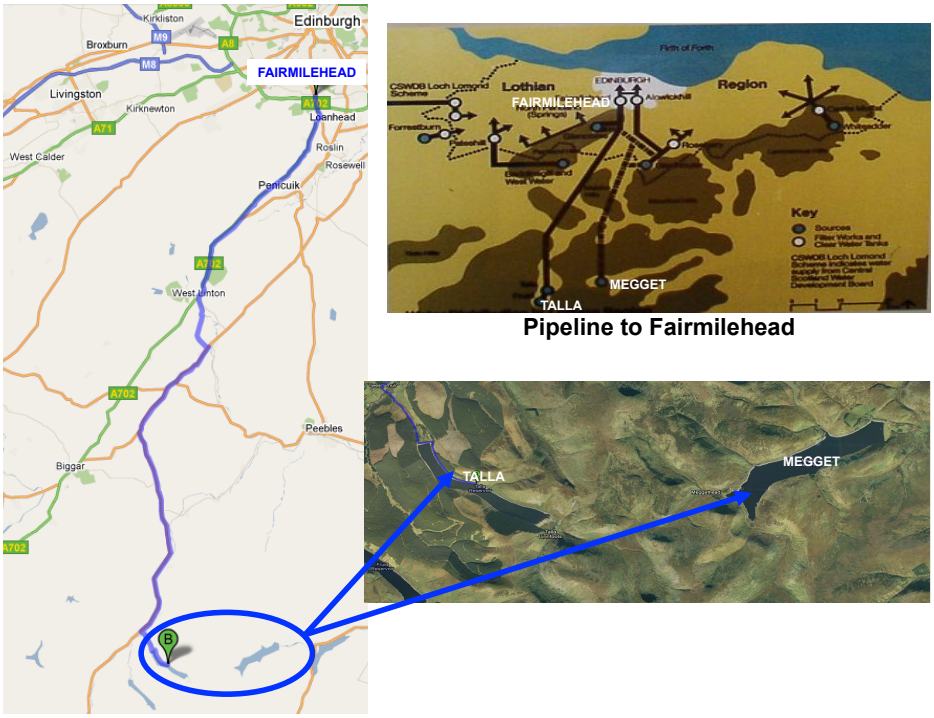


Figure 4.2: Maps showing the location of the Talla and Megget reservoirs in relation to Fairmilehead water treatment works.

19th century, opening in 1899, and the Megget reservoir was built in the 1980s to supplement Edinburgh's water supply. Like the Talla, the Megget resevoir is a man-made water body constructed by damming the valley at its narrowest point; this 259 hectares reservoir is held back by the largest earth dam in Scotland and collects water from the Tweedsmuir Hills. Once the reservoir water has percolated through the slow sand filters the effluent is chlorinated and then distributed via Scottish water's distribution system to the consumers. The Fairmilehead site supplies 160 million Litres/day.

4.2.2 Filter Bed Sand Characterisation

The sand comprising the filter bed in the sand filters was analysed to determine the effective grain size distribution (ES) and uniformity coefficient (UC). To determine these values, 300g of sand were dried and the ES was determined through sieve analysis by plotting the cumulative weight of sand against the sieve mesh diameter. In the water industry the suitability of sand for slow sand filtration is determined by the ES_{10} (sieve mesh by which 10% of the sand weight passes) and the UC (degree of sand uniformity, calculated by dividing ES_{60} by the ES_{10}). Thus if there are large amounts of fine or coarse grains in the sand this will affect the values of ES_{10} and ES_{60}, and hence the UC. For optimal drinking water production

using SSF an Es_{10} between 0.2mm and 0.25mm and a UC lower than 3 (ideally under 2) is recommended [Van Dijk et al., 1978].

4.2.3 Sampling the Filter Beds

Sampling entailed collecting one 50cm sand core from each side of both filter beds using a Multi-stage sediment sampler (AMS, America Falls, USA) (Figure 4.4). Cores were taken from the three separate accessible sides of the filters (Figure 4.3). The multi-stage sampler does not have an auger head to disturb the sand during emplacement, and so must be pushed vertically into the sand bed and then hammered deeper into the filter bed using the attached jack-hammer (Figure 4.5). The sampler (Figure 4.4) is constructed from stainless steel and features a vented top cap with a one way valve, extendable body and coring tip with a butterfly valved retainer. The sampler is assembled by threading (male to female) the core tip onto the lower liners and then threading this onto the extension rods and finally the 20lb slide hammer. Once assembled the sampler is inserted into the filter and by using the slide hammer it can reach the necessary depth (50cm deep in the filter bed). The sampler is recovered by reverse hammering with the slide hammer; once out of the filter the top cap is removed using the slip wrench on the sampler body and an adjustable wrench on the top cap threaded extension connection. The liner and butterfly retainer are then removes from the Multi-Stage Base Section and plastic caps are placed on either end of the liner to secure the sand section.

These undisturbed cores were sectioned at eight depths (0, 4, 10, 15, 20, 30, 40 and 50 cm) and 0.5 g of each subsample was used for DNA extraction using the FastDNA spin kits for soil (MP Bio-Medical, Cambridge, UK); for details of this procedure, see Appendix B. Extracted DNA was used for Illumina 16S amplicon sequencing and qPCR to identify and quantify the various members of the microbial community. It should be noted that in order to design specific probes and primers for qPCR a clone library representing the entire sand core was created.

4.2.4 Water Quality Analysis

At each sampling occasion 2L influent and effluent water were collected from the two filters, and pH and water temperature were measured on site with portable meters. Water sam-

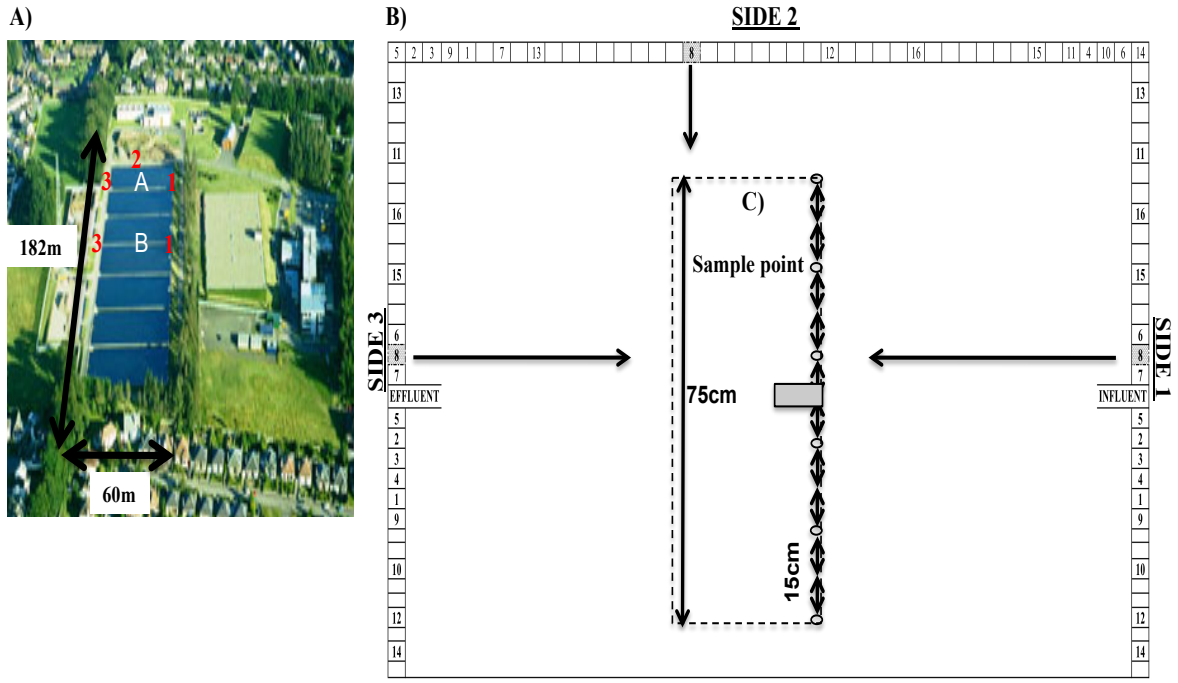


Figure 4.3: Schematic of Fairmilehead filter plant ***A)****, sampling locations within the filter bed, numbers represent the sampling order* ***B)*** *and mesoscale sampling locations* ***C)****.*

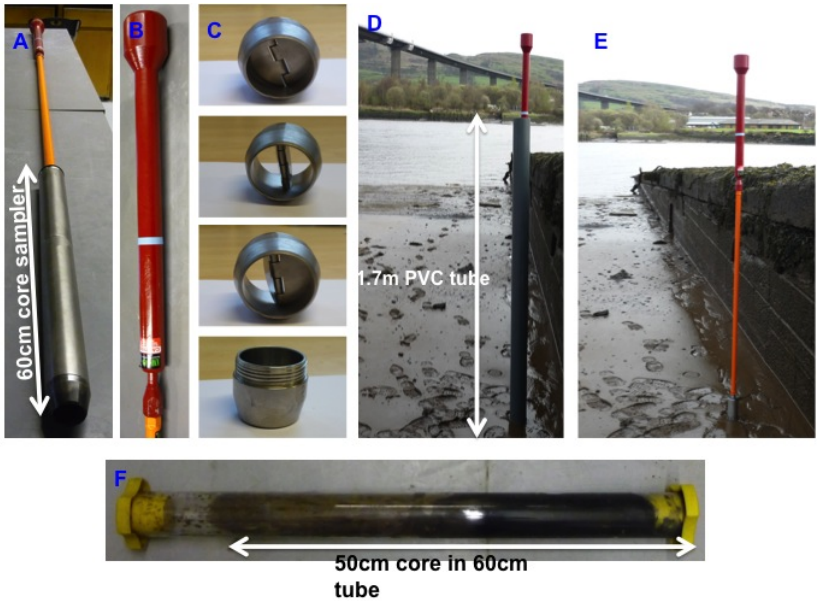


Figure 4.4: Photos of the multi-stage sampler. (A) shows the whole sampler. (B) shows the 20lb slide hammer. (C) shows the butterfly valved core tip. (D and E) illustrate how the sampler will be used at the SSF site; inserted into a hollow PVC tube which helps to guide the sampler down to the filter bed. Once the blue mark on the slide hammer is just above the PVC tubing the sampler has reached a depth of 50cm. (F) shows the plastic liner which the sample is contained within.

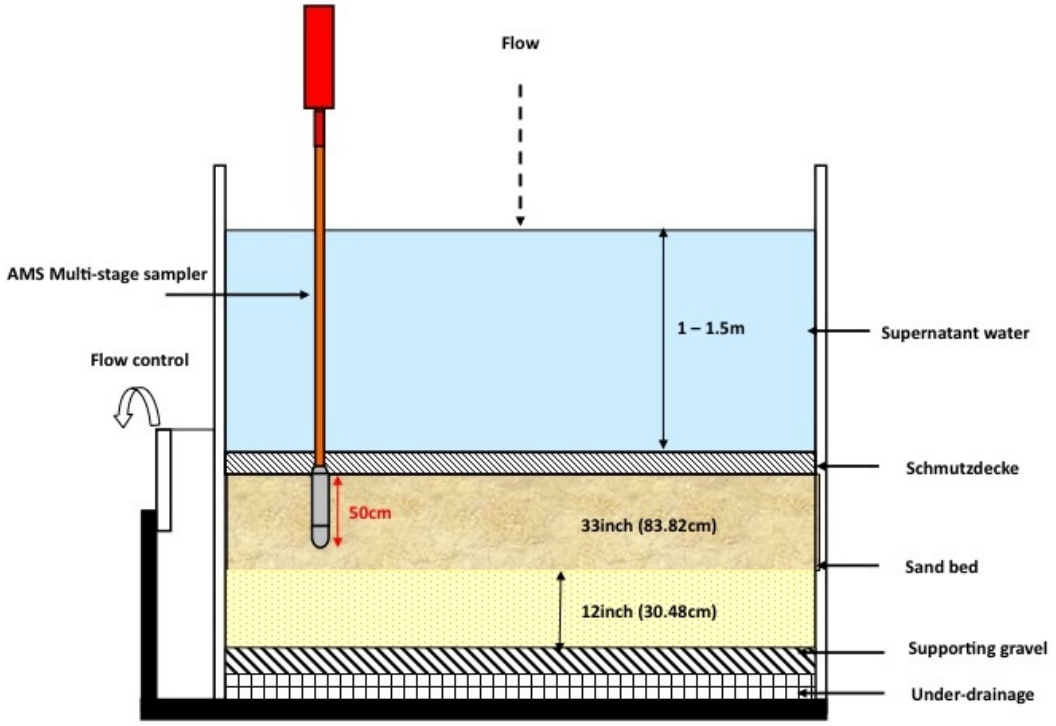


Figure 4.5: Schematic of a slow sand filter and the location of AMS's Multi-stage sampler during sampling

ples were processed in triplicate for turbidity, dissolved organic carbon, specific ultraviolet absorbency, chemical oxygen demand, nitrate, nitrite, ammonia and phosphate. Coliform counts, and total viable bacteria counts at 30°C and 13°C, were measured five times each (Table 4.1) using the methodology found in Appendix A. Environmental temperature, water temperature, environmental light intensity and light intensity 5cm submerged into the filters were constantly measured from May to November 2011, and logged using Hobo Data Loggers (Onset Computer Corporation, UK). To evaluate overall filter performance a new parameter called "Performance metric" ($\bigtriangledown$) was created. This parameter assigned the effluent of each filter a number from 0 - 10 based on the number of the top 10 water quality parameters outlined by European Union Council Directive [1998] it fulfilled (Table 4.1). The ranking is as follows: 0 - 4 is designated as poor performance, 5 - 6 as average, 7 - 8 as good and 9 - 10 as excellent performance.

4.2.5 DNA Sequencing

A 16S rRNA gene clone library was constructed to identify likely important groups and to provide positive controls for subsequent qPCR assay development. Further, in order to char-

Table 4.1: Measured water quality parameters in this study, alongside the maximum concentration allowed, as designated by European Union Council Directive [1998], all analytical methods used come from the Standard Methods used by the water industry [Clesceri et al., 2005].

Parameter	Analysis Technique	Concentration	Standard Method
Ammonia$*$	Salicylate digestion	0.05mg/L	P4500-NH_3A
COD$*$	Closed Reflux, titrimetric method	30	P5220C
Coliforms $*$	Serial dilutions on MLSB	0/100ml	9222
DOC$*$	Samples filtered and processed using Hach TOC kit	10mg/L	P5310C
Nitrate$*$	Cadium Reduction	50mg/L	P4500-NO_3_E
Nitrite$*$	Diazotization	0.5mg/L	P4500-NO_2_B
Orthophosphate$*$	Ascorbic Acid	0.1mg/L	P4500-PE
pH$*$	Hydrogen ion concentration	6.5-9.5	
SUVA$*$	Division of UV_{254nm} by DOC measurement	3.5	5910B
TVB	Serial dilutions on plate count agar.		Bahgat et al. [1999]
Turbidity$*$	Nephelometric measurement	4NTU	P2130B
UV_{254nm}	Absorbance at 254nm after filter (0.45μl filter)		5910

$*$ denotes parameters which are used to assign $\bigtriangledown$. Abbreviations: COD: Chemical Oxygen Demand, DOC: Dissolved Organic Carbon, SUVA: Specific UV Absorbance, and TVB: Total Viable Bacteria at 30°C

acterise, and calculate abundances of the whole microbial community found within SSFs, Illumina 16S rRNA amplicon sequencing was also performed.

4.2.5.1 Clone library

A 16S rRNA clone library was created using the universal prokaryotic primers (27F: [5'-GAGTTTGATCCTGGCTCAG-3'] and 1392R: [5'-ACGGGCGGTGTGTRC-3']) on sand from a mixed core (i.e., sand from all depths) from the industrial SSFs using the TOPO TA kit (Invitrogen) following manufacturers instructions. One hundred clones were screened using amplified ribosomal DNA restriction analysis, with the restriction enzyme HAEIII (Promega,UK). Operational taxonomic units (OTUs) were identified, based on restriction cleavage patterns, and clones representing the different OTUs were sequenced (Genepool, Edinburgh). Chimera checking was performed using Bellerophon [Huber et al., 2004]. For details of the exact protocol used see Appendix B.

4.2.5.2 Illumina 16S Amplicon Sequencing

The 16S rRNA gene amplicon of 674 full-scale SSF samples (56 water and 618 sand samples), representing different depths, filters, filter ages and levels of filter performance were

processed by the Earth Microbiome Project (EMP) using the 515F [5'-GTGCCAGCMGCCGCGGTAA-3'] and 806R [5'-GGACTACHVGGGTWTCTAAT-3'] primers following the protocol outlined by Caporaso et al. [2012]. Amplified samples were then pooled together (equimolar concentrations) and sequenced on Illumina HiSeq 2000. Sequenced samples were then quality filtered and demultiplexed. Taxonomy for all samples was assigned against the Greengenes database [DeSantis et al., 2006] which pre-clustered at $\geq$97% OTU identity. Additionally, two mock communities processed in triplicate were also included to act as positive controls. Raw and processed reads can be found at: www.microbio.me/emp/.

4.2.6 qPCR

To calculate the changes in the absolute concentrations of the dominant phyla and classes (based on indicative results from the 16S rRNA clone library) qPCR assays were done, both over time and depth within SSFs. Eight phyla- and four class-specific qPCR primers were designed. Primer information can be found in Table 4.2. All samples were processed in triplicate along with two negative controls and standards.

qPCR assays were conducted in polypropylene 96-well plates on a CFX96 Real-Time PCR Detection System (Bio-Rad). Each 10μl reaction contained the following: 5μl of SsoFast EvaGreen Supermix (Bio-Rad), 0.8μl of each primer (0.4μM; Eurofins), 2.4μl water and 1μl of template DNA (1ng μl^{-1}). PCR conditions for total bacterial 16S, *Acidobacteria, Bacteroidetes, and Betaproteobacteria* were 15 mins at 95°C, followed by 40 cycles of 95°C for 1 min, 30 s at the annealing temperature, and 72°C for 1 min; all other assays were performed for 35 cycles (Table 4.2). Melting curve analysis of the PCR products was conducted following each assay to confirm that fluorescence signal originated from specific PCR products and not from primer dimers or other artefacts. Standard curves were generated using triplicate 10-fold dilutions of linearised (EcoR1) plasmid DNA (generated from the clone library, representatives of the different groups). DNA concentrations ranged from 8×10^8 to 8×10^1ng/μl. For all the qPCR assays, there was a linear relationship between the log of the plasmid DNA copy number and the calculated threshold cycle value across the specified concentration range (R^2 >0.99 in all cases). Amplification efficiencies, calculated using the method described by [Pfaffl, 2001], varied from 1.8-2.1 across the 13 qPCR assays; these values are consistent with those reported in other studies [Fierer et al., 2005, Castillo et al.,

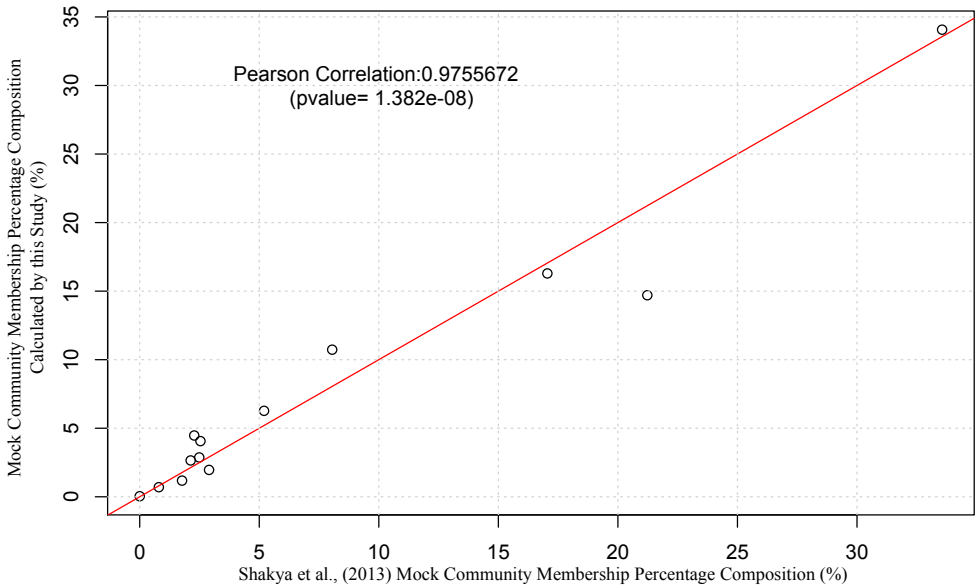


Figure 4.6: Scatter plot depicting the correlation between the percentages of the different phyla calculated by this study's qPCR assays and the known abundances [Shakya et al., 2013].

2006].

To ascertain the accuracy of the qPCR assays, a blind test was performed for each phylum / class with a previously quantified mock community [Shakya et al., 2013], results of which can be seen in Figure 4.6. Calculating Pearson's correlation coefficient between the previously determined abundances in this mock community against our measured abundances resulted in a very strong (0.976) positive correlation, which confirms the accuracy of the qPCR assays and subsequent phyla abundances in unknown samples.

4.2.7 Statistical Analysis

Correlations between water quality parameters were explored using the nonparametric Kendall τ procedure and seasonal effects were examined using a seasonal Mann-Kendall test. Additionally, the functional relationships between water quality parameters and the absolute abundance of eight bacterial phyla and four bacterial classes were analysed using stepwise multivariate forward/reverse regression analysis.

Taxonomic and OTU tables generated for the Illumina samples were used to calculate pair wise similarities among samples based on the Bray-Curtis similarity index. The resulting matrices were examined for temporal and spatial patterns in bacterial community structure

Table 4.2: qPCR primers for SSF

Primer Name	Target Taxon	Sequence (5'-3')	Paired Primers	Annealing Temperature (°C)	Amplicon Length (bp)	Reference
Acid31F	Acidobacteria	GATCCTGGCTCAGAATC	Eub518R	50	500	Fierer et al. [2005]
Eub518R	Acidobacteria	ATTACCGCGGCTGCTGG	Acid31F, Gem440	50	500	Fierer et al. [2005]
Act920F3	Actinobacteria	TACGGCCGCAAGGCTA	Act1200R	60.5	170	De Gregoris et al. [2011]
Act1200R	Actinobacteria	TCRTCCCCACCTTCCTCCG	Act920F3	60.5	170	De Gregoris et al. [2011]
cfb319	Bacteroidetes	GTACTGAGACACGGACCA	Eub518R	65	200	Fierer et al. [2005]
1055YF	Chloroflexi	ATGGYTGTCGTCAGCT	CFX1223R	55	170	Yoon et al. [2010]
FlexR	Chloroflexi	CCATTGTAGCGTGTGTGTMG	1055YF	55	170	adapted from Yoon et al. [2010]
Gem677	Gemmatimonadetes	TTCCSGGTGTAGCGGTGG	Eub518R	59.2	165	adapted from Morales and Holben [2009]
Nspra-675f	Nitrospira	GCGGTGAAATGCGTAGAKATCG	Nspra746R	58	70	Graham et al. [2007]
Nspra746R	Nitrospira	TCAGCGTCAGRWAYGTTCCAGAG	Nspra-675f	58	70	Graham et al. [2007]
Pln930R	Planctomycetes	CTCCACCGCTTGTGTGA	PlnF	61	215	adapted from Blackwood et al. [2005]
PlnF	Planctomycetes	AGGAACRCCGGWGGC	Pln930R	61	215	adapted from Blackwood et al. [2005]
α682F	α Proteobacteria	CNAGTGTAGAGGTGAAATTC	908αR	61.5	180	De Gregoris et al. [2011]
908αR	α Proteobacteria	CCCCGTCAATTCCTTTGAGTT	α682F	61.5	180	De Gregoris et al. [2011]
Eub338	β Proteobacteria	ACTCCTACGGGAGGCAGCAG	β680R	60	360	Fierer et al. [2005]
β680R	β Proteobacteria	TCACTGCTACACGYG	Eub338	60	360	Fierer et al. [2005]
γ1080F	γ Proteobacteria	TCGTCAGCTCGTGTYGTGA	γ1202R	60.5	170	De Gregoris et al. [2011]
γ1202R	γ Proteobacteria	CGTAAGGGCCATGATG	γ1080F	60.5	170	De Gregoris et al. [2011]
Del979F	δ Proteobacteria	GGTGGAGCATGTGGTT	DelR1434	56	360	Adapted from Nechitaylo et al. [2010]
DelR1434	δ Proteobacteria	GACTTCTGGAGCAAYYG	DelF972p	56	360	Nechitaylo et al. [2010]
VMB537F	Verrucomicrobia	GCCAGCAGCCGCGGTAATACA	VMB1295r	63.5	600	adapted from O'Farrell and Janssen [1999]
VMB1295r	Verrucomicrobia	GCAGMCTICAATCTGAACTGRGC	VMB537f	63.5	600	adapter from O'Farrell and Janssen [1999]
Eub338	All bacteria	ACTCCTACGGGAGGCAGCAG	Eub518R	53	200	Fierer et al. [2005]

by using non-metric multidimensional scaling (NMDS) as implemented in the Vegan package [Oksanen et al., 2012]. Significant differences in the microbial community composition between filters, age, depth, location of core, season and addition of chlorine were determined by using the adonis function, which performs a nonparametric multivariate analysis of variance (MANOVA) [Anderson, 2001]. To determine the contribution of individual taxon to differences in filter performance SIMPER (similarity percentage) analysis [Clarke, 1993] was used. SIMPER analysis is a useful measure of the magnitude of difference, however, in order to decide whether a taxon differed significantly pairwise t-tests (kendall non-parametric) adjusted for multiple comparisons using Benjamini-Hochberg false-discovery method [Benjamini and Yekutieli, 2001] were performed. Only taxa with a false-discovery rate of less than 5% were reported. Shannon diversity indices, Chao's richness, Pielou's evenness and rarefaction curves were calculated using the Vegan package on rarefied samples at the 3% genetic distance. The relationships between environmental variables and patterns in bacterial community structure were examined by canonical correspondence analysis with significance tested by analysis of variance tests (ANOVA) after reducing the overall suite of environmental variables using step-wise Akaike's Information Criterion (AIC) model. All statistical analysis was performed in R [R Development Core Team, 2011].

4.3 Results

Within this section the results of the study, alongside discursive analysis will be presented, with an overall discussion being provided in Section 4.4.

4.3.1 Sand Characterisation

Referring to Figure 4.7 the uniformity coefficient (UC) of the sand at Fairmilehead was calculated as 2.8, which is within the guidelines for SSF performance [Huisman et al., 1974]. However, it can be seen that the Es_{10} is double the optimal value stated in the guidelines and that the average particle size is 1300μm, four times greater than the guide-lined size. Irrespective of these finding this sand has been successfully used for centuries in providing high quality drinking water at this site and further emphasises the flexibility in the sand type which can be used for slow sand filtration.

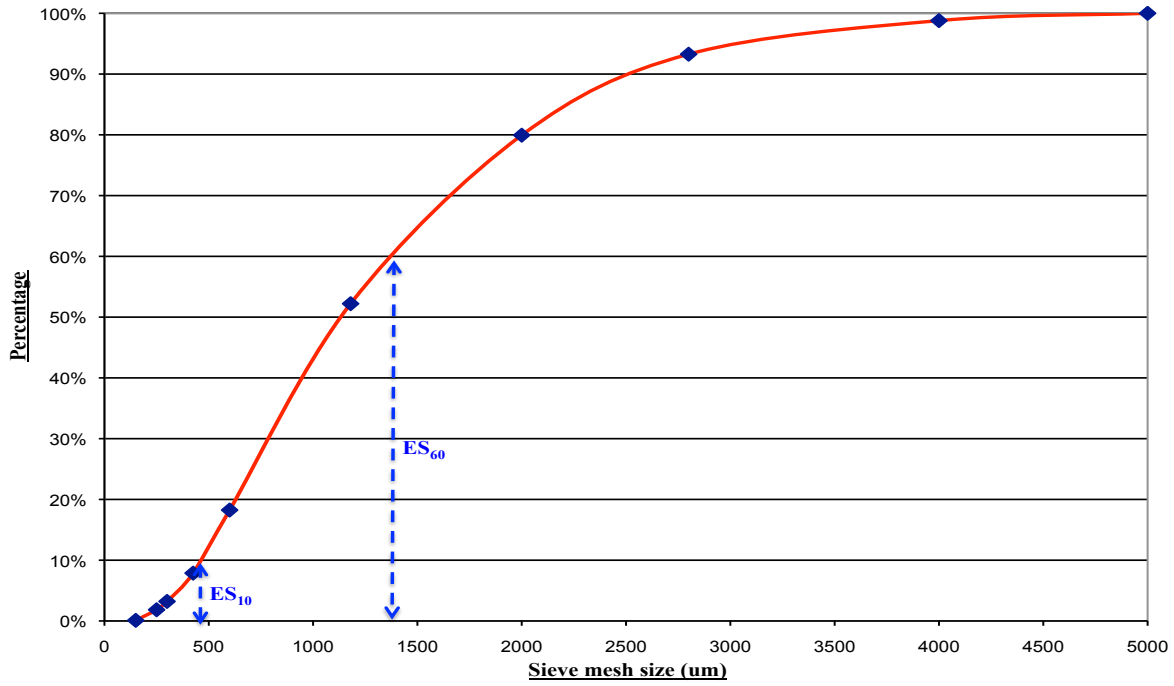


Figure 4.7: Effective particle size distribution graph for the sand taken from Fairmilehead

4.3.2 Water quality

Two industrially operated full-scale SSFs were sampled routinely from April 2011 until they were decommissioned in November 2011, equating to a total of 16 sampling trips. Both industrial sand filters performed extremely well in terms of the Water Supply, Water Quality, Scotland Regulations 2001 (part of the Water Scotland Act 1980). Influent, effluent and percentage removal results can be seen in Tables 4.3, 4.4 and 4.5. Overall, the filters failed to meet only one drinking water requirement; the coliform levels. However it should be noted that these filters are not a single point of purification, with effluent from the filters being chlorinated before being distributed, a process which would remove the low levels of coliforms present in the effluent. Additionally, in terms of performance there was no statistically significant difference (p-value 0.08, Wilcox Test) in filter performance between filters A and B.

Referring to Figure 4.8 it can be seen that all of the water quality parameters significantly correlate with at least six other parameters, with DOC correlating the least and NH_4 and DO correlating with every parameter. Additionally, several parameters (Table 4.6) showed varying strengths of correlation with the age of the filters, with coliform removal showing the strongest positive correlation, with optimum coliform removal occurring after seven weeks. This is consistent with operators reports of SSF performance increasing with age / matu-

Table 4.3: Summary of physical and chemical characteristics of influent water. Age: number of days since filters were scraped. Units of measurement are: TEMP:°C, PH:pH, NTU:turbidity, DO (Dissolved Oxygen): mg/L, DOC (Dissolved Organic Carbon): mg/L, UV: absorbance at 254nm, SUVA (Specific UV Absorbance): L/mg-M, COD (Chemical Oxygen Demand): mg/L, PO4 (Phosphate): mg/L, NO3 (Nitrate): mg/L, NO2 (Nitrite): mg/L, NH4 (Ammonium): mg/L, Total Viable Bacteria grown at 30°C (TVB) and 13°C (TVB13) and coliforms: cfu/ml

Average (based on 5 replicas) Water Quality Parameters for the Influent water supplying Two Industrial Slow Sand Filters

FILTER	Additional Info	Sampled	AGE	TEMP	PH	NTU	DO	DOC	UV	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB 13oC	Coliforms
A	In service	19-Apr-11	393	12.6	7.97	1.18	8.8	4.8	0.11	2.19	8.05	0.24	0.01	0	0.26	2800	1000	900
A	In service	03-May-11	407	12.4	7.84	1.19	10	5.1	0.2	3.88	0.32	0.13	0.02	0	2.12	124	1332	3.9
A	In service	10-May-11	414	12.7	7.99	1.56	9.7	5.05	0.21	4.2	2.44	0.65	0.04	0	0.01	88	3370	21
A	In service	17-May-11	421	12.6	7.99	3.82	9.2	4.93	0.22	4.46	4.72	0.1	0.03	0	0.16	350	112	20.8
A	In service	31-May-11	435	12.8	8.44	1.3	10	5.02	0.21	4.14	9.13	5.91	0.01	0	2.63	52.8	15.4	13.8
A	Drained	07-Jun-11																
A	Not in service	14-Jun-11	7	13.5	8.06	0.89	10	3	0.2	6.5	1.66	4.04	0	0	3	249	108	21.6
A	Not in service	21-Jun-11	14	14.3	7.3	0.94	9.7	1.9	0.18	9.56	1.8	7.89	0	0	2.64	180	36	52.8
A	In service	28-Jun-11	21	15.1	7.22	0.54	10	5.21	0.2	3.8	2.42	0.01	0.01	0	0.04	58	4	25.8
A	In service	06-Jul-11	29	13.6	7.2	0.58	9.7	5.12	0.22	4.3	1	0.04	0.01	0.01	0.04	288	6	89.8
A	In service	12-Jul-11	35	14.9	7.32	0.67	10	5.23	0.21	4.05	0.33	6.03	0.01	0	0.02	47	8	20.4
A	In service	19-Jul-11	42	16.7	7.1	0.71	8.7	5	0.2	4	13.93	0.03	0.01	0	0.03	111	15.8	38.4
A	In service	26-Jul-11	49	17.4	7.1	0.64	9.5	6.1	0.24	3.87	7.13	6.67	0.01	0	0.01	190	112	81.8
A	In service	24-Aug-11	78	15.7	7.12	0.24	9.3	4.42	0.11	2.49	6.15	7.87	0.02	0.01	0.01	132	76	42.8
A	Chlorine added	20-Sep-11	105	15.6	7.4	0.69	9.2	1.91	0.11	5.81	2.97	0.01	0	0	0.04	634	126	36.2
A	Chlorine added	18-Oct-11	133	15.6	7.21	0.58	9.5	4.71	0.18	4.05	4.85	2.95	0.01	0	0.03	208.57	49.69	47.89
B	In service	19-Apr-11	158	13	7.96	0.93	8.7	4.2	0.23	4.93	18.86	0.15	0.04	0	0.24	2800	57.3	236
B	In service	03-May-11	172	13.7	7.29	1.68	9.2	4.25	0.2	3.75	0.17	0.04	0.08	0	0.92	218	64.58	14
B	Not in service	10-May-11	7	12.6	8	1.3	9.8	4.95	0.22	4.51	1.95	0.18	0.06	0	0	218	596	334
B	Not in service	17-May-11	14	12.3	7.89	1.12	9.2	4.97	0.23	4.53	4.45	0.05	0.05	0	3.27	218	104	26.4
B	Not in service	31-May-11	28	12.6	5.86	0.97	10	5.17	0.22	4.2	4.87	0.22	0.02	0	3.2	45.2	10.6	10.6
B	In service	07-Jun-11	35	13.5	7.78	0.83	9.8	3.1	0.12	3.87	5.01	4.16	0.01	0	0.01	320	32	1.8
B	In service	14-Jun-11	42	13.7	7.78	0.79	10	2.7	0.11	4.11	2.27	5.64	0	0.01	0.01	1860	16	20.6
B	In service	21-Jun-11	49	14	6.98	1.62	9.6	2.9	0.21	7.11	2.1	0.11	0	0.01	0.93	585	0	35.8
B	In service	28-Jun-11	56	14.9	7.26	0.56	10	4.95	0.22	4.43	2.35	0.08	0.01	0.01	0.01	254	33.6	33.6
B	In service	06-Jul-11	64	13.4	7.08	0.6	10	5.14	0.22	4.28	11.4	0.06	0.02	0	0.01	268	0	100
B	In service	12-Jul-11	70	14.7	7.22	0.68	10	5.29	0.21	4.03	0.21	1.67	0.01	0	0.01	43.6	6	30.2
B	In service	19-Jul-11	77	16.8	7.1	0.8	8.9	5	0.22	4.4	10.03	7.2	0.02	0	0.04	75.2	56.6	47.6
B	In service	26-Jul-11	84	16.9	7.1	0.8	9.7	7.1	0.25	3.54	5.18	0.07	0.03	0	0.01	1500	183	24.8
B	In service	24-Aug-11	113	15.4	6.84	0.35	9.6	4.61	0.11	2.38	5.11	0.56	0.04	0	0.01	586	236	16
B	Chlorine added	20-Sep-11	140	13.1	7.68	1.6	10	2.92	0.12	4.01	3.3	0.19	0.01	0	0.03	1530	636	51
B	Chlorine added	18-Oct-11	168	9.22	6.9	0.34	10	2.32	0.12	5	1	5.15	0.06	0	0.02	342	236	2.4

Table 4.4: Summary of physical and chemical characteristics of effluent water. Age: number of days since filters were scraped. Units of measurement are: TEMP: °C, PH:pH, NTU:turbidity, DO (Dissolved Oxygen): mg/L, DOC (Dissolved Organic Carbon): mg/L, UV: absorbance at 254nm, SUVA (Specific UV Absorbance): L/mg-M, COD (Chemical Oxygen Demand): mg/L, PO4 (Phosphate): mg/L, NO3 (Nitrate): mg/L, NO2 (Nitrite): mg/L, NH4 (Ammonium): mg/L, Total Viable Bacteria grown at 30°C (TVB) and 13°C (TVB13) and coliforms: cfu/ml and Performance Metric (▽): score assigned based on overall performance (explained in Section 4.2.4).

Average (based on 5 replicas) Water Quality Parameters for the Effluent of Two Industrial Slow Sand Filters

FILTER	Additional Info	Sampled	AGE	TEMP	PH	NTU	DO	DOC	UV	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB 13oC	Coliforms	Performance Metric
A	In service	19-Apr-11	393	10.6	7.84	0.23	8.5	4.73	0.2	4.29	13.72	0.16	0.08	0	0.01	480	22	2	7
A	In service	03-May-11	407	12.2	7.79	0.3	9.6	5	0.16	3.28	0.37	0.09	0.13	0	0.96	6	38	0	9
A	In service	10-May-11	414	13	7.81	0.22	9.2	5.08	0.18	3.5	5.39	0.24	0.13	0	0.01	18	34	0.2	9
A	In service	17-May-11	421	12.2	7.04	0.58	9	4.94	0.18	3.64	7.92	6.63	0.11	0	3	912	466	0.2	7
A	In service	31-May-11	435	12.2	7.96	0.82	8.4	5.08	0.18	3.46	6.13	0.11	0.06	0	2.5	35.6	1.6	0.6	7
A	Not in service	14-Jun-11	7	14	7.38	2.24	8.6	2.9	0.18	6.21	1.25	0.23	0	0	4.14	200	10.6	13.2	6
A	Not in service	21-Jun-11	14	14.5	7.12	0.8	9	1.81	0.17	9.3	1.62	0.48	0	0	2.3	134	14	5.2	6
A	In service	28-Jun-11	21	14.3	7.3	0.3	9	4.51	0.19	4.23	2.2	0.01	0	0	0.02	42	0	5.6	8
A	In service	06-Jul-11	29	13.7	6.94	0.26	8.4	5	0.2	3.94	11.67	0.04	0.01	0	0.02	136	2	4.4	8
A	In service	12-Jul-11	35	14.4	7	0.29	8.9	4.92	0.18	3.66	0.31	3.62	0.01	0	0.02	36.4	0.8	2.2	7
A	In service	19-Jul-11	42	16.5	6.94	0.22	7.9	4.9	0.18	3.71	14.1	0.03	0.04	0	0.04	36.4	2	3	8
A	In service	26-Jul-11	49	16.9	7	0.28	7.9	5.06	0.21	4.11	1.22	0.02	0.03	0	0.02	804	74	6	8
A	In service	24-Aug-11	78	16.4	6.9	0.21	7.7	4.45	0.1	2.22	6.06	0.07	0.05	0	0.02	544	184	17.2	9
A	Chlorine added	20-Sep-11	105	14.9	7.56	0.47	8.7	1.92	0.11	5.63	3.01	0.03	0.01	0	0.03	1290	164	19.4	8
A	Chlorine added	18-Oct-11	133	10.9	6.48	0.27	8.2	1.9	0.15	7.89	2.11	0.07	0.06	0	0.03	1021	116	21.4	7
B	In service	19-Apr-11	158	10	6.44	0.22	8.8	6.3	0.21	3.76	11.22	0.12	0.1	0	0	128	111	2	6
B	In service	03-May-11	172	12.3	6.19	0.17	0	6	0.18	3.99	0.03	0.13	0.11	0	1.07	22	108	2	5
B	Not in service	10-May-11	7	12.9	7.74	0.8	9.4	5.1	0.19	3.71	5.98	0.23	0.18	0	0.02	186	540	366	7
B	Not in service	17-May-11	14	11.8	6.87	0.83	9.9	4.96	0.19	3.77	8.29	0.12	0.06	0	5.03	250	76.8	4.4	6
B	Not in service	31-May-11	28	12.3	6.19	0.4	9.8	5	0.2	4	4.6	0.07	0.02	0	2.93	25.8	2.4	2.2	7
B	In service	07-Jun-11	35	13.1	7.58	0.28	8.9	2.9	0.1	3.52	5	0.13	0.02	0	0	6120	90	0.2	8
B	In service	14-Jun-11	42	12.7	7.22	0.39	9.4	1.81	0.1	5.52	1.41	1.21	0.02	0.01	0.01	960	7.6	0.8	8
B	In service	21-Jun-11	49	13.5	7.1	0.29	9.2	2.75	0.19	6.72	2	0.03	0.02	0.01	0.3	21.9	34	9.4	8
B	In service	28-Jun-11	56	13.5	7.22	0.2	9.2	4.82	0.2	4.04	2.22	0.02	0	0	0.03	36	0	0.8	9
B	In service	06-Jul-11	64	13.4	7	0.29	9.3	4.97	0.2	4.02	3.2	0.04	0.01	0.02	0.03	172	0	0.2	9
B	In service	12-Jul-11	70	14.1	7.08	0.18	10	4.9	0.19	3.84	0.21	1.19	0	0	0.05	19.6	0.8	0.8	8
B	In service	19-Jul-11	77	16.3	7.1	0.18	8.7	4.6	0.18	4	8.43	0.35	0.02	0	0.04	311.2	5.6	0.4	8
B	In service	26-Jul-11	84	16.2	7.1	0.2	8.8	4.7	0.22	4.74	4.07	0.04	0.02	0	0.03	2490	112	0.4	9
B	In service	24-Aug-11	113	14.9	7	0.14	9.5	5.08	0.1	1.97	5.26	0.02	0.03	0	0.03	550	276	1.4	9
B	Chlorine added	20-Sep-11	140	12.9	7.22	0.22	9.8	2.91	0.11	3.81	4.42	0.01	0.01	0	0.01	1160	226	0.8	9
B	Chlorine added	18-Oct-11	168	9.48	6.7	0.26	9.8	6.1	0.11	1.8	12.26	0.31	0.06	0	0.02	50	296	1.2	8

Table 4.5: Summary of the percentage removal of the physical and chemical characteristics. Age refers to the number of days since the filter was scraped. Units of measurement are percentage removal, positive number mean removed, negative number indicate an increase of the parameter

Average (based on 5 replicas) Percentage Removal of Various Water Quality Parameters for Two Industrial Slow Sand Filters

FILTER	Additional Info	Sampled	AGE	TEMP	PH	NTU	DO	DOC	UV	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB 13oC	COLIFORM
A	In service	19-Apr-11	393	15.87	1.63	80.51	3.07	1.46	-93.33	-95.89	-70.44	32.23	-700.00	66.67	96.15	82.86	97.80	99.78
A	In service	03-May-11	407	1.94	0.64	74.96	4.29	1.96	17.17	15.46	-17.14	29.55	-712.50	100.00	54.91	95.16	97.10	100.00
A	In service	10-May-11	414	-2.36	2.25	85.90	5.35	-0.59	16.04	16.53	-120.78	63.25	-251.35	-200.00	0.00	79.55	98.99	99.05
A	In service	17-May-11	421	3.34	11.89	84.90	1.87	-0.20	18.18	18.35	-67.73	-6738.14	-224.24	-100.00	-1742.33	-160.57	-316.07	99.04
A	In service	31-May-11	435	4.54	5.69	37.08	18.80	-1.20	15.38	16.38	32.85	98.09	-500.00	100.00	5.05	32.60	89.60	95.65
A	Not in service	14-Jun-11	7	-3.86	8.44	-151.12	15.47	3.33	7.69	4.51	24.88	94.24	100.00	66.67	-38.00	19.70	90.20	38.89
A	Not in service	21-Jun-11	14	-1.40	2.47	14.68	7.25	5.04	7.69	2.79	10.00	93.92	75.00	100.00	13.09	25.56	61.11	90.15
A	In service	28-Jun-11	21	5.55	-1.11	44.55	9.64	13.39	3.54	-11.32	9.09	23.08	100.00	100.00	45.95	27.59	100.00	78.29
A	In service	06-Jul-11	29	-0.29	3.61	55.92	13.97	2.34	10.45	8.31	-1066.70	6.98	0.00	20.00	48.65	52.78	66.67	95.10
A	In service	12-Jul-11	35	3.35	4.37	57.40	11.36	5.93	15.09	9.63	5.83	40.00	30.00	100.00	0.00	22.55	90.00	89.22
A	In service	19-Jul-11	42	0.96	2.25	69.21	9.07	2.00	9.00	7.25	-1.22	0.00	-300.00	33.33	-42.86	67.21	87.34	92.19
A	In service	26-Jul-11	49	2.87	1.41	56.21	16.93	17.05	11.86	-6.25	82.89	99.70	-200.00	-33.33	-200.00	-323.16	33.93	92.67
A	In service	24-Aug-11	78	-4.32	3.09	10.64	17.19	-0.68	10.00	10.84	1.46	99.15	-150.00	87.50	-216.67	-312.12	-142.11	59.81
A	Chlorine added	20-Sep-11	105	4.49	-2.16	30.95	5.87	-0.52	2.70	3.21	-1.35	-200.00	100.00	50.00	39.53	-103.47	-30.16	46.41
A	Chlorine added	18-Oct-11	133	30.08	10.11	53.51	13.39	59.68	18.41	-95.16	56.46	97.63	-500.00	100.00	-20.00	-389.52	-133.47	55.31
B	In service	19-Apr-11	158	23.20	0.25	76.34	-0.92	-50.12	6.55	23.73	40.51	19.74	-150.00	0.00	98.31	95.43	-93.72	99.15
B	In service	03-May-11	172	10.22	1.37	89.77	99.90	-41.29	10.50	-6.37	80.92	28.57	-37.50	100.00	-16.56	89.91	-67.23	100.00
B	Not in service	10-May-11	7	-2.38	3.25	38.35	3.70	-3.03	15.25	17.74	-206.67	-24.04	-215.79	0.00	-566.67	14.68	9.40	-9.58
B	Not in service	17-May-11	14	3.90	12.93	26.03	-7.90	0.20	16.89	16.72	-86.42	-155.32	-34.04	0.00	-53.77	-14.68	26.15	83.33
B	Not in service	31-May-11	28	2.70	-5.63	58.72	5.75	3.29	7.83	4.70	5.49	66.82	0.00	0.00	8.34	42.90	77.40	79.25
B	In service	07-Jun-11	35	3.55	2.57	65.82	8.63	6.45	15.00	9.14	0.08	96.81	-100.00	100.00	70.00	-1810.00	-181.00	88.90
B	In service	14-Jun-11	42	7.30	7.20	50.63	7.89	32.96	9.91	-34.39	37.67	78.60	-100.00	22.22	0.00	48.40	52.50	96.12
B	In service	21-Jun-11	49	3.85	-1.72	82.43	4.20	4.97	10.19	5.50	4.76	71.96	-900.00	12.50	67.64	96.26	-100.00	73.74
B	In service	28-Jun-11	56	9.65	0.55	63.96	8.23	2.51	10.96	8.80	5.53	75.90	100.00	75.00	-170.00	85.83	100.00	97.62
B	In service	06-Jul-11	64	-0.30	1.13	52.40	6.74	3.31	9.09	5.98	71.93	36.51	56.52	-300.00	-160.00	35.82	0.00	99.80
B	In service	12-Jul-11	70	4.21	1.94	73.38	0.50	7.37	11.74	4.71	0.96	28.79	100.00	0.00	-300.00	55.05	86.67	97.35
B	In service	19-Jul-11	77	2.98	0.00	77.03	2.69	8.00	16.36	9.09	15.95	95.14	0.00	0.00	11.36	-313.83	90.11	99.16
B	In service	26-Jul-11	84	4.14	0.00	75.12	9.37	33.80	11.16	-34.21	21.43	42.86	33.33	-33.33	-420.00	-66.00	38.80	98.39
B	In service	24-Aug-11	113	3.63	-2.34	58.62	0.63	-10.20	10.71	17.23	-2.94	96.41	25.00	50.00	-366.67	6.14	-16.95	91.25
B	Chlorine added	20-Sep-11	140	1.37	5.99	86.14	3.72	0.34	5.13	4.80	-33.94	94.82	0.00	100.00	64.52	24.18	64.47	98.43
B	Chlorine added	18-Oct-11	168	-2.82	2.90	24.33	3.93	-162.93	5.17	63.93	-1126.00	93.98	0.00	0.00	-18.75	85.38	-25.42	50.00

Table 4.6: Significant correlations of age against percentage removal of water quality parameters (red text denotes positive and blue negative correlations), based on 470 samples

Water Quality Parameter	Correlation	P-value
Ammonium	-0.283	3.89×10^{-10}
Coliforms	0.537	1.80×10^{-4}
Dissolved Organic Carbon	-0.285	3.11×10^{-10}
Nitrate	-0.606	2.20×10^{-16}
Nitrite	-0.171	1.90×10^{-4}
Performance Metric ($\bigtriangledown$)	0.475	4.00×10^{-3}
Phosphate	-0.411	2.20×10^{-16}
Total viable bacteria (TVB) at 30°C	0.243	1.90×10^{-3}
TVB at 13°C	0.117	1.09×10^{-3}

rity. Statistical analysis of seasonality using the Mann-Kendall test showed only a significant seasonal effect for temperature and turbidity, suggesting performance improvements are not due to seasonal effects. This is not surprising as temperature is known to follow a natural seasonal trend in Scotland. Likewise, turbidity is known to fluctuate with rainfall levels, as higher rainfall leads to increased agricultural runoff and thus more turbid influent water supplying the filters. Therefore these results suggest that the improvement in filter performance is due to maturity in the microbial community and not seasonal variability.

4.3.3 Clone Library

To determine the most abundant phyla and bacterial classes found within SSFs and to create taxon-specific qPCR primers to quantify temporal concentration changes, a 16S rRNA gene clone library was constructed. The SSF clone library was composed of 100 clones that grouped into 35 phylotypes, representing 10 different phyla; results of the sequencing can be seen in (Table 4.7). From Figure 4.9 it can be seen that the bacterial phylotypes can be divided into ten different phyla: *Proteobacteria* (56% of the total clones), *Planctomycetes* (16%), *Bacteroidetes* (8%), *Actinobacteria* (6%), *Acidobacteria* (4%), *Nitrospira* (4%), *Chloroflexi* (1%), *Gemmatmonadetes* (1%), *Verrucomicrobia* (1%) and Unclassified at 95% cutoff with RDP classifier (3%). *Proteobacteria* were the dominant phylum, with *Betaproteobacteria* and *Deltaproteobacteria* representing 45% of the clone library, with the dominant phylotype (20% of the clone library) being affiliated with the *Deltaproteobacteria*.

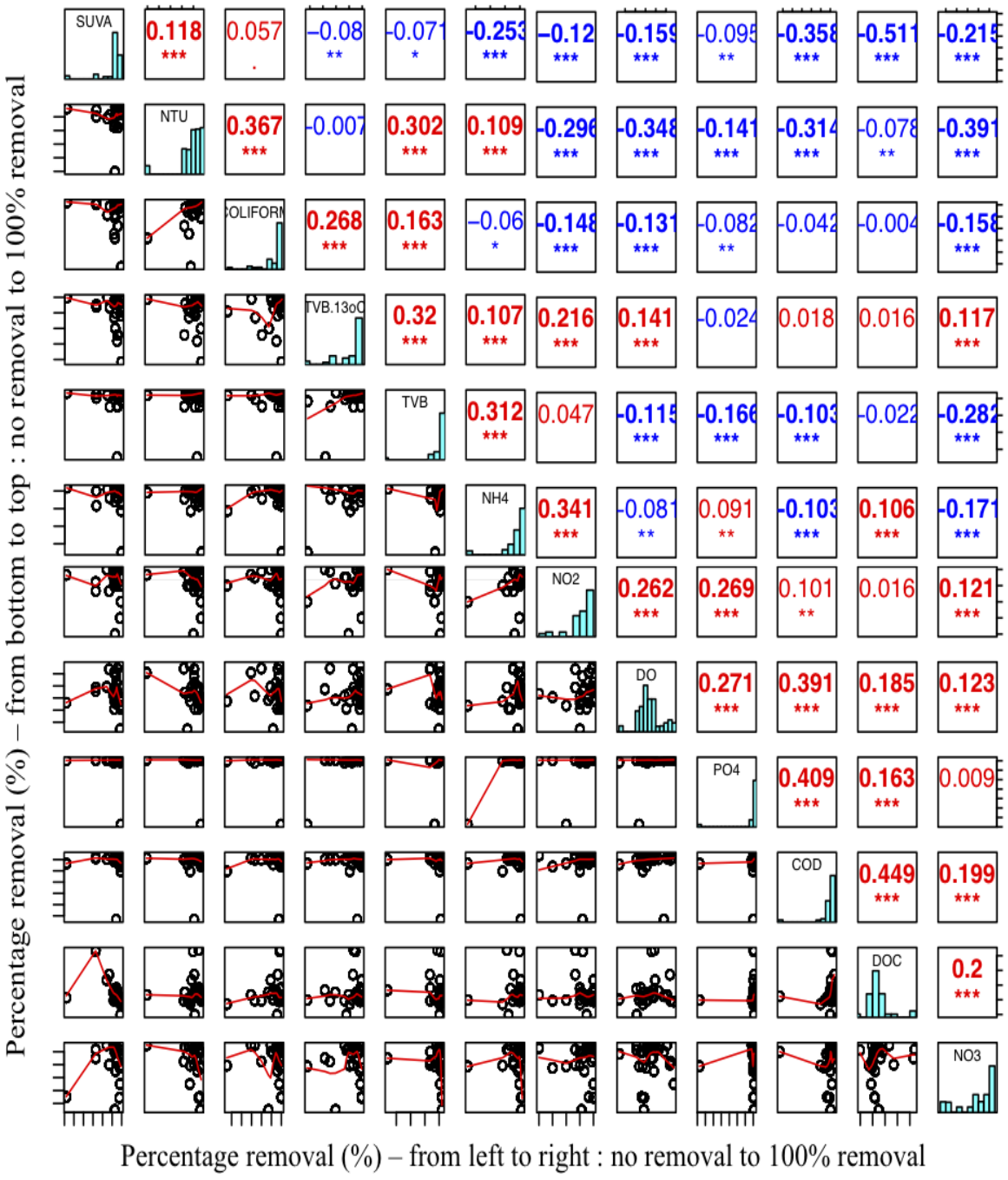


*Figure 4.8: Kendall correlation tests showing the relationships between the percentage removal of the various water quality parameters. Red indicate positive correlations and blue negative correlations. P-values are denoted by . = 0.5 , * = 0.1, ** = 0.001, *** = 0. Temperature, pH and UV were left out of the analysis.*

Table 4.7: Results of SSF clone library sequencing, (NCBI results only given if match greater than 96%)

OTU	Frequency	NCBI Match	RDP Phylum (% match)	RDP Class (% match)	RDP Order (% match)	RDP Family (% match)	RDP Genus (% match)
1	1%	AY922039.1 -Uncultured Gemmatimonadetes (97%)	Gemmatimonadetes (100)	Gemmatimonadetes (100)	Gemmatimonadales (100)	Gemmatimonadaceae (100)	Gemmatimonas (100)
2	1%		Proteobacteria (71)	Deltaproteobacteria (39)	Myxococcales (17)	Cystobacteraceae (17)	Anaeromyxobacter (16)
3	3%	GQ500790.1 -Uncultured bacterium (97)	Proteobacteria (100)	Betaproteobacteria (100)	Methylophilales (100)	Methylophilaceae (100)	Methylobacillus (40)
4	1%		Acidobacteria (100)	Acidobacteria Gp5 (100)			Gp5 (100)
5	4%		Planctomycetes (100)	Planctomycetacia (100)	Planctomycetales (100)	Planctomycetaceae (100)	Pirellula (90)
6	4%	GQ500737.1 -Uncultured bacterium (97)	Proteobacteria (100)	Alphaproteobacteria (100)	Rhizobiales (94)	Rhodobiaceae (55)	Parvibaculum (28)
7	1%		Proteobacteria (100)	Deltaproteobacteria (100)	Myxococcales (100)	Polyangiaceae (87)	Byssovorax (49)
8	1%		Proteobacteria (100)	Deltaproteobacteria (99)	Desulfuromonadales (55)	Geobacteraceae (53)	Geothermobacter (33)
9	1%		Proteobacteria (100)	Betaproteobacteria (100)	Methylophilales (100)	Methylophilaceae (100)	Methylobacillus (29)
10	1%		Planctomycetes (100)	Planctomycetacia (100)	Planctomycetales (100)	Planctomycetaceae (100)	Planctomyces (63)
11	2%	DQ066983.1 -Uncultured bacterium (97)	Bacteroidetes (39)	Flavobacteria (35)	Flavobacteriales (35)	Cryomorphaceae (25)	Fluviicola (20)
12	1%	AB630824.1 -Uncultured bacterium (98)	Chloroflexi (100)	Caldilineae (100)	Caldilineales (100)	Caldilineaceae (100)	Caldilinea (100)
13	4%	JN038608.1 -Uncultured Nitrospira sp. (99)	Nitrospira (100)	Nitrospira (100)	Nitrospirales (100)	Nitrospiraceae (100)	Nitrospira (100)
14	5%	JF139715.1 -Uncultured bacterium (98)	Actinobacteria (100)	Actinobacteria (100)	Actinomycetales (100)	Sporichthyaceae (42)	Sporichthya (42)
15	2%		Proteobacteria (100)	Betaproteobacteria (100)	Burkholderiales (77)	Oxalobacteraceae (10)	Oxalicibacterium (10)
16	1%		Actinobacteria (100)	Actinobacteria (100)		Acidimicrobidae incertae sedis (93)	Ilumatobacter (93)
17	20%		Proteobacteria (94)	Deltaproteobacteria (93)	Syntrophobacterales (24)	Syntrophaceae (19)	Smithella (16)
18	1%	HQ864095.1 -Uncultured bacterium (97)	Bacteroidetes (100)	Sphingobacteria (100)	Sphingobacteriales (100)	Chitinophagaceae (100)	Terrimonas (98)
19	2%	GQ127745.1 -Uncultured bacterium (97)	Acidobacteria (100)	Acidobacteria Gp4 (100)		Gp4 (100)	
20	1%		Proteobacteria (62)	Deltaproteobacteria (57)	Desulfobacterales (34)	Desulfobacteraceae (14)	Desulfatiferula (9)
21	3%		Proteobacteria (100)	Alphaproteobacteria (100)	Rhizobiales (100)	Hyphomicrobiaceae (100)	Hyphomicrobium (78)
22	1%		Verrucomicrobia (100)	Verrucomicrobiae (100)	Verrucomicrobiales (100)	Verrucomicrobiaceae (100)	Verrucomicrobium (97)
23	1%	GQ339129.1 -Uncultured bacterium (98)	Proteobacteria (100)	Gammaproteobacteria (100)	Xanthomonadales (96)	Sinobacteraceae (96)	Steroidobacter (96)
24	2%		Bacteroidetes (100)	Sphingobacteria (45)	Sphingobacteriales (45)	Sphingobacteriaceae (41)	Solitalea (38)
25	1%	GQ500774.1 -Uncultured bacterium (98)	Proteobacteria (100)	Alphaproteobacteria (100)	Rhizobiales (97)	Hyphomicrobiaceae (35)	Prosthecomicrobium (28)
26	15%	GQ302535.1 -Uncultured beta proteobacterium (98)	Proteobacteria (100)	Betaproteobacteria (100)	Burkholderiales (56)	Burkholderiales incertae sedis (34)	Thiobacter (34)
27	1%	AB240511.1 -Uncultured bacterium (98)	Bacteroidetes (100)	Sphingobacteria (97)	Sphingobacteriales (97)	Cytophagaceae (71)	Sporocytophaga (62)
28	4%	EF020203.1 -Uncultured planctomycete (99)	Planctomycetes (100)	Planctomycetacia (100)	Planctomycetales (100)	Planctomycetaceae (100)	Blastopirellula (64)
29	7%	FN984830.1 -Uncultured bacterium (99)	Planctomycetes (100)	Planctomycetacia (100)	Planctomycetales (100)	Planctomycetaceae (100)	Pirellula (85)
30	2%	GQ397080.1 -Uncultured bacterium (97)	Bacteroidetes (100)	Sphingobacteria (100)	Sphingobacteriales (100)	Chitinophagaceae (100)	Terrimonas (100)
31	2%		Chlamydiae (36)	Chlamydiae (36)	Chlamydiales (36)	Parachlamydiaceae (36)	Parachlamydia (31)
32	1%	AM935240.1 -Uncultured beta proteobacterium (98)	Proteobacteria (100)	Betaproteobacteria (100)	Rhodocyclales (73)	Rhodocyclaceae (73)	Azovibrio (73)
33	2%	GQ127745.1 -Uncultured bacterium (98)	Acidobacteria (100)	Acidobacteria Gp4 (100)		Gp4 (100)	
34	1%	FM200950.1 -Uncultured bacterium (98)	Proteobacteria (64)	Epsilonproteobacteria (20)	Campylobacterales (18)	Helicobacteraceae (18)	Wolinella (18)

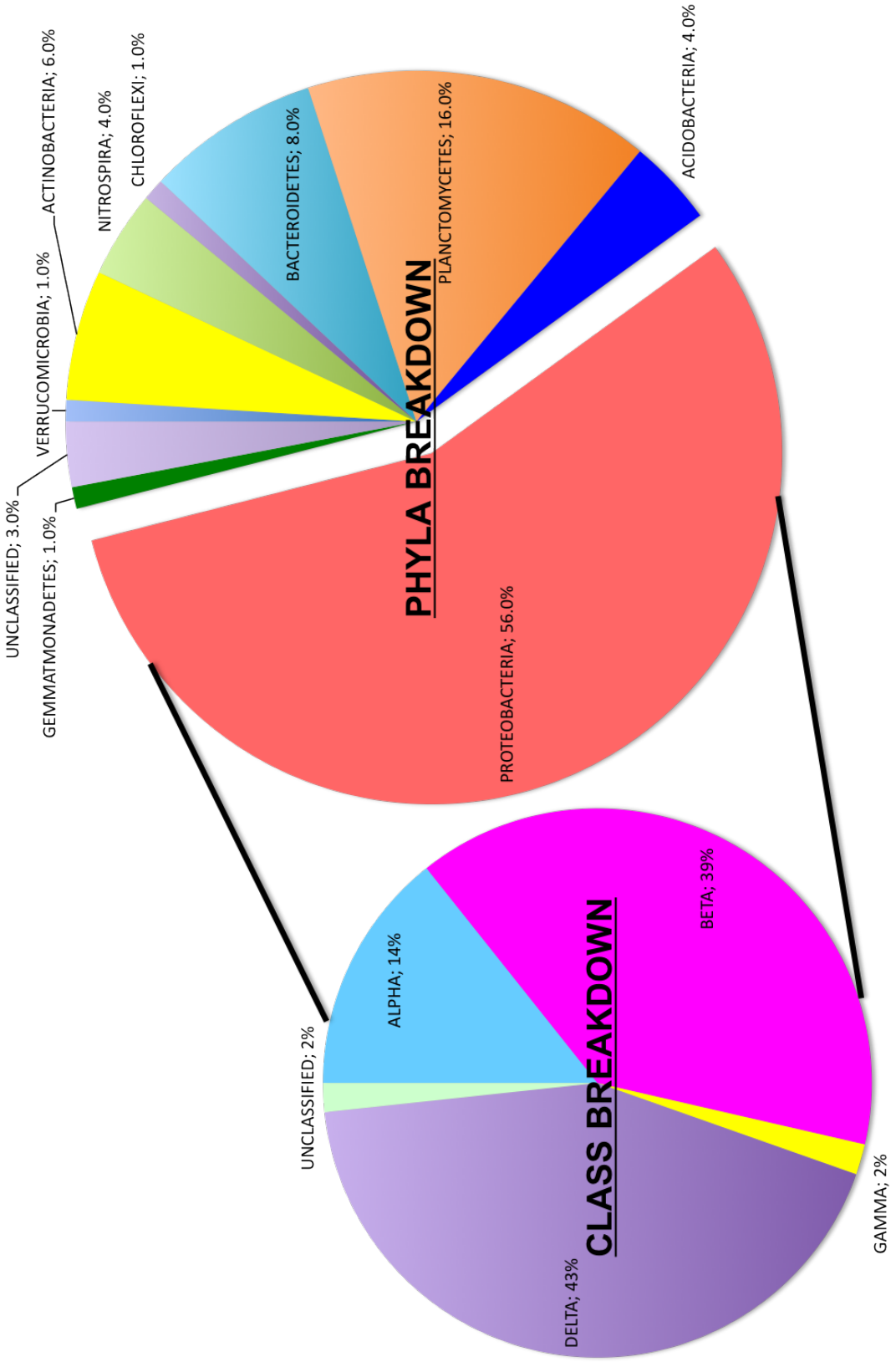


Figure 4.9: Pie charts of the proportion of each phylum in the 16S rRNA SSF clone library, smaller pie chart on lefthand side is the proportion of each class of Proteobacteria

Since there are known biases associated with DNA extraction and PCR amplification, the abundances of a phylotype on a clone library does not necessarily reflect its true abundance in the sample and hence in the environment. However, this SSF clone library does have similarities to other soil and sediment clone libraries; in particular there is great similarity between this SSF clone library and an Arctic perennial Spring sediment clone library from Canada [Perreault et al., 2007] and a German drinking water reservoir [Röske et al., 2012], with respect to the dominance of *Proteobacteria*, specifically *Deltaproteobacteria*.

4.3.4 Distinct Microbial Community Composition Between Samples from Sand, Influent and Effluent

A total of 26,163,232 sequences were generated from Illumina sequencing with an average number of 38,566 reads for each sample (Figure 4.10). To account for differences in read number, and therefore diversity, samples were rarefied to the lowest read number within the dataset (5,909). Rarefied samples were classified below domain level, being affiliated to: 36 phyla, 82 classes, 126 orders, 239 families, 688 genera and 11,026 OTUs. *Proteobacteria* were the dominant phylum in all samples as shown in Figure 4.11, accounting for, on average 51% of the community, which is consistent with the clone library results. Overall, sand from operational SSFs contained the greatest number of OTUs (8,319, of which 2,312 were unique to sand), which was almost double that found in drained SSF sand (4,482, with 3 unique OTUs) and both the influent (4,504) and effluent (3,947) samples. This coincided with significant differences in species diversity and evenness; with operational SSF sand having greater species diversity and evenness than drained sand samples (Wilcoxon test p-value: 0.0021 and 0.0004, respectively).

Influent water samples possess more OTUs than effluent samples, alongside a significantly higher species diversity index (p-value = 0.001); however, there was no difference between species evenness values (0.65 and 0.63, respectively). Interestingly, sand samples from operational filters only shared 55% and 73% of their OTUs with influent and effluent water samples, respectively. This highlights that other environmental factors such as wildlife and weather, in addition to the supply water, are important in shaping the microbial community in SSFs. It should, however, be noted that the OTU numbers reported above are likely to be over-estimates of the true diversity, given that an over-estimation of OTU number in dupli-

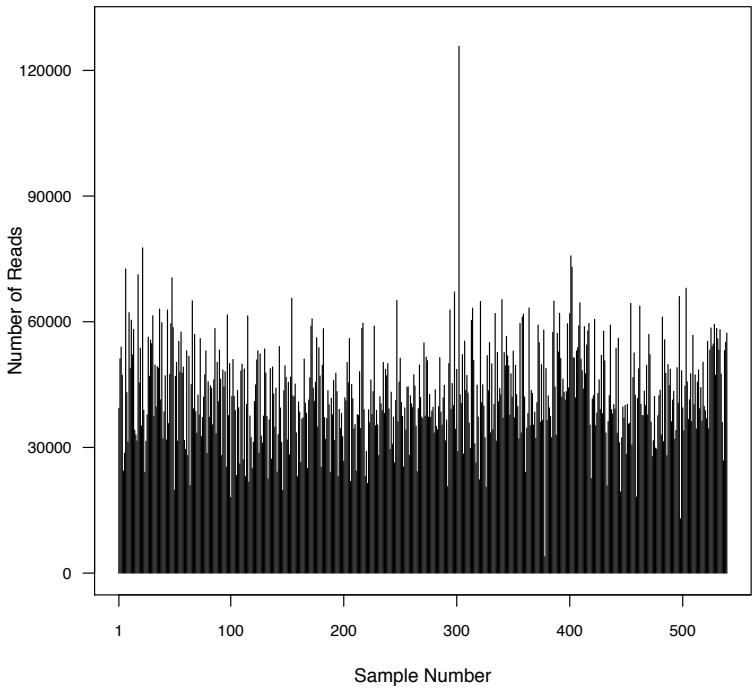


Figure 4.10: Number of reads associated with each sample sequenced by Illumina HiSeq 2000.

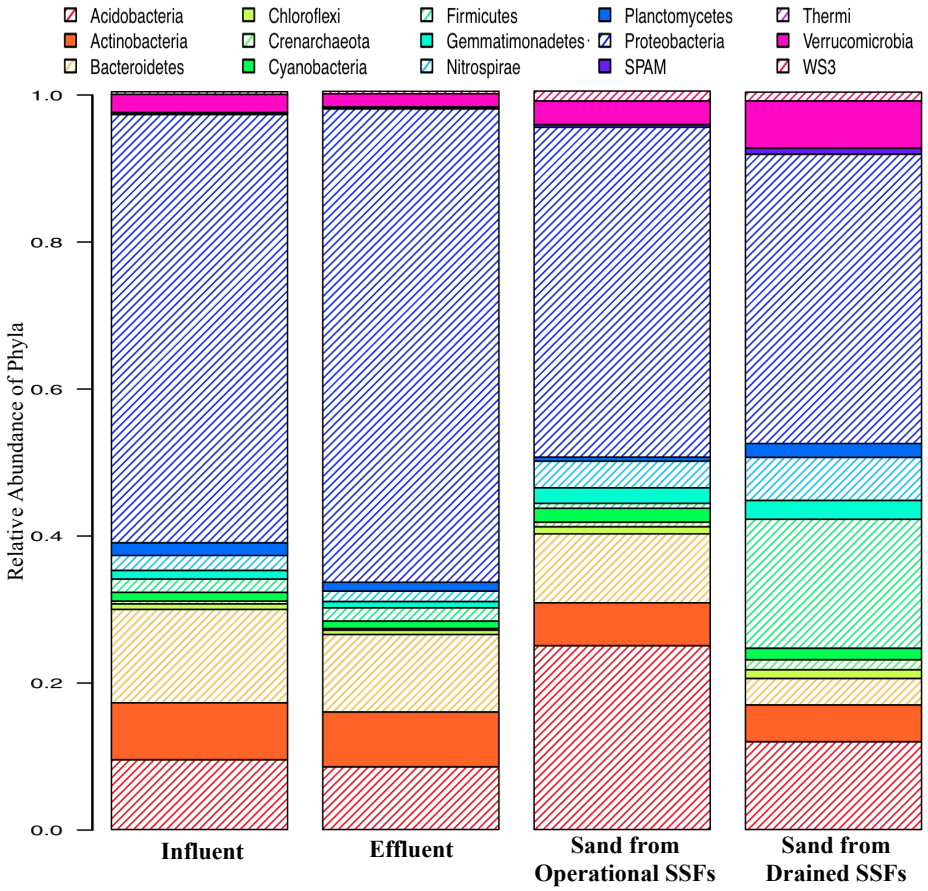


Figure 4.11: Average relative abundance of the top 15 phyla based on Illumina 16S rRNA amplicon sequencing in: influent, effluent, operational SSF sand and drained SSF sand.

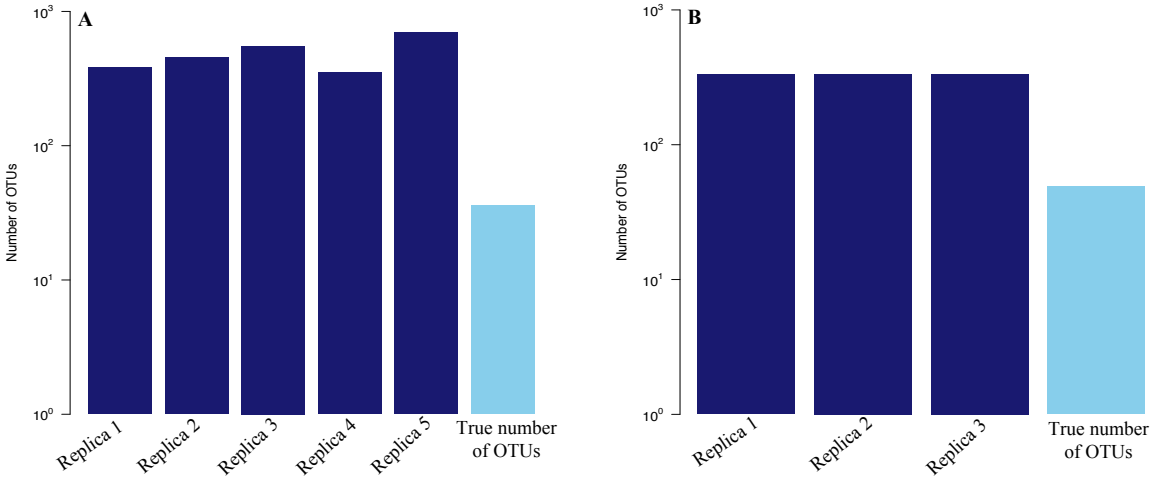


Figure 4.12: Barplots showing the actual number of OTUs in light blue and the sequenced number of OTUs in dark blue for two mock communities. A: A slow sand filter mock community created from the clone library mentioned in Section 4.3.3. B: The mock community created by Shakya et al. [2013].

cate analysis of positive control mock communities was observed (Figure 4.12). Based on comparisons to the known OTU number of these two mock communities it can be concluded that this Illumina sequencing over-estimates the true OTU number by on average 6.80-13.11 times.

NMDS analysis [Oksanen et al., 2012] (Figure 4.13) revealed that all samples clustered into four groups: influent water, effluent water, sand from operational filters and drained sand from two SSFs. Adonis analysis confirmed that the four groups were significantly different (p-value = 0.001). The SIMPER procedure was used to identify the top ten OTUs responsible for the dissimilarities between water and sand samples (Table 4.8). Referring to Figure 4.13 and Table 4.8 the extreme difference in community composition of drained SSFs compared to operational SSFs is apparent. In particular, *Bacillaceae*, members of the *Firmicutes* phylum appear to be responsible for the greatest proportion of the difference, being more than 16,000 times more abundant in drained samples than in operational SSFs.

4.3.5 Spatial and Temporal Community Diversity in Sand Samples

The spatial and temporal patterns of the absolute number of *Acidobacteria*, *Actinobacteria* α*-Proteobacteria*, *Bacteroidetes*, β*-Proteobacteria*, *Chloroflexi*, δ*-Proteobacteria*, γ*-Proteobacteria*, *Gemmatimonadetes*, *Nitrospira*, *Planctomycetes*, *Verrucomicrobia* and the

Table 4.8: SIMPER analysis identifies top ten taxa (at 97% OTU level) that account for the most of the dissimilarities between the water (influent and effluent) and sand (sand from functional and drained SSF) SSF samples

Taxon	Genus/Family	Contribution (%)	Avg.Sand (%)	Avg.Drained (%)
Firmicutes	*Bacillus*	5.84	<0.001	10.27
Proteobacteria	*Massilia*	2.41	4.12	0.42
Firmicutes	*Lysinibacillus*	1.85	<0.001	3.31
Nitrospirae	*Nitrospira*	1.79	3	1.21
Proteobacteria	*Oxalobacteraceae*	1.71	3.06	<0.001
Firmicutes	*Solibacillus*	1.53	<0.001	2.66
Proteobacteria	*Acinetobacter*	1.31	<0.001	2.59
Actinobacteria	*Arthrobacter*	1.3	<0.001	2.29
Proteobacteria	*Klebsiella*	1.17	<0.001	2.07
Bacteroidetes	*Flavobacterium*	1.05	1.88	<0.001

Taxon	Genus/Family	Contribution (%)	Avg.Influent (%)	Avg.Effluent (%)
Proteobacteria	*Pseudomonas*	6.04	5.59	5.98
Proteobacteria	*Massilia*	2.61	1.93	3.35
Bacteroidetes	*Flexibacteraceae*	2.25	3.22	1.28
Proteobacteria	*Nevskia*	2.11	0.31	3.19
Proteobacteria	*Polynucleobacter*	1.73	2.56	1.88
Proteobacteria	*Oxalobacteraceae*	1.68	0.59	2.71
Proteobacteria	*Pseudomonas*	1.6	0.25	2.4
Proteobacteria	*Oxalobacteraceae*	1.56	0.38	2.53
Proteobacteria	*Syntrophobacteraceae*	1.52	2.39	0.17
Proteobacteria	*Comamonadaceae*	1.37	1.97	1.25

Taxon	Genus/Family	Contribution (%)	Avg.Sand (%)	Avg.Influent (%)
Proteobacteria	*Pseudomonas*	3.76	0.04	5.59
Proteobacteria	*Massilia*	3.39	4.12	1.93
Proteobacteria	*Oxalobacteraceae*	2.33	3.06	3.22
Bacteroidetes	*Flexibacteraceae*	2.16	<0.001	3.22
Proteobacteria	*Syntrophobacteraceae*	1.9	0.79	2.39
Nitrospirae	*Nitrospira*	1.71	3	1.32
Proteobacteria	*Polynucleobacter*	1.68	<0.001	2.56
Proteobacteria	*Comamonadaceae*	1.32	<0.001	1.97
Bacteroidetes	*Flavobacterium*	1.24	1.88	0.29
Actinobacteria	*Arthrobacter*	1.14	1.76	0.18

Taxon	Family	Contribution (%)	Avg.Sand (%)	Avg.Effluent (%)
Proteobacteria	*Pseudomonas*	3.56	0.04	5.98
Proteobacteria	*Massilia*	3.51	4.12	3.35
Proteobacteria	*Oxalobacteraceae*	1.99	3.06	1.11
Proteobacteria	*Nevskia*	1.91	0.02	3.19
Nitrospirae	*Nitrospira*	1.64	3	0.74
Proteobacteria	*Oxalobacteraceae*	1.6	0.07	2.71
Proteobacteria	*Pseudomonas*	1.57	0.35	2.4
Proteobacteria	*Oxalobacteraceae*	1.49	0.05	2.53
Bacteroidetes	*Flavobacterium*	1.17	1.88	0.5
Proteobacteria	*Polynucleobacter*	1.09	0.09	1.88

total bacterial community were studied using qPCR in the two ISSFs, which were divided into three sections based on the side (location in filter i.e., beside the influent or effluent pipe) of the filters (Figure 4.14). As expected, the highest abundance was observed for the total bacterial community, with densities ranging from 2.13 x 10^7 to 2.21 x 10^{10} 16S rRNA genes/g of extracted sand. A strong positive correlation was found between total 16S rRNA genes and age of the filters (0.43, p-value: 0.01) and a mild negative correlation between the total 16S rRNA genes and depth (-0.165, p-value: 0.001), which is consistent with other studies [Koizumi et al., 2003]. Overall, clear changes can be seen in the absolute phyla abundances over time and depth (Figure 4.14), with *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Planctomycetes*, *δ-Proteobacteria* and *Verrucomicrobia* accounting for 85% of the population on average, with the abundance of *Actinobacteria*, *Bacteroidetes*, *Proteobacteria* and *Verrucomicrobia* fluctuating most over time.

