DOCTOR OF PHILOSOPHY

Distribution, Diversity, and Function of Filterable Microorganisms in the Conwy River

Ghuneim, Lydia-Ann

Award date:
2020

Awarding institution:
Bangor University

Distribution, Diversity, and Function of Filterable Microorganisms in the Conwy River

Lydia-Ann John Ghuneim

(2019)

PRIFYSGOL
BANGOR
UNIVERSITY

A thesis submitted to Bangor University in candidature for the degree Philosophiae Doctor

School of Natural Sciences, Bangor University,
Bangor, Wales, LL57 2UW
United Kingdom

Declaration and Consent

Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

Author Name: Lydia-Ann John Ghuneim

Title: Distribution, Diversity, and Function of Filterable Microorganisms in the Conwy River

Supervisor/Department: Peter Golyshin/SNS, Olga Golyshina/SNS, and David Jones/SNS

Funding body (if any):

.....

Qualification/Degree obtained: PhD

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use

(http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

Statement 1:

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …….. (candidate)

Date ……..

Statement 2:

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed ……. (candidate)

Date …….

Statement 3:

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed ……. (candidate)

Date …….

NB: Candidates on whose behalf a bar on access has been approved by the Academic Registry should use the following version of **Statement 3:**

Statement 3 (bar):

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loans and for electronic storage (subject to any constraints as defined in statement 4), after expiry of a bar on access.

Signed …… (candidate)

Date ……

Statement 4:

Choose **one** of the following options

a) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.	x
b) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted.	
c) I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.	

Options B should only be used if a bar on access has been approved by the University.

In addition to the above I also agree to the following:

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: ……

Date : …….

Table of Contents

Thesis Summary

The Conwy River in North Wales, UK, is a large freshwater system that has been the subject of numerous studies over many years. Even so, the river's microbiome is poorly understood. It is widely known that in this system, as with all lotic systems, dissolved organic matter (DOM) is the primary source of energy for these resident microorganisms. Therefore, understanding how these species utilize DOM can allow scientists to make better predictions concerning the river's water quality.

However, there are two major scientific challenges that must be considered. Firstly, due to the anthropogenic inputs (from agriculture and wastewater treatment) in tandem with the variation in flow rate and weather events, make it inherently difficult to properly model this system. Secondly, the vast majority of bacteria and archaea cannot be cultured under typical *in vitro* conditions. Evidence also suggests that many microbial species in aquatic systems have evaded detection due to their ability to pass through ultra-small filters (<0.45 μm pore sizes), i.e. filterable microorganisms. The term filterable microorganism can refer to one of the following: (1) small-bodied cells (less than 0.1 μm^3 volume), (2) shrunken cells (due to limited nutrients or senescence), and (3) large cells that squeeze through small filters (<0.45 μm pore sizes). Their exact role in freshwater systems remain largely unknown.

The purpose of this thesis was to uncover the taxonomic identity, overall function, and role in DOM cycling of filterable species residing in the Conwy River while also comparing them to the native lotic community (i.e. unfiltered population). We utilised 16S rRNA single amplicon sequencing and shotgun sequencing to conduct a phylogenetic analysis of both ultra-filtered (passed through a 0.22 μm pore size filter) and unfiltered river water to understand the phylogenies and relative phyla distributions as well as determining which clusters of orthologous groups (COGs) were present. The distribution of COGs of both microcosms were compared to other environments and bacterial genomes to (1) assess similarity, and (2) determine if organism complexity is related to environment (i.e. are more complex organisms found in nutrient rich environments, etc). Next, we examined how either microbial community utilised dissolved organic carbon (DOC) via multi-omics and ^{14}C radio-isotope tracking in order to determine whether DOC influenced these populations or whether the residing species showed any particular preference to a DOC type.

The major findings indicated that, the dominant phyla (listed in decreasing abundance) in the whole community were *Proteobacteria*, *Bacteriodetes*, *Actinobacteria*, and *Firmicutes*. Whereas, the filtered community contained more *Firmicutes* than *Bacteriodetes* and *Actinobacteria*. We also we detected the presence of several candidate phyla, most notably "*Candidatus* Parcubacteria". There were more COGs in the filtered community that fall under the functional categories of replication, recombination, repair, and cell wall/membrane/envelope biogenesis comparatively to the entire population. Clustering metagenomes against single genomes revealed that the filtered community's COG distribution was closely related to COG distribution of organisms with limited/streamlined genomes. The filtered microbiome also metabolized DOC at a slower rate than the whole community and was confirmed to be a taxonomically unique subset within the greater system. Changes within each community were not influenced by the addition of DOC and neither system had a preference in DOC type.

Overall, the results obtained from this body of work demonstrate that the filtered microcosm was a unique population nestled within the general microbial community. They differ in taxonomic makeup and their usage of low-molecular weight DOC, suggested that they may have different functional roles in freshwater ecosystems. By exploring the complex microbiome of the Conwy River, researchers can gain a better understanding of water quality, ecosystem management, and the nature of filterable microorganisms.

This work is dedicated to Dr. Martha May-John Madar Ghuneim (September 20th 1949-September 20th 2009). Love you mom.

اسم الله عليكي بزيادة تقلع عين الحسادة، اسم الله عليكي قالت أمك عين الحسود تطلع منك

(Translation)

In God's name, (to you, in excess) the evil eye will be plucked out. In God's name, your mother said to the evil eye, leave.

Acknowledgements

First and foremost, I would like to thank my supervisors, in no particular order, Professor Davey Jones, Professor Peter Golyshin, and Dr. Olga Golyshina. Because of your guidance, assistance, support, and overall putting up with my shenanigans, I was able to evolve into the scientist I am today. Thank you for challenging me to think more critically about my work and giving indispensable knowledge that I will carry for the rest of my life.

To my fellow colleagues and comrades in arms, thank you for helping me not only in my PhD, but also survival in unfamiliar territory. Considering that prior to this I have never been to Europe, it was a tremendous help to have a group of individuals explain what the heck Eurovison is and the cultural significance of Colin the Caterpillar. To my comrades that do not hail from the British Isles, thank you for graciously sharing your culture with me. Each and every one of you helped me feel less lonely in an alien world. Unfortunately, there are too many of you to list here individually! You know who you are.

To my friends and family, thank you for being you. Each of you has taught me something and I will never forget your kindness. As stated before, I have limited space to thank every single person, but know this, it is truly a privilege to know each and every one of you.

A special shout out to my friend, Dr. Akiko Kochi. I look forward to our daily exchanges and you keep me sane (well mostly sane). The advice you have given me throughout these years has been truly indispensable. Proof that a simple conversation about cute shoes (specifically my patchwork chucks) can transform into a friendship, lasting years that even continents cannot separate. Although, as I recall that conversation it also involved excellent smelling cologne and possibly taste in anime.

On a more personal note, thank you to my siblings (and acquired siblings) Dr. George Ghuneim, Sarah Ghuneim, Dr. Angela Ghuneim, Dr. Alexander Ghuneim, and Sarah Sidlow. How lucky are we to be part of the same family and be friends. We may quibble and get on each other's nerves from time to time (as do all siblings on this planet), but know that I will always be there for you as you have been for me. We have been through hell and back with each other and I wouldn't trade it for anything. To my niece and nephews, Anastasia Rose Ghuneim, Gabriel George Ghuneim, and John George Ghuneim, just know that Auntie loves you very much. See your faces makes my heart sing.