To statistically determine the parameters shaping the microbial community in SSF sand, two ISSFs were sampled at various times of year; ages; locations; depths and distances from both the influent and effluent pipes; and the data were analysed using canonical correspondence analysis (CCA). A combination of these factors (all factors in Table 4.9 except filter) ex-

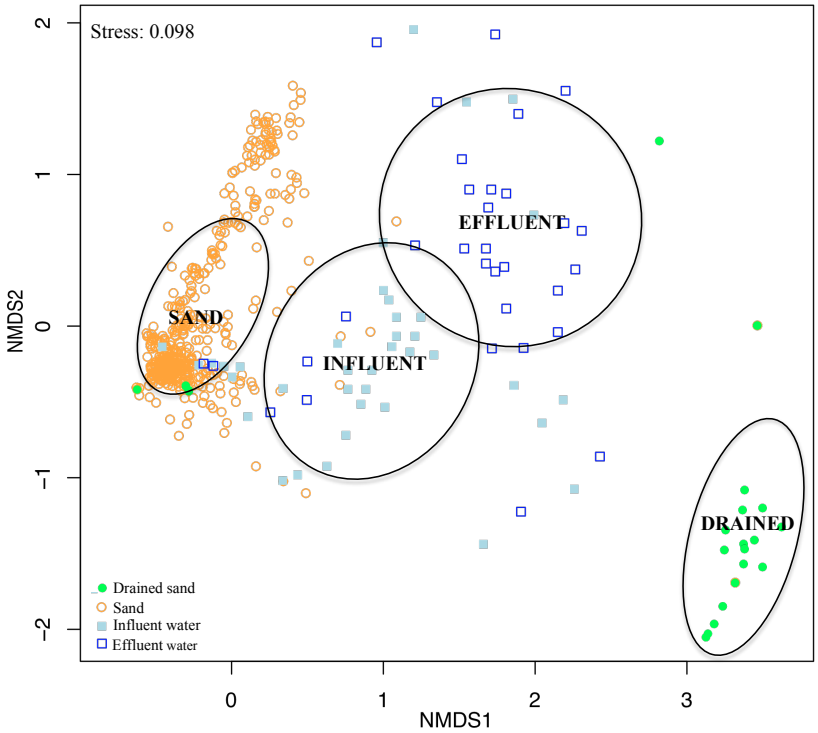


Figure 4.13: NMDS ordination for the microbial community structure for all samples in the Illumina dataset at the 97% OTU level. Ellipses designate the 95% confidence intervals for the four groups.

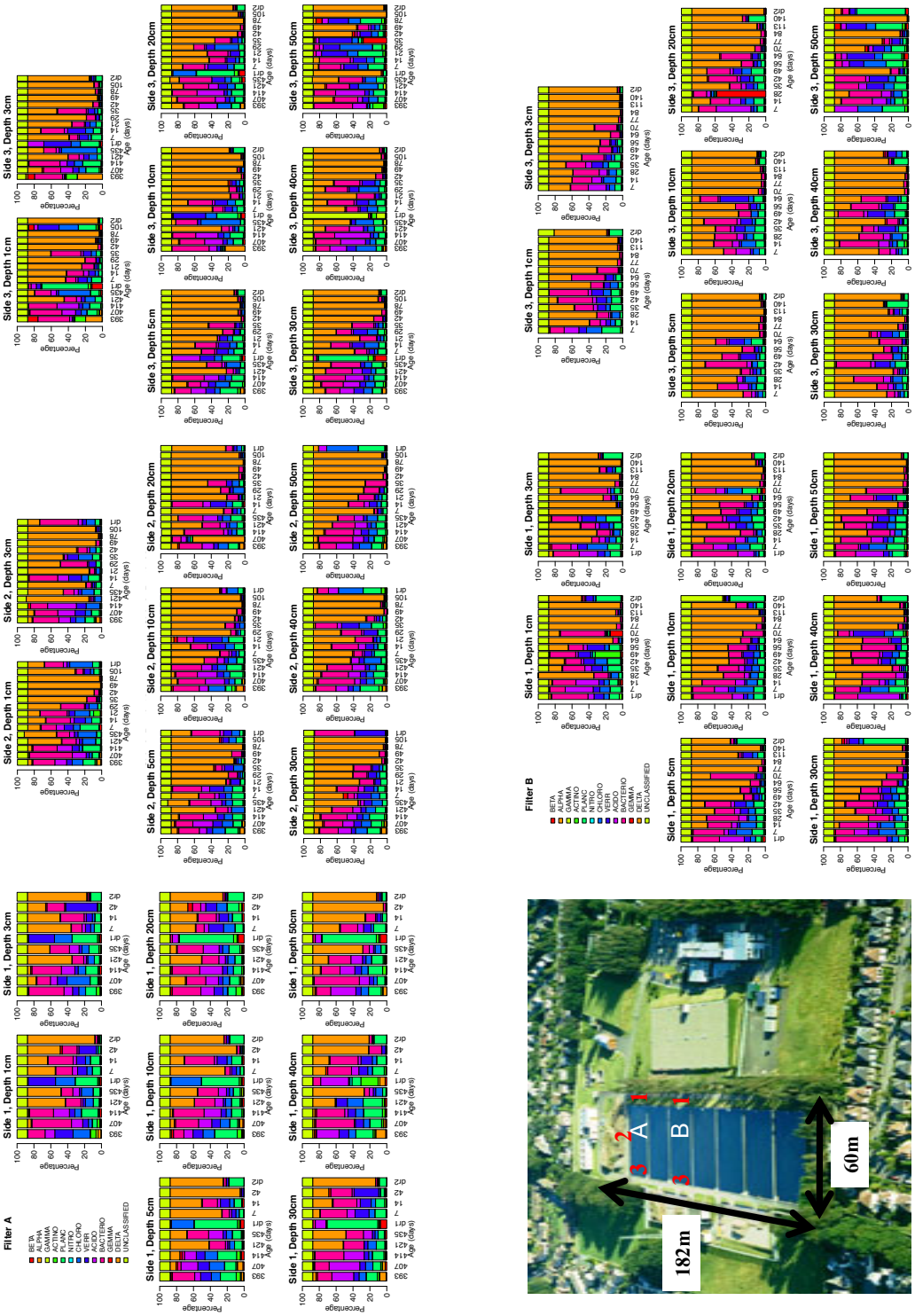


Figure 4.14: Stacked barplots depicting percentage abundance of each phylum (determined by qPCR) at different depths and location of filter A (top) and B (bottom) over time

Table 4.9: Canonical correspondence analysis of the relative abundances of bacterial OTUs and filter parameters and characteristics in 406 sand samples from two SSFs

Parameter	Degrees of Freedom	χ^2	F value	Number of permutations	Pr(>F)
Month	3	0.1575	6.4583	99	0.01
Age category	2	0.1359	10.4997	99	0.02
Side	1	0.0691	8.5006	99	0.01
Distance from Effluent Pipe	1	0.0246	3.0248	99	0.04
Distance from Influent Pipe	1	0.0234	2.8756	99	0.01
Age	1	0.0577	7.0844	99	0.01
Depth	1	0.0131	1.6080	99	0.05
Filter	1	0.0093	1.1394	99	0.64
Residual	190	1.5352			

plained the SSF sand microbial community structure (p-value: 0.005). Of the characteristics evaluated, age (both individual ages and age categories: early, middle and late); the side of the filter sampled; and the month at which a sample was taken; were the major drivers for the bacterial community structure, with age being the most significant factor (Figure 4.15). Subtle differences at specific age categories (early:0-4weeks, mid: 5-8weeks and late:$\geq$9 weeks) with respect to the abundance of the top 18 families can be seen in Figure 4.16, in particular differences between: *Flavobacteriaceae*, *Micrococcaceae*, *Nitrospiraceae* and *Oxalobacteraceae*. Further analysis revealed that there is a strong positive correlation in the total number of OTUs and the total 16S counts with the age of the filters (early: 4790, mid: 5234 and late: 6798 OTUs), corresponding to average 16S counts of 5.62×10^7, 2.14×10^8 and 4.39×10^8 16S/g of sand). Therefore, it can be concluded that older filters possessed a greater number and density of OTUs than earlier stages, which is consistent with previous studies [Ramond et al., 2013]. SIMPER analysis corroborated that significant differences in community composition at the various age categories was mainly due to members in the: *Flavobacteriaceae*, *Micrococcaceae*, *Nitrospiraceae* and *Oxalobacteraceae* families. However, *Ruminococcaceae*, a less abundant family, was likewise found to explain a significant amount of the community variation, with percentage abundances of this family being significantly higher in older filters (early abundance: $<$0.001%, mid: 0.0013% and late: 0.268%). Further, there were nine families (Table 4.10), ranging in relative abundance from $<$0.0001-0.027%, which were only present in the oldest filters.

Surprisingly, depth was only a marginally significant parameter (p-value: 0.05) in explaining differences between sand samples. This is unusual as chemical gradients are hypothesised

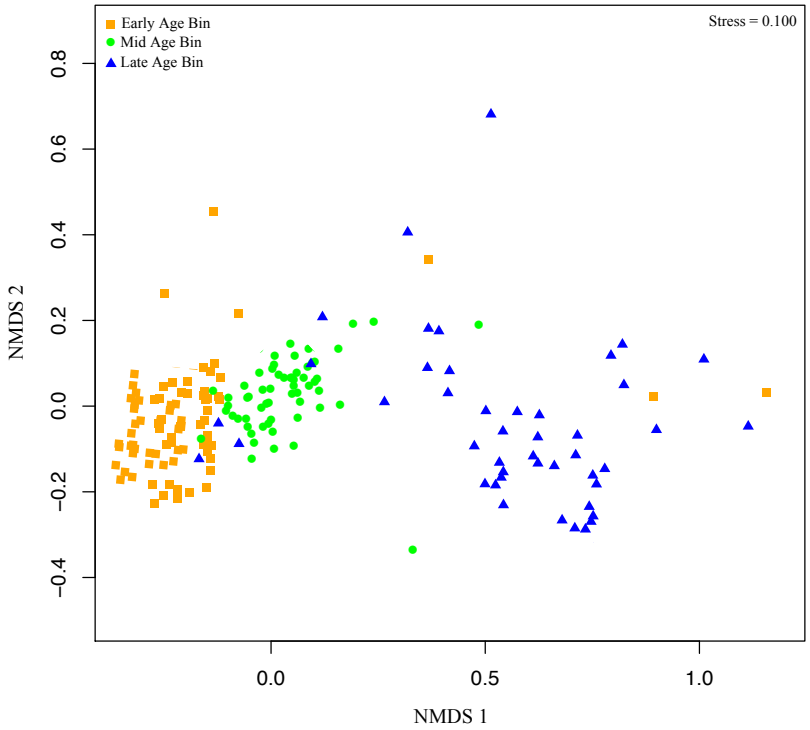


Figure 4.15: Non-metric multidimensional scaling plot of the bacterial OTU community structures in 406 sand samples taken from two SSFs. Samples are coloured according to their age bin category: early = 0-4 weeks, mid = 5-8 weeks and late = >9weeks after scraping

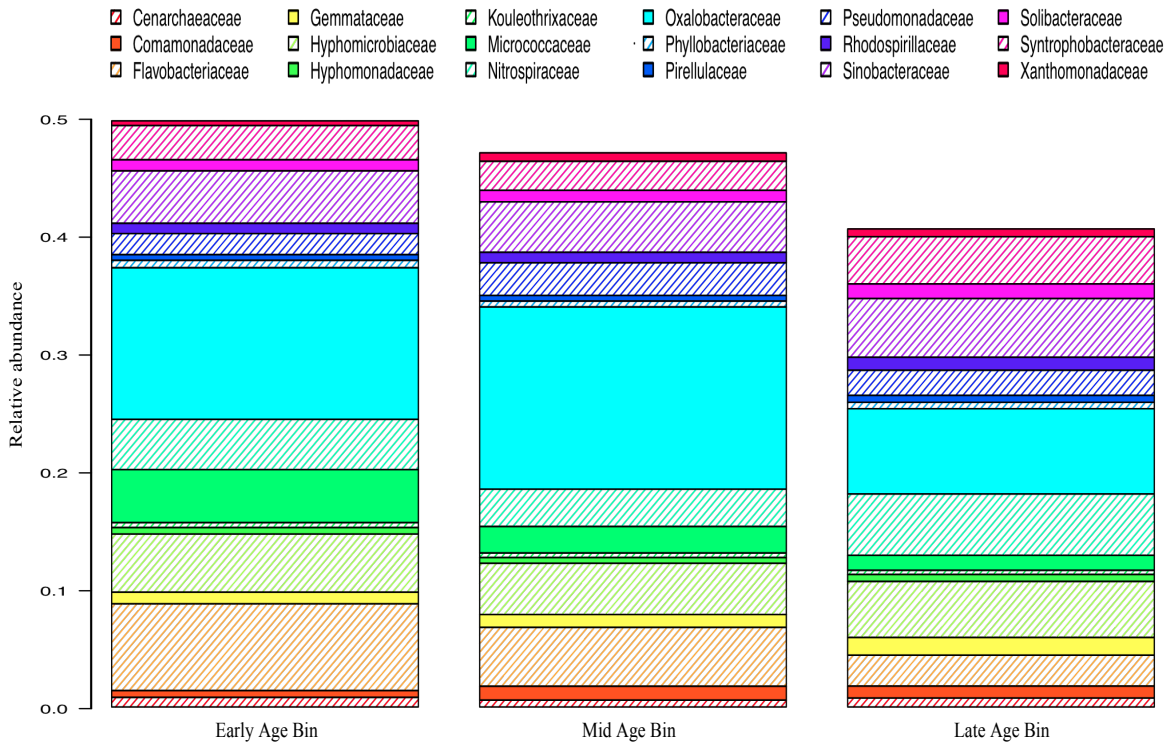


Figure 4.16: Average relative abundance of the top 18 families in SSFs at different age bins (early: 0-4 weeks, mid: 5-8 weeks and late: >10weeks after scraping), data generated from Illumina amplicon sequencing

Table 4.10: Average percentage relative abundance of the nine families only found in late age-binned SSFs. ∗ denotes a member of the archae

Family	Average Relative Abundance (%)
Alteromonadaceae	2.01E-05
Beutenbergiaceae	9.71E-06
Elusimicrobiaceae	1.15E-03
Listeriaceae	5.66E-05
*Methanocorpusculaceae**	1.61E-03
Mycoplasmataceae	4.32E-05
Rikenellaceae	2.71E-02
vadinHA31	3.35E-04
Victivallaceae	2.83E-04

to exist with SSFs and, therefore, community composition was expected to be explained in a large part by depth. However, laterally (side of the filter sampled), there were significant differences in the microbial community. SIMPER analysis revealed that side 1 was most similar to side 2 (47% similarity) but less similar to side 3 (41% similarity), and that side 2 was 43% similar to side 3. The majority of the differences between the microbial community composition at the different sides was due to *Acidobacteria* and various orders of *Proteobacteria* (Figure 4.17). Furthermore, Adonis analysis showed that the microbial community in both filters A and B were statistically indiscriminate (p-value = 0.093). Therefore, with respect to reproducibility and homogeneity between microbial communities, the filters were indistinguishable.

4.3.5.1 The Impact of draining and Chlorination on the Microbial Community

Species evenness indices revealed that as the filters matured the microbial community becomes more even, with evenness being consistent irrespective of the depth in the sand bed. However, during the decomissioning of the site, chlorine was added to the filters and, as in the study by Wang et al. [2013], chlorination was shown to significantly lower species evenness (Filter A before = 0.558 ± 0.090 and after chlorination = 0.502 ± 0.077, Filter B before = 0.556 ± 0.070 and after chlorination = 0.448 ± 0.090). This effect was also seen during the time when filters were drained and scraped (Figure 4.18), with a more profound effect on the top depths than lower depths. Referring more closely to Figure 4.14, a staggering difference in community composition can be seen at all depths when the filters are drained;

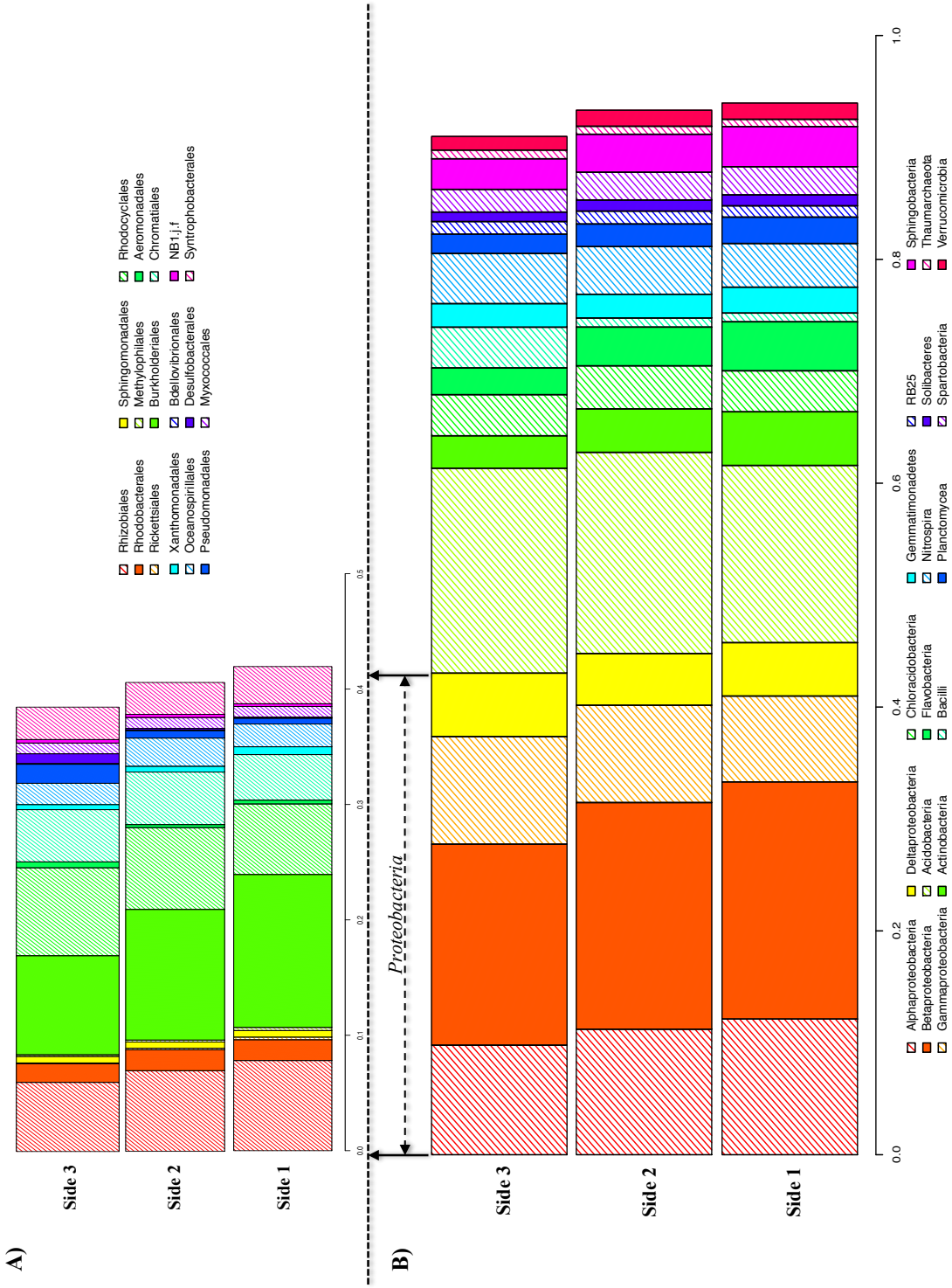


Figure 4.17: Stacked barplots depicting average percentage abundance of top 18 taxon: (A) classes, (B)orders of Proteobacteria at the three different sides of SSFs

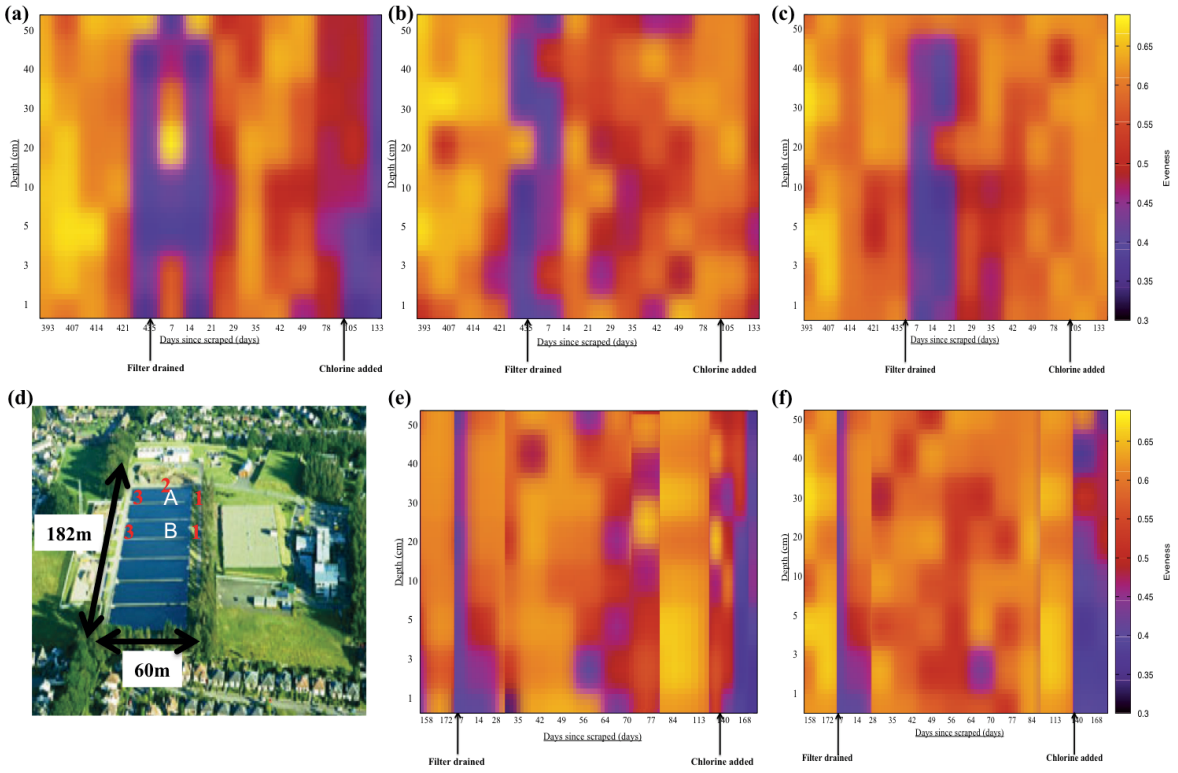


Figure 4.18: Heatmaps showing the temporal and spatial changes in species evenness in SSF A and B, ***(a)*** *Filter A Length 1* ***(b)*** *Filter A Length 2,****(c)*** *Filter A Length 3,* ***(d)*** *Aerial picture of filter showing length and filter locations,* ***(e)*** *Filter B Length 1 and* ***(f)*** *Filter B Length 3.*

specifically, a large increase was observed in the proportion of *Chloroflexi*, *Planctomycetes* and *Verrucomicrobia*, which coincided with a decrease in *Acidobacteria*, *Bacteroidetes* and *Deltaproteobacteria*.

Diversity indices (shannon index) also showed significant reduction during these periods (Filter A before chlorination = 1.55 $\pm$ 0.13 and 1.27 $\pm$ 0.20 after chlorination, Filter B before = 1.50 $\pm$ 0.16 and 1.08 $\pm$ 0.15 after chlorination). This reduction in evenness and diversity is explained by the large change in the microbial composition of the filters, specifically the dominance of *Deltaproteobacteria*, which had increased from an average abundance of 24.87% to 63.50%. Although chlorination did impact species evenness and diversity it should be noted that the level of effect was different depending upon the side of the filter, with side 1 (side of the filter where the influent pipe is located) being the first and most severely affected. This is not surprising as it is the closest location to where chlorine delivery occurs.

4.3.6 Mesoscale Spatial Variation

Previously, the importance of the side and distance from the influent and effluent pipe on explaining differences in the microbial community has been discussed. However, in order to gain a better understanding of this spatial variation on a more microbially realistic scale, six cores at each side of both industrial SSFs were taken on the 21st of June 2011 (Figure 4.3C). Referring to the canonical correspondence analysis plot (Figure 4.19), the sand samples from the mesoscale experiment form three distinct clusters those: distance from the influent pipe, distance from effluent pipe and distance from effluent corner. The depth and distance from the influent pipe correlated with CA1 and explained 33.97% of the variation and the distance from the effluent pipe and effluent corner correlated with CA2 and explained 16.89%. Adonis analysis confirmed that there were significant differences in the microbial community within and between groups (between groups p-value = 0.009; within distance from the: influent pipe p-value = 0.034; effluent pipe p-value = 0.018; effluent corner p-value = 0.053). Such differences in community can be attributed to chemical gradients which likely exist within the filters. SIMPER analysis revealed that the abundance of the *Massilia* genus increased with distance from the influent pipe (average abundance at 0.2m = 1.36%, 0.42m = 14.44% and 0.68m = 23.10%). *Massilia* sp. have been isolated from various environmental samples from many sources, including air, dust, soil, roots and drinking water [Gallego et al., 2006], however, the reason for their dominance away from the influent pipe is unclear. Interestingly, total 16S numbers were highest at the middle length (0.42m) and lowest at the furthest length (0.68m), which coincided with the largest and smallest number of OTUs (3275 and 2898 respectively). However, in direct contrast the total number of 16S rRNA genes was greatest beside the effluent pipe (albeit half the number found beside the influent pipe).

4.3.7 Correlation Between Community Members and Water Quality

Stepwise multivariate regression showed that the relative proportion of several bacterial families correlated strongly with the removal of certain water quality parameters (Table 4.11). These correlations are consistent with the findings of other studies. Additional multivariate regression analysis showed that the water quality performance of SSFs significantly correlated with both the age and species evenness of the filters (Age p-value: 0.022, Evenness p-value: 1.620×10^{-4}), with species evenness being more important than age (Figure 4.20).

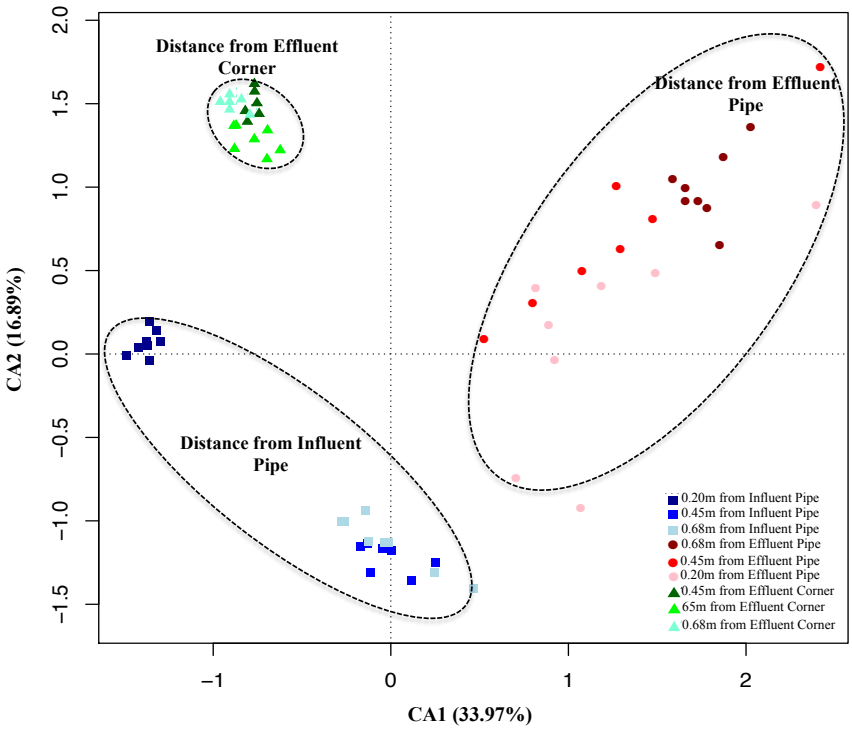


Figure 4.19: Canonical Correspondence Analysis (CCA) of mesoscale spatial variability in the microbial community in sand samples taken from several locations from the influent and effluent pipes supplying two ISSFs.

The importance of species evenness in relation to better performance has been previously documented [Werner et al., 2011]. Wittebolle et al. [2009] explained that higher species evenness implies greater robustness and functional stability and therefore a greater ability to adapt to new and fluctuating parameters.

Significant differences (p-value: 0.01) in the community composition at different performance levels (poor, average, good and excellent) were determined by Adonis analysis, with the major organisms that contribute to these dissimilarities being determined through SIMPER analysis (Table 4.12). In particular, it can be seen that for excellent performance, an evenly distributed community is required with no overly abundant organisms. On further analysis it is apparent that poor performance is due to an uneven community structure. In particular, an over-abundance of *Acidovorax* and *Sphingobium*, and an under-abundance of *Halomonas* and *Sphingomonas* (Figure 4.21), as well as the complete absence of *Naxibacter*, *Streptopyta* and *Acinetobacter* compared with good and excellent performing filters.

Table 4.11: Stepwise multivariate regression of water quality parameters and family abundances.

Parameter	Model P-value	Adjusted R^2	Family	P-value	Relationship with Removal
Ammonium	1.562×10^{-5}	0.4133	*CL500.29*	0.0399	+
			Cellulomonadaceae	0.0784	-
			Mycobacteriaceae	0.0464	-
			Nocardiaceae	0.0635	-
			Carnobacteriaceae	0.0040	+
			Rhizobiaceae	0.0041	+
			Leuconostocaceae	0.0094	-
			Pseudomonadaceae	0.0286	-
Coliform	1.837×10^{-6}	0.5265	*Erysipelotrichaceae*	0.0877	-
			Carnobacteriaceae	0.0653	-
			Fusobacteriaceae	0.0418	-
			Isosphaeraceae	8.23×10^{-5}	+
			Planctomycetaceae	0.0207	-
			Desulfobacteraceae	0.0507	-
			Sinobacteraceae.1	0.0251	+
			Opitutaceae	0.0005	+
			Verrucomicrobia subdivision3	0.0011	-
			Enterobacteriaceae	0.0435	-
Dissolved Organic Carbon	2.2×10^{-16}	0.8583	*Nocardiaceae*	0.0140	+
			Promicromonosporaceae	3.24×10^{-7}	-
			Propionibacteriaceae	6.59×10^{-9}	-
			Bifidobacteriaceae	3.00×10^{-7}	+
			Solirubrobacteraceae	0.0431	+
			Alicyclobacillaceae	0.0029	+
			Pasteuriaceae	0.0126	-
			Carnobacteriaceae	0.0045	-
			Leuconostocaceae	7.80×10^{-5}	-
			Sphingomonadaceae	1.44×10^{-8}	+
			Rhodocyclaceae	0.0066	-
Nitrate	1.469×10^{-8}	0.5524	*Brevibacteriaceae*	1.96×10^{-5}	-
			Dermacoccaceae	0.0030	-
			FW	0.0002	-
			Rhodobiaceae	0.1071	+
			Mycoplasmataceae	0.0006	-
Nitrite	0.008712	0.1952	*Thermodesulfovibrionaceae*	0.0699	-
			Planctomycetaceae	0.0503	+
			Hyphomicrobiaceae	0.0298	-
			Phyllobacteriaceae	0.0193	+
			Rhodobacteraceae	0.0291	-
			Xanthobacteraceae	0.0534	+
Performance (▽)	1.726×10^{-9}	0.6219	*Holophagaceae*	0.0171	+
			CL500.29	0.0004	+
			Kineosporiaceae	0.0034	-
			Micrococcaceae	5.78×10^{-8}	-
			Fusobacteriaceae	0.0001	-
			Rhodobiaceae	0.0197	-
			Shewanellaceae	0.0001	-
			Sphingomonadaceae	0.0875	+
pH	0.000205	0.307	*Dietziaceae*	0.0143	-
			Microbacteriaceae	0.0646	+
			Micrococcaceae	0.0002	+
			Saprospiraceae	0.0448	-
			Moraxellaceae	0.0122	-
Phosphate	9.567×10^{-6}	0.3917	*Flavobacteriaceae*	1.21×10^{-5}	-
			Sphingobacteriaceae	0.0079	+
			Alicyclobacillaceae	0.0717	+
			Carnobacteriaceae	0.0025	+
			Leuconostocaceae	0.0009	-
Turbidity	2.014×10^{-11}	0.6599	*Actinomycetaceae*	6.55×10^{-6}	+
			Fusobacteriaceae	1.55×10^{-11}	-
			Isosphaeraceae	0.0020	+
			Bradyrhizobiaceae	0.0007	-
			Shewanellaceae	0.0076	-
			Peptococcaceae	0.0831	+
Total Viable Bacteria	2.2×10^{-16}	0.8407	*Catenulisporaceae*	1.88×10^{-9}	-
			Rivulariaceae	$2. \times 10^{-16}$	-
			Thermodesulfovibrionaceae	0.0014	+
			Gemmataceae	0.0020	-
			Pirellulaceae	0.0067	+
			Procabacteriaceae	0.0331	-

Table 4.12: SIMPER analysis of the top 15 taxa accounting for majority of the dissimilarities between SSFs producing different levels of water quality.

Taxon	Genus	Contribution (%)	Avg. Poor (%)	Avg. Excellent (%)
Alphaproteobacteria	*Sphingobium*	12.61	14.96	2.97
Gammaproteobacteria	*Pseudomonas*	4.55	5.72	3.52
Betaproteobacteria	*Acidovorax*	10.27	12.23	4.37
Bacteroidetes	*Flavobacterium*	6.64	6.51	6.07
Alphaproteobacteria	*Methylobacterium*	2.37	1.81	1.14
Gammaproteobacteria	*Halomonas*	5.72	1.62	6.34
Alphaproteobacteria	*Sphingomonas*	3.11	1.89	2.93
Betaproteobacteria	*Naxibacter*	3.63	3.18	2.55
Betaproteobacteria	*Massilia*	1.44	0.7	1.4
Betaproteobacteria	*Polynucleobacter*	0.005	1.88	0.84
Alphaproteobacteria	*Novosphingobium*	2.17	1.98	1.53
Bacteroidetes	*Arcicella*	1.4	0.7	1.02
Gammaproteobacteria	*Nevskia*	0.002	0.13	0.36
Streptophyta	*Streptophyta*	0.19	1.11	0.11
Gammaproteobacteria	*Acinetobacter*	0.86	0.89	0.24

Taxon	Genus	Contribution (%)	Avg. Average (%)	Avg. Excellent (%)
Alphaproteobacteria	*Sphingobium*	17.16	24.33	38.73
Gammaproteobacteria	*Pseudomonas*	12.61	14.96	2.97
Betaproteobacteria	*Acidovorax*	4.55	5.72	3.52
Bacteroidetes	*Flavobacterium*	10.27	12.23	4.37
Alphaproteobacteria	*Methylobacterium*	6.64	6.51	6.07
Gammaproteobacteria	*Halomonas*	2.37	1.81	1.14
Alphaproteobacteria	*Sphingomonas*	5.72	1.62	6.34
Betaproteobacteria	*Naxibacter*	3.11	1.89	2.93
Betaproteobacteria	*Massilia*	3.63	3.18	2.55
Betaproteobacteria	*Polynucleobacter*	1.44	0.7	1.4
Alphaproteobacteria	*Novosphingobium*	0.005	1.88	0.84
Bacteroidetes	*Arcicella*	2.17	1.98	1.53
Gammaproteobacteria	*Nevskia*	1.4	0.7	1.02
Streptophyta	*Streptophyta*	0.002	0.13	0.36
Gammaproteobacteria	*Acinetobacter*	0.19	1.11	0.11

Taxon	Genus	Contribution (%)	Avg. Good (%)	Avg. Excellent (%)
Alphaproteobacteria	*Sphingobium*	4.14	3.11	2.97
Gammaproteobacteria	*Pseudomonas*	6.41	8.68	3.52
Betaproteobacteria	*Acidovorax*	5.5	5.65	4.37
Bacteroidetes	*Flavobacterium*	6.13	5.51	6.07
Alphaproteobacteria	*Methylobacterium*	3.82	4.31	1.14
Gammaproteobacteria	*Halomonas*	5.64	2.75	6.34
Alphaproteobacteria	*Sphingomonas*	4.57	2.83	2.93
Betaproteobacteria	*Naxibacter*	2.82	2.21	2.55
Betaproteobacteria	*Massilia*	2.92	2.85	1.4
Betaproteobacteria	*Polynucleobacter*	1.33	1.26	0.84
Alphaproteobacteria	*Novosphingobium*	1.78	1.12	1.53
Bacteroidetes	*Arcicella*	2.04	1.44	1.02
Gammaproteobacteria	*Nevskia*	1.35	1.71	0.36
Streptophyta	*Streptophyta*	0.85	0.98	0.11
Gammaproteobacteria	*Acinetobacter*	1.32	1.19	0.24

Performance designated by using $\bigtriangledown$: 0-4:Poor, 5-6:Average, 7-8:Good and 9-10:excellent.

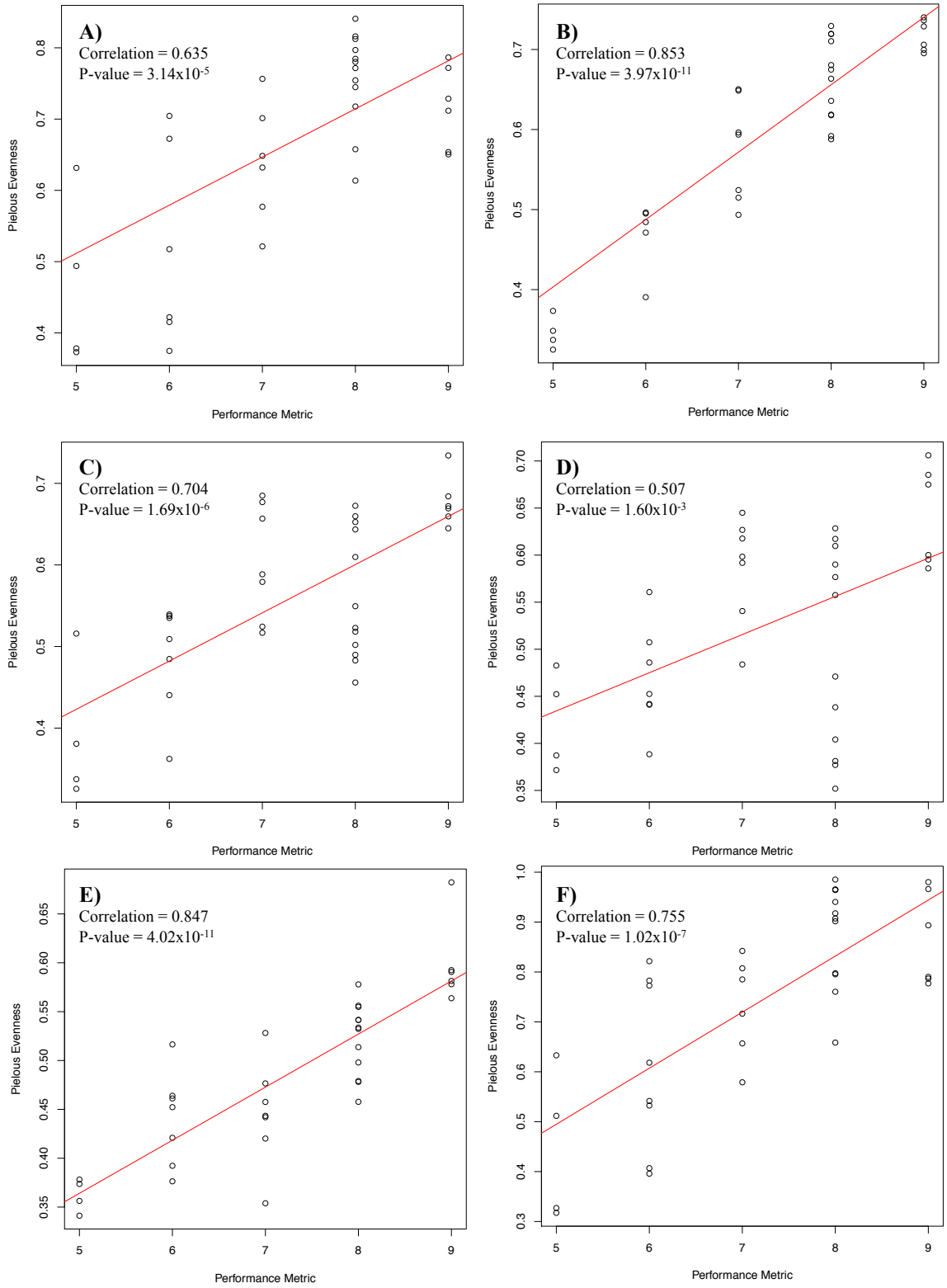


Figure 4.20: Scatter plot showing the correlation between sand filter performance ($\bigtriangledown$) and species evenness at different levels of classification. (A) Phyla, (B) Class, (C) Order, (D) Family, (E) Genus and (F) OTU. Higher $\bigtriangledown$ corresponds to better water quality performance. Based on Illumina sequencing of 618 sand samples.

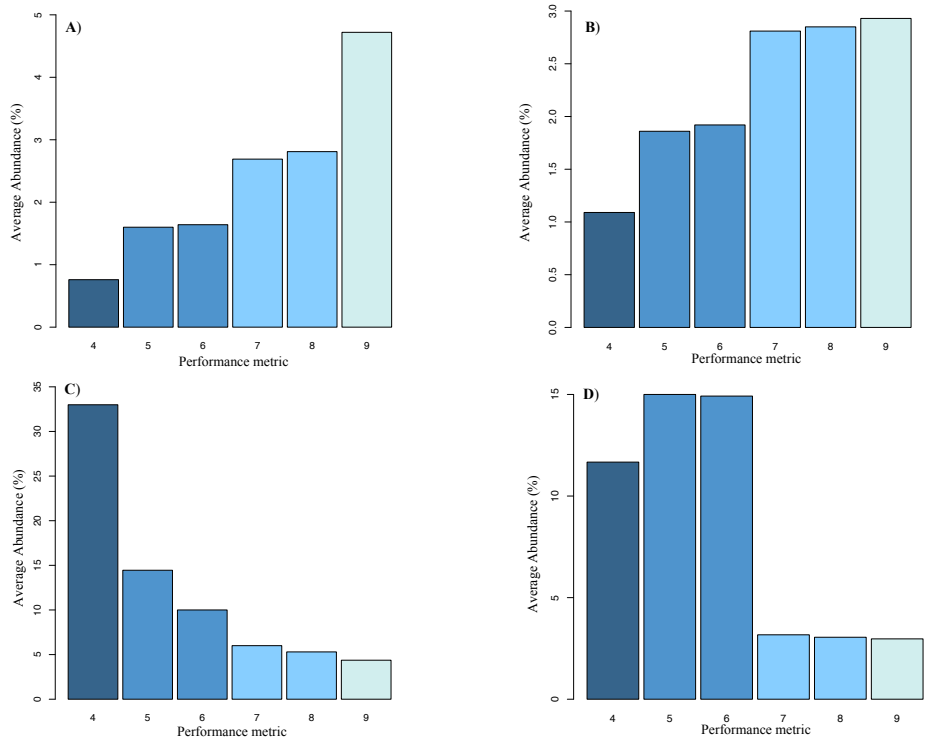


Figure 4.21: Barplots of the average percentage abundance of 4 key genera: A) Halomonas, B) Sphingomonas, C) Acidovorax and D) Sphingobium, at different levels of water quality performance ($\bigtriangledown$). Note: different y-axis scale. Based on Illumina sequencing of 618 sand samples.

4.4 Discussion

4.4.1 Slow Sand Filters Host Diverse Bacterial Communities

The first major work into the characterisation of the microbial ecology of SSFs began several decades ago [Brink, 1967, Lloyd, 1974]. In these studies, the diversity in the bacterial community of these filters was deemed to be very low. However, the work of Brink and Lloyd was based on using conventional plating and isolation techniques, which are known to under-estimate the true diversity. Since this initial work, several studies have been published [Eighmy et al., 1992, Bahgat et al., 1999, Calvo-Bado et al., 2003, Wakelin et al., 2011, Ramond et al., 2013], which have begun to utilise modern molecular methods to answer the same questions of Brink and Lloyd. These studies have all found that SSF communities are extremely diverse both metabolically and phylogenetically [Eighmy et al., 1992]. However, all of these studies have been carried out in SSFs used to purify wastewater or storm water, rather than drinking water, as in this study, or on only samples from the schmutzdecke and not various depths, as in this study.

In this thesis, the microbial diversity of two ISSFs was found to be far larger than previously reported with 36 phyla and 239 families found, compared to the 21 phyla and 149 families found by Wakelin et al. [2011] in an Australian SSF. Such differences in diversity may be explained by the different methodological approaches and primers used, as well as different water sources, and perhaps more significantly, the different depths sampled within the SSF (Wakelin et al. [2011] used storm water and only sampled the schmutzdecke). However, like Petry-Hansen et al. [2006] and Wakelin et al. [2011], *Proteobacteria* were found to be the dominant phylum in this study. Additionally, the microbial community in the influent and effluent water was found to be significantly different to the sand community, both in terms of composition and density. This highlights that other environmental factors such as wildlife and weather are important in shaping the microbial community found in SSFs and should be considered for future studies.

4.4.2 Reproducibility of Filter Performance and Microbial Community

The microbial community composition of the SSFs were significantly different depending upon the status (operational or drained), age, location, month, distance from the influent and effluent pipe, and depth from which samples were taken (Table 4.9). This was a previously uknown. The age of the filter was the most significant parameter in explaining both changes in the microbial community and water quality variables. This is not surprising as it is widely documented by operators that SSF performance improves with maturity [Huisman et al., 1974]. Additionally, the increase in the abundance of *Ruminococcaceae*, and the presence of the nine other families (Table 4.10) in older filters might be explained by the fact that they are all either facultative or strict anaerobes commonly found in wastewater and sewage [Whitman et al., 2012]. Their increased abundance is likely due to prolonged exposure to faeces from wildlife (i.e., birds) surrounding the filters and similar exposure at the reservoir feeding the filters.

Surprisingly, filter identity was not a significant factor in explaining the differences in the microbial community, suggesting that the communities within SSFs at this site are indistinguishable. This alongside the similarity in water quality production between the filters, implies that both the microbial community and performance are highly reproducible between filters at this site. Another unexpected finding was the marginal significance of depth in ex-

plaining differences in community composition. This is strange as in many other freshwater studies [Lin et al., 2012] depth has been shown to be extremely important, as chemical gradients are known to exist, causing changes in community composition. Although surprising, this is not the first study to find depth a marginally significant variable. Recently, Röske et al. [2012] showed that depth was an insignificant variable in explaining community composition in sediment from the Saidenbach drinking water reservoir in Germany. Regardless of this, future work should focus on determining whether such water chemistry gradients exist and if they affect or shape the microbial community of SSFs.

Although vertical (depth) spatial differences in the SSF microbial community were marginal, lateral (side and distance from influent and effluent pipe) differences were highly significant. Both can be a consequence of habitat heterogeneities imposed by differences in the physicochemical characteristics [Deschesne et al., 2007], such as partially filled or unfilled voids between sand grains that would disperse nutrients and microbes or the dilution of components away from the influent pipe, creating nutritional gradients. Perhaps such dispersal of nutrients occurs faster and easier along the surface of SSFs rather than vertically and thus may account for the higher significance.

4.4.3 Species Evenness is Critical to Performance

Stepwise multivariate regression showed that the water quality performance of SSFs significantly correlates with both the age and species evenness of the filters (Table 4.11), with better performing filters having higher evenness values. This is the first study to the author's knowledge to correlate bacterial species evenness to the differing levels of performance of water filters (Figure 4.20). Greater evenness has been linked to greater robustness and functional stability [Wittebolle et al., 2009], and therefore, the ability to adapt to new and fluctuating parameters, such as those brought by weather events (e.g., storms), which would impact the composition of the influent water feeding the filters. Therefore, the increased species evenness and richness found in excellently performing filters is additional confirmation of the "insurance hypothesis" conceived by Yachi and Loreau [1999], which hypothesises that both functional redundancy and evenness are necessary for functionally robust ecosystems.

The importance of species evenness is further emphasised by the dramatic effect seen during

draining events compared to operational times, in particular the over-abundance of *Firmicutes* and *Planctomycetes*. This dominance may be directly related to the fact that *Firmicutes* are known to produce endospores during periods of starvation or stress (i.e., during SSF draining periods when organisms in the sand are exposed to temperature, pH, oxygen, nutritional and UV fluctuations). The dominance of *Planctomycetes* can be explained by the increased exposure to sunlight (due to reduced depth of water), resulting in heightened algal growth, which has been shown to promote increased *Planctomycetes* growth [Pizzetti et al., 2011]. Likewise, a similar effect occurs during chlorination, with the decline in evenness being attributed to an increase in *Deltaproteobacteria*. *Deltaproteobacteria* are widely documented as being capable of reductive dechlorination, or halorespiration (the process of using halogenated compounds such as sodium hypochlorite, as terminal electron acceptors in anaerobic respiration) [Richardson, 2013]. Therefore, their dominance after chlorination is not wholly surprising.

NMDS analysis of variance revealed that at different levels of performance (excellent, average and poor, defined in Section 4.2.4), the composition of the microbial community is different. During periods of excellent performance there is a higher relative abundance of *Sphingomonas* and *Acinetobacter*, two genera known to be capable of biodegradation and metabolism of a wide range of chemicals (e.g., polycyclic aromatic hydrocarbons (PAHs), cyanotoxins, endocrine disruptors and herbicides) [Shi et al., 2001, Bending et al., 2003, Valeria et al., 2006, Fang et al., 2007, Williams and Ray, 2008], all of which would likely be present in the reservoir water feeding the SSFs. Moreover, Innerebner et al. [2011] recently showed that *Sphingomonas* has a striking plant-protective effect by suppressing disease symptoms and diminishing pathogen growth. Therefore, within SSFs this recent finding may help to explain why their abundance is higher in excellent performing filters, which typically have no, or low, pathogen counts. Likewise, the increased abundance of *Halomonas* in filters with excellent performance may be explained by the recent discovery that several members of this genus can produce bioflocculants with a power of $>80\%$ turbidity reduction [Cosa et al., 2013].

Conversely, in poorly performing filters the over-abundance of *Acidovorax* and *Sphingobium* may be explained by niche competition; both of these genera are known to be capable of

similar processes to *Sphingomonas* and *Acinetobacter*, which are found in greater abundance in filters achieving excellent performance. The antagonistic effects of such competition between members of the Sphingomonad family (*Sphingomonas*, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*) have been examined [Cunliffe and Kertesz, 2006] and shown to influence removal performance of PAHs in contaminated soils. However, it is important to note that it is impossible to determine whether differences in filter performance are due to the microbial community, or if it is the performance of the filters that shapes the community. Speculatively, the former seems most plausible as SSFs predominantly function via biological mechanisms. Additionally, referring to Table 4.3, little change in the characteristics of the influent water feeding the filters can be seen and, therefore, performance differences must be attributed to the microbial community. Overall, these findings show that higher species evenness is integral to excellent SSF performance, and for the first time, associate specific genera with differing levels of water quality production and ISSF performance.

4.5 Conclusions

In summary, the results of this study show that the microbial diversity of SSFs is far greater than previously documented and that, in terms of community composition and performance, the two SSFs sampled were indistinguishable and highly reproducible. Both filters produced high quality drinking water, with quality improving as the filters matured. The month, age, side, distance from the influent and effluent pipe and depth from which the samples were taken significantly impacted the microbial community in SSFs, with age being the most significant variable. As filters aged both the number and density of OTUs increased, as did species evenness. Further, Illumina sequencing indicated that the abundance of various members of the microbial community, specifically *Acidovorax*, *Halomonas*, *Sphingobium* and *Sphingomonas* were important for performance. More significantly, it was found that increased species evenness was critical for excellent filter performance. Decreased species evenness indices were found in drained and juvenile SSFs, coinciding with increased abundance of *Planctomycetes*, possibly induced by additional exposure to sunlight. Future work should investigate the impact of reducing the drainage period, or the effects of covering filters during draining and scraping events, on species evenness and the abundance of *Planctomycetes*. Such work could significantly reduce the downtime (the period of time SSFs are

non-operational due to poor performance) of SSFs.

Together the results of this study provide the most detailed characterisation of the functional microbial community found in SSFs and provide a framework for future ecological and physiological microbial research in these systems. To conclude, this study is the first to provide insight into the importance of specific taxa to performance. However, the extent of their importance, and other abiotic and biotic factors, requires additional field-based study as well as ecophysiological study under carefully controlled laboratory conditions.

Chapter 5

Mimicking Full-Scale Industrial SSFs in the Laboratory

"Happy families are all alike; every unhappy family is unhappy in its own way."

Leo Tolstoy, (Anna Karenina)

Previous laboratory-scale studies to characterise the functional microbial ecology of slow sand filters have suffered from methodological limitations that could compromise their relevance to full-scale systems. Therefore, to ascertain if laboratory-scale slow sand filters (LSSFs) can replicate the microbial community and water quality production of industrially operated full-scale slow sand filters (ISSFs), eight LSSFs were constructed and were used to treat water from the same source as the ISSFs. Half of the LSSF sand beds were composed of sterilised sand (sterile) from the industrial filters and the other half with sand taken directly from the same industrial filter (non-sterile). All filters were operated for 10 weeks, with the microbial community and water quality parameters sampled and analysed weekly. To characterise the microbial community phylum-specific qPCR assays and 454 pyrosequencing of 16S rRNA genes were used in conjunction with an array of statistical techniques. The results demonstrate that it is possible to mimic both water quality production and the structure of the microbial community of full-scale filters in the laboratory, allowing subsequent LSSF experimentation to be directly comparable to full-scale units. Further, it was found that the sand type composing the filter bed (non-sterile or sterile), the water quality produced, the age

A condensed version of this chapter is published: Haig, S. Quince, C. Davies, R.L. Dorea, C.C. and Collins, G. (2014). Validating the Performance and Microbial Community of Laboratory-Scale Slow Sand Filters with respect to Full-Scale Industrial Filters. ***Water Research*****, 61, 141-151**

of the filters and the depth of sand samples were all significant factors in explaining observed differences in the structure of the microbial consortia. This study is the first to the author's knowledge that demonstrates that scaled-down slow sand filters can accurately reproduce the water quality and microbial consortia of full-scale slow sand filters.

5.1 Introduction

As outlined in Chapter 3 and reviewed in Haig et al. [2011] there have been a number of studies that have attempted to characterise the purification mechanisms in SSFs and the microbes responsible for them, however, all previous studies have suffered from limitations of approach or the techniques available. Even recently, many of these investigations have been constrained by a focus on specific elements of the filter, such as the schmutzdecke, a biological layer formed on the top of the SSF bed [Campos et al., 2002, Rooklidge et al., 2005, Unger and Collins, 2008, Wakelin et al., 2011] or on specific biological processes, such as denitrification [Aslan and Cakici, 2007] and predation [Lloyd, 1996, Weber-Shirk and Dick, 1999, Lee and Oki, 2013]. A common drawback of these studies is their use of specific microcosms under carefully controlled conditions that may not fully represent the presumably complex and diverse microbial communities under-pinning real-world SSFs [Haig et al., 2011]. There is, therefore, a need to verify whether scaled-down, laboratory SSFs (LSSFs) can adequately represent full-scale industrially operated slow sand filters (ISSFs) with respect to microbial community structure and water quality performance. This type of verification will allow more complex questions pertaining to removal capabilities and the effects of abiotic and biotic parameters to be explored under controlled laboratory conditions, with the knowledge that conclusions drawn are directly comparable to full-scale filters. Further it will allow better understanding of the factors and mechanisms driving microbial community assembly, which will be of great benefit to the field of microbial ecology.

In this study, microbial diversity and water treatment performance were investigated in LSSFs to determine the feasibility of mimicking the microbial consortia and water quality production of ISSFs in the laboratory. We hypothesise that the age, scale and type of filter bed sand used (sterile - autoclaved, non-sterile directly taken from an operational ISSF or full-scale) are important factors in explaining community structure.

Hypotheses

Hypothesis 5.1 *Temporal and spatial changes in the microbial community structure of full-scale SSFs can be reproduced at the laboratory-scale.*

Hypothesis 5.2 *Laboratory-scale SSFs can replicate the performance of full-scale SSFs across the majority of the water quality metrics measured.*

Hypothesis 5.3 *Laboratory-scale SSFs containing biologically active filter media can replicate the performance and community structure of the full-scale SSFs from which they were seeded faster than laboratory-scale SSFs with a filter bed consisting of sterile filter media, despite being operated identically.*

5.2 Materials and Methods

In order to mimic a SSF in the laboratory it is important to keep the sand and water depths the same as found in industrial SSFs (ISSFs). This will allow us to accurately reproduce ISSFs conditions, with the aim of reproducing the microbial community. There have been few previous studies that have constructed laboratory-scale SSFs. Within this chapter LSSFs (Figure 5.1 and Figure 5.2) have been designed and constructed based on adapted versions of two previous models [Weber-Shirk and Dick, 1997a,b] and [Calvo-Bado et al., 2003]. This is a proof of concept study, which aims to determine if it is possible to replicate the microbial community and water quality production of ISSFs in LSSFs. In order to do this two sets of four replicate LSSF rigs were constructed, each supplied from the same untreated water source supplying the industrial filters discussed in Chapter 4. LSSFs were operated and sampled for ten weeks, with comparisons being made to ISSFs of the same age.

5.2.1 Design and Construction of Lab-scale SSFs

Filters were constructed using a flange-based design made with 2 x 1.25 m of 54 mm diameter polyvinyl chloride with a wall thickness of 5 mm (Figure 5.1). Filter set 1: ϵ, ζ, η and θ have 8 x 3, 11 mm diameter sampling ports down the length (0, 5, 10, 15, 20, 30, 45, 70 cm) and around the circumference of the sand bed (Figure 5.2), whereas Filter set 2: α, β, γ, and δ do not have these sampling chambers. This was to verify that the presence of sampling ports have no effect on the bacterial community and water quality production. Filters have

1m sand bed and 6.5 cm of under-drainage (0.85 - 1.4 mm gravel) to allow free drainage of filtered water from the columns. The water flow rate for all experiments was set at 0.15 $m^3m^{-2}h^{-1}$ (height of water column passing per hour) which is consistent with the ISSFs. Water was pumped from the storage tank which contained the untreated water supplying the industrial filters to the header tank (5 L) at a continuous rate. This storage tank was refilled weekly with the water supplying the industrial filters and was kept constantly aerated and homogenised. The sand used in all experiments was taken from the industrial sand filter site sampled in Chapter 4 and had an effective size of 0.38 mm and a uniformity coefficient of 2.78. In order to ascertain if microbial community assembly and water quality performance differed and could be reached quicker in the laboratory half of the columns were filled with sand from a freshly sampled operational industrial SSFs (non-sterile sand) and the remaining LSSFs were filled with sand which has be autoclaved at 121 °C for 20 minutes (sterile sand).

5.2.2 Sampling and Water Quality Testing

Each week, 2 L of influent and effluent water were collected from all the filters and tested for the same water parameters presented in Table 4.1 following the methods outlined in Appendix A. Additionally, sand samples were taken weekly from one of the sampling ports at each depth of filter set 1 and used for DNA extraction (Appendix B). Once extracted DNA was used for next-generation sequencing and qPCR (as performed in Section 4.2.6). In order to compare the effects of the presence of sampling ports on the microbial community sand samples from all filters were taken at the end of the experiments and final community composition compared.

5.2.3 qPCR

Performed as described in Section 4.2.6

5.2.4 454 Pyrosequencing

The V4-V5 region of the 16S rRNA gene (Figure 3.1) was PCR amplified and sequenced on the 454 GS FLX platform for 226 samples representing different depths, filters and filter ages. Note that ISSF samples matching the same ages as LSSFs were sequenced. PCR amplification was carried out in a Gene Pro thermal cycler (Bioer Technology, UK).

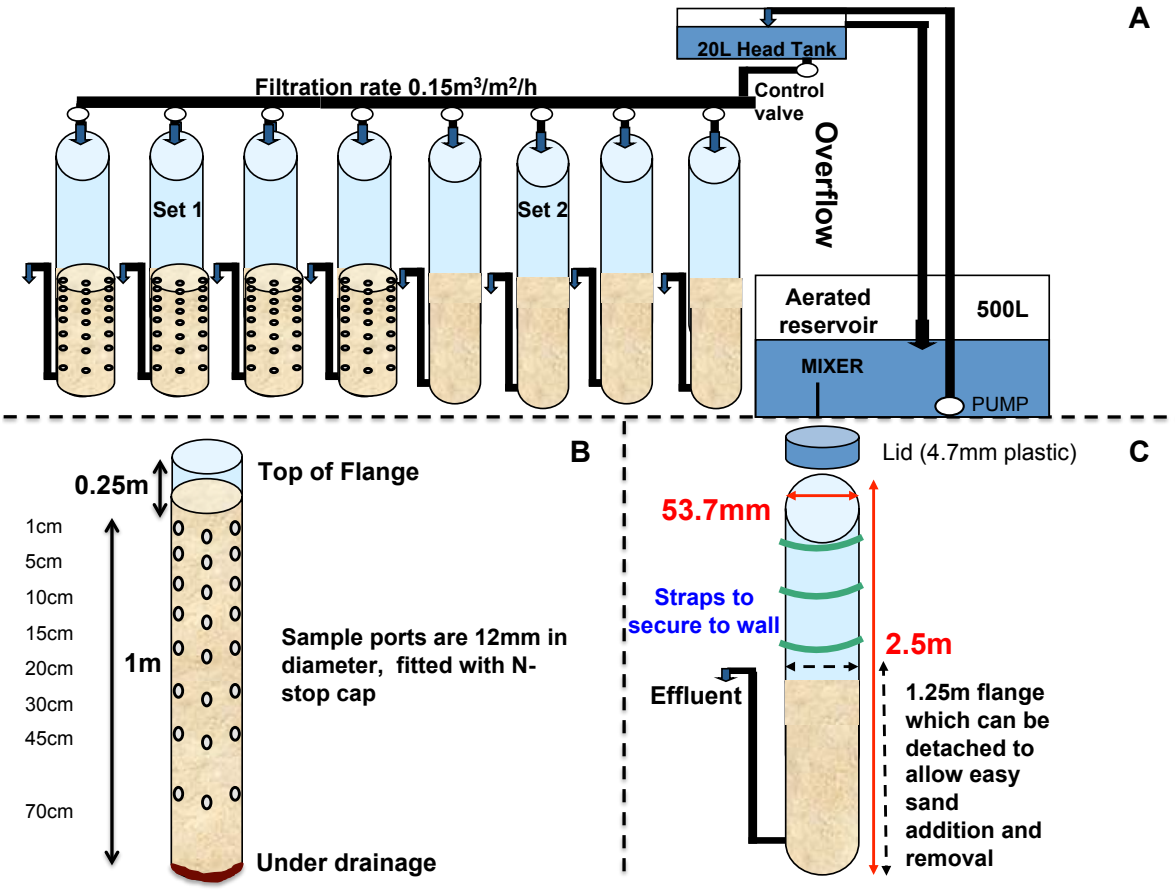


Figure 5.1: Schematic of laboratory-scale slow sand filters. ***A.*** *depicts the overall view of the LSSF setup with 2 sets of 4 filters, each 2.5m high by 54mm diameter.* ***B.*** *depicts the design of set 1 which have a series of holes at different heights around the circumference of the filter. These holes are to allow for investigation into spatial distribution.* ***C.*** *depicts the design of set 2 which are the same as set 1 with the exception of the omission of the holes.*

Extracted DNA was used to amplify the V4-V5 regions of the 16S rRNA gene in eight replicates with the universal bacterial primers 515F [GTGCCAGCMGCCGCGGTAA] and 926R [CCGTCAATTCMTTTRAGT] with different multiplex identifiers (MIDs) attached using Phusion polymerase (Finnzymes, Espoo, Finland) with the GC buffer and 2.5 % dimethyl sulfoxide (DMSO). Cycling conditions consisted of an initial denaturation at 98 °C for 30s, 23 cycles of 98 °C for 10s, 55 °C for 30s, and 72 °C for 10s, and a final extension at 72 °C for 5 min. After amplifications, the replicates were pooled and purified with the Qiagen Gel Cleanup kit. The size and quality of the PCR products were determined with a Bioanalyzer 2100 (Agilent Technologies) and the DNA concentrations were measured with a Qubit fluorometer (Invitrogen, UK). Purified and quantified products were sent to the Centre for Genomic Research (Liverpool University, UK) where emulsion PCR and sequencing was

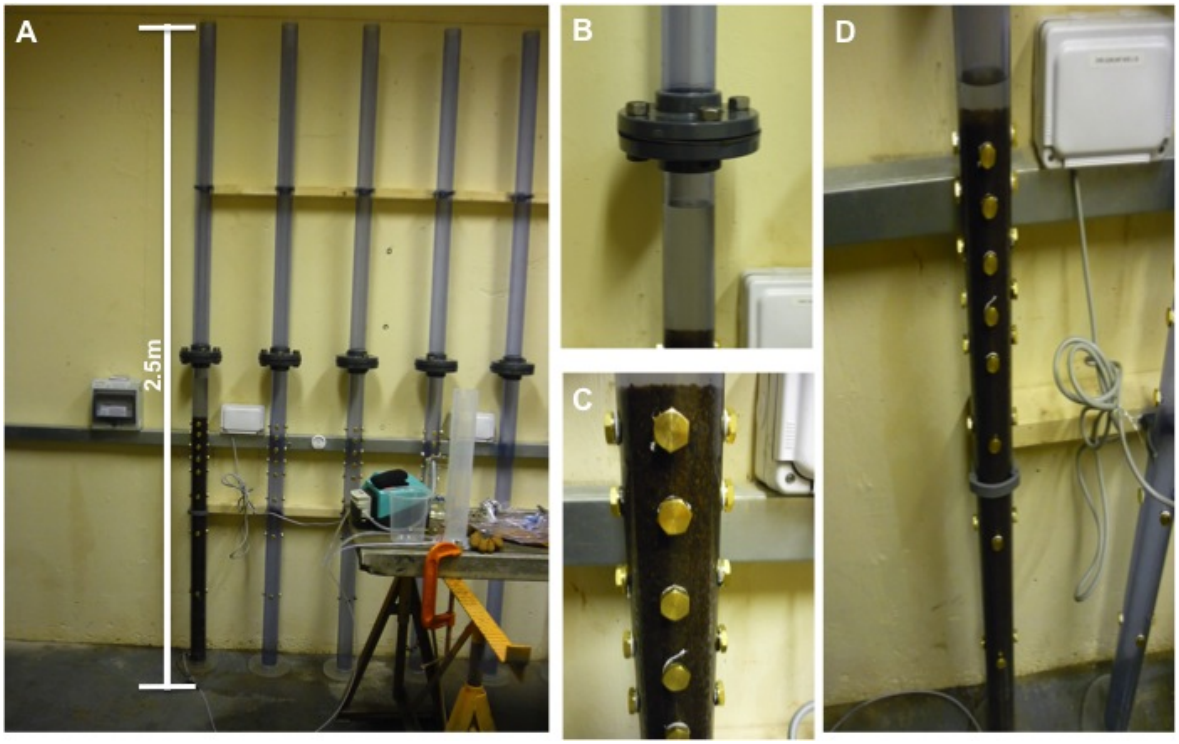


Figure 5.2: Photograph of laboratory-scale SSFs A. depicts the overall view of 4 of the LSSFs. B. depicts the flange used to connect the two pipes used to make the filter. C and D. depict the design of set two of the LSSFs (sampling ports).

performed.

454 Data Processing

Pyrosequencing reads contain a substantial number of errors (noise), which include sequencing errors from the inclusion or deletion of single bases in homopolymer runs of $\geq$3 bp, PCR single base substitutions and PCR chimeras [Quince et al., 2009]. Therefore prior to generating OTUs this noise was removed using the AmpliconNoise pipeline [Quince et al., 2011]; this comprises filtering, flowgram and sequence clustering steps and has been shown to reduce noise by 50% in environmental data sets. Subsequently, chimeras were identified using Perseus [Quince et al., 2011]. This algorithm generates a chimera index (CI) for each read that is $\geq$0 with higher values corresponding to reads that are most likely to be chimeric. Perseus identifies the likely parent sequences and breaking-points of such candidate sequences through pairwise alignments. Logistic regression is then used to classify chimeras so that the pyrosequencing data output lists chimeric and non-chimeric sequences. The lower the probability of the sequence evolving naturally, the higher the CI. After denoising and chimera checking taxonomic classification was assigned using the RDP classifier

[Cole et al., 2009] with a 80% confidence threshold.

5.2.5 Statistical Analysis

To allow robust comparisons among samples containing different numbers of sequences, sample diversity was calculated based on samples rarefied to contain 1500 sequences. The taxonomic and OTU tables generated for the samples were used to calculate pair-wise similarities among samples based on the Bray-Curtis similarity index. The resulting matrices were examined for temporal and spatial patterns in bacterial community structure by using non-metric multidimensional scaling (NMDS) as implemented in the Vegan package in R [Oksanen et al., 2012]. Significant differences in the microbial community composition between filter types (ISSFs, sterile and non-sterile LSSFs), location (industry or laboratory), presence and absence of sample ports and filter age were determined using the Adonis function in Vegan, which performs a permutation or nonparametric multivariate analysis of variance (MANOVA) [Anderson, 2001]. The contribution of individual taxa to overall community dissimilarity was determined using SIMPER (similarity percentage) analysis [Clarke, 1993]. Shannon diversity indices, Chao's diversity index, Pielou's evenness and rarefaction curves were calculated using Vegan on rarefied samples at the 3% nucleotide distance. The relationships between environmental variables and patterns in bacterial community structure were examined by canonical correspondence analysis (CCA) with significance tested by analysis of variance tests (ANOVA) after reducing the overall suite of environmental variables using step-wise variable selection based on Akaike's Information Criterion (AIC).

5.3 Results

Within this section the results of the study alongside discursive analysis will be presented, with an overall discussion being provided in Section 5.4.

5.3.1 Water Quality

Analysis of the water quality parameters and overall water quality performance ($\bigtriangledown$) showed that, as with the full-scale filters (discussed in Chapter 4), the LSSFs produce good quality water (all removal results can be seen in Tables 5.1 and 5.2). The laboratory-scale filters

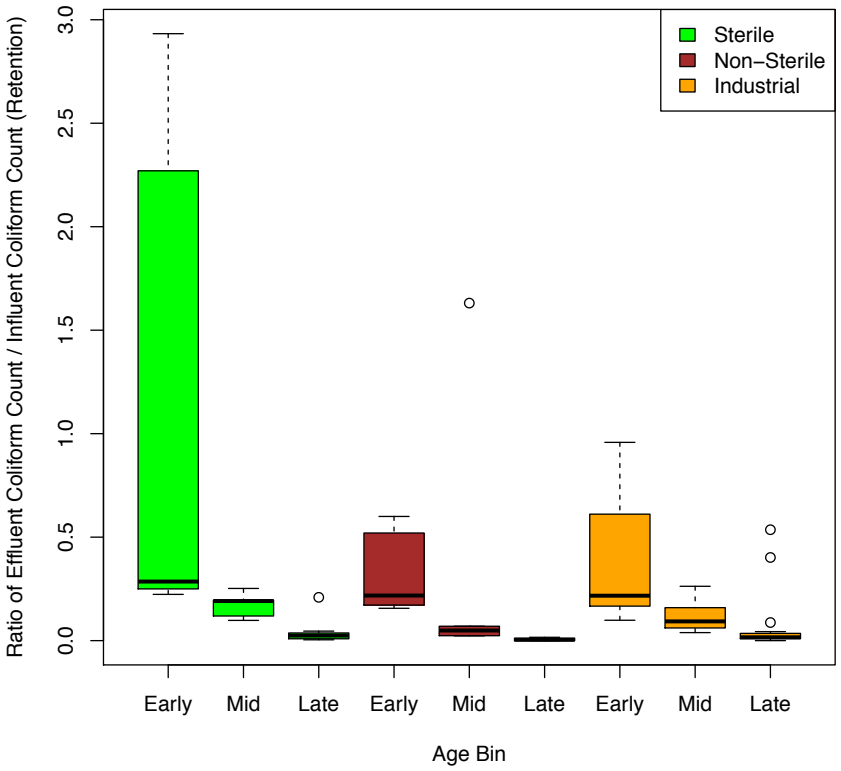


Figure 5.3: Boxplot depicting the fraction of coliforms retained by laboratory-scale and industrial SSFs over time. Early age bin: 0-3 weeks, Mid: 4-6 weeks and Late: 7-10 weeks.

met all of the quality parameters set out in the European Union Directive: 98/83/EC (1998), except coliform removal targets (0/100ml). However, a significant negative correlation was found between coliform retention and age of the filters, where both LSSFs and ISSFs displayed the same trend of improved removal with age (Figure 5.3). Additionally, multivariate linear regression analysis showed that laboratory-scale filters do mimic the ISSFs in terms of overall performance ($\bigtriangledown$) when the age of the filters is taken into consideration (R^2: 0.43, p-value: 2.038×10^{-8}), with non-sterile LSSFs mimicking ISSFs faster than sterile LSSFs. Further, no significant difference was found in water quality production between filters with and without sample ports.

5.3.2 Bacterial Diversity and Richness

Rarefaction curves, richness and diversity estimates were identified for OTUs at the 3% nucleotide distance, which was used to approximate species [Schloss and Handelsman, 2005]. In general, rarefaction curves (Figure 5.4) showed under-sampling, implying that the true diversity of ISSFs and LSSFs samples are likely to be underestimated. However comparison to the known diversity of the two mock communities (Figure 5.5) has shown that diversity esti-

Table 5.1: Age refers to the number of days in operation. Units of measurement are percentage removal, positive number mean removed, negative number indicate an increase of the parameter

Average (based on 5 replicas) Percentage Removaal of Various Water Quality Parameters for Sterile Labscale SSFs

TYPE	FILTER	AGE	NTU	DO	COD	PO4	NO3	NO2	NH4	TVB	COLIFORM	DOC	UVA	SUVA	Performance Metric
STERILE	BETA	6	-28.05	4.29	-32.47	91.72	-329.00	100.00	0.00	-32.92	77.65	3.39	3.19	-0.20	5
STERILE	BETA	13	-2.75	7.76	-134.14	96.61	-200.00	-200.00	6.25	-53.06	72.60	7.01	11.06	4.35	6
STERILE	BETA	20	46.53	9.73	-119.53	95.13	-35.00	80.00	-10.00	30.27	70.31	8.21	-15.71	-26.06	6
STERILE	BETA	27	67.49	15.38	-58.35	22.54	-335.00	0.00	20.00	55.84	74.78	10.48	-49.09	-66.54	6
STERILE	BETA	34	9.14	9.51	-20.00	92.50	-50.00	42.86	-6200.00	76.75	80.43	71.17	17.07	-187.64	8
STERILE	BETA	41	31.67	2.84	-22.40	37.50	0.00	50.00	-5900.00	66.18	89.29	36.19	3.03	-51.97	8
STERILE	BETA	49	41.30	2.50	-11.67	97.53	0.00	55.56	-200.00	31.84	97.32	14.79	1.00	-16.18	8
STERILE	BETA	55	15.91	0.81	6.85	99.22	0.00	22.22	44.90	-61.18	99.40	10.07	6.54	-3.92	9
STERILE	BETA	63	66.33	1.54	27.40	98.89	53.49	50.00	53.33	27.38	97.79	19.97	18.18	-2.23	8
STERILE	EPSILON	6	20.29	5.20	18.42	98.76	-329.00	-533.00	0.00	-34.32	69.39	6.66	2.13	-1.71	7
STERILE	EPSILON	13	23.85	10.08	-134.93	92.67	-200.00	-200.00	1.03	-61.22	-290.00	7.01	12.44	3.55	5
STERILE	EPSILON	20	44.15	8.81	-134.38	93.40	-15.00	80.00	85.71	57.35	-415.63	14.20	-19.29	-34.55	7
STERILE	EPSILON	27	-32.62	11.43	-131.99	97.59	-150.00	40.00	85.00	50.21	-28.70	11.34	-50.91	-71.33	7
STERILE	EPSILON	34	49.94	11.70	-20.00	92.50	-50.00	57.14	-900.00	77.28	85.87	54.78	17.07	-17.17	8
STERILE	EPSILON	41	36.00	3.71	-8.94	37.50	0.00	62.50	-1000.00	-2.69	-202.38	34.78	4.04	-48.05	8
STERILE	EPSILON	49	58.02	7.50	-16.69	98.65	0.00	55.56	-100.00	6.23	-75.10	9.75	0.50	-12.25	8
STERILE	EPSILON	55	47.40	11.11	28.77	99.67	0.00	88.89	44.90	-90.59	83.23	6.36	5.14	-5.02	7
STERILE	EPSILON	63	59.17	2.53	10.52	99.03	53.49	50.00	46.67	97.23	79.07	22.64	17.27	-4.11	7
STERILE	GAMMA	6	10.46	5.31	-42.97	-3.64	-429.00	100.00	0.00	-36.84	-363.27	3.77	2.13	-1.71	7
STERILE	GAMMA	13	36.80	6.60	-183.48	94.21	-300.00	-150.00	7.19	-20.41	-127.00	9.22	12.44	3.55	5
STERILE	GAMMA	20	66.78	8.61	-134.38	64.28	-50.00	80.00	-271.43	55.64	75.00	11.35	-47.27	-85.35	6
STERILE	GAMMA	27	53.66	6.12	-122.49	93.22	-185.00	-20.00	11.25	45.98	80.87	11.92	-49.09	-34.64	7
STERILE	GAMMA	34	49.06	5.96	-20.00	90.06	-50.00	42.90	-2400.00	69.74	90.22	51.95	17.07	-17.17	8
STERILE	GAMMA	41	30.00	1.75	-8.94	37.50	0.00	50.00	-2300.00	32.81	88.10	35.19	3.03	3.03	8
STERILE	GAMMA	49	56.48	2.50	-15.43	98.88	50.00	55.60	-2400.00	5.66	95.40	11.36	1.00	1.00	8
STERILE	GAMMA	55	-48.66	3.92	26.03	99.01	0.00	44.40	44.90	-89.12	98.80	9.68	6.54	217.25	8
STERILE	GAMMA	63	58.83	1.54	2.73	99.30	53.49	50.00	46.67	45.37	97.18	20.54	18.18	-1105.51	8
STERILE	THETA	8	64.49	8.81	-71.88	78.31	35.00	80.00	0.00	18.29	-415.63	2.51	-15.71	-18.69	6
STERILE	THETA	15	65.13	11.53	-147.82	96.56	-35.00	80.00	77.50	9.75	-51.30	10.19	-38.18	-53.86	7
STERILE	THETA	22	51.44	11.08	-120.00	92.50	0.00	42.86	-1400.00	32.46	50.00	15.21	3.41	-13.91	9
STERILE	THETA	29	13.83	2.29	-53.00	46.25	-15.00	25.00	-100.00	-64.91	59.52	35.59	14.14	-33.30	8
STERILE	THETA	37	62.97	10.33	-25.47	81.17	-150.00	-566.67	-100.00	20.57	55.94	37.15	6.50	-48.76	7
STERILE	THETA	43	54.96	19.58	-5.48	99.26	0.00	-211.11	44.90	-0.29	83.83	13.97	1.87	-14.06	8
STERILE	THETA	51	56.17	12.87	25.71	99.30	53.49	50.00	40.00	62.47	99.60	6.25	6.36	0.12	8

Table 5.2: Age refers to the number of days in operation. Units of measurement are percentage removal, positive number mean removed, negative number indicate an increase of the parameter

Average (based on 5 replicas) Percentage Removaal of Various Water Quality Parameters for Non-sterile Labscale SSFs

TYPE	FILTER	AGE	NTU	DO	COD	PO4	NO3	NO2	NH4	TVB	COLIFORM	DOC	UVA	SUVA	Performance Metric
NON-STERILE	ALPHA	8	-7.12	9.94	-39.84	96.02	-35.00	60.00	14.29	-55.64	84.38	2.79	6.43	3.74	6
NON-STERILE	ALPHA	15	-10.17	15.70	-186.62	89.78	-150.00	40.00	85.00	-92.65	74.78	1.27	0.91	-0.36	7
NON-STERILE	ALPHA	22	-37.17	10.76	-72.00	90.62	0.00	42.86	-1400.00	78.20	93.48	0.93	-12.20	-13.25	9
NON-STERILE	ALPHA	29	-88.00	5.02	-55.45	-16.25	0.00	50.00	-1500.00	-16.96	-63.10	-0.64	-13.13	-12.41	8
NON-STERILE	ALPHA	37	-19.45	8.80	-50.56	98.43	-50.00	55.56	-1600.00	6.27	97.70	2.30	11.00	8.91	8
NON-STERILE	ALPHA	43	5.35	15.72	27.40	98.97	0.00	44.44	40.82	-60.29	98.80	3.43	6.07	2.74	8
NON-STERILE	ALPHA	51	-19.33	3.19	62.47	92.53	53.49	50.00	46.67	-19.75	98.39	6.82	10.00	3.41	7
NON-STERILE	DELTA	6	-147.54	3.73	-55.09	98.50	-343.00	-433.33	-16.67	-7.81	81.63	33.99	5.32	-43.43	7
NON-STERILE	DELTA	13	5.61	25.05	-206.97	97.97	-200.00	-1400.00	99.69	-53.06	40.00	29.66	21.66	-11.37	7
NON-STERILE	DELTA	20	64.58	14.86	-128.13	4.45	0.00	60.00	100.00	-32.30	82.81	-95.55	-42.86	26.95	8
NON-STERILE	DELTA	27	34.99	14.86	-52.81	98.13	0.00	40.00	100.00	25.74	93.04	-49.68	-81.82	-21.47	8
NON-STERILE	DELTA	34	71.21	14.73	-32.00	92.50	-50.00	57.14	-1400.00	41.67	96.74	33.86	50.24	24.77	9
NON-STERILE	DELTA	41	-92.33	4.37	-76.25	50.00	0.00	50.00	-1400.00	12.75	97.62	-0.64	48.99	49.32	9
NON-STERILE	DELTA	49	13.48	6.63	-19.20	98.88	-50.00	55.56	-1200.00	28.02	100.00	6.33	6.00	-0.35	9
NON-STERILE	DELTA	55	14.33	9.71	-36.99	99.47	0.00	88.89	40.82	-59.71	100.00	41.47	19.63	-37.33	9
NON-STERILE	DELTA	63	53.67	0.99	26.49	99.30	53.49	50.00	46.67	93.02	99.60	35.79	23.64	-18.92	8
NON-STERILE	ETA	8	-67.97	11.68	-69.53	98.27	35.00	40.00	0.00	-11.67	54.69	6.50	6.43	3.74	6
NON-STERILE	ETA	15	-50.83	12.26	-137.53	90.48	-15.00	40.00	85.00	-42.03	86.09	-1.04	0.91	-0.36	8
NON-STERILE	ETA	22	-2.38	12.33	-50.00	92.50	-50.00	57.14	-900.00	50.00	89.13	1.07	-12.20	-13.25	9
NON-STERILE	ETA	29	-36.67	5.35	-46.88	33.75	0.00	50.00	-1000.00	-18.13	97.62	-4.87	-13.13	-12.41	8
NON-STERILE	ETA	37	63.14	13.04	-38.02	98.27	0.00	55.56	-1500.00	11.32	97.70	3.71	11.00	8.91	8
NON-STERILE	ETA	43	-58.11	20.44	93.15	99.26	0.00	88.89	40.82	-188.82	100.00	4.02	6.07	2.74	9
NON-STERILE	ETA	51	-31.17	12.21	80.39	99.03	53.49	25.00	40.00	93.15	99.40	7.20	10.00	3.41	8
NON-STERILE	ZETA	6	-59.43	4.41	-68.01	89.39	-314.00	-233.33	0.00	-7.22	67.35	35.91	4.26	-49.40	6
NON-STERILE	ZETA	13	28.03	19.99	-226.55	98.35	-200.00	100.00	99.78	-23.47	20.00	30.26	22.58	-11.01	7
NON-STERILE	ZETA	20	69.83	15.88	-134.38	98.54	0.00	80.00	100.00	44.75	93.75	-92.70	-44.29	25.13	9
NON-STERILE	ZETA	27	84.87	11.84	-128.82	96.56	0.00	40.00	100.00	70.81	97.39	-47.38	-70.00	-15.35	7
NON-STERILE	ZETA	34	75.97	12.43	-62.00	93.06	-50.00	42.90	0.00	78.86	97.83	34.99	51.22	24.96	9
NON-STERILE	ZETA	41	-8.83	3.17	-71.36	53.75	0.00	50.00	0.00	49.71	98.81	0.76	48.99	48.60	9
NON-STERILE	ZETA	49	70.14	3.15	-22.96	99.10	35.00	55.60	0.00	4.72	100.00	7.13	10.00	3.09	9
NON-STERILE	ZETA	55	62.83	2.85	-61.64	99.22	0.00	55.60	40.82	-45.00	99.70	24.70	18.69	-7.98	8
NON-STERILE	ZETA	63	61.17	1.43	5.32	99.03	53.49	50.00	46.67	90.99	99.80	36.74	26.36	-16.40	8

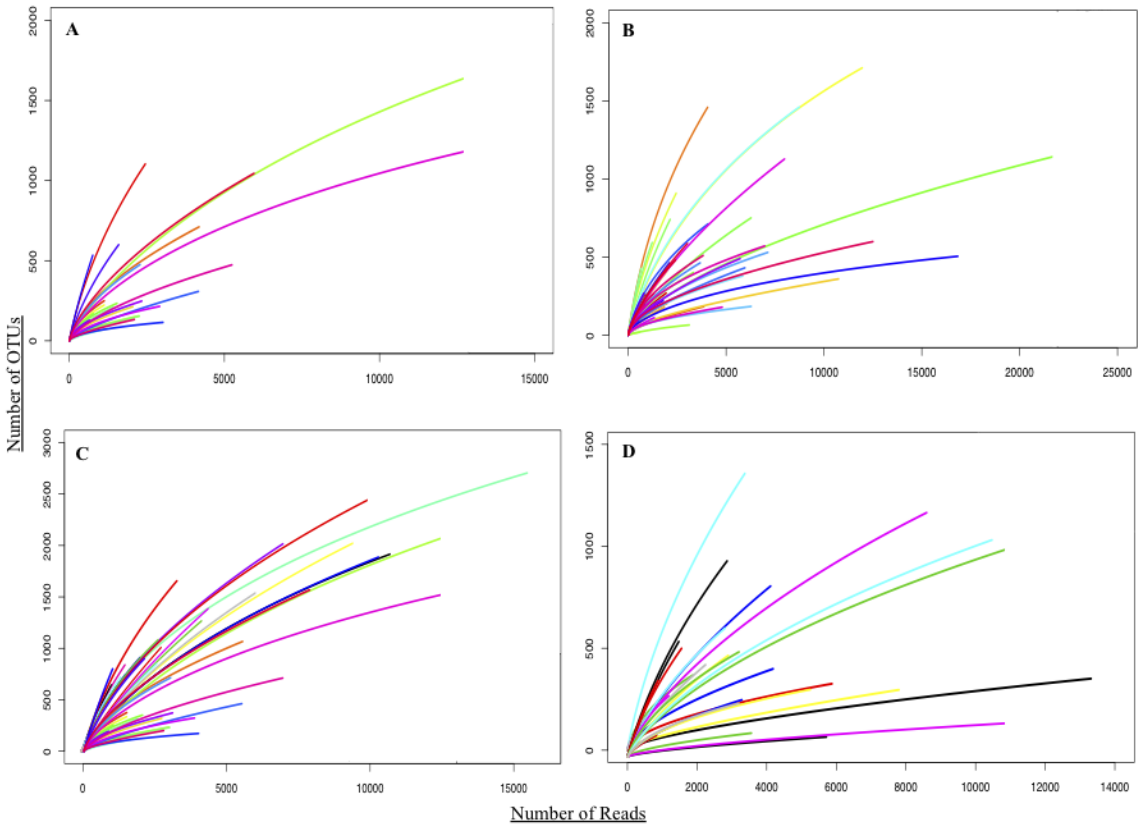


Figure 5.4: Rarefaction Curves. A: Influent water, B: Effluent water, C: Industrial SSFs, and D: Labscale SSFs

mates are very accurate, overestimating the true diversity on average by 1.82 times thus being more accurate than that found by the Illumina results discussed in Chapter 4. The Shannon index of diversity (H') was determined for all samples and averages calculated for the type of filter (industrial, sterile and non-sterile LSSFs) and age profile (early: 0-3 weeks, mid: 4-6 weeks and late: 7-10 weeks of operation) (Table 5.3). These diversities were consistent with previous studies [Hunter et al., 2012]. Overall the Shannon index ranged from 0.74 found in the influent water feeding the filters to 6.71 in the oldest industrial SSF. Comparison of the mean H' of the different types of filter and age profiles revealed that the highest diversity was found in middle-aged industrial filters. Further, there was statistically significant differences in H' between groups during the first six weeks of operation. However, by the late age phase ($\geq$7weeks) of filter operation there was no statistically significant difference in H' between industrial and laboratory-scale filters (p-value: 0.172). Furthermore, the diversity of all the filters converged by the conclusion of the experiment (Figure 5.6).

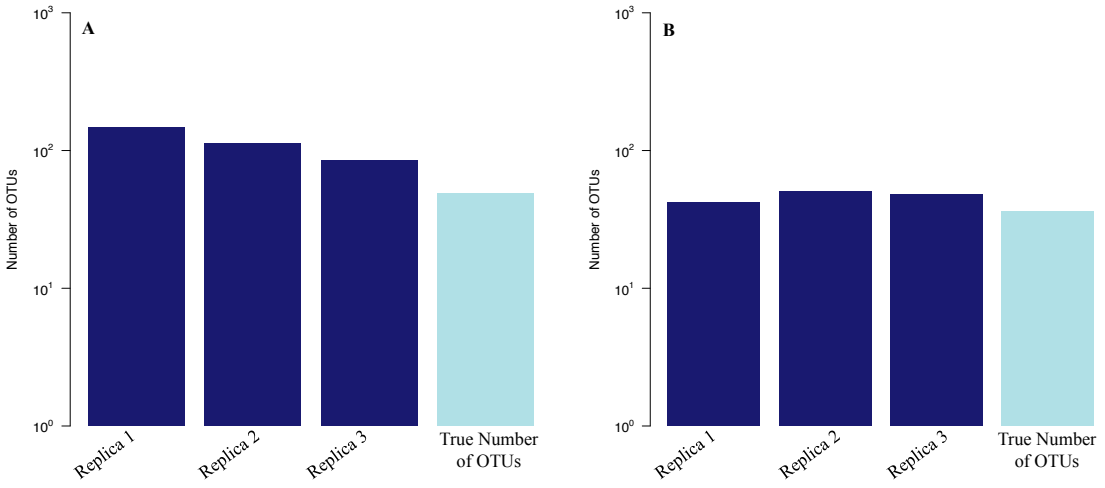


Figure 5.5: Barplots showing the actual number of OTUs in light blue and the sequenced number of OTUs in dark blue for two mock communities. ***A)****:mock community created by Shakya et al. [2013] and* ***B)****:SSF mock community created in Chapter 4.*

The trend of increasing diversity with age is consistent with the increasing number of 16S rRNA gene copies (Figure 5.7). Using pairwise Wilcoxon signed-rank tests no significant difference in the 16S rRNA gene abundance between non-sterile LSSFs and ISSFs (p-value: 0.1209) was found. However, sterile LSSFs possessed a significantly lower number of 16S rRNA copies at the early (p-value: 0.0174) and late (p-value: 0.000406) age bins compared to both ISSFs and non-sterile SSFs. Regardless of these points, laboratory-scale and full-scale SSFs display similar levels of diversity, OTU number and 16S rRNA copy number, further evidence that the richness and diversity of environmental microbial communities can be replicated in the laboratory. In terms of shaping and influencing the sand microbial community, the influent water feeding the ISSFs and LSSFs plays a large role, with 37% of the OTUs from ISSFs sand samples and 52% of OTUs from LSSFs sand samples being shared with their source water (Table 5.4). Interestingly, there was a marginally significant (p-value: 0.07889) difference in the number of shared OTUs between ISSFs and LSSFs,

Table 5.3: Average species richness, diversity and evenness for industrial and laboratory-scale SSFs at different age bins (early: 0-3weeks, mid: 4-6weeks and late: 7-10weeks).

	INDUSTRIAL			STERILE LAB			NON-STERILE LAB				
	EARLY	MID	LATE	EARLY	MID	LATE	EARLY	MID	LATE	INFLUENT	EFFLUENT
3% OTU											
Observed OTUs	481	569	1056	282	348	974	777	799	1119	405	451
Choa's Richness	1105	2880	1080	220	752	1023	465	1672	1845	1816	3033
Shannon Diversity Index	6.050	6.279	6.138	2.330	4.219	5.907	4.218	5.513	6.095	3.699	3.758
Pielou's Evenness	0.975	0.906	0.910	0.417	0.758	0.815	0.865	0.815	0.893	0.545	0.640

Table 5.4: Table displaying the percentage of shared OTUs in each of the industrial and laboratory-scale SSFs with the influent and effluent water sources.

		% Shared OTUs		
Filter	Location	Influent and Sand	Effluent and Sand	Influent and Effluent
ISSF A	Industry	42.50	72.9	31.71
ISSF B	Industry	31.36	54.40	25.55
Alpha	Non-sterile Laboratory-scale	45.26	82.50	27.60
Delta	Non-sterile Laboratory-scale	45.60	65.60	24.94
Eta	Non-sterile Laboratory-scale	52.40	70.19	22.96
Zeta	Non-sterile Laboratory-scale	55.29	72.00	24.51
Beta	Sterile Laboratory-scale	68.17	76.25	23.79
Epsilon	Sterile Laboratory-scale	35.09	72.16	17.83
Gamma	Sterile Laboratory-scale	65.68	21.25	26.63
Theta	Sterile Laboratory-scale	46.64	56.98	19.10

with laboratory-scale filters having greater variation in the percentage shared OTUs with the source water compared to ISSFs.

Similarly to the ISSFs in Chapter 4, LSSFs display a significant positive correlations between species evenness and $\bigtriangledown$, however, only at the phylum (p-value: 0.0001), class (p-value: 0.0001), order (p-value: 0.0039) and OTU (p-value: 0.0180) levels of classification, compared to all levels in ISSFs. Compounding the previous conclusions in Chapter 4, that filters with a greater species evenness produce a higher standard of water quality than filters of lower species evenness.

5.3.3 Differences and Similarities in Community Structure Between Labscale and Industrial SSFs

A total of 468,773 sequences were classified below domain level being affiliated to 30 phyla, 61 classes, 76 orders, 191 families, 591 genera and 28,612 OTUs. The dominant phyla in sand samples were: *Proteobacteria, Bacteriodetes, Acidobacteria, Actinobacteria, Planctomycetes, Verrucomicrobia*, and *Gemmatimonadetes* representing 60.80, 9.27, 5.17, 4.67, 4.42, 1.78, and 1.15 % of the reads respectively (Figure 5.9). This richness in diversity is consistent with a previous SSF study [Wakelin et al., 2011] although there are discrepancies

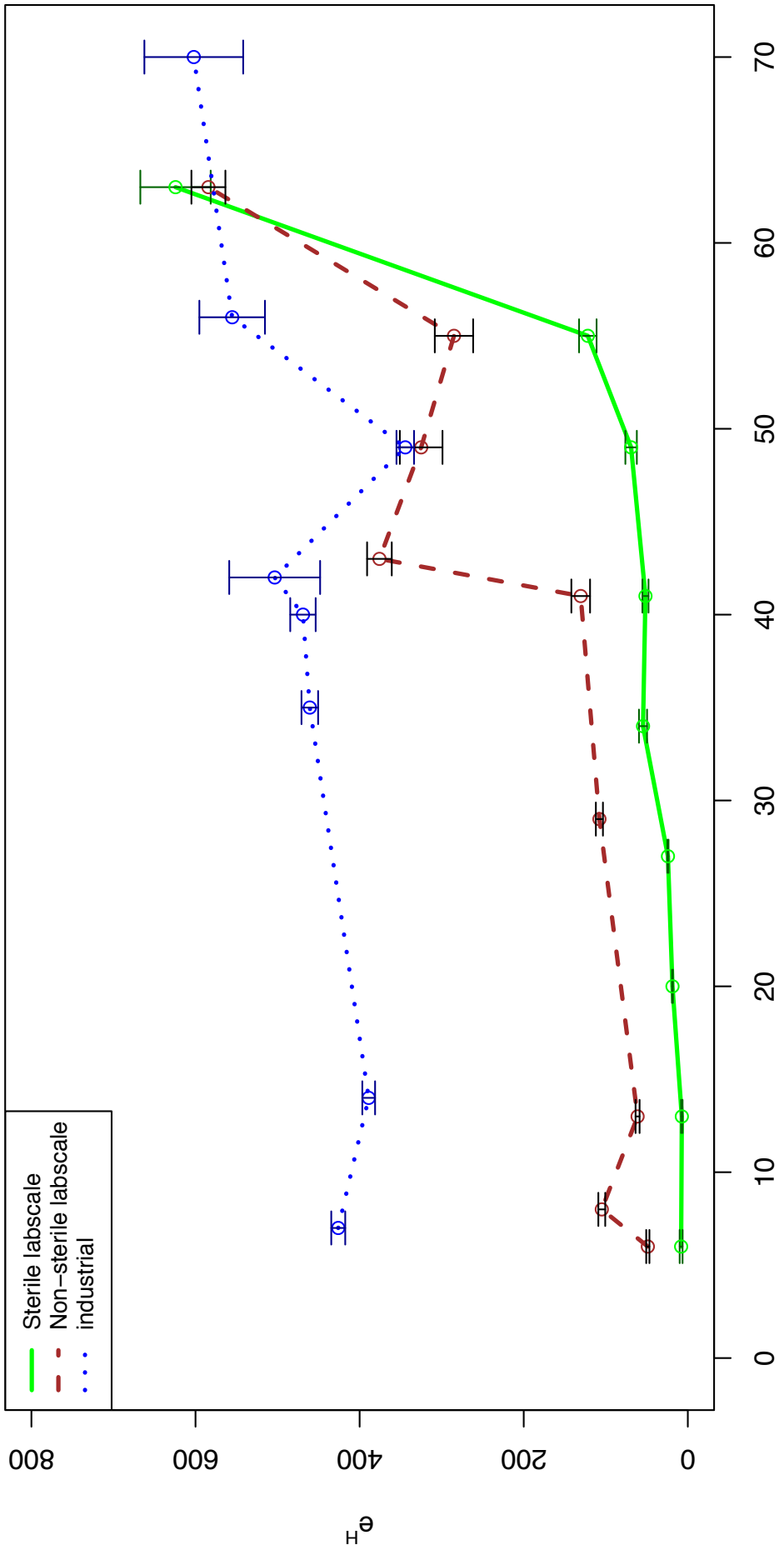


Figure 5.6: Scatter plot showing how the shannon index changes as laboratory-scale and industrial SSFs age. Points based on the average within each filter set with bars representing the standard error. Data generated from 454 16S rRNA amplicon sequencing.

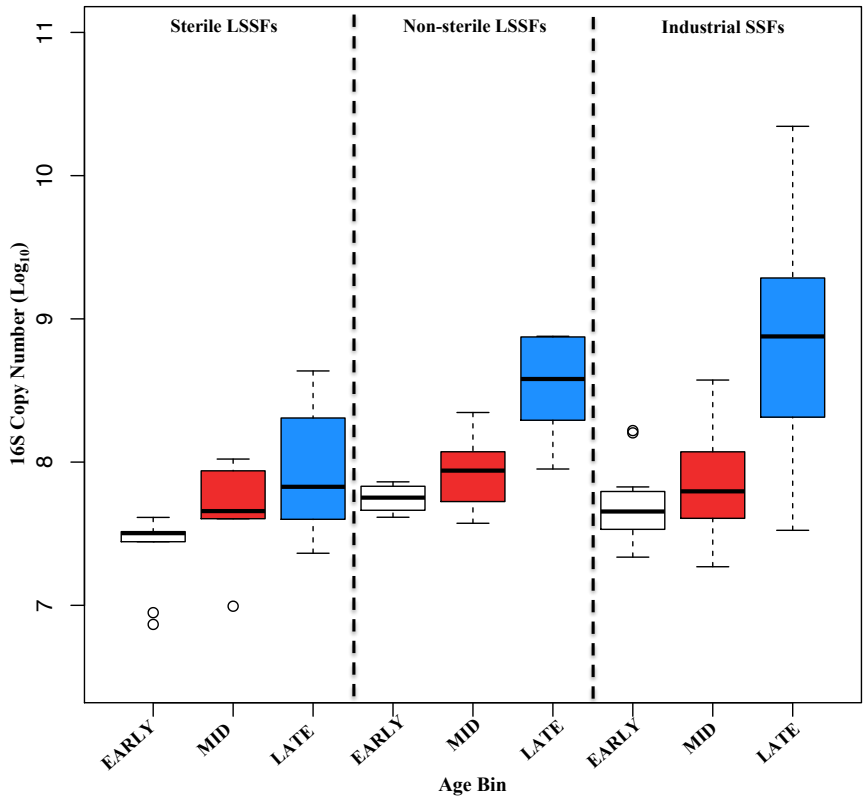


Figure 5.7: Box-and-whisker plot comparing the abundance of the 16S rRNA gene in labscale and industrial SSFs at three age bins. The top and bottom boundaries of the boxes show the 75th and 25th percentile and the ends of the whiskers show the maximum and minimum values within 1.5 of the interquartile range of the upper and lower quartiles (respectively). Bold lines within the boxes represent median values (50th percentile).

between dominant taxa which could be attributed to the different methodological approach adopted. Referring more closely to Figure 5.9 it can be seen that the relative abundance of dominant phyla changed with the age of the filters, with *Proteobacteria*, specifically *Alphaproteobacteria* and *Betaproteobacteria* dominating ISSF samples, which is in accordance with findings from a Chinese sand filter [Bai et al., 2013]. However, a more stark difference is seen in the rare sand phyla ($\leq$1 % of all classified sequences) where phyla abundance and presence change dramatically (Figure 5.10). Overall, the LSSFs microbial community are indistinguishable from industrial SSF at the phylum and class level by the end of the experiment, with the main difference being the abundance of unclassified organisms. These conclusions are further confirmed by qPCR analysis, where vector maps (Figure 5.8) show temporal changes in the microbial community composition with non-sterile LSSFs mimicking the microbial consortia of ISSFs faster than sterile LSSFs. More impressively Figure

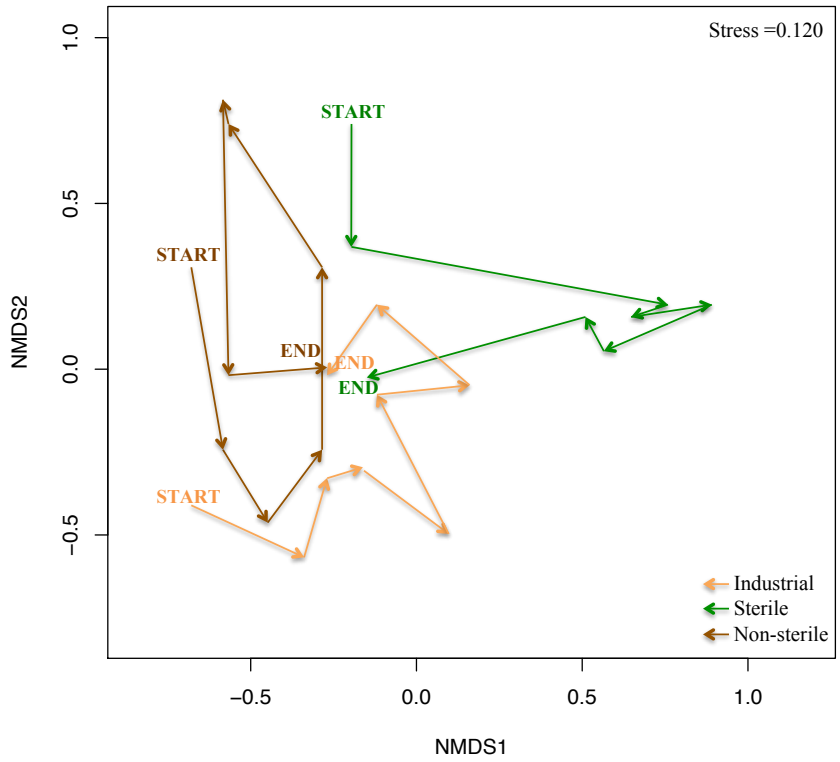


Figure 5.8: Non-metric multidimensional scaling plot with vector maps illustrating the change in the microbial community composition (determined by qPCR) of industrial and labscale SSFs with time.

5.8 depicts the convergence of LSSFs and ISSFs microbial communities by the end of the experimental period.

Initial exploratory analysis (Figure 5.11) showed that the microbial community from sand and water samples clustered into four distinct groups (industrial filters, sterile LSSFs, non-sterile LSSFs and water). As shown in Figure 5.9 the community of non-sterile filters more closely represents the ISSFs community than sterile LSSFs. Using permutation ANOVA it was determined that at both the early (0-3weeks) and mid age bin (4 to 6 weeks) there were significant differences in community composition between labscale and industrial filters (early p-values: 0.001 and R^2: 0.795, Mid p-value: 0.001 and R^2: 0.518). However, by the late age bin ($\geq$ 7 weeks of operation) there was no statistically significant differences between community composition both with 454 and qPCR results (p-value: 0.115 and 0.074 respectively) (Table 5.6) between LSSFs and ISSFs. To identify which taxa contributed most to the differences between the filter types at early and mid age bins SIMPER analysis (Table 5.5) was used and clearly showed that the majority of the differences (>60%) between filter types and age bins are due to unclassifiable organisms, *Comamonadaceae* and *Sph-*

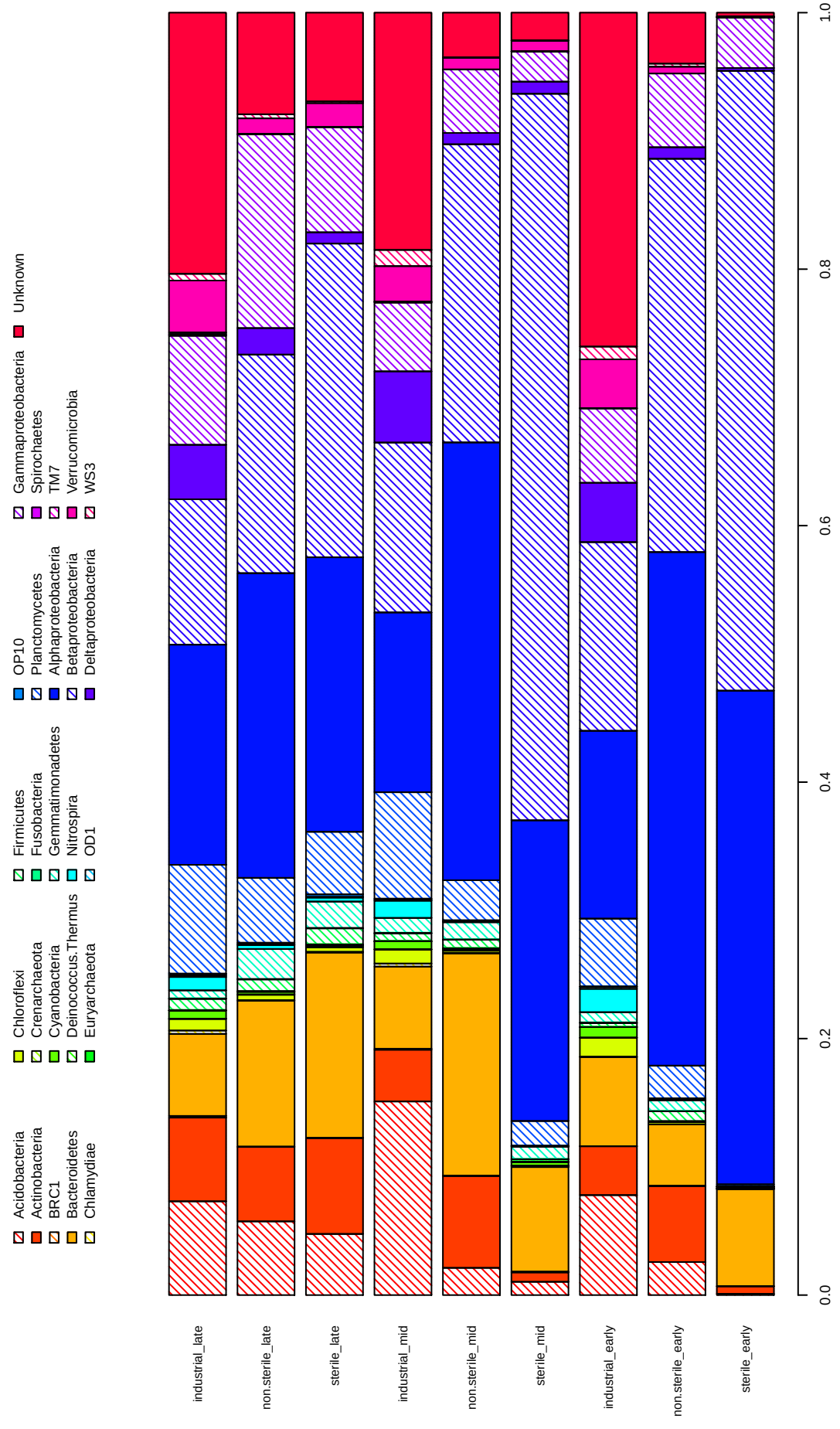


Figure 5.9: Stacked barplots depicting percentage abundance of all the phyla and classes of Proteobacteria present in labscale and industrial filters over time

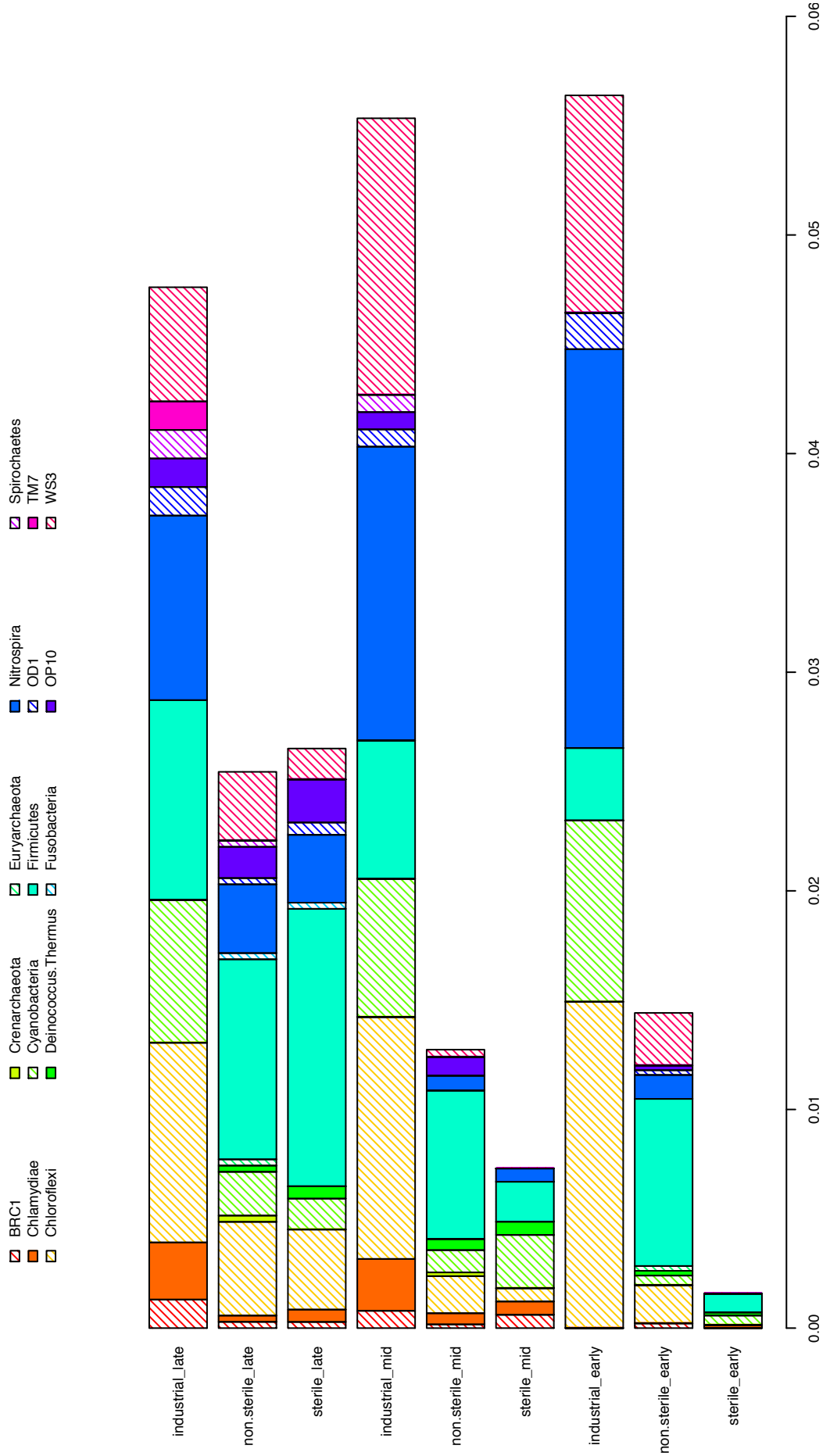


Figure 5.10: Stacked barplot depicting the phyla which makeup ≤ 1 % of the community in labscale and industrial SSF at different ages.

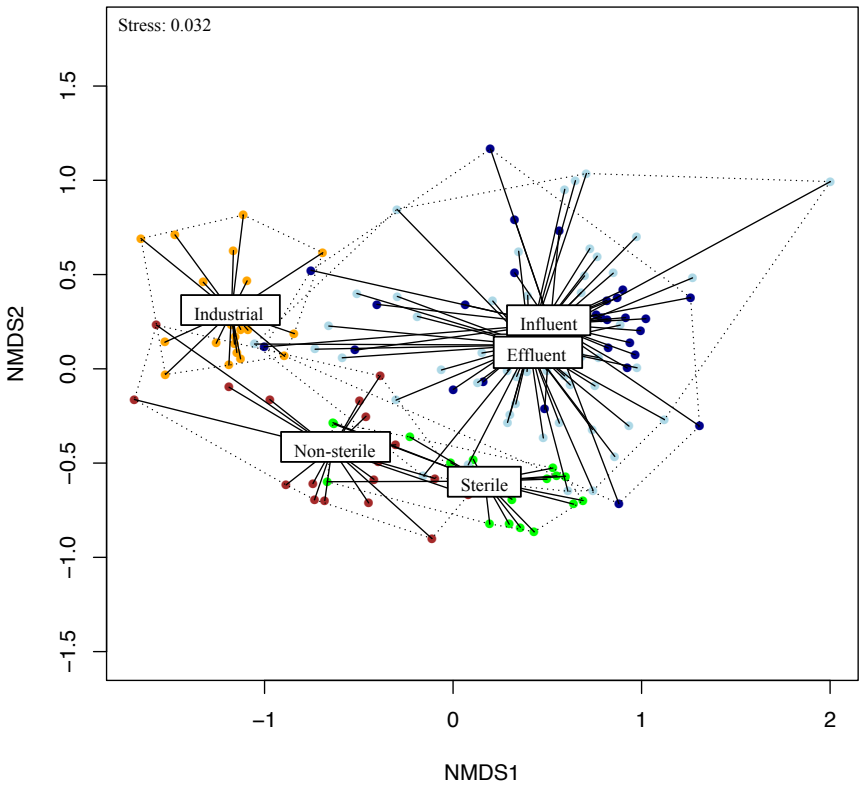


Figure 5.11: Non-metric multidimensional scaling ordination for the microbial community structure for all samples in the 454 dataset at the 97% OTU level.

ingomonadaceae families. Interestingly, these two families and several others in Table 5.5 were identified in Chapter 4 as being key taxa involved with performance.

It should, however, be stressed, that although at higher levels of taxon classification LSSFs and ISSFs communities become indistinguishable once ≥ 7 weeks of age, at the OTU level this is not the case. Referring to Figure 5.12 it can be seen that at the OTU level of classification there are significant differences between groups at all age bins (p-values: 0.001, 0.014 and 0.006 for early, mid and late age bin respectively). Interestingly, however there was no statistically significant difference in community composition at the late age bin between the two types of laboratory-scale filters (p-value: 0.175).

5.3.4 Impact of Filter Identity, Type and Location on the Microbial Community

In order to determine which factors explain the differences between LSSFs and ISSFs microbial communities, permutation ANOVAs were performed with phyla, family and OTU classified data. The results of these and canonical correspondence analysis (Figure 5.13a)

Table 5.5: SIMPER analysis showing the top 10 taxa responsible for differences between community composition in industrial and labscale SSFs at different agebins

Taxon	Family	EARLY AGE			
		Contribution (%)	**Avg. Industrial (%)**	**Avg.Sterile Lab (%)**	**Avg. Non-Sterile Lab (%)**
Unknown	Unknown	37.79	41.51	4.76	20.67
Betaproteobacteria	*Comamonadaceae*	17.77	20.4	34.13	19.98
Alphaproteobacteria	*Sphingomonadaceae*	21.35	7.9	36.75	31.86
Betaproteobacteria	*Oxalobacteraceae*	3.29	0.36	7.47	2.12
Bacteroidetes	*Flavobacteriaceae*	2.89	1.25	7.35	2.28
Planctomycetes	*Planctomycetaceae*	2.71	5.28	0.19	2.27
Bacteroidetes	*Chitinophagaceae*	1.09	2.35	0.84	2.88
Gammaproteobacteria	*Xanthomonadaceae*	0.8	0.41	0.57	3.38
Betaproteobacteria	*Burkholderiales_incertae_sedis*	0.46	1.04	1.36	0.99
Gammaproteobacteria	*Halomonadaceae*	0.11	0.2	0.25	0.13

Taxon	Family	MID AGE			
		Contribution (%)	**Avg. Industrial (%)**	**Avg.Sterile Lab (%)**	**Avg. Non-Sterile Lab (%)**
Unknown	Unknown	36.01	57	21	35
Betaproteobacteria	*Comamonadaceae*	14.85	5	22.65	11.61
Alphaproteobacteria	*Sphingomonadaceae*	10.43	4.5	10.43	12.57
Betaproteobacteria	*Burkholderiales_incertae_sedis*	4.1	0.85	8.78	2.12
Planctomycetes	*Planctomycetaceae*	3.04	6.95	3.06	5.3
Bacteroidetes	*Flavobacteriaceae*	3.03	5.95	6.1	2.53
Bacteroidetes	*Chitinophagaceae*	2.38	3	4.07	4.82
Betaproteobacteria	*Oxalobacteraceae*	1.53	1.12	4.95	0.85
Gammaproteobacteria	*Xanthomonadaceae*	1.14	0.54	1.02	2.14
Gammaproteobacteria	*Halomonadaceae*	0.87	0.96	0.96	0.16

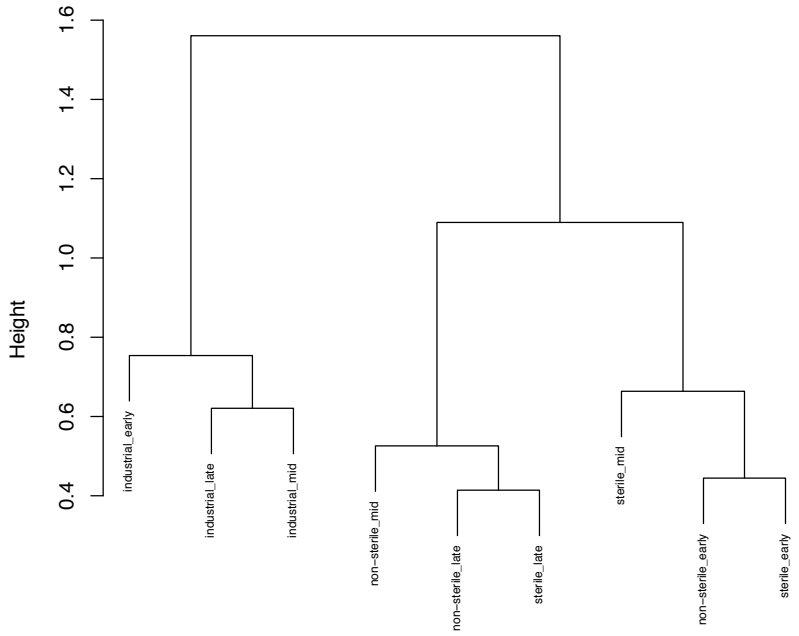


Figure 5.12: Hierarchical cluster dendrogram of the microbial community at the OTU level of classification in industrial and labscale SSFs at different agebins.

with phyla and family data revealed that the type of filter (sterile, non-sterile or industrial) explained the biggest proportion of the bacterial community composition, whereas individual filter identity (i.e., α, β, γ), location (industry or laboratory) and the presence of sample ports were not significant variables, implying that community composition is not random. Further referring to Table 5.7A, age bin and the depth of the sample are also significant factors in explaining differences in the microbial community. However, at the OTU level of classification the location of the filters (laboratory or industry) was the most significant factor in explaining the variation in the community (Figure 5.13b and Table 5.7B), with both filter identity and the presence of sample ports remaining insignificant variables. Such differences

Table 5.6: Results from MANOVA analysis using 454 and qPCR data, showing the level of community composition similarity between labscale and industrial SSFs at various ages.

Analysis	**Factor**	***qPCR***		***454 – Phyla and Proteobacteria***	
		R^2	***P* value**	R^2	***P* value**
Early filter age (0-3 weeks) by type of filter (Sterile, Non-sterile and Industrial)	Early age by type	0.068	0.001	0.795	0.001
Intermediate filter age (4-6 weeks) by type of filter (Sterile, Non-sterile and Industrial)	Intermediate age by type	0.05	0.043	0.518	0.001
Late filter age (7-11 weeks) by type of filter (Sterile, Non-sterile and Industrial)	Late age by type	0.03	0.074	0.197	0.115

Table 5.7: Canonical correspondence analysis of the relative abundance of bacterial: (A) phyla and (B) OTUs with filter parameters. ∗ denotes significant parameters.

A) Parameter	Degrees of Freedom	χ^2	F value	Number of permutations	Pr(>F)
Type	2	0.1799	19.0187	99	0.01*
Age bin	2	0.0307	3.2508	99	0.01*
Depth	1	0.0121	4.6392	99	0.02*
Performance	1	0.0116	2.4576	99	0.02*
Filter Identity	6	0.0281	0.9960	99	0.47
Location	1	0.0079	0.8480	99	0.51
Sample ports	1	0.0007	0.1492	99	0.4
Residual	52	0.2459			

B) Parameter	Degrees of Freedom	χ^2	F value	Number of permutations	Pr(>F)
Location	1	0.6438	2.7589	99	0.01*
Age bin	1	0.7259	1.5530	99	0.01*
Depth	1	0.3826	1.6397	99	0.01*
Type	2	0.2972	1.2737	99	0.01*
Performance	1	0.2760	1.1828	99	0.04*
Filter Identity	7	1.1416	0.9785	99	0.45
Sample ports	1	0.2478	1.0621	99	0.35
Residual	47	10.9671			

at the OTU level of classification due to location might explained by the lack of wildlife and climatic effects in the laboratory, which were shown in Chapter 4 to add significant diversity to ISSFs. Interestingly, when only mature samples (those from the late age bin) were analysed (Figure 5.14), only the depth of the sample was significant in explaining the variation in the microbial community; both filter type and location were no longer significant. Further, individual filters in each group showed little variability in community composition yet displayed clear clustering within their individual group. This is surprising as in Chapter 4 depth was found to be a marginally significant parameter, implying that in LSSFs there is more vertical spatial variation than in ISSFs. This might be induced by the relatively narrow diameter of the LSSFs or the reduced height to width ratio compared with the ISSFs.

5.4 Discussion

This study showed that it is possible to replicate the water quality performance and the microbial community of ISSFs in the laboratory. Although the diversity of the SSFs in this study was found to be high and reproducible between laboratory-scale and industrial filters, significant under sampling was found. This alongside the much higher diversity found in Chapter 4 maybe due to the different sequencing technologies used as Illumina HiSeq 2000 can produce six billion paired end reads [Illumina, 2013], whereas Roche's 454 FLX + only produces one million [Roche, 2013]. Regardless of the differing extents of diversity found

between these results and those of Chapter 4, patterns pertaining to community assembly and performance were the same. For all the water quality parameters tested, results showed that the laboratory SSFs do mimic ISSFs when age is taken into consideration (Figure 5.3).

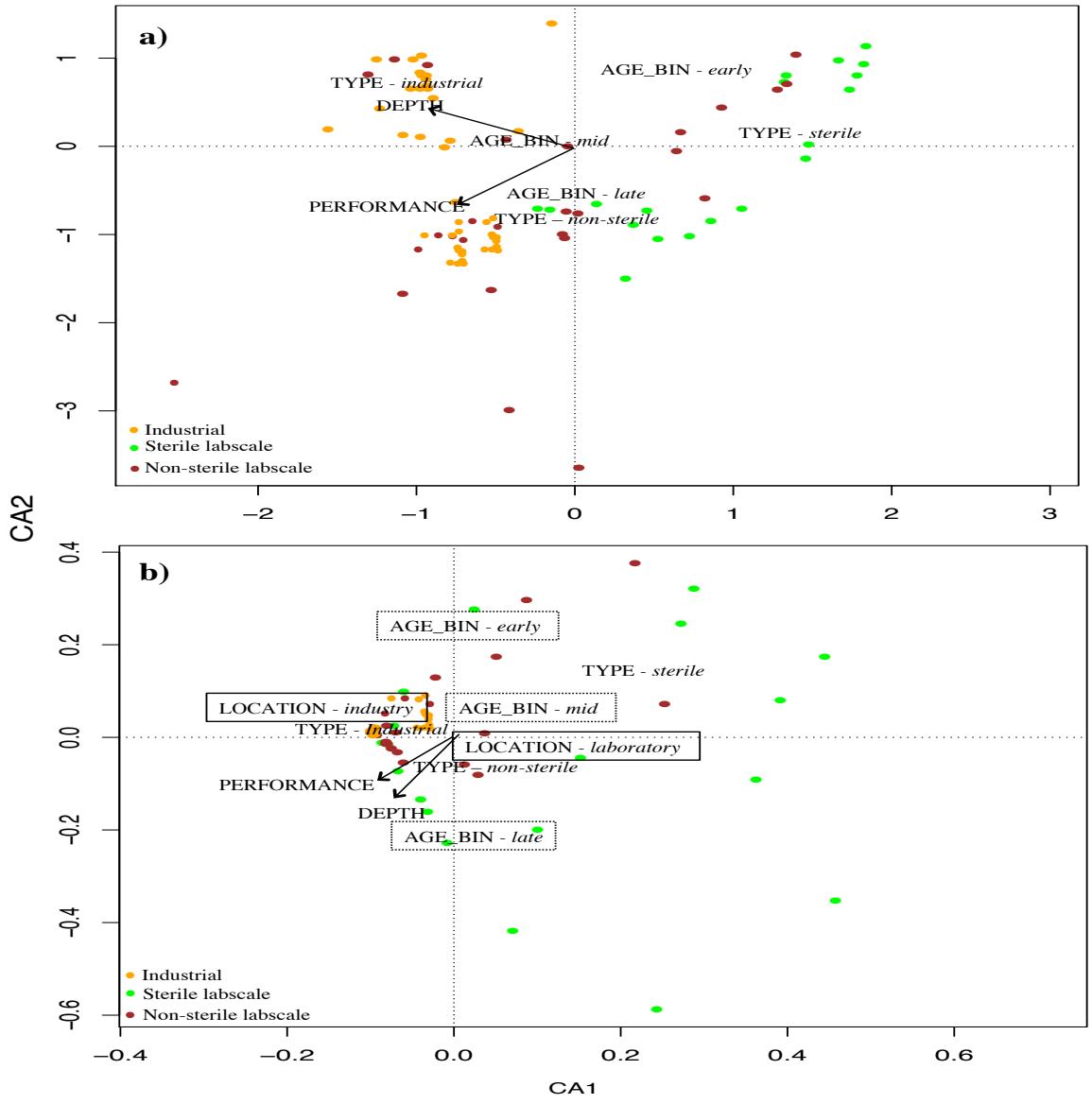


Figure 5.13: Canonical correspondence analysis of labscale and industrial SSFs at: a) phylum, and b) OTU levels of classification.

This shows that slow sand filtration can be adequately reproduced in the laboratory and with performance comparable to industrial SSFs. Similar to Werner et al. [2011] and the findings of Chapter 4, communities with greater evenness had a higher level of water quality performance, implying greater robustness and functional stability [Wittebolle et al., 2009].

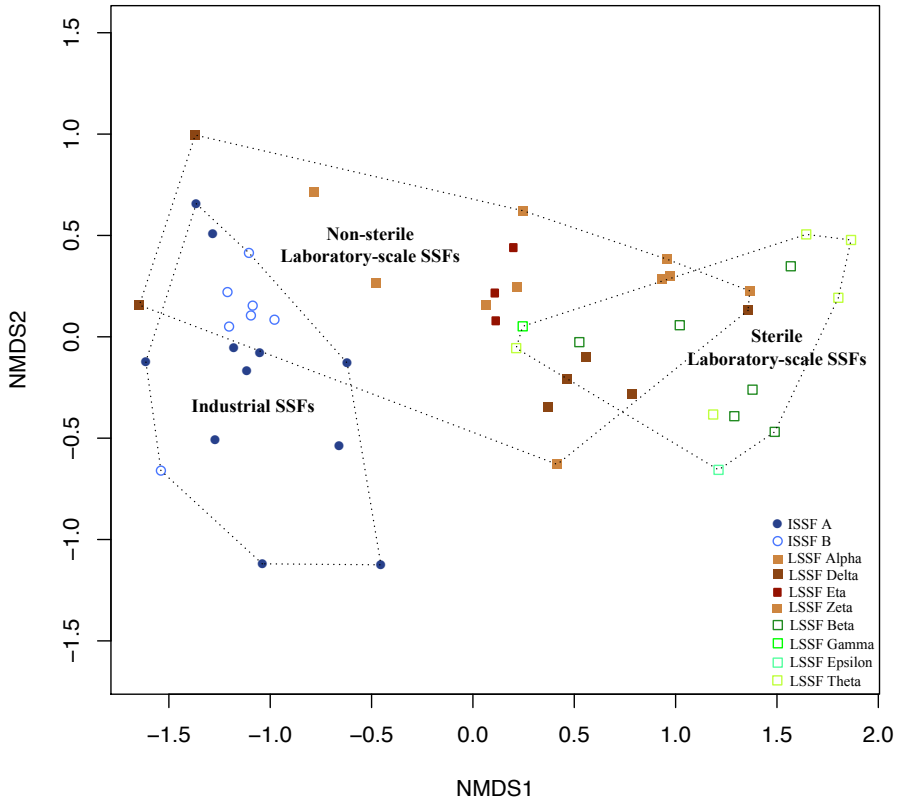


Figure 5.14: NMDS plots of OTUs of mature filters ≥7 weeks of age)

Although differences between non-sterile and sterile laboratory SSFs were found, with non-sterile filters mimicking full-scale filters performance three weeks before sterile LSSFs, the findings nonetheless confirm that a mature microbial community is necessary for optimum operation. Such differences in initial performance can be attributed to the differences in community composition, as depicted by factorial analysis, Shannon indices and the stacked barplots (Figure 5.9), which all show that the community composition between laboratory-scale and full-scale filters are quite different at the beginning of operation (early-age bin). Intriguingly, however, the communities in both the full-scale and laboratory-scale filters are highly similar at phyla and class level of classification by the late-age bin (≥7 weeks of operation). Such convergence in the microbial community composition between labscale and industrial filters is staggering and suggests as Massol-Deya et al. [1997] concluded in their study of fixed film reactors, that the microbial community converges towards a structure which is best for optimal performance. This is however, the first example to the author's knowledge that demonstrates that functional microbial communities from an engineered system can be reproduced in the laboratory. Further, the reappearance from the ISSFs discussed in Chapter 4 of the importance of the *Comamonadaceae* and *Sphingomonadacea* families

when filters age, further strengthens conclusions of their importance in performance and the ability of LSSFs to replicate the functional community of ISSFs. Although there are obvious differences in abundance between these families in the LSSFs and ISSFs, the ratio of individual genera to one another may be important, as was shown for *Bacteriodetes, Gammaproteobacteria,* and *Nitrospira* to *Betaproteobacteria* by Garrido et al. [2014] in varying qualities of wastewater. Future work should investigate through qPCR the absolute number, abundance and ratios of the *Comamonadaceae* and *Sphingomonadacea* families to individual members and the whole community, in particular *Acidovorax, Sphingomonas* and *Sphingobium* in order to confirm their importance to overall filter performance ($\bigtriangledown$).

Although good reproduction of the microbial community composition of ISSFs was observed in LSSFs at higher levels of classification, this was not true at the OTU level, where there were significant differences at all time points. Although filter location (laboratory or industry) was observed to be an insignificant variable at higher taxonomic classification, it was found to be the most important factor in explaining differences in community structure at the OTU level. The significance of location may be explained by the lack of environmental variables (e.g., wildlife and climatic effects) in the laboratory that may add significantly more diversity to ISSFs. This finding is supported by the higher proportion of shared OTUs between influent water and the sand community in LSSFs compared to full-scale slow sand filters. Additionally, differences in shared OTUs may also be an age-dependent effect as full-scale filters have been exposed to their water source for a longer period of time compared to LSSFs. Hence, ISSFs have a higher initial diversity and share fewer OTUs with the current water supply. This may explain the higher proportion of shared OTUs between sterile LSSFs and the influent water source compared to non-sterile LSSFs. Further, another factor which may explain differences at OTU level is the reduced width to height ratio of the laboratory-scale filters compared to ISSFs. It is widely known that different bacterial species (OTUs) survive and occupy specific niches, some of which can be very small (i.e., the voids between two sand grains). It is also known that within microbial communities, micro-, meso- and macro-scale variation exists [Nunan et al., 2002] which is often driven by environmental and chemical gradients [Nunan et al., 2003]. Therefore, it is possible that the reduced similarity at the OTU level between LSSFs and ISSFs may be due to the reduced dimensions of the laboratory-scale filters, therefore producing differences in chemical gradi-

ent distribution, and hence niche formation. Further attention could be aimed to examine the effect on the microbial community of various width-to-height ratios of LSSFs to determine if convergence in community composition can be achieved at genus and OTU level with LSSFs with a bigger diameter.

Such differences in community composition at higher levels of classification are not surprising and alongside the finding that there was no difference in composition between individual filters, implies that community assembly is not random. This provides further evidence of niche theory [Sloan et al., 2007]. However, the divergence of community composition at the OTU level supports neutral theories [Sloan et al., 2007], which state that different OTUs can perform the same functions. This, in-conjunction with the support for niche theory at higher taxonomic classification, underscores the complexity of microbial community assembly and implies that a combined neutral and niche theory may best explain microbial community assembly in SSFs. Although there is growing support for the combined neutral and niche theory explanation for microbial community assembly, in order to determine which mechanism is more important within SSFs community assembly, more experimentation is needed. In particular, by combining the data presented in this chapter and in chapter 4 with additional characterisation surveys a large dataset encompassing multiple environmental variables will be produced, which will allow robust statistical analysis and interpretation to be gained. Such analysis will be able to determine the extent to which niche and neutral processes explain community composition. For example if neutral processes dominate we would predict that:

1. The SSF microbial species abundances will fit the zero-sum multinomial (ZSM) distribution [Hubbell, 2001].

2. Changes in the SSF community composition will be related to the distance between samples, indicating the effects of dispersal limitation.

3. There will be no relationship between the SSF community and either water or sand properties.

However, if niche-based mechanisms dominate we would predict that:

1. The SSF microbial species abundances will fit a log-normal or other niche-based species abundance distribution [McGill et al., 2007].

2. Changes in the SSF community composition will be related to changes in either the water or sand properties [Jongman et al., 1995]

3. The SSF microbial community will not be structured by distance effects [Dumbrell et al., 2010].

5.5 Conclusions

The bacterial communities of slow sand filters are extremely rich in taxa; not dominated by any particular phylogenetic group; and exhibit spatial and temporal changes. This study is the first to demonstrate that it is possible to recreate this complex and rich SSF community in the laboratory at phylum, class and order levels. Based on this study, it is now possible to use these laboratory scale SSFs to ask more complex questions relating to water quality and community assembly. However, it should be noted that although conclusions drawn from LSSFs are applicable to ISSFs they might not accurately reflect OTU dynamics. Whilst the laboratory and industrial filters appear identical at a coarse taxonomic level, when OTUs (species proxies) are considered they differ consistently. Given that they also appear equivalent in terms of function, a degree of redundancy is suggested. Neutral theory proposes that different species level community compositions are possible even in communities with the same distribution of families and with the same functional capacity. It is unclear what is driving the differences at the OTU level, but it is likely environmental differences such as temperature and the presence / absence of wildlife between the laboratory and the industrial filters have a part to play. However, this requires further study. Ultimately, by utilising the laboratory-scale SSFs designed in this study, further investigation into the individual SSF community members and their relationship with water quality performance can be explored. This will allow optimised and tailored operation and design of full-scale slow sand filters for specific water quality needs and requirements. Furthermore, this provides a paradigm for future microbial ecology studies aimed at understanding and modelling microbial community assembly.

Chapter 6

Shedding Light on Pathogen Removal in SSFs

"Tell me what you eat, and I shall tell you what you are".

Jean Anthelme Brillat-Savarin, (Physiologie du goût – The Physiology of Taste)

Stable-isotope probing (SIP) and metagenomics were applied to study samples taken from laboratory-scale SSFs (described in Chapter 5) 0.5, 1, 2, 3, and 4h after challenging with ^{13}C-labelled *E.coli*, in order to determine the mechanisms and organisms responsible for coliform removal. Further, this work focuses on the effects of light on the microbiome and functioning of slow sand filters. In order to do this the laboratory-scale filters described in Chapter 5 were supplied with influent water from the River Kelvin for 70 days. Half the filters were in total darkness and the other half were exposed to a 12h daylight 12h darkness period mirroring light intensities found in summer at the industrial SSF site discussed in Chapter 4. After 35 days the filters were challenged for 1h with ^{13}C labelled *E.coli*. In order to determine the mechanisms responsible for *E.coli* removal, sand samples were taken from different depths and multiple time points after challenging and used for SIP and metagenomics analysis. Direct counts and qPCR assays revealed a clear predator-prey response between protozoa and *E.coli*. The importance of top-down trophic-interactions was confirmed by metagenomic analysis, identifying several protozoan and viral species connected to *E.coli* attrition, with protozoan grazing responsible for the majority of the removal. In ad-

A condensed version of this chapter has been accepted by the ISME journal: Haig, S. Schirmer, M. D'Amore, R. Gibbs, J. Davies, R.L. Collins, G. and Quince, C. Stable-Isotope Probing and Metagenomics Reveal Predation by Protozoa Drives *E.coli* Removal in Slow Sand Filters

dition to top-down mechanisms, indirect mechanisms such as algal reactive oxygen species induced autolysis, and mutualistic interactions between algae and fungi, were suggested to be associated with coliform removal. Further, the top upper depths of the SSFs were found to be responsible for the majority of the coliform removal in both light and dark filters, however, at different depths, coliform removal occurs more quickly in light filters in comparison to dark conditions where *E.coli* still appears after 96h. The differences in the abundance of labeled *E.coli* at various depths in light and dark filters implies that a light environment is more hostile in terms of *E.coli* survival. These findings significantly further our understanding of the processes and trophic interactions underpinning *E.coli* removal. This study provides an example for similar studies, and the opportunity to better understand, manage and enhance *E.coli* removal by allowing the creation of more complex trophic interaction models.

6.1 Introduction

In Chapter 5 laboratory-scale filters have been compared to full-scale units using phylum-specific qPCR primers and both 454 and Illumina sequencing (Earth Microbiome Project). This revealed that the microbial communities underpinning slow sand filters are extremely complex, with specific organisms correlating with certain water quality parameters (e.g., $\bigtriangledown$). However, such complexity and performance was shown to be reproducible in the laboratory and now allows more pertinent questions pertaining to human health and microbial ecology to be addressed. Therefore, from a functional perspective, understanding how pathogenic microorganisms (e.g., *E.coli*) are removed is a critical question. This will address one of the primary tasks of modern ecology; linking the biotic interactions of organisms within an ecosystem to their functional performance [Mikola and Setälä, 1998]. Determining the mode(s) of removal would be highly advantageous and potentially allow water companies to control *E.coli* levels by managing the slow sand filter microbial community. Further by exploring how abiotic factors such as the presence of sunlight affect such performance will provide great insight and knowledge into how these "real world" microbial food-webs survive and perform in differing environments.

The need to remove pathogens and understand the mechanisms responsible for pathogen

removal in potable water supplies is a well-recognised issue, emphasised by the fact that approximately 3.4 million people each year die from water-related disease [World Health Organisation, 2004]. Determining and understanding these mechanisms would be highly advantageous and would vastly improve the implementation of drinking water technologies in developing countries, including household systems. In addition, it could allow water companies in developed countries to control pathogen levels by directly managing the slow sand filter community. Further, by determining the trophic mechanisms and interactions involved in *E.coli* removal in a "real world" food-web, great insight and knowledge for general microbial ecology will be obtained. This will provide a paradigm for similar studies and the opportunity to create more realistic trophic interaction models in the future.

Previous SSF studies have examined the ability of specific organisms (e.g., Chrysophyte) to remove pathogenic bacteria [Weber-Shirk and Dick, 1999], or the overall pathogen removal efficiency of SSFs [Bomo et al., 2004, Grobe et al., 2006, Hijnen et al., 2007, Elliott et al., 2008]. However, these studies are limited by their specificity. Further, based on these studies, and knowledge from marine and terrestrial environments, both top-down (predation by protozoa and viral lysis) and bottom-up (nutrient / resource availability) mechanisms have been suggested to be important for the regulation of microbial mortality [Lloyd, 1973, Hunter and Price, 1992, Pace and Cole, 1994, Weber-Shirk and Dick, 1999, Rosemond et al., 2001]. Additionally, theoretical models and empirical surveys have indicated that majority of the mortality is due to grazing by protists, and to a lesser extent viral lysis [Pernthaler, 2005]. However, abiotic factors, such as UV radiation and reactive oxygen species (ROS)-associated lysis, have also been hypothesised as potential lysis routes for microbes / pathogens [Curtis et al., 1992, Alonso-Sáez et al., 2006, Liu et al., 2007, Kadir and Nelson, 2014].

Although these studies are informative, they are also unrealistic as they have been performed in microcosms, focussing on one or a small group of organisms and hence over-simplify and potentially provide inaccurate or biased conclusions on regulatory mechanisms. Currently, no study, to the author's knowledge, exists which aims to determine the mechanisms responsible for pathogen removal in a real biological system without prior knowledge as to which removal mechanisms or organisms to target. However, the complexity of real communities requires an untargeted approach capable of quantifying the importance of all trophic groups

simultaneously. Likewise various studies have looked at the performance effects of covered and uncovered SSFs (i.e. exclusion of light) [Campos et al., 2002, Abrahamsson and Dromberg, 2006]; however such studies have never looked at differences and similarities within the microbial community. Here, we develop such an approach by combining stable-isotope probing with metagenomics [Sul et al., 2009] and apply it to the tractable, though complex system in SSFs, allowing all mechanisms and organisms involved in the removal of non-pathogenic *E.coli K12* to be determined. We will use this organism, a commonly used faecal indicator, as a proxy for true pathogens, such as other *E.coli* strains, making the assumption that the removal processes will be the same. The experiment was used to test the hypothesise that the principal modes of removal will be top-down removal mechanisms, such as predation by protozoa and viral lysis, although the extent of these processes is expected to differ throughout time and in different filter conditions.

Hypotheses

Hypothesis 6.1 *Filters exposed to light will perform better in terms of operation ability (fulfil more of the water quality parameters outlined in the European Union Directive) quicker than those in darkness.*

Hypothesis 6.2 *The community of microbes in covered SSFs will be less diverse and have lower biomass than the non-covered SSFs.*

Hypothesis 6.3 *Viral lysis is the key mode of E.coli removal in SSFs*

Hypothesis 6.4 *Majority of the E.coli removal will occur at the top of the SSFs.*

6.2 Materials and Methods

6.2.1 Filter Set-up and Operation

The same slow sand filter set-up (eight filters of 2.5m in height and 54mm in diameter) and operational procedures, as employed in Section 5.2 were used in this study. The only difference was the addition of high-power LED lights fitted with a cool white (240 lumens or 5.4W) bulb, erected above half of the SSFs to simulate daylight conditions, similar to those found at the full-scale SSF site (Chapter 4). These lights functioned on a 12h light /

12h dark cycle for the duration of the experiment, with times being regulated by a digital electronic timer, these filters will be referred to as "uncovered SSFs". The remaining filters were covered in metal foil and black-out curtains to prevent any light reaching them and will be referred to as "covered SSFs". The source of water feeding each of the filters was the River Kelvin in Glasgow, and was supplied at a constant rate of $0.15 m^3 m^{-2} h^{-1}$, which is consistent with full-scale SSF operation. The sand used in all filters was sourced from the full-scale SSF discussed in Chapter 4 and was sterilised by autoclaving at 121°C for 20min prior to being put into the eight laboratory-scale SSFs.

6.2.2 Spiking the Filters with Isotopically-Labelled *E.coli*

After seven weeks of operation, each of the filters were spiked with isotopically-labelled *E.coli K12* (Section 6.2.2.1 for strain details) following the protocol outlined in Marley et al. [2001]. Briefly, *E.coli K12* was grown overnight in M9 minimal medium (Table 6.1) with 20ml of filter-sterilised 20% (w/v) ^{13}C-glucose (Sigma) as the sole carbon source at 37°C, with shaking at 200rpm. The overnight culture was then centrifuged at 3000*g* for 10min and washed twice with sterile PBS before being resuspended in autoclaved river water to a density of 300cfu/ml, 5 min before spiking into the SSFs. Spiking entailed feeding the isotopically-labelled *E.coli* to all filters for one hour at the same filtration rate used previously ($0.15 m^3 m^{-2} h^{-1}$), after which time normal filter operation resumed with non-spiked, non-autoclaved river water. The concentration of *E.coli* used was approximately ten times the normal concentration found in the river water and was chosen to mimic levels found during pollution events and storm run-off events.

6.2.2.1 *E.coli* Strain

The *E.coli* strain used for this work is the non-pathogenic K12 strain (TOP10) which is commonly used for cloning and for which the full genome is available[1]. The genomic composition of TOP10 is: F^- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 ΔrecA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG. The genome contains the following defective prophages: CP4-6, DLP12, E14, Fels-1, Rac, Qin, and λ80 [Durfee et al., 2008]. Once transformed with the pCR4-TOPO plasmid the strain additionally becomes Ampicillin and Kanamycin resistant. In order to verify that the strain retained its antibiotic

[1] http://www.genome.jp/dbget-bin/www_bget?refseq+NC_010473

Table 6.1: Chemical Composition of M9 Minimal Media: (A) Composition of M9 Minimal Media (B) Composition of 5x M9 Salts

A)

Component	Amount in 1L
5X M9 Salts	200ml
1M $MgSO_4$	2ml
1M $CaCl_2$	0.1ml
Water	777.9ml
Ampicillin (5mg/ml)	20ml
Kanamycin (5mg/ml)	10ml
Streptomycin (5mg/ml)	5ml

B)

Components	Quantity in 1L
$Na_2HP0_4.7H_2O$	64g
KH_2P0_4	15g
NaCl	2.5g
NH_4Cl	5g

resistance for the entirety of the experiment, the transformed *E.coli* strain was placed in the same autoclaved river water used for the experiment. Viability was tested on membrane lauryl sulphate agar (MLSB) and LB agar plates supplemented with all three antibiotics every 12h for 120h. The transformed strain was found to be viable for $>$120h, this alongside the absence of growth of any other water dwelling bacteria on LB, Plate Count Agar or MLSB containing the three antibiotics possessed by the transformed strain (Ampicillin, Kanamycin, and Streptomycin) confirms that any bacterial counts observed on this media are the transformed strain.

6.2.3 Sampling Spiked Filters

To determine the mechanisms responsible for *E.coli* removal, sand was sampled from the filters at depths (1, 5, 10, 15 cm) and times of 0.5h, 1h, 2h, 3h, and 4h after spiking begun. Additionally, all depths (0, 5, 10, 15, 20, 30, 45, 70 cm) were sampled from the filters 24h and 96h after spiking. Sand samples (0.5g wet weight) were used for; direct *E.coli* plate counts (Section 6.2.3.1) on MLSB containing: 100 μgml^{-1} ampicillin, 50 μgml^{-1}, kanamycin, and 25 μgml^{-1} streptomycin (Life Technologies, UK); direct protozoa quantification (Section 6.2.3.2); and DNA extraction (Appendix B) with the downstream application of qPCR (Section 6.2.4), and stable-isotope probing in conjunction with metagenomic sequencing (Section 6.2.6).

Table 6.2: Additional qPCR primers used for SIP study

Primer	Target	Sequence (5'-3')	Annealing Temp (°C)	Amplicon Length (bp)	Reference
EcoliF	E.coli 16S	CAATTTTCGTGTCCCCTTCG	58	450	Khan et al. [2007]
EcoliR	E.coli 16S	GTTAATGATAGTGTGTCGAAAC	58	450	Khan et al. [2007]
Euk567F	All 18S	GTTAAARVGYTCGTAGTY	56	517	Bass [2013]
Euk1084r	All 18S	CCGTCAATTHCTTYAART	56	517	Bass [2013]

6.2.3.1 Quantifying Labelled Coliforms

In order to determine the survival and extent of the viability of the spiked *E.coli* in the different depths of the SSFs, 0.5g sand samples were placed in 1ml of PBS and centrifuged at 3000g for 10 min. The resulting supernatant was used for plate counts on MLS agar with 100μgml^{-1} ampicillin, 25μgml^{-1} streptomycin and 50μgml^{-1} kanamycin, each sample was performed in replicas of five. Additionally, after 12h, 24h and 96h after spiking effluent water samples from all the filters were tested for *E.coli* viability in the same manner.

6.2.3.2 Quantifying the Abundance of Protozoa

As protozoan grazing is believed to be a major factor influencing bacterial populations in water systems, it is appropriate that we monitor changes in their abundance during the experiment. This was achieved by adapting the procedure outlined by [Dehority, 1984]. Briefly, sand samples were fixed with 10% formalin (1:1 weight:volume) for 2h at room temperature. After fixation samples were centrifuged at 500g for 10min and the resulting supernatant was stained with glycerol buffered methyl green (1:20 final dilution) for 12h. Three 100μl stained samples were then loaded into a haemocytometer and visualised on an inverted optical microscope (Inverted Olympus IX71) at 100x magnification, results were expressed as cells/g (wet weight). Additionally, the same procedure was performed for effluent water samples, with the omission of the centrifugation step, results were expressed as cells ml^{-1}

6.2.4 qPCR

All samples were processed in triplicate for the 13 different qPCR assays described previously (Section 4.2.6), with the addition of two assays for *E.coli* specific 16S and total 18S rRNA (Table 6.2), both of which were performed using 35 cycles.

6.2.5 DNA-Stable-Isotope Probing (DNA-SIP)

To separate the labelled (^{13}C) from the unlabelled (^{12}C) DNA the procedure of Neufeld et al. [2007] was used, complete details of which can be found in Appendix C. Briefly, separation was achieved by using density gradient fractionation of the total DNA extract (50μL) on a CsCl gradient with a buoyant density of 1.725 g/ml that was subjected to ultracentrifugation in a Sorvall 100SE Ultracentrifuge (Thermo Scientific) at 44,100rpm for 40h at 20°C. The density gradient was fractionated into 12 aliquots (approximately 400μL each) by a drop-wise collection method, where fractions were taken from the bottom of the ultracentrifugation tube by pumping water into the top of the tube with a constant-flow (500μL/min) syringe pump (Gilson's Miniplus 2 peristaltic pump). The resulting fractions density were measured with an AR200 refractometer (Reichert) and ranged from 1.47 to 1.73g/ml with a median density of 1.68g/ml. Fractions were precipitated using a polyethylene glycol solution and dissolved in 30μL of TE buffer, and used for qPCR quantification of 18S rRNA, total 16S and *E.coli* specific 16S rRNA genes (Table 6.2). Based on qPCR and density profiles of the samples compared with ^{12}C and ^{13}C controls, two fractions from each sample, one representing labelled (density: $>$1.68g/ml) DNA and one representing non-labelled (density $<$1.68g/ml) DNA, were chosen for metagenomic library construction and analysis (Figures 6.1 and 6.2).