Last but definitely not least to my father, Dr. John George Ghuneim, there are no words to fully express my gratitude. From delivery room to now you have always been there. Even though we are thousands of miles apart, you are in my heart and I will never stop loving you. I love you infinity times infinity to the infinity power. This can be represented via the following equation, where *L* stands for love: $L = (\infty \times \infty)^{\infty}$

List of Abbreviations

AAI	Average Amino acid Identity
AMD	Acid Mine Drainage
AMP	Adenosine Monophosphate
ANOVA	Analysis of Variance
ARMAN	Archaeal Richmond Mine Acidophilic Nanoorganism
ATP	Adenosine Triphosphate
BLAST	Blast Local Alignment Search Tool
BSTFA	N,O-Bis(trimethylsilyl)trifluoroacetamide
CARD-FISH	Catalyzed Reporter Deposition-Fluorescence *In Situ* Hybridization
CBB	Calvin-Benson-Bassham cycle
CoA	Coenzyme A
COG(s)	Clusters of Orthologous Group(s)
Cyro-TEM	Cryogenic Transmission Electron Microscopy
DDH	DNA-DNA Hybridization
dH_2O	Distilled Water
DNA	Deoxyribonucleic Acid
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
DOP-PCR	Degenerate Oligonucleotide PCR
dsDNA	Double Stranded Deoxyribonucleic Acid
eDNA	Environmental Deoxyribonucleic Acid
EI	Electron Ionization
FACS	Fluorescence-activated cell sorting
FAME	Fatter Acid Methyl Ester
Fe	Iron
FISH	Fluorescence *in situ* Hybridization
***g* (RCF)**	Relative Centrifugal Force
G+C	Guanine and Cytosine
GC	Gas Chromatography

GC-EI-Q-TOF-MS	Gas Chromatography-Electron Ionization-Quadpole-Time of Flight-Mass Spectrometry
GC-MS	Gas Chromatography-Mass Spectrometry
GGD	Genome-To-Genome Distance
GTDB	Genome Taxonomy Database
LC-MS	Liquid Chromatography–Mass Spectrometry
LMW	Low Molecular Weight
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionization- Time-Of-Flight Mass Spectrometry
Mbp	Mega Base Pair
MDA	Multiple Displacement Amplification
MLSA	Multi-Locus Sequence Analysis
MRC	Marine *Roseobacter* Clade
MW	Molecular Weight
NA	Not Available/Not Applicable/No value
NAD	Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NanoSIMS	Stable Isotope Imaging
NCBI	National Center for Biotechnology Information
NMDS	Non-metric Multidimensional Scaling
nr/nt	Nucleotide
OTU(s)	Operational Taxonomic Unit(s)
PBS	Phosphate Buffer Solution
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PEP	Primer Extension Preamplification
PERMANOVA	Permutational Multivariate Analysis of Variance
POC	Particulate Organic Carbon
POM	Particulate Organic Matter
PVDF	Hydrophilic Polyvinylidene Fluoride
RNA	Ribonucleic Acid
RO	Reverse Osmosis
rRNA	Ribosomal Ribonucleic Acid
RuBisCO	Ribulose-1, 5-biphosphate carboxylase-oxygenase
SA	Surface Area

SEM	Scanning Electron Microscopy
SIP	Stable Isotope Probing
SNS	School of Natural Science
SSU	Small Subunit
TCA	Tricarboxylic Acid Cycle
TMCS	Chlorotrimethylsilane
TEM	Transmission Electron Microscopy
TNF-α	Tumor Necrosis Factor alpha
tRNA	Transfer Ribonucleic Acid
UV	Ultra-Violet
V	Volume
WGA	Whole Genome Amplification
Xgal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Research significance

Freshwater systems play a vital role in terrestrial global carbon (C) cycling (e.g. biotic fixation, consumption, transport, storage), even though they only contain 0.009% of the Earth's total water (Cole et al., 2007; Okafor, 2011). It is estimated that inland freshwater ecosystems, such as lentic and lotic systems, receive over 1.9 Pg of C from terrestrial systems per year, with approximately 0.2 Pg of this C subsequently becoming locked up in aquatic sediments. A further 0.8 Pg of C is returned to the atmosphere via gas exchange and the remaining 0.9 Pg is delivered to marine systems (Cole et al., 2007; Battin et al., 2008). This makes aquatic systems one of the largest active reservoirs of dissolved organic C (DOC) in the biosphere (Benner and Amon, 1996). Much of this C (including dissolved organic matter; DOM) is utilized by heterotrophic microorganisms (Azam et al., 1983; Benner and Amon, 1996; Meyer, 1994; Münster and Chróst, 1990; Stutter and Cains, 2016). The utilisation of DOM is one of the key functions of the 'microbial loop', i.e. the microbial food web (Fig. 1.1) (Azam et al., 1983; Fenchel, 2008).

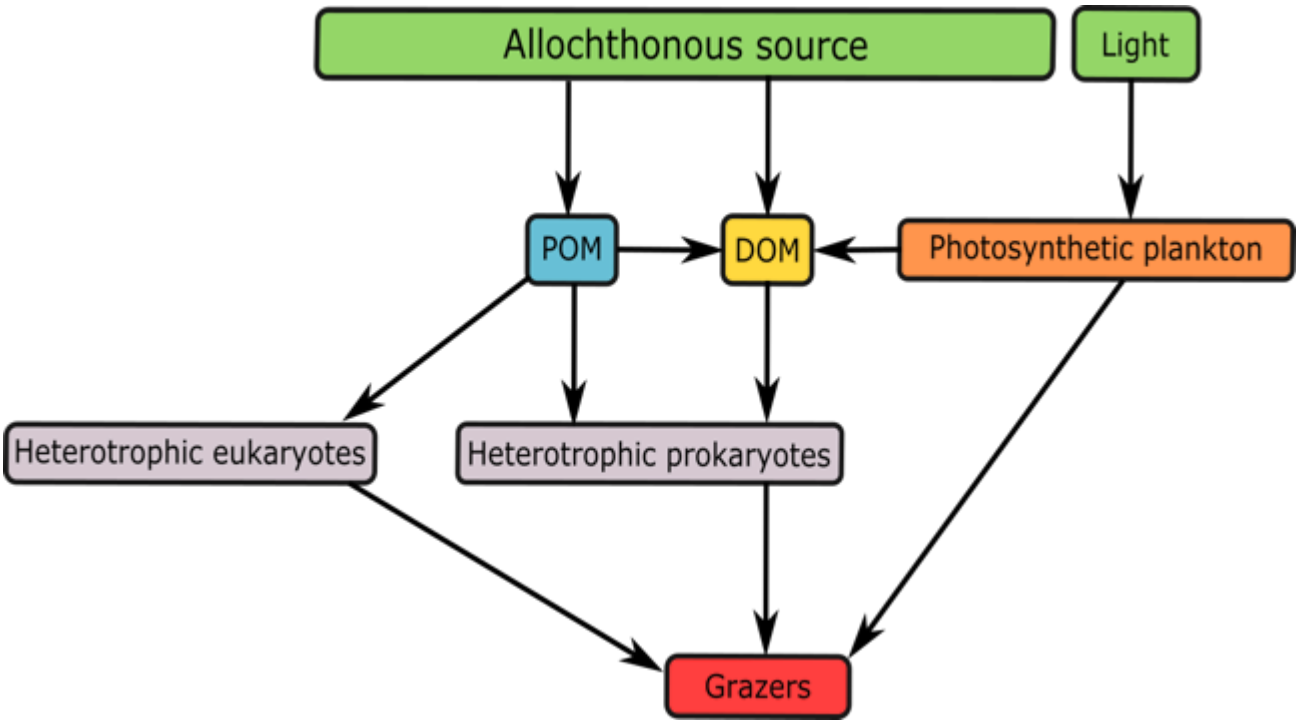


Figure 1.1-Unicellular microbial loop in lotic systems. Grey boxes are heterotrophic (benthic) microorganisms. The orange box is photosynthetic plankton and the red box is grazers (which can be heterotrophic eukaryotes). The blue box is POM (particulate organic matter). The yellow Box is DOM. The green boxes are the external source of nutrients from either an allochthonous source (runoff and other detritus) or from light. The external source is either in the form of or broken down into POM and DOM. POM is taken up by heterotrophic eukaryotes and prokaryotes or is further broken down to DOM. Which then heterotrophic prokaryotes use as their primary source of nutrients. Heterotrophic microorganisms are prey of larger grazers

Microalgae and bacteria are major components of the microbial loop and it is estimated that lentic and lotic populations range between 10^3 -10^4 to 10^6-10^7 cells per mL (Jones, 2001). Archaea, which are usually associated with extreme environments, are found in freshwater and can participate in nutrient cycling (Wells et al., 2006; Bomberg et al., 2008; Herfort et al., 2009; Cavicchioli, 2011). They normally constitute less than 10% of the microbial community in freshwater systems, but there are some systems, such as high mountain oligotrophic lakes in Spain and Crater Lake in the USA, where they make up to 37% of the total community (Wells et al., 2006; Bomberg et al., 2008; Herfort et al., 2009; Cavicchioli, 2011). Although freshwater microorganisms are extremely abundant, it can be

challenging to study them for a range of reasons including: (1) large temporal variation in freshwater chemistry and flow over both short and long timescales, (2) wide spatial variation in biogeochemistry within and between river systems, (3) anthropogenic perturbation, and (4) the general difficulties in the isolation and characterisation of microorganisms.

The first challenge noted above involves the intrinsic dynamic conditions of freshwater ecosystems. Freshwater ecosystems are exposed to numerous external factors (e.g. variations in temperatures and solar radiation, flow volumes, inputs from agricultural runoff sand soil erosion) that cause transient or permanent changes in physical and chemical parameters (Böckelmann et al., 2000; Jones, 2001; Beman et al., 2005; Allan and Castillo, 2007). These changes can alter biological activity as well as the size, composition and functioning of the microbial community (Böckelmann et al., 2000; Jones, 2001). For instance, lotic systems are largely dependent on allochthonous (external) sources of C and the hydrologic flow can easily alter the concentration, composition and bioavailability of DOC, inorganic nutrients and xenobiotics in the system (Meyer, 1994; Sigee, 2005). In response to this, heterotrophic benthic species dominate, residing in both the surface and subsurface regions, and are responsible for many key processes that include organic matter breakdown and the provision of C for consumers within the food chain (Fig. 1.1)(Meyer, 1994; Sigee, 2005).

Another major challenge is the presence of microbial dark matter, which is the colloquial term used to described microorganisms that remain largely uncharacterized due to our inability to culture them in the laboratory (Stewart, 2012; Lever et al., 2015). The reason for this limitation is that these microorganisms have stringent growth requirements (i.e. growth requirements and their dependencies on other community members) which cannot easily be replicated with current technologies (Stewart, 2012). Large metagenomic studies have uncovered many species which in turn has led to genome reconstructions. This is

exemplified by the discovery of "*Ca.* Pelagibacter ubique" which was initially detected, via a metagenomic investigation in pelagic water from the Sargasso sea, with later studies proving it to be one of the most abundant organisms on Earth (Rappé et al., 2002; Carini et al., 2012; Tripp, 2013; Zhao et al., 2017).

Therefore, to combat the challenges outlined above, a combination of system-level (holistic) approaches are required. These include studies of their abundance, growth, metabolic activity, function, spatial distribution, taxonomy, relationships with other organisms, and persistence using techniques like fluorescence in situ hybridization (FISH), metagenomics, isotopic tracers etc. (Böckelmann et al., 2000).

The Conwy catchment is located in North Wales, UK and is rather unique in that it contains a wide range of land cover, soil types, topography, agricultural land use types, hydrological dynamics, and climate types (Emmett et al., 2016). The main artery of the catchment, the Conwy River, at 55 km long with a tidal limit 20 km inland, discharges into the Irish Sea with an average flow rate of 18.59 m^3/s (National River Flow Archive; Simpson et al., 2001). It (and the wider catchment) has been the subject of numerous studies which includes and not limited to: viromics, DOM cycling, pollutant flow (e.g. pesticides, metals), eDNA tracking, sediment flow, estuarine-river dynamics, inorganic nutrient cycling (e.g. N, P, C) (Mudge and Norris, 1997; Simpson et al., 2001; Williams et al., 2012; Brailsford et al., 2017; Farkas et al., 2018).

Interestingly, there has been few major studies that have focused on the general microbial ecology of this river. However, a recent government-sponsored research programme within the catchment has facilitated investigation of microbial populations. The impetus for these investigations is that the Conwy suffers from known microbial/pathogenic contamination that threatens the EU Blue Flag status of local tourist beaches and has also been associated with foodborne disease outbreaks originating from local shellfisheries (Lees

et al., 1995; Adriaenssens et al., 2018; Farkas et al., 2018). Some of these previous studies have therefore focused on specific microbiological pollutants uses for assessing water quality (e.g. faecal coliforms, Norovirus), whilst others have focused on the general ecology of the river (e.g. the effect of filtration, temperature, and acidification on the utilisation of DOM by the microbial community) (Quilliam et al., 2014; Brailsford et al., 2017; Adriaenssens et al., 2018; Farkas et al., 2018). In one study it was found that freshwater passed through a 0.22 μm pore size filter showed clear consumption of ^{14}C-labelled DOC (Brailsford et al., 2017). It was concluded that this was due to the presence of filterable microorganisms. This group of organisms are vastly understudied in freshwater ecosystems and therefore formed the basis for this study.

Filterable microorganisms are difficult to define, however, this term typically refers to: (1) nano-sized microorganisms, small-bodied microorganisms that always have dimensions of 50-400 nm; (2) microorganisms with larger cells that have the capability to squeeze through filters with pore sizes of less than 0.45 μm, and (3) small-cell variants of microorganisms with larger cell sizes (e.g. dormant or senescent forms) (Velimirov, 2001; Panikov, 2005; Duda et al., 2012; Proctor et al., 2018). The scientific community has known about these entities since the latter half of the twentieth century, and remained largely understudied due to technical limitations (Oppenheimer, 1952). As the development of culture independent techniques rose, so did further exploration into the nature of these microorganisms. Even with all the advancements, the ecological role of filterable and nano-sized microorganisms is largely unknown. However, from what is known it appears that they have variable functions, ranging from prey for protist grazers to participation in geochemical processes (Salcher et al., 2013; Dang and Lovell, 2016). A further in depth review of filterable microorganisms can be found in Ch.2 of this thesis.