6.2.6 Illumina Metagenomic Library Preparation on SIP Samples

Thirty-six Illumina libraries (eighteen pairs of labelled ^{13}C and non-labelled ^{12}C fraction from various filters and time-points) were prepared using the Nextera XT kits (Illumina, UK), following the manufacturer's instructions. Briefly, 5μL (0.2ng/μL) of extracted DNA were tagmented and then subjected to PCR using specific index primers and common adaptors (P5 and P7). Amplified libraries were cleaned-up using the AMPure XP beads (Beckman Coulter, UK) and eluted in a final volume of 12μL TE. Libraries were checked for their fragment size distribution and concentration using a Bioanalyzer 2100 (Agilent, UK), and appropriate libraries were size selected (500-800bp) using a Pippin Prep (Sage Science, USA) using a 1.5% cassette. Size selected libraries were pooled using equimolar quantities in order to obtain the desired number of reads for each sample. The pool was sequenced on a HiSeq 2000 (Illumina, UK) at the Centre for Genomic Research (Liverpool).

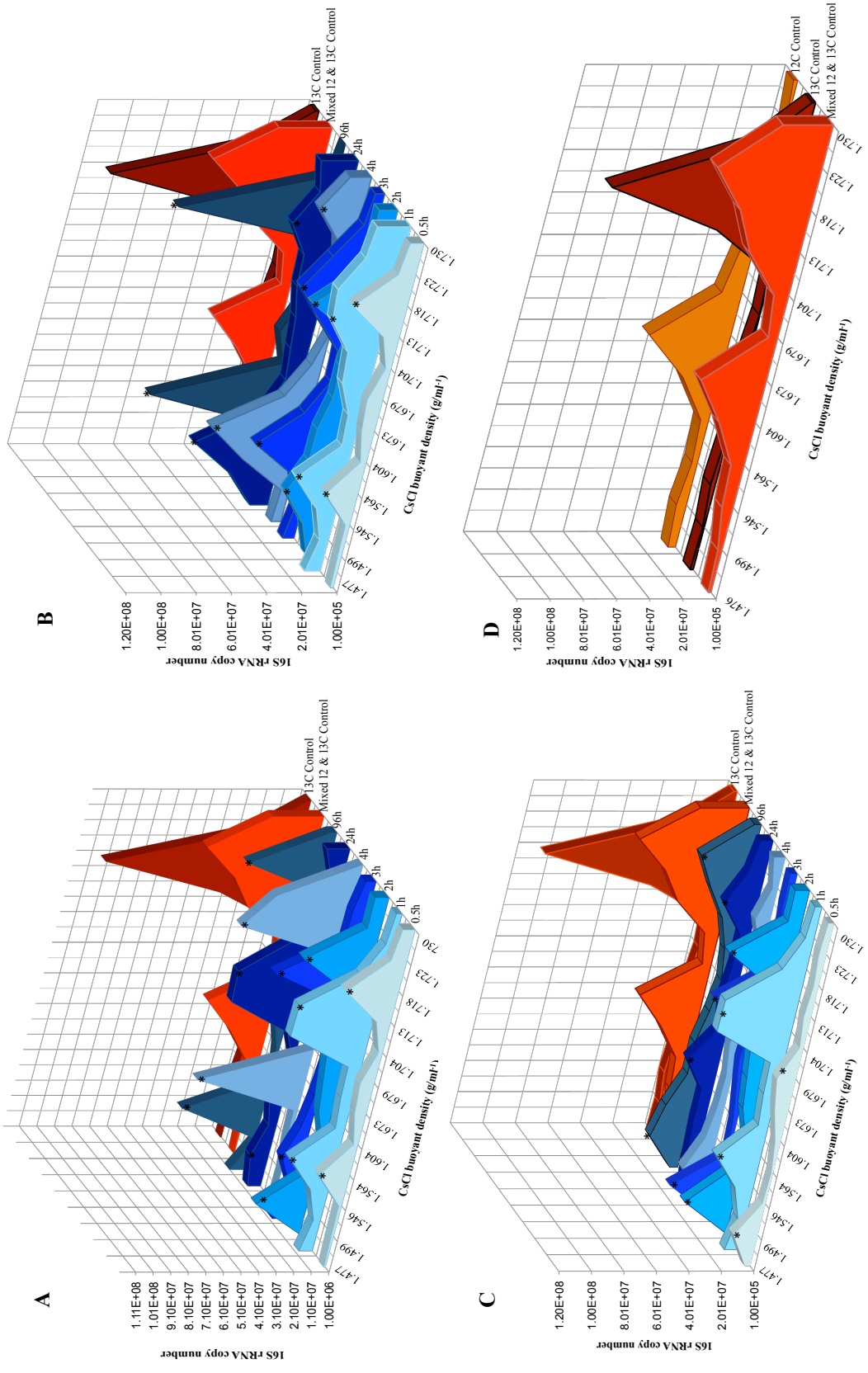


*Figure 6.1: Quantitative profiles of total 16S rRNA distribution in SIP fractions 0.5, 1, 2, 3, 4, 24 and 96h after exposure to ^{13}C labelled E.coli. **A:** SSF C, **B:** SSF D, **C:** SSF G and **D:** Controls. ∗ designate fractions chosen for metagenomics.*

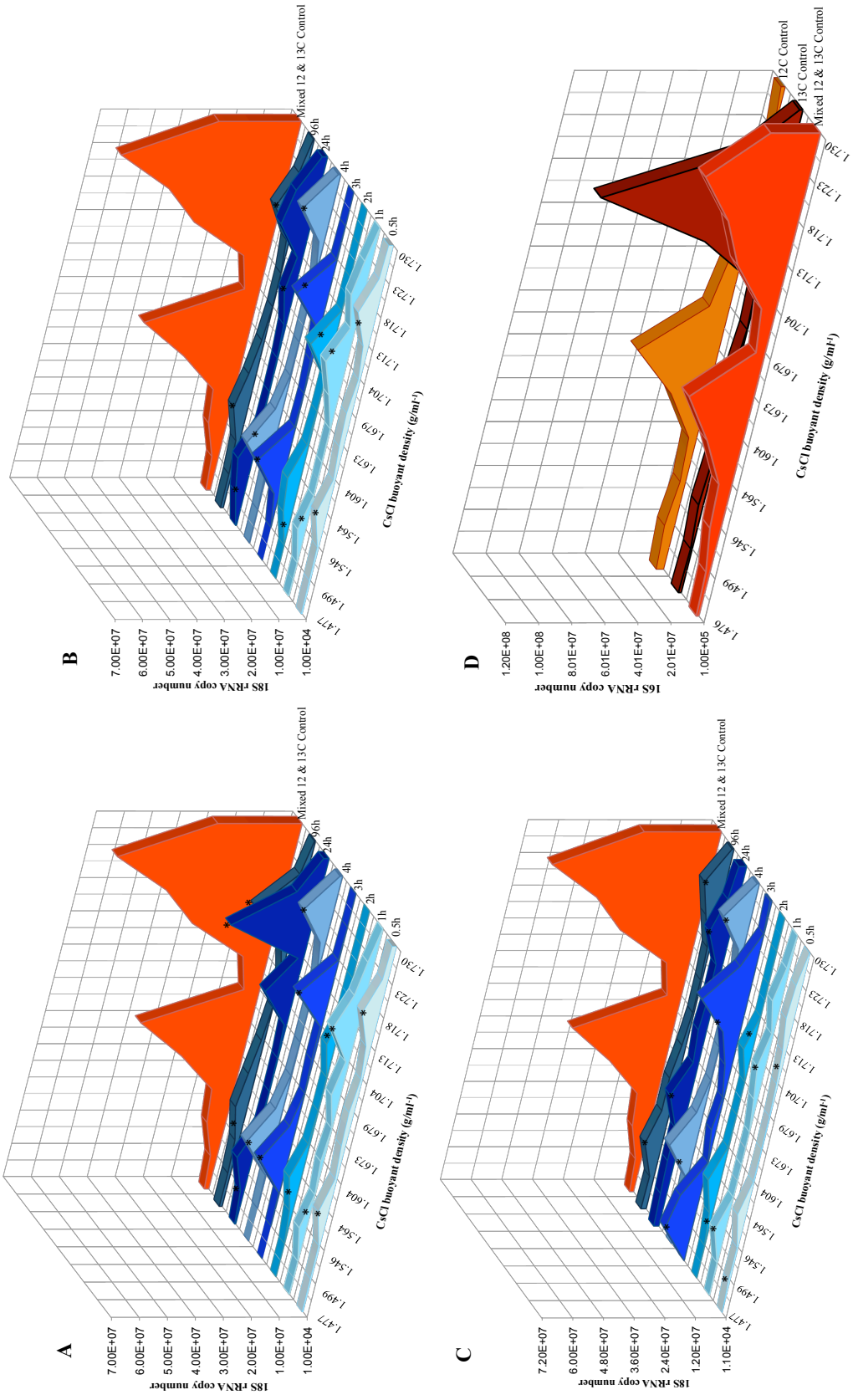


*Figure 6.2: Quantitative profiles of total 18S rRNA distribution in SIP fractions 0.5, 1, 2, 3, 4, 24 and 96h after exposure to ^{13}C labelled E.coli. **A:** SSF C, **B:** SSF D, **C:** SSF G and **D:** Controls. $*$ designate fractions chosen for metagenomics.*

6.2.7 Metagenomic Sequence Analysis

Sequenced reads for each sample were quality trimmed using sickle[2]. Quality profiles were constructed with FastQC [Andrews, 2010], which revealed a non-uniform distribution of nucleotides at the start of the reads, indicating the possible partial remainder of adapter or transposon sequence. Therefore the first 20bp of the MiSeq and 16bp of the HiSeq reads were trimmed and reads were filtered based on a minimum read length of 80bp and 40bp for MiSeq and HiSeq respectively (originally MiSeq: 150bp and HiSeq: 99bp). Resulting samples contained on average 8,157,287.

For the taxonomic classification, paired-end reads were converted to a format suitable for analysis using MEGAN [Huson et al., 2011]. LAST [Frith et al., 2010] was used to align the reads (max. of 20 matches) against a customised subset of the NCBI database containing the microbial, protozoan, viral and fungal databases to achieve a more time-efficient analysis. The output was converted into "blast format" and piped into MEGAN where the lowest common ancestor (LCA) was assigned to each read (LCA parameters: max-matches=100, minscore=35.0, toppercent=10.0, winscore=0.0, minsupport=1, mincomplexity=0.3). Occurrence tables of the taxonomic assignment were generated using a custom designed script[3] in which the last column of the MEGAN output files was converted into the corresponding taxid and the taxonomic path was inferred by utilising the perl library Bio::LITE::Taxonomy. Directly exporting taxonomic paths with MEGAN caused problems in particular for eukaryotes and viruses as not all taxonomic levels were defined. Additionally, Bio::LITE::Taxonomy was unable to resolve issues due to synonyms present in the database. Therefore, the taxonomic paths of all "taxids" which were unresolved by Bio::LITE::Taxonomy were directly inferred from the NCBI taxonomy database (names.dmp). Furthermore, the full taxonomic paths of several organisms (e.g., *Monosiga brevicollis* and *Dictyostelium discoideum*) were added to the database.

6.2.8 Statistical Analysis

Significant differences in the microbial community composition determined by qPCR analysis between covered and uncovered filters, age, and depth were identified by using permu-

[2] www.github.com/najoshi/sickle

[3] Designed by Melanie Schirmer

tational multivariate analysis of variance tests (MANOVA) [Anderson, 2001]. The contribution of individual taxon to overall community dissimilarity was determined using SIMPER analysis [Clarke, 1993]. SIMPER analysis is a useful measure of the magnitude of difference, however in order to decide whether taxa differed significantly pairwise t-tests (kendall non-parametric) adjusted for multiple comparisons using the Benjamini-Hochberg false-discovery method were performed. Only taxa with a false-discovery rate of less than 5% [Benjamini and Yekutieli, 2001] were reported. The relationships between environmental variables and patterns in bacterial community structure were examined by canonical correspondence analysis with significance tested by ANOVA after reducing the overall suite of environmental variables using a step-wise AIC model. To determine which organisms played a significant role in *E.coli* removal, pair-wise similarities among samples based on the Bray-Curtis similarity index were calculated. The resulting matrices were examined for temporal patterns and differences between ^{13}C (labelled) and ^{12}C (non-labelled) samples using NMDS and CCA analysis as implemented in the Vegan package for R [Oksanen et al., 2012]. Significant differences in the metagenomic community composition between different time points (0.5-4h) and carbon sources (labelled - ^{13}C and non-labelled - ^{12}C) after spiking with *E.coli* were determined using an MANOVA. To determine individual contributions from each taxon to the differences between labelled and non-labelled samples, and for the various time points, SIMPER analysis alongside pairwise t-tests adjusted for multiple comparisons using the Benjamini-Hochberg false-discovery method were performed.

6.3 Results

In this section the results of the study will be presented alongside discursive analysis, with an overall discussion being provided in Section 6.4.

6.3.1 Water Quality of Covered and Non-covered SSFs

Analysis of the water quality parameters and overall water quality performance ($\bigtriangledown$) has shown that like the industrial filters (discussed in Chapter 4), and labscale filters (discussed in Chapter 5) the covered and non-covered LSSFs studied in this chapter produce good quality water (Tables 6.3 - 6.5 show the influent and the effluent from the filters). No significant difference was found between covered and non-covered LSSFs' $\bigtriangledown$ (p-value:0.7806). This

Table 6.3: Summary of the physical and chemical characteristics of the influent water (from the River Kelvin) supplying covered and non-covered LSSFs.

Average (bases on 5 replicas) water quality parameters for the influent water feeding all labscale SSFs

Date Sampled	Additional Info	TEMP	PH	NTU	DO	DOC	UVA	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB_13	COLIFORM
30/01/2012		14.88	7.48	3.780	12.020	5.700	0.325	5.702	15.867	0.140	0.490	0.035	0.238	8000	3570	102
06/02/2012		14.8	7.5	2.456	12.000	4.626	0.265	5.720	6.620	0.100	0.507	0.018	0.026	14000	340	40
13/02/2012		15.06	7.6	3.560	11.000	4.933	0.247	5.000	9.817	1.150	0.640	0.030	0.021	8860	580	58
20/02/2012		15.4	7.2	2.340	10.500	6.500	0.307	4.718	10.767	0.080	0.340	0.013	0.083	5540	380	14
27/02/2012		14.6	7.6	4.808	9.500	4.867	0.308	6.329	33.233	0.090	0.233	0.018	0.101	13500	3840	20
05/03/2012		16.4	7.4	3.410	8.524	3.870	0.306	7.916	14.500	0.183	0.300	0.028	0.026	11660	1140	84
12/03/2012	Spiked with 13C labelled *E.coli*	15.7	7.6	5.720	9.560	4.000	0.308	7.708	9.580	0.100	0.290	0.008	0.012	14660	2950	490
19/03/2012		14.84	7.8	1.680	10.150	4.200	0.284	6.762	7.450	0.090	0.150	0.002	0.020	12880	340	24
26/03/2012		15.36	7.8	3.930	10.196	4.283	0.297	6.934	11.600	0.390	0.150	0.003	0.068	6380	240	36
03/04/2012		17.24	7.5	1.772	9.282	4.307	0.302	7.005	16.300	1.057	0.293	0.036	0.015	45200	440	26

Table 6.4: Summary of the physical and chemical characteristics of the effluent water from non-covered LSSFs.

Average (bases on 5 replicas) water quality parameters for the effluent of non-covered (exposed to 12h light: 12h darkness) labscale SSFs

Filter	Additional Info	Sampled	Age	TEMP	PH	NTU	DO	DOC	UVA	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB_13	COLIFORM	Performance Metric (▽)
C		30/01/2012	4	15.58	7.5	1.556	11.180	7.467	0.323	4.330	17.067	0.073	0.290	0.040	0.412	2454	2496	608	7
C		06/02/2012	11	14.98	7.7	1.770	11.000	4.430	0.274	6.180	26.533	0.110	0.687	0.524	0.296	27600	12920	836	5
C		13/02/2012	18	15	7.8	0.818	10.000	4.140	0.257	6.208	11.503	0.183	0.507	0.022	0.051	12380	780	242	6
C		20/02/2012	25	15.62	7.6	0.718	8.923	4.287	0.311	7.247	15.067	0.100	0.293	0.010	0.077	20180	16180	120	7
C		27/02/2012	32	14.7	7.8	1.532	9.210	4.853	0.298	6.147	13.900	0.060	0.187	0.030	0.015	14840	760	26	8
C		05/03/2012	39	16.6	7.6	1.030	8.416	8.833	0.293	3.313	12.733	0.220	0.273	0.007	0.008	6800	84	8.2	8
C	^{13}C labelled *E.coli added*	12/03/2012	46	15.8	7.6	0.625	9.000	8.963	0.297	3.317	19.700	0.120	0.263	0.070	0.099	11400	740	8.8	7
C		19/03/2012	53	15.08	7.88	0.572	9.907	6.033	0.295	4.895	6.363	0.560	0.200	0.044	0.018	7780	92	7.2	7
C		26/03/2012	60	15.66	7.74	0.646	10.000	6.100	0.299	4.896	11.200	0.117	0.140	0.018	0.002	5580	74	0.4	7
C		03/04/2012	68	17.52	7.6	0.336	9.000	6.000	0.299	4.983	14.667	0.250	0.200	0.015	0.015	43000	160	0	8
D		30/01/2012	4	15.88	7.5	1.556	11.220	7.533	0.325	4.314	17.000	0.200	0.293	0.044	0.412	2924	2828.8	532	6
D		06/02/2012	11	15.02	7.92	1.668	11.080	4.450	0.264	5.928	26.433	0.263	0.807	0.571	0.297	26000	30240	800	5
D		13/02/2012	18	14.9	7.8	0.820	10.000	4.140	0.259	6.264	11.267	0.153	0.460	0.085	0.050	22920	12900	112	7
D		20/02/2012	25	15.6	7.6	0.719	9.000	4.297	0.309	7.192	15.100	0.117	0.303	0.076	0.076	21600	6460	12	6
D		27/02/2012	32	14.7	7.8	1.070	9.200	4.850	0.299	6.158	13.767	0.060	0.180	0.008	0.015	22280	3260	10	8
D		05/03/2012	39	17	7.7	1.080	8.416	4.407	0.294	6.679	13.867	0.100	0.273	0.017	0.008	7300	320	6.8	8
D	^{13}C labelled *E.coli added*	12/03/2012	46	15.72	7.9	0.921	9.100	4.430	0.300	6.780	20.100	0.103	0.270	0.033	0.026	5100	600	10.8	7
D		19/03/2012	53	15.34	7.96	0.730	9.950	4.313	0.297	6.886	3.803	0.847	0.060	0.010	0.006	1880	92	8	7
D		26/03/2012	60	15.6	7.66	0.645	10.010	4.307	0.300	6.966	11.233	0.110	0.140	0.006	0.009	7540	86	0.4	7
D		03/04/2012	68	17.62	7.66	0.607	9.013	4.387	0.300	6.847	14.000	0.160	0.267	0.012	0.016	44800	360	1.8	7
E		30/01/2012	4	15.4	7.6	1.968	11.300	7.533	0.320	4.248	17.233	0.050	0.340	0.040	0.348	2790	2570	422	7
E		06/02/2012	11	14.84	7.6	1.650	11.060	4.440	0.263	5.910	26.400	0.107	1.013	0.545	0.296	28480	16540	578	5
E		13/02/2012	18	14.9	7.7	0.605	10.200	4.150	0.260	6.257	11.447	0.127	0.430	0.026	0.051	10000	840	92	6
E		20/02/2012	25	15.5	7.4	0.618	8.943	4.303	0.296	6.886	14.933	0.100	0.363	0.052	0.077	10000	6300	246	7
E		27/02/2012	32	14.7	7.8	1.054	9.210	4.850	0.298	6.151	13.867	0.060	0.183	0.024	0.015	8400	4020	9.8	8
E		05/03/2012	39	16.9	7.4	1.200	8.400	3.850	0.296	7.680	14.167	0.100	0.270	0.011	0.018	24400	16900	64	8
E	^{13}C labelled *E.coli added*	12/03/2012	46	16	7.9	0.976	9.000	3.930	0.297	7.549	20.000	0.100	0.280	0.040	0.026	9320	5420	126	8
E		19/03/2012	53	15.3	8	0.708	5.964	4.500	0.180	4.000	4.610	0.163	0.280	0.010	0.017	3600	1700	0	8
E		26/03/2012	60	15.6	7.8	0.649	10.020	4.440	0.299	6.727	11.767	0.287	0.140	0.001	0.009	6580	520	0	8
E		03/04/2012	68	17.94	7.76	0.566	9.030	4.333	0.299	6.900	13.800	0.220	0.280	0.037	0.018	44800	320	0	8
F		30/01/2012	4	15.52	7.6	1.494	11.280	7.567	0.325	4.295	17.367	0.053	0.283	0.040	0.350	2404	2612	492	7
F		06/02/2012	11	14.92	7.84	1.738	11.080	4.390	0.265	6.020	26.433	0.173	0.860	0.519	0.296	30000	17780	562	5
F		13/02/2012	18	15.2	7.7	0.713	10.600	4.160	0.260	6.242	11.460	0.157	0.457	0.024	0.051	6260	780	60	6
F		20/02/2012	25	15.44	7.64	0.467	8.877	4.327	0.299	6.911	15.067	0.087	0.363	0.011	0.078	5020	2640	34	7
F		27/02/2012	32	14.84	7.8	0.931	9.230	4.760	0.288	6.057	13.767	0.060	0.180	0.022	0.012	6340	1640	8.6	8
F		05/03/2012	39	17	7.7	1.140	8.420	3.897	0.296	7.588	13.767	0.090	0.300	0.008	0.021	3900	540	6	8
F	^{13}C labelled *E.coli added*	12/03/2012	46	15.9	8	0.670	8.967	3.953	0.293	7.403	19.867	0.100	0.270	0.038	0.027	8000	1280	31.6	8
F		19/03/2012	53	15.34	7.96	0.646	9.960	4.500	0.298	6.615	7.727	0.073	0.280	0.010	0.017	4840	640	0.6	8
F		26/03/2012	60	15.7	7.86	0.403	10.007	4.493	0.300	6.669	11.733	0.190	0.123	0.000	0.009	3880	200	0	8
F		03/04/2012	68	17.88	7.84	0.286	9.043	4.417	0.301	6.815	13.867	0.140	0.280	0.013	0.017	43200	280	0	8

Table 6.5: Summary of the physical and chemical characteristics of the effluent water from covered LSSFs.

Average (bases on 5 replicas) water quality parameters for the effluent of covered (kept in complete darkness) labscale SSFs

Filter	Additional Info	Sampled	Age	TEMP	PH	NTU	DO	DOC	UVA	SUVA	COD	PO4	NO3	NO2	NH4	TVB	TVB 13	COLIFORM	Performance Metric (∇)
G		30/01/2012	4	15.56	7.7	3.636	11.640	8.867	0.296	3.335	23.600	0.227	0.267	0.043	0.410	6890	5684	1992	7
G		06/02/2012	11	14.8	7.5	1.870	11.200	4.390	0.290	6.600	28.167	0.103	0.967	0.504	0.283	32240	34000	1608	5
G		13/02/2012	18	15.16	7.6	0.577	9.900	3.740	0.248	6.640	9.910	0.110	0.540	0.020	0.021	11740	10820	448	7
G		20/02/2012	25	15.5	7.46	0.664	8.890	4.607	0.308	6.679	12.067	0.067	0.357	0.027	0.024	12800	7200	54	8
G		27/02/2012	32	14.8	7.7	1.062	9.230	4.760	0.300	6.296	14.767	0.280	0.130	0.022	0.009	11040	540	8.4	7
G		05/03/2012	39	16.84	7.7	1.180	8.390	4.707	0.295	6.268	12.867	0.240	0.270	0.014	0.009	6180	500	6.4	7
G	^{13}C labelled *E.coli added*	12/03/2012	46	15.76	7.7	0.650	8.900	4.513	0.293	6.484	17.767	0.080	0.273	0.057	0.104	164000	9440	10	7
G		19/03/2012	53	15.4	7.8	0.594	5.976	4.477	0.295	6.597	6.670	0.090	0.240	0.008	0.017	5540	1160	3.6	8
G		26/03/2012	60	15.74	7.72	0.528	9.983	4.483	0.299	6.669	11.100	0.113	0.090	0.003	0.019	3040	152	2	7
G		03/04/2012	68	17.84	7.8	0.342	9.037	4.513	0.300	6.647	13.800	4.600	0.287	0.013	0.005	44200	340	0	8
H		30/01/2012	4	15.62	7.7	2.924	11.160	8.900	0.299	3.363	22.267	0.130	0.290	0.043	0.410	5956	5816	1986	7
H		06/02/2012	11	14.96	7.96	1.392	11.280	4.400	0.290	6.598	28.100	0.130	1.040	0.511	0.284	28100	28000	1462	5
H		13/02/2012	18	15.2	7.6	0.744	10.300	3.800	0.245	6.447	9.920	0.120	0.550	0.019	0.021	7960	10960	462	7
H		20/02/2012	25	15.4	7.9	0.581	8.907	4.597	0.307	6.686	12.167	0.070	0.360	0.017	0.022	5580	980	50	8
H		27/02/2012	32	14.8	7.5	0.801	9.240	4.750	0.303	6.379	14.767	0.060	0.143	0.010	0.012	6630	940	9.6	8
H		05/03/2012	39	17.02	7.8	0.714	8.400	4.757	0.294	6.174	12.600	0.090	0.273	0.018	0.026	5460	400	3.6	8
H	^{13}C labelled *E.coli added*	12/03/2012	46	15.8	7.9	0.407	8.980	4.643	0.296	6.368	17.733	0.140	0.030	0.129	0.039	5240	560	7	7
H		19/03/2012	53	15.5	7.7	0.400	9.967	4.707	0.296	6.296	6.663	0.107	0.200	0.000	0.016	3100	140	3.8	7
H		26/03/2012	60	15.8	7.8	0.421	10.367	4.503	0.299	6.632	11.167	0.093	0.097	0.012	0.048	3200	132	2.4	8
H		03/04/2012	68	17.94	7.7	0.318	9.093	4.580	0.299	6.528	15.300	1.830	0.260	0.026	0.001	44200	180	0	8
I		30/01/2012	4	15.62	7.8	2.900	11.200	8.867	0.297	3.346	20.733	0.100	0.283	0.042	0.411	4432	4354	1852	8
I		06/02/2012	11	14.96	7.66	1.304	11.240	4.400	0.291	6.590	28.333	0.130	1.000	0.524	0.284	34400	31200	1460	5
I		13/02/2012	18	15.3	7.7	0.623	9.980	3.777	0.246	6.514	9.913	0.137	0.543	0.052	0.020	24000	30500	496	7
I		20/02/2012	25	15.5	7.6	0.581	8.653	4.593	0.307	6.691	12.133	0.110	0.360	0.033	0.023	7680	3020	46	7
I		27/02/2012	32	14.82	7.8	0.960	9.220	4.757	0.296	6.223	14.733	0.200	0.167	0.006	0.007	9620	1680	24.8	7
I		05/03/2012	39	17.02	7.86	1.120	8.400	4.723	0.294	6.231	12.200	0.083	0.280	0.008	0.009	8580	740	17.8	8
I	^{13}C labelled *E.coli added*	12/03/2012	46	15.96	7.98	0.592	9.017	4.703	0.297	6.322	17.800	0.110	0.280	0.060	0.040	7120	720	8.8	7
I		19/03/2012	53	15.48	8	1.300	9.960	4.687	0.294	6.266	5.577	0.083	0.253	0.034	0.015	8220	1120	2.6	8
I		26/03/2012	60	15.8	7.8	0.518	10.300	4.680	0.300	6.410	10.733	0.147	0.090	0.009	0.051	16140	3200	2	6
I		03/04/2012	68	18	7.8	0.500	9.053	4.607	0.299	6.498	15.067	0.280	0.327	0.007	0.006	5320	860	1.6	7
J		30/01/2012	4	15.26	7.8	3.514	11.200	8.867	0.317	3.579	30.667	0.087	0.287	0.041	0.411	6708	5612	1928	6
J		06/02/2012	11	14.96	7.84	1.320	11.200	4.400	0.290	6.598	28.500	0.137	1.007	0.521	0.284	35980	31940	960	5
J		13/02/2012	18	15.32	7.88	0.790	10.020	3.793	0.247	6.503	9.913	0.170	0.543	0.063	0.021	13740	6120	298	7
J		20/02/2012	25	15.4	7.6	0.484	8.567	4.600	0.308	6.696	12.100	0.083	0.248	0.017	0.023	10480	9680	34	8
J		27/02/2012	32	14.82	7.8	1.370	9.230	4.773	0.298	6.236	14.867	0.110	0.167	0.022	0.010	8880	4820	26	7
J		05/03/2012	39	17	7.94	1.444	8.394	3.967	0.296	7.454	12.267	0.087	0.277	0.016	0.002	10760	3920	64	8
J	^{13}C labelled *E.coli added*	12/03/2012	46	15.96	7.9	0.501	8.973	4.010	0.297	7.398	17.867	0.103	0.280	0.067	0.040	19580	16520	200	7
J		19/03/2012	53	15.48	7.9	0.363	9.943	4.683	0.292	6.242	5.540	0.073	0.267	0.041	0.006	6820	1040	2.4	8
J		26/03/2012	60	15.7	7.8	0.516	10.367	4.690	0.300	6.397	10.533	0.147	0.090	0.000	0.054	4320	2900	8	6
J		03/04/2012	68	18.08	7.7	0.801	9.020	4.587	0.299	6.526	15.033	0.167	0.327	0.030	0.008	5620	1020	3.6	7

is noteworthy as in comparison to the previous studies the water feeding these filters came from river water that is much higher in organic matter. Therefore the fact that these filters produced water of the same quality as filters purifying reservoir water shows the versatility of SSFs to treat a range of water sources.

6.3.1.1 Coliform Removal in Covered and Non-covered SSFs

Similar to the filters in chapters 4 and 5, these filters fulfil all of the water quality parameters except coliform removal targets set out in the Water (Scotland) Act 1980. However, unlike the SSFs in previous chapters the filters within this chapter remove significantly fewer coliforms (approximately 24 times less) at the early age bin (p-value: 6.839×10^{-6}). Further, significant negative correlations were found between coliform retention and age of the filters, with both covered and non-covered LSSFs displaying the same trend of improved removal with age (Figure 6.3). Additionally, Wilcoxon tests revealed that there are significant differences in coliform retention / removal between covered and non-covered filters at the early age bin (p-value: 0.0085); with non-covered filters removing $>$ 2.5 times more coliforms than covered LSSFs. However, by the mid and late age bins ($>$ 42 days old) there was no difference is coliform removal ability between the filter types (p-values: 0.4245 and 0.3126 respectively) even during the spiking event.

6.3.2 Impact of Light on the SSF Microbial Community

Initial exploratory NMDS and MANOVA analysis (Figure 6.4) revealed that the microbial communities present in covered and non-covered LSSFs were significantly different (p-value: $>$0.001), with *Gammaproteobateria*, *Acidobacteria* and unclassified bacteria dominating non-covered LSSFs (12h light / 12h darkness), and *Deltaproteobacteria* and *Planctomycetes* dominating covered LSSFs (Figure 6.5). Additionally, significant differences in the abundance of 16S rRNA genes were found between covered and non-covered filters (p-value: 0.0014), with non-covered filters possessing on average four times more copies than covered LSSFs. Further, in accordance with the filters in Chapter 4 and 5 a positive correlation was found between between age and 16S rRNA copy number (0.15, p-value: 0.007) and a negative correlation was found between depth and 16S rRNA copy number (0.19, p-value: 0.0005). Overall, mean 16S rRNA copy numbers were consistent with the filters of previous chapters (covered filters: 5.48×10^7 and non-covered filters: 2.02×10^8 16S rRNA copies

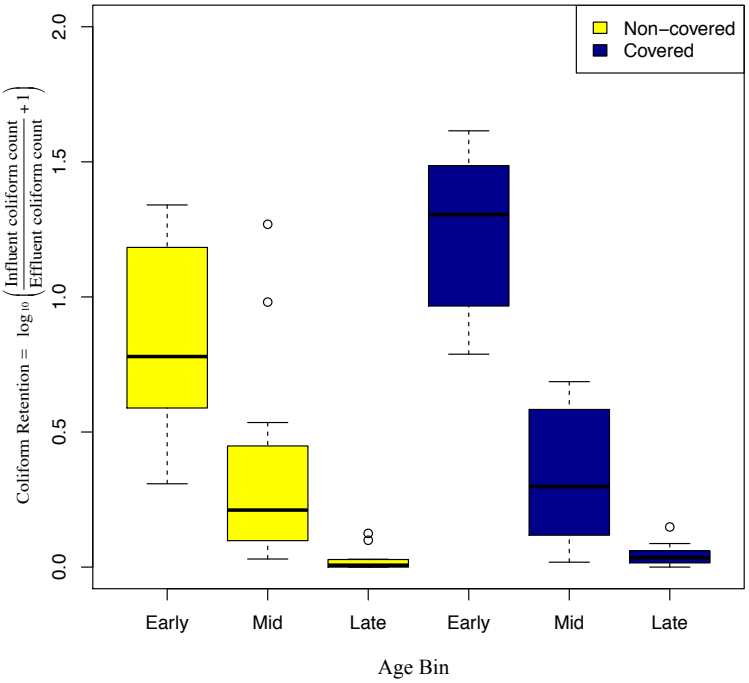


Figure 6.3: Box-and-whisker plot comparing the coliform retention in covered and non-covered LSSFs at three age bins. Early age bin: 0-3 weeks, Mid: 4-6weeks and Late: 7-10 weeks. The top and bottom boundaries of the boxes show the 75th and 25th percentile and the ends of the whiskers show the maximum and minimum values within 1.5 of the interquartile range of the upper and lower quartiles (respectively). Bold lines within the boxes represent median values (50th percentile). Note: values below 1 correspond to coliform removal and values above 1 signify retention / no removal.

per gram of extracted sand).

Overall, clear changes can be seen in the absolute phyla abundances over time and depth (Figure 6.5) with the *Proteobacteria* phylum dominating both covered and non-covered LSSFs as seen previously in labscale and industrial SSFs. However, interestingly the dominance of the various classes of *Proteobacteria* are extremely different both between covered and non-covered LSSFs and between previous labscale and industrial SSFs. Previously, *Alphaproteobacteria* and *Betaproteobacteria* accounted for the biggest proportion of the *Proteobacteria* phyla in both labscale and industrial SSFs (Figure 5.9), however in the labscale filters in this chapter *Gammaproteobacteria* and *Deltaprotoebacteria* are the dominant classes. Reasons for such differences are likely due to the different water sources (previously reservoir water and now river water) which have already been shown to strongly influence the SSF sand microbial community and can be seen in Figure 6.6 to do the same for the filters here.

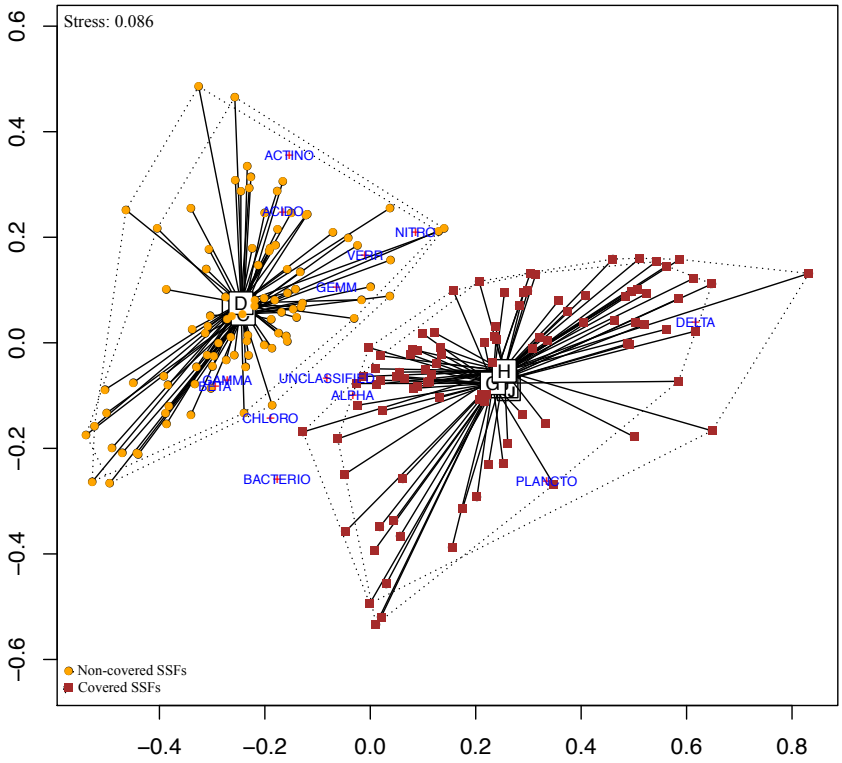


Figure 6.4: NMDS analysis of phlya abundance in covered and non-covered SSFs. Capital letters denote filter names, non-covered: C, D, E, and F and covered: G, H, I, and J.

In order to determine which factors explain the differences between covered and non-covered LSSFs microbial communities, MANOVA and canonical correspondence analysis was performed. This analysis revealed (Table 6.6) that the condition of the filter (covered or non-covered) explained the biggest proportion (39.5%) of the difference in bacterial community composition, with age, depth and the individual filter identity (i.e., C, D, E), also being highly significant. Interestingly, individual filter identity was not a significant variable in explaining the differences between bacterial community composition in previous chapters; a potential reason for its importance now could be due to the different influent water. Within

Table 6.6: CCA analysis of bacterial phyla and class abundances against various filter parameters. ∗ denotes significant parameters.

Parameter	Degrees of Freedom	χ^2	F value	Number of permutations	Pr(>F)
Condition (Covered / non-covered)	1	0.3166	137.8915	99	0.01*
Age	1	0.0679	29.5647	99	0.01*
Depth	1	0.0181	7.9029	99	0.01*
Filter Identity	6	0.0444	3.2230	99	0.01*
Residual	342	0.7853			

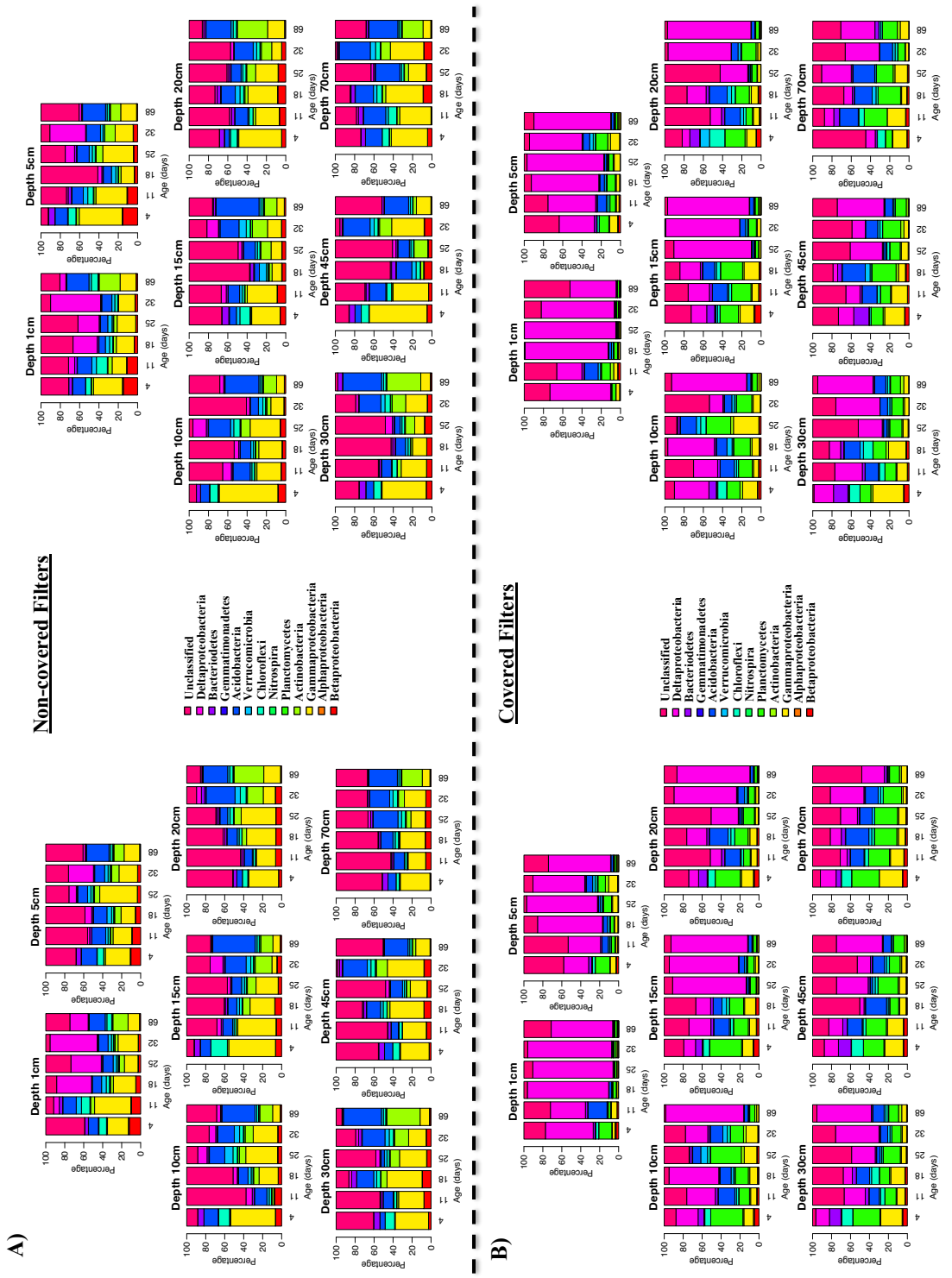


Figure 6.5: Stacked barplots of the percentage abundance of each phyla at different depths in: A) non-covered (filters subjected to sunlight) and B) covered (filters in complete darkness) SSFs

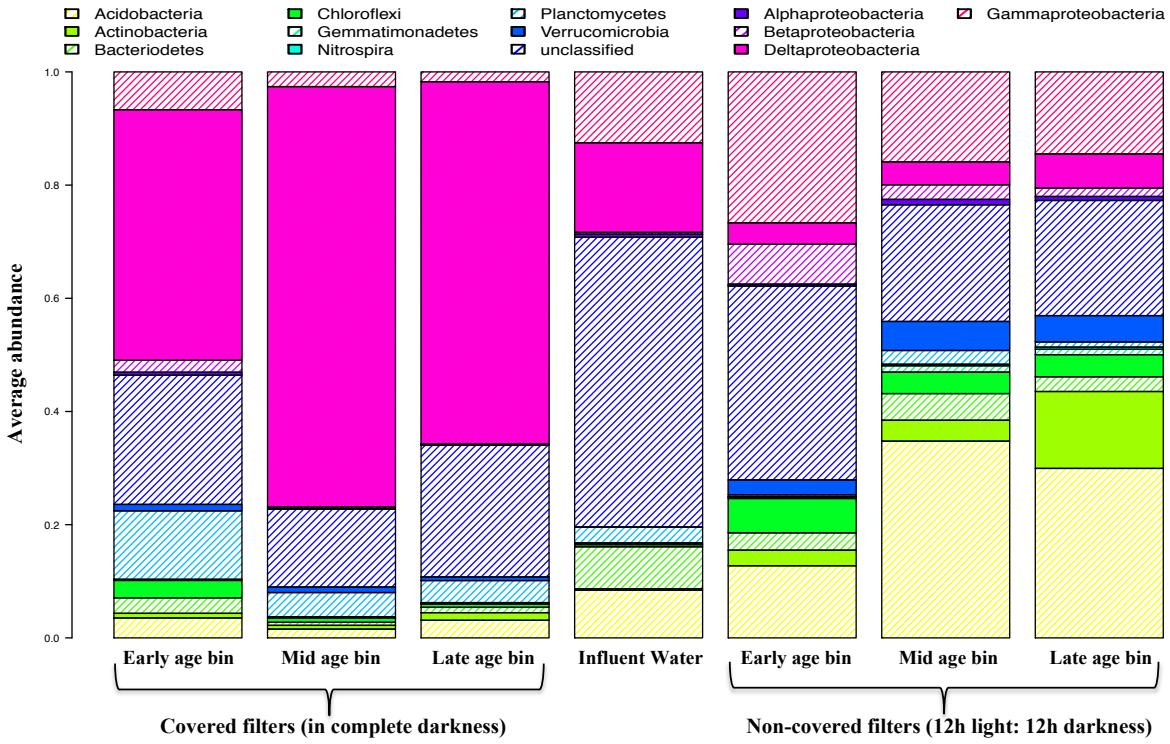


Figure 6.6: Stacked barplots depicting the average fractional abundance of all phyla and classes assessed by qPCR in influent water, covered and non-covered LSSFs at different age bins. Early age bin: 0-3 weeks, mid: 4-6 weeks and late: 7-10weeks.

this chapter, water from the River Kelvin was used which can be seen (Figure 6.7) to have much higher 16S rRNA copy numbers than the water feeding the industrial SSFs and lab-scale SSFs of chapter 5. Such exposure to increased diversity alongside the knowledge of neutral theory (multiple organisms can perform the same function) may explain individual filter variability both within and between covered and non-covered filters.

Determining the Mechanisms of *E.coli* Removal

To determine the magnitude with which protozoa and other eukaryotes affect *E.coli* removal; direct recombinant *E.coli* and protozoa counts, total and *E.coli* specific 16S rRNA, and 18S rRNA, qPCR assays were performed on samples taken from SSFs challenged with isotopically labelled (^{13}C) *E.coli*. To resolve which organisms (bacteria, eukaryotes and viruses) were responsible for *E.coli* removal, the different carbon densities (^{12}C and ^{13}C) in the samples were separated (DNA-SIP) and used for metagenomic analysis. Increased abundance of any organism in ^{13}C labelled samples is indicative of potential involvement in *E.coli* removal.

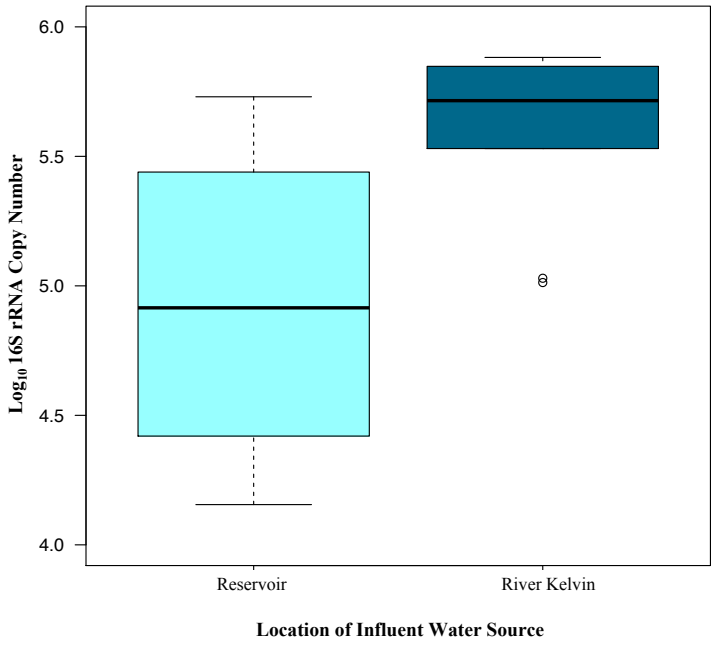


Figure 6.7: Box-and-whisker plot comparing the number of 16S rRNA copies (determined by qPCR) in the water supplies used in Chapter 4 and 5 (Reservoir) and this chapter (River Kelvin).

6.3.3 Protozoan predator-prey response - direct counts and qPCR

Direct counts of *E.coli* and total protozoa revealed a clear predator-prey relationship (Figure 6.9), with most removal occurring at the top of the filters, 2-3h after spiking with ^{13}C labelled *E.coli* (Figure 6.8). Both filter types achieved 100% *E.coli* removal after 24h, however at different sand depths coliform removal occurs more quickly in uncovered filters in comparison to covered SSFs, with *E.coli* still appearing after 96h in covered SSFs at lower depths (Figure 6.8). This implies that an environment with light exposure is more hostile in terms of *E.coli* survival, yet has no affect on water quality production. However, it should be noted that increased retention of viable *E.coli* in covered filters at lower depths could result in the release of pathogens at a later point into the effluent of the covered SSFs. Therefore in a real world situation both types of SSFs deal with heavy pollution and run-off events extremely well resulting in reduced pathogen numbers. However, referring back to Tables 6.4 and 6.5 it can be seen that coliform removal in the three weeks after the spiking event is significantly lower in covered SSFs compared to non-covered, insinuating that covered filters incur longterm effects to increased *E.coli* exposure which have a knock-on effect to coliform removal.

The importance of protozoan grazing was further emphasised by the gradual decrease over

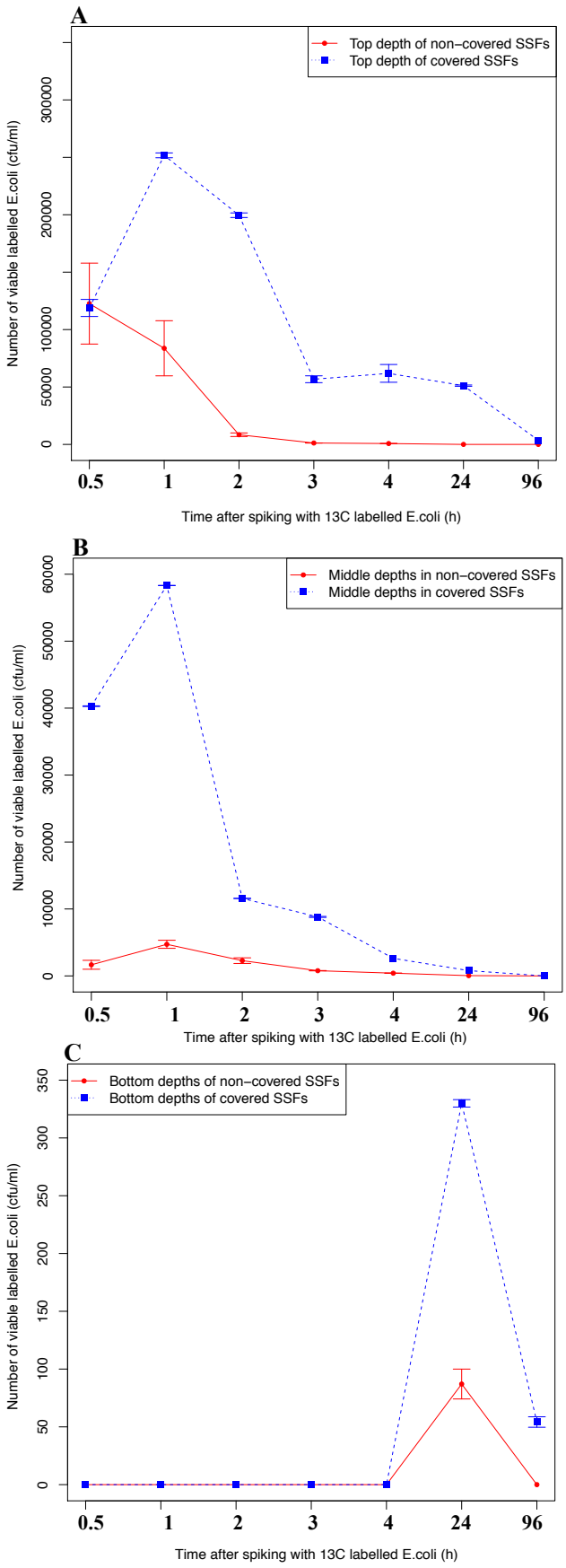


Figure 6.8: Direct counts of viable ^{13}C labelled E.coli at various timepoints and depths within SSFs. ***A:*** *Top depths (0-5cm),* ***B:*** *Middle depths (15-30cm) and* ***C:*** *Bottom depths (40-70cm). Note the different y-axes.*

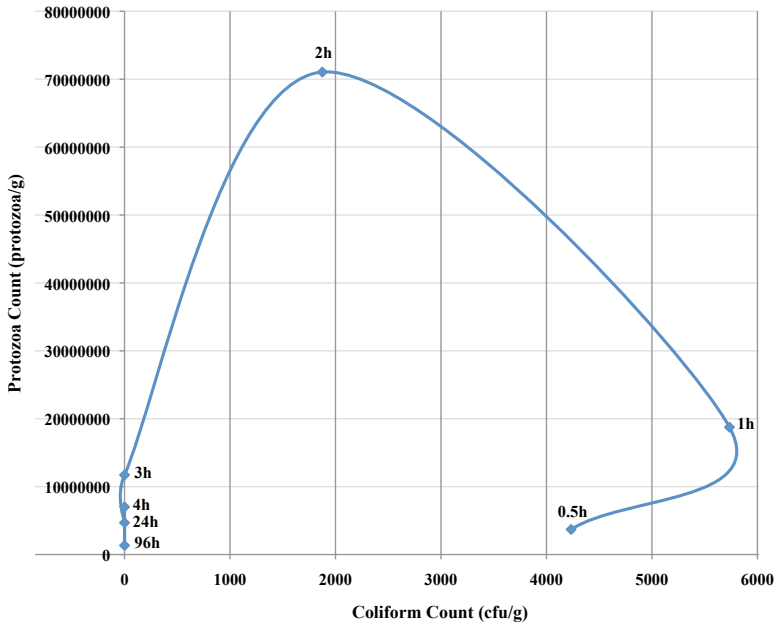


Figure 6.9: Scatter plot showing the predator-prey response of protozoa on coliforms at the top depth in SSFs

time in the abundance of ^{13}C labelled *E.coli*, as well as the peak in the number of labelled (^{13}C) 18S rRNA copies in qPCR assays on samples 3 and 24 hours after spiking (Figure 6.10). However, as the ^{12}C 18S rRNA results showed a similar trend compared to labelled 18S rRNA, it can be assumed that incomplete ^{13}C incorporation has occurred i.e., samples with a density resembling normal carbon (^{12}C) may have started to incorporate labelled ^{13}C but have not incorporated enough into their biomass to cause a density change. Additionally, the increased occurrence of 18S rRNA copies at early time-points (2 and 3h after spiking) is surprising due to the on average longer replication rate of eukaryotes, however some protists and various fungi are known to incorporate carbon into their genomes and replicate within two hours after exposure to a food source [Bååth, 2001].

6.3.4 All Domains of Life are Important for *E.coli* Removal

To resolve which organisms (bacteria, eukaryotes and viruses) were responsible for *E.coli* removal, DNA-SIP in conjunction with metagenomic analysis was employed. Metagenomics unlike qPCR and conventional sequencing approaches, does not rely on prior knowledge of the organisms of interest and, thus organism-specific primers are not required [Wooley et al., 2010]. However, metagenomics does still suffer from biases introduced during DNA extraction, enzymatic cutting during library preparation and PCR amplification. MANOVA

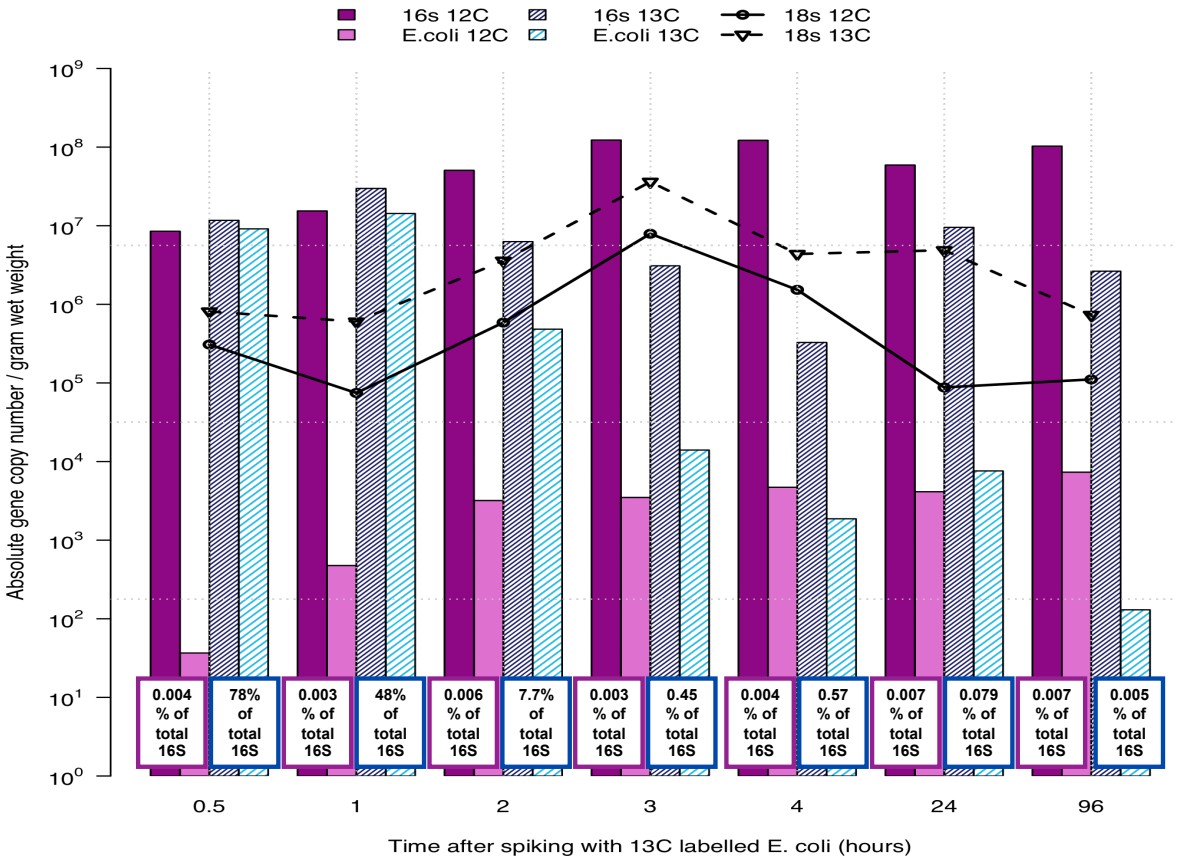


Figure 6.10: Absolute numbers of 16S, 18S and E.coli specific 16S rRNA in ^{12}C and ^{13}C fractions, determined by qPCR assays on the top depth of sand.

revealed significant differences in the metagenomic communities at all levels of taxonomic classification between: carbon types (^{13}C or ^{12}C), and times after spiking, of the 36 samples taken from four SSFs after adding the ^{13}C labeled *E.coli*. Time was the most significant variable explaining 15.4% (p-value: 0.01) of the variance followed by type (11.3% p-value: 0.001) and condition (5.1% p-value: 0.02). Canonical correspondence analysis (Figure 6.11) alongside SIMPER analysis (Table 6.7) on individual taxa proportions adjusted for multiple comparisons identified ten orders (two bacterial, six eukaryotic and two viral) as being statistically significant in explaining differences between ^{12}C and ^{13}C metagenomic communities.

Overall, SIMPER analysis revealed that the viral order *Caudovirales* accounted for the biggest difference between ^{13}C and ^{12}C communities, insinuating that viral lysis may be a significant mode for *E.coli* removal. Further, SIMPER analysis revealed that the *Enterobacteriales* order accounted for a large difference between labelled and non-labelled samples, this is not surprising as *E.coli* is a member of this order and differences here are due to the excessive quantify of labelled *E.coli* added during the experiment. Therefore as all three

Table 6.7: Relative abundances of statistically significant Orders of organism which explain differences between ^{13}C and ^{12}C samples. Adjusted p-values calculated using the Benjamini-Hochberg method, average percentages are relative to the kingdom the order belongs to.

Order	Kingdom	t	p-value	Adjusted p-value	SIMPER Calculated Contribution to Difference (%)	Average Percentage Abundance 12C	13C
Peniculida	Eukaryota	2.68182	0.01173	0.015076	0.0700	0.3295	0.2305
Herpesvirales	Viral	2.51466	0.01656	0.022064	9.6000	29.9375	14.0812
Caudovirales	Viral	-2.5104	0.01671	0.022266	9.9500	49.9682	85.8644
Eurotiales	Fungi	-2.4689	0.02777	0.032573	1.5201	3.7049	6.2315
Enterobacteriales	Bacterial	-2.2899	0.04082	0.04742	4.1696	3.5448	16.8301
Vibrionales	Bacterial	-2.2864	0.04108	0.047727	0.6103	0.4350	2.3151
Kinetoplastida	Eukaryota	-1.5084	0.04205	0.050322	1.3890	4.9205	6.5473
Perkinsida	Eukaryota	-1.437	0.04775	0.049914	1.2152	1.0862	2.9243
Sordariales	Fungi	-0.4374	0.05068	0.05858	2.3166	4.7333	10.5973
Saccharomycetales	Fungi	-0.5542	0.05074	0.060559	2.3438	9.1490	10.2895

domains of life (Viruses, Eukaryotes and Bacteria) have been identified as accounting for significant differences between ^{13}C and ^{12}C communities and hence *E.coli* removal, each kingdom was analysed independently at higher levels of taxonomic resolution. It should be noted that the number of reads associated with each time point (Figure 6.12) within each group (^{12}C and ^{13}C) was not significantly different (p-value: 0.9356), and both fractions had similar overall community structures (^{12}C: 0.09% Viruses, 9.18% Eukaryotes and 90.73% Prokaryotes; ^{13}C: 0.08% Viruses, 8.50% Eukaryotes and 91.42% Prokaryotes).

6.3.5 The Importance of Viral Lysis for *E.coli* Removal

Twenty-two viral species (Table 6.8) were identified by pairwise t-tests with adjusted p-values based on Benjamini-Hochberg false discovery tests as being significantly different between ^{12}C and ^{13}C metagenomic communities. None were prophages found within the *E.coli K12* genome. Overall, fifteen of these viruses were present at higher abundances in ^{13}C communities compared to ^{12}C samples and hence are involved directly or indirectly in *E.coli* removal. Collectively these accounted for 22% of the dissimilarity between ^{13}C and ^{12}C samples. Visually referring to the 22 significant viral species (Figure 6.13) the importance of *Enterobacteria Phages* is apparent, and in particular *Enterobacteria Phage Lambda, Enterobacteria Phage cdtl* and *Enterobacteria Phage N15*, which account for over 14% of the difference between labelled and non-labelled samples (Table 6.8). In particular, the abundance of *Enterobacteria Phage Lambda* was over 117 times more abundant in ^{13}C samples than ^{12}C communities, implying its importance in *E.coli* removal. Further their increased

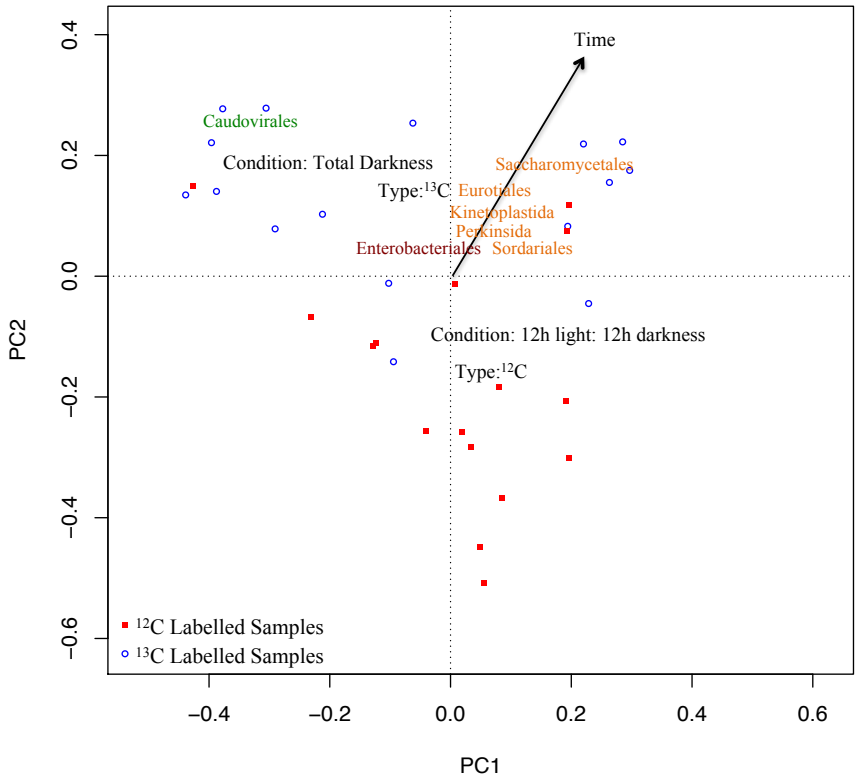


Figure 6.11: Principle component analysis of all orders of organism. Green text represents viral orders, brown text represents bacterial orders and orange text represents eukaryotic orders.

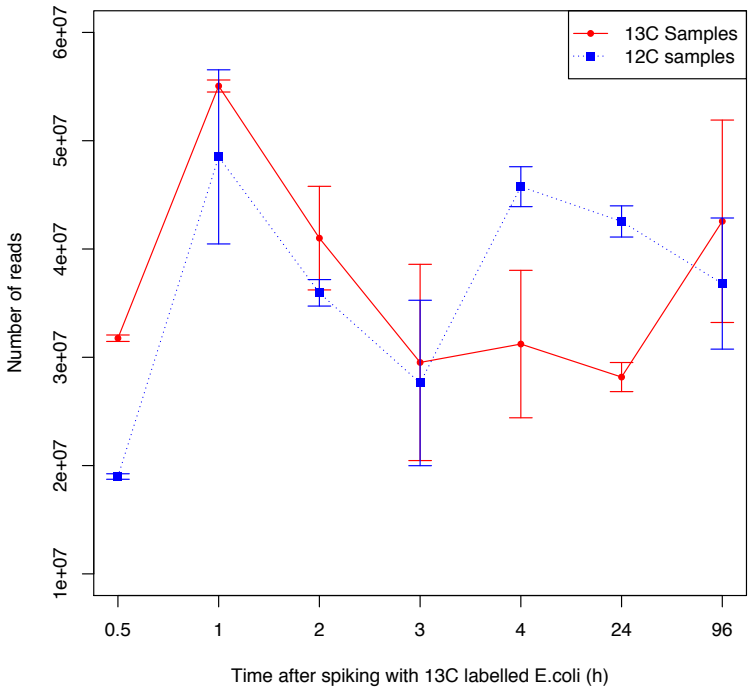


Figure 6.12: Line plot displaying the total number of reads classified to species level at each time-point within ^{12}C and ^{13}C metagenomic samples. Data points represent the mean and bars represent the standard error.

Table 6.8: Relative abundances of 22 viral species which are statistically significant between ^{13}C and ^{12}C samples. Adjusted p-values calculated using the Benjamini-Hochberg method.

			Adjusted	SIMPER Calculated	Average Percentage Abundance	
Viral Species	**t**	**p-value**	**p-value**	**Contribution to Difference (%)**	**12C**	**13C**
Ictalurid.herpesvirus.1	2.6716	0.0094	0.0195	2.5004	4.2004	1.4493
Enterobacteria.phage.lambda	-2.9675	0.0105	0.0123	10.3949	0.1162	13.6214
Glypta.fumiferanae.ichnovirus	2.6974	0.0106	0.0153	4.3017	6.7856	3.1190
Microbacterium.phage.Min1	-2.9207	0.0107	0.0125	1.1951	0.5574	1.8756
Human.herpesvirus.6	2.9802	0.0107	0.0163	1.3084	2.0376	0.6672
Stenotrophomonas.phage.S1	-2.7294	0.0172	0.0196	0.0600	0.0112	0.0854
Enterobacteria.phage.YYZ.2008	-2.6555	0.0210	0.0235	0.4300	0.0055	0.5960
Rhodococcus.phage.REQ3	-2.5802	0.0217	0.0256	0.3785	0.0604	0.2577
Stx2.converting.phage.1717	-2.6298	0.0220	0.0247	0.8700	0.0192	1.1474
Phage.Gifsy.2	-2.5886	0.0237	0.0268	1.0210	0.0100	1.0228
Rhodococcus.phage.REQ1	-2.4596	0.0257	0.0309	1.0004	0.2396	0.5613
Enterobacteria.phage.N15	-2.5010	0.0278	0.0318	0.5100	0.0189	0.6695
Pseudomonas.phage.phi297	-2.4469	0.0286	0.0337	0.5902	0.3174	0.9253
Mycobacterium.phage.L5	-2.4226	0.0288	0.0344	0.2400	0.1299	0.3619
Mycobacterium.phage.SWU1	-2.3819	0.0308	0.0369	0.2300	0.1322	0.3619
Enterobacteria.phage.cdtI	-2.4057	0.0332	0.0382	3.6571	0.0450	4.7744
Enterobacteria.phage.SfV	-2.2911	0.0341	0.0407	0.6100	0.0033	0.0141
Salmonella.phage.Fels.1	-2.3826	0.0346	0.0399	0.6996	0.0091	0.9147
Enterobacteria.phage.P2	2.0842	0.0369	0.0580	0.0300	0.0441	0.0142
Phage.Gifsy.1	-2.3279	0.0379	0.0440	0.7810	0.0962	2.4099
Mycobacterium.phage.Cjw1	-2.2296	0.0436	0.0515	0.5700	0.0934	0.3841
Propionibacterium.phage.PAD20	1.5668	0.0438	0.0509	5.4700	7.1897	2.3280

abundance alongside *Phage gifsy 1* and 2 and *Stx2 Converting Phage 1717* are not surprising as they are known phages of *E.coli*. However their over expression at 1, 4 and 96h and lower abundance at the remaining time points suggests that these viruses are fluctuating between states of pseudolysogeny and lysogeny to lytic pathways (Figure 6.15). Such behaviour is widely documented for environmental phages [Abedon, 2008], with different pathways being chosen based on the "health" of the host bacteria (i.e., availability of nutrients and exposure to environmental stress such as UV).

Referring more closely to the significant viral species (Table 6.8) it was observed that $>15\%$ of the dissimilarity between ^{13}C and ^{12}C is due to members of the *Lambda-like-viruses* (Figure 6.14). Furthermore, it can be seen that the abundance of the *Lambda-like-viruses* dramatically increases at 1h, 4h and 96h after spiking with labelled *E.coli*. Additionally it can be seen that two members of the *Myoviridae* family; *T4-like-viruses* and *L3-like-viruses* increase in abundance at the time-points where *Lambda-like-viruses* have reduced abundance (2h, 3h, and 24h). These differences in labelled viral genera at different time-points imply that various viruses are important for *E.coli* removal and important in downstream labelled carbon metabolism. Additionally such differences in the abundance of different genera may

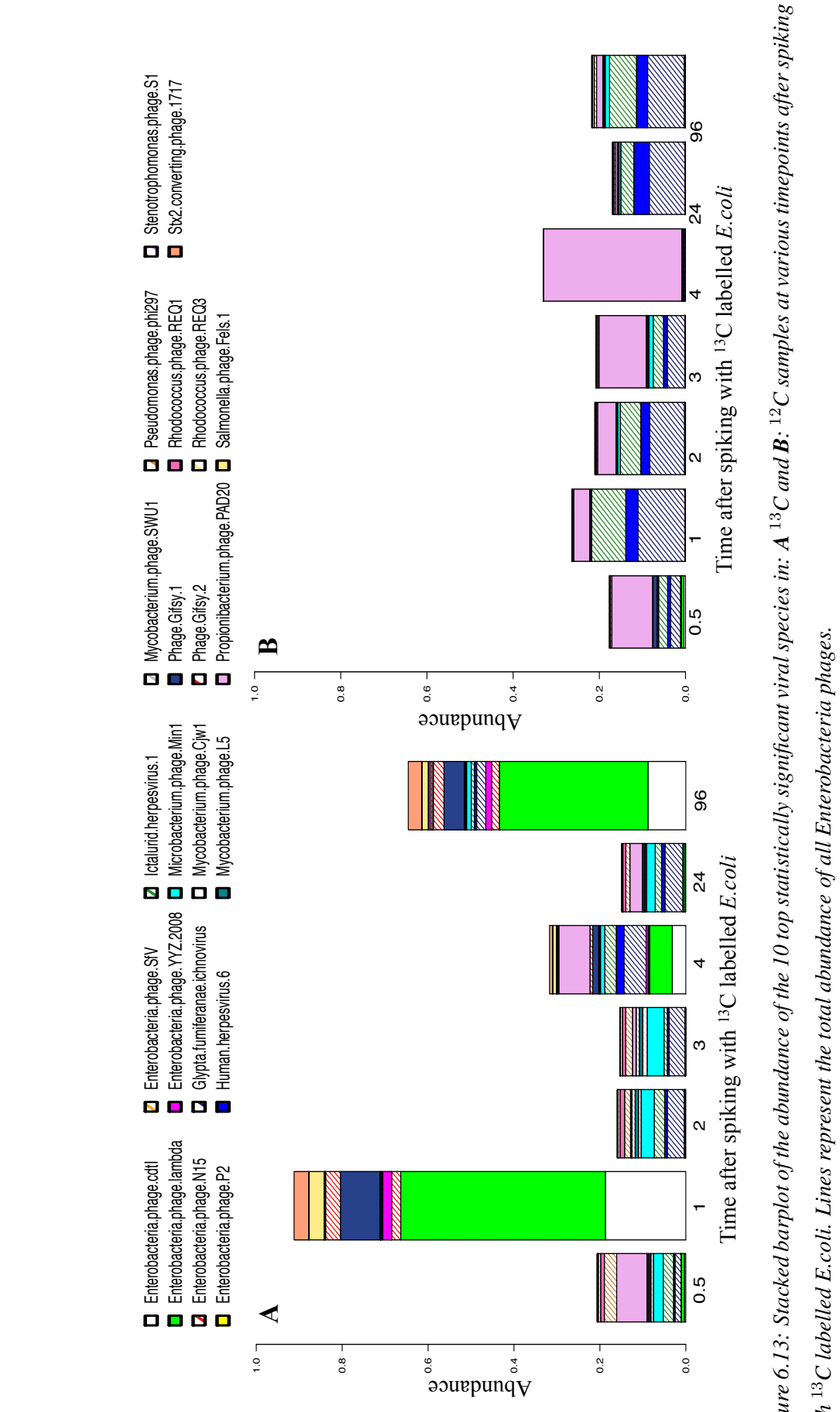


*Figure 6.13: Stacked barplot of the abundance of the 10 top statistically significant viral species in: **A** ^{13}C and **B**: ^{12}C samples at various timepoints after spiking with ^{13}C labelled E.coli. Lines represent the total abundance of all Enterobacteria phages.*

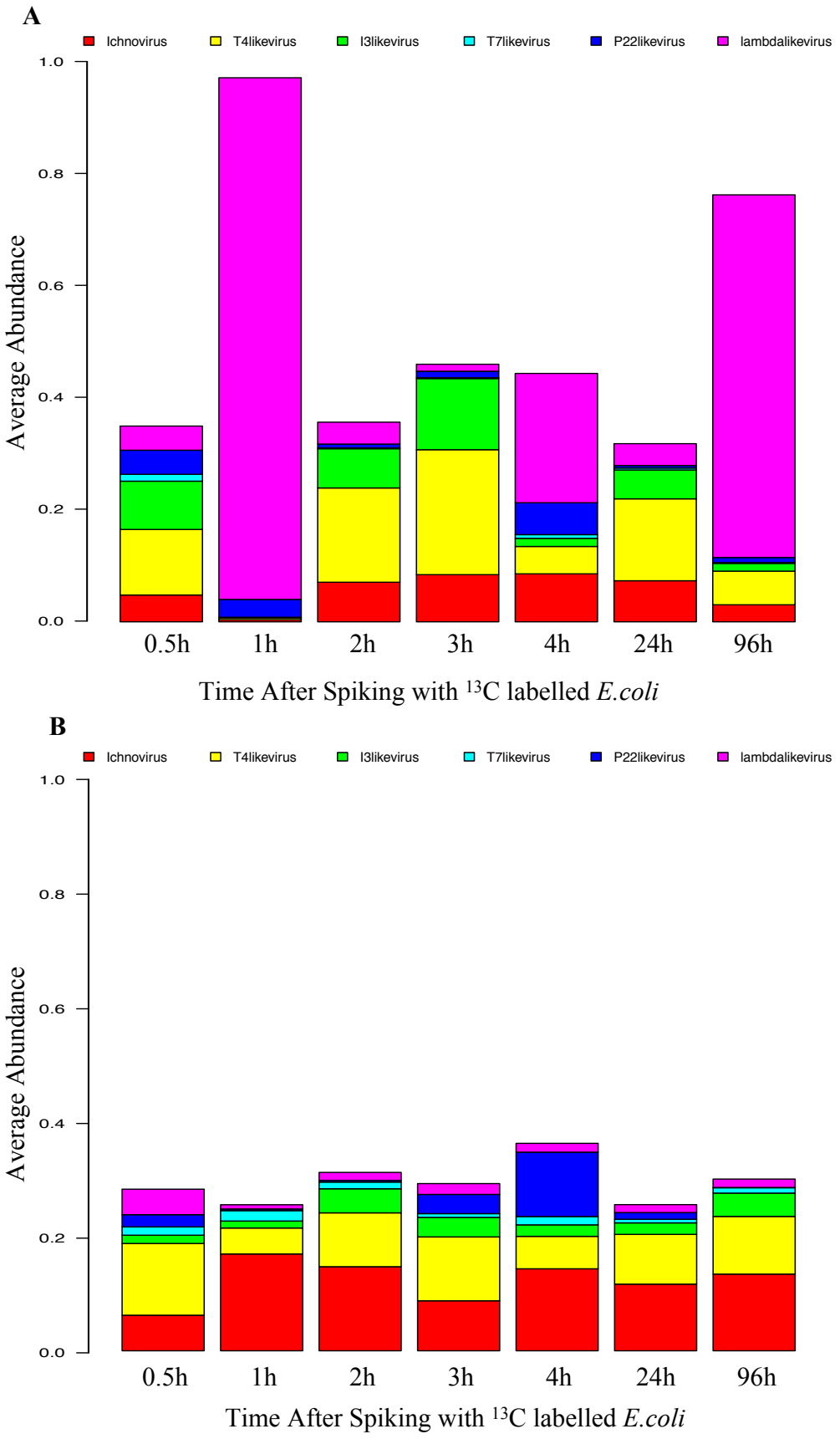


Figure 6.14: Stacked barplot showing the abundance of the top 6 viral genera in: A: ^{13}C (labelled) samples and B: 12C (non-labelled) samples at various timepoints after spiking with ^{13}C labelled E.coli

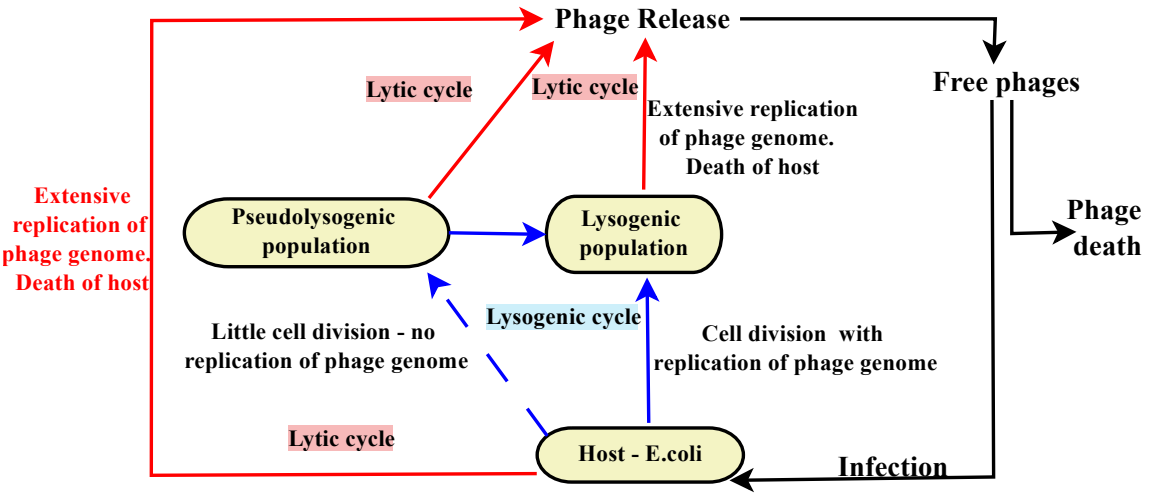


Figure 6.15: The progressive relationship between host bacterial cells and different bacteriophage life cycles, adapted from Abedon [2008]. Red lines indicate lytic pathways and blue lines indicate lysogenic pathways.

be due to differences in genome size and hence replication rate. For example *Lambda-like-viruses* have a smaller genome (48500 nucleotides on average) compared to an average of 101800 nucleotides for viruses in the *Myoviridae* family (*T4-like-viruses* and *L3-like-viruses*). Additionally, the increased abundance of *Lambda-like-viruses* at 96 hours may be due to the fact that unlike members of the *Myoviridae* family *Lambda-like-viruses* have the potential to be lysogenic (integrate into the genome of their host and remain dormant until they desire to leave the host (Figure 6.15)).