1.2. Purpose, aims, and plan of research

The work presented in this thesis aims to further characterize the microbial community of the Conwy River and its filterable fraction in terms of taxonomic identity, potential functionality (predicted functionalities), and DOM (or DOC) usage. The first aim will be to determine which microbial taxa, particularly filterable organisms, reside within the Conwy lotic system using a combination of shotgun sequencing and single amplicon sequencing (Fig. 1.2 and 1.3). Once the community members have been identified, the second aim will be to explore their function. This will be achieved by tracing the uptake and fate of common low molecular weight ^{14}C-labelled and unlabelled organic compounds (i.e. amino acids, sugars and organic acids) at representative concentrations (Fig. 1.4). This isotopic tracer study will be combined with mass-spectrometry (GC-MS) and shotgun metagenome sequencing to determine the changes in taxonomic compositions of both, total microbial community and its filterable fraction (Fig. 1.4).

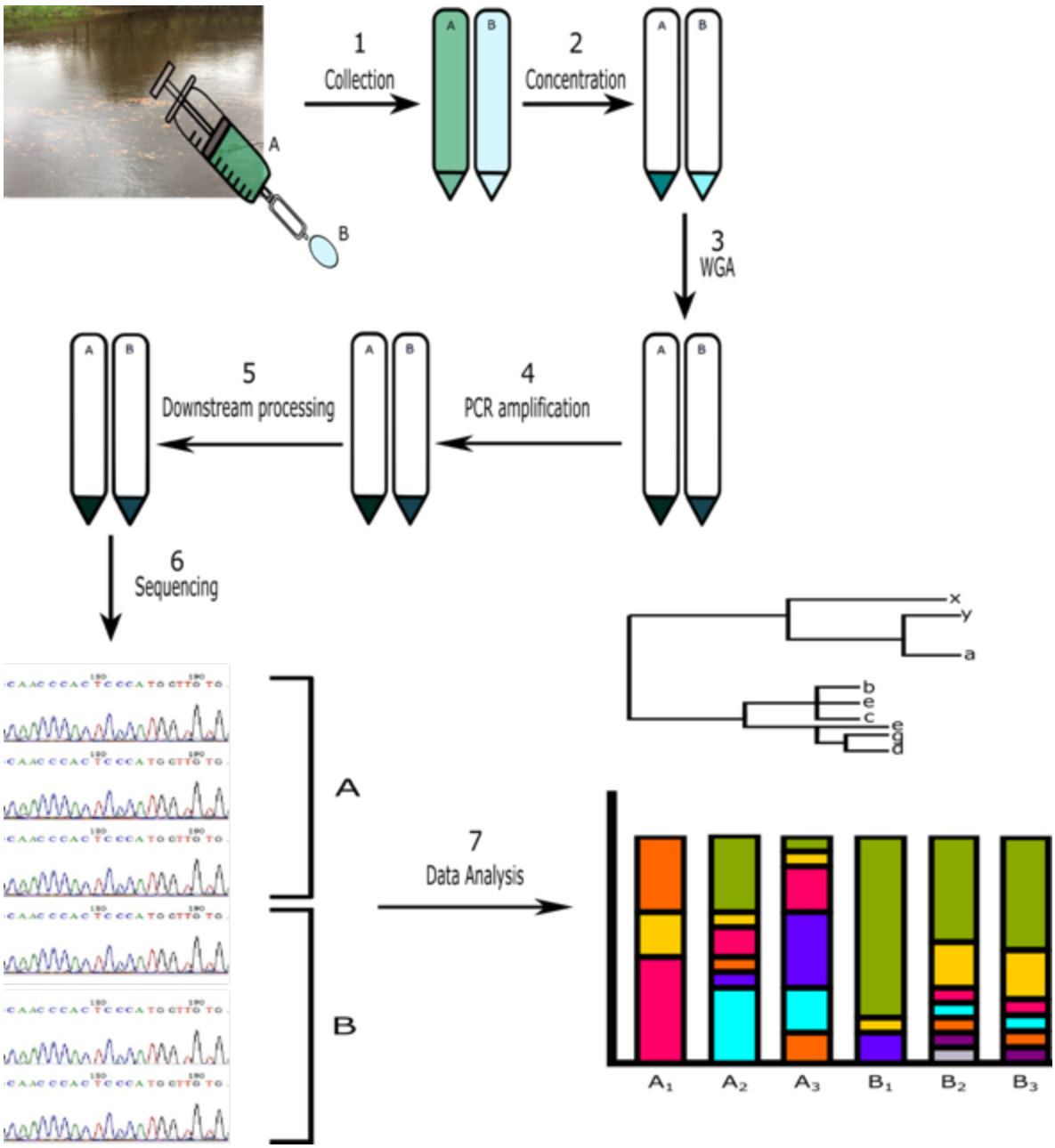


Figure 1.2- Sampling and sequencing workflow. (1) Water was collected from the Conwy River. Both the (A) unfiltered and (B) filtered fractions were analysed. (2) Then the samples were concentrated via high speed centrifugation. (3) Supernatant was removed and then cellular pellets were subjected to WGA (Whole genome amplification). (4) PCR amplification using primers 27F and 1492R, clone library preparation and sequencing; V3 and V4 regions for Illumina amplicon sequencing and Oxford Nanopore primers for 16S workflow; (5) Downstream processing is done in the form of prepping amplified regions for sequencing (i.e. further purification, gel extraction, TOPO-TA cloning). (6) Sequencing is done for 16S rRNA gene amplicons or clones and shotgun sequencing using Illumina MiSeq protocols. (7) Resultant sequences and contigs are finally read, trimmed, assembled, and identified for further data analysis (taxa identification and distribution, gene distribution, diversity measures).

1.3. Microbial community's taxonomic composition

The first aim of this work was to determine which microbial species were residing in the Conwy River. I used the workflow outlined in figures 1.2 and 1.3 to answer the following questions:

- What is the taxonomic community composition of the Conwy River?
- Which microorganisms are filterable (i.e. cells that can pass through a 0.22 μm filter)?
- What are the differences, in terms of community makeup, between 0.22 μm filtered freshwater fraction and the unfiltered water?
- Can we predict the presence of small genome microorganisms within an environment by the clusters of orthologous groups of proteins (COGs)?
- Is the microbial community of the Conwy River unique when compared to similar freshwater environments?
- How does the addition of low molecular weight compounds (amino acids, sugars, and organic acids) in nanomolar concentrations change the makeup of the community?

I began by collecting approximately 50 mL of aqueous sample from both the unfiltered and filtered fraction (Fig. 1.2). The filtered fraction was obtained via ultrafiltration with a 0.22 μm pore sized filter. Subsequently, these were concentrated by high speed centrifugation followed by whole genome amplification (WGA) to amplify the DNA present (Figs. 1.2 and 1.3). After the DNA was amplified, taxonomic analysis was carried out in the form of amplicon sequencing (via examination of the 16S rRNA genes via Sanger sequencing of PCR amplicon libraries and Nanopore™ amplicon) and shotgun sequencing.