6.3.6 The Importance of Eukaryotes for *E.coli* Removal

Following the same approach taken to identify significant viral species involved with *E.coli* removal and ^{13}C metabolism, 52 eukaryotic species were identified (Table 6.9), of which twenty were protozoa (Figure 6.16), fifteen algae (Figure 6.17) and seventeen fungi (Figure 6.18). The presence of significant species from all eukaryotic kingdoms further underscores the complex mechanisms involved with *E.coli* removal.

6.3.6.1 The Importance of Protozoan Grazing for *E.coli* Removal

The twenty significant protozoan species represented fifteen different genera and members from the flagellate, ciliate and amoeboid groups, all of which are known predators of *E.coli* [Weekers et al., 1993, Fleck et al., 2000, Fey et al., 2007, Cassidy-Hanley, 2012, Yue et al.,

Table 6.9: Eukaryotic species relative abundances which are statistically significant between ^{13}C and ^{12}C samples. Samples are from the top depth of the SSF bed and represent averages based on all time points (0.5-96h post spiking).

Eukaryotic Species	t	p-value	Adjusted p-value	SIMPER Calculated Contribution to Difference (%)	Average Abundance (%) 12C	13C
Paramecium.tetraurelia	3.57762	0.0017	0.00174	0.0896	0.3115	0.1591
Tetrahymena.thermophila	2.75045	0.0103	0.01268	0.1990	0.5367	0.3526
Postia.placenta	-2.73381	0.0153	0.01829	0.0808	0.1343	0.2669
Tetrahymena.malaccensis	-2.54231	0.0216	0.02607	0.0130	0.0173	0.0358
Desmarestia.viridis	-2.55943	0.0219	0.02619	0.0126	0.0168	0.0352
Aspergillus.terreus	-2.53551	0.0244	0.02859	0.1161	0.1236	0.3122
Neosartorya.fischeri	-2.53955	0.0245	0.02855	0.2753	0.2594	0.7184
Reclinomonas.americana	-2.46792	0.0266	0.03166	0.0519	0.0661	0.1428
Malawimonas.jakobiformis	-2.43917	0.0277	0.03310	0.0790	0.1040	0.2189
Pylaiella.littoralis	-2.41217	0.0285	0.03429	0.0105	0.0137	0.0287
Blastocystis.sp..subtype.3	-2.38201	0.0297	0.03577	0.0104	0.0150	0.0292
Saccharina.coriacea	-2.38071	0.0302	0.03629	0.0121	0.0173	0.0340
Saccharina.diabolica	-2.37830	0.0304	0.03655	0.0123	0.0175	0.0346
Saccharina.ochotensis	-2.37830	0.0304	0.03655	0.0124	0.0175	0.0346
Saccharina.longipedalis	-2.37773	0.0304	0.03659	0.0123	0.0175	0.0346
Saccharina.religiosa	-2.35847	0.0314	0.03781	0.0194	0.0172	0.0336
Rhizopus.oryzae	-2.36037	0.0319	0.03828	0.0305	0.0415	0.0853
Saccharina.angustata	-2.33985	0.0325	0.03909	0.0116	0.0169	0.0327
Saccharina.japonica.x.latissima	-2.33277	0.0335	0.04016	0.0116	0.0164	0.0324
Eremothecium.gossypii	-2.33713	0.0342	0.04076	0.0721	0.1106	0.2116
Paulinella.chromatophora	-2.34440	0.0343	0.04074	0.0824	0.1044	0.2257
Glaucocystis.nostochinearum	-2.30653	0.0343	0.04123	0.0199	0.0304	0.0572
Chaetomium.globosum	-2.33314	0.0354	0.04193	0.1467	0.2187	0.4357
Schizophyllum.commune	-2.33412	0.0355	0.04199	0.1342	0.1655	0.3741
Polysphondylium.pallidum	-2.30359	0.0361	0.04313	0.0154	0.0226	0.0438
Cafeteria.roenbergensis	-2.28885	0.0361	0.04333	0.0199	0.0293	0.0561
Chattonella.marina	-2.28340	0.0368	0.04409	0.0164	0.0238	0.0458
Ochromonas.danica	-2.26917	0.0376	0.04507	0.0177	0.0272	0.0509
Tetrahymena.pigmentosa	-2.24581	0.0378	0.04517	0.0124	0.0193	0.0355
Vermamoeba.vermiformis	-2.26055	0.0379	0.04544	0.0275	0.0381	0.0714
Laminaria.digitata	-2.26860	0.0380	0.04552	0.0113	0.0165	0.0319
Sordaria.macrospora	-2.28253	0.0381	0.04542	1.6880	2.9028	5.3071
Naumovozyma.castellii	2.22123	0.0382	0.03448	0.2902	1.0603	0.6671
Gigaspora.rosea	-2.28863	0.0384	0.04551	0.0106	0.0126	0.0287
Monosiga.brevicollis	1.53508	0.0390	0.04320	1.7310	6.5174	8.2966
Gigaspora.margarita	-2.27833	0.0391	0.04638	0.0106	0.0126	0.0287
Chromerida.sp..RM11	-2.27704	0.0391	0.04645	0.0066	0.0808	0.1779
Tetrahymena.pyriformis	-2.22485	0.0397	0.04748	0.0135	0.0207	0.0382
Dictyota.dichotoma	-2.24800	0.0400	0.04773	0.0159	0.0226	0.0447
Tetrahymena.paravorax	-2.21393	0.0401	0.04780	0.0128	0.0197	0.0363
Phytophthora.sojae	-2.23975	0.0403	0.04818	0.0191	0.0290	0.0547
Euglena.gracilis	-2.24004	0.0404	0.04829	0.0539	0.0782	0.1509
Phytophthora.andina	-2.23832	0.0405	0.04837	0.0197	0.0291	0.0551
Ichthyophthirius.multifiliis	-2.21725	0.0407	0.04865	0.0135	0.0207	0.0385
Paramecium.aurelia	-2.13811	0.0408	0.05814	0.0100	0.0135	0.0275
Phytophthora.phaseoli	-2.23284	0.0409	0.04885	0.0189	0.0287	0.0542
Phytophthora.mirabilis	-2.23246	0.0409	0.04889	0.0194	0.0293	0.0555
Acanthamoeba.castellanii	-2.12744	0.0411	0.05064	0.0101	0.0141	0.0288
Cyanophora.paradoxa	-2.23258	0.0413	0.04932	0.0923	0.1337	0.2592
Phytophthora.ipomoeae	-2.22309	0.0416	0.04970	0.0193	0.0293	0.0552
Pythium.ultimum	-2.21143	0.0421	0.05032	0.0455	0.0711	0.1311
Dictyostelium.fasciculatum	-2.09591	0.0433	0.05033	0.0108	0.0177	0.0318

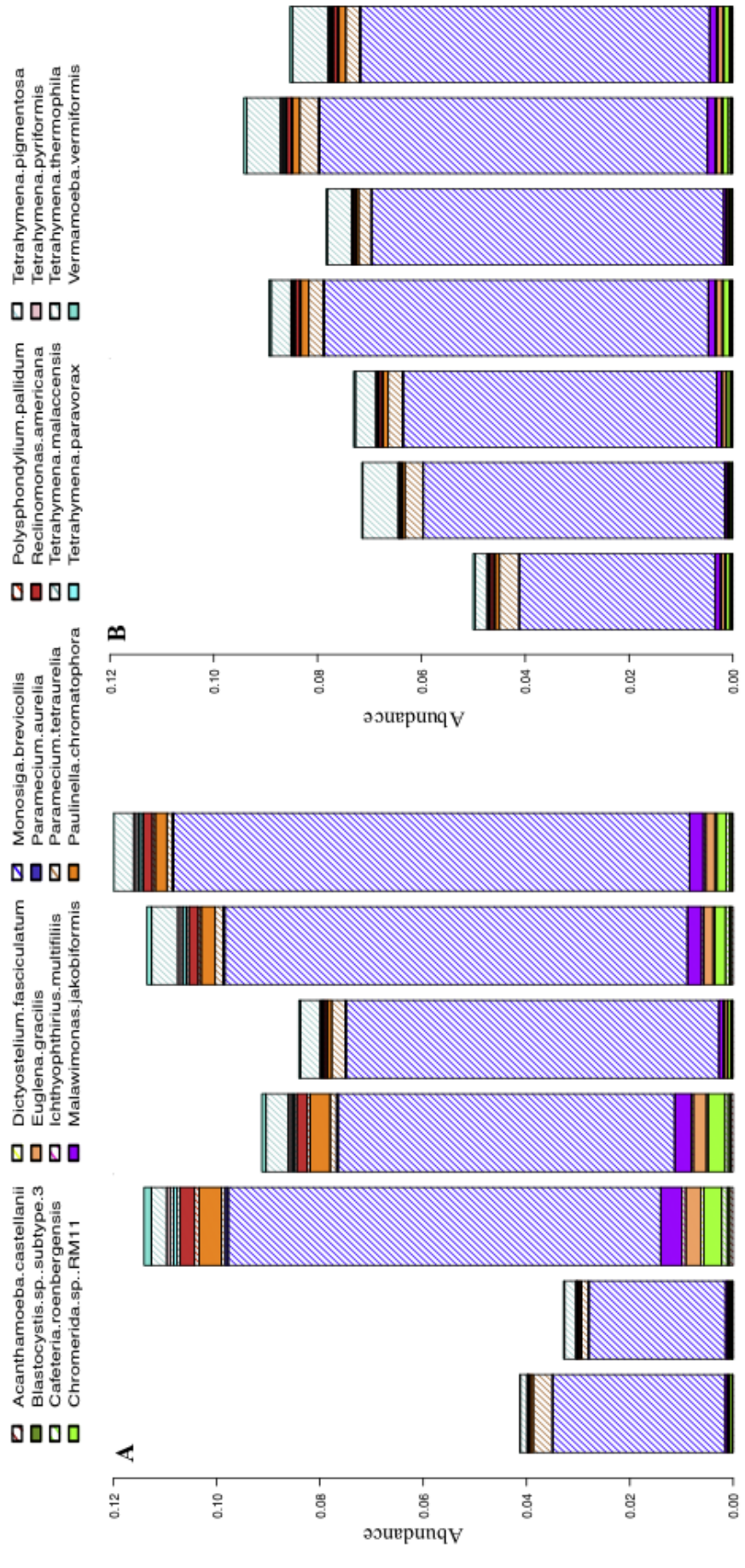


Figure 6.16: Stacked barplots showing the abundance of the statistically significant protozoan species: **A:** *^{13}C and* **B:** *^{12}C samples at various time points after spiking with ^{13}C labelled E.coli.*

2013]. Referring to Figure 6.16A it can be seen that the proportion of significant protozoan species increases over time, compared with the relatively stable proportion of 7% in non-labelled (^{12}C) samples (Figure 6.16B). Further, the biggest difference between ^{13}C and ^{12}C communities are due to fluctuations in the populations of: *Chromerida RM11, Euglena gracilis, Malawimonas jakobiformis, Monosiga brevicollis, Paulinella chromatophora, Reclinomonas americana, Tetrahymena paravorax* and *Vermamoeba vermiformis*, (Figure 6.16) all of which are highly motile and possess voracious appetites. Specifically, the importance of protozoan grazing on labelled *E.coli* and hence *E.coli* removal, is apparent two hours post-spiking, where a large increase in protozoan population is observed. In particular a large increase in the proportion of *Monosiga brevicollis* and *Tetrahymena* spp. were observed, with both genera collectively being responsible for 2% of the dissimilarity between ^{12}C and ^{13}C communities (Table 6.9).

6.3.6.2 The Mutualistic Relationship of Fungi and Algae in the Removal of *E.coli*

The abundances of seventeen fungal, and fifteen algal, species were identified as being significantly different between ^{12}C and ^{13}C samples, of which only the fungal species *Naumovozyma castelli* was present in greater proportion in non-labelled samples. Therefore, the remaining species appear to be involved in *E.coli* removal and / or ^{13}C labelled carbon metabolism. On initial analysis of the algal and fungal abundances, a staggering similarity in community dynamics can be seen, with both communities increasing and decreasing in abundance at the same time points (Figure 6.17 and Figure 6.18); Kendall correlation tests of all significant fungal and algal species confirmed this to be a significant relationship (tau 0.81952, p-value: 0.01071). This mirrored behaviour is indicative of a mutualistic relationship. Such symbioses between algae and fungi have been widely documented in various environments [Harte and Kinzig, 1993, Danger et al., 2013].

The fifteen significant algal species represented nine different genera, with *Saccharina spp.* accounting for 47% of the significant algal species. Clear shifts in the abundance of significant algae can be seen between ^{12}C and ^{13}C samples at all time points (Figure 6.17), in particular at 2, 3, 24 and 96 hours after spiking with ^{13}C labelled *E.coli*, where the average abundance tripled in ^{13}C samples compared to the non-labelled samples. Additionally, fungal species followed the same trend with *Sordaria macrospora* and *Chaetomium globosum*

showing the largest increase (Figure 6.18).

6.3.6.3 Calculating the Importance of Protozoa and Viruses

To approximate the importance of protozoa and viruses in *E.coli* removal the following assumptions were made:

1. The added ^{13}C labelled *E.coli* does not undergo replication
2. Only protozoa and viruses are responsible for *E.coli* removal
3. All protozoa and viruses are grazing / infecting *E.coli* at a constant rate
4. *Enterobacteria Phage Lambda* is used to represent all viruses (since it is the most abundant and significant virus identified in ^{13}C samples)
5. *Monosiga brevicollis* and *Tetrahymena thermophila* are used to average the abilities of protozoa (as they are the two most significant and abundant protozoa in ^{13}C samples)
6. For a protozoan to become ^{13}C labelled, 50% of its grazing consumption must be from ^{13}C labelled *E.coli*

Based on these assumptions and taking genome size and progeny production into consideration, it was concluded (Appendix D) that protozoan grazing appeared to be the major driving force behind *E.coli* removal (99.86%), with *Monosiga brevicollis* accounting for the biggest proportion of *E.coli* removal (24.83%), followed by 4.68% achieved by *Tetrahymena thermophila*. Conversely, viral associated lysis was responsible for 0.14% removal of which *Enterobacteria Phage Lambda* was responsible for 0.076%, which was 326 times smaller than the removal achieved by *Monosiga brevicollis*. However, it is important to note that these calculations do not take into consideration, the effects that viruses entering a lysogenic life cycle may have on *E.coli* removal and the ability of protozoa to be able to graze on other organisms (not *E.coli*) which may have been labelled through the consumption of released biomass.

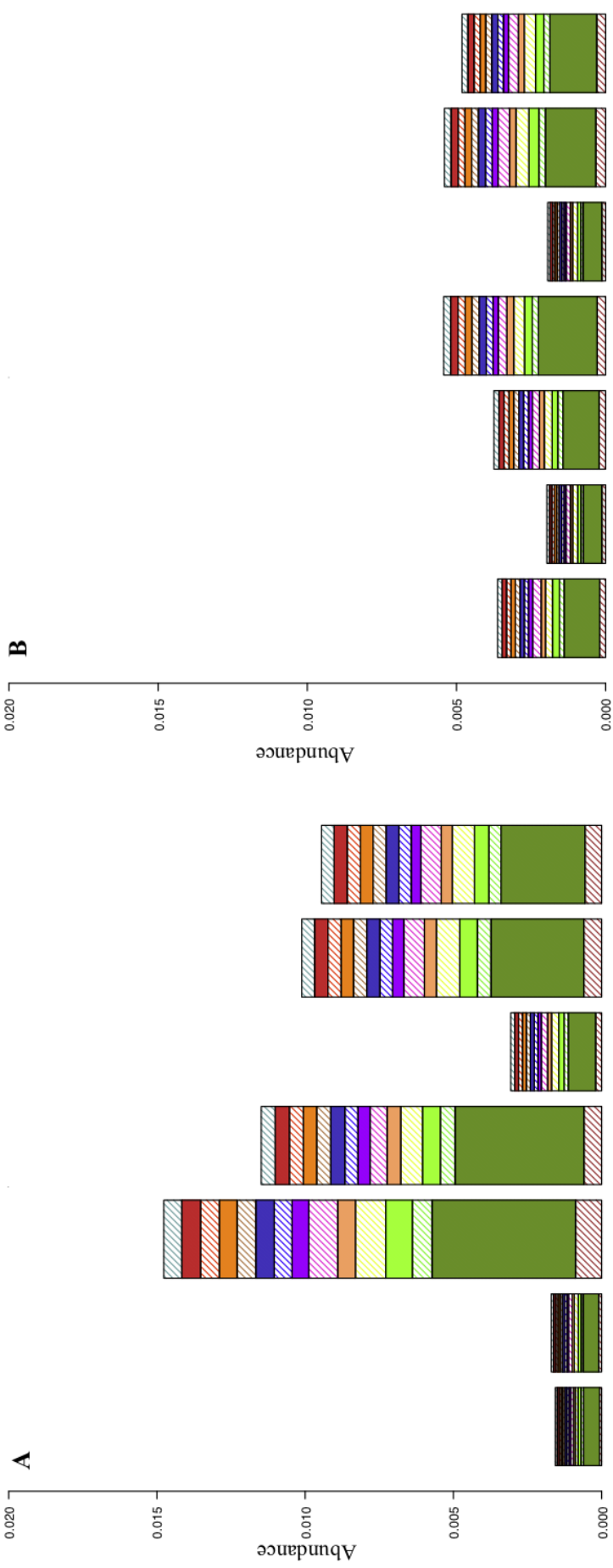


Figure 6.17: Stacked barplots showing the abundance of the statistically significant algal species: **A:** *^{13}C and* **B:** *^{12}C samples at various time points after spiking with ^{13}C labelled E.coli.*

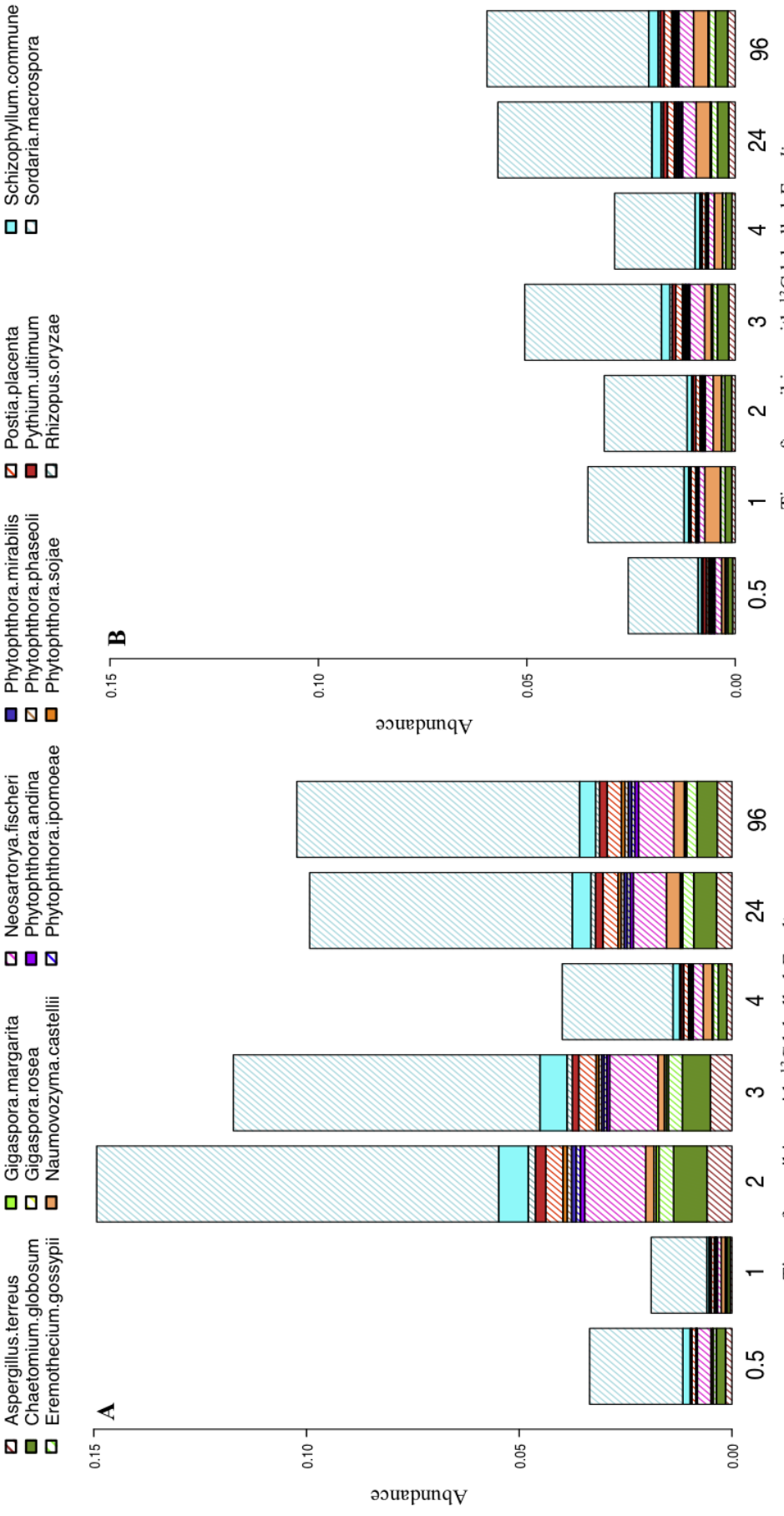


*Figure 6.18: Stacked barplots showing the abundance of the statistically significant fungal species: **A:** ^{13}C and **B:** ^{12}C samples at various time points after spiking with ^{13}C labelled E.coli.*

6.4 Discussion

6.4.1 Light Affects the Microbial Community but not Performance

Light exposure had significant spatial and temporal effects on the composition of the microbial community of laboratory-scale slow sand filters. Non-covered (exposed to light) LSSFs displayed a more even community composition in comparison to covered (no light exposure) LSSFs which were generally dominated by *Deltaproteobacteria*. Such dominance is not surprising as *Deltaproteobacteria* are either chemoorganotrophic or chemolithotrophic, hence do not require the presence of light to survive [Garrity et al., 2004]. Likewise the dominance of *Gammaproteobacteria* in LSSFs exposed to light can be explained by the fact that majority of this class are highly resistant to solar radiation [Alonso-Sáez et al., 2006]. In addition filters exposed to light displayed a biomass which was four times greater than that of covered LSSFs. This is unsurprising as light / dark environments are known to support and facilitate a wider range of metabolisms than environments of complete darkness. These results support the findings of Campos et al. [2002] who found that covered SSFs had a significantly lower overall biomass concentration than non-covered. A potential reason for this increased abundance is likely due to the presence of photochemical oxidation of dissolved organic matter (DOM) which has aided in releasing and making more organic matter available for bacterial metabolism and hence supports more growth and diversity [Judd et al., 2007].

Interestingly, differences in the microbial community composition and abundance had little effect on overall filter performance ($\bigtriangledown$). Such little difference in performance between covered and non-covered filters is in accordance with previous work by Campos et al. [2002] which found little difference in organic carbon removal between filter types. The work in this chapter is however, the first study to the author's knowledge which has found no difference in the removal capabilities of various water quality parameters between covered and non-covered slow sand filters. Furthermore, this study is the first to show that the microbial community of covered and non-covered SSFs are extremely divergent in composition and abundance yet achieve the same level of excellent filter performance, further evidence of neutral theory (theory that the bacterial community composition is a product of random events in connection to the recruitment of functionally equivalent bacterial taxa from the "metacommunity" [Langenheder et al., 2006]). It should however be noted that at the early

age bin (0-3 weeks) there was a significant difference between covered and non-covered LSSFs ability to remove coliforms, with non-covered filters removing 24 times more than covered filters. This in conjunction with the increased retention of viable *E.coli* in covered filters at all depths insinuates that light exposure facilitates *E.coli* removal. Numerous studies have shown that sunlight can improve *E.coli* removal; for example, UV has been shown to damage the cytoplasmic membrane, making bacteria more sensitive to the effects of other factors such as pH [Curtis et al., 1992]. Further, photo-oxidation [Curtis et al., 1992, Mezrioui et al., 1995] increased algal growth and subsequent algal toxin and reactive oxygen species (ROS) production have likewise been shown to cause bacterial death [Maynard et al., 1999].

6.4.2 Top-down Trophic Interactions are Essential for *E.coli* Removal

Identifying and unpicking trophic interactions, particularly those involved with pathogen removal, is an extremely complex and important question. Previously, mathematical models and work in simplified microcosms have shown the individual importance of viral lysis, protozoan grazing, and endogenous and exogenous reactive oxygen species in *E.coli* removal [Curtis et al., 1992, Bomo et al., 2004, Grobe et al., 2006, Liu et al., 2007, Kadir and Nelson, 2014]. Within this study the level of involvement in *E.coli* removal of each kingdom has been approximated from the metagenomic analysis and several assumptions. Furthermore, to optimise for the success of the DNA-SIP approach, i.e. obtain enough genomic material for metagenomic sequencing, a higher concentration of *E.coli* than normally found in surface water was spiked into the SSFs. However, this is the first study, to the author's knowledge, to examine and identify the ecosystem-wide trophic interactions and mechanisms responsible for pathogen removal in a "real" system, without prior bias as to which organisms and mechanisms to target. Additionally, this study is the first to show that it is possible to isotopically label *E.coli* and follow its removal through an ecosystem.

Based on the direct counts, qPCR (Figures 6.9 and 6.10) and DNA-SIP metagenomic sequencing analysis, the importance of top-down regulatory mechanisms for *E.coli* removal is apparent. Among a consortium of phages, protozoa, fungi and algae, which were ^{13}C-labelled (hence involved in *E.coli* removal / metabolism) and identified as highly significant via statistical analysis, *Enterobacteria Phage Lambda, Monosiga brevicollis* and *Tetrahy-*

mena spp. were identified as the main organisms responsible for *E.coli* removal. Based on our calculations (Appendix D) and direct count observations, protozoan grazing was responsible for more than 99% of the *E.coli* removal, which is consistent with previous investigations into pathogen removal in constructed wetlands [Weber and Legge, 2008] and estuaries [McCambridge and McMeekin, 1980]. From the twenty statistically significant protozoan species (Table 6.9) *Monosiga brevicollis* was predicted to be responsible for the majority (24.83%) of the *E.coli* removal once factors such as replication rate, progeny production, grazing rates and genome size were taken into consideration. This equated to approximately 326 times more *E.coli* removal than that achieved by *Enterobacteria Phage Lambda*. This lowered involvement of viral associated lysis for *E.coli* removal is also consistent with previous work [Withey et al., 2005].

The importance, and dominance, of *Monosiga brevicollis* from the beginning to the end of the experiment is not surprising as it is known to play a critical role in marine global carbon cycling [Yue et al., 2013]. Further, its dominance over other protozoan grazers including *Tetrahymena spp.* (responsible for 4.68% of *E.coli* removal) may be explained by their very short doubling time of 4.6h [Christaki et al., 2005] and fast grazing rates of 196 bacterial cells^{1}h [Parry, 2004], which in theory would allow them to outcompete other identified significant protozoan species. Although the feeding rate of *M.brevicollis* is slower than *Tetrahymena spp.*, such dominance could be explained by *Monosiga brevicollis*;

1. Possessing a microvilli collar, which hold bacteria from the water flow and allows them to be engulfed at a later time [Yue et al., 2013], hence providing energy storage for less plentiful times.

2. Possessing six oxidative stress genes; four of which are algal in origin - two ascorbate peroxidases and two metacaspases - that help protect the protozoa from various algal-mediated reactive oxygen species [Nedelcu et al., 2008].

Although protozoan grazing appeared to be the major route for *E.coli* removal, the role of viral-associated lysis cannot be overlooked, especially as *Enterobacteria Phage Lambda* was more abundant in ^{13}C labelled samples (Figure 6.13) and was identified by SIMPER analysis to contribute to more than 10% of the dissimilarity between ^{12}C and ^{13}C communities (Table

6.8).

Overall, with the exception of the *Microbacterium, Pseudomonas, Salmonella* and *Rhodococcus phages* the remaining eight statistically-significant phages are all known prophages of *E.coli* and have been shown to reduce *Enterobacteria* (including *E.coli*) numbers. However, unlike the protozoa the viral abundances changed dramatically with time (Figure 6.13), with increased abundance at 1, 4 and 96h and lower abundance at the remaining time points. Such behaviour suggests that these viruses are fluctuating between states of pseudolysogeny / lysogeny and lytic pathways (Figure 6.15), behaviours widely documented for environmental phages [Abedon, 2008]. This choice of life-cycle has been shown to aid in the regulation of bacterial biomass [Ripp and Miller, 1997], which is not taken into consideration in our calculations of removal. Additionally, the extremely high abundance of phages one hour after incubation with ^{13}C labelled *E.coli* was surprising; however, Zeng and Golding [2011] showed that *Enterobacteria Phage Lambda* can infect, replicate and enter the lysogenic cycle within *E.coli* after only 80 minutes. Therefore, although viral lysis has been calculated to be responsible for 0.14% of the *E.coli* removal (Appendix D) it is likely that the phages are supporting the regulations of the population (by allowing the *E.coli* population to recover) to ensure sufficient hosts for subsequent viral infection [Abedon, 2008]. In addition, the rapid appearance of *E.coli* phages in this experiment may reflect the heightened metabolic state of introduced *E.coli* at the start of the spiking period due to glucose availability.

6.4.3 Ecosystem-Wide Associations are Needed for *E.coli* Removal

Although top-down regulatory trophic interactions, such as protozoan grazing and viral lysis are the major mechanisms for *E.coli* removal, the importance of indirect and abiotic mechanisms cannot be overlooked. For example, previous work has shown algae to be actively involved in *E.coli* removal by the production of ROS [Curtis et al., 1992, Maynard et al., 1999, Feng et al., 2011], which causes lysis of *E.coli* and other bacterial species. In particular, extensive work has shown that *Chattonella marina* (one of the algae which showed significant increase in labelled samples (Figure 6.17)) produces several ROS species [Liu et al., 2007] known to significantly reduce coliform numbers. Therefore, it is conceivable that algae are actively participating in *E.coli* removal by indirect mechanisms (Figure 6.19). Furthermore, as eleven out of the fifteen significant algal species are mixotrophs (obtain en-

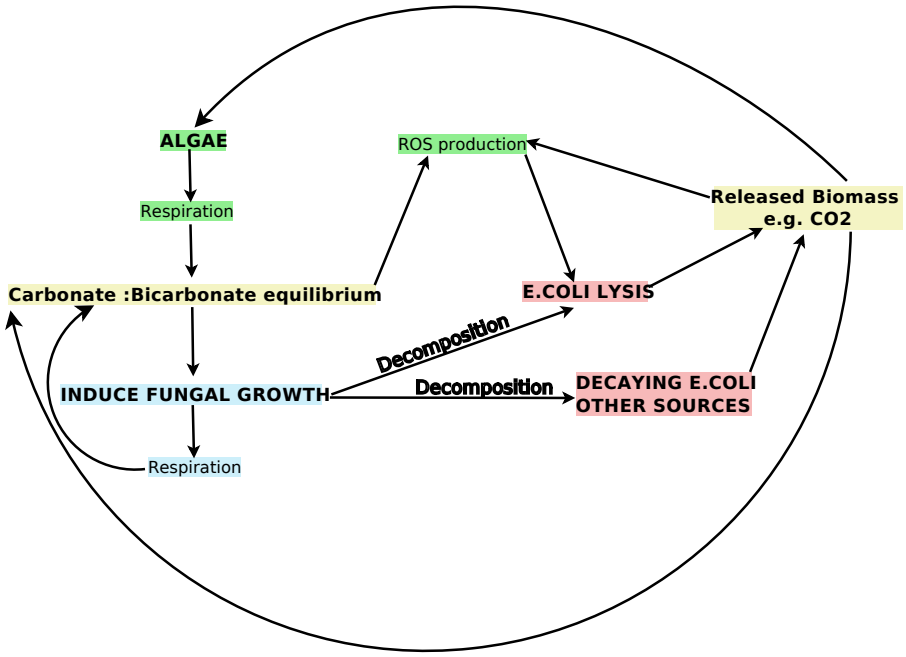


Figure 6.19: The mutulistic relationship of algae and fungi in the removal of E.coli

ergy from the sun or organic sources and obtain carbon from both inorganic and organic sources) [Nasr et al., 1968, Semple and Cain, 1996, Aamot, 2011] it is likely they accessed carbon from the released biomass from the ^{13}C labelled *E.coli* (via viral lysis and protozoan grazing).

In addition to indirect bacterial autolysis induced by algal ROS, these products help to explain the dominance of *M.brevicollis* (which contains ROS protection genes). Such autolysis, alongside protozoa and viral lysis of labelled *E.coli*, will have increased the amount of free ^{13}C labelled biomass components available for fungal degradation, explaining the dominance of the fast growing saprotrophs *Sordaria macrospora* and *Chaetomium globosum* [Kavak, 2012] which dominated the fungal ^{13}C community. This is likely to have amplified changes in the carbonate-bicarbonate equilibrium induced by algal respiration, which has been show to induce elevated growth rates and fruiting body formation in *Sordaria spp.* [Elleuche and Pöggeler, 2010]. This, in turn, results in further elevated CO_2 levels due to fungal respiration causing an additional imbalance in the carbonate-bicarbonate equilibrium and induces a knock-on effect to the water pH, which has been shown to inducing elevated algal ROS production [Liu et al., 2007]. Such an association helps to explain the apparent mutulistic relationship displayed by fungi and algae during *E.coli* removal (Figure 6.19). Nonetheless,

in addition to biological removal mechanisms, physical removal mechanisms such as straining, sedimentation and absorption also play an important part in pathogen removal in SSFs [Haarhoff and Cleasby, 1991].

In summary, it was possible to ascertain that *E.coli* removal was achieved by both direct (protozoan grazing and viral lysis) and indirect (lysis induced by algal ROS production and fungal degradation of released biomass) mechanisms (Figure 6.20). These mechanisms appeared to occur simultaneously with the involvement of species from various kingdoms, in particular Fungi and Algae, which exhibited mutualistic interactions. The highest removal of *E.coli* occurred between one and three hours after spiking. This level of removal at these time-points is consistent with the following characteristics of the ^{13}C communities:

1. Phages peaked in abundance at 1h, with extensive replication as part of their lytic pathway, resulting in reduced *E.coli* numbers after 2h;

2. Protozoa numbers peaked at 2-3h, allowing extensive grazing on ^{13}C labelled *E.coli* prior to this;

3. Algal abundance peaked at 2 and 3h, which was likely due to the increased availability of ^{13}C labelled CO_2 and other inorganics, created during viral lysis and protozoan grazing;

4. Fungal abundance peaked at 2 and 3h, when extensive reduction in *E.coli* numbers occurred, hence releasing biomass for decomposition and resulting in changes in carbonate-bicarbonate equilibrium inducing algal ROS production and further autolysis of *E.coli*

6.5 Conclusions

While various studies have shown the individual importance of viral lysis, protozoan grazing and endogenous, and exogenous, ROS in *E.coli* removal [Curtis et al., 1992, Bomo et al., 2004, Grobe et al., 2006, Liu et al., 2007, Kadir and Nelson, 2014], this is the first study, to our knowledge, to indicate the importance and interactions of all of these mechanisms for *E.coli* removal. Further, our approach enabled us to identify that the majority of the *E.coli* removal is due to top-down trophic interactions, such as protozoan grazing by *Monosiga brevicollis* and *Tetrahymena spp.* and viral lysis by *Enterobacteria phages*. Additionally,

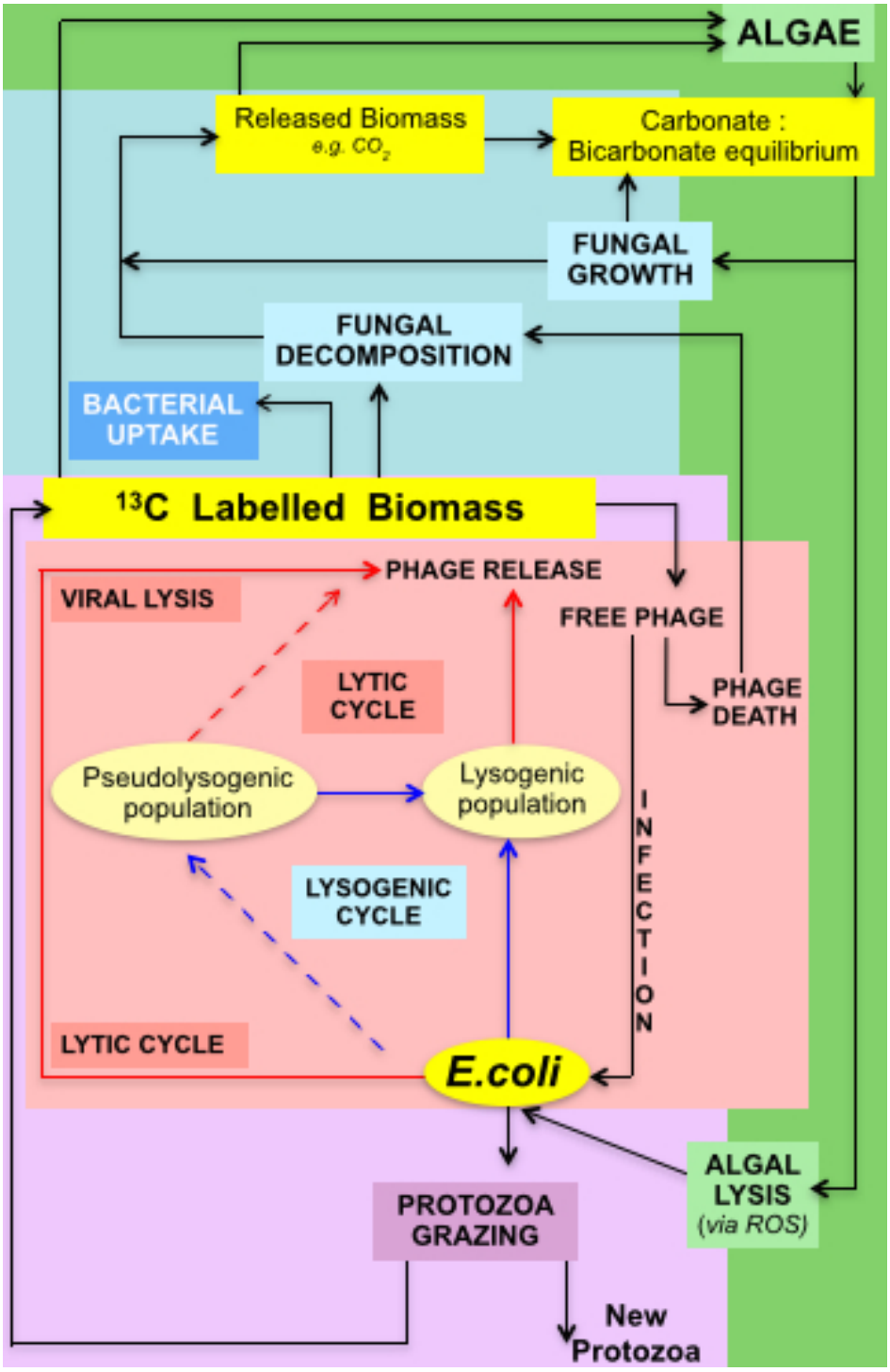


Figure 6.20: Food-web showing ecosystem wide involvement and trophic interactions in E.coli removal. Red shaded text represents viral lysis, purple shading represents protozoan grazing and blue and green shading corresponds to the mutulistic interactions between algae and fungi.

although *E.coli* K12 was used in this study it is highly likely that the mechanisms of removal of pathogenic strains of *E.coli* would follow similar routes. The protozoan grazers identified are non-specific grazers, affected only by the size of the prey community and the phages identified are *Enterobacteria*-specific, rather than species specific. However more work is required to determine if these removal mechanisms are similar for pathogenic and environmentally-persistent bacterial species.

This study has shown that SSFs provide an ideal laboratory-scale system to study relevant and functional food webs. By applying cutting-edge molecular methods to these systems we have furthered our understanding of the processes, mechanisms and organisms responsible for *E.coli* removal. The work and methodology adopted in this study will provide both a paradigm for similar studies and the opportunity to:

1. Design and improve pathogen removal and overall performance of new and existing water purification systems by managing the community;
2. Predict *E.coli* removal rates in natural treatment systems that have biological components, in particular during pollution and weather events;
3. Further our understanding of complex food webs and trophic interactions;
4. Create more complex and realistic trophic interaction models.

Future work should aim to develop more sophisticated trophic interaction models using data generated from DNA-SIP studies. These models should be further integrated into pathogen prediction models to allow better pathogen tracking and removal prediction. Finally, the conclusion that ecosystem-wide associations are essential for complete *E.coli* removal may help to explain the reduced performance of household purification systems, which support a less diverse ecosystem. Therefore, future work should investigate the ability and benefit that introducing a more complex community may create.

Chapter 7

Bioaugmentation of Slow Sand Filters with Estrogen Metabolisers

"All the world's a stage,
And all the men and women merely players;
They have their exits and their entrances;
And one man in his time plays many parts"
William Shakespeare, (As You Like It)

Exposure to endocrine-disrupting chemicals (EDCs), such as estrogens, is a growing issue for human and wildlife health as they have been shown to be the main causative agents for reproductive and developmental abnormalities in wildlife and plants and have been linked to increasing incidents of testicular cancer and other male infertility disorders in humans. With the rise of intensive farming, together with the increase in weather events such as storms, flash flooding, and land slides, the risk of estrogen exposure (via agricultural or wastewater runoff) into waterways typically used to supply slow sand filters is high. Therefore effective removal strategies are required. In this chapter the potential to improve natural estrogen (estrone, estradiol and estriol) removal and overall SSF performance by bioaugmentation with estrogen degrading bacteria was explored. In order to ensure optimal survival of the augmented bacterial strains within LSSFs, strains were isolated from full-scale SSFs. The

A condensed version of some of the work from this chapter is published: S. Haig, C. Gauchotte-Lindsay, G. Collins, R. Davies, and C. Quince. (2014). ***Progress in Slow Sand and Alternative Biofiltration Processes: Further Developments and Applications.*****, Chapter 28: Bioaugmentation Reduces Negative Effect of Estrogens on Coliform Removal in SSFs. IWA Publishing**

genomes of the estrogen degraders (E1-rm, E2-rm and E3-rm) used to augment the LSSFs all possessed several known enzymes involved in estrogen degradation. Concentrations of the natural estrogens determined by gas chromatography coupled with mass spectrometry (GC-MS) revealed that augmented filters reduced the overall estrogenic potency of the supplied water by 26% on average and removed significantly more estrone and estradiol than non-augmented filters. However this removal was significantly lower than that achieved in pure culture experiments. Additionally, a negative correlation was found between coliform removal and estrogen concentration in non-augmented filters. This was hypothesised to be due to the toxic inhibition of protozoa, suggesting the functional implications (impaired coliform removal) that high estrogen concentrations might have in SSFs. Consequently, we suggest that high estrogen concentrations could impact significantly on water quality production, and in particular on pathogen removal.

7.1 Introduction

The water industry faces a huge challenge in supplying a sustainable and safe supply of drinking water to a growing world population. Increasing demand has led to the reuse of various water sources of varying quality, including wastewater [Falconer et al., 2006]. However, along with increasing urbanisation and changes in agricultural practices this has led to reduced water quality mainly through anthropogenic contamination. Common and emerging contaminants include: various metals; carcinogenic organic compounds; synthetic chemicals; pharmaceuticals; veterinary growth stimulators; ingredients in personal care products; and food supplements [Kolpin et al., 2002, Caliman and Gavrilescu, 2009]. Today there is a growing body of scientific research indicating that these substances, in particular natural estrogens (estrone (E1), 17β-estradiol (E2) and estriol (E3)), may interfere with the normal function of the endocrine system of humans and wildlife by: (i) mimicking and/or antagonising the effect of endogenous hormones; (ii) disrupting the synthesis and metabolism of endogenous hormones and hormone receptors, resulting in various reproductive and developmental abnormalities and disorders [Goldman et al., 2000, Conroy et al., 2007, Caserta et al., 2008, Caliman and Gavrilescu, 2009, Alvarez et al., 2013]

Furthermore, since estrogens are excreted by all humans and animals, they can enter the

environment via several routes, particularly from sewage treatment works discharge (incomplete removal) and agricultural runoff. It is, therefore, unsurprising that recent surveys revealed broad occurrences of up to 85ng/L of E1, E2 and E3 in surface waters in the U.S.A., Pan-European area, and Asia (Table 7.3) [Ternes et al., 1999, Kuch and Ballschmiter, 2001, Matthiessen et al., 2006, Zhao et al., 2009, Wang et al., 2011]. Therefore, due to increasing concerns about the adverse health effects posed by natural estrogens, the US EPA recently added E1, E2, and E3 onto its Contaminant Candidate List 3 [U.S. Environmental Protection Agency, 2009b]. Likewise, the European Union's Water Framework Directive added E2 as a "Hazardous" substance, meaning that Member States must include removal measures for it from surface water and wastewater discharge by 2015 and meet the defined environmental quality standards by 2021 [European Commission Scientific Committee on Health and Environmental Risks, 2011].

Regardless of the potential danger posed by estrogens little research has been conducted into their effect and impact on the biological engineered systems used to remove them. Therefore, as reclaimed wastewater and other surface waters are likely to be needed to supplement the existing drinking water supply, urgent research is required to determine how estrogen degradation can be improved or introduced into existing biological water purification systems.

Numerous studies have been carried out to determine how to remove estrogens from wastewater using highly energy-intensive processes. However, to our knowledge the ability of SSFs to transform or remove such contaminants has never been investigated. Previous studies in wastewater treatment systems have shown that EDC removal can be enhanced by the introduction / augmentation of specific strains of degrading bacteria into the systems (bioaugmentation) [Iasur-Kruh et al., 2011]. For example Hashimoto et al. [2009] and Roh and Chu [2011] demonstrated that bioaugmentation of conventional activated sludge systems and sequencing batch reactors, respectively, with *Novosphingobium sp.* strain JEM-1 and *Sphingomonas* strain KC8, increased estradiol removal significantly. Furthermore, bioaugmentation has been shown to be a successful approach for the removal of various contaminants such as steroid hormones [Iasur-Kruh et al., 2011], petroleum hydrocarbons [Mishra et al., 2001] and toluene [Kim et al., 2004] in various environments, including wastewater purification systems [Runes et al., 2001, Lorah and Voytek, 2004, Ren et al., 2007, Hashimoto and

Murakami, 2009, Silva et al., 2012, Yu et al., 2013]. Unlike other water treatment methods, slow sand filtration requires little energy input (as discussed in Chapter 2). This, along with their extremely diverse microbial community hence catabolic potential, makes them an attractive candidate for bioaugmentation with EDC degraders, potentially resulting in effective and economical EDC removal.

In Chapters 4 and 5, laboratory-scale filters were compared to full-scale SSFs using phylum-specific qPCR primers, and both 454 and Illumina sequencing, revealing that the microbial communities underpinning SSFs are extremely complex, and phylogenetically and metabolically diverse. However, from a functional perspective, determining whether removal of emerging contaminants (e.g., estrogens) can be achieved and/or enhanced through bioaugmentation of the SSF microbial communities is important. Furthermore, analysis of the functional impact of exposure to these contaminant on existing treatment systems is required, especially as the list of contaminants included in water quality guidelines expands. Therefore, this chapter will focus on understanding the functional effect that estrogen exposure induces on SSFs performance, and determine whether bioaugmentation with known estrogen metabolisers can improve estrogen removal and overall filter performance.

Hypotheses

Hypothesis 7.1 *Full-scale SSFs do contain estrogen degrading bacterial species.*

Hypothesis 7.2 *Bioaugmentation of SSFs with estrogen degrading bacteria will successfully enhance the estrogen removal capacity of SSFs.*

Hypothesis 7.3 *The estrogen removal capacity of the laboratory-scale SSFs augmented with estrogen degraders will be lower than the estrogen removal capacity of the estrogen degrading bacteria in pure culture.*

7.2 Background

In this section a detailed literature review providing an overview of the sources, fate in the environment, and techniques for monitoring natural estrogens is provided.

Table 7.1: Daily excretion (μg) of estrogenic hormones, adapted from Ying et al. [2002]

Category	E1	E2	E3	EE2
Males	3.9	1.6	1.5	-
Menopausal females	4	2.3	1	-
Menstruating females	8	3.5	4.8	-
Pregnant women	600	259	6000	-
Women using oral contraceptives	-	-	-	35

7.2.1 Natural Estrogens

Naturally occurring estrogens (estrone (E1), 17β-estradiol (E2) and estriol (E3)) are predominantly female hormones, which are important for maintaining healthy reproductive tissues, breasts, skin and brains and are released daily in urine and faeces in varying amounts (Table 7.1). Natural estrogens are collectively known as steroid hormones, a group of biologically active compounds that are synthesised from cholesterol and possess a cyclopentan-o-perhydrophenanthrene ring [Ying et al., 2002]. Each carbon in this fused-ring structure is assigned a number identifier, and each ring, a letter (Figure 7.1), the natural estrogens differ only in the number of hydroxyl groups they possess (Table 7.2). In humans and animals, estrogens undergo various transformations mainly in the liver, where they are frequently oxidised, hydroxylated, deoxylated and methylated prior to final conjugation with glucuronic acid.

Due to the physicochemical properties of E1, E2 and E3 (Table 7.2) they are considered hydrophobic, low-volatile chemicals with varying degrees of potency (amount needed to induce transcriptional activation). E2 is considered the most potent of the natural estrogens with 1.47ng inducing estrogen receptor transcriptional activation in humans, although this

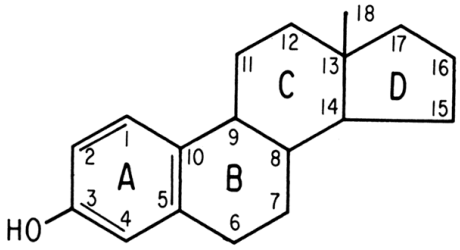
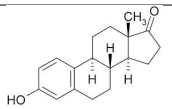
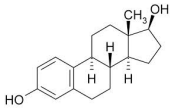
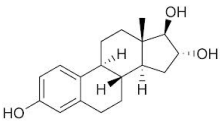


Figure 7.1: Basic estrogen molecule. All natural estrogen molecules have 18 carbon atoms and differ only in the number of hydroxyl groups.

Table 7.2: Physiochemical properties of natural estrogens, adapted from Silva et al. [2012]

Estrogen	MW[1]	pKa	log K_{OW}[2]	EC_{50}[3]	REA[4]	Structure
Estrone (E1)	270.4	10.3 - 10.8	3.43	122ng	0.20	
17β-estradiol (E2)	272.4	10.5 - 10.7	3.94	1.47ng	1	
Estriol (E3)	288.4	10.4	2.81	10.3ng	0.01	

[1]Molecular Weight; [2]Octanol / water partition coefficient; [3]half maximal effective concentration required to induce fish endocrine receptor gene transcription, adapted from Alvarez et al. [2013]; [4]Relative (to E2) Estrogenic Activity based on the Yeast Estrogen Screen [Conroy et al., 2007]

is considerably lower than that of the other estrogens [Alvarez et al., 2013]. It is important to recognise that E1 and E2 can be readily converted to each another and E1 can be further converted into E3, with many other polar metabolites, such as 16-hydroxy-estrone, 16-ketoestrone or 16-epiestriol being formed and excreted with varying levels of potency.

7.2.2 Estrogen in the Environment

In aquatic ecosystems, the major sources of estrogens are considered to be from discharge after incomplete treatment removal from wastewater treatment plants (WWTs) [Silva et al., 2012]; and agricultural runoff [Falconer et al., 2006] (in particular from concentrated animal feeding operations) as steroid hormones are frequently used to control the reproductive systems of livestock [Ying et al., 2002] (Figure 7.2). Therefore, the increase in intensive farming and incomplete removal by WWTs has led to reports of estrogenic steroids being found in surface water, ground water and drinking water in the low ng/L range (Table 7.3) [Shore et al., 1995, Ternes et al., 1999, Kuch and Ballschmiter, 2001, Ying et al., 2002, Quintana et al., 2004, Langston et al., 2005, Zhao et al., 2009].

The occurrence of endocrine-disrupting chemicals (EDCs) in the environment is a growing issue for human and wildlife development and reproduction [Shore et al., 1995, Belfroid et al., 1999, Ternes et al., 1999, Jobling et al., 2006, Zhou et al., 2010]. In particular, EDCs alongside synthetic xeno-estrogens, such as 17β-ethynylestradiol (EE2), bisphenol A,

4-nonylphenol, 4-tert-octylphenol and the more potent natural estrogens (E1, E2 and E3) [Ternes et al., 1999, Langston et al., 2005, Zhou et al., 2010] have been shown to be the main causative agents for the feminisation of fish, and reproductive and developmental abnormalities, and disorders, in wildlife and plants at pg/L concentrations [Shore et al., 1995, Ternes et al., 1999, Jobling et al., 2006, Zhou et al., 2010]. Further, in humans they have been linked to decreasing sperm counts and increasing incidents of testicular cancer and other male infertility disorders [Ternes et al., 1999]. Despite the range of studies that have examined the effects of estrogens on fish, other aquatic organisms and to a lesser extent, humans [Caldwell et al., 2010], no study to date has examined the impact of estrogen exposure on the performance and microbial community of biological engineered water treatment systems.

7.2.3 Degradation and Removal of Estrogen

Estrogens in water can been removed using physiochemical treatment processes, such as membrane separation [Cartinella et al., 2006, Xue et al., 2010], advanced oxidation [Liu et al., 2009, Stalter et al., 2011, Silva et al., 2012] and granular activated carbon [Le Noir et al., 2007, Stalter et al., 2011], with varying levels of removal success (19-92%) [Iasur-Kruh et al., 2011]. However, from the numerous studies undertaken to investigate the fate and transport of estrogens in the natural environment [Yu et al., 2013], biodegradation has been shown to be the primary removal mechanism. Collectively these studies have shown that both an increased hydraulic retention time and the presence of estrogen-degrading bacteria are necessary for optimal removal [Roh and Chu, 2011].

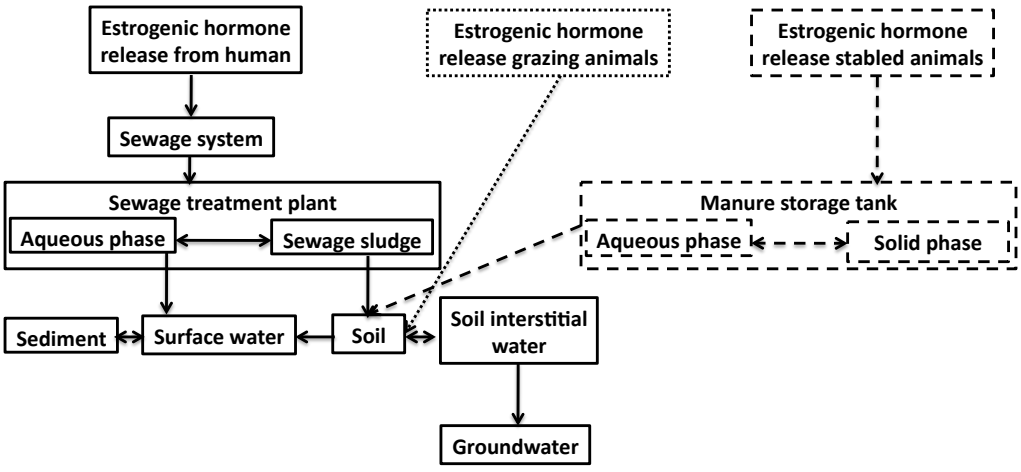


Figure 7.2: Routes by which estogenic hormones enter the environment

Table 7.3: Maximum and minimum concentration of natural estrogens found in surface waters and sewage treatment plant effluents from around the world.

Location	Sample type	E1 (ng/L)	E2 (ng/L)	E3 (ng/L)	Reference
China	Pearl and Liao River	0-78.7	0.9-7.7	NM	[Zhao et al., 2009, Wang et al., 2011]
Germany	River water (31 samples)	0.1-18	0.15-5.2	NM	[Kuch and Ballschmiter, 2001]
Japan	Tama River	0.01-85.6	0.01-12.3	NM	[Wang et al., 2011]
Netherlands	Coastal/Estuarine water (11 samples)	0.1-3.4	0.3-5.5	0.1-4.3	[Ying et al., 2002]
UK	River water (28 sites)	0.07-3.33	0-0.9	0-1.2	[Xiao et al., 2001]
USA	River water (139 sites)	<5-112	<5-93	<5-51	[Kolpin et al., 2002]
USA	Mississippi River	0.02-4.7	0.02-4.5	NM	[Zhang et al., 2007]
USA	Drinking water (10 samples)	0.03-0.1	0.01-0.02	0.01-0.02	[Wise et al., 2011]
Taiwan	Wulo Creek - downstream of CAFO[1] (54 samples)	7.4-1267	9.3-313.6	0-210	Chen et al. [2010]
UK	River downstream of farm runoff (25 samples)	1.46-9.31	0.69-3.62	NM	[Matthiessen et al., 2006]
UK	Poultry litter runoff	32.5-3500	NM	NM	[Langston et al., 2005]
USA	8 CAFO[1] (swine lagoon)	1100-17400	194-3900	47-6290	[Hutchins et al., 2007]
USA	8 CAFO[1] (cattle runoff)	20-102	8-200	ND	[Hutchins et al., 2007]
Canada	Effluent from STW[2]	2.5-82.1	0.44-3.3	0.43-18	[Baronti et al., 2000]
Germany	Effluent from STW[2]	1-70	<1-3	NM	[Ternes et al., 1999]
Germany	15 Rivers subjected to STW[2] discharge	<0.1-1.6	NM	NM	[Ternes et al., 1999]
Spain	Effluent from STW[2]	<0.2-1.8	ND	NM	[Kuster et al., 2008]
Spain	Llobregat River subjected to STW[2] discharge	0.7-1	ND	NM	[Kuster et al., 2008]
UK	Effluent from 3 STW[2]	<0.4-12.2	<0.4-4.3	NM	[Williams et al., 2003]
UK	River subjected to STW[2] discharge	<0.4-2.5	<0.4-2.3	NM	[Williams et al., 2003]

NM: Not Measured; ND: Not Detected; [1]Concentrated Animal Feeding Operation; [2]Sewage Treatment Works

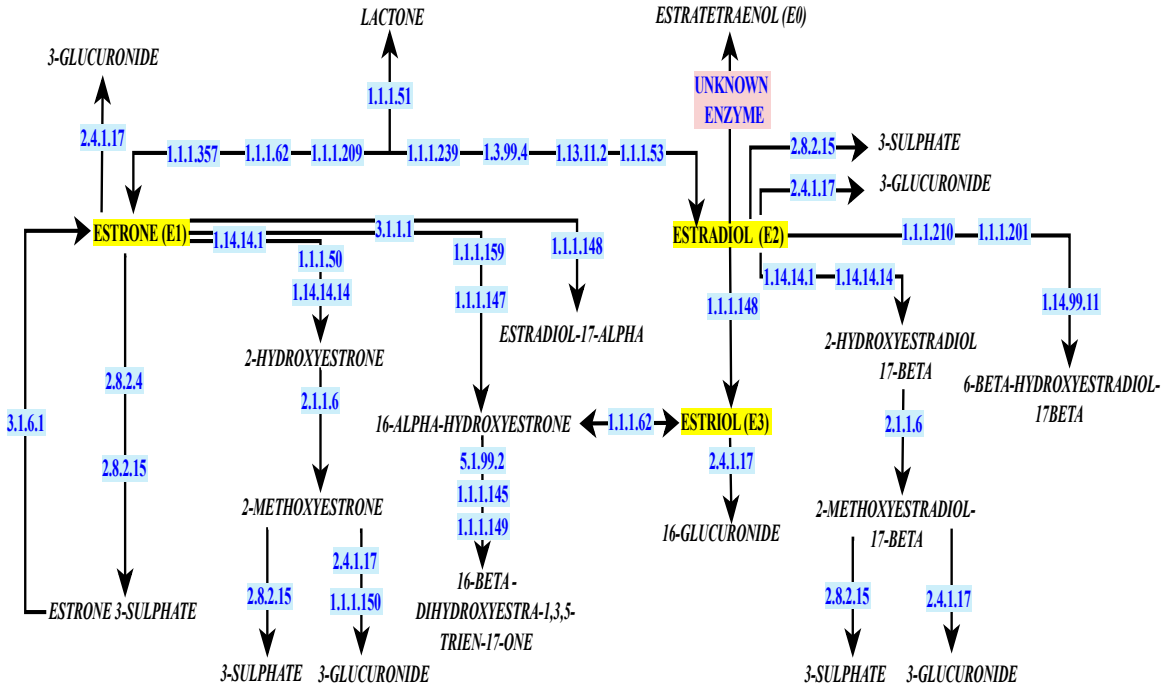
The ability of bacteria to degrade aromatics as their sole carbon source was first demonstrated in 1908 by Stromer, who isolated *Bacillus hexacarbovorum*, which grew on toluene and xylene. This discovery fuelled research into isolating other microorganisms capable of degrading other aromatic compounds, such as estrogens. Biodegradation is a *"process by which microbial organisms transform or alter (through metabolic or enzymatic action) the structure of chemicals introduced into the environment"* [U.S. Environmental Protection Agency, 2009c]. Since the pioneering work of Stromer there has been extensive research into the biodegradation of estrogens in activated sludge treatment, membrane bioreactors, fixed bed reactors and fluidized bed reactors [Ternes et al., 1999, Ren et al., 2007, Hashimoto and Murakami, 2009, Silva et al., 2012]. Although biological estrogen removal has not been investigated in the context of drinking water treatment, studies monitoring the performance of wastewater treatment systems have shown varying degrees of both natural and synthetic estrogen removal. However, these systems either require energy (electricity) or chemical input to facilitate removal and hence are extremely costly. The need for cheaper water purification has led to recent studies, in which existing systems have been bioaugmented with EDC

degrading microbes (*Rhodococcus zopfii, Pseudomonas aeruginosa, Sphingobacterium sp. JCR5, Phyllobacterium myrsinacearum, Sphingomonas KC8*) with varying levels of success [Pauwels et al., 2008, Hu et al., 2011, Silva et al., 2012].

Microorganisms can degrade steroid hormones using two possible degradation mechanisms: growth-linked (metabolic) and non-growth-linked (co-metabolic) [Yu et al., 2007]. Growth-linked degradation utilises the steroid hormones as the energy / carbon source, whereas during co-metabolism the steroid hormone is degraded by an enzyme or co-factor produced during metabolism of another carbon source (e.g. glucose) [Madigan et al., 2011]. As co-metabolic degradation yields no carbon or energy benefits to the degradative microorganisms, a primary growth substrate is needed for sustainable bacterial growth and is required to induce the expression of degradative enzymes. Previous studies have shown that E2 can be co-metabolically degraded by heterotrophic bacteria [Yu et al., 2007, Pauwels et al., 2008] with increased removal being observed during the presence of non-specific monooxygenases, such as ammonia monooxygenase [Shi et al., 2004, Yu et al., 2013]. Generally, details pertaining to the exact pathways and enzymes involved in estrogen degradation are limited, however some of the identified enzymes are shown in Figure 7.3. In general, the first stages of estrogen degradation can be divided into four groups: hydroxylation of the A ring at C4, hydroxylation of the saturated ring, dehydration of the D ring at C17 and dehydrogenation of the D ring at C17 [Overbeek et al., 2005, Hu et al., 2011, García-Gómez et al., 2012, Lloret et al., 2012, Yu et al., 2013]. Additionally, many microbes known to survive in high steroid hormone environments possess several multidrug efflux pumps (e.g. AcrAB-TolC and EmrAB-TolC) of which steroids are the substrate [Elkins and Mullis, 2006, García-Gómez et al., 2012].

7.2.4 Analysing and Measuring Estrogens in Environmental Samples

Accurately detecting estrogens in environmental samples is challenging due to the low concentration at which they are commonly found. Therefore several analytical methods such as, gas chromatography / mass spectrometry (GC/MS); GC/MS/MS; high performance liquid chromatography (HPLC); and liquid chromatography (LC-MS/MS) have been developed to detect and quantify estrogens [Liu et al., 2011] to the low and sub-nanogram-per-litre scale [Briciu et al., 2009]. However, these methods on their own are often insufficiently sensitive



EC Number	Name	Reference for Involvement with Estrogens
1.1.1.50	Hydroxysteroid dehydrogenase	Hu *et al.,* (2011)
1.1.1.51	3(or 17)Beta-hydroxysteroid dehydrogenase	Hu *et al.,* (2011)
1.1.1.53	3Alpha(or 20beta)-hydroxysteroid	Overbeek *et al.*, (2005)
1.1.1.62	17-Beta-estradiol dehydrogenase	Overbeek *et al.*, (2005)
1.1.1.145	Delta5-3-beta-hydroxysteroid dehydrogenase	Overbeek *et al.*, (2005)
1.1.1.147	16-Alpha-hydroxysteroid dehydrogenae	Hu *et al.,* (2011)
1.1.1.148	Estradiol 17alpha-dehydrogenase	Overbeek *et al.*, (2005)
1.1.1.149	20-Alpha-hydroxysteroid dehydrogenae	Hu *et al.,* (2011)
1.1.1.150	21-hydroxysteroid dehydrogenae	Hu *et al.,* (2011)
1.1.1.159	7Alpha-hydroxysteroid dehydrogenae	Overbeek *et al.*, (2005)
1.1.1.201	7Beta-hydroxysteroid dehydrogenae	Hu *et al.,* (2011)
1.1.1.209	3Alpha-hydroxysteroid dehydrogenae	Overbeek *et al.*, (2005)
1.1.1.210	3Beta-hydroxysteroid dehydrogenae	Hu *et al.,* (2011)
1.1.1.239	3Alpha(17beta)-hydroxysteroid dehydrogenase	Hu *et al.,* (2011)
1.1.1.357	3Alpha-hydroxysteroid 3-dehydrogenase	Garcia-Gomez *et al.,* (2013)
1.3.99.4	3-ketosteroid-delta1-dehydrogenase	Hu et al., (2011) and Sang *et al.,* (2011)
1.13.11.2	Catechol 2,3-dioxygenase	Hu *et al.,* (2011)
1.14.14.1	Xenobiotic monooxygenase	Overbeek *et al.*, (2005)
1.14.14.14	Aromatase	Overbeek *et al.*, (2005)
1.14.99.11	Estradiol 6beta-hydroxylase	Overbeek *et al.*, (2005)
2.1.1.6	Catechol O-methyltransferase	Overbeek *et al.*, (2005)
2.4.1.17	Estradiol glucuronosyltransferase	Overbeek *et al.*, (2005)
2.8.2.15	Steroid sulfotransferase	Overbeek *et al.*, (2005)
2.8.2.4	Estrone sulfotransferase	Overbeek *et al.*, (2005)
3.1.1.1	Carboxylesterase	Sang *et al.,* (2011)
3.1.6.1	Arylsulfatase	Overbeek *et al.*, (2005)
5.1.99.2	16-hydroxysteroid epimerase	Overbeek *et al.*, (2005)
1.14.99.39	Ammonium monoxygenase[1]	Shi *et al.,* (2004)
1.10.3.2	Laccase[2]	Eibes et al., (2012)
NA	AcrAB-TolC Multi drug efflux pump[3]	Garcia-Gomez *et al.,* (2013)
NA	EmrAB-TolC Multi drug efflux pump[4]	Elkins and Mullis, (2006)

Figure 7.3: Estrogen metabolism pathways, adapted from the KEGG estrogen metabolism pathway, blue text represent the EC number of the enzymes involved. The bottom four features of the list cannot be found in the pathway diagram, however have been shown to be important for estrogen metabolism or for survival in high estrogen concentrated environments: [1] *and* [2] *have been shown to improve the rate of estrogen metabolism,* [3] *and* [4] *are multidrug efflux pumps which have been shown to provide resistance to steroid hormones.*

and, thus, pre-concentration and purification steps, such as solid phase extraction (SPE) are required.

GC/MS has been widely used for the analysis of natural estrogens and other EDCs in urine, meat and environmental samples [Shore et al., 1995, Kuch and Ballschmiter, 2001, Quintana et al., 2004, Liu et al., 2011], although it has lost popularity in recent years to LC-MS, which has greater accuracy at the lower limits of detection [Nováková and Vlčková, 2009]. Gas chromatography mass spectrometry is an instrumental technique, comprising a gas chromatograph (GC) coupled to a mass spectrometer (MS). GC/MS allows complex mixtures of chemicals to be separated (due to differences in molecular weight and boiling point), identified and quantified (due to specific ion fingerprints, generated from ionisation). It should, however, be noted that for a compound to be analysed by GC/MS it must be sufficiently volatile and thermally stable. Therefore, in order to make the low volatile estrogens more compatible with GC/MS (improved stability; improved chromatographic separation; and the production of more favourable diagnostic fragmentation patterns) derivatisation is necessary [Shareef et al., 2006, Briciu et al., 2009].

N,O-bis(trimethylsilyl) trifluoroacetamide (BSTFA), N-methyl-N-(trinethylsilyl) trifluoroacetamide (MSTFA) and N-(tert-butyldimethylsilyl)-N-methyl trifluoroacetamide (MTBSTFA) are the most commonly used derivatisation reagents for estrogens [Quintana et al., 2004]. These agents all derivatise via silylation, which is the most common form of derivatisation. Silylation is the process by which the active proton in organic compounds containing active hydrogen atoms (e.g. -OH, -NH, -NH_2, -SH, -COOH) are displaced by the silyl group.

7.3 Materials and Methods

7.3.1 Enrichment of Estrogen Metabolising Bacteria

In order to optimise the survival and success of bioaugmentation with estrogen degrading bacteria, hence estrogen removal in SSFs, it was decided that the best approach was to augment with bacteria which originate from full-scale SSFs. Therefore, in order to isolate such bacterial strains a traditional enrichment based isolation procedure was chosen.

7.3.1.1 Minimal Medium Composition

To obtain pure cultures of estrogen-metabolising bacteria from SSFs, an enrichment culture procedure was adopted. This entailed using a minimal medium (Table 7.4) adapted from Bold's basal medium [Shelton and Tiedje, 1984]. After medium preparation, 100μg/L of either E1, E2 or E3 and 0.01g/L ammonium chloride were dissolved in autoclaved 1L bottles of the minimal medium as the sole carbon and nitrogen sources. In addition, trace elements and vitamin stocks were added to the medium, which was then adjusted to a pH of 7.5. Prepared medium was aliquoted (9ml) into 20ml vials and inoculated with 1g of sand taken from the full-scale SSFs described in Chapter 4. Each enrichment culture was set up in complete darkness at both room temperature and 4°C. Every three weeks over a year, 1ml of each culture was subcultured into fresh medium following the enrichment procedure described by Madigan et al. [2011]. Enrichments were routinely Gram stained to check for purity.

7.3.1.2 Isolation of Estrogen Metabolising Bacteria

After a year of subculturing, the final vial of each enriched culture condition were used as a source for the isolation of E1, E2 or E3 degrading isolates. One millilitre of these was used to streak on minimal media agar (1.5%) plates composed of the same components as described in Table 7.4. The plates were incubated at the temperature that the original culture was enriched from (room temperature or 4°C) in total darkness for seven days. This was repeated twice, after which morphologically distinct colonies were selected from the room temperature and 4°C, E1, E2 and E3 cultures. The six isolated strains were named, E1-rm,

Table 7.4: Chemical composition of minimal medium. Bold font denotes ingredient added after preparation. Adapted from Shelton and Tiedje [1984]

Component	Amount in stock
$MgSO_4.7H_2O$	100mg/L dH_2O
K_2HPO_4	0.35g/L dH_2O
KH_2PO_4	0.27g/L dH_2O
$MgCl_2$	100mg/L dH_2O
$CaCl_2.2H_2O$	75mg/L dH_2O
$FeSO_4.7H_2O$	14mg/L dH_2O
$MnCl_4.4H_2O$	0.05mg/L dH_2O
$NaHCO_3$	**1.2g/L dH_2O**

Nitrogen concentrations based on those used by Suneethi and Joseph [2010]

E2-rm, E3-rm, E1-4°C, E2-4°C and E3-4°C based on the estrogen they degrade, and the temperature they were isolated from, respectively. These six isolates were grown overnight in liquid medium and used to prepare cryogenically (80°C) preserved stocks of the strains (15% of glycerol (v/v)).

7.3.1.3 Phylogenetic Identification of Enriched Isolates

The DNA of the six enriched isolates (E1-rm, E2-rm, E3-rm, E1-4°C, E2-4°C and E3-4°C) were obtained using FastDNA Spin Kit for soil (MP Biomedcal, Cambridge, UK) according to manufacturers' instructions. The 16S rRNA gene fragments were obtained by amplifying the 16S rRNA gene with the following reaction mix: 200 ng DNA, 25 μl Bioline PCR mix and 12.5 pmol of both the forward, 27F (5'-GAGTTTGATCCTGGCTCAG-3') and reverse primer 1392R (5'-ACGGGCGGTGTGTRC-3') in a Gene Pro thermal cycler (Bioer Technology, UK). PCR amplification was carried out under the conditions described by McHugh et al. [2004]. DNA sequencing of the PCR fragments was carried out by Source Bioscience, LifeSciences (Nottingham). After chimera checking, taxonomic classifications were assigned using the RDP classifier [Cole et al., 2009] using an 80% confidence threshold.

7.3.1.4 Whole-Genome Assembly and Annotation

Total genomic DNA from the pure cultures of the three room temperature estrogen metabolising enrichment cultures (E1-rm, E2-rm and E3-rm) were used to make Illumina metagenomic library preparations. Illumina libraries were prepared using the Nextera Standard kits (Illumina), following the same methodology discussed in Section 6.2.6, with the exception that 25μL (2ng/μL) of extracted DNA were tagmented. Sequenced reads for each sample were trimmed to remove adaptors and quality filtered (using their Phred score, ≤ 25 are removed). Trimmed and quality checked paired end sequences were then used for *de novo* assembly using the program Velvet [Zerbino and Birney, 2008]. Velvet works by converting reads into k-mers using a hash table and then assembling overlapping k-mers into contigs (velvetOptimiser was used to choose the optimal parameters for assembly)[1]. Genome assembly was then performed by running the newly generated contigs through the Rapid Annotation using Subsystem Technology (RAST) program [Aziz et al., 2008] against suitable reference genomes (based upon their 16S rRNA characterisation (Section 7.3.1.3)).