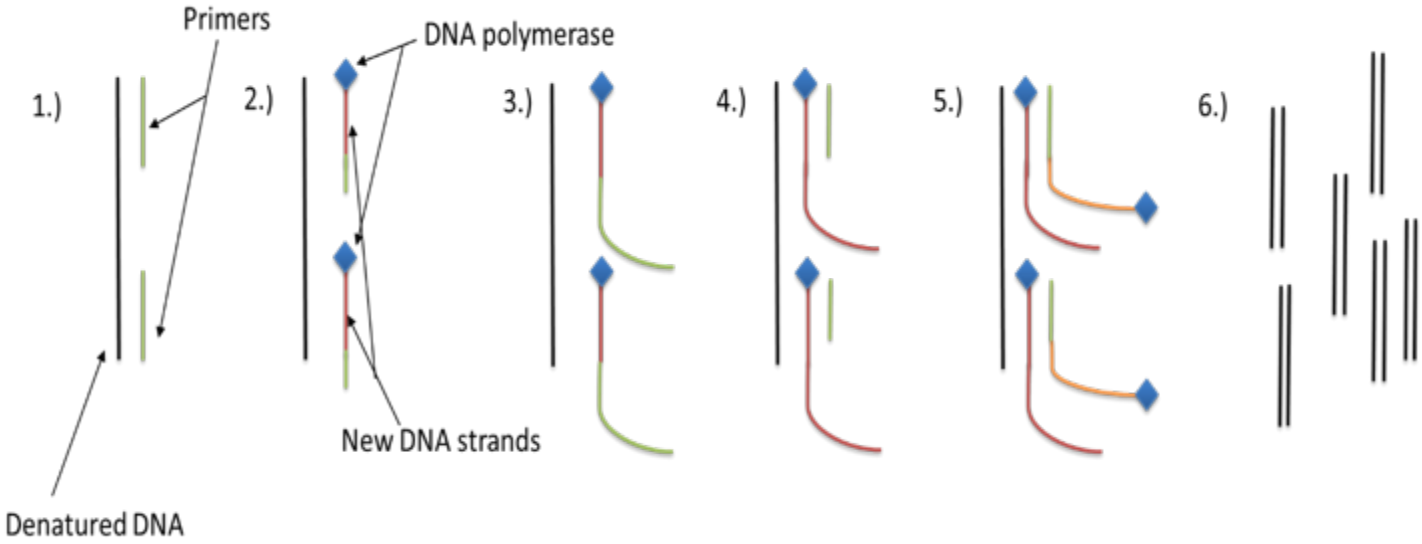


Figure 1.3- Schematic of the WGA process. This figure was adapted from Figure 1 of the IllustraTM Ready-To-GoTM GenomiPhiTM V3 DNA Amplification Kit product manual (GE Healthcare UK Limited, Little Chalfont, Buckinghamshire, UK). The WGA process is as follows: (1) The primers bind to denatured (single stranded) template DNA; (2) DNA polymerase starts the polymerization, which in turn (3) displaces the strands, and (4) new primers bind to new DNA. (5) This polymerization cycle continues until (6) the DNA polymerase is inactivated at 65 °C.

1.4. DOC utilisation and function prediction

The next phase of this work was to predict genomic backgrounds, model functionality, and to evaluate for real-time utilisation of common, DOC compounds by the filtered and unfiltered microbial communities and to via from the results of the through shotgun metagenome sequencing via COGs (Fig. 1.2). In addition, I examined the fate of added ^{14}C-labelled DOC and the transformation of unlabelled substrates in laboratory-based microcosms (Fig. 1.4). The questions that were addressed here were:

- Can the filtered microbial fraction utilize low molecular weight DOC (amino acids, organic acids, and sugars) and how does this compare to the raw unfiltered fraction?
- Is there any substrate preference in the lotic microbial community and is there a difference between lotic microorganisms residing in filtered versus the general community?
- What are the differences in whole-metagenome metabolic reconstructions of filtered freshwater microbial community fraction and that of the total microbial community?

The microbial community composition dynamics was examined via 16S rRNA amplicon metabarcoding and shotgun sequencing. Whole-community with consequent metabolic reconstruction was achieved via shotgun sequencing. In addition, radioisotopic methods ($^{14}CO_2$ respiration and, ^{14}C incorporation into the biomass, and ^{14}C labelled substrate consumption) and targeted metabolomics (GC-MS) were used to determine rate of consumption of low molecular weight DOC compounds (Fig. 1.4).