[1]These stages were conducted by Dr. Umer Ijaz, University of Glasgow

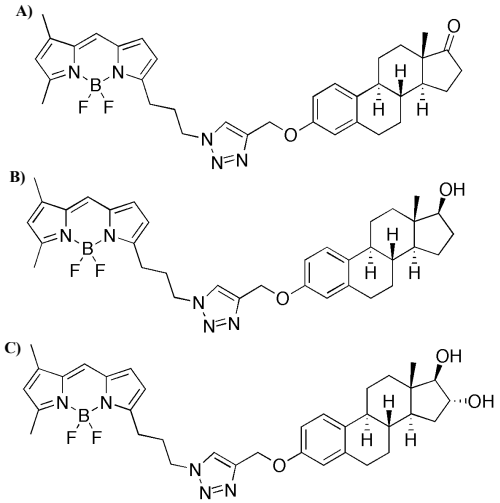


Figure 7.4: Chemical structure of the flourescently tagged estrogens, produced via "click" chemistry using an alkyne linker between the estrogen (A:Estone, B: Estradiol and C: Estriol) and the BODIPY-FL fluorophore (Life technologies, UK), produced by Dr. Alan Sewell, University of Glasgow.

7.3.2 Growth Kinetics of Estrogen Metabolisers

The three room-temperature estrogen enrichment cultures were grown in 250 ml minimal medium (Section 7.3.1.1) with either: 5 μg/L estrone, 3.9 μg/L estriol or 1.2 μg/L estradiol as their sole carbon source for 168 h at 18°C in darkness. E1, E2 and E3 concentrations were measured following the timeframe used in previous E2 bacterial-degrading studies [Pauwels et al., 2008] (0, 2.5, 5, 12, 24, 48, 72, 96, 120 and 168h) by collecting 10ml from each culture. This 10ml sample was filtered (0.45 μm cellulose filter) and then used for solid phase extraction, followed by GC/MS quantification (Section 7.3.3). Similarly the same procedure was repeated using the more environmentally relevant concentration of 500ng/L for each of the estrogens, as performed by Pauwels et al. [2008]. In addition to this, in order to visually detect estrogen metabolism, fluorescently labelled E1, E2 and E3 (Figure 7.4) at the same concentrations as mentioned above, were also cultured with the enriched isolates. All cultures were grown in triplicate, alongside negative (sterile media) and positive controls (two known E1 and E2 metabolsing *Acinetobacter* and *Pseudomonas* strains obtained from Pauwels et al. [2008]).

To determine biomass growth, and uptake of the fluorescent estrogens, the OD_{600nm} was measured alongside fluorescence microscopy. Epifluorescence microscopy was performed following an adaptation of the procedure used by Bomo et al. [2004]. Briefly, 1ml of the experimental cultures were filtered through a 0.1μm Anodisc 25-mm membrane filter (What-

man, Maidstone, UK) with a slight vacuum (20 kPa). The Anodisc membrane was filtered to dryness, placed on a microscope slide and mounted in Vectashield (Vector Laboratories, UK). Digital images of green fluorescent bacteria on the membrane were captured using an inverted optical microscope (Inverted Olympus IX71) at 100x magnification.

7.3.3 Quantifying Estrogen Via GC/MS

GC/MS analysis of natural estrogens was performed following an adapted procedure of that used by Quintana et al. [2004]. Briefly, 10 ml of filtered medium from the growth kinetic cultures were forced through a 60mg HLB Oasis SPE cartridge (Waters, UK) (approximately at 15-20 ml/min) that had been sequentially pre-conditioned with 3ml ethyl acetate and 3ml Milli-Q water adjusted to the same pH as the sample. After finishing the concentration step, cartridges were dried under a stream of nitrogen for 30 min and eluted with 1ml of ethyl acetate. Prior to GC/MS analysis the resulting analytes were further concentrated by evaporating to dryness under a constant nitrogen stream and resuspended in 100μl of ethyl acetate, which were derivatised with 200μl of MSTFA (Fisher Scientific, UK) at 85°C for 100min. Derivatised samples were analysed using GC (Agilent 7890A) system equipped with a HP-5MS capillary column (30m $\times$ 0.25mm $\times$ 0.25μm) coupled to a single quadrupole (Agilent 3157 5975c Inert XL EI/CI MSD) run in splitless full scan mode (mass range 50-500mz). Compounds were separated using the following oven program: 1min at 50°C, first ramp at 10°C/min to 220°C, second ramp at 5°C/min to 280°C (held for 10min). The GC/MS interface temperature was set to 250°C. Quantification of each of the estrogens including the fluorescently tagged versions was achieved using MSD ChemStation (Agilent, UK) using single ion extracts (estrone = 342 m/z, estradiol = 416 m/z and estriol = 504 m/z). All GC/MS runs were performed in triplicate alongside six standards, negative controls and blanks. Calibration curves were built by plotting the ratio of analyte peak area versus the analyte concentration.

7.3.4 Slow Sand Filter Operation and Sampling

SSFs were operated as previously described in Section 5.2 with the exception that half (four) of the filters were bioaugmented (Section 7.3.5) with the three enriched estrogen metabolisers (E1-rm, E2-rm and E3-rm). All water quality analyses, as described in previous chapters,

were done weekly. Additionally, every week the influent and effluent from all eight filters were analysed for E1, E2 and E3 concentration using the GC/MS method documented in Section 7.3.3 with the only alteration being that 1L samples were filtered.

As in previous chapters, sand samples were retrieved weekly from every sampling port to assess microbial community structure using qPCR. Further, in order to monitor the abundance of the three enriched organisms, custom qPCR primers and assays were designed (Section 7.3.6). To assess the SSFs effectiveness in removing high concentrations of estrogens, filters were spiked with E1 (50ng/L), E2 (12ng/L) and E3 (39ng/L) at weeks 5 and 8, effluent samples from all filters were collected 12h post spiking to ensure complete passage of the water through the filters. Concentrations were chosen based on averages from Fine et al. [2003] and Chen et al. [2010] and simulate agricultural or wastewater pollution events.

7.3.5 Bioaugmentation of SSFs with Estrogen Metabolising Bacteria

Bioaugmentation of the SSFs was achieved following the method outlined by Kim et al. [2004]. Briefly, at the beginning of the study 20ml (1×10^8cfu/ml [OD_{600nm} of 0.6]) of each of the three room temperature estrogen enrichments (E1-rm, E2-rm and E3-rm) grown to exponential phase were added to the top of the filter-bed of four LSSF every day for four days, equating to 5 hydraulic retention times. After which normal filter operation resumed as described in Section 5.2.

7.3.6 qPCR of Estrogen Metabolising Enrichment Cultures

To accurately quantify the abundance of the three enriched estrogen metabolisers specific qPCR primers, targeting the 16S rRNA gene sequence of each organism were designed. Based on the constructed consensus 16S rRNA sequences obtained from the Sanger sequencing (Section 7.3.1.3) primers were designed using CODEHOP [Rose et al., 2003] and tested for their uniqueness using the Silva TestPrime database [Klindworth et al., 2013]. Primer information can be found in Table 7.5. qPCR assays were conducted in triplicate as previously described (Section 4.2.6) alongside two no-template controls, negative controls (*E.coli*) and six different concentrations of linearised 16S rRNA from each of the enriched isolates.

Table 7.5: qPCR primers used for the enriched and isolated estrogen metabolising bacteria

Primer	Target	Sequence (5'-3')	Paired Primers	Anneal. Temp. (°C)	Amplicon Size (bp)
6-8-RT-F	E1 metaboliser	ACAGAGGGATAGCCCAGAGA	6RT-R and 8-RT-R	52	205
6RT-R	E1 metaboliser	TAATAGTGGCTTCATGCGAA	6-8-RT-F	52	205
7-RT-F	E2 metaboliser	TTTAAAGGGTGCGTAGGTGG	7-RT-R	54.3	257
7-RT-R	E2 metaboliser	GTCCTCATCGTTTACGGCAT	7-RT-F	54.3	257
8-RT-R	E3 metaboliser	TAATATAATCCTGATGCCAG	6-8-RT-F	52.3	127

7.3.7 Statistics

Significant differences in the microbial composition as determined by qPCR analysis between bioaugmented and non-augmented filters, age, and depth were identified by using permutational multivariate analysis of variance tests [Anderson, 2001]. Significant differences in water quality production and estrogen removal efficiencies between bioaugmented and non-augmented SSFs were tested using Wilcoxon tests. Estrogen removal efficiencies were calculated relative to the influent water supplying the filters and total estrogenic potency was determined using equation 7.1, where REA represents the relative estrogenic activity (Table 7.2). Coliform retention (number of coliforms not removed) was calculated as a ratio (effluent coliform counts / influent coliform counts), where higher number correspond to greater retention / low removal.

$$\begin{aligned} Estrogenic\,potency &= \frac{E1\,Concentration}{\sum\ E1,\ E2\ and\ E3\ concentration} \times\ REA\ of\ E1\ (0.2) \\ &+ \frac{E2\,Concentration}{\sum\ E1,\ E2\ and\ E3\ concentration} \times\ REA\ of\ E2\ (1.0) \\ &+ \frac{E3\,Concentration}{\sum\ E1,\ E2\ and\ E3\ concentration} \times\ REA\ of\ E3\ (0.01) \end{aligned} \tag{7.1}$$

7.4 Results

In this section the results of the study will be presented alongside discursive analysis, with an overall discussion being provided in Section 7.5.

7.4.1 Characterisation of Enrichment Cultures

Six estrogen-degrading bacteria, designated as, E1-rm, E2-rm, E3-rm, E1-4°C, E2-4°C and E3-4°C, were successfully isolated from the estrogen enrichment cultures. The analysis of the near full-length (1300bp) 16S rRNA genes of the isolates revealed they were phylogenetically distinct (Table 7.6), belonging to either the *Proteobacteria* or *Bacteriodetes* phyla, with a clear temperature effect also apparent. All room temperature strains were highly similar to three previously identified steroid-degrading bacterial species and represent the two most widely known steroid-degrading bacterial genera - *Acidovorax* and *Pseudomonas*.

7.4.2 Whole-Genome Metagenomic Analysis

The final assembly of the E1, E2 and E3 metabolising room temperature enrichment culture genomes (Figure 7.5), consisted of 123, 837 and 607 contigs with total lengths of 5,200,896 bp, 4,809,037 bp and 4,985,812 bp, respectively.

7.4.2.1 Estrone-Metaboliser

The estrone (E1) enrichment isolate (E1-rm) genome contained a single circular chromosome with an average GC content of 61.9% and shared a 93% sequence homology to *Acidovorax sp.* strain KKS102. E1-rm possessed 3961 bp more than the KKS102 strain [Ohtsubo et al., 2012], potentially relating to estrogen metabolism. Additionally, searches within the genome revealed twelve known estrogen metabolism and degradation enzymes alongside five known steroid-removing efflux pumps (Figure 7.5). All the identified enzymes are involved with the metabolic breakdown of estrone to either estradiol, 16-α-hydroxyestrone or estrone-3-sulphate (Figure 7.5).

7.4.2.2 Estradiol-Metaboliser

The estradiol (E2) enrichment isolate (E2-rm) genome contained a single chromosome with an average GC content of 60.57% and shared closest sequence homology (90%) to *Pseudomonas sp.* strain UW4. UW4 is a plant growth-promoting bacteria found in rhizospheres of reeds and is known for its extreme environmental stress resistance. Unlike the UW4 strain, E2-rm was 1,374,351 bp smaller [Duan et al., 2013]. In addition, searches within the genome revealed the possession of 21 known estrogen metabolism and degradation enzymes,

Table 7.6: Phylogenetic classification based on the 16S rRNA gene of estrogen enrichment cultures, (NCBI results only given if match greater than 96%)

Name	Temperature	NCBI Match	RDP Phylum	RDP Class	RDP Order	RDP Family	RDP Genus	Species
E1-4°C	4°C	NR025923.1	Bacteroidetes (100)	Flavobacteria (97)	Flavobacteriales (97)	Flavobacteriaceae (97)	Flavobacterium (97)	
E1-rm	room	CP003872.1	Proteobacteria (100)	Betaproteobacteria (99)	Burkholderiales (99)	Comamonadaceae (99)	Acidovorax (99)	Acidovorax sp. KKS102 (99)
E2-4°C	4°C	JF496261.1	Bacteroidetes (100)	Sphingobacteria (100)	Sphingobacteriales (100)	Cytophagaceae (100)	Dyadobacter (100)	
E2-rm	room	CP003880.1	Proteobacteria (100)	Gammaproteobacteria (100)	Pseudomonadales (99)	Pseudomonadaceae (97)	Pseudomonas (95)	Pseudomonas sp. UW4 (95)
E3-4°C	4°C	EF377722.1	Bacteroidetes (100)	Flavobacteria (99)	Flavobacteriales (99)	Flavobacteriaceae (99)	Flavobacterium (75)	
E3-rm	room	CP000094.2	Proteobacteria (100)	Gammaproteobacteria (100)	Pseudomonadales (99)	Pseudomonadaceae (97)	Pseudomonas (97)	Pseudomonas fluorescenes (97)

Brackets designate the percentage match and yellow highlighted rows represent enrichments used for bioaugmentation

alongside steroid-removing efflux pumps (Figure 7.5). Analysis of the possessed enzymes reveals that estradiol metabolism / degradation appears to occur by conversion to estrone and then subsequent estrone metabolism (Figure 7.5). Additionally, and unlike the other two estrogen metabolising enrichment isolates, the estradiol degrader possesses an ammonium monoxygenase enzyme which has been linked to increased estrogen degradation [Shi et al., 2004].

7.4.2.3 Estriol-Metaboliser

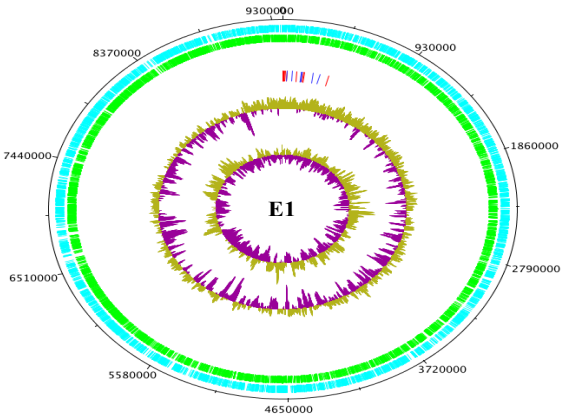
The estriol enrichment isolate (E3-rm) genome contained a single chromosome with an average GC content of 56.05% and shared closest sequence homology (90%) to *Pseudomonas fluorescenes* strain Pf0-1. Pf0-1 like the UW4 strain is a plant growth-promoting bacteria and is commonly found in soil. Unlike the Pf0-1 strain, E3-rm possesses 1,452,593 bp less [Silby et al., 2009], however has seven estrogen metabolism enzymes. Interestingly, these enzymes are associated with the breakdown of estradiol to estrone and not estriol. No known enzymes involved with estriol metabolism or degradation were found within the genome (Figure 7.5).

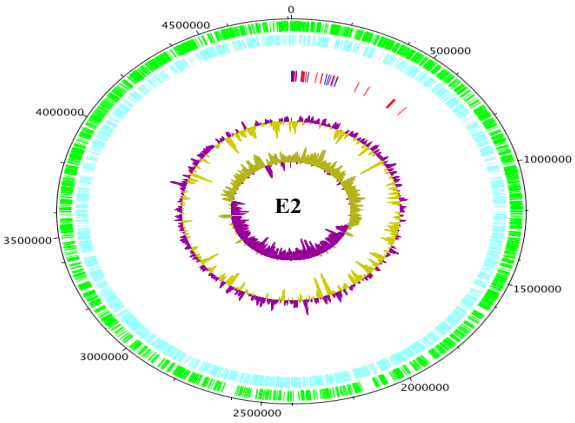
7.4.3 Growth Kinetics of the Estrogen Degrading Isolates

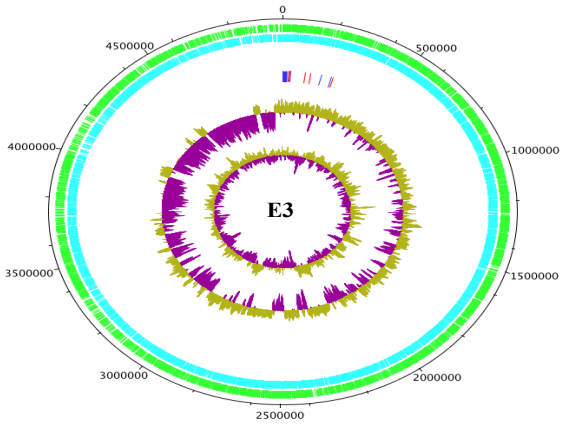
Growth experiments with the three isolated estrogen degraders with two different concentrations of fluorescently-tagged E1, E2 and E3 (Figure 7.6) indicated the impact of both the carbon concentration and the fluorophore on the growth of the isolates. Overall, referring to the non-tagged cultures, the E3-degrader was the quickest growing strain (23.75h) and the E1-degrader the slowest (36.23h). These results are reliable as the growth rates of the two positive control organisms are consistent with the findings of Pauwels et al. [2008] (Figure 7.6).

Surprisingly, unlike the other two strains the E1-degrading isolate grown in the presence of the fluorescently tagged E1 had a doubling time six hours faster than its non-tagged control. This could potentially be due to a 3D conformational change of the tagged estrone making the active sites of degradation enzymes more accessible. However, the E2 and E3 degrading isolates grew slower in the presence of fluorescently tagged estrogen (Figure 7.6B, C and E), implying an inhibitory effect of the fluorophore. Further, comparing the two isolates, the E2-degrading isolate was much more severely affected by the fluorophore (increase of

FEATURE	COLOUR
Forward CDS	
Reverse CDS	
Enzymes associated with Estrogen metabolism / degradation	
Efflux pumps associated with Estrogens	
Above average GC content	
Below average GC content	







ENZYME / FEATURE	EC NUMBER	NUMBER OF CDS REGIONS E1 METABOLISER	E2 METABOLISER	E3 METABOLISER
3Alpha(or 20beta)-hydroxysteroid	1.1.1.53	3	2	0
7Alpha-hydroxysteroid dehydrogenase	1.1.1.159	1	3	0
3Alpha-hydroxysteroid dehydrogenase	1.1.1.209	1	0	0
3-ketosteroid-delta1-dehydrogenase	1.3.99.4	0	5	1
Catechol 2,3-dioxygenase	1.13.11.2	0	4	0
Carboxylesterase	3.1.1.1	5	6	6
Arylsulfatase	3.1.6.1	2	0	0
Steroid sulfotransferase	2.8.2.15	0	1	0
Ammonium monoxygenase	1.14.99.39	0	1	0
AcrAB-TolC Multi drug efflux pump	NA	2	4	2
EmrAB-TolC Multi drug efflux pump	NA	3	4	3

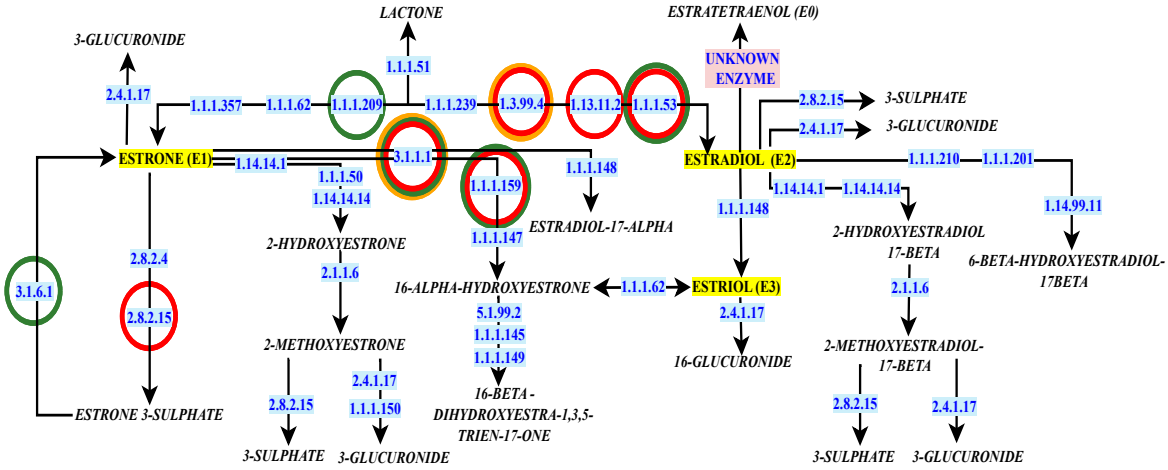
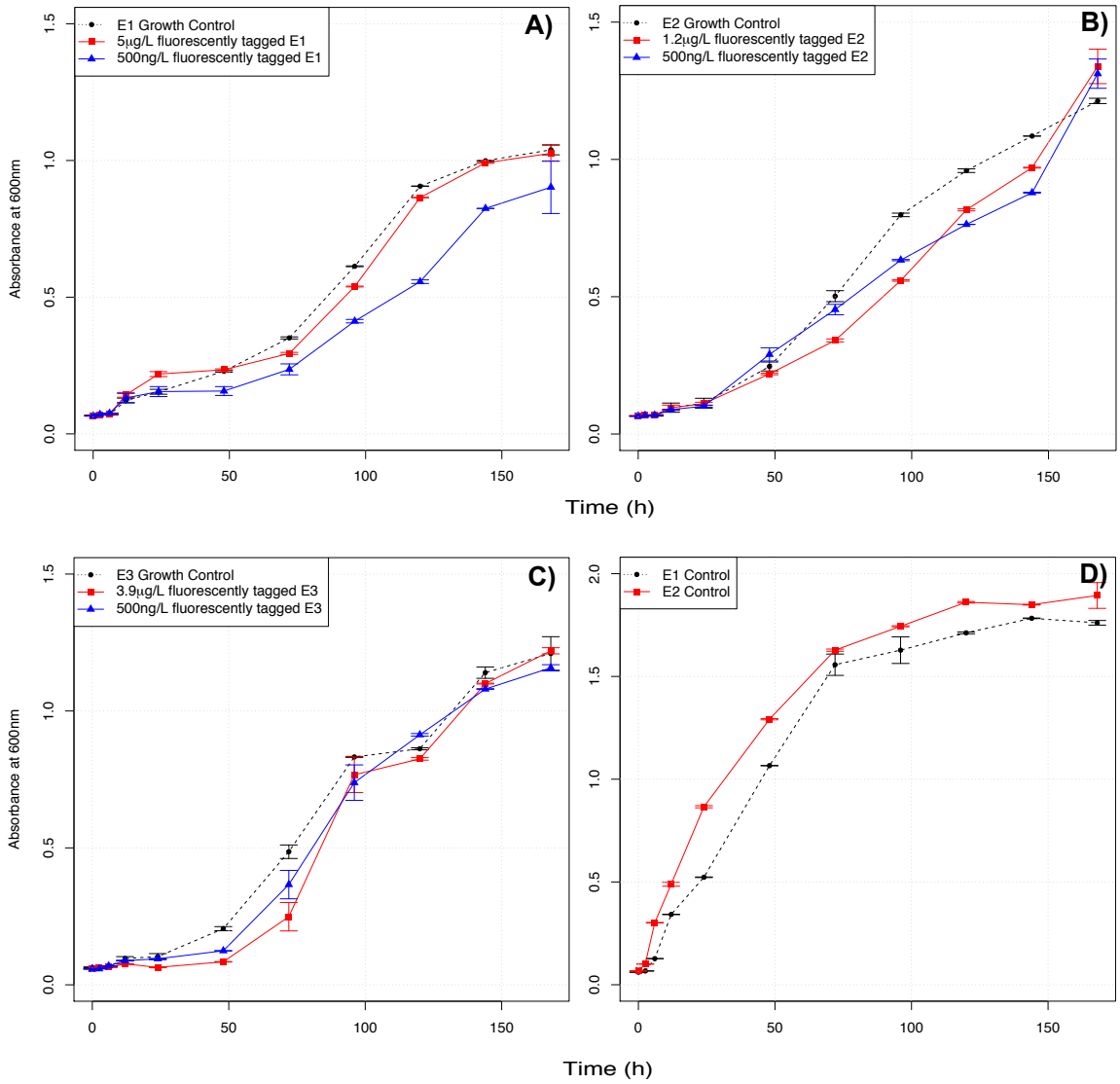


Figure 7.5: Annotated genomes of the three estrogen metabolising isolates, alongside the number and identity of known enzymes or genes involved with estrogen metabolism and their location in the metabolic pathway. Green rings correspond to enzymes possessed by E1-rm, red rings to E2-rm and orange to E3-rm. Genome annotation diagrams were created using DNAPlotter [Carver et al., 2009].



E) Culture Identity	Estrogen Culture Condition	Growth Rate -k (gen/hr)	Generation Time -tgen (h)
Acidovorax sp.	E1 control[1] 5µg/L	0.0276	36.23
Acidovorax sp.	E1 5µg/L	0.0324	30.86
Acidovorax sp.	E1 500ng/L	0.0251	39.84
Pseudomonas sp. UW4	E2 control[1] 1.2µg/L	0.0354	28.25
Pseudomonas sp. UW4	E2 1.2µg/L	0.0262	38.17
Pseudomonas sp. UW4	E2 500ng/L	0.0201	49.75
Pseudomonas fluorescenes	E3 control[1] 3.9µg/L	0.0421	23.75
Pseudomonas fluorescenes	E3 3.9µg/L	0.0387	25.84
Pseudomonas fluorescenes	E3 500ng/L	0.0399	25.06
Previously characterised E1 and E2 metabolisers (Pauwels *et al.,* 2008)			
Pseudomonas sp.	E1	0.0328	30.48
Acinetobacter sp.	E2	0.0379	26.38

[1]non fluorescently tagged estrogen

Figure 7.6: Growth curves of the mean and standard error for the estrone ***A)****, estradiol* ***B)*** *and estriol* ***C)*** *degrading isolates at two concentrations of fluorescently tagged estrogen.* ***D)*** *shows the growth curves of two known estrone- and estradiol-degrading bacteria [Pauwels et al., 2008].* ***E)*** *shows the growth rate and generation times for each isolate.*

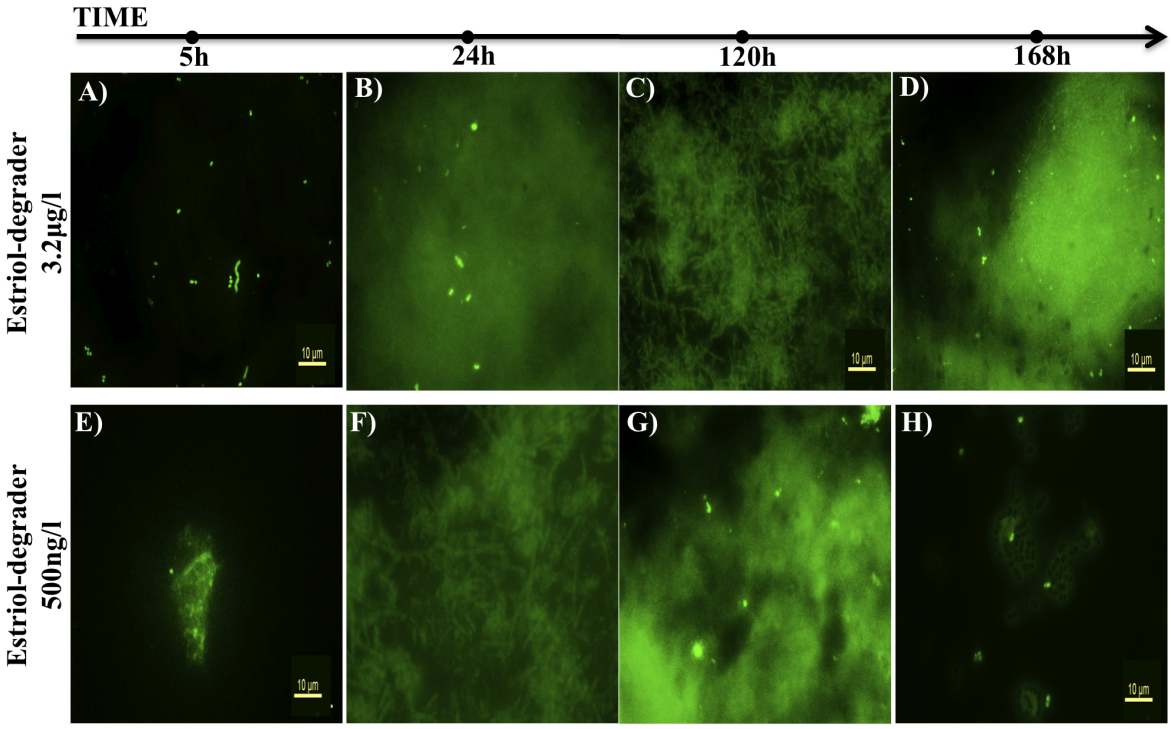


Figure 7.7: Time series fluorescent microscopy images of estriol metabolising isolate. A-D: estriol isolate at 3.2µg/L, E-H: estriol isolate at 500ng/L. Time series intervals are: 5h, 24h, 120h, 168h.

over 35% in its doubling time) than the E3-degrader which showed an increase of 8%. In addition, the hypothesised inhibitory effect of the fluorophore to the E2 and E3-degrading isolates was supported by morphological changes (filamentation) and an absence of fluorescent bacteria at the end of the experiment (Figure 7.7C and D and Figure 7.8K,L and P). This filamentation is possibly due to stress-induced inhibition of SulA and MinCD, which prevent bacterial division at the septum [Bi and Lutkenhaus, 1993]. Furthermore, the mat-like biofilm formation (Figure 7.9) is additional evidence of stress, and the distinct absence of fluorescent bacteria at the end of the study could be due to death or to the removal of the fluorophore by efflux pumps (Figure 7.5).

7.4.4 Estrogen Degradation Capacity of Enriched Isolates

After 48h the average removal efficiencies for the six cultures grown with fluorescently tagged estrogens ranged from: 34-86% for E1, 25-84% for E2 and 97-98% for E3. In all cases, except for the E3 isolate, the highest removal efficiency was achieved with the higher estrogen concentration. The removal efficiencies are in accordance with the non-tagged controls (Figure 7.10F). The degradation profiles indicate considerably slower degradation than described by Pauwels et al. [2008], but are consistent with those described by Yu et al.

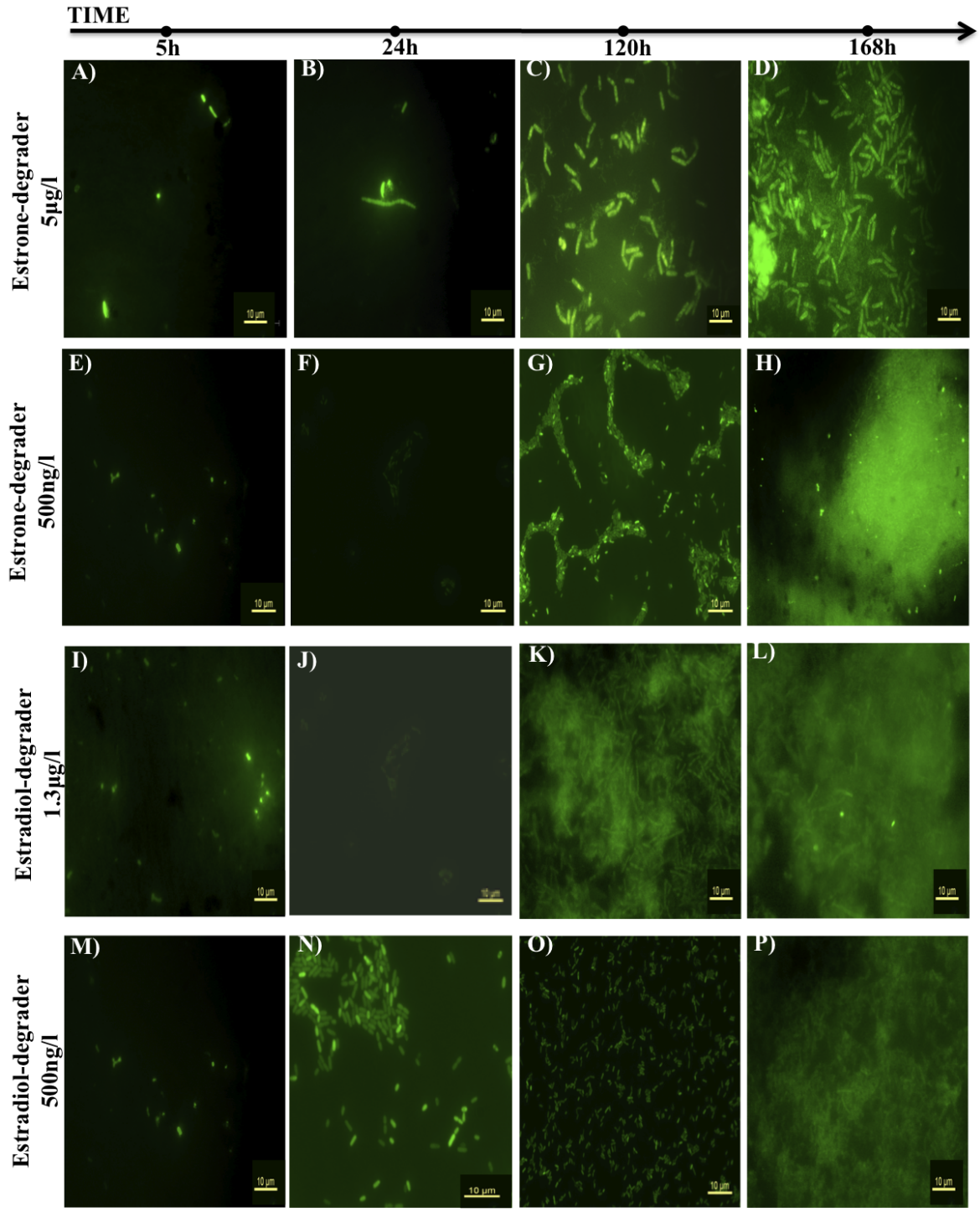


Figure 7.8: Time series fluorescent microscopy images of estrone and estradiol metabolising isolates. A-D: estrone isolate at 5µg/L, E-H: estrone isolate at 500ng/L, I-L: estradiol isolate at 1.3µg/L and M-P: estradiol isolate at 500ng/L. Time series intervals are: 5h, 24h, 120h, 168h.

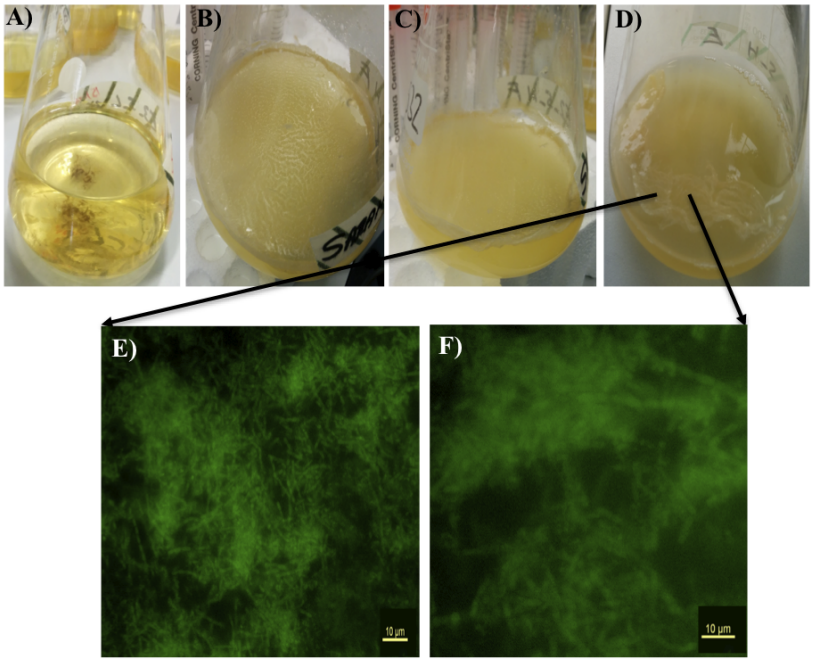


Figure 7.9: The developmental stages of the estradiol-metabolising isolate. A-D: time series of biofilm development, (left to right: 24h, 96h, 120h and 168h), E-F: fluorescence microscopy images of a section of the biofilm mat taken from the end of the study.

[2007], and this is likely due to the differences in bacterial taxa used.

By the end of the study all enriched isolates removed at least 63% of their estrogen source, with the E2 isolate being the poorest degrader. Degradation in the case of the E1 and E2 isolates occurred via some conversion of the estrogen source into the other estrogens (Figure 7.10) and various metabolites (Figure 7.11). In both cases more significant estrogen conversion was seen with lower (500ng/L) estrogen concentrations than with higher (5000ng/L E1 and 1300ng/L E2) concentrations. Unlike the E1 and E2-degrading isolates, the E3-degrader at both the high and low concentration achieved 100% E3 removal after 96 and 72h respectively; however, an increase in E3 was subsequently detected (Figure 7.10E). This was not seen in the E3 non-tagged control (Figure 7.10F) and suggests a toxic effect induced by the fluorophore.

All isolates showed complex degradation routes, with accumulation of metabolites (various estrogen degradation products) being seen at the end of the study (Figure 7.11). Some of these products have been identified as known estrogen degradation products (dehydroepiandrosterone and 17-hydroxypregneonolone) and bacterial cellular components (palmitic

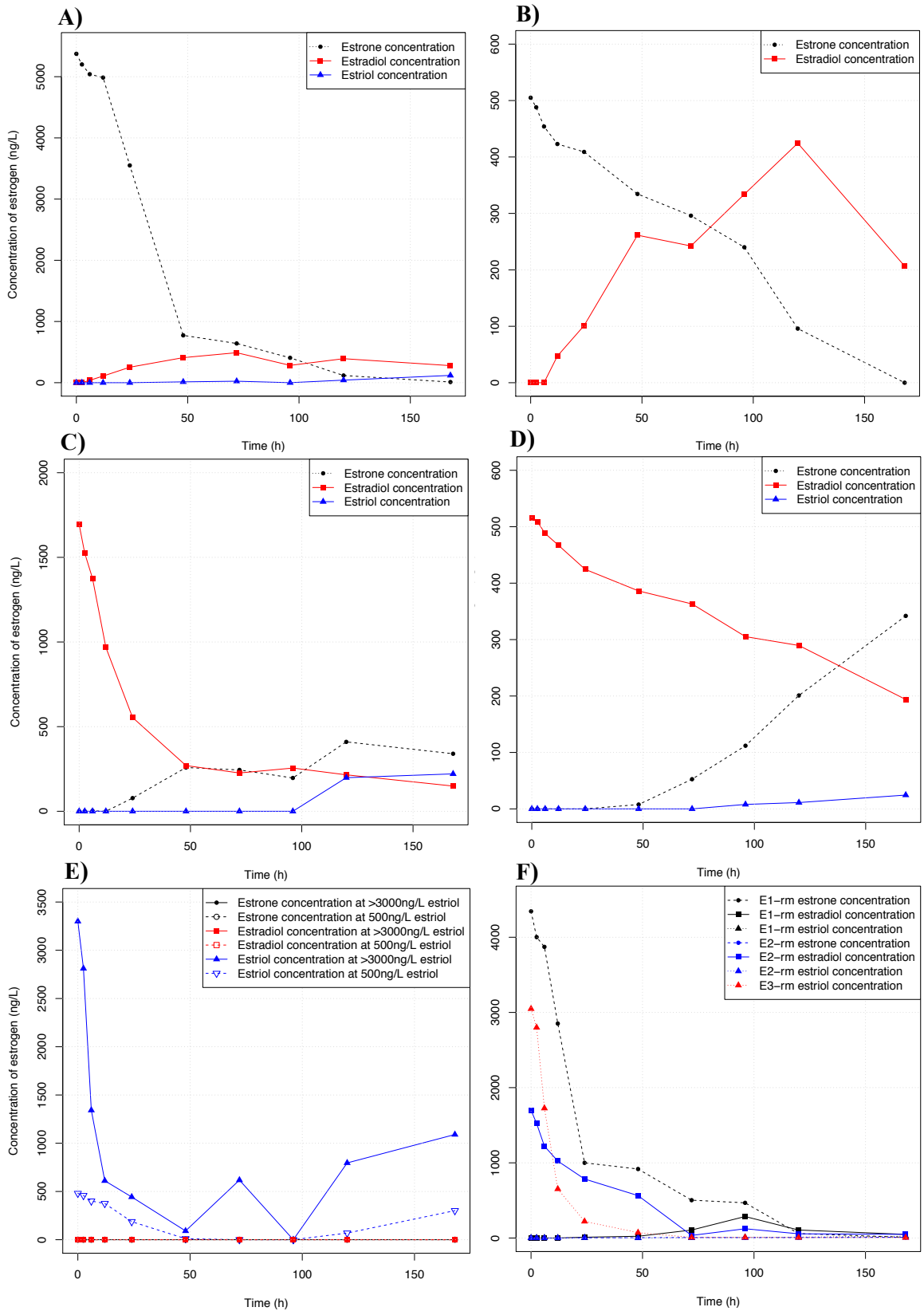


*Figure 7.10: Estrogen degradation graphs: **A)** and **B)** are the estrone enriched isolate at 5000ng/L and 500ng/L respectively, **C)** and **D)** are the estradiol enriched isolate at 1300ng/L and 500ng/L respectively and **E)** is the estriol enriched isolate. **F)** depicts the three enriched isolates with non-fluorescently tagged estrone (5000ng/L), estradiol (1300ng/L) and estriol (3900ng/L)*

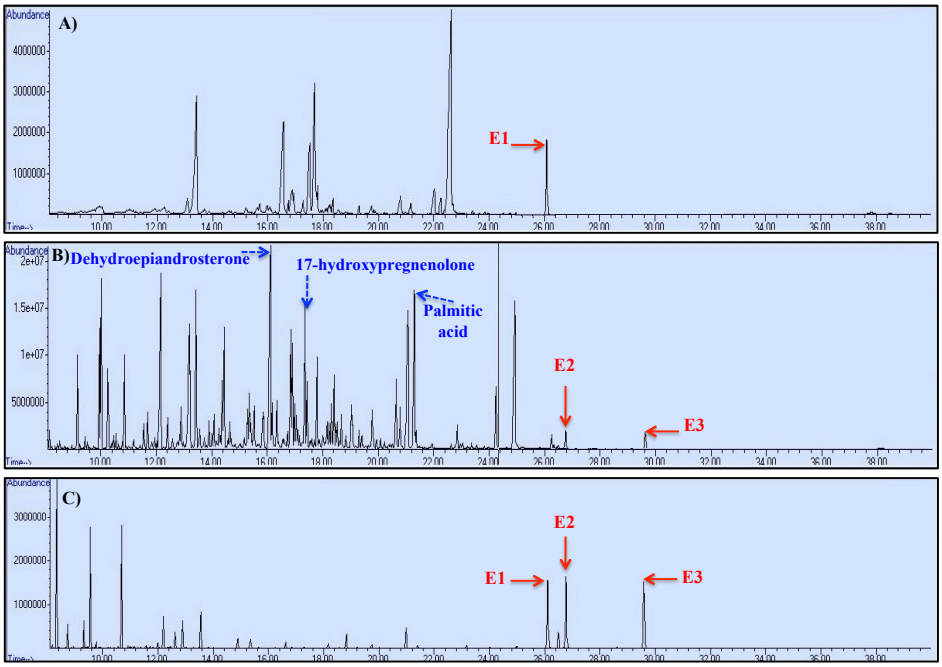


*Figure 7.11: Chromatographs of the E1 metaboliser depicting estrone removal and metabolite production. **A)** beginning of study, **B)** end of study and **C)** standards. Blue text: predicted metabolites.*

acid and various amino acids). The presence of palmitic acid, the major fatty acid found in membrane phospholipids [Madigan et al., 2011] alongside various amino acids (L-tyrosine and L-lysine) may be further evidence that bacterial cell death occurred at the end of the experiment with fluorescently tagged estrogen.

7.4.5 Effects of Bioaugmentation on SSF Functionality

Analysis of the water quality parameters and overall water quality performance ($\bigtriangledown$) showed that, as with the industrial filters (discussed in Chapter 4) and lab-scale filters (discussed in Chapters 5 and 6), the augmented and non-augmented LSSFs studied in this chapter produce good quality water (Table 7.7 and Table 7.8). Significant differences were found between augmented and non-augmented LSSFs $\bigtriangledown$ (p-value: 0.042), which are solely attributed to differences in coliform removal capabilities. Interestingly, when data pertaining to the two spiking events is removed there were only marginal significant differences in filter performance. Additionally, the average total estrogen concentration of the raw water - influent (Table 7.9) supplying the filters was 23.86ng/L, more than four times greater than the previously described highest concentration found in UK rivers [Xiao et al., 2001].

Estrogen degradation capacity results indicate that augmented filters removed significantly

Table 7.7: Summary of the physical and chemical characteristics of the effluent from the bioaugmented LSSFs

FILTER	ADDITIONAL INFO	SAMPLED	AGE	TEMP	PH	NTU	DO	COD	PO4	NO3	NO2	NH4	TVB	TVB_13	COLIFORM	ESTRONE	ESTRADIOL	ESTRIOL	TOTAL ESTROGEN
1		19/03/2013	7	14.36	7.99	0.503	8.552	2.307	1.017	1.000	0.314	0.309	3540	1160	484	1.00	1.36	0.08	2.44
1		26/03/2013	14	14.16	7.89	0.808	8.552	4.647	0.620	0.393	0.019	0.007	9200	3700	130	0.34	6.38	1.84	8.55
1		02/04/2013	21	14.64	8.07	0.661	8.606	5.993	0.600	0.260	0.049	0.049	4260	1020	34	0.63	3.99	0.77	5.38
1		09/04/2013	28	14.40	8.01	0.736	8.716	0.013	0.260	0.420	0.045	0.129	22240	14760	18	3.63	34.51	15.62	53.77
1	Spiked with estrogen	15/04/2013	35	14.70	8.15	0.349	9.152	0.443	0.270	0.500	0.002	0.068	6460	4080	190	5.66	3.91	24.59	34.17
1		23/04/2013	42	14.82	7.91	0.653	9.134	0.028	0.160	0.080	0.000	0.012	5840	540	60	2.52	0.38	26.90	29.80
1		30/04/2013	49	14.82	8.11	0.433	9.152	0.013	0.270	0.010	0.000	0.038	11200	380	8	0.00	1.91	21.92	23.83
1	Spiked with estrogen	07/05/2013	56	14.82	7.91	0.653	9.134	0.428	0.293	0.010	0.009	0.038	4780	3420	6	3.67	33.41	43.31	80.39
1		14/05/2013	63	14.84	7.61	0.347	9.112	1.020	0.140	0.240	0.002	0.009	38280	1060	4	0.00	4.31	0.53	4.84
1		21/05/2013	70	14.66	7.42	0.463	9.030	0.873	0.140	0.260	0.000	0.036	5780	50	0	2.80	1.86	6.52	11.18
5		19/03/2013	7	14.24	8.16	0.515	8.500	2.523	1.147	1.027	0.277	0.309	2260	540	352	0.50	1.11	0.00	1.61
5		26/03/2013	14	14.32	7.95	1.194	8.488	4.440	0.690	0.260	0.027	0.007	8000	1120	180	0.00	8.27	4.17	12.44
5		02/04/2013	21	14.74	8.05	2.722	8.650	8.270	0.510	0.190	0.015	0.046	9620	2880	74	3.64	3.79	3.41	10.85
5		09/04/2013	28	14.70	7.99	0.812	8.658	0.083	0.510	0.493	0.004	0.108	40000	5960	166	3.99	1.90	8.91	14.80
5	Spiked with estrogen	15/04/2013	35	14.80	8.21	0.390	8.714	0.508	0.423	0.500	0.001	0.052	30080	8360	189	7.12	5.40	16.76	29.29
5		23/04/2013	42	14.88	8.05	0.377	8.720	0.033	0.190	0.097	0.000	0.012	11200	720	70	13.15	0.38	33.84	47.37
5		30/04/2013	49	14.98	8.05	0.455	9.488	0.017	0.380	0.030	0.000	0.039	6520	380	6	0.80	6.37	1.32	8.48
5	Spiked with estrogen	07/05/2013	56	14.90	8.05	0.455	9.510	0.526	0.200	0.030	0.009	0.038	2840	20	4	2.12	41.76	52.88	96.76
5		14/05/2013	63	14.94	7.72	0.222	9.142	1.003	0.220	0.247	0.014	0.009	3960	84	0	0.00	3.17	18.06	21.23
5		21/05/2013	70	14.68	7.86	0.443	9.060	0.750	0.087	0.263	0.004	0.039	16300	20	0	2.56	1.24	5.98	9.78
6		19/03/2013	7	14.44	7.99	0.643	8.400	2.990	1.427	0.927	0.319	0.308	4500	1800	296	0.50	1.11	0.00	1.61
6		26/03/2013	14	14.56	7.96	0.635	8.380	5.757	0.770	0.270	0.026	0.007	5040	1120	126	9.30	8.94	2.25	20.49
6		02/04/2013	21	14.68	8.01	0.435	8.680	8.307	0.907	0.453	0.010	0.045	5760	3320	64	3.64	3.78	13.65	21.07
6		09/04/2013	28	14.68	7.98	0.377	8.694	0.091	0.313	0.507	0.003	0.091	18840	7660	94	3.99	1.53	17.28	22.80
6	Spiked with estrogen	15/04/2013	35	14.70	8.31	0.341	9.472	1.320	0.310	0.570	0.005	0.056	21540	16140	70	7.12	5.01	24.38	36.52
6		23/04/2013	42	14.88	8.16	1.200	9.546	0.040	1.170	0.113	0.000	0.013	29820	9520	50	13.15	0.38	25.71	39.24
6		30/04/2013	49	14.78	8.12	1.224	9.620	0.015	0.370	0.010	0.000	0.037	17600	4820	38	0.00	7.57	15.61	23.18
6	Spiked with estrogen	07/05/2013	56	14.98	8.04	0.540	9.640	1.130	0.240	0.013	0.014	0.037	6120	60	10	2.13	41.76	47.67	91.56
6		14/05/2013	63	15.06	7.82	1.854	9.142	1.010	0.240	0.240	0.007	0.009	12820	1920	10	0.00	2.73	18.06	20.79
6		21/05/2013	70	14.82	7.89	0.520	9.064	0.743	0.090	0.260	0.015	0.038	11100	420	0	2.56	3.72	17.66	23.94
8		19/03/2013	7	14.46	8.14	3.524	8.434	2.303	1.287	1.000	0.323	0.309	3840	1140	460	3.76	1.12	2.07	6.95
8		26/03/2013	14	14.52	7.87	1.698	8.426	4.473	0.690	0.420	0.023	0.008	5900	1660	122	1.53	8.03	2.25	11.82
8		02/04/2013	21	14.56	7.89	0.705	8.658	5.990	0.620	0.377	0.071	0.048	2840	880	40	4.75	3.88	0.64	9.26
8		09/04/2013	28	14.86	8.10	4.512	8.646	0.013	0.537	0.587	0.005	0.099	46640	5320	158	6.03	1.57	17.73	25.33
8	Spiked with estrogen	15/04/2013	35	14.64	8.19	0.341	9.110	0.592	0.310	0.590	0.000	0.060	4560	580	4	6.11	4.46	0.95	11.52
8		23/04/2013	42	14.96	7.91	0.530	9.112	0.030	0.170	0.167	0.001	0.010	24880	10220	28	8.42	0.38	22.66	31.46
8		30/04/2013	49	15.04	7.90	0.710	9.132	0.015	0.187	0.033	0.000	0.038	8100	56	14	1.18	14.51	15.93	31.61
8	Spiked with estrogen	07/05/2013	56	15.06	7.99	0.767	9.124	0.591	0.530	0.030	0.011	0.038	18040	140	0	2.08	43.85	49.55	95.48
8		14/05/2013	63	14.98	7.55	0.810	9.106	1.370	2.700	0.247	0.003	0.009	10560	100	28	0.00	19.26	18.85	38.11
8		21/05/2013	70	14.78	7.74	0.258	9.070	0.780	0.170	0.253	0.000	0.037	11940	52	0	3.17	2.25	5.49	10.91

Spiking events occurred 14h prior to sampling to ensure filtration through the SSFs

Table 7.8: Summary of the physical and chemical characteristics of the effluent from the non-augmented LSSFs

FILTER	ADDITIONAL INFO	SAMPLED	AGE	TEMP	PH	NTU	DO	COD	PO4	NO3	NO2	NH4	TVB	TVB_13	COLIFORM	ESTRONE	ESTRADIOL	ESTRIOL	TOTAL ESTROGEN
2		19/03/2013	7	14.24	7.99	0.430	8.628	1.383	0.103	0.947	0.334	0.290	2960	1120	414	2.10	1.63	1.79	5.52
2		26/03/2013	14	14.06	7.85	0.613	8.616	2.650	0.240	0.370	0.030	0.042	4680	1680	116	1.19	56.55	1.67	59.41
2		02/04/2013	21	14.60	8.06	0.720	8.640	8.640	0.040	0.160	0.006	0.057	2560	1320	388	0.86	1.83	22.46	25.15
2		09/04/2013	28	14.68	8.06	0.797	8.684	0.019	0.140	0.270	0.007	0.090	50000	21000	66	0.71	14.01	14.78	29.50
2	Spiked with estrogen	15/04/2013	35	14.78	8.26	0.470	0.000	1.420	0.250	0.290	0.001	0.108	12640	4940	350	50.16	41.06	38.86	130.08
2		23/04/2013	42	14.82	8.02	1.020	0.000	0.213	0.033	0.030	0.000	0.035	6160	680	46	0.82	0.78	33.02	34.61
2		30/04/2013	49	14.90	8.09	0.577	0.000	0.020	0.070	0.053	0.000	0.045	2580	88	22	0.99	13.85	16.29	31.13
2	Spiked with estrogen	07/05/2013	56	14.74	8.04	0.780	9.338	0.621	0.150	0.060	0.007	0.046	2800	36	14	29.19	26.41	0.00	55.60
2		14/05/2013	63	15.02	8.03	0.661	9.134	1.403	0.057	0.280	0.002	0.015	16280	420	4	0.43	2.30	0.11	2.84
2		21/05/2013	70	14.84	7.59	0.297	9.042	1.387	0.573	0.070	0.002	0.040	3920	68	3	12.33	4.25	6.08	22.66
3		19/03/2013	7	14.42	8.01	0.539	8.688	1.433	0.103	0.947	0.339	0.287	4720	2100	434.8	2.26	1.59	1.76	5.61
3		26/03/2013	14	14.20	7.81	0.533	8.680	2.707	0.233	0.357	0.031	0.044	8000	1760	118	1.28	55.44	1.63	58.35
3		02/04/2013	21	14.42	8.16	0.669	8.660	8.680	0.030	0.160	0.000	0.054	2920	480	380	1.18	1.80	22.02	24.99
3		09/04/2013	28	14.68	8.07	0.817	8.676	0.026	0.123	0.210	0.006	0.089	59800	27760	56	0.76	13.74	14.49	28.99
3	Spiked with estrogen	15/04/2013	35	14.70	8.25	0.474	9.170	1.440	0.240	0.290	0.001	0.110	12720	4160	426	45.60	40.25	38.10	123.95
3		23/04/2013	42	14.92	8.08	1.014	9.140	0.264	0.010	0.030	0.000	0.037	4240	198	24	0.88	0.25	32.37	33.50
3		30/04/2013	49	14.98	8.09	0.573	9.170	0.021	0.050	0.053	0.000	0.044	2220	80	6	1.13	13.58	15.97	30.68
3	Spiked with estrogen	07/05/2013	56	14.78	8.03	0.769	9.336	0.578	0.143	0.050	0.006	0.044	2480	24	12	25.84	25.89	0.00	51.73
3		14/05/2013	63	14.98	8.04	0.657	9.136	1.403	0.053	0.199	0.002	0.014	13360	360	0	0.46	2.26	0.10	2.83
3		21/05/2013	70	14.78	7.56	0.293	9.048	0.987	0.600	0.073	0.002	0.041	4100	56	2	13.26	4.17	5.96	23.38
4		19/03/2013	7	14.36	8.01	0.331	8.546	1.877	0.070	1.060	0.339	0.285	5320	3620	352	3.60	7.72	1.48	12.80
4		26/03/2013	14	14.36	7.95	1.254	8.424	2.867	0.157	0.337	0.022	0.047	2040	174	308	1.42	9.81	1.90	13.13
4		02/04/2013	21	14.76	7.96	0.828	8.670	8.680	0.020	0.260	0.008	0.055	4380	2640	142	2.75	33.66	7.59	44.00
4		09/04/2013	28	14.72	7.97	0.697	8.664	0.019	0.040	0.287	0.003	0.091	29000	11760	18	0.32	14.37	16.29	30.98
4	Spiked with estrogen	15/04/2013	35	14.70	8.04	0.873	9.116	0.502	0.783	0.370	0.002	0.111	20080	11540	226	18.66	33.28	77.72	129.67
4		23/04/2013	42	15.04	7.86	0.902	9.140	0.102	0.040	0.030	0.000	0.032	37400	23200	40	2.66	3.36	29.28	35.30
4		30/04/2013	49	14.82	8.11	0.433	9.152	0.013	0.270	0.010	0.000	0.038	11200	380	8	0.26	3.03	16.47	19.76
4	Spiked with estrogen	07/05/2013	56	14.82	7.91	0.653	9.134	0.428	0.293	0.010	0.009	0.038	10960	140	4	29.55	26.71	45.12	101.37
4		14/05/2013	63	15.16	8.06	0.246	9.088	1.393	0.010	0.290	0.004	0.014	7040	440	16	0.33	0.23	0.12	0.68
4		21/05/2013	70	14.62	7.82	0.243	9.042	0.757	0.050	0.293	0.000	0.042	7560	420	0	13.17	0.80	5.11	19.08
7		19/03/2013	7	14.44	7.84	0.477	8.216	1.720	0.137	0.877	0.315	0.287	2000	900	360	1.80	5.07	1.92	8.78
7		26/03/2013	14	14.48	7.84	0.447	8.184	2.233	0.160	0.290	0.034	0.043	1960	184	130	0.25	13.84	2.40	16.49
7		02/04/2013	21	14.74	7.95	0.366	8.624	6.990	0.447	0.467	0.051	0.055	6120	1640	802	0.78	34.29	13.65	48.72
7		09/04/2013	28	14.76	7.91	0.507	8.620	0.022	0.267	0.287	0.001	0.095	12840	1800	22	16.89	11.09	20.79	48.77
7	Spiked with estrogen	15/04/2013	35	14.62	8.26	0.398	9.060	1.047	0.037	0.410	0.000	0.103	16780	14220	250	37.62	24.23	24.89	86.73
7		23/04/2013	42	14.84	8.19	0.420	9.208	0.102	0.047	0.030	0.000	0.033	13000	820	18	2.26	3.37	33.25	38.87
7		30/04/2013	49	15.04	8.06	0.571	9.220	0.023	0.060	0.053	0.001	0.047	8800	200	28	3.19	14.00	15.93	33.12
7	Spiked with estrogen	07/05/2013	56	15.02	8.05	1.146	9.234	1.020	0.270	0.050	0.012	0.046	9780	8	6	15.02	144.70	53.36	213.08
7		14/05/2013	63	15.14	7.90	0.836	9.130	1.387	0.813	0.283	0.003	0.014	5640	52	28	0.47	5.73	233.64	239.85
7		21/05/2013	70	14.80	7.89	0.370	9.048	0.880	0.127	0.303	0.000	0.043	5460	36	0	3.78	7.94	4.19	15.92

Spiking events occurred 14h prior to sampling to ensure filtration through the SSFs

Table 7.9: Summary of the physical and chemical characteristics of the influent water (River Kelvin) supplying bioaugmented and non-augmented LSSFs

SAMPLED	ADDITIONAL INFO	AGE	TEMP	PH	NTU	DO	COD	PO4	NO3	NO2	NH4	TVB	TVB 13	COLIFORM	ESTRONE	ESTRADIOL	ESTRIOL	TOTAL ESTROGEN
19/03/2013		7	12.98	7.68	11.184	8.122	1.170	0.077	0.650	0.209	0.300	8300	1040	406	3.50	9.37	2.00	14.87
26/03/2013		14	12.96	8.03	13.612	8.164	2.217	0.290	0.343	0.017	0.002	8780	5440	90	1.07	5.54	1.25	7.87
02/04/2013		21	13.60	8.15	19.494	9.326	12.207	0.330	0.463	0.026	0.058	10240	7300	742	1.41	35.92	11.01	48.33
09/04/2013		28	13.92	8.19	23.400	9.472	5.003	0.320	0.633	0.017	0.089	24840	10080	196	6.42	10.46	9.00	25.88
15/04/2013	Spiked with estrogen	35	14.38	8.37	14.840	9.435	3.520	0.280	0.500	0.012	0.078	22440	13900	192	58.33	28.57	50.80	137.69
23/04/2013		42	14.58	7.93	43.240	9.378	0.107	1.427	0.113	0.010	0.018	17160	12560	528	4.09	3.61	14.71	22.41
30/04/2013		49	14.56	7.98	20.900	9.312	0.139	0.450	0.113	0.021	0.078	16880	5840	334	0.71	5.22	14.61	20.55
07/05/2013	Spiked with estrogen	56	14.54	7.88	29.080	9.360	2.257	0.660	0.120	0.010	0.078	37100	5780	180	59.10	20.88	47.64	127.62
14/05/2013		63	14.56	7.60	22.120	9.288	2.810	0.270	0.360	0.018	0.018	5860	4060	210	0.79	2.54	0.79	4.11
21/05/2013		68	14.64	7.83	30.720	9.230	2.003	0.320	0.413	0.009	0.078	14720	6480	78	14.97	5.63	10.64	31.24

Spiking events occurred 14h prior to sampling to ensure filtration through the SSFs. Age: number of days since filters were scraped. Units of measurement are: TEMP: °C, PH:pH, NTU:turbidity, DO (Dissolved Oxygen): mg/L, COD (Chemical Oxygen Demand): mg/L, PO4 (Phosphate): mg/L, NO3 (Nitrate): mg/L, NO2 (Nitrite): mg/L, NH4 (Ammonium): mg/L, Total Viable Bacteria grown at 30°C (TVB) and 13°C (TVB13), coliforms: cfu/ml, and estrogens: ng/L

Table 7.10: Average percentage removal of estrogens in augmented and non-augmented filters.

Estrogen	Average percentage removal by filter type		p-value
	Augmented	Non-augmented	
Estrone	30.1 ± 11.02	-130 ± 23.0	* 0.00784
Estradiol	5.2 ± 0.20	-90 ± 37.0	* 0.02285
Estriol	-69 ± 19.2	-282 ± 74.4	0.104

∗Significant differences tested using Wilcoxon tests. Positive percentages correspond to removal. Negative percentage correspond to an increase in the concentration of estrogens i.e. no removal.

more estrone and estradiol than non-augmented filters (Table 7.10). However, there was no significant difference in estriol removal. These removal capabilities are significantly lower than seen in the pure culture growth kinetic experiments (Section 7.4.4) and may likely be due to the inability to form mat-like biofilms (Figure 7.9) within the SSFs due to shearing between sand grains. Although estrogen removal was lower than predicted, bioaugmentation did significantly reduce the overall estrogenic potency (p-value: 6.889×10^{-5}) of the purified water by 26% on average, compared to no reduction in non-augmented SSFs (Table 7.11).

Irrespective of the reduced estrogen removal in SSFs, estrogen concentration was found to negatively correlate with coliform removal in non-augmented filters (correlation = -0.34, p-value = 0.0379) (Figure 7.12), which appears to be due to inhibition of coliform-grazing protozoa (Figure 7.13). Referring more closely to Figure 7.13, a reduction in the number of eukaryotes (i.e., protozoa) present during the two spiking periods of estrogen can be seen in both augmented and non-augmented filters. However, this reduction was significantly lower in augmented filters (55% and 53% respectively) compared to non-augmented filters (93% and 85% respectively). This alongside the interesting finding that coliform removal in augmented filters was less affected by estrogen concentration than non-augmented filters, implies that augmentation reduces the toxic effect of estrogen on the coliform-grazing, protozoan community.

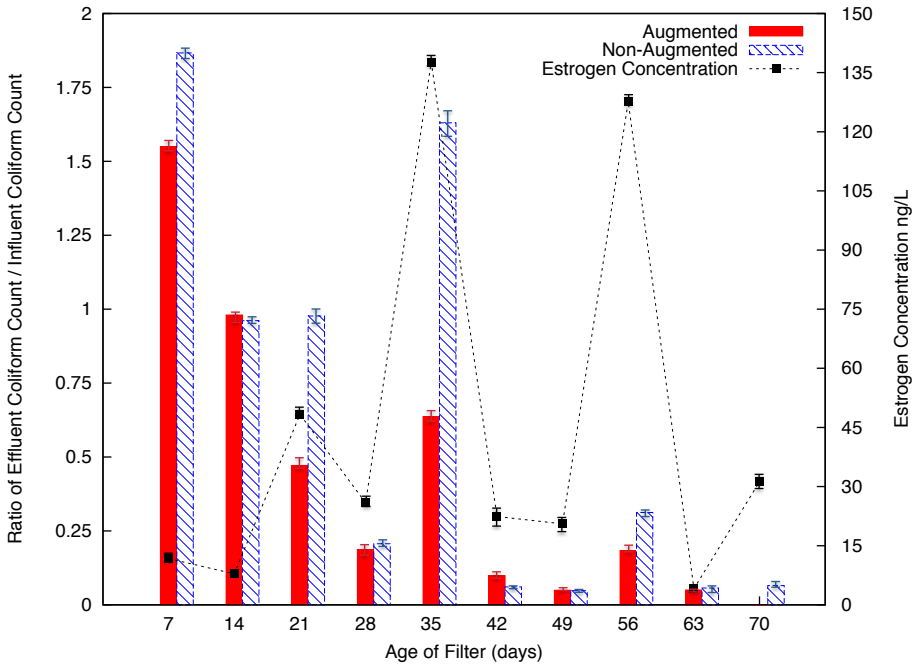


Figure 7.12: Coliform retention in augmented and non-augmented LSSFs, alongside the corresponding combined estrogen (estrone, estradiol and estriol) concentration of the influent supplying all filters. Ratio below 1 correspond to coliform removal and ratios above 1 correspond to an increase in the number of coliforms in respect to the concentration found in the influent water.

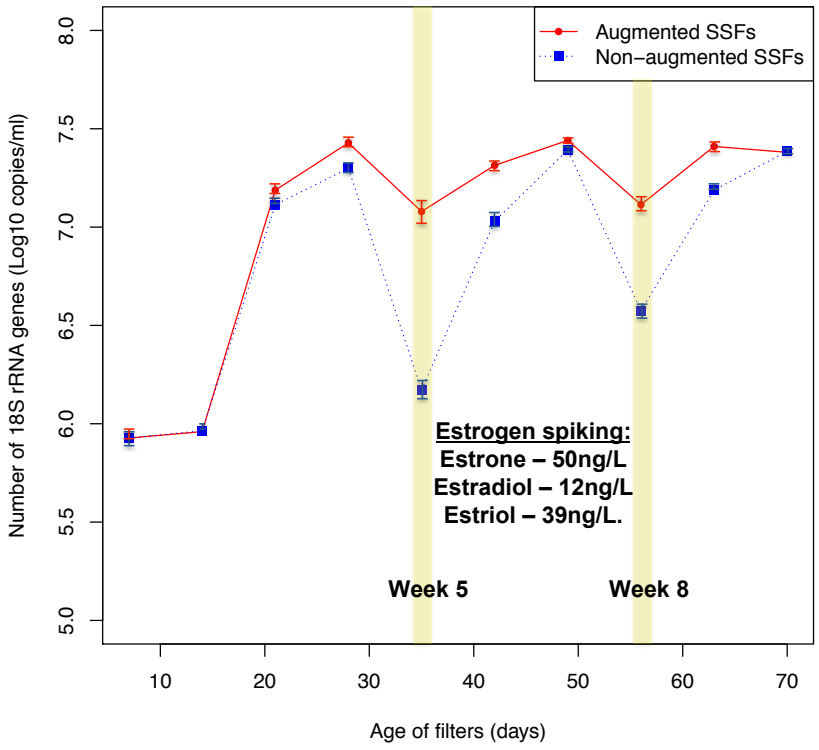


Figure 7.13: Concentration of 18S rRNA genes in augmented and non-augmented slow sand filters. Yellow highlighted columns correspond to estrogen spiking time points.

7.4.6 Effect of Bioaugmentation of Filter Community

Initial exploratory NMDS and MANOVA analysis revealed that the microbial communities present in augmented and non-augmented LSSFs were significantly different (p-value: 0.003), with *Gammaproteobateria*, and unclassified bacteria dominating augmented LSSFs, and *Bacteriodetes* and unclassified bacteria dominating non-augmented LSSFs (Figure 7.14). The dominance of *Gammaproteobacteria* is unsurprising as two of the bacterial isolates used to augment the filters belong to this taxonomy. Additionally, significant differences in the number of 16S rRNA copies were found between augmented and non-augmented filters (p-value: 0.0087), with augmented filters possessing on average three times more copies than non-augmented LSSFs.

In order to determine which factors explain the differences between augmented and non-augmented LSSFs microbial communities, MANOVA and canonical correspondence analysis analysis was performed. This analysis revealed (Table 7.12) that the type of filter (augmented or non-augmented) explained the biggest proportion (38%) of the difference in bacterial community composition, with age and the occurrence of estrogen spiking also being highly significant. Interestingly, individual filter identity and depth were not significant variables in explaining the differences between bacterial community composition as seen in Chapters 4 and 5.

Table 7.11: Total estrogenic potency within the influent and effluent of bioaugmented and non-augmented SSFs

		Total Estrogenic Potency			
DATE	**Influent**	**Augmented**	**Retention**	**Non-augmented**	**Retention**
19/03/2013	0.68	0.47	0.69	0.72	1.06
26/03/2013	0.73	0.64	0.87	0.78	1.06
02/04/2013	0.75	0.39	0.52	0.46	0.61
09/04/2013	0.46	0.37	0.82	0.33	0.72
15/04/2013	0.30	0.22	0.75	0.34	1.15
23/04/2013	0.20	0.07	0.33	0.07	0.33
30/04/2013	0.27	0.36	1.34	0.38	1.42
07/05/2013	0.26	0.35	1.36	0.54	2.08
14/05/2013	0.66	0.35	0.54	0.69	1.04
21/05/2013	0.28	0.19	0.67	0.29	1.05
AVERAGE	**0.46**	**0.34**	**0.74**	**0.46**	**1.00**

Potency calculated using the formula in equation 7.1.

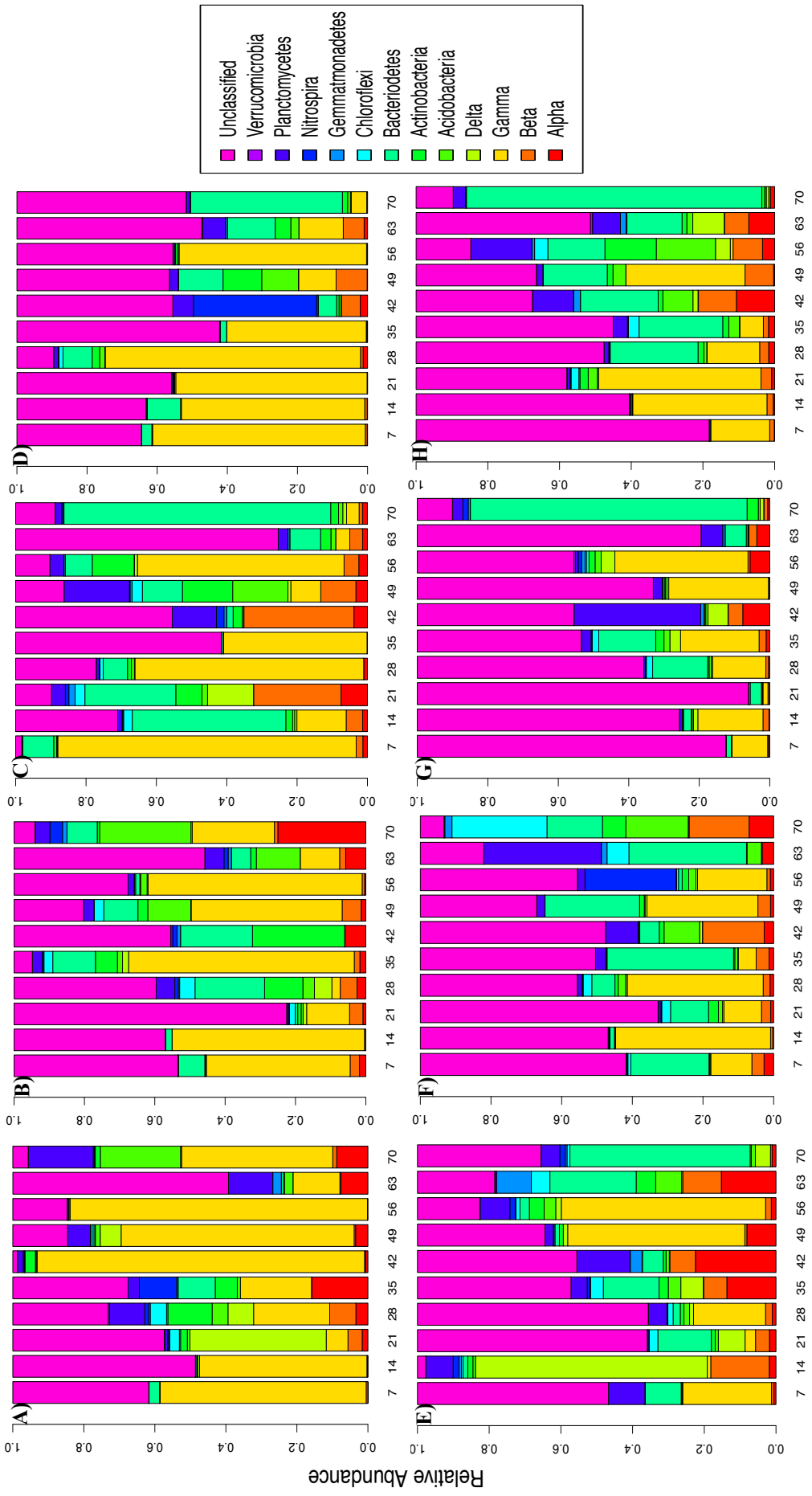


*Figure 7.14: Stacked barplots of the relative abundance of each phylum at different depths in: **A-D** (augmented filters) and **E-H** (non-augmented filters), from left to right depths are: 0cm, 5-15cm, 20-45cm and 75cm as assessed by qPCR assays.*

To determine whether bioaugmentation with the three enrichment estrogen degraders was successful and sustainable, specifically-designed qPCR primers (Table 7.5) were used to enumerate their presence. In bioaugmented SSFs the initial number of the E1, E2 and E3-degraders were 38, 25 and 16 times greater than that found in the non-augmented SSFs. It is however, important to note that non-augmented filters did possess the three enriched estrogen degraders, which originated from the river; confirmed by qPCR assays with DNA extracted from the influent river water. Although bioaugmentation was shown to be successful and sustainable it is important to note that only between 0.01% and 1.3% of the initial inocula (1 $\times$ 10^8cfu/ml) for each isolate remained at the top depth of the SSFs by the end of the study in augmented SSFs. This reduction is likely due to natural competition between other bacteria. Throughout the study the three enriched isolates were present in greater abundance in the bioaugmented filters (Figure 7.15), with greatest abundance in both types of filter being seen during and after the two estrogen spiking periods.

7.5 Discussion

7.5.1 Estrogen-Degrading Enriched Bacterial Strains

An estrone (E1)-degrading bacterial strain (E1-rm) capable of metabolising and removing E1 and E2 was enriched and isolated from a full-scale SSF. Likewise estradiol (E2) and estriol (E3)-degrading bacterial strains (E2-rm and E3-rm) capable of degrading only their enriched estrogen source were isolated. This is the first study to isolate estrogen degrading bacterial strains from a SSF. Consistent with the literature and this study, the ability to degrade estradiol was more widespread than the ability to degrade estrone and estriol [Yu et al., 2007,

Table 7.12: Canonical-correspondence analysis analysis of bacterial phyla and class abundances against various parameters from augmented and non-augmented filter.

Parameter	Degrees of Freedom	χ^2	F value	Number of permutations	Pr(>F)
Age	1	0.1093	18.0847	99	0.01*
Type	1	0.5730	94.8510	99	0.005*
Spiked	1	0.0229	3.7903	99	0.02*
Residual	216	0.7894			

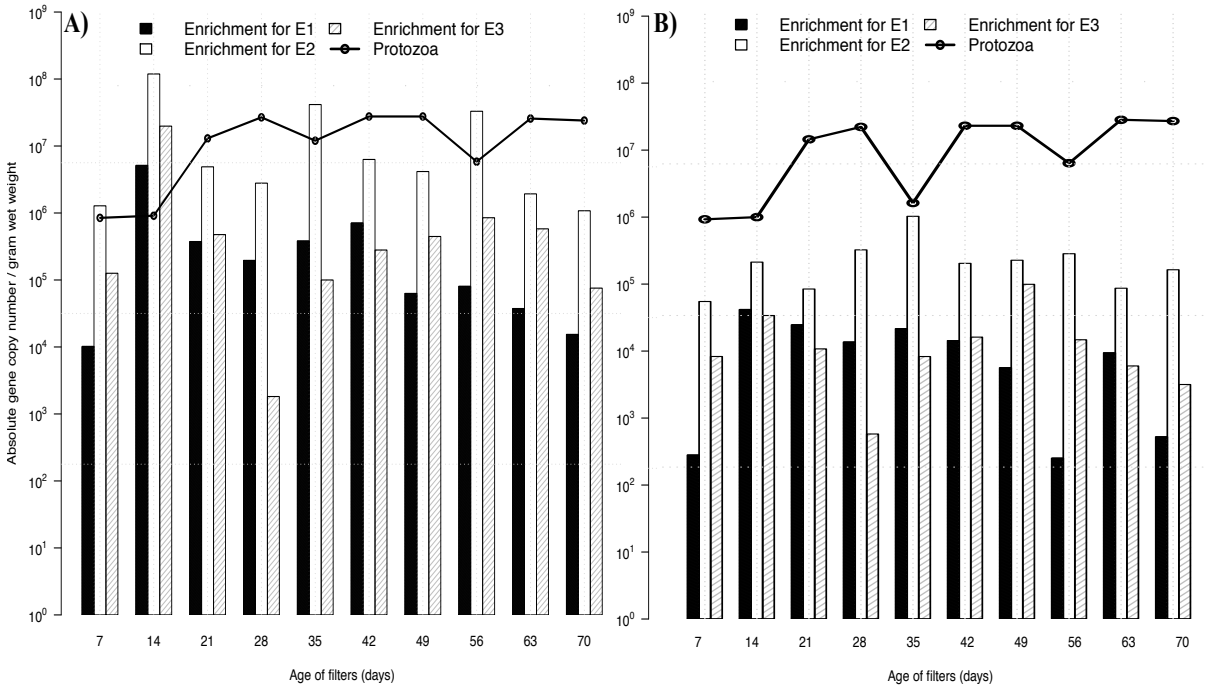


Figure 7.15: Abundance of estrogen-degrading enriched isolates and protozoa at the top depth of: ***A)*** *augmented and* ***B)*** *non-augmented SSFs*

Pauwels et al., 2008, Iasur-Kruh et al., 2011]. Such conversion of estradiol to either estrone, estriol or various intimidates is extremely beneficial as E1 is five times less potent than estradiol [Conroy et al., 2007]. Therefore from an environmental and water quality perspective this reduces the overall estrogenicity [Iasur-Kruh et al., 2011].