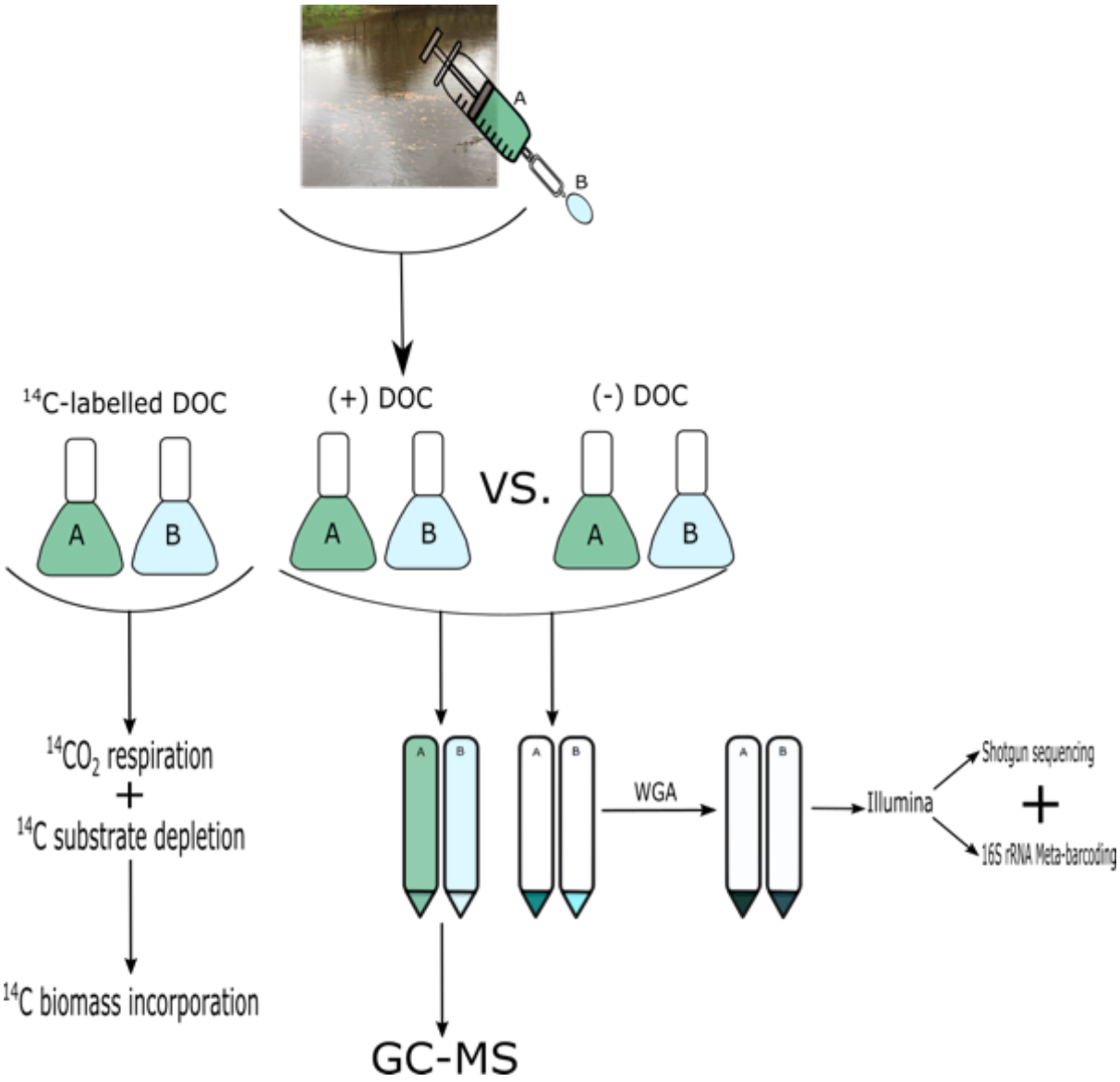


Figure 1.4-Workflow of microcosm experiment. Water is collected from the Conwy River. Both the (A) unfiltered and (B) filtered fractions are collected. Filtered and unfiltered fraction samples are spiked with ^{14}C-labelled DOC or unlabelled DOC (compare to non-spiked sample) and were allowed to incubate for 3 weeks. Targeted metabolomics (via GC-MS), community (taxa) dynamics (via 16S rRNA metabarcoding and shotgun sequencing), whole-community metabolic reconstruction (shotgun sequencing), and $^{14}CO_2$ respiration, and ^{14}C substrate depletion) were measured periodically throughout the incubation period. Subsequently ^{14}C biomass incorporation was then calculated via the difference from the respiration and depletion values.

1.5. Chapter information

All chapters are prepared in the style of a journal article manuscript. This includes authors, contributions, manuscript progress (published, submission, accepted, not yet submitted). Below is a summary of the chapters and appendices presented within the thesis.

- Chapter 2 (Literature Review): "Nano-sized and Filterable Archaea and Bacteria: Biodiversity and Function" (Published)
- Chapter 3: "Assessment of methodologies of DNA extraction and whole genome amplification (WGA) from freshwater and seawater samples for consecutive genomic analysis"
- Chapter 4:" Prokaryotic diversity of filtered (0.22 μm) and unfiltered fractions residing in the Conwy River"
- Chapter 5: "Utilization of low molecular weight organic compounds by microbial communities residing in the Conwy River"
- Chapter 6: "Discussion"
- Appendix I: "Method development for long term ^{14}C-labelled DOC isotopic experiments"
- Appendix II: "Supplemental Material for Chapter 5"

1.6. References

Adriaenssens, E. M., Farkas, K., Harrison, C., Jones, D., Allison, H. E., and Mccarthy, A. J. (2018). Viromic analysis of wastewater input 2 to a river catchment reveals a diverse 3 assemblage of RNA viruses. 3, 1–39. doi:10.1101/248203.

Allan, J. D., and Castillo, M. M. (2007). *Stream Ecology: Structure and function of running waters*. Springer Netherlands Available at: https://books.google.co.uk/books?id=4tDNEFcQh7IC.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer - Reil, L. A., and Thingstad, F. (1983). The ecological role of water - column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi:10.3354/meps010257.

Battin, T. J., Kaplan, L. A., Findlay, S., Hokinson, C. S., Marti, E., Packman, A. I., et al. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* 1, 95–100. doi:10.1038/ngeo602.

Beman, J. M., Arrigo, K. R., and Matson, P. A. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434, 211–214. doi:10.1038/nature03370.

Benner, R., and Amon, R. M. W. (1996). Bacterial Utilization of Different Size Classes of Dissolved Organic Matter. *Limnol. Oceanogr.* 41, 41–51.

Böckelmann, U., Manz, W., Neu, T. R., and Szewzyk, U. (2000). Characterization of the microbial community of lotic organic aggregates ('river snow') in the Elbe River of Germany by cultivation and molecular methods. *FEMS Microbiol. Ecol.* 33, 157–170. doi:10.1016/S0168-6496(00)00056-8.

Bomberg, M., Montonen, L., Münster, U., and Jurgens, G. (2008). Diversity and function of archaea in freshwater habitats. *Curr. Trends Microbiol.* 4, 1–89. Available at:

http://www.researchtrends.net/tia/abstract.asp?in=0&vn=4&tid=41&aid=2693&pub=2008&type=3.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S. J. (2012). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J.* 7, 592–602. doi:10.1038/ismej.2012.122.

Cavicchioli, R. (2011). Archaea - Timeline of the third domain. *Nat. Rev. Microbiol.* 9, 51–61. doi:10.1038/nrmicro2482.

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., et al. (2007). Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems* 10, 172–185. doi:10.1007/s10021-006-9013-8.

Dang, H., and Lovell, C. R. (2016). Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* 80, 91–138. doi:10.1128/MMBR.00037-15.Address.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Farkas, K., Cooper, D. M., McDonald, J. E., Malham, S. K., de Rougemont, A., and Jones, D.

L. (2018). Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. *Sci. Total Environ.* 634, 1174–1183. doi:10.1016/j.scitotenv.2018.04.038.

Fenchel, T. (2008). The microbial loop - 25 years later. *J. Exp. Mar. Bio. Ecol.* 366, 99–103. doi:10.1016/j.jembe.2008.07.013.

Herfort, L., Kim, J., Coolen, M., Abbas, B., Schouten, S., Herndl, G., et al. (2009). Diversity of Archaea and detection of crenarchaeotal amoA genes in the rivers Rhine and Têt. *Aquat. Microb. Ecol.* 55, 189–201. doi:10.3354/ame01294.

Jones, G. J. (2001). Freshwater ecosystems - Structure and response. *Ecotoxicol. Environ. Saf.* 50, 107–113. doi:10.1006/eesa.2001.2079.

Lees, D. ., Henshilwood, K., Green, J., Gallimore, C. I., and Brown, D. W. G. (1995). Detection of small round structured viruses in shellfish by reverse transcription-PCR. *Appl. Environ. Microbiol.* 61, 4418–4424.

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., et al. (2015). Life under extreme energy limitation: A synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* 39, 688–728. doi:10.1093/femsre/fuv020.

Meyer, J. L. (1994). The microbial loop in flowing waters. *Microb. Ecol.* 28, 195–199. doi:10.1007/BF00166808.

Mudge, S. M., and Norris, C. E. (1997). Lipid biomarkers in the Conway Estuary (North Wales, UK): a comparison between fatty alcohols and sterols. *Mar. Chem.* 57, 61–84.

Münster, U., and Chróst, R. J. (1990). "Origin, Composition, and Microbial Utilization of Dissolved Organic Matter," in (Springer, New York, NY), 8–46. doi:10.1007/978-1-4612-3382-4_2.

National River Flow Archive Available at: https://nrfa.ceh.ac.uk/.

Okafor, N. (2011). "Ecology of Microorganisms in Freshwater," in *Environmental Microbiology of Aquatic and Waste Systems* (Dordrecht: Springer Netherlands), 111–122. doi:10.1007/978-94-007-1460-1_5.

Oppenheimer, C. H. (1952). The membrane filter in marine microbiology. *J. Bacteriol.* 64, 783–786.

Panikov, N. S. (2005). Contribution of Nanosized Bacteria to the total biomass and activity of a soil microbial community. *Adv. Appl. Microbiol.* 57, 245–296.

Proctor, C. R., Besmer, M. D., Langenegger, T., Beck, K., Walser, J.-C., Ackermann, M., et al. (2018). Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J.* doi:10.1038/s41396-018-0070-8.