Results from the pure culture growth kinetics (Figure 7.10) and examination of the isolates' genomes suggest that the transformation of E1 and E2 can be explained by the mechanisms proposed by Yu et al. [2007]. For example E2 was likely degraded to a ketone following the oxidation of the secondary alcohol on C17. The degradation of E1, which was the major metabolite of E2 biodegradation, suggests that the conversion of E2 to secondary metabolites might be the rate-limiting step. Furthermore, the increased E1 production during metabolism of 500ng/L of E2 suggests a potential enzymatic negative feedback system which inhibits E2 conversion at high concentrations of E1 (Figure 7.10C and D). These mechanisms are supported by many field studies involving wastewater treatment [Baronti et al., 2000], in which efficient E2 removal was achieved whereas E1 removal was not. In addition, enzyme searches within the genomes showed that strain E2-rm possessed an ammonia monooxygenase gene, which has been linked to cometabolic reactions of a wide range of organics including estrogens [Shi et al., 2004, Yu et al., 2013]. Therefore it is likely that both metabolic

and co-metabolic degradation of E2 and E1 are achieved by E2-rm, however, more experiments are needed to test this hypothesis.

Unlike the E1 and E2 degraders, the route of E3 transformation is unknown as secondary metabolites could not be identified and searches within the genome revealed no known estriol degradative enzymes. However, it should be noted that although no significant difference between degradation rates using fluorescently tagged and non-tagged estrogen were observed, the degradation rate of secondary metabolites (i.e. E1 from E2 metabolism) were different (Figure 7.10). Metabolic stress induced by the presence of the fluorophore is the likely reason for the decreased removal and increased time required for doubling (Figure 7.6), although metatranscriptomic research would be required to verify this.

7.5.2 Impact of Bioaugmentation on SSF Performance and Community

In addition to the single-species growth kinetics, E1-rm, E2-rm and E3-rm were incorporated into sterile SSFs (bioaugmentation). The success of bioaugmentation is determined by two major principles: the ability of the integrated bacterium to survive in the augmented environment and its ability to degrade the target pollutant [El Fantroussi and Agathos, 2005]. The three bacterial strains were able to satisfy both criteria, removing more estrone (E1) and estradiol (E2) than non-augmented SSFs (Table 7.10). Recently, several studies have shown enhanced estradiol degradation by bioaugmentation into activated sludge systems [Yu et al., 2007] and constructed wetlands [Iasur-Kruh et al., 2011], but this is the first study to the author's knowledge to show successful augmentation into a drinking water system.

With respect to estrogen removal efficiency, the isolates achieved far greater degradation in pure culture than once augmented into the laboratory-scale SSFs. This is unsurprising as pure culture experiments with the isolates led to mat-like biofilm formation (Figure 7.9), which would be unlikely to occur within the SSFs due to shearing pressure between sand grains. Additionally, the inability to remove estriol (E3) in augmentation experiments could be due to toxicity induced by estriol and other estrogen degradation metabolites, which may explain the absence of fluorescent bacteria at the end of pure culture experiments (Figure 7.7). This toxicity is further supported by the evidence of filamentous morphology, increased concentrations of fluorophore and cellular components observed at the latter stages of pure culture

work which was potentially due to lysis induced by stress.

Although bioaugmentation was achieved, only a small percentage of the original inocula survived in the SSFs by the end of the study (Figure 7.15). This may also help explain the reduced estrogen removal compared with pure culture assays. Recently, several studies have shown that bioaugmentation can be improved by augmentation at different times and through nutrient management. For example, Iasur-Kruh et al. [2011] successfully bioaugmented an E2-degrading bacterium into a mature wetland biofilm, resulting in complete E2 removal and Gallego et al. [2001] showed that a 50% increase in diesel oil degradation could be achieved by managing the carbon/nitrogen and carbon/phosphorus ratios in soil microcosms. Therefore, future work should also aim to increase bioaugmented bacterial survival in SSFs by either augmentation at different stages of the SSF lifecycle (e.g. once a mature biofilm has formed) or through biostimulation of the estrogen-degrading bacteria. In addition, more work is needed to optimise physical removal mechanisms within the filters. For example anion exchange resin coated sand could help absorb some of the potentially toxic metabolites which may have induced bacterial death, hence if used in conjunction with bioaugmentation could result in increased estrogen removal.

7.5.3 Estrogen Exposure Affects Coliform Removal

Interestingly, although estrogen removal within SSFs was lower than expected, augmentation not only significantly improved estrogen removal (Table 7.10), but significantly increased the overall water quality performance of augmented SSFs. This increased performance was solely due to differences in coliform removal capabilities, with estrogen concentration negatively correlating with coliform removal in non-augmented filters effluents (Figure 7.12). This is the first example to the author's knowledge in any system, to show that estrogen concentration induces a negative effect on pathogen removal. The reduced coliform removal was hypothesised to be due to the inhibition of coliform-grazing protozoa (Figure 7.13). This, alongside the interesting finding that coliform removal in augmented SSFs was less affected by estrogen concentration than non-augmented SSFs, implies that augmentation reduced the toxic effect of estrogen on the coliform-grazing protozoan community. This is the first study to show that bioaugmentation not only improves the removal of the target chemical, but reduces the negative impact on treatment efficiency that such contaminants induce on water

quality production.

7.6 Conclusions

In this chapter, six estrogen-degrading bacterial strains were isolated from the industrial SSFs discussed in Chapters 4 and 5. Three of these isolates were used to augment laboratory-scale SSFs to assess whether estrogen removal and overall filter performance could be improved. To the best of the author's knowledge, this is the first study to isolate estrogen-degrading bacteria from SSFs and successfully use them to augment filter performance. To summarise, the main conclusions of this chapter are:

1. Bioaugmentation is possible and significantly improved the removal of estrone and estradiol in experimental slow sand filters;

2. Estrogen concentration is negatively correlated with coliform removal in non-augmented filters;

3. High estrogen concentrations cause reduced protozoan growth (18S rRNA qPCR measurements) and death resulting in reduced coliform removal. However, these effects were significantly lower in augmented filters compared to non-augmented filters.

With respect to current and future implications, the negative effects of estrogen exposure on coliform removal in SSFs pose questions for the operation of SSFs and other biological water filters (both waste water and drinking water). These include understanding the effects of natural estrogens on the performance, and micro- and macro- community of biological filters. This is important, especially due to increases in intensive farming and extreme weather phenomena, all of which result in elevated estrogen levels reaching various water sources. This is the first study to discover that estrogen exposure induces a negative effect on coliform removal in SSFs, potentially by the toxic inhibition of protozoan grazers. Future work should aim to determine the level of toxicity induced by estrogen exposure to protozoan grazers. Further, this work highlights the potentially serious consequences of unmonitored estrogen exposure, not only for water quality production but for the health of aquatic ecosystems in water sources such as reservoirs and rivers - which are used as water sources for SSFs.

Chapter 8

Differential Toxicity of Estrogens to Protozoan Species

"To every action there is always opposed an equal reaction: or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts".

Isaac Newton, (Principia Mathematica - Newton's Third Law)

Exposure to EDCs such as estrogens are now widely accepted to partially explain declines in the biodiversity of aquatic ecosystems, due to their adverse effects on the endocrine, reproductive and immune systems of aquatic organisms. Furthermore, recent study (Chapter 7) hypothesised that the impaired pathogen removal capabilities of SSFs seen during high estrogen exposure may be due to toxicity induced by estrogen exposure to the protozoa grazing community. Here, for the first time, the impact (of varying concentrations) of natural estrogens (estrone, estradiol and estriol) on three subgroups of protozoa (amoeboids, ciliates and flagellates), was examined through population impairment growth assays. Differing levels and extents of toxicity were found depending upon the subgroup of protozoa, with *Dictyostelium discoideum* (an amoeboid) being completely unaffected by both high and low concentrations of estrogens. In contrast, high estrogen concentrations were toxic to the growth of both *Euglena gracilis* (a flagellate) and *Tetrahymena pyriformis* (a ciliate), with the latter also being severely impeded at lower concentrations. The differential toxic effect induced by natural estrogens to the three subgroups of protozoa explains the reduced pathogen removal seen in Chapter 7. It also indicates the potential negative effects that estrogen exposure may have on protozoan populations in waterways, shifting the protozoa community makeup, with potential impacts both to higher and lower trophic levels. This highlights the

serious ramifications such exposure may have on the health, survival and functional performance of aquatic ecosystems and engineered systems.

8.1 Introduction

As discussed in Chapter 7, the occurrence of EDCs in the environment is a growing issue for human and wildlife development and reproduction [Belfroid et al., 1999, Ternes et al., 1999, Jobling et al., 2006, Zhou et al., 2010]. In particular, EDCs have recently been shown to be the main causative agents for the feminisation of fish, and for reproductive and developmental damage in wildlife, plants and humans [Jobling et al., 2006, Shore et al., 1995, Ternes et al., 1999]. Furthermore, in Chapter 7 estrogen exposure was for the first time shown to have a detrimental effect (reduced pathogen removal) on the performance of SSFs. This impaired functional performance was hypothesised to be due to the toxic inhibition of the protozoan grazers.

Despite the range of studies examining the effects of estrogens on fish, other aquatic organisms, and to a lesser extent humans [Caldwell et al., 2010], no study to date has investigated the impact of estrogens on environmental protozoa, which are one of the most important groups of organisms in aquatic environments [Martín-González et al., 2006]. Furthermore, within Chapter 6, protozoa, specifically *Monosiga brevicolis*, *Tetrahymena sp.* and various ciliates were shown to be essential for optimal pathogen removal (responsible for more than 99% of the *E.coli* removal) in SSFs. Ciliated, flagellated and amoeboid protozoa are found in SSFs and various other aquatic ecosystems and occupy the lower trophic levels within these ecosystems, either as primary producers or bacteriovores. Together they play essential roles in the transfer of matter and energy within the microbial loop and are integral to the health and survival of aquatic ecosystems. Moreover, due to their sensitivity to environmental changes they are considered excellent biological indicators of water quality and pollution [Martín-González et al., 2006, Valster et al., 2011, Shi et al., 2012].

The presence of estrogens in the environment is believed to come from two main sources (Figure 7.2); incomplete removal by sewage and wastewater treatment plants, and animal husbandry waste [Chen et al., 2010, Wang et al., 2011]. Recent surveys from around the

globe have revealed the broad occurrence of varying levels of E1, E2, E3 and EE2 in surface, ground and drinking waters [Shore et al., 1995, Ternes et al., 1999, Kuch and Ballschmiter, 2001, Ying et al., 2002, Williams et al., 2003, Quintana et al., 2004, Zhao et al., 2009]). Additionally, estrogens are persistent contaminants in sediments and soil [Ying et al., 2002, Wise et al., 2011] providing an additional source of estrogen leakage into surface waters.

In this study three different types of protozoa representing the three most dominant protozoan genera (the amoeba *Dictyostelium discoideum*, the ciliate *Tetrahymena pyriformis* and the flagellate *Euglena gracilis*) found in Chapter 6 to be involved with *E.coli* removal were used to assess the differential effects that exposure to natural estrogens may induce on growth. Results of which may help to explain the reduced coliform removal observed in Chapter 7 during periods of high estrogen exposure. Further, these three organisms were chosen to best represent the diverse groups of protozoa found within SSFs; additionally, the full genome sequences are available for all three organisms.

Hypotheses

Hypothesis 8.1 *Estrogen is toxic to protozoa.*

8.2 Materials and Methods

8.2.1 Cell Cultures and Estrogen Exposure

All toxicological assays used axenic cultures of: *Dictyostelium discoideum* B10 strain DBS0304514, *Tetrahymena pyriformis* strain CCAP 1630/1W and *Euglena gracilis* strain CCAP 1224/5Z. *Dictyostelium discoideum* were obtained from DictyBase, Northwestern University, USA, while both *Tetrahymena pyriformis* and *Euglena gracilis* were obtained from Strains of Culture Collection of Algae and Protozoa, UK. Details pertaining to strain-specific culturing can be found in sections 8.2.2-8.2.4.

Briefly, the density of protozoa was adjusted to 10^4 cells/ml in fresh growth media (specific to the protozoa). Natural estrogens (estrone, estradiol and estriol, purchased from Sigma-Aldrich, UK) were added to the cells at a final concentration of 50ng/L estone, 12ng/L estra-

diol and 39ng/L estriol (i.e., a total of 101ng/L) and at a 1 in 10 dilution of these concentrations (10.1ng/L), these concentrations corresponded to a total estrogenic potency of 0.23 (Equation 7.1). The higher concentration of 101ng/L was chosen to best represent, and simulate, agricultural or wastewater pollution events (based on averages from Chen et al. [2010]). Additionally, a non-exposed culture and a culture exposed to the solvent (ethyl acetate) used to dissolve the estrogens were set up. All assays were performed in triplicate in 25 ml culture vessels. Effects on growth were determined by cell counts using a haemocytometer under phase-contrast microscopy using a 100× optic, and three separate samples were counted at each time point.

8.2.2 Culturing *Dictyostelium discoideum*

Dictyostelium discoideum is an amoeba often referred to as a "social amoeba" or "slime mould", which has the ability during times of starvation to form a self-preserving multicelluar structure. To obtain an axenic culture of *D. discoideum*, the protocol outlined by Fey et al. [2007] was followed. Briefly, the protozoa were initially grown with *Klebsiella aerogenes* on SM plates at 22°C for 2 days. Then 5 x 10^3 freshly grown *D. discoideum* cells were inoculated from the SM plates into several 200ml flasks containing Developmental Buffer (5 mM Na_2HPO_4, 5 mM KH_2PO_4, 1 mM $CaCl_2$ and 2 mM $MgCl_2$) and *Escherichia coli* (adjusted to an OD_{600nm} of 8) and grown at 22°C, with shaking at 180rpm.

8.2.3 Culturing *Tetrahymena pyriformis*

The ciliated protozoan *Tetrahymena pyriformis* is an amicronucleate strain often used for toxicological studies. Growth conditions and procedure followed the protocol outlined by Cassidy-Hanley [2012]. Briefly, protozoa were grown axenically without shaking at 25°C in proteose peptone yeast extract media (PPY) and were maintained in exponential growth phase by reseeding in PPY liquid medium.

8.2.4 Culturing *Euglena gracilis*

Euglena gracilis is a unicellular flagellated protozoan that is both photoautotrophic and chemoheterotrophic. Growth conditions and procedure used can be found in Fleck et al. [2000]. Briefly, *E.gracilis* were grown axenically in several 10ml cultures of Proteose Pep-

tone media (PP) overnight at 23°C with 100 rpm. Cells were maintained in exponential growth phases by reseeding in fresh PP media.

8.2.5 Population Growth Impairment and Generation Time Determination

Effects on growth and generation time were determined following the protocol outlined by Dias et al. [2003]. Briefly, three aliquots of 100μl were immediately taken (T0) from the controls and the exposed protozoan cultures, and subsequently at 4, 8, 12, 24 and 48h. The samples were diluted in distilled water and visualised using a haemocytometer under phase contrast on an inverted optical microscope (Inverted Olympus IX71) at 100$\times$ magnification. Effects on protozoan population growth were characterised by their generation time (g) required for doubling the population. Generation time was calculated by the formulae in Equation 8.1 and Equation 8.2:

$$Number\,of\,generations\,(n) = \frac{logN_1 - logN_0}{log2} \tag{8.1}$$

where N_1 = number of cells at 24h and N_0 = The number of cells at T_0

$$Generation\,time\,(g) = \frac{Time\,of\,growth}{n} \tag{8.2}$$

where Time of growth = 24h

8.2.6 Statistics

In each assay, the experimental data represent the mean of three independent assays. Significant differences in protozoan growth between the controls and experimental assays were determined by analysis of variance (ANOVA) tests performed in the statistical package R [R Development Core Team, 2011]. Significance between groups was determined by p-values of less than 0.05.

8.3 Results

8.3.1 Population Growth Impairment of *Dictyostelium discoideum*

At the beginning of the experiment (T0) the cell density of *Dictyostelium discoideum* in the culture condition was $1.18 \times 10^4 \pm 9.62 \times 10^2$ cells/ml, with a normal generation time of 10.6 h. Addition of natural estrogens had little effect on growth and generation time (Table 8.1) of *D. discoideum*. Even the highest concentration (101ng/L) of estrogens had only a marginally significant effect on generation time (p-value: 0.0508, Figure 8.1A). However, after 48h of exposure at the high concentration there was a significant difference in population growth (p-value: 0.03). From these results the predicted lethal concentration resulting in the death of 50% of the population (LC_{50}) of *D. discoideum* is likely to be >101ng/L total estrogens.

8.3.2 Population Growth Impairment of *Euglena gracilis*

At the beginning of the experiment (T0) the cell density of *Euglena gracilis* in the culture condition was $6.60 \times 10^4 \pm 1.02 \times 10^3$ cells/ml, with a normal generation time of 19.9h. Addition of natural estrogens at lower concentrations resulted in no significant reduction in generation time (Table 8.1 and Figure 8.1B). However, at higher concentrations a significant effect can be seen on population growth (Figure 8.1B), with less than 50% of the original population of *E.gracilis* being viable after 24h, with this further decreasing to 34% after 48h. Taking these results into consideration the likely LC_{50} value for natural estrogens on *E. gracilis* is within the range of 10.1-101ng/L.

8.3.3 Population Growth Impairment of *Tetrahymena pyriformis*

At the beginning of the experiment (T0) the cell density of *Tetrahymena pyriformis* in the culture condition was $3.12 \times 10^4 \pm 1.10 \times 10^3$ cells/ml, with a normal generation time of 11.7h. The addition of natural estrogens had a significant effect on growth and generation time, with a clear concentration-dependent effect (Table 8.1 and Figure 8.1C). Low estrogen concentration resulted in a doubling of generation time compared with controls and no viable cells were detected after 24h of exposure at the higher concentration. Taking these results into consideration the likely LC_{50} value for natural estrogens on *T. pyriformis* is <101ng/L

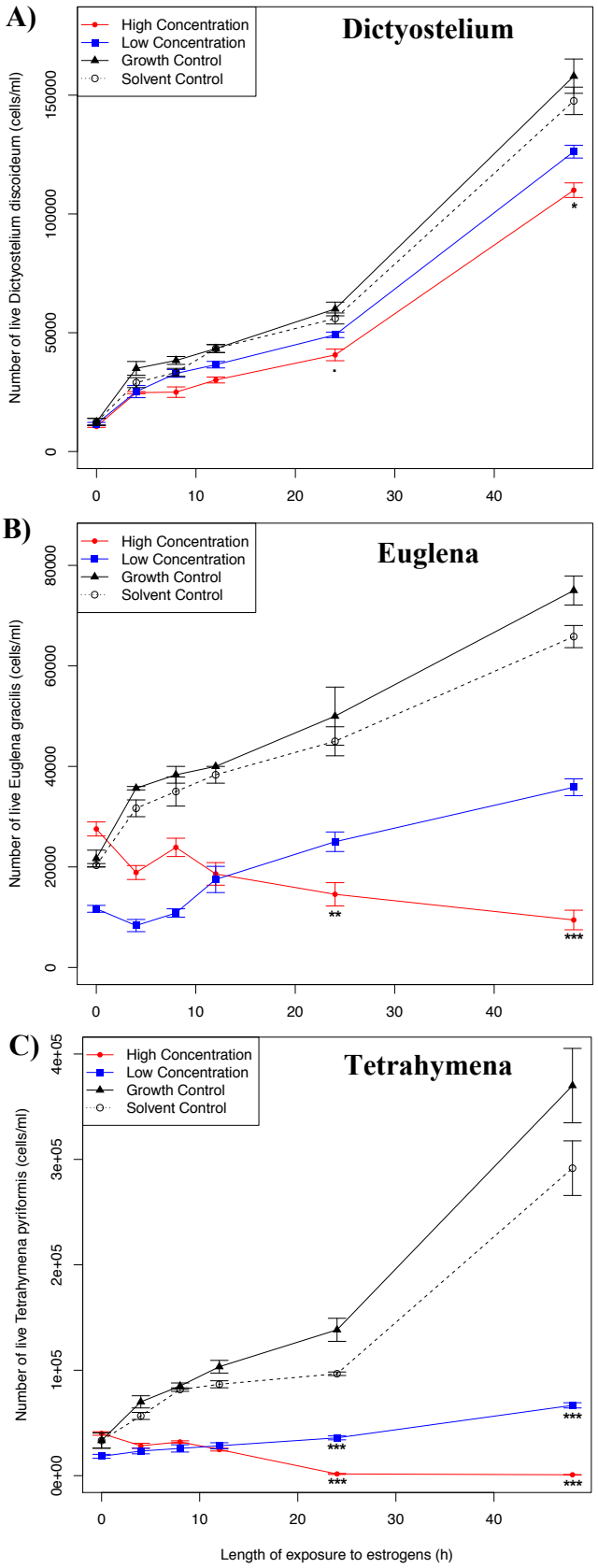


*Figure 8.1: Effect of natural estrogens on: **A)** D.discoideum, **B)** E.gracilis and **C)** T.pyriformis population growth during 48h. Concentrations of estrogens used were: high (101ng/L) and low (10.1ng/L). Data points represent the mean of triplicate assays and bars represent the standard error. ***Indicates significant differences (P-value: <0.0005), *(P-value: 0.05) and .(P.value: 0.03) from the control values.*

Table 8.1: The effect of natural estrogens on the generation time after 24 hours of growth on D.discoideum, T.pyriformis and E.gracilis.

Protozoa	Estrogen Concentration (ng/L)	Generation Time (h)	Increase in Generation Time (%)
T.pyriformis	Growth Control	11.67 ± 1.08	-
T.pyriformis	Solvent Control	12.77 ± 1.51	9.43
T.pyriformis	101	-	-
T.pyriformis	10.1	22.74 ± 1.49**	94.86
D.discoideum	Growth Control	10.61 ± 0.88	-
D.discoideum	Solvent Control	11.09 ± 0.76	4.52
D.discoideum	101	12.44 ± 0.82*	17.25
D.discoideum	10.1	11.57 ± 0.69	9.05
E.gracilis	Growth Control	19.89 ± 0.94	-
E.gracilis	Solvent Control	20.94 ± 1.36	5.28
E.gracilis	101	-	-
E.gracilis	10.1	23.86 ± 2.39	19.96

Increase in generation time is in respect to the control.

but $>$10.1ng/L.

8.4 Discussion

This is the first study to the author's knowledge to examine the effect of natural estrogens (E1, E2 and E3) on the growth of protozoa and link the findings to the reduced pathogen removal performance found in chapter 7. In this chapter it has been shown that there are differential effects to growth depending upon the protozoa and the concentration of estrogen used. Growth of both the flagellate (*Tetrahymena pyriformis*) and the ciliate (*Euglena gracilis*) were severely impeded, and in the latter case complete cell death was observed at high estrogen concentrations. However, at lower concentrations only the growth of the ciliate was significantly affected. These observations help to clarify, and provide explanations for the reduced coliform removal observed in Chapter 7. This, alongside the absence of an effect on the growth of the amoeba, highlights the importance of testing multiple organisms to ascertain the true toxicity of a chemical in the environment. Such differential responses to EDCs are not surprising, as similar findings arose in human protozoan parasite studies [Nava-Castro et al., 2012], and in aquatic vertebrate studies [Segner et al., 2003]. Reasons for such effects on growth can only be speculated due to the lack of knowledge pertaining to protozoan endocrinology. However, in various other EDC studies involving invertebrates, estrogens have been found to alter the phosphorylation state of mitogen-activated protein ki-

nases (MAPKs) and lysosomal membrane integrity [Canesi et al., 2004] resulting in various detrimental effects. It is therefore conceivable that similar processes may explain the effects in protozoa. Furthermore, based on the reduced coliform removal performance observed in Chapter 7, during high estrogen exposure, it is possible that EDC reduce the grazing capacity of protozoa; however, ingestion rate studies would be required to confirm this.

The reasons for the differences in effect between the three groups of protozoa are unclear. Possible explanations may be related to the differing binding affinity of estrogens to the receptors of the different protozoa, as mentioned previously. Additionally, potential reasons for the reduced effects in the amoeba could be due to the complete absence of either the ERα or ERβ estrogen receptor or homologues, unlike in *T. pyriformis* [Kohidai and Csaba, 2003] and *E. gracilis* [Cann, 2004]. However, it should be noted that although *D. discoideum* was unaffected by estrogens over the first 24h of exposure, there was significant impediment to growth after 48h. This implies that although the amoeba is not as sensitive as both the ciliate and flagellate to initial estrogen exposure, extended exposure is detrimental. A possible reason for such a delayed effect may be the critical buildup of ROS or other damaging radicals (e.g. hydroxyl radicals), produced in the mitochondria when oxygen is reduced along the electron transport chain. Internal ROS accumulation induced by TiO_2 and fullerene exposure has been shown to interfere with the correct functioning of various cofactors and enzymes, resulting in the disruption of protein structure and function, and in reduced growth rates and survival [Jafar and Hamzeh, 2013].

Hypothetically, if an environmental pollution event such as large-scale agricultural runoff into waterways was to arise, these results suggest that the survival of these three organisms would be questionable. The results of this would be fatal as *E. gracilis* is a key primary producer in aquatic environments [Fleck et al., 2000], and both *T. pyriformis* and *D. discoideum* are key members of the base of the microbial and detritivore food webs, playing important roles as grazers of microbes in aquatic environments and controlling bacterio-plankton production. Together these three organisms form the basis of the first trophic level in aquatic environments, and their disruption could have serious knock-on effects to higher trophic levels. In addition to the general environmental impact, increased estrogen exposure into waterways; typically used to supply slow sand filters would have serious knock-on-effects

for optimal SSF performance, in particular for effective *E.coli* removal (Chapter 7). Therefore, from a functional perspective the toxicity of these protozoa in particular *Tetrahymena spp.*, which were shown in Chapter 6 to be very important for *E.coli* removal, would have serious implications for the pathogen removal performance of SSFs and potentially other water purification systems with biological components.

8.5 Conclusions

Due to the alarming and differential effects of estrogens on protozoa and the knock-on-effect to SSF performance discovered in this study, serious issues for the functional performance of SSFs and survival and health of aquatic food webs and ecosystems are posed. Further work is required to determine the individual toxicity and LC_{50} of E1, E2 and E3 to various protozoan members, alongside the characterisation of the mode(s) of toxicity. This, alongside further studies in LSSFs and river microcosms (which will more closely mimic the true aquatic environment) [Gerhardt et al., 2010] and SSFs with varying concentrations, periods of exposure and estrogenic potencies, will allow more accurate conclusions to be drawn about the potential deleterious effects of estrogen exposure. To conclude, this study is the first to show that natural estrogen exposure has negative effects on protozoan population growth and survival and that the extent of such effects are protozoan specific. Further, it is the first study to explain the functional implications (reduced pathogen removal) that estrogen exposure has on engineered systems performance.

Chapter 9

Conclusions and Future Work

"Knowledge is of no value unless you put it into practice."

Anton Chekhov

The supply of clean and safe drinking water, free from any substances or organisms, which pose a danger to human health, is an objective of the European Union Drinking Water Directive (98/83/EC) and World Health Organisation (WHO) [World Health Organisation, 2004]. For over 200 years slow sand filtration has been an effective means of treating water for the control of microbiological and chemical contaminants in both small and large community water supplies [Huisman et al., 1974, Haig et al., 2011]. This capability and efficiency to remove various contaminants has underpinned SSF deployment in various areas outside of drinking water purification including: aquaculture [Arndt and Wagner, 2004], horticulture [Calvo-Bado et al., 2003], storm-water purification [Urbonas, 1999] and food and drink waste management [Ramond et al., 2013]. However, despite their adoption and use in energy-efficient production of high-quality water, little is understood about the biological mechanisms and organisms responsible for producing the diverse and efficient functional capacity offered by SSFs [Haig et al., 2011]. This lack of knowledge has and will continue to hinder optimisation in design, management and operation of these systems.

The focus of this thesis was to build on the limited ecological knowledge of SSFs using techniques to identify the culturable and non-culturable microorganisms, and to relate diversity and specific taxa to water quality production. Understanding these dynamics alongside performance-induced effects associated with operational differences and exposure to novel contaminants will promote optimised SSF design, maintenance and operation, creating more

efficient and environmentally sustainable filters. Specifically, this thesis aimed to address the following questions:

1. Which microorganisms are present in full-scale industrially operated slow sand filters and which roles do they perform?

2. Does the microbial community structure change temporally and spatially in slow sand filters?

3. Can a laboratory-scale slow sand filter be constructed to mimic the performance and microbial community of full-scale, industrially operated slow sand filters?

4. What is the impact of light on the microbial community and filter performance?

5. Which mechanisms are responsible for the removal of the human pathogen *E.coli* in slow sand filters?

6. How effective are slow sand filters at removing estrogen and can their performance be improved by bioaugmentation?

To address these complex questions a two-pronged approach was taken: firstly the microbial community of two full-scale slow sand filters (Chapter 4) was surveyed for eight months, representing all stages of the filters lifecycle (scraped, ripened, clogged and drained). Secondly, to better understand pathogen removal mechanisms, as well as the impact of light exposure (Chapter 6) and of estrogen exposure (Chapter 7-8), eight laboratory-scale SSFs were designed, constructed and operated (Chapter 5). Overall, this thesis has shown that SSFs are phylogenetically and metabolically diverse systems capable of producing high-quality water, with the ability to adapt to remove novel contaminants.

The remainder of this chapter is structured as follows; Section 9.1 lists the contributions of this thesis. Section 9.2 outlines directions for future work, describing improvements that can be made to the work presented here, and explores the wider implications of this work and how it fits into other areas of research; and Section 9.3 provides a summary of this chapter, and concludes the thesis.

9.1 Contributions

The contributions of this thesis are as follows:

Detailed Characterisation of the Microbial Community of Full-Scale SSFs

Chapter 4 describes the periodic sampling of two industrially operated slow sand filters was presented in order to study the spatial and temporal structure of the bacterial community comprising the filters, and to determine how specific microbial groups were related to overall filter performance. The results of this chapter showed that the microbial diversity of SSFs is far greater than previously documented and that the month; age; side; distance from the influent and effluent pipe; and depth from which the samples were taken significantly impacted the microbial community in SSFs, with age being the most significant variable. As filters aged, both the number and density of OTUs increased, as did species evenness. Further, Illumina 16S rRNA amplicon sequencing indicated that the abundance of various members of the microbial community, specifically *Acidovorax, Halomonas, Sphingobium* and *Sphingomonas*, were important for performance. Moreover it was found that increased species evenness was critical for excellent filter performance, with decreased species evenness being found in drained and early-stage SSFs coinciding with increased abundance of *Planctomycetes*. Although this chapter presents the most detailed microbial SSF survey to date, additional sampling extending through multiple filter life-cycles would provide more statistically robust taxa-function information. This, however, was not possible as the site was decommissioned.

Verification that Full-scale SSFs can be Replicated in the Laboratory

Chapter 5 describes eight laboratory-scale SSFs that were designed, constructed, operated and applied to demonstrate, for the first time, that the water quality and microbial community of full-scale SSFs could be replicated in the laboratory. Phylum-specific qPCR assays and 454 pyrosequencing of the 16S rRNA gene revealed that whilst the laboratory and industrial filters appeared identical at a coarse taxonomic level, when OTUs (species proxies) were considered they differ consistently. Given that they also appear equivalent in terms of function, a degree of redundancy is suggested. It is unclear what is driving the differences at the OTU level, but it is likely environmental differences, such as temperature and

the presence or absence of wildlife between the laboratory and the industrial filters have a part to play. However, this requires further study. In addition to the parameters (age, depth from which samples were taken and water quality produced) identified in Chapter 4 to explain differences in the microbial community of full-scale SSFs, the sand type (sterile or non-sterile) was also found to be a significant factor. However, irrespective of the type of sand within lab-scale SSFs the microbial consortia converged in all filters after seven weeks of operation. Such convergence is staggering and suggests that the microbial community converges towards a structure which is best for optimal performance. In particular the reappearance from the ISSFs discussed in Chapter 4 of the importance of the *Comamonadaceae* and *Sphingomonadacea* families when filters age - further strengthens conclusions of their importance in performance and the ability of LSSFs to replicate the functional community of ISSFs. Furthermore, although water quality parameters were measured in filter effluents in order to accurately relate microbial taxa to specific function, the use of depth specific probes would provide more accurate data. Based on this proof of concept study, it is now possible to use these laboratory-scale SSFs to ask more complex questions relating to water quality and community assembly and relate any findings directly to full-scale units.

Ecosystem-wide Responses are Involved in *E.coli* Removal in SSFs

In Chapter 6, the laboratory-scale SSFs first described in Chapter 5 were used to determine the mechanisms responsible for *E.coli* removal, using stable-isotope-probing in conjunction with metagenomics. This study is the first to follow the fate of an isotopically-labelled pathogen through a biological system using metagenomics. Direct counts and qPCR assays revealed a clear predator-prey response between protozoa and *E.coli*. Furthermore, the importance of top-down trophic-interactions were confirmed by metagenomic analysis, identifying several protozoa (*Monosiga brevicollis* and *Tetrahymena spp.*) and viral species (*Enterobacteria phages*) connected to *E.coli* attrition, with protozoan grazing responsible for $>$99% of the removal. In addition to top-down mechanisms, indirect mechanisms such as algal reactive oxygen species (ROS)-induced lysis, and mutualistic interactions between algae and fungi, were suggested to also be associated with coliform removal. Although complex interactions involved in *E.coli* removal were determined in this chapter, more detailed food webs may be created by sampling at smaller time intervals and performing mass balance calculations. Furthermore, in order to determine the specific mechanisms deployed by or-

ganisms RNA-based sequencing and metatranscriptomics would be required. To summarise, whilst various studies have shown the individual importance of viral lysis, protozoan grazing and endogenous, and exogenous, ROS in *E.coli* removal, chapter 6 presents the first study, to the author's knowledge, to show the importance and interactions of all of these mechanisms for pathogen removal in SSFs.

Functional Implications of Covering SSFs (Operation in Darkness)

Within Chapter 6, the effects of light exposure on the performance and microbial community in SSFs was explored using qPCR and water quality analysis; from an engineering perspective, this was to determine if there were differences between covered (e.g. used in the Netherlands) and uncovered filters (e.g. used in the UK and USA). Overall, light exposure had significant spatial and temporal effects on the composition of the microbial community of LSSFs, with non-covered (exposed to light) LSSFs developing a more even community composition compared to covered (no light exposure) LSSFs. Interestingly, differences in the microbial community composition and abundance had little effect on overall filter performance ($\bigtriangledown$). This study is the first to show that the microbial community of covered and non-covered SSFs are extremely divergent in composition and abundance yet achieve the same level of excellent filter performance. Although, overall, there was no difference in $\bigtriangledown$ between covered and non-covered SSFs, at early time-points (0-3 weeks) in the filter lifecycle there were significant differences in coliform removal potential, with non-covered filters removing 24 times more than covered SSFs. This, along with the increased retention of viable *E.coli* in covered filters at all depths, implies that light exposure facilitates *E.coli* removal. Furthermore, the reduced capacity for coliform removal in covered filters in the three weeks after the *E.coli* spiking event suggests covered SSFs incur long-term effects from prolonged *E.coli* exposure which have a knock-on effect to coliform removal.

Estrogen-Degrading Bacteria Exist in SSFs

Within Chapter 7 the potential to improve estrogen removal and overall filter performance by bioaugmentation with estrogen-degrading bacteria was explored. In order to promote optimal survival of the introduced bacterial strains within LSSFs, strains were isolated from the full-scale SSFs described in Chapters 4 and 5. Overall, six bacterial strains capable of

degrading at least one of the natural estrogens (E1, E2 or E3) at room temperature and 4°C were isolated. The genomes of the three room-temperature isolates: E1-rm, E2-rm and E3-rm all possessed several known enzymes involved in estrogen degradation. However, the E3-rm isolate unlike the other two isolates did not possess any known degradation enzymes involved in its enriched estrogen source (E3) metabolism, despite its 100% removal in pure culture experiments. A limitation of this study was the absence of RNA sequencing which would have revealed the functionally relevant enzymes which were being used by the isolates. This is the first study to show that SSFs support the survival of estrogen degraders and to obtain pure culture isolates.

Bioaugmentation Improves Estrogen Removal in SSFs

Within Chapter 7 LSSFs were bioaugmented with the three room-temperature estrogen degrading isolates enriched from full-scale SSFs. Bioaugmentation was shown by qPCR to be successful due to the survival of the organisms within the filters throughout the study. Functionally, augmentation significantly improved the removal of E1 and E2 and reduced the overall estrogenic potency of the influent water by 26%, compared to non-augmented filters. However, this removal was lower than expected based on pure culture growth kinetic experiments, which is likely due to the inability of the isolates to form their mat-like biofilm formation between sand grains and due to natural competition from other microorganisms within the filters. To the best of the author's knowledge this is the first example of augmentation in a drinking water system with estrogen degrading microbes and the first demonstration that bioaugmentation can improve SSF performance.

High Estrogen Exposure Causes Reduced Pathogen Removal

In addition to the improved estrogen removal due to the bioaugmentation discussed in Chapter 7, estrogen concentration was found to negatively correlate with coliform removal in non-augmented filters. This was hypothesised and shown within chapter 8 to be due to the toxic inhibition of coliform-grazing protozoa. Interestingly, coliform removal in augmented filters was less affected by estrogen concentration than non-augmented filters, implying that augmentation reduced the toxic effect of estrogen on the coliform-grazing, protozoan community. This study underscores the importance of investigating the effects of emerging con-

taminants on the micro- and macro-community of SSFs and further shows that bioaugmentation of SSFs with estrogen degraders is beneficial for both chemical and microbiological removal. This is the first study, to the author's knowledge, to demonstrate the negative effect of estrogens on pathogen removal in any system and demonstrate the toxicity of estrogen exposure to protozoa.

9.2 Future Directions

The work within this thesis has demonstrated that the microbial community of SSFs are metabolically and phylogenetically rich, influenced by various operational parameters (exposure to light, and estrogen concentration) and capable of producing high-quality drinking water. Furthermore, this thesis has shown links between various taxa and optimal filter performance and revealed that multi-trophic interactions appear to be responsible for *E.coli* removal. Additionally, it has demonstrated that SSFs can be augmented with estrogen degrading bacteria which promotes improved estrogen removal and confers protection to the coliform-grazing protozoan community which were inhibited by estrogen exposure. However, the work presented in the preceding chapters provide several opportunities for future research, which will be described in this section.

9.2.1 Comparison Between Geographic Areas and Technologies

One obvious direction for future work is to expand the microbial ecology survey, to study other SSFs in different geographic regions in order to determine if there is a universal optimal SSF community. Additionally, it would be interesting to compare the communities between different treatment systems e.g. Manz filters, household SSFs, rapid sand filters, and GAC filters to allow functionally relevant taxa to be identified and in the future monitor and manage them. Alongside the community characterisation, functional screens could be performed using standard and custom designed EcoPlates (Biolog, UK) to determine the metabolic capabilities of the SSF community.

9.2.2 Filter Design, Maintenance and Operation

As discussed in Chapter 2, extensive knowledge about the physical mechanisms within SSFs is understood, however the future of biological filter technology lies in optimising both the physical and biological mechanisms. Therefore, future work should aim to manage and boost the survival and metabolic degradation profile of SSFs in addition to pairing this with studies comparing the impact of different filter bed materials. For example a SSF bed composed of a layer of granular activated carbon may allow absorption of certain contaminants which are either not or incompletely removed via microbial degradation. Such synergistic studies will potentially provide tailored water purification systems specific to the needs of the user. Additionally from Chapter 5 the functional implications of covering SSFs in regards to reduced performance during heavy coliform exposure suggest that non-covered SSFs are optimal for the UK. However, it would be interesting to compare this functional effect in covered and non-covered filters in other geographical locations.

In Chapter 4 the decreased species evenness and increased abundance of *Planctomycetes* found in drained and early-stage SSFs was likely due to additional sunlight exposure, these findings pose several avenues for future work focussed on improving SSF maintenance. For example, future work should investigate the impact of reducing the drainage period or the effects of covering filters during draining and scraping events on species evenness and the abundance of *Planctomycetes*. Such work could significantly reduce the period of time SSFs are non-operational due to poor performance and hence have economical benefits.

Regarding the operation of SSFs, there are several areas of improvement and future work which have come to light from this thesis. Firstly, the major disadvantage of SSFs compared to other systems is the slow production of purified water, therefore additional work is required into developing an "intermediate sand filter" i.e., a sand filter operating at a filtration rate between that of rapid and slow sand filters. Secondly, from the work within this thesis it was shown that filter performance (produce better water quality) improves with age, however it is unclear as to whether this is due to changes in the community or due to increased exposure to contaminants. From Chapter 6 it was noted that during periods of high coliform number exposure, filters showed increasing removal capabilities in correlation to the number of times these events occurred. Likewise in Chapter 7, during the two periods

of high estrogen exposure both filter types showed improved removal and non-augmented filters additionally displayed an increased protective effect to the protozoan community during the second estrogen exposure, in-line with the augmented filters. These two phenomena are suggestive that the SSF community can be primed to respond rapidly to contaminants through increased exposure. Therefore future work might aim to research "vaccination like" procedures i.e., expose SSFs (following the same approach used in Chapters 6 and 7) to the contaminant of choice in various quantities, at different points in the filters lifecycle (particularly during early stages) and for different durations in order to determine if improved exposure can induce an adaptive immune like response in the SSFs.

9.2.3 Predictive Water Quality Modelling

Another avenue for further development is the creation of predictive water quality models based on species evenness information and qPCR quantified abundances of *Halomonas, Sphingomonas, Acidovorax* and *Sphingobium* which were found within this study to give accurate predictions of filter performance. Such models would allow operators to determine the water quality production of a filter in a faster time-frame than that achieved using tradition water quality analysis methods. Additionally, integrating this model with the functional effects of differing estrogen concentrations will allow more reliable predictions of coliform numbers to be determined in filter effluents. In addition to these water quality models, mathematical models and food-web interaction models based on the *E.coli* removal mechanisms discovered in Chapter 6 should be created. However, additional work is required to integrate physical removal mechanisms into such a model. Further, the DNA-SIP study discussed within this thesis should be replicated using various pathogenic strains of *E.coli* to determine if the removal mechanisms are the same.

9.2.4 Metabolic Limits

Due to the increasing number of contaminants being added to the EU Drinking Water Directive (98/83/EC), there is a need to find solutions for current and emerging contaminant removal. Within this thesis the ability of SSFs to remove estrogens has been shown, however future work should focus on determining the metabolic limit of SSFs, through metagenomic, metatranscriptomic and traditional enrichment culture based studies. Additionally, this thesis

has shown that SSFs are amenable to bioaugmentation. Therefore future work should aim to harness this potential by further laboratory-scale studies trialling different augmentation and biostimulation procedures. For example the survival of augmented organism could be boosted by encapsulating them in alginate, hence allowing the diffusion of products in and out but reducing the risk of predation and washout in the filters.

9.2.5 Integration with Other Systems

Although SSFs have been shown within this thesis to be effective purification systems, they are known to struggle with high turbidity and lack the ability to remove colour. Therefore in order to be effectively utilised to purify a range of water sources (grey-water, storm water, waste water, and water obtained from sustainable urban drainage (SUDS)) they will need to be linked to other technologies. For example SSFs could be paired with portable constructed wetlands to be used in nurseries and greenhouses as an economical and ecologically based approach for treating runoff containing nutrients, pesticides, pathogens and other organic and biological contaminants. Such a setup would potentially allow nutrient / resource recovery which may have financial benefits. Furthermore SSFs could be used to purify road-runoff collected in SUDS, hence providing sustainable, economical and environmentally friendly treatment technologies to recycle water or to assure compliance with increasingly stringent environmental regulations regarding the discharge of non-point-source pollutants.

9.3 Closing Remarks

The work presented within this thesis has provided a better understanding of the spatial, temporal and functionally induced changes in SSF microbial communities. It has also shown that SSFs are phylogenetically and metabolically diverse systems capable of producing high quality water, with the ability to adapt to remove novel contaminants. Using the information gathered within this thesis alongside future research will allow improvements to filter maintenance and operation to be achieved. It is my belief that the future of drinking water purification does not lie in developing a new technology, but in improving the way current systems such as SSFs are operated and managed. By optimising the synergy between physical and biological removal mechanisms, SSFs could be the most adaptable, efficient, sustainable and cost-effective treatment solution of the future.

Bibliography

I.A. Aamot. *How Photosynthesis in Laminaria digitata and Saccharina latissima is Affected by Water Temperature*. PhD thesis, Norwegian University: Science and Technology, 2011.

S. Abedon. *Bacteriophage Ecology: Population Growth, Evolution and Impact of Bacterial Viruses. Advances in Molecular and Cellular Microbiology*, chapter 5. Contribution of lysogeny, pseudolysogeny and starvation to phage ecology. Cambridge University Press, 2008.

J Abrahamsson and P Dromberg. Covering slow sand filters; qualitative and operational aspects. In *Recent Progress in Slow Sand and Alternative Biofiltration Processes*, page 231. IWA Publishing, 2006.

S.D Allison and J.B.H. Martiny. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences*, 105(1):11512–11519, 2008.

L. Alonso-Sáez, J.M Gasol, T. Lefort, J. Hofer, and R. Sommaruga. Effect of natural sunlight on bacterial activity and differential sensitivity of natural bacterioplankton groups in northwestern Mediterranean coastal waters. *Applied and Environmental Microbiology*, 72 (9):5806–5813, 2006.

DA Alvarez, NW Shappell, LO Billey, DS Bermudez, VS Wilson, DW Kolpin, SD Perkins, N Evans, WT Foreman, JL Gray, et al. Bioassay of estrogenicity and chemical analyses of estrogens in streams across the United States associated with livestock operations. *Water Research*, 47(10):3347–3363, 2013.

R.I. Amann, W. Ludwig, and K.H. Schleifer. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiology and Molecular Biology Reviews*, 59(1):143– 165, 1995.

L.A. Amaral-Zettler, E.A. McCliment, H.W. Ducklow, and S.M. Huse. A method for studying protistan diversity using massively parallel sequencing of v9 hypervariable regions of small-subunit ribosomal rna genes. *PLoS One*, 4(7):e6372, 2009.

M.J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, 2001.

S Andrews. FastQC A Quality Control tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*, 2010.

R.E Arndt and E.J Wagner. Rapid and slow sand filtration techniques and their efficacy at filtering triactinomyxons of *Myxobolus cerebralis* from contaminated water. *North American Journal of Aquaculture*, 66(4):261–270, 2004.

K.E. Ashelford, N.A. Chuzhanova, J.C. Fry, A.J. Jones, and A.J. Weightman. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, 71(12):7724–7736, 2005.

S. Aslan. Biological nitrate removal in a laboratory-scale slow sand filter. *Water S. A.*, 34 (1):99–106, 2008.

S. Aslan and H. Cakici. Biological denitrification of drinking water in a slow sand filter. *Journal of hazardous materials*, 148(1-2):253–258, 2007.

R.K Aziz, D. Bartels, A.A Best, M. DeJongh, T. Disz, R.A Edwards, and et al. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1):75, 2008.

E. Bååth. Estimation of fungal growth rates in soil using ^{14}C-acetate incorporation into ergosterol. *Soil Biology and Biochemistry*, 33(14):2011–2018, 2001.

M. Bahgat, A. Dewedar, and A. Zayed. Sand-filters used for wastewater treatment: buildup and distribution of microorganisms. *Water Research*, 33(8):1949–1955, 1999.

Y. Bai, R. Liu, J. Liang, and J. Qu. Integrated Metagenomic and Physiochemical Analyses to Evaluate the Potential Role of Microbes in the Sand Filter of a Drinking Water Treatment System. *PloS one*, 8(4):e61011, 2013.

M.N. Baker and M.J. Taras. *Quest for Pure Water*. American Water Works Association, Lancaster; PA, 1948.

C. Baronti, R. Curini, G. D'Ascenzo, A. Di Corcia, A. Gentili, and R. Samperi. Monitoring natural and synthetic estrogens at activated sludge sewage treatment plants and in a receiving river water. *Environmental Science & Technology*, 34(24):5059–5066, 2000.

J.M.S Bartlett and D Stirling. A short history of the polymerase chain reaction. In *PCR protocols*, pages 3–6. Springer, 2003.

A.K Bartram, M.D.J Lynch, J.C Stearns, G. Moreno-Hagelsieb, and J.D Neufeld. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology*, 77 (11):3846–3852, 2011.

D. Bass. Private communication: Protozoan specific 18s primers, Feburary 2013. Private Communication: Protozoan specific 18S primers.

C.J Bauer. Results of Chilean water markets: Empirical research since 1990. *Water Resources Research*, 40(9):1–11, 2004.

R. Bauer, H. Dizer, I. Graeber, K.H. Rosenwinkel, and J.M. López-Pila. Removal of Bacterial Fecal Indicators, Coliphages and Enteric Adenoviruses from Waters with High Fecal Pollution by Slow Sand Filtration. *Water Research*, 45:439–452, 2011.

AC Belfroid, A Van der Horst, AD Vethaak, AJ Schäfer, GBJ Rijs, J Wegener, and WP Cofino. Analysis and occurrence of estrogenic hormones and their glucuronides in surface water and waste water in The Netherlands. *Science of the Total Environment*, 225 (1):101–108, 1999.

T. Bell, J.A Newman, B.W Silverman, S.L Turner, and A.K Lilley. The contribution of species richness and composition to bacterial services. *Nature*, 436(7054):1157–1160, 2005.

W.D. Bellamy, D.W. Hendricks, and G.S. Logsdon. Slow sand filtration: influences of selected process variables. *Journal of the American water works association*, 77(12):62, 1985.

G.D Bending, S.D Lincoln, S.R Sørensen, J.A.W Morgan, J. Aamand, and A. Walker. In-field spatial variability in the degradation of the phenyl-urea herbicide isoproturon is the result of interactions between degradative *Sphingomonas* spp. and soil pH. *Applied and Environmental Microbiology*, 69(2):827–834, 2003.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, pages 1165–1188, 2001.

E. Bi and J Lutkenhaus. Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *Journal of bacteriology*, 175(4):1118–1125, 1993.

C.B. Blackwood, A. Oaks, and J.S. Buyer. Phylum-and class-specific pcr primers for general microbial community analysis. *Applied and environmental microbiology*, 71(10):6193, 2005.

A.M.S. Bomo, T.K. Hovi, I. Hanssen, and J. Fredrik. Bacterial Removal and Protozoan Grazing in Biological Sand Filters. *Journal of Environmental Quality*, 33:1041–1047, 2004.

N. Boon, C. Marle, E.M. Top, and W. Verstraete. Comparison of the spatial homogeneity of physico-chemical parameters and bacterial 16S rRNA genes in sediment samples from a dumping site for dredging sludge. *Applied microbiology and biotechnology*, 53(6):742–747, 2000.

D.G. Bourne, R.L. Blakeley, P. Riddles, and G.J. Jones. Biodegradation of the cyanobacterial toxin microcystin lr in natural water and biologically active slow sand filters. *Water research*, 40(6):1294–1302, 2006.

R.D Briciu, A. Kot-Wasik, and J. Namiesnik. Analytical challenges and recent advances in the determination of estrogens in water environments. *Journal of Chromatographic Science*, 47(2):127–139, 2009.

N. Brink. Ecological studies in biological filters. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 52(1):51–122, 1967.

A.J Brook. The attached algal flora of slow sand filter beds of waterworks. *Hydrobiologia*, 7(1-2):103–117, 1955.

NP Burman. Slow sand filtration. H_2O, 11(16):348–351, 1978.

NP Burman and J. Lewin. Microbiological and operational investigation of relative effects of skimming and in situ sand washing on two experimental slow sand filters. *Journal of the Institution of Water Engineers*, 15:355–367, 1961.

D.J Caldwell, F. Mastrocco, E. Nowak, J. Johnston, H. Yekel, D. Pfeiffer, and et al. An assessment of potential exposure and risk from estrogens in drinking water. *Environmental health perspectives*, 118(3):338–344, 2010.

F.A. Caliman and M. Gavrilescu. Pharmaceuticals, personal care products and endocrine disrupting agents in the environment–a review. *CLEAN–Soil, Air, Water*, 37(4-5):277–303, 2009.

L.A. Calvo-Bado, T.R. Pettitt, N. Parsons, G.M. Petch, J.A.W. Morgan, and J.M. Whipps. Spatial and temporal analysis of the microbial community in slow sand filters used for treating horticultural irrigation water. *Applied and environmental microbiology*, 69(4): 2116 – 2125, 2003.

LC Campos, MFJ Su, NJD Graham, and SR Smith. Biomass development in slow sand filters. *Water research*, 36(18):4543–4551, 2002.

L.C. Campos, S.R. Smith, and N.J.D. Graham. Deterministic-based model of slow sand filtration. ii: Model application. *Journal of Environmental Engineering*, 132:887, 2006.

L. Canesi, L.C. Lorusso, C. Ciacci, M. Betti, M. Zampini, and G. Gallo. Environmental estrogens can affect the function of mussel hemocytes through rapid modulation of kinase pathways. *General and comparative endocrinology*, 138(1):58–69, 2004.

M. Cann. Mini-theme critical review on nucleotidyl cyclases Mini-theme Critical Review on Nucleotidyl Cyclases: Bicarbonate Stimulated Adenylyl Cyclases. *IUBMB Life*, 56(9): 529–534, 2004.

J.G Caporaso, C.L Lauber, W.A Walters, D. Berg-Lyons, J. Huntley, N. Fierer, and et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, 6(8):1621–1624, 2012.

E. Cardenas and j.m. Tiedje. New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology*, 19(6):544 – 549, 2008.

J.L Cartinella, T.Y Cath, M.T Flynn, G.C Miller, K.W Hunter, and A.E Childress. Removal of natural steroid hormones from wastewater using membrane contactor processes. *Environmental Science & Technology*, 40(23):7381–7386, 2006.

T Carver, N Thomson, A Bleasby, M Berriman, and J Parkhill. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25(1):119–120, 2009.

D Caserta, L Maranghi, A Mantovani, R Marci, F Maranghi, and M Moscarini. Impact of endocrine disruptor chemicals in gynaecology. *Human reproduction update*, 14(1):59–72, 2008.

D.M Cassidy-Hanley. Tetrahymena in the Laboratory: Strain Resources, Methods for Culture, Maintenance, and Storage. *Methods in cell biology*, 109:237–276, 2012.

M. Castillo, S.M. Martin-Orue, E.G. Manzanilla, I. Badiola, M. Martín, and J. Gasa. Quantification of total bacteria, enterobacteria and lactobacilli populations in pig digesta by real-time pcr. *Veterinary microbiology*, 114(1-2):165–170, 2006.

T-S. Chen, T-C Chen, K-J.C Yeh, H-R. Chao, E-T. Liaw, C-Y. Hsieh, K-C. Chen, and et al. High estrogen concentrations in receiving river discharge from a concentrated livestock feedlot. *Science of the Total Environment*, 408(16):3223–3230, 2010.

U. Christaki, E. Vázquez-Domínguez, C. Courties, and P. Lebaron. Grazing impact of different heterotrophic nanoflagellates on eukaryotic (Ostreococcus tauri) and prokaryotic picoautotrophs (Prochlorococcus and Synechococcus). *Environmental Microbiology*, 7 (8):1200–1210, 2005.

R Clarke, K. Non-parametric multivariate analyses of changes in community structure. *Australian journal of ecology*, 18(1):117–143, 1993.

L.S. Clesceri, A.E. Greenberg, A.D. Eaton, and American Public Health Association, editors. *Standard methods for the examination of water and wastewater*. American Public Health Association Washington, DC, 21st edition, 2005.

JR Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, RJ Farris, and et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37(S1):D141–D145, 2009.

G. Collins, S. Kavanagh, S. McHugh, S. Connaughton, A. Kearney, O. Rice, and et al. Accessing the black box of microbial diversity and ecophysiology: recent advances through polyphasic experiments. *Journal of Environmental Science and Health Part A*, 41(5):897–922, 2006.

O. Conroy, A.E Sáez, D. Quanrud, W. Ela, and R.G Arnold. Changes in estrogen/anti-estrogen activities in ponded secondary effluent. *Science of the Total Environment*, 382 (2):311–323, 2007.

KE Cooksey and B. Wigglesworth-Cooksey. Adhesion of bacteria and diatoms to surfaces in the sea: a review. *Aquatic Microbial Ecology*, 9(1):87–96, 1995.

S. Cosa, A.M Ugbenyen, L.V Mabinya, K. Rumbold, and A.I Okoh. Characterization and flocculation efficiency of a bioflocculant produced by a marine *Halobacillus*. *Environmental Technology*, 34(17):2671– 2679, 2013.

K.O. Cranston and A. Amirtharajah. Improving the initial effluent quality of dual-media filter by coagulants in backwash. *Journal American Water Works Association*, 79(12): 50–63, 1987.

M. Cunliffe and M.A Kertesz. Effect of *Sphingobium yanoikuyae* B1 inoculation on bacterial community dynamics and polycyclic aromatic hydrocarbon degradation in aged and freshly PAH-contaminated soils. *Environmental Pollution*, 144(1):228–237, 2006.

T.P Curtis, D.D Mara, and S.A Silva. Influence of pH, oxygen, and humic substances on ability of sunlight to damage fecal coliforms in waste stabilization pond water. *Applied and Environmental Microbiology*, 58(4):1335–1343, 1992.

H. Dang and C.R. Lovell. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rrna gene restriction analysis and sequence analysis of 16s rrna genes. *Applied and environmental microbiology*, 66(2):467, 2000.

M. Danger, J. Cornut, E. Chauvet, P. Chavez, A. Elger, and A. Lecerf. Benthic algae stimulate leaf litter decomposition in detritus-based headwater streams: a case of aquatic priming effect? *Ecology*, 94(7):1604 – 1613, 2013.

T.B. De Gregoris, N. Aldred, A.S. Clare, and J.G. Burgess. Improvement of phylum-and class-specific primers for real-time PCR quantification of bacterial taxa. *Journal of Microbiological Methods*, 86(3):351–356, September 2011.

B.A. Dehority. Evaluation of subsampling and fixation procedures used for counting rumen protozoa. *Applied and Environmental Microbiology*, 48(1):182–185, 1984.

E.F DeLong, G.S Wickham, and N.R Pace. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science*, 243(4896):1360–1363, 1989.

T.Z DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E.L Brodie, T.H. Keller, and et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069–5072, 2006.

A. Deschesne, C. Pallud, and G.L Grundmann. Spatial distribution of bacteria at the microscale in soil. In *The spatial distribution of microbes in the environment*, chapter 4, pages 87–107. Springer, 2007.

E. Devadhanam, Joubert and B. Pillay. Visualisation of the microbial colonisation of a slow sand filter using an environmental scanning electron microscope. *Electronic Journal of Biotechnology*, 11(2):119–125, 2008.

N. Dias, R.A Mortara, and N. Lima. Morphological and physiological changes in *Tetrahymena pyriformis* for the in vitro cytotoxicity assessment of Triton X-100. *Toxicology in vitro*, 17(3):357–366, 2003.

E.A. Dinsdale, R.A. Edwards, D. Hall, F. Angly, M. Breitbart, J.M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, et al. Functional metagenomic profiling of nine biomes. *Nature*, 452(7187):629–632, 2008.

CC Dorea. Slow sand filtration pre-treatment with alum is efficient, but is it effective? *Journal of Water, Sanitation and Hygiene for Development*, 3(2):106–111, 2013.

R.E Drenovsky, G.N Elliott, K.J Graham, and K.M Scow. Comparison of phospholipid fatty acid (PLFA) and total soil fatty acid methyl esters (TSFAME) for characterizing soil microbial communities. *Soil Biology and Biochemistry*, 36(11):1793–1800, 2004.

J. Duan, W. Jiang, Z. Cheng, J.J Heikkila, and B.R Glick. The complete genome sequence of the plant growth-promoting bacterium *Pseudomonas sp.* UW4. *PloS one*, 8(3):e58640, 2013.

M.G Dumont and J.C Murrell. Stable isotope probinglinking microbial identity to function. *Nature Reviews Microbiology*, 3(6):499–504, 2005.

A. Duncan. *Slow sand filtration: recent developments in water treatment technology*, volume 1, chapter 3. Biological aspects: The ecology of slow sand filters, pages 163–179. John Wiley and Sons, 1988.

M. Dunthorn, J. Klier, J. Bunge, and T. Stoeck. Comparing the Hyper-Variable V4 and V9 Regions of the Small Subunit rDNA for Assessment of Ciliate Environmental Diversity. *Journal of Eukaryotic Microbiology*, 59(2):185–187, 2012.

T. Durfee, R. Nelson, S. Baldwin, G. Plunkett, V. Burland, B. Mau, and et al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *Journal of bacteriology*, 190(7):2597–2606, 2008.

T.T Eighmy, R. M. Collins, S.K. Spanos, and J. Fenstermacher. Microbial populations, activities and carbon metabolism in slow sand filters. *Water Research*, 26(10):1319–1328, 1992.

M. Eisenstein. Oxford Nanopore announcement sets sequencing sector abuzz. *Nature biotechnology*, 30(4):295–296, 2012.

S. El Fantroussi and S.N Agathos. Is bioaugmentation a feasible strategy for pollutant removal and site remediation? *Current opinion in microbiology*, 8(3):268–275, 2005.

C.A Elkins and L.B Mullis. Mammalian steroid hormones are substrates for the major RND- and MFS-type tripartite multidrug efflux pumps of *Escherichia coli*. *Journal of bacteriology*, 188(3):1191–1195, 2006.

S. Elleuche and S. Pöggeler. Carbonic anhydrases in fungi. *Microbiology*, 156(1):23–29, 2010.

MA Elliott, CE Stauber, F. Koksal, FA DiGiano, and MD Sobsey. Reductions of e. coli, echovirus type 12 and bacteriophages in an intermittently operated household-scale slow sand filter. *Water research*, 42(10-11):2662–2670, 2008.

KV Ellis and WE Wood. Slow sand filtration. *Critical Reviews in Environmental Science and Technology*, 15(4):315–354, 1985.

European Commission Scientific Committee on Health and Environmental Risks. Opinion on: Chemicals and the water framework directive: environmental quality standards: ethinylestradiol (EE2), 2011.

European Union Council Directive. 98/83/EC of 3 November 1998 relative to the quality of drinking water. *Official Journal of the European Communities*, pages 330–332, 1998.

I.R Falconer, H.F Chapman, M.R Moore, and G. Ranmuthugala. Endocrine-disrupting compounds: A review of their challenge to sustainable and safe water supply and water reuse. *Environmental toxicology*, 21(2):181–191, 2006.

H.H.P Fang, D. Liang, and T. Zhang. Aerobic degradation of diethyl phthalate by *Sphingomonas* sp. *Bioresource Technology*, 98(3):717–720, 2007.

Y. Feng, X. Lin, J. Zhu, and Z. Jia. A phototrophy-driven microbial food web in a rice soil. *Journal of Soils and Sediments*, 11(2):301–311, 2011.

P. Fey, A.S Kowal, P. Gaudet, K.E Pilcher, and R.L Chisholm. Protocols for growth and development of *Dictyostelium discoideum*. *Nature protocols*, 2(6):1307–1316, 2007.

N. Fierer, J.A. Jackson, R. Vilgalys, and R.B. Jackson. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Applied and Environmental Microbiology*, 71(7):4117, 2005.

D.A. Fike, C.L. Gammon, W. Ziebis, and V.J Orphan. Micron-scale mapping of sulfur cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-FISH approach. *ISME*, 2(7):749–759, 2008.

D.D Fine, G.P Breidenbach, T.L Price, and S.R Hutchins. Quantitation of estrogens in ground water and swine lagoon samples using solid-phase extraction, pentafluorobenzyl/trimethylsilyl derivatizations and gas chromatography–negative ion chemical ionization tandem mass spectrometry. *Journal of Chromatography A*, 1017(1):167–185, 2003.

R.A Fleck, E.E Benson, D.H Bremner, and J.G Day. Studies of free radical-mediated cryoinjury in the unicellular green alga *Euglena gracilis* using a non-destructive hydroxyl radical assay: a novel approach for developing protistan cryopreservation strategies. *Free Radical Research*, 32(2):157–170, 2000.

D. Fogel, J. Isaac-Renton, R. Guasparini, W. Moorehead, and J. Ongerth. Removing giardia and cryptosporidium by slow sand filtration. *American water works association*, 85:77–84, 1993.

W.M Freeman, S.J Walker, and K.E Vrana. Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26:112–125, 1999.

M.C Frith, M. Hamada, and P. Horton. Parameters for accurate genome alignment. *BMC bioinformatics*, 11(1):80, 2010.

J.L.R Gallego, J. Loredo, J.F Llamas, F. Vázquez, and J. Sánchez. Bioremediation of diesel-contaminated soils: Evaluation of potential in situ techniques by study of bacterial degradation. *Biodegradation*, 12(5):325–335, 2001.

V. Gallego, C. Sánchez-Porro, M. García, and A. Ventosa. Massilia aurea sp. nov., isolated from drinking water. *International journal of systematic and evolutionary microbiology*, 56(10):2449–2453, 2006.

E. García-Gómez, B. González-Pedrajo, and I. Camacho-Arroyo. Role of sex steroid hormones in bacterial-host interactions. *BioMed research international*, 2013, 2012.

L Garrido, O Sánchez, I Ferrera, N Tomàs, and J Mas. Dynamics of microbial diversity profiles in waters of different qualities. Approximation to an ecological quality indicator. *The Science of the total environment*, 468:1154–1161, 2014.

G.M Garrity, J.A Bell, and T.G Lilburn. *Taxonomic outline of the prokaryotes. Bergey's manual of systematic bacteriology*. Springer-Verlag, New York, 2nd edition, 2004.

A. Gerhardt, A. Ud-Daula, and K-W. Schramm. *Tetrahymena* spp. (Protista, Ciliophora) as test species in rapid multilevel ecotoxicity tests. *Acta Protozoologica*, 49(4):271–280, 2010.

F.B Gich, E. Amer, J.B Figueras, Charles A., M.D Balaguer, and M. Poch. Assessment of microbial community structure changes by amplified ribosomal dna restriction analysis (ardra). *International Microbiology*, 3(2):103–106, 2010.

R Gimbel and R Collins, editors. *Advances in slow sand and alternative biological filtration*. TJ International Ltd., Cornwall, UK, 2006.

J.M Goldman, S.C Laws, S.K Balchak, R.L Cooper, and R.J Kavlock. Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. *CRC Critical Reviews in Toxicology*, 30(2):135–196, 2000.

D.W. Graham, C.W. Knapp, E.S. Van Vleck, K. Bloor, T.B. Lane, and C.E. Graham. Experimental demonstration of chaotic instability in biological nitrification. *The ISME journal*, 1(5):385–393, 2007.

N.J.D. Graham. Removal of humic substances by oxidation/biofiltration processes: A review. *Water Science and Technology*, 40(9):141–148, 1999.

ND Gray, R. Howarth, RW Pickup, J.G. Jones, and IM Head. Use of combined microautoradiography and fluorescence in situ hybridization to determine carbon metabolism in mixed natural communities of uncultured bacteria from the genus achromatium. *Applied and environmental microbiology*, 66(10):4518 – 4522, 2000.

E.A. Grice, H.H. Kong, G. Renaud, A.C. Young, G.G. Bouffard, R.W. Blakesley, T.G. Wolfsberg, M.L. Turner, and J.A. Segre. A diversity profile of the human skin microbiota. *Genome research*, 18(7):1043 – 1050, 2008.

S. Grobe, H. Petry-Hansen, M. Uhl, and H.C. Flemming. *Slow sand filtration: recent developments in water treatment technology*, volume 3, chapter Retention of hygienically relevant microorganisms from storm water effluents by slow sand filters, pages 474–479. John Wiley and Sons, 2006.

J. Haarhoff and JL Cleasby. Biological and physical mechanisms in slow sand filtration. In *Slow Sand Filtration*, pages 19–68. American Society of Civil Engineers, New York, 1991.

SJ Haig, G. Collins, RL Davies, CC Dorea, and C. Quince. Biological aspects of slow sand filtration: past, present and future. *Water Science & Technology: Water Supply*, 11(4): 468–472, 2011.

SJ Haig, C. Quince, R.L. Davies, C.C Dorea, and G Collins. Validating the Performance and Microbial Community of Laboratory-Scale Slow Sand Filters with respect to Full-Scale Industrial Filters. *Water Research*, 61:141–151, 2014.

J. Harte and A. Kinzig. Mutualism and competition between plants and decomposers: implications for nutrient allocation in ecosystems. *American Naturalist*, pages 829–846, 1993.

T. Hashimoto and T. Murakami. Removal and degradation characteristics of natural and synthetic estrogens by activated sludge in batch experiments. *Water research*, 43(3):573–582, 2009.

T. Hashimoto, K. Onda, T. Morita, B.S Luxmy, K. Tada, A. Miya, and et al. Contribution of the estrogen-degrading bacterium *Novosphingobium sp.* strain JEM-1 to estrogen removal in wastewater treatment. *Journal of Environmental Engineering*, 136(9):890–896, 2009.

S.A Hashsham, A.S Fernandez, S.L Dollhopf, F.B Dazzo, R.F Hickey, J.M Tiedje, and et al. Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. *Applied and Environmental Microbiology*, 66(9):4050–4057, 2000.

R. Hatzenpichler, E. Lebedeva, E. Spieck, K. Stoecker, A. Richter, H. Daims, and et al. A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proceedings of the National Academy of Sciences*, 105(6):2134–2139, 2008.

W.A.M. Hijnen, Y.J. Dullemont, J.F. Schijven, A.J. Hanzens-Brouwer, M. Rosielle, and G. Medema. Removal and fate of *Cryptosporidium parvum, Clostridium perfringens* and small-sized centric diatoms (*Stephanodiscus hantzschii*) in slow sand filters. *Water research*, 41(10):2151–2162, 2007.

H. Holtan, L. Kamp-Nielsen, and AO Stuanes. Phosphorus in soil, water and sediment: an overview. *Hydrobiologia*, 170(1):19–34, 1988.

A. Hu, J. He, K.H. Chu, and C.P. Yu. Genome sequence of the 17β-estradiol-utilizing bacterium sphingomonas strain kc8. *Journal of bacteriology*, 193(16):4266–4267, 2011.

W.E. Huang, K. Stoecker, R. Griffiths, L. Newbold, H. Daims, and et al Whiteley. Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Environmental Microbiology*, 9 (8):1878–1889, 2007.

T. Huber, G. Faulkner, and P. Hugenholtz. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, 20(14):2317–2319, 2004.

J Huggett, K Dheda, S Bustin, and A Zumla. Real-time rt-pcr normalisation; strategies and considerations. *Genes and immunity*, 6(4):279–284, 2005.

L. Huisman, WE Wood, and (Geneva) World Health Organization. *Slow sand filtration*. World Health Organisation Geneva, 1974.

M.D Hunter and P.W Price. Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology*, 73(3):723–732, 1992.

P.J. Hunter, L.A. Calvo-Bado, N.R. Parsons, T.R. Pettitt, G.M. Petch, E. Shaw, and et al. Variation in microbial communities colonizing horticultural slow sand filter beds: implications for filter function. *Irrigation Science*, pages 1–12, 2012.

P.J. Hurd and C.J. Nelson. Advantages of next-generation sequencing v.s. the microarray in epigenetic research. *Briefings in Functional Genomics and Proteomics*, 8:174–183, 2009.

D.H Huson, S. Mitra, H-J Ruscheweyh, N. Weber, and S.C Schuster. Integrative analysis of environmental sequences using MEGAN4. *Genome research*, 21(9):1552–1560, 2011.

S.R Hutchins, M.V White, F.M Hudson, and D.D Fine. Analysis of lagoon samples from different concentrated animal feeding operations for estrogens and estrogen conjugates. *Environmental Science & Technology*, 41(3):738–744, 2007.

L. Iasur-Kruh, Y. Hadar, and D. Minz. Isolation and bioaugmentation of an estradiol-degrading bacterium and its integration into a mature biofilm. *Applied and Environmental Microbiology*, 77(11):3734–3740, 2011.

Illumina. HiSeq Systems Comparison. http://www.illumina.com/systems/hiseq_comparison.ilmn, October 2013. URL $http://www.illumina.com/systems/hiseq_comparison.ilmn$.

G. Innerebner, C. Knief, and J.A Vorholt. *Sphingomonas* strains protect *Arabidopsis thaliana* against leaf pathogenic *Pseudomonas syringae* in a controlled model system. *Applied and Environmental Microbiology*, 77(10):3202–3210, 2011.

Invitrogen Corporation. Real-Time PCR: From Theory To Practice. Hand book, 2008.

Gohargani Jafar and Ghasemi Hamzeh. Ecotoxicity of Nanomaterials in Soil. *Annals of Biological Research*, 4(1):86–92, 2013.

N. Jehmlich, F. Schmidt, M. Von-Bergen, H-H. Richnow, and C. Vogt. Protein-based stable isotope probing reveals active species within anoxic mixed cultures. *ISME*, 2(11):1122–1133, 2008.

CO Jeon, W Park, P Padmanabhan, C DeRito, JR Snape, and EL Madsen. Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proceedings of the National Academy of Sciences*, 100(23): 13591–13596, 2003.

S. Jobling, R. Williams, A. Johnson, A. Taylor, M. Gross-Sorokin, M. Nolan, and et al. Predicted exposures to steroid estrogens in UK rivers correlate with widespread sexual disruption in wild fish populations. *Environmental health perspectives*, 114(S1):32, 2006.

M Johnson, DD Ratnayaka, and MJ Brandt. *Twort's Water Supply*, chapter 8. Butterworth-Heinemann, Burlington-Oxford, 2009.

K.E Judd, B.C Crump, and .W Kling. Bacterial responses in activity and community composition to photo-oxidation of dissolved organic matter from soil and surface waters. *Aquatic Sciences*, 69(1):96–107, 2007.

K. Kadir and K.L Nelson. Sunlight mediated inactivation mechanisms of *Enterococcus faecalis* and *Escherichia coli* in clear water versus waste stabilization pond water. *Water research*, 50(1):307–317, 2014.

H. Kavak. Some Biological Parmaters In *Sordaria Fimicola*. *Pakistan Journal of Botany*, 44 (3):1079–1082, 2012.

I.UH Khan, V. Gannon, R. Kent, W. Koning, D.R Lapen, J. Miller, and et al. Development of a rapid quantitative PCR assay for direct detection and quantification of culturable and non-culturable *Escherichia coli* from agriculture watersheds. *Journal of Microbiological Methods*, 69(3):480–488, 2007.

H. Kim, P.R. Jaffé, and L.Y. Young. Simulating biodegradation of toluene in sand column experiments at the macroscopic and pore-level scale for aerobic and denitrifying conditions. *Advances in water resources*, 27(4):335–348, 2004.

J.L Kirk, L.A Beaudette, M. Hart, P. Moutoglis, J.N Klironomos, H. Lee, and et al. Methods of studying soil microbial diversity. *Journal of Microbiological Methods*, 58(2):169–188, 2004.

A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research*, 41(1):e1, 2013.

J. Kohidai, L. Katona and G. Csaba. Effects of steroid hormones on five functional parameters of *Tetrahymena*: evolutionary conclusions. *Cell Biochemistry and Function*, 21(1): 19–26, 2003.

Y. Koizumi, H. Kojima, and M. Fukui. Characterization of depth-related microbial community structure in lake sediment by DGGE of amplified 16S rDNA and reversely transcribed 16S rRNA fragments. *FEMS Microbiology Ecology*, 46(2):147–157, 2003.

D.W Kolpin, E.T Furlong, M.T Meyer, E.M Thurman, S.D Zaugg, and L. et al. Barber. Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999-2000: A national reconnaissance. *Environmental Science & Technology*, 36(6): 1202–1211, 2002.

A Konopka. *Encyclopedia of microbiology*, volume 1, chapter Ecology, Microbial, pages 91–106. Academic Press, 2009.

K. Koskinen, J. Hultman, L. Paulin, P. Auvinen, and H. Kankaanpää. Spatially differing bacterial communities in water columns of the northern baltic sea. *FEMS microbiology ecology*, 75(1):99–110, 2011.

J. Kozdroj and J.D van Elsas. Structural diversity of microorganisms in chemically perturbed soil assessed by molecular and cytochemical approaches. *Journal of Microbiological Methods*, 43(3):197–212, 2001.

H.M. Kuch and K. Ballschmiter. Determination of endocrine-disrupting phenolic compounds and estrogens in surface and drinking water by HRGC-(NCI)-MS in the picogram per liter range. *Environmental science & technology*, 35(15):3201–3206, 2001.