Quilliam, R. S., Jamieson, J., and Oliver, D. M. (2014). Seaweeds and plastic debris can influence the survival of faecal indicator organisms in beach environments. *Mar. Pollut. Bull.* 84, 201–207. doi:10.1016/J.MARPOLBUL.2014.05.011.

Rappé, M. S., Connon, S. a, Vergin, K. L., and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633. doi:10.1038/nature00917.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907. doi:10.1038/ismej.2012.162.

Sigee, D. C. (2005). *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment*. Wiley Available at: https://books.google.co.uk/books?id=AJ8nIkoB1o0C.

Simpson, J. H., Vennell, R., and Souza, A. J. (2001). The salt fluxes in a tidally-energetic estuary. *Estuar. Coast. Shelf Sci.* 52, 131–142. doi:10.1006/ecss.2000.0733.

Stewart, E. J. (2012). Growing Unculturable Bacteria. *J. Bacteriol.* 194, 4151–4160.

Stutter, M. I., and Cains, J. (2016). The mineralisation of dissolved organic matter recovered from temperate waterbodies during summer. *Aquat. Sci.* 78, 447–462. doi:10.1007/s00027-015-0446-z.

Tripp, H. J. (2013). The unique metabolism of SAR11 aquatic bacteria. *J. Microbiol.* 51, 147–153. doi:10.1007/s12275-013-2671-2.

Velimirov, B. (2001). Nanobacteria, Ultramicrobacteria and Starvation Forms: A Search for the Smallest Metabolizing Bacterium. *Microbes Environ.* 16, 67–77. doi:10.1264/jsme2.2001.67.

Wells, L. E., Cordray, M., Bowerman, S., Miller, L. A., Vincent, W. F., and Deming, J. W. (2006). Archaea in particle-rich waters of the Beaufort Shelf and Franklin Bay, Canadian Arctic: Clues to an allochthonous origin? *Limnol. Oceanogr.* 51, 47–59. doi:10.4319/lo.2006.51.1.0047.

Williams, A. P., Quilliam, R. S., Thorn, C. E., Cooper, D., Reynolds, B., and Jones, D. L. (2012). Influence of land use and nutrient flux on metabolic activity of E. coli O157 in river water. *Water. Air. Soil Pollut.* 223, 3077–3083. doi:10.1007/s11270-012-1090-z.

Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., Mcintosh, J. R., and Nicastro, D. (2017). Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium "Candidatus Pelagibacter ubique. *Appl. Environ. Microbiol.* 83, 1–14.

Chapter 2

Nano-sized and filterable Bacteria and Archaea: Biodiversity and Function

This manuscript has been published in Frontiers in Microbiology

Ghuneim L-AJ, Jones DL, Golyshin PN and Golyshina OV (2018) "Nano-Sized and Filterable Bacteria and Archaea: Biodiversity and Function." *Front. Microbiol*. 9:1971. doi: 10.3389/fmicb.2018.01971

Abstract

Nano-sized and filterable microorganisms are thought to represent the smallest living organisms on earth and are characterized by their small size (50-400 nm) and their ability to physically pass through <0.45 μm pore size filters. They appear to be ubiquitous in the biosphere and are present at high abundance across a diverse range of habitats including oceans, rivers, soils and subterranean bedrock. Small-sized organisms are detected by culture-independent and culture-dependent approaches, with most remaining uncultured and uncharacterized at both metabolic and taxonomic levels. Consequently, their significance in ecological roles remain largely unknown. Successful isolation, however, has been achieved for some species (e.g. *Nanoarchaeum equitans* and "*Candidatus* Pelagibacter ubique"). In many instances, small-sized organisms exhibit a significant genome reduction and loss of essential metabolic pathways required for a free-living lifestyle, making their survival reliant on other microbial community members. In these cases, the nano-sized prokaryotes can only be co-cultured with their 'hosts'. This paper analyses the recent data on small-sized microorganisms in the context of their taxonomic diversity and potential functions in the environment.

Key words: nano-sized microorganisms, ultramicrocells, filterable microorganisms, unculturable, oligotrophy and copiotrophy

2.1. Introduction

Recent technological advances in microbiology have helped to reveal the enormous diversity of prokaryotic life on our planet (Kuczynski et al., 2010; Caporaso et al., 2011; Thompson et al., 2017). While this has enabled us to characterize and map prokaryote populations across a diverse array of ecosystems, the functional role of most of these organisms remains unknown, due to our inability to culture, and study them in the laboratory. Nevertheless, using culture-independent approaches, e.g. metagenomics, many new candidate taxa that include nano-sized and filterable organisms have been discovered.

Nano-sized microorganisms are termed ‘ultra-micro bacteria’, ‘ultra-micro cells’, ‘dwarf cells’, ‘ultra-small bacteria’, ‘nanoorganisms’, ‘nanobacteria’, nanoarchaea and ‘nanobes’ (Velimirov, 2001; Baker et al., 2010; Duda et al., 2012). The term nanoarchaea only relates to the phylum *Nanoarchaeota* (Huber et al., 2002), although it is commonly erroneously used within the literature. The exact definition of these terms is widely debated and no clear set of guidelines currently exists, however, it is considered that the microorganism must be in the “nano-range” (i.e. 50 to 400 nm) in size. It should also be noted that in regards to aquatic systems, these ultra-small-sized organisms are not part of nanoplankton (2.0-20 μm in size), but instead reside in the picoplankton (0.2-2.0 μm) or femtoplankton (0.02-0.2 μm) communities (Sieburth et al., 1978; Fenchel, 1982; Azam et al., 1983).

Previous studies have focused on detection of ultra-small-sized organisms in a wide range of environmental conditions including: acid mine drainage settings (AMD) (Baker et al., 2006), glacial ice (Miteva and Brenchley, 2005), permafrost (Suzina et al., 2015), freshwater (Fedotova et al., 2012; Ma et al., 2016; Nakai et al., 2016), subterranean bedrock (Wu et al., 2015), hypersaline lakes (Narasingarao et al., 2012), the open ocean (Venter et al., 2004; Giovannoni et al., 2005; Glaubitz et al., 2013; Rogge et al., 2017), and the human body (Kajander and Ciftcioglu, 1998; Kajander et al., 2003; He et al., 2015). The predictions from genomic data from these environments suggest that there are many microorganisms that contain small genomes and either are present as free-living organisms or form a symbiotic relationship with other life forms, which adds another level of complexity to assess their functional role in the environment.

As the review of Duda et al. (2012) discusses a number of issues related with ultramicrobacteria, the aim of present review was to highlight the latest discoveries related to (1) taxonomic diversity, (2) biogeography, (3) current experimental approaches to characterize

these organisms and (4) potential role of ultra-small Bacteria and Archaea within a contrasting range of environments.

2.1.1 Overview of Terminology

When considering ultra-small or nano-sized organisms, it is important to note the significance of the terminology. There is no singular definition of what a nano-sized organism is (ultra-small bacteria, ultra-micro bacteria, nanobes, nanoforms, ultramicrocells, etc.) and consequently a variety of interpretations exists. Many of the terms are either synonymous, as in the case of ultra-small and ultra-micro (Velimirov, 2001), or can be classified as separate organisms, as in the case of nanobacterium and nanobe (Duda et al., 2012). Here we consider three scenarios for their denotation (Fig. 2.1).