M. Kuster, M. López de Alda, M.D Hernando, M. Petrovic, J. Martín-Alonso, and D. Barceló. Analysis and occurrence of pharmaceuticals, estrogens, progestogens and polar pesticides in sewage treatment plant effluents, river water and drinking water in the Llobregat river basin (Barcelona, Spain). *Journal of hydrology*, 358:112–123, 2008.

M.M. Kuypers. Sizing up the uncultivated majority. *Science*, 317:1510–1511, 2007.

S. Langenheder, E.S Lindström, and L.J Tranvik. Structure and function of bacterial communities emerging from different sources under identical conditions. *Applied and Environmental Microbiology*, 72(1):212–220, 2006.

J. Langlet, L. Ogorzaly, J-C. Schrotter, C. Machinal, F. Gaboriaud, J.FL Duval, and et al. Efficiency of MS2 phage and Qβ phage removal by membrane filtration in water treatment: Applicability of real-time RT-PCR method. *Journal of membrane science*, 326(1): 111–116, 2009.

WJ Langston, GR Burt, BS Chesman, and CH Vane. Partitioning, bioavailability and effects of oestrogens and xeno-oestrogens in the aquatic environment. *Journal of the Marine Biological Association of the United Kingdom*, 85(1):1–31, 2005.

P.A Lara-Martín, A. Gómez-Parra, T. Köchling, J. Sanz, R. Amils, and E. González-Mazo. Anaerobic degradation of linear alkylbenzene sulfonates in coastal marine sediments. *Environmental Science & Technology*, 41(10):3573–3579, 2007.

A.I. Laskin, J.W. Bennett, and G.M. Gadd. *Advances in applied microbiology*, volume 57. Academic Press, San Diego, 2004.

K. Lautenschlager, N. Boon, Y. Wang, T. Egli, and F. Hammes. Overnight stagnation of drinking water in household taps induces microbial growth and changes in community composition. *Water Research*, 44:4868–4877, 2010.

M. Le Noir, A-S. Lepeuple, B. Guieysse, and B. Mattiasson. Selective removal of 17βestradiol at trace concentration using a molecularly imprinted polymer. *Water Research*, 41(12):2825–2831, 2007.

E. Lee and L.R Oki. Slow sand filters effectively reduce *Phytophthora* after a pathogen switch from *Fusarium* and a simulated pump failure. *Water research*, 47(14):5121–5129, 2013.

X. Lin, J. McKinley, C.T Resch, R. Kaluzny, C.L Lauber, and et al. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *ISME*, 6(9): 1665–1676, 2012.

R.L Lindeman. The trophic-dynamic aspect of ecology. *Ecology*, 23(4):399–417, 1942.

W. Liu, D.W. Au, D.M. Anderson, P.K. Lam, and R.S. Wu. Effects of nutrients, salinity, pH and light: dark cycle on the production of reactive oxygen species in the alga *Chattonella marina*. *Journal of Experimental Marine Biology and Ecology*, 346(1):76–86, 2007.

Z-H. Liu, Y. Kanjo, and S. Mizutani. Removal mechanisms for endocrine disrupting compounds (EDCs) in wastewater treatmentphysical means, biodegradation, and chemical advanced oxidation: a review. *Science of the Total Environment*, 407(2):731–748, 2009.

Z-H. Liu, J.A. Ogejo, A. Pruden, and K.F Knowlton. Occurrence, fate and removal of synthetic oral contraceptives (SOCs) in the natural environment: A review. *Science of the Total Environment*, 409(24):5149–5161, 2011.

L Lloret, G Eibes, G Feijoo, MT Moreira, and JM Lema. Degradation of estrogens by laccase from *Myceliophthora thermophila* in fed-batch and enzymatic membrane reactors. *Journal of hazardous materials*, 213:175–183, 2012.

B. Lloyd. *The functional microbial ecology of slow, sand filters.* PhD thesis, University of Surrey, 1974.

B. Lloyd. The construction of a sand profile sampler: its use in the study of the vorticella populations and the general interstitial microfauna of slow sand filters. *Water Research*, 7 (7):963–973, 1973.

BJ Lloyd. The significance of protozoal predation and adsorption for the removal of bacteria by slow sand filtration. In *Advances in Slow Sand and Alternative Biological Filtration*, pages 129–137. John Wiley & Sons, UK, 1996.

M.M Lorah and M.A Voytek. Degradation of 1, 1, 2, 2-tetrachloroethane and accumulation of vinyl chloride in wetland sediment microcosms and in situ porewater: biogeochemical controls and associations with microbial communities. *Journal of contaminant hydrology*, 70(1):117–145, 2004.

R.A. Luna, L.R. Fasciano, S.C. Jones, B.L. Boyanton Jr, T.T. Ton, and J. Versalovic. Dna pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *Journal of clinical microbiology*, 45(9):2985 – 2992, 2007.

M.T Madigan, J.M Martinko, D. Stahl, and D.P Clark. *Brock Biology of Microorganisms 13th edition*, volume 13. Pearson Education, 2011.

S. Malik, M. Beer, M. Megharaj, and R. Naidu. The use of molecular techniques to characterize the microbial communities in contaminated soil and water. *Environment international*, 34(2):265–276, 2008.

J. Marley, M. Lu, and C. Bracken. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20(1):71–75, 2001.

KC Marshall. Biofilms: An overview of bacterial adhesion, activity, and control at surfaces. control of biofilm formation awaits the development of a method to prevent bacterial adhesion. *ASM American Society for Microbiology News*, 58(4):202–207, 1992.

A. Martín-González, S. Díaz, S. Borniquel, A. Gallego, and J.C. Gutiérrez. Cytotoxicity and bioaccumulation of heavy metals by ciliated protozoa isolated from urban wastewater treatment plants. *Research in Microbiology*, 157(2):108–118, 2006.

A Massol-Deya, R Weller, L Rios-Hernandez, JZ Zhou, RF Hickey, and JM Tiedje. Succession and convergence of biofilm communities in fixed-film reactors treating aromatic hydrocarbons in groundwater. *Applied & Environmental Microbiology*, 63(1):270–276, 1997.

P Matthiessen, D Arnold, AC Johnson, TJ Pepper, TG Pottinger, and KGT Pulman. Contamination of headwater streams in the United Kingdom by oestrogenic hormones from livestock farms. *Science of the Total Environment*, 367(2):616–630, 2006.

HE Maynard, SK Ouki, and SC Williams. Tertiary lagoons: a review of removal mechanisms and performance. *Water Research*, 33(1):1–13, 1999.

J McCambridge and TA McMeekin. Relative effects of bacterial and protozoan predators on survival of *Escherichia coli* in estuarine water samples. *Applied and Environmental Microbiology*, 40(5):907–911, 1980.

L.K. McConnell, R.C. Sims, and B.B. Barnett. Reovirus removal and inactivation by slow-rate sand filtration. *Applied and environmental microbiology*, 48(4):818–825, 1984.

S. McHugh, M. Carton, G. Collins, and V. O'Flaherty. Reactor performance and microbial community dynamics during anaerobic biological treatment of wastewaters at 16–37 c. *FEMS microbiology ecology*, 48(3):369–378, 2004. ISSN 1574-6941.

J.E McMahon and S.K. Price. Water and energy interactions. *Annual Review of Environment and Resources*, 36:163–191, 2011.

N Mezrioui, B Baleux, and M Troussellier. A microcosm study of the survival of *Escherichia coli* and *Salmonella typhimurium* in brackish water. *Water Research*, 29(2): 459–465, 1995.

J. Mikola and H. Setälä. No evidence of trophic cascades in an experimental microbial-based soil food web. *Ecology*, 79(1):153–164, 1998.

S. Mishra, J. Jyot, R.C. Kuhad, and B. Lal. Evaluation of inoculum addition to stimulate in situ bioremediation of oily-sludge-contaminated soil. *Applied and Environmental Microbiology*, 67(4):1675–1681, 2001.

S.E. Morales and W.E. Holben. Empirical testing of 16s rrna gene pcr primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. *Applied and environmental microbiology*, 75(9):2677–2683, 2009.

A.M. Moreno, C. Matz, S. Kjelleberg, and M. Manefield. Identification of ciliate grazers of autotrophic bacteria in ammonia-oxidizing activated sludge by rna stable isotope probing. *Applied and environmental microbiology*, 76(7):2203–2211, 2010.

G. Muyzer and K. Smalla. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek*, 73(1):127–141, 1998.

G. Muyzer, E.C. de Waal, and A.G. Uitterlinden. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, 59(3):695, 1993.

AH Nasr, Inaam A Bekheet, and RK Ibrahim. The effect of different nitrogen and carbon sources on amino acid synthesis in Ulva, Dictyota and Pterocladia. *Hydrobiologia*, 31(1): 7–16, 1968.

K. Nava-Castro, R. Hernández-Bello, S. Muñiz-Hernández, I. Camacho-Arroyo, and J. Morales-Montor. Sex steroids, immune system, and parasitic infections: facts and hypotheses. *Annals of the New York Academy of Sciences*, 1262(1):16–26, 2012.

T.Y. Nechitaylo, M.M. Yakimov, M. Godinho, K.N. Timmis, E. Belogolova, B.A. Byzov, and et al. Effect of the earthworms Lumbricus terrestris and Aporrectodea caliginosa on bacterial diversity in soil. *Microbial Ecology*, 59(3):574–587, 2010.

AM Nedelcu, IH Miles, AM Fagir, and K Karol. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *Journal of Evolutionary Biology*, 21(6):1852–1860, 2008.

J.D. Neufeld, J. Vohra, M.G. Dumont, T. Lueders, M. Manefield, M.W. Friedrich, and J.C. Murrell. Dna stable-isotope probing. *Nature protocols*, 2(4):860–866, 2007.

L. Nováková and H. Vlčková. A review of current trends and advances in modern bio-analytical methods: chromatography and sample preparation. *Analytica Chimica Acta*, 656(1):8–35, 2009.

N Nunan, K Wu, IM Young, JW Crawford, and K Ritz. *In situ* spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microbial Ecology*, 44 (4):296–305, 2002.

N. Nunan, K. Wu, I.M Young, J.W. Crawford, and K. Ritz. Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiology Ecology*, 44(2):203–215, 2003.

K.A. O'Farrell and P.H. Janssen. Detection of verrucomicrobia in a pasture soil by pcr-mediated amplification of 16s rrna genes. *Applied and environmental microbiology*, 65 (9):4280–4284, 1999.

Y. Ohtsubo, F. Maruyama, H. Mitsui, Y. Nagata, and M. Tsuda. Complete genome sequence of *Acidovorax sp.* strain KKS102, a polychlorinated-biphenyl degrader. *Journal of bacteriology*, 194(24):6970–6971, 2012.

CSP Ojha and NJD Graham. Computer-aided simulation of slow sand filter performance. *Water research*, 28(5):1025–1030, 1994.

J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R. B. O'Hara, and et al. *Vegan: Community Ecology Package*, 2012. URL `http://CRAN.R-project.org/package=vegan`. R package version 2.0-3.

R. Overbeek, T. Begley, R.M Butler, J.V Choudhuri, H. Chuang, M. Cohoon, and et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17):5691–5702, 2005.

S.O Oyola, T.D Otto, Y. Gu, G. Maslen, M. Manske, and S. et al. Campino. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC genomics*, 13(1):1, 2012.

ML Pace and JJ Cole. Comparative and experimental approaches to top-down and bottom-up regulation of bacteria. *Microbial Ecology*, 28(2):181–193, 1994.

N.R. Pace. A molecular view of microbial diversity and the biosphere. *Science*, 276(5313): 734–740, 1997.

J.D. Parry. Protozoan grazing of freshwater biofilms. *Advances in applied microbiology*, 54: 167–196, 2004.

B. Pauwels, K. Wille, H. Noppe, H. De Brabander, T. Van de Wiele, W. Verstraete, and N. Boon. 17α-ethinylestradiol cometabolism by bacteria degrading estrone, 17β-estradiol and estriol. *Biodegradation*, 19(5):683–693, 2008.

V.J. Pereira, D. Fernandes, G. Carvalho, M.J. Benoliel, M.V. San Romão, and M.T. Barreto Crespo. Assessment of the presence and dynamics of fungi in drinking water sources using cultural and molecular methods. *Water Research*, 44(17):4850 – 4859, 2010.

A. Pernthaler. Identification of Environmental Microorganisms by Fluorescence in-situ Hybridization. In Kenneth N. Timmis, editor, *Handbook of Hydrocarbon and Lipid Microbiology*, chapter 59, pages 4127–4135. Springer Berlin Heidelberg, 2010.

J. Pernthaler. Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology*, 3(7):537–546, 2005.

N.N. Perreault, D.T. Andersen, W.H. Pollard, C.W. Greer, and L.G. Whyte. Characterization of the prokaryotic diversity in cold saline perennial springs of the canadian high arctic. *Applied and environmental microbiology*, 73(5):1532, 2007.

J.F. Petrosino, S. Highlander, R.A.L. Richard, and A.G.J. Versalovic. Methagenomic pyrosequrincing and microbial identification. *Clinical chemistry*, 55(5):856–866, 2009.

H. Petry-Hansen, H. Steele, M. Grooters, J. Wingender, and H.C. Flemming. *Slow sand filtration: recent developments in water treatment technology*, volume 3, chapter 17. Feacal contamination indicator organisms in slow sand filters, pages 143–151. John Wiley and Sons, 2006.

M.W. Pfaffl. A new mathematical model for relative quantification in real-time rt–pcr. *Nucleic acids research*, 29(9):e45–e45, 2001.

L. Philippot, D. Bru, N. Saby, J. Čuhel, D. Arrouays, M. Šimek, and S. Hallin. Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environmental Microbiology*, 11(12):3096–3104, 2009.

I. Pizzetti, B.M Fuchs, G. Gerdts, A. Wichels, K.H Wiltshire, and R. Amann. Temporal variability of coastal *Planctomycetes* clades at Kabeltonne station, North sea. *Applied and Environmental Microbiology*, 77(14):5009–5017, 2011.

H. Ploug, B. Adam, N. Musat, T. Kalvelage, G. Lavik, D. Wolf-Gladrow, and et al. Carbon, nitrogen and O2 fluxes associated with the cyanobacterium Nodularia spumigena in the Baltic Sea. *The ISME Journal*, 5(9):1549–1558, 2011.

SFB Poynter and JS Slade. The removal of viruses by slow sand filtration. In *Eighth International Conference on Water Pollution Research, Sydney, Australia, 17-22 October, 1976*, pages 75 – 88. Pergamon Press, 1978.

K.J. Purdy, P.J. Hurd, J. Moya-Larano, M. Trimmer, Brian B. Oakley, and Guy Woodward. *Integrative Ecology: From Molecules to Ecosystems*, volume 43, chapter System biology for ecology: from molecules to ecosystems, pages 88–149. Academic Press, 2010.

C. Quince, A. Lanzén, T.P. Curtis, R.J. Davenport, N. Hall, I.M. Head, L.F. Read, and W.T. Sloan. Accurate determination of microbial diversity from 454 pyrosequencing data. *nature methods*, 6(9):639–641, 2009.

C. Quince, A. Lanzen, R.J. Davenport, and P.J. Turnbaugh. Removing Noise from Pyrosequenced Amplicons. *BMC bioinformatics*, 12(1):38, 2011.

JB Quintana, J. Carpinteiro, I. Rodrıguez, RA Lorenzo, AM Carro, and R. Cela. Determination of natural and synthetic estrogens in water by gas chromatography with mass spectrometric detection. *Journal of chromatography A*, 1024(1):177–185, 2004.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

J-B Ramond, P.J Welz, M.I Tuffin, S. G Burton, and D.A Cowan. Assessment of temporal and spatial evolution of bacterial communities in a biological sand filter mesocosm treating winery wastewater. *Journal of applied microbiology*, 115(1):91–101, 2013.

MS Rappé and SJ Giovannoni. The uncultured microbial majority. *Annual Reviews in Microbiology*, 57:87–90, 2003.

Y.X. Ren, K. Nakano, M. Nomura, N. Chiba, and O. Nishimura. Effects of bacterial activity on estrogen removal in nitrifying activated sludge. *Water research*, 41(14):3089–3096, 2007.

R.E. Richardson. Genomic insights into organohalide respiration. *Current opinion in biotechnology*, 24(3):498–505, 2013.

S. Ripp and R.V. Miller. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology*, 143(6):2065–2070, 1997.

Roche. Performance Information: GS FLX+ System, October 2013. URL $http : //454.com/products/gs - flx - system/$.

H. Roh and K-H. Chu. Effects of solids retention time on the performance of bioreactors bioaugmented with a 17β-estradiol-utilizing bacterium, *Sphingomonas* strain KC8. *Chemosphere*, 84(2):227–233, 2011.

S.W. Roh, G.C.J. Abell, K-H. Kim, Y-D. Nam, and J-W. Bae. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in biotechnology*, 28(6):291–299, 2010.

W.F.M Röling, M. Ferrer, and P.N Golyshin. Systems approaches to microbial communities and their functioning. *Current opinion in biotechnology*, 21(4):532–538, 2010.

T. Romeo. *Bacterial biofilms*. Springer Verlag, Berlin, 2008.

S.J. Rooklidge, E.R. Burns, and J.P. Bolte. Modeling antimicrobial contaminant removal in slow sand filtration. *Water research*, 39(2-3):331–339, 2005.

T.M Rose, J.G Henikoff, and S. Henikoff. CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR primer design. *Nucleic Acids Research*, 31(13):3763–3766, 2003.

A.D Rosemond, C.M Pringle, A. Ramírez, and M.J Paul. A test of top-down and bottom-up control in a detritus-based food web. *Ecology*, 82(8):2279–2293, 2001.

K. Röske, R. Sachse, C. Scheerer, and I. Röske. Microbial diversity and composition of the sediment in the drinking water reservoir Saidenbach (Saxonia, Germany). *Systematic and applied microbiology*, 35(1):35–44, 2012.

DB Roszak and RR Colwell. Survival strategies of bacteria in the natural environment. *Microbiology and Molecular Biology Reviews*, 51(3):365–379, 1987.

H Runes, J Jenkins, and P Bottomley. Atrazine degradation by bioaugmented sediment from constructed wetlands. *Applied microbiology and biotechnology*, 57(3):427–432, 2001.

R. Sadiq, M.A. Al-Zahrani, A.K. Sheikh, T. Husain, and S. Farooq. Performance evaluation of slow sand filters using fuzzy rule-based modelling. *Environmental Modelling & Software*, 19(5):507–515, 2004.

S. Saleh-Lakha, M. Miller, R.G Campbell, K. Schneider, P. Elahimanesh, M.M Hart, and J.T Trevors. Microbial gene expression in soil: methods, applications and challenges. *Journal of Microbiological Methods*, 63(1):1–19, 2005.

A. Schippers, L.N. Neretin, J. Kallmeyer, T.G. Ferdelman, B.A. Cragg, R.J. Parkes, and et al. Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Letters to nature*, 433(24):861–864, 2005.

M. Schirmer, W.T. Sloan, and C. Quince. Benchmarking of viral haplotype reconstruction programmes: an overview of the capacities and limitations of currently available programmes. *Briefings in bioinformatics*, 2012.

P.D. Schloss and J. Handelsman. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, 71(3):1501–1506, 2005.

A. Schmalenberger, F. Schwieger, and C.C. Tebbe. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Applied and Environmental Microbiology*, 67 (8):3557 – 3563, 2001.

M.A. Schneegurt, S.Y. Dore, and C.F. Kulpa Jr. Direct extraction of DNA from soils for studies in microbial ecology. *Current issues in molecular biology*, 5(1):1–8, 2003.

H Segner, K Caroll, M Fenske, CR Janssen, G Maack, D Pascoe, and et al. Identification of endocrine-disrupting effects in aquatic vertebrates and invertebrates: report from the European IDEA project. *Ecotoxicology and environmental safety*, 54(3):302–314, 2003.

K.T Semple and R.B Cain. Biodegradation of phenols by the alga *Ochromonas danica*. *Applied and Environmental Microbiology*, 62(4):1265–1273, 1996.

K. Sen and N.J Ashbolt. *Environmental Microbiology: Current Technology and Water Applications*, chapter 7. Horizon Scientific Press, 2011.

M. Shakya, C. Quince, J.H Campbell, Z.K Yang, C.W Schadt, and M. Podar. Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. *Environmental microbiology*, pages 1–18, 2013.

A. Shareef, M.J. Angove, and J.D. Wells. Optimization of silylation using N-methyl-N-(trimethylsilyl)-trifluoroacetamide,N,O-bis-(trimethylsilyl)-trifluoroacetamide and N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide for the determination of the estrogens estrone and 17α-ethinylestradiol by gas chromatography-mass spectrometry. *Journal of Chromatography A*, 1108(1):121–128, 2006.

A.J. Sharpe, P.A. Quern, and R.A. Currier. A consultant's point of view. In M.R. Collins and N.J.D. Graham, editors, *Slow Sand Filtration*. American Water Works Association, 1994.

D.R. Shelton and J.M. Tiedje. General method for determining anaerobic biodegradation potential. *Applied and Environmental Microbiology*, 47(4):850–857, 1984.

J. Shendure and H. Ji. Next-generation dna sequencing. *Nature biotechnology*, 26(10): 1135–1145, 2008.

J. Shi, S. Fujisawa, S. Nakai, and M. Hosomi. Biodegradation of natural and synthetic estrogens by nitrifying activated sludge and ammonia-oxidizing bacterium *Nitrosomonas europaea*. *Water Research*, 38(9):2323–2330, 2004.

T Shi, J.K Fredrickson, and D.L Balkwill. Biodegradation of polycyclic aromatic hydrocarbons by *Sphingomonas* strains isolated from the terrestrial subsurface. *Journal of Industrial Microbiology and Biotechnology*, 26(5):283–289, 2001.

X. Shi, X. Liu, G. Liu, Z. Sun, and H. Xu. An approach to analyze spatial patterns of protozoan communities for assessing water quality in the Hangzhou section of Jing-Hang Grand Canal in China. *Environmental Science and Pollution Research*, 19(3):739–747, 2012.

S. Shokralla, J.L Spall, J.F Gibson, and M. Hajibabaei. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8):1794–1805, 2012.

L.S. Shore, Y. Kapulnik, M. Gurevich, S. Wininger, H. Badamy, and M. Shemesh. Induction of phytoestrogen production in *Medicago sativa* leaves by irrigation with sewage water. *Environmental and Experimental Botany*, 35(3):363–369, 1995.

M.W Silby, A.M Cerdeño-Tárraga, G.S Vernikos, S.R Giddens, R.W Jackson, G.M Preston, X-X Zhang, and et al. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome biology*, 10(5):R51, 2009.

C.P. Silva, M. Otero, and V. Esteves. Processes for the elimination of estrogenic steroid hormones from water: A review. *Environmental Pollution*, 165:38–58, 2012.

M Silva, S. McLellan, and J. Li. *Slow sand filtration: recent developments in water treatment technology*, volume 3, chapter 40. The removal of green fluorescent labelled *Escherichia coli* by pilot scale drinking water biofilters, pages 337–345. John Wiley and Sons, 2006.

C. Simon and R. Daniel. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology*, 77(4):1153–1161, 2011.

FR Slater, CR Johnson, LL Blackall, RG Beiko, and PL Bond. Monitoring associations between clade-level variation, overall community structure and ecosystem function in enhanced biological phosphorus removal (EBPR) systems using T-RFLP. *Water Research*, 44:4908 – 4923, 2010.

W.T. Sloan, S. Woodcock, M. Lunn, I.M Head, and T.P Curtis. Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microbial Ecology*, 53(3):443–455, 2007.

E. Smit, P. Leeflang, and K. Wernars. Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiology Ecology*, 23(3):249–261, 1997.

C.J. Smith and A.M. Osborn. Advantages and limitations of quantitative pcr (q-pcr)-based approaches in microbial ecology. *FEMS microbiology ecology*, 67(1):6–20, 2009.

C.J Smith, D.B Nedwell, L.F Dong, and A.M Osborn. Diversity and abundance of nitrate reductase genes (narG and napA), nitrite reductase genes (nirS and nrfA), and their transcripts in estuarine sediments. *Applied and Environmental Microbiology*, 73(11):3612–3622, 2007.

F.R Spellman. *Water and Wastewater Infrastructure: Energy Efficiency and Sustainability*. Number 9-11. CRC Press Llc, 2013.

D. Stalter, A. Magdeburg, M. Wagner, and J. Oehlmann. Ozonation and activated carbon treatment of sewage effluents: Removal of endocrine activity and cytotoxicity. *Water Research*, 45(3):1015–1024, 2011.

J.F Stolz. *Structure of microbial mats and biofilms*, pages 1–8. Springer, 2000.

K.L Straub, M. Benz, and B. Schink. Iron metabolism in anoxic environments at near neutral pH. *FEMS Microbiology Ecology*, 34(3):181–186, 2001.

W.J. Sul, J. Park, J.F. Quensen, J.L.M. Rodrigues, L. Seliger, T.V. Tsoi, G.J. Zylstra, and et al. DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Applied and Environmental Microbiology*, 75(17):5501–5506, 2009.

S. Suneethi and K. Joseph. Batch culture enrichment of anammox populations from anaerobic and aerobic seed cultures. *Bioresource technology*, 102:585–591, 2010.

S. Suzuki, N. Ono, C. Furusawa, B-W. Ying, and T. Yomo. Comparison of sequence reads obtained from three next-generation sequencing platforms. *PLoS One*, 6(5):e19534, 2011.

T.A. Ternes, M. Stumpf, J. Mueller, K. Haberer, R.D. Wilken, and M. Servos. Behavior and occurrence of estrogens in municipal sewage treatment plantsi. investigations in germany, canada and brazil. *Science of the Total Environment*, 225(1):81–90, 1999.

P. Thayanukul, K. Zang, T. Janhom, F. Kurisu, I. Kasuga, and H. Furumai. Concentration-dependent response of estrone-degrading bacterial community in activated sludge analyzed by microautoradiography-fluorescence in situ hybridization. *Water Research*, 44: 4878 – 4887, 2010.

J.W. Thomas. Sus'ruta Samhita the Hindu system of medicine according to Sus'ruta. Translated from Sanskrit by Udoy Chand Dutt. *Royal Asiatic Society of Bengal*, 95(45):215–222, 1883.

J.M Tiedje, S. Asuming-Brempong, K. Nüsslein, T.L Marsh, and S.J Flynn. Opening the black box of soil microbial diversity. *Applied Soil Ecology*, 13(2):109–122, 1999.

P.J. Turnbaugh, R.E. Ley, M. Hamady, C.M. Fraser-Liggett, R. Knight, and J.I. Gordon. The human microbiome project. *Nature*, 449(7164):804–810, 2007.

G.W. Tyson, J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, and et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43, 2004.

UK Land Directory. Understanding UK Land Prices, 2014. URL `http://www.uklanddirectory.org.uk/land-prices.asp`.

M. Unger and M.R. Collins. Assessing escherichia coli removal in the schmutzdecke of slow-rate biofilters. *Journal American Water Works Association*, 100(12):60–73, 2008.

United Nations General Assembly. General assembly adopts resolution recognizing access to clean water, sanitation as human right. sixty-fourth general assembly plenary 108th meeting. United Nations, 2010.

B.R. Urbonas. Design of a sand filter for stormwater quality enhancement. *Water Environment Research*, 71(1):102–113, 1999.

U.S. Environmental Protection Agency. First National Expert and Stakeholder Workshop on Water Infrastructure Sustainability and Adaptation to Climate Change; EPA 600/R-09/010, 2009a.

U.S. Environmental Protection Agency. Drinking Water Contaminant Candidate List 3-Final: E9-24287. *Federal Register: U.S. Environmental Protection Agency, Code of Federal Regulations*, 74(194):51850–51862, October 2009b.

U.S. Environmental Protection Agency. Glossary of technical terms, 2009c.

A.M. Valeria, E.J. Ricardo, P. Stephan, and W.D Alberto. Degradation of microcystin-RR by *Sphingomonas* sp. CBA4 isolated from San Roque reservoir (Córdoba–Argentina). *Biodegradation*, 17(5):447–455, 2006.

R.M. Valster, B.A Wullings, R. van den Berg, and D. van der Kooij. Relationships between free-living protozoa, cultivable *Legionella* spp., and water quality characteristics in three drinking water supplies in the Caribbean. *Applied and Environmental Microbiology*, 77 (20):7321–7328, 2011.

JC Van Dijk, J. Oomen, and WHO International Reference Centre for Community Water Supply. *Slow sand filtration for community water supply in developing countries: a design and construction manual*. WHO International Reference Centre for Community Water Supply, 1978.

FXR Van Leeuwen. Safe drinking water: the toxicologist's approach. *Food and Chemical Toxicology*, 38:S51–S58, 2000.

J.T. Visscher, R. Paramasivam, A. Raman, HA Heijnen, et al. Slow sand filtration for community water supply; planning, design, construction, operation and maintenance. Technical Report 24, The Hague, The Netherlands, 1987.

S. Wakelin, D. Page, P. Dillon, P. Pavelic, GCJ Abell, AL Gregg, and et al. Microbial community structure of a slow sand filter schmutzdecke: a phylogenetic snapshot based on rRNA sequence analysis. *Water Science & Technology: Water Supply*, 11(4):426–436, 2011.

G.C.Y. Wang and Y. Wang. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiology*, 142 (5):1107–1114, 1996.

H. Wang, M.A Pryor, M.A. Edwards, J.O Falkinham III, and A. Pruden. Effect of GAC pretreatment and disinfectant on microbial community structure and opportunistic pathogen occurrence. *Water research*, 47(15):5760–5772, 2013.

L. Wang, G-G. Ying, J-L. Zhao, S. Liu, B. Yang, L-J. Zhou, and et al. Assessing estrogenic activity in surface water and sediment of the Liao River system in northeast China using combined chemical and biological tools. *Environmental Pollution*, 159(1):148–156, 2011.

K.P Weber and R.L Legge. *Pathogen removal in constructed wetlands*, chapter 5: Wetlands: Ecology, Conservation and Restoration, pages 176–211. Nova Publishers: New York, 2008.

M.L. Weber-Shirk and R.I. Dick. Bacterivory by a chrysophyte in slow sand filters. *Water research*, 33(3):631–638, 1999.

M.L. Weber-Shirk and R.I. Dick. Biological mechanisms in slow sand filters. *Journal-American Water Works Association*, 89(2):72–83, 1997a.

M.L. Weber-Shirk and R.I. Dick. Physical-chemical mechanisms in slow sand filters. *Journal-American Water Works Association*, 89(1):87–100, 1997b.

P.H.H. Weekers, P.L.E. Bodelier, J.P.H. Wijen, and G.D. Vogels. Effects of grazing by the free-living soil amoebae *Acanthamoeba castellanii, Acanthamoeba polyphaga*, and *Hartmannella vermiformis* on various bacteria. *Applied and Environmental Microbiology*, 59 (7):2317–2319, 1993.

J.J Werner, D. Knights, M.L Garcia, N.B Scalfone, S. Smith, K. Yarasheski, and et al. Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proceedings of the National Academy of Sciences*, 108(10):4158–4163, 2011.

S.G. Wesley and S. Satheesh. Temporal variability of nutrient concentration in marine biofilm developed on acrylic panels. *Journal of Experimental Marine Biology and Ecology*, 379(1-2):1–7, 2009.

W.B Whitman, M. Goodfellow, P. Kämpfer, H-J. Busse, M.E. Trujillo, W. Ludwig, and et al. *Bergeys manual of systematic bacteriology of systematic bacteriology*, volume 5. Springer, 2012.

C. Will, A. Thurmer, A. Wollherr, H. Nacke, N. Herold, M. Schrumpf, Gutknecht, and et al. Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Applied and Environmental Microbiology*, 76(20):6751 – 6759, 2010.

P.A Williams and C.M Ray. Catabolism of Aromatic Compounds by *Acinetobacter*. *Acinetobacter Molecular Biology*, 1:99–117, 2008.

R.J. Williams, A.C. Johnson, J.J.L. Smith, and R. Kanda. Steroid estrogens profiles along river stretches arising from sewage treatment works discharges. *Environmental science & technology*, 37(9):1744–1750, 2003.

C. Winderl, B. Anneser, C. Griebler, R.U. Meckenstock, and T. Lueders. Depth-resolved quantification of anaerobic toluene degraders and aquifer microbial community patterns in distinct redox zones of a tar oil contaminant plume. *Applied and Environmental Microbiology*, 74(3):792–801, 2008.

E. Windle-Taylor. The removal of viruses by slow sand filtration. *Rep. Results Bac. Chem. Biol. Exam. London Waters 1969–1970*, 44:52, 1969.

V Wintzingerode, Friedrich, U.B Göbel, and E. Stackebrandt. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS microbiology reviews*, 21(3):213–229, 1997.

A. Wise, K. OBrien, and T. Woodruff. Are Oral Contraceptives a Significant Contributor to the Estrogenicity of Drinking Water? *Environmental science & technology*, 45(1):51–60, 2011.

S Withey, Elise Cartmell, LM Avery, and Tom Stephenson. Bacteriophages potential for application in wastewater treatment processes. *Science of the Total Environment*, 339(1): 1–18, 2005.

L. Wittebolle, M. Marzorati, L. Clement, A. Balloi, D. Daffonchio, K. Heylen, and et al. Initial community evenness favours functionality under selective stress. *Nature*, 458(7238): 623–626, 2009.

C.R. Woese. Bacterial evolution. *Microbiological reviews*, 51(2):221– 271, 1987.

GM Wolfaardt, JR Lawrence, RD Robarts, SJ Caldwell, and DE Caldwell. Multicellular organization in a degradative biofilm community. *Applied and environmental microbiology*, 60(2):434–446, 1994.

G. Woodward. *Ecological Networks*, volume 42. Academic Press, 2010.

J.C. Wooley, A. Godzik, and I. Friedberg. A Primer on Metagenomics. *PLoS Compututational Biology*, 6(2):1–13, 2010.

World Health Organisation. Water, Sanitation and Hygiene Links to Health: Facts and Figures, Updated November 2004. *Geneva: WHO*, 2004.

RS Wotton. Water purification using sand. *Hydrobiologia*, 469(1):193–201, 2002.

X-Y. Xiao, D.V. McCalley, and J. McEvoy. Analysis of estrogens in river water and effluents using solid-phase extraction and gas chromatography–negative chemical ionisation mass spectrometry of the pentafluorobenzoyl derivatives. *Journal of Chromatography A*, 923 (1):195–204, 2001.

W. Xue, C. Wu, K. Xiao, X. Huang, H. Zhou, H. Tsuno, and et al. Elimination and fate of selected micro-organic pollutants in a full-scale anaerobic / anoxic / aerobic process combined with membrane bioreactor for municipal wastewater reclamation. *Water Research*, 44(20):5999–6010, 2010.

S. Yachi and M. Loreau. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences*, 96(4): 1463–1468, 1999.

G.G. Ying, R.S. Kookana, and Y.J. Ru. Occurrence and fate of hormone steroids in the environment. *Environment international*, 28(6):545–551, 2002.

D.N. Yoon, S.J. Park, S.J. Kim, C.O. Jeon, J.C. Chae, and S.K. Rhee. Isolation, characterization, and abundance of filamentous members of caldilineae in activated sludge. *The Journal of Microbiology*, 48(3):275–283, 2010.

C-P. Yu, H. Roh, and K-H. Chu. 17β-Estradiol-degrading bacteria isolated from activated sludge. *Environmental science & technology*, 41(2):486–492, 2007.

C-P. Yu, R.A. Deeb, and K-H. Chu. Microbial degradation of steroidal estrogens. *Chemosphere*, 91:1225–1235, 2013.

J. Yue, G. Sun, X. Hu, and J. Huang. The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate *Monosiga brevicollis*. *BMC genomics*, 14(1):729, 2013.

L. Zeng and I. Golding. Following Cell-fate in *E.coli* After Infection by Phage Lambda. *Journal of visualized experiments: JoVE*, 1(56):3363, 2011.

D.R. Zerbino and E. Birney. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5):821–829, 2008.

S. Zhang, Q. Zhang, S. Darisaw, O. Ehie, and G. Wang. Simultaneous quantification of polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and pharmaceuticals and personal care products (PPCPs) in Mississippi river water, in New Orleans, Louisiana, USA. *Chemosphere*, 66(6):1057–1069, 2007.

J.L. Zhao, G.G. Ying, L. Wang, J.F. Yang, X.B. Yang, L.H. Yang, X. Li, et al. Determination of phenolic endocrine disrupting chemicals and acidic pharmaceuticals in surface water of the Pearl Rivers in South China by gas chromatography, negative chemical ionization and mass spectrometry. *Science of the total environment*, 407(2):962–974, 2009.

X. Zheng, M. Ernst, and M. Jekel. Pilot-scale investigation on the removal of organic foulants in secondary effluent by slow sand filtration prior to ultrafiltration. *Water research*, 44(10): 3203–3213, 2010.

J. Zhou, Z-H. Cai, and X-S. Zhu. Are Endocrine Disruptors Among the Causes of the Deterioration of Aquatic Biodiversity? *Integrated Environmental Assessment and Management*, 6(3):492–498, 2010.

B. Zhu, X. Wang, and L. Li. Human gut microbiome: the second genome of human body. *Protein & Cell*, 1(8):718–725, 2010.

Appendix A

Water quality testing

A.1 Ammonia (NH_3)

Procedure: Performed using Hach Test 'N Tube Vials, method based on Salicylate method (Standard Method 4500-NH_3 A

1. Add 2ml of sample to one Test N Tube vials and 2ml ammonia free water to another.
2. Add the contents of one ammonia salicylate reagent pillows to each vial.
3. Add the contents of one ammonia cyanurate reagents pillows to each vial.
4. Cap vials and shake thoroughly to dissolve powder.
5. Start the instrument timer (20min)
6. Put blank into the instrument and zero.
7. Put sample in and measure.

A.2 Chemical Oxygen Demand (COD)

Analysis by the closed reflux titrimetric method (standard method 5220C)

In environmental chemistry, the chemical oxygen demand (COD) test is commonly used to indirectly measure the amount of organic compounds in water. Most applications of COD determine the amount of organic pollutants found in surface water (e.g. lakes and rivers), making COD a useful measure of water quality. It is expressed in milligrams per litre (mg/L), which indicates the mass of oxygen consumed per litre of solution.

COD is defined as the amount of a specified oxidant that reacts with the sample under controlled conditions. The quantity of oxidant consumed is expressed in terms of oxygen equivalence. The oxidant used is the dichromate ion ($Cr_2O_7^{2-}$) which is reduced to the chromic ion (Cr^{3+}). In the standard COD tests a mixture of potassium dichromate and sulfuric acid are used to oxidise organic matter, with silver (Ag^+) added as the catalyst. A simplified example of this reaction is illustrated below.

$$Cr_2O_7^{2-} + 14H^+ + 6e^- \xrightarrow{heat+Ag^+} 2Cr^{3+} + 7H_2O \tag{A.1}$$

A known amount of a solution of potassium dichromate in moderately concentrated sulphuric acid is added to a measured amount of sample and the mixture is boiled in air. In this reaction the oxidising agent, hexavalent chromium is reduced to trivalent chromium. After boiling the remaining hexavalent chromium is titrated against a reducing agent (ferrous ammonium sulphate). The difference between the initial amount of chromium added to the sample and the remaining chromium after the organic matter has been oxidised is proportional to the COD.

A.3 Coliforms

To isolate and enumerate coliforms and *E.coli* from water samples the recommended standard method [Clesceri et al., 2005] deployed uses the membrane filter procedure with the media Membrane Lauryl Sulphate Broth (MSLB). This method works on the basis that lactose fermenting organisms produce yellow colonies on MLSB when incubated at 35°C for 22 -24 h. Dilutions of sample performed in distilled water. These colonies are counted and are known as thermotolerant coliforms, i.e. bacteria resembling and including *E.coli.* : Four vials containing 9ml water were autoclaved at 120°C for 20 minutes and allowed to cool. The first vial was inoculated with 1ml *E.coli* sample and mixed thoroughly using a rotary mixer. The second vial was then inoculated with 1ml of the new sample, mixed and the procedure repeated to produce dilutions of 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}. Petri dishes containing a sterile Millipore pad soaked in membrane lauryl sulphate broth medium (MLSB) were prepared. Using sterile filter membranes, 1ml of each neat, 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}dilutions was filtered using a pump and filter system. The Filter membranes were then placed on the MLSB medium soaked pads, labeled and incubated at 35°C for 22 -24 h.

A.4 Dissolved Organic Carbon (DOC)

Dissolved organic carbon (DOC) is defined as the organic matter that is able to pass through a filter (filters generally range in size between 0.7 and 0.22 um). DOC are important components in the carbon cycle and serve as a primary food sources for aquatic food webs. In addition, DOC alters aquatic ecosystem chemistries by contributing to acidification in low-alkalinity, weakly buffered, freshwater systems. Furthermore, DOC forms complexes with trace metals, creating water-soluble complexes which can be transported and taken up by organisms. Finally, organic carbon, as well as other dissolved and particulate matter, can affect light penetration in aquatic ecosystems, which is important for the ecosystem's phototrophs that need light to subsist.

Procedure

Using the Hach kit for TOC analysis (following Heated-Persulfate Oxidation Method, Standard methods 5310 C) and TOC standards for precision.

1. Filter 10ml of the sample through a 0.45μm filter
2. Turn on the DRB200 reactor and select the TOC program
3. Pour the 10ml filtered sample into 50ml Erlenmeyer flask and add a flee
4. Add 0.4ml of the buffer solution (check pH is 2)
5. Place flask on stir plate and allow to stir on medium speed for 10min
6. Label 2 acid digestion vials with blank and sample
7. Use a filter funner to add contents of Persulfate pillow to each vial (colourless)
8. Add 3ml of organic free water to the blank vial and 3ml of the sample to sample vial.
9. Rinse 2 blue indicator ampules with deionised water and wipe clean
10. Lower one of the unopened ampules into the acid digestion vial. When the score marks on the ampule is level with the top of acid digestion vial, snap the top off and allow it to drop into the digestion vial.
11. Cap the vials and insert into DRB200 and close lid and start program (2h at 103-105°C)
12. Remove vials from machine and place in testube rack, allow to cool (1h)
13. Select Organic Carbon test on spec. Insert the blank and zero the machine.
14. Insert sample and press read, note down reading.

A.5 Nitrates (NO_3^-)

In water and wastewater the forms of nitrogen of greatest interest are in order of decreasing oxidation state; nitrate, nitrite, ammonia and organic nitrogen. All of these forms are biochemically interconvertable and are components of the nitrogen cycle. Total oxidised nitrogen is the sum of nitrate and nitrite. Nitrate generally occurs in trace quantities in surface waters but may attain high levels in some ground water. In excessive amounts it contributes to the illness methemoglobinemia and due to this a limit of 10mg nitrate/L has been imposed for drinking water. Nitrate is found only in small amounts in fresh domestic wastewater but in the effluent of nitrifying biological treatment plants may be fond in concentrations up to 30mg/L.

Principle

NO_3^- is reduced almost quantitatively to nitrite (NO_2^-) in the presence of cadmium (Cd). This method uses commercially available Cd granules treated with copper sulphate. The NO_2^- produced is determined by diazotizing with sulfanilamide and coupling with N-(1-naphthyl)-ethylenediamine dihydrochloride to form a highly coloured azo dye that is measured colourmetrically.

Procedure: Performed using Hach Nitrate powder pillows (21061-69) method based on Cadmium reduction (Standard Methods 4500-NO_3^- E)[1]

1. Measure 2 x 15ml of sample water
2. Add contents of NitraVer 6 Reagent Powder pillow to the sample. Put a stopper over the sample and shake vigorously for 3min
3. Leave the sample to sit for 2min
4. Pour 10ml of the sample into a clean sample cell (not cadmium particles)
5. Add the contents of the NitriVer 3 Nitrite reagent powder pillow to the sample cell.
6. Cap the sample and shake for 30sec (Pink colour will develop) then leave for 15min
7. Fill another sample cell with 10ml of the original sample (for blank)
8. Insert blank and zero instrument (DR 2800)

[1]Note: Nitrate standard (0.4mg/L) were made to confirm accuracy

9. Insert sample and press read, note down the results (mg/L NO_3^- -N) do 5 readings for each sample and average.

A.6 Nitrites (NO_2^-)

Procedure: Performed using Hach Nitrite powder pillows method based on Diazotization method (Standard Methods 4500-NO_2^- B

1. Measure 2 x 10ml of sample water
2. Add the contents of the NitriVer 3 Nitrite reagent powder pillow to the sample cell.
3. Swirl bottle to help powder dissolve (Pink colour will develop) & leave for 20min
4. Fill another sample cell with 10ml of the original sample (for blank)
5. Insert blank and zero instrument (DR 2800)
6. Insert sample and press read, note down the results (mg/L NO_2^- -N) do 5 readings for each sample and average.

A.7 pH: measured using the Hachs portable pH meter

pH will affect the bacteria present and will affect their activity as certain enzymes are affected by pH, it will also give an idea of what kind of reactions are occurring in the water i.e. nitrification will produce more alkaline pH.

A.8 Phosphate (PO_4^{3-})

Phosphorus occurs in nature only in the form of chemical compounds, either as inorganic orthophosphate ($HPO_4{}^{2-}$, $H_2PO_4{}^-$) or in organic compounds. Total phosphorus can be subdivided into particulate phosphorus and soluble phosphorus. Furthermore, soluble phosphorus can be divided into soluble reactive phosphorus and soluble unreactive phosphorus.

Particulate phosphorus consists of adsorbed, exchangeable phosphorus, organic phosphorus, precipitates, reaction products with Ca^{2+} , Fe^{2+} , Al^{3+} and other cations as well as

crystalline minerals and amorphous phosphorus. The soluble form of phosphorus is normally though to consist of orthophosphate, inorganic polyphosphates and dissolved organic phosphorus [Holtan et al., 1988]. The distribution of different species of orthophosphate (H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, or PO_4^{3-}) is pH-dependent. A large part of the identified organic phosphorus fraction is represented by inositol phosphates, phospholipids, nucleic acids, organic acids and phosphate esters. Organic phosphorus can be hydrolysed to inorganic forms through chemical and/or biological reactions [Holtan et al., 1988] or by reactions driven by UV-radiation. Phosphorus combined to biological material (bacteria, phytoplankton) can comprise a large fraction of the total phosphorus in lake water.

Principle

Chemical phosphorus analysis has two steps: 1) conversion of phosphorus compounds to dissolved orthophosphate, and 2) colorimetric determination of the dissolved orthophosphate. According to the Standard Methods, phosphorus is analysed by the ascorbic acid method, where ammonium molybdate and potassium antimonyl tartrate react in an acid medium with orthophosphate to form phosphomolybdic acid, which then reacts with ascorbic acid forming the coloured compound molybdenum blue, which can be analysed with a spectrophotometer.

Procedure: Following Hach protocol based on Standard Methods 4500-P E: Ascorbic Acid Method

1. Make a 2mg/L standards to test accuracy (process as if sample).
2. Fill a sample cell with 10ml of sample
3. Add contents of one PhosVer 3 phosphate pillow to the sample and shake vigorously for 30sec
4. Fill another cell with 10ml (Blank)
5. Put the blank into the cell and zero instrument
6. Put sample in and press read, note down the measurement.

A.9 Specific Ultraviolet Absorption (SUVA)

The SUVA calculation requires both the DOC and UVA measurement. The SUVA is calculated by dividing the UV absorbance of the sample (in cm^{-1}) by the DOC of the sample (in mg/L) and then multiplying by 100 cm/M. SUVA is reported in units of L/mg-M. A high SUVA indicates that a large portion of the organics present in the water are aromatic. Since aromatic organics have a greater tendency to react with disinfectants to create DBPs, a high SUVA indicates there is a high potential for the formation of DBPs.

$$
\begin{aligned}
SUVA \text{ (L/mg-M)} = \frac{UVA(cm^{-})}{DOC(mg/L)} \times 100cm/M \text{(as metre means x 10)} \\
UVACalculation : UVA = A/d \\
\text{where:} \\
\text{UVA = The calculated UV absorbance of the sample} \\
A \text{ = The measured UV absorbance of the sample} \\
d \text{ = The quartz cell path length}
\end{aligned}
\tag{A.2}
$$

A.10 Temperature

Temperature effects the type of bacteria present and the level of their activity. This was measured in the field (at the filter site, in the filters) by inserting digital logging devices equipped with a thermometer into the filter and reading the temperature.

A.11 Total Viable bacteria

To be performed on Plate Count Agar (PCA) following the method of Bahgat et al. [1999], Lautenschlager et al. [2010]. Briefly, dilutions of the sample (total 1ml) are aliquoted and spread onto the media and left to grow for 3d at 30°C, colonies are counted after this time.

A.12 Turbidity

Turbidity in water is caused by suspended and colloidal matter such as clay, silt, finely divided organic and inorganic matter, and plankton and other microscopic organisms. Tur-

bidity is an expression of the optical property that causes light to be scattered and absorbed rather than transmitted with no change in direction or flux level through the sample. Historically, the standard method for determination of turbidity has been based on the Jackson candle turbidimeter; however, the lowest turbidity value that can be measured directly on this device is 25 Jackson Turbidity Units (JTU). Because turbidities of water treated by conventional fluid-particle separation processes usually fall within the range of 0 to 1 unit,indirect secondary methods were developed to estimate turbidity. Electronic nephelometers are the preferred instruments for turbidity measurement. Its precision, sensitivity, and applicability over a wide turbidity range make the nephelometric method preferable to visual methods. Report nephelometric measurement results as nephelometric turbidity units (NTU).

Principle: Based on American Public Health Association protocol 2130 B.
This method is based on a comparison of the intensity of light scattered by the sample under defined conditions with the intensity of light scattered by a standard reference suspension under the same conditions. The higher the intensity of scattered light, the higher the turbidity. Formazin polymer is used as the primary standard reference suspension. The turbidity of a specified concentration of formazin suspension is defined as 4000 NTU.

Procedure

1. Nephelometer calibration: Follow the manufacturers operating instructions. Run at least one standard in each instrument range to be used. Make certain the nephelometer gives stable readings in all sensitivity ranges used.
2. Measurement of turbidity: Gently agitate sample. Wait until air bubbles disappear and pour sample into cell and measure turbidity (on Turbidimeter 2100N Hach)

A.13 $\mathbf{UV}_{254nm}$

The UVA procedure requires that the sample be passed through a 0.45μl filter and transferred to a quartz cell. It is then placed in a spectrophotometer to measure the UV absorbance at 254 nm and is reported in cm^{-1}, Zheng et al. [2010] provides useful interpretation advice.

Appendix B

16S rRNA Clone Library Construction

Sand sample

In order to investigate the sand filter associated bacterial diversity, different levels of a core from a sand filter (Fairmilehead) were mixed together for total genomic DNA extraction. For an overview of the processing involving in creating a clone library see Figure B.1

DNA Extraction

This procedure has been optimised for sand / sediment samples using MP BIO FastDNA Spin Kit for soil (6560-200)

1. Add up to 500mg of soil to a Lysing matrix E tube.
2. Add 978μl Sodium Phosphate buffer to the sample tube
3. Add 122μl MT buffer
4. Homogenise in the FastPrep machine for 30s at speed setting 6.0
5. Centrifuge at 13,000g for 10 min to pellet debris.
6. Add 250μl PPS to clean 2.0ml microcentrifuge tube and transfer supernatant into. Mix by inverting by hand 10x.
7. Centrifuge at 14000g for 5min to pellet precipitate. Transfer supernatant to a clean 2ml tube.
8. Add 1ml binding matrix suspension to supernatant in 2ml tube.

9. Invert by hand for 2min to allow DNA binding.
10. Place tubes in a rack for 3min to allow silica matrix to settle.
11. Remove and discard 600μl of supernatant
12. Resuspend binding matrix (shake) in the remaining amount of supernatant and transfer 600μl of mixture to Spin Filter and centrifuge at 14000g for 1min.
13. Empty catch tube and turn filter to prevent clogging, add remaining mixture to the filter and recentrifuge as before and empty.
14. Add 500μl of SEWS-M (remember to have added 100ml 100% ETOH to the bottle) to the tube to resuspend the pellet.
15. Centrifuge at 13000g for 1min and empty catch tube.
16. Centrifuge again at 13000g for 2min without any liquid being added.
17. Remove filters and put in kit supplied catch tubes, allow to air dry the filter for 5min at RT
18. Gently re-suspend binding matrix (above spin filter) in 50l DNase pyrogen free water (pipette into centre of filter).
19. Centrifuge at 13000g for 1min to bring eluted DNA into clean catch tube. Discard spin filter.
20. Qubit
21. Run 5μl of DNA extract with 2μl loading buffer on a gel to check purity.
22. Run at 100v for 50min.
23. Store DNA extraction in -20 freezer until needed for PCR

Always best to run extractions on a 1% gel (1g agarose to 100ml TAE or TBE + 1μl / 10ml sybr safe), to do load 7.5μl (5μl template mixed with 2.5μl

PCR Amplification

Amplification of the 16s rRNA was performed by pulling the DNA extracted from 10 separate 500mg samples of Sand with the Universal prokaryotic 16S primers:
27F (5'-GAGTTTGATCCTGGCTCAG-3') and 1392R (5'-ACGGGCGGTGTGTRC-3').
The amplification reaction consisted of

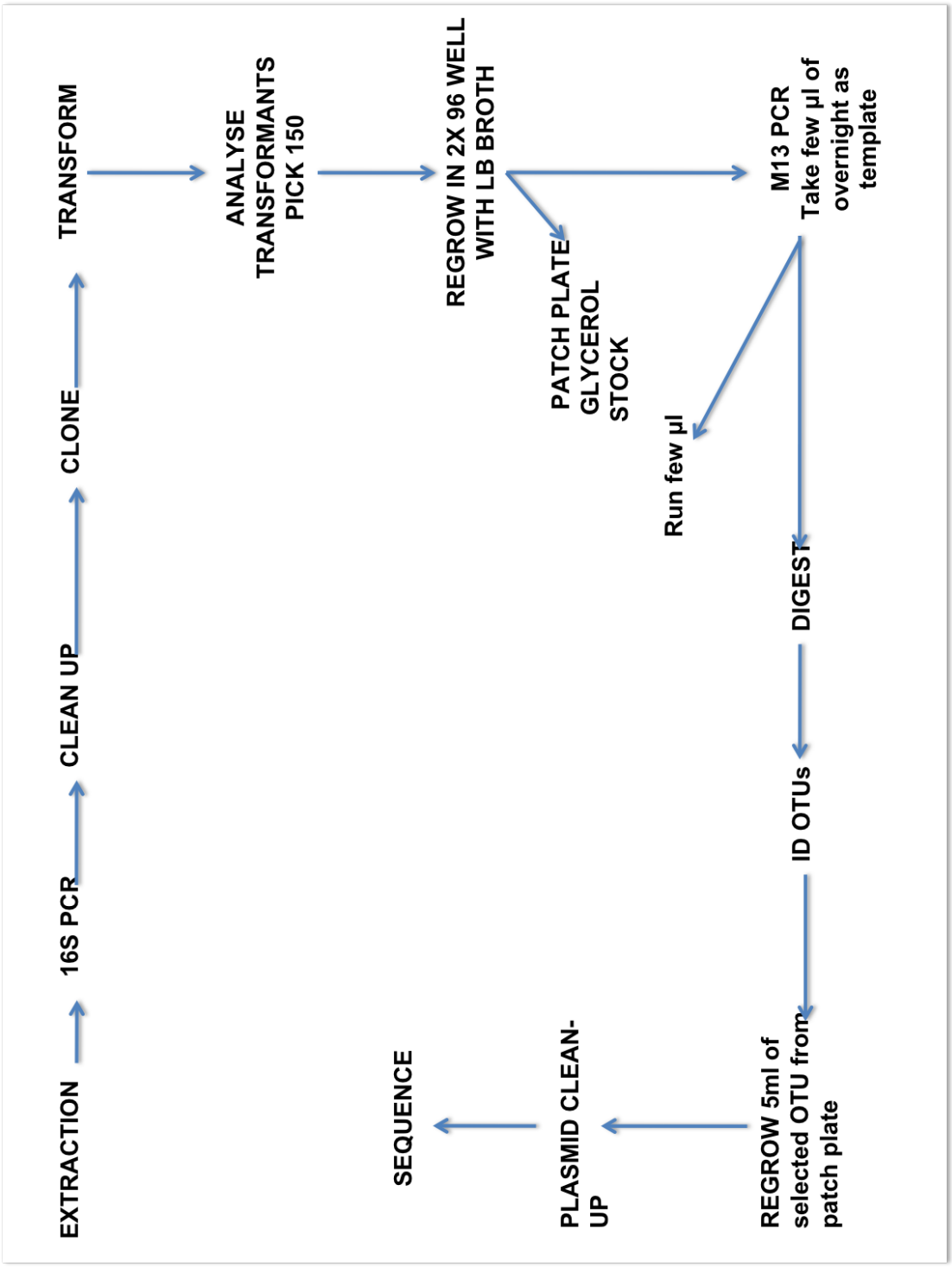


Figure B.1: Work flow showing procedures involved in making clone library

2μl (200ng) DNA template (1:10 dilution)

25μl Bioline PCR Mix

1μl each primer (12.5pmol each)

21μ Water

Total Volume 50μl

The amplification was carried out using the following conditions: initial denaturation (95°C for 5 min) followed by 30 cycles of denaturation (94°Cfor 1min), annealing (62°C for 1min), and extension (72°C for 1min) and a final extension of 72°C for 10 min [McHugh et al., 2004]. All 10 separate amplifications were pooled and then purified using QIAquick PCR Purification Kit.

PCR Product Clean-Up

The PCR product was run on a 1% gel and checked for purity. The PCR products were then purified using the QIAquick PCR Purification Kit (250 reactions)
Important points before starting

- Add ethanol (96-100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature.
- Add 1:250 volume pH indicator I to Buffer PB (i.e., add 120 μl pH indicator I to 30 ml Buffer PB or add 600μl pH indicator I to 150 ml Buffer PB). The yellow color of Buffer PB with pH indicator I indicates a pH of 7.5.
- Add pH indicator I to entire buffer contents. Do not add pH indicator I to buffer aliquots.
- If the purified PCR product is to be used in sensitive microarray applications, it may be beneficial to use Buffer PB without the addition of pH indicator I.

Procedure

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix. It is not necessary to remove mineral oil or kerosene. (For example, add 500 μl of Buffer PB to 100μl PCR sample (not including oil)).

2. If pH indicator I has been added to Buffer PB, check that the colour of the mixture is yellow. If the colour of the mixture is orange or violet, add 10μl of 3 M sodium acetate pH 5.0, and mix. The colour of the mixture will turn to yellow.
3. Place a QIAquick spin column in a provided 2 ml collection tube. 4.To bind DNA, apply the sample to the QIAquick column and centrifuge for 30-60 s. 5. Discard flow-through. Place the QIAquick column back into the same tube. Collection tubes are re-used to reduce plastic waste.
4. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30-60s.
5. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min.
   IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.
6. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.
7. To elute DNA, add 50μl Buffer EB (10 mM TrisCl, pH 8.5) or water (pH 7.0-8.5) to the center of the QIAquick membrane and centrifuge the column for 1 min. Alternatively, for increased DNA concentration, add 30μl elution buffer to the centre of the QIAquick membrane, let the column stand for 1 min, and then centrifuge.[1]
8. If the purified DNA is to be analysed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel.

Library construction

Cloning PCR Product into Vectors for Sequencing

Cloning will be performed using the TOPO TA kit (K4575-02 Invitrogen)

Making Kanamycin stock

[1]Elution efficiency is dependent on pH. The maximum elution efficiency is achieved between pH 7.0 and 8.5. When using water, make sure that the pH value is within this range, and store DNA at 20°C as DNA may degrade in the absence of a buffering agent. The purified DNA can also be eluted in TE buffer (10 mM TrisCl, 1 mM EDTA, pH 8.0), but the EDTA may inhibit subsequent enzymatic reactions.

Kanamycin is usually made at a stock of 50μg/ml (1mg/ml = 1μg / μl).Dissolve 1g of KAN into 20mL of ddH_2O and syringe filter into 0.45μl and aliquot into 1ml eppendorf tube and freeze at -20°C.

Making the LB-Agar Plates with Kanamycin

1. 10g Trypton
2. 5g yeast extract
3. 10g $NaCl_2$
4. 950ml ddH_2O
5. 15g Agar
6. Adjust pH with NaOH to 7.3
7. Autoclave
8. Once cooled to touch add 1ml Kanamycin (50mg/ml) using sterile technique to the media and pour plates sterilely and store at 4°C

Making LB with Kanamycin

In order to make the LB broth follow the procedure above except omit the agar.

Cloning PCR product into Vector using TOPO TA Kit

Add the following ingredient together

1. 2μl PCR product
2. 1μl Salt solution
3. 1μl TOPO Vector
4. 2μl Water
5. 6μl Final Volume
6. Note should use the control template amplified with appropriate primers to test efficiency.
7. Incubate at room temperature (22-25°C) for 30min
8. Place reaction on ice and proceed to transformation
9. NB reaction can be stored at -20°C if necessary

Transforming One Shot TOP 10

1. Thaw TOP10 cells on ice
2. Add 2μl of TOPO cloning reaction to a vial of TOPO10 competent cells and mix gently (**Flick the tube gently, don't pipette up and down**).
3. Incubate on ice for 30min
4. Heat shock the cells for 30sec at 42°C (without shaking).
5. Immediately transfer tubes to ice
6. Add 250μl of room temperature S.O.C medium
7. Cap the tubes tightly and shake the tubes horizontally (200rpm) at 37°C for 1hour.
8. Spread 100μl of mixture on each plate and spread using a glass spreader. Additionally spread the remaining 150μl
9. Incubate at 37°C overnight
10. An efficient TOPO cloning reaction should produce several hundred colonies, pick 150 (white) for analysis and proceed to analysing transformants.

Analysing Transformants

Pick 150 colonies (growth of *E.coli* has Kanamycin resistance gene and insert in vector has spliced the ccd gene (lethal) which is fused to the C-terminus of LacZ) and culture in 200μl of LB broth (produces as above) overnight at 37°C in 96 well nunc plates[2]

Amplification of Insert in Transformants

Amplification products, generated using the vector-specic primers
M13F (5'GTTTTCCCAGTCACGAC-3') and M13R (5'-CAGGAAACAGCTATGAC-3'), were obtained from clones following the same recipe as before with 2μl of the overnight culture as the template source. 2μl of each sample were analysed on a 1% agarose gel and only inserts of the correct size were subjected to digestion.

[2]Remember to make a copy plate which contains LB broth with 15% glycerol for long term storage. Mix 0.85 ml of culture with 0.15 ml of sterile glycerol and transfer to a cryovial and store at -80C. Alternatively fill the wells of a 96 well plate up with 200μl of LB + ampicillin + 15% glycerol and transfer a loopful of each clone into the wells.

Digestion

ARDRA (Amplified Ribosomal DNA Restriction Analysis)

Digested using the tetrameric restriction endonuclease HaeIII (Promega) at 37°C 4h following the manufacturers protocol:
sterile, deionized water 7.3μl (change depending on template concentration)
RE 10X Buffer 2μl
Acetylated BSA, (10μg/μl) 0.2μl
DNA, (1μg/μl) 10μl
Mix by pipetting, then add: Restriction Enzyme, (10μg/μl) 0.5μl
final volume 20μl
We will digest the remaining M13 amplified amplicon (48μl)

Resultant DNA fragments (ran all 20μl) were separated electrophoretically in 0.7% TAE high-resolution agarose gels, containing syber safe. Operational taxonomic units (OTUs) were identified, based on restriction cleavage patterns and clones representing the OTUs selected for sequencing. Using the Quick Plasmid Miniprep Kit kit plasmid DNA from overnight cultures of different OTUs were prepared and sent for sequencing.

Plasmid purification

Before Starting: Overnight cultures of the selected clones are to be grown in 3ml of LB broth + kanamycin (in falcon tubes).

- Add RNase A to Resuspension Buffer (R3) according to instructions on the label. Mix well. Mark on the label that RNase A is added. Store buffer at 4°C.
- Add 96 - 100% ethanol to Wash Buffer (W9) and Wash Buffer (W10) according to instructions on each label. Mix well. Mark on the labels that ethanol is added. Store both wash buffers with ethanol at room temperature.
- If the Lysis Buffer (L7) contains salt precipitates, warm the buffer in a 37°C water bath for a few minutes until precipitates dissolve. Do not shake the buffer.

Preparing Cell Lysate

1. Pellet 1-5 ml of an overnight culture. Thoroughly remove all medium from the cell pellet.
2. Completely resuspend the pellet in 250μl Resuspension Buffer (R3) with RNase A. No cell clumps should remain.
3. Add 250μl Lysis Buffer (L7) to cells. Mix gently by inverting the capped tube 5 times. Do not vortex.
4. Incubate the tube for 5 minutes at room temperature. Do not exceed 5 minutes.
5. Add 350μl Precipitation Buffer (N4). Mix immediately by inverting the tube until the solution is homogeneous. For large pellets shake more vigorously. Do not vortex.
6. Centrifuge the mixture at 12,000 x g for 10 minutes at room temperature using a microcentrifuge to clarify the lysate from lysis debris.
7. Load the supernatant from Step 6 onto a Spin Column.
8. Place the Spin Column with supernatant from Step 7 of Preparing Cell Lysate (front page) into a 2-ml Wash Tube.
9. Centrifuge at 12,000 x g for 1 minute. Discard the flow- through and place the column back into the Wash Tube.
10. Add 500μl Wash Buffer (W10) with ethanol to the column. Incubate for 1 minute at room temperature. Centrifuge at 12,000 x g for 1 minute. Discard the flow-through and place column back into the Wash Tube.
11. Add 700μl Wash Buffer (W9) with ethanol to the column.
12. Centrifuge the column at 12,000 x g for 1 minute. Discard the flow-through and place the column back into the Wash Tube.
13. Centrifuge the column at 12,000 x g for 1 minute to remove any residual Wash Buffer (W9). Discard the Wash Tube with the flow-through.
14. To Elute DNA place the Spin Column in a clean 1.5-ml Recovery Tube.
15. Add 75μl of preheated TE Buffer (TE) to the centre of the column.
16. Incubate the column for 1 minute at room temperature.
17. Centrifuge at 12,000 x g for 2 minutes.
18. The Recovery Tube contains your purified plasmid DNA. Discard the column.

Appendix C

Stable Isotope Probing (SIP)

SIP Protocol: adapted from Neufeld et al. [2007]

Reagents

TrisHCl (1 M, pH 8.0) Dissolve 121.1 g Tris base in 800 ml of milliQ water. Adjust to pH 8.0 with HCl and add milliQ water to make up the volume to 1.0 litre.

70% ethanol

GB Add 50ml of 1MTrisHCl, 3.75g KCl and 1ml of0.5M EDTA to 400ml of milliQ water. Dissolve KCl, then add milliQ water to 500 ml. Filter-Sterilise (0.2 μm). Autoclave. The final solution is 0.1 M Tris, 0.1 M KCl and 1 mM EDTA.

TE buffer: Prepare a solution of 10 mM TrisHCl (pH 8.0) and 1 mM EDTA (pH 8.0) in milliQ water. Autoclave.

CsCl solution: Combine 250 ml of water with 250 g CsCl (this solution is referred to as 1 g ml^{-1}). Warming the solution to 30 1C can help dissolve the CsCl. Store solution in sealed aliquots (to prevent evaporation and crust formation around bottle caps). The final density of this solution, once combined with EtBr, is cited as being 1.55 g ml^{-1}

PEG solution: Combine 150 g of polyethylene glycol 6000 with 46.8 g NaCl. Dissolve in milliQ water to a total volume of 500 ml. Filter-sterilize (0.2 mm). Autoclave. The solution is 30% PEG 6000 and 1.6 M NaCl.

TE-saturated 1-butanol: Add TE to 100 ml of 1-butanol with intermittent mixing until two phases form. Butanol is less dense and becomes the upper phase.

Method

1. Extract DNA and quantify
2. Mix a volume of DNA (510 ηg) with GB to a final volume of 1.2 ml in a 15-ml screw-cap tube.
3. The mixed volume of GB/DNA will vary depending on the density of the CsCl stock solution. The volume of GB/DNA mix to be added to the CsCl stock solution can be calculated using the following formula:

   $Volume =$ ((CsCl stock density - desired final density) $\times$ (CsCl volume added)$\times$ 1.52) (C.1)

4. In a 15-ml screw-cap tube, combine 4.80 ml of CsCl stock solution and the 1.20 ml GB/DNA solution. **CRITICAL STEP**: Before using the CsCl stock, ensure that a precipitate of CsCl has not formed in the bottom of the bottle. If some precipitation has occurred, shaking the bottle for several minutes dissolves the CsCl back into solution
5. Mix by gentle inversion
6. Slowly add the mixture to an ultracentrifuge tube with a Pasteur pipette or a syringe and needle. **CRITICAL STEP**: Solution should be flush with the base of the tube neck. Tubes should be balanced to within 10mg.
7. Seal the tube with a heat sealer (POSSIBLE PAUSE POINT, DNA STABLE IN CsCl FOR WEEKS) and place tubes in the rotor with balanced pairs opposite each other. Make sure caps are tightly screwed on and check they are hanging ok by pulling down on the buckets.
8. Spin conditions are 44,100 rpm in the AH-650 swinging rotor at 20°C for 40 h with vacuum, maximum acceleration and without brake.
9. Carefully remove tubes from rotor with forceps and place in rack
10. Prepare the top displacement gradient fractionator by connecting tubing into the fractionator hood to a HPLC pump drawing from mineral oil. Prime the line until a single drop emerges out of the fractionator hood.
11. Place centrifuge tube in fraction recovery system. Lower the fraction recovery top onto the top of the open tube, ensuring tight seal, and lock in position
12. Pierce bottom of tube by inserting fraction collection needle

13. Set pump to 500μL/min (2.0 on pump), switch on and start a stop watch. Collect a fraction every 30 secs in a sterile 1.5ml tube.
14. Clean fractionators by removing hood and pipetting 2ml of 0.1m NaOH, then 2ml absolute ethanol through tube
15. Check density of fractions using a refractometer
16. Precipitate DNA by adding 10μl glycogen (5mg/ml) for every 150μl of DNA and 2 volumes of PEG (Glycogen help to recover low DNA concentrations). Mix well by inversion.
17. Leave at RT for 2h to precipitate DNA (can store overnight at this stage if necessary)
18. Centrifuge at 13000g for 5 min, discard supernatant.
19. Wash with 500μl of 70% ETOH
20. Centrifuge at 13000g for 5 min and remove supernatant and allow to air dry for at least 15 min
21. Resuspend pellet in 30μl sterile water
22. 1μL of each sample will then be used for 16s, 18s and *E.coli* specific qPCR assays
23. Based on density and qPCR results a heavy (13C) and a light (12C) fraction will be chosen for each time point and used for metagenomic library creation (Nextera XT)

Appendix D

Calculating the Importance of Protozoa and Viruses in *E.coli* Removal

All calculations are based on the SIP data up to 4h as this is the longest doubling time of one of the protozoa. These calculations are approximations and based on the following several assumptions[1].

- **Protozoa 1** (**P1**) - *Monosiga brevicollis*; grazing rate (gr_1) = 196 bacteria/ h, genome size = 41.6MB
- **Protozoa 2** (**P2**) - *Tetrahymena*; grazing rate (gr_2) = 1161 bacteria/ h, genome size = 104MB
- **Virus 1** - *Enterobacteria Phage Lambda*; burst size (bs) = 161viruses/ h, genome size = 48KB

E.coli was fed to the system at 300cfu/ml at flow rate of $0.15 m^3 m^{-2} h^{-1}$ corresponding to 850ml = 255000 *E.coli*. After 4h only 750cfu left - therefore 254,249 *E.coli* removed (=E1)

X = # Total organisms in the population

[1]These calculations were formulated with the assistance of Melanie Schirmer, University of Glasgow

Protozoa 1

$$\text{P1}_{Lab}\text{(Proportion of labelled protozoa 1)} = \frac{\text{Number of reads for labelled protozoa 1}}{\text{Total number of reads}} \tag{D.1}$$

$= 109706/3653980 = 0.0300237$

To account for the differences in genome size (P1 genome size is 867 times larger than V1 genome) normalisation is performed as follows:

N-P1$_{Lab}$ $=$ **P1**$_{Lab}$ **/ 867 = 0.0000346**

X.P1$_{Lab}$: Labelled P1

P1$_{Prot}$**:** Proportion of significant labelled protozoa made up by protozoa 1

P1$_{Prot}$**:** $= 109706/256425 = 0.4278288$ or 42.78288%

E1: number of *E.coli* eaten by P1

After 4h Protozoa 1 will have eaten 4 $\times$ gr1 (gr1 = grazing rate) E.coli. We assume that only 50% of the new carbon must be labeled to cause a density change in the protozoa and hence become labeled. Thus we multiply by half.

$$E1_e = \frac{4 \times gr}{2}$$
$$E1_e = 392 \tag{D.2}$$

$$P1_P = \frac{109706.33}{1850465}$$
$$P1_P = 0.05928\ (5.93\%) \tag{D.3}$$

α_1**:** In order to account for 2 protozoa species competing for *E.coli* we must average, therefore each protozoa gets 50% of the *E.coli*.

$$\alpha_1 = \frac{50}{P1_{Prot}} \tag{D.4}$$
$$\alpha_1 = 1.1686918 \tag{D.5}$$

Therefore the number of labelled *E.coli* eaten by protozoa 1:

$$\alpha_1(X \times \text{N-P1}_{Lab1} \times E1) = 0.0158512 \times X \tag{D.6}$$

Protozoa 2 Variable Names

Analogous for protozoa 2, we get:

$$\text{P2}_{Lab}(\text{Proportion of labelled protozoa 2}) = \frac{\text{Number of reads for labelled protozoa 1}}{\text{Total number of reads}} \tag{D.7}$$

$= 8461 / 3653980 = 0.0023156 \quad (0.23\%)$

Normalised by genome size (P2 genome is 2,167 times larger than V1 genome).

N-P2$_{Lab} = \text{P2}_{Lab} / 2167 = 0.0000011$

X.P2$_{Lab}$: Labelled P2

P2$_{Prot}$**:** Proportion of all significant protozoa made up by protozoa 2

P2$_{Prot}$**:** $= 8461/256425 = 0.03299600$ (3.299600%)

E2: number of *E.coli* eaten by P2

After 4h Protozoa 2 will have eaten 4 × gr2 (gr2 = grazing rate) E.coli. We assume that only 50% of the new carbon must be labeled to cause a density change in the protozoa and hence become labeled. Thus we multiply by half.

$$E2_e = \frac{4 \times gr}{2}$$
$$E2_e = 2322 \tag{D.8}$$

α_2**:** In order to account for 2 protozoa species competing for *E.coli* we must average, therefore each protozoa gets 50% of the *E.coli*.

$$\alpha_2 = \frac{50}{P2_{Prot}} \tag{D.9}$$
$$\alpha_2 = 15.1533519 \tag{D.10}$$

Therefore the number of labelled *E.coli* eaten by protozoa 2:

$$\alpha_2(X \times \text{N-P2}_{Lab2} \times E2) = 0.0387047 \times X \tag{D.11}$$

Virus 1 Variable Names

X = number of organisms in the population

$$V1_{Lab}(\text{Proportion of labelled virus 1}) = \frac{\text{Number of reads for labelled virus 1}}{\text{Total number of reads}} \tag{D.12}$$

= 24378 / 3653980 = 0.0066716

X.V1$_{Lab}$: Labelled V1

bs: Burst size (161 viruses / h)

$$\text{V1}_{vir} = \frac{\text{labeled V1}}{\text{All significant labeled viruses}} \tag{D.13}$$

= 24378 / 44339 = 0.5498094 (54.98094%)

β: *Multiplication factor to approximate the whole virus population with V1:*

β = 100/V1$_{vir}$

= 1.8188121

Therefore the number of labelled *E.coli* lysed by all viruses:

$$\beta(X \times \frac{\text{V1}_{Lab}}{bs}) = 0.0000754 \tag{D.14}$$

Final Calculation

As 254,249 *E.coli* are known to be removed after four hours by the system and assuming that this is achieved collectively by protozoa 1, protozoa 2 and virus 1 the following calculations can be used to approximate the proportion of *E.coli* removed by each organism.

$$(\alpha_1(X \times \text{N-P1}_{Lab2} \times E1)) + (\alpha_2(X \times \text{N-P2}_{Lab2} \times E2_e)) + (\beta(X \times \frac{\text{V1}_{Lab}}{bs})) = 254249$$
$$0.0546313 \times X = 254249$$
$$X = 4653907.1923970 \tag{D.15}$$

Therefore replacing X with the solved number for each organism and adjusting for the proportion they make-up of their community we can conclude:

$$P1 = \frac{0.0158512 \times X}{\alpha_1}$$
$$P1 = 63121.8715560 \approx 63,122 \textit{ E.coli} \tag{D.16}$$

$$P2 = \frac{0.0387047 \times X}{\alpha_2}$$
$$P2 = 11887.0123693 \approx 11,888 \textit{ E.coli} \tag{D.17}$$

$$V1 = \frac{0.0000754 \times X}{\beta}$$
$$V1 = 192.9306509 \approx 193 \textit{ E.coli} \tag{D.18}$$

Therefore based on the above assumptions P1 approximately removes 326 times more E.coli than V1 and 5.3 times more than P2. Overall P1 accounts for 24.83% of the removal, P2 accounts for 4.68% and V1 0.08%.

All significant protozoa v.s. all significant phages:

Protozoa: 0.0158512X + 0.0387047X = 0.0545559X

Viruses : 0.0000754X

Thus Protozoa consumed 253,898.0953977 ≈ 253,899 *E.coli* and

Viruses consumed 350.9046023 ≈ 351 *E.coli*

Therefore based on the approximations:

All significant protozoa collectively remove 254,193 (99.86%) *E.coli*

All significant viruses collectively remove 57 (0.14%) *E.coli*