The first scenario that these microorganisms originated from known species, whose cell size decreases over time due to either internal and/or external factors such as lack of nutrients or ageing (Velimirov, 2001; Panikov, 2005; Duda et al., 2012). Such ability of bacteria and archaea to change size in response to external stress is a well-studied phenomenon. For example, under low nutrient conditions, *Staphylococcus aureus* reduced its size by 40% (Watson et al., 1998; Chien et al., 2012), while the transfer of *Pseudomonas syringiae* from laboratory culture media to plant leaves, induced the 50% reduction in cell size (Monier and Lindow, 2003). This size reduction is an attribute of dwarf cells, midget cells, ultra-small, ultramicro (Velimirov, 2001; Duda et al., 2012). For these cases, we advocate for the term 'ultramicrocells' *sensu* Duda et al. (2012).

The second scenario conjunctures that some distinct taxa, independently of growth conditions, nutrients' availability or age of their culture do constantly exhibit small cell sizes. One source describes these organisms in the following way: the microorganisms must be 0.1 μm^3 or smaller (<0.05-0.40 μm in diameter); the size must stay consistent under environmental

stressors and life cycles; and finally, its genome size must be within the range 0.58 Mbp to 3.2 Mbp (Duda et al., 2012). Under this definition, nano-sized microorganisms are associated with terms like ultra-small, ultramicronanoarchaea, nanoforms, nanoorgansims, and nanobacteria (Schut et al., 1995; Kajander and Ciftcioglu, 1998; Velimirov, 2001; Huber et al., 2002; Miteva and Brenchley, 2005; Panikov, 2005; Comolli et al., 2009; Duda et al., 2012; Fedotova et al., 2012; Luef et al., 2015; Giovannoni, 2017; Rogge et al., 2017). However, many standard-sized microorganisms (i.e. cell volumes >0.1 μm^3) also possess small genomes (1.5-2.0 Mbp) and would therefore fall into the 'ultra-small' category if based on these criteria alone.

The third scenario are microorganisms that have the ability to pass through membrane filter pores with small diameters (0.45 or 0.22 μm) despite having larger cell sizes (above the dimensions of 50-400 nm previously mentioned) (reviewed in Duda et al., 2012). This is often due to the absence of a rigid cell wall, which allows these microorganisms to effectively squeeze through small pores and as a result are commonly confused with nano-sized or ultramicro-sized. 'Filterable' microorganisms is the most appropriate term to define such microorganisms.

In this review, a unified definition for nano-sized organisms is proposed. We define them as microorganisms that exhibit constant dimensions of 50-400 nm (volume $\leq$ 0.1 μm^3). All microorganisms with synonymous names that fall under the definition provided are considered nano-sized organisms. Viruses and prions, which are smaller than 50 nm in size, are not considered to be living organisms (Fig. 2.2; Table 2.1). In aquatic systems, nano-sized organisms are a part of the picoplankton and femtoplankton communities, along with viruses (Venter et al., 2004; Tringe et al., 2005; Sieburth et al., 1978; Salcher, 2014).

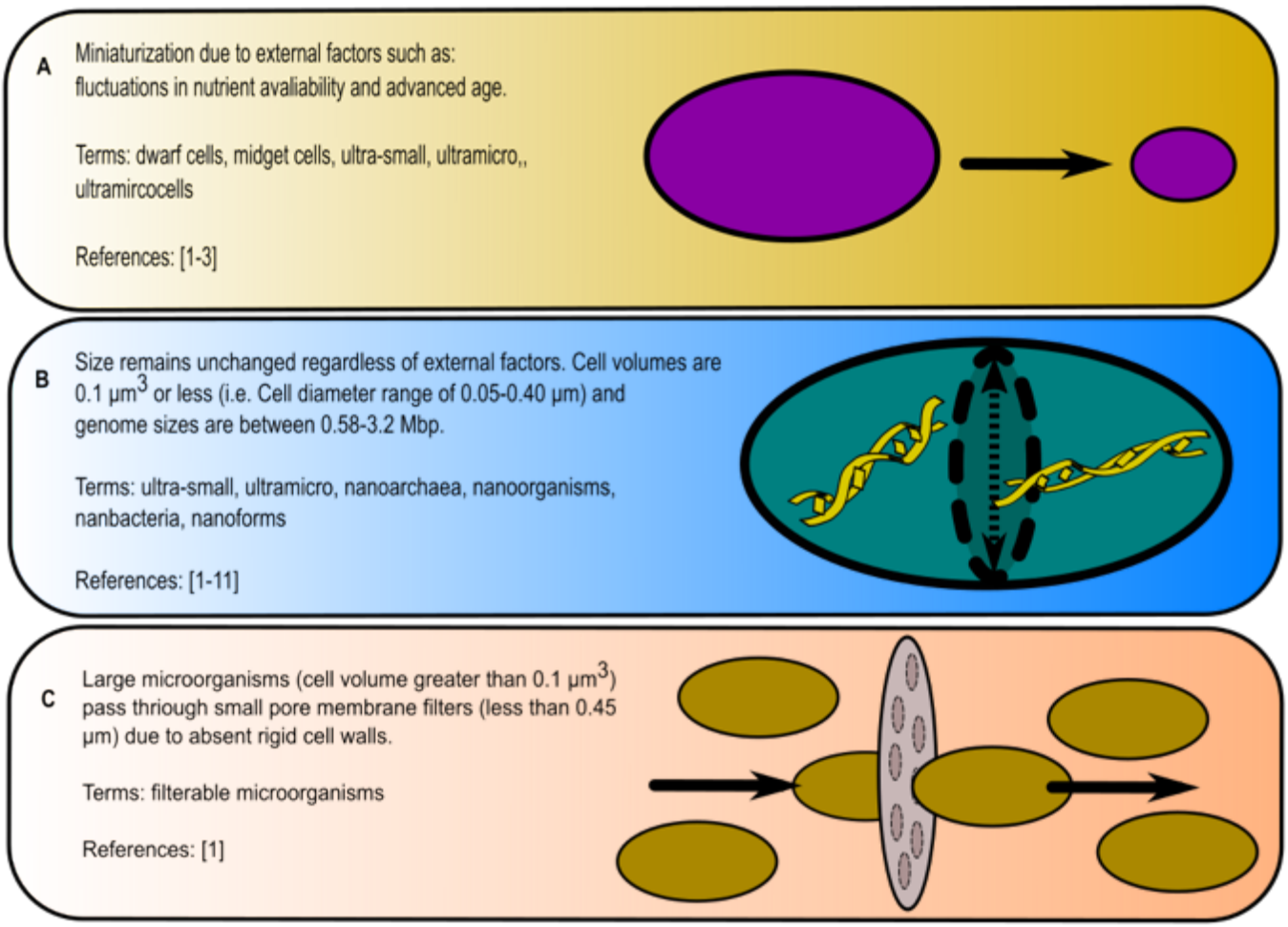


Figure 2.1- Summary of definitions used to describe nano-sized organisms. References are the following: [1] Duda et al. (2012), [2] Verlmirov et al. (2001), [3] Panikov (2005), [4] Shut et al. (1995), [5] Miteva and Brenchley (2005), [6] Luef et al. (2015), [7] Huber et al. (2002), [8], Rogge et al. (2017), [9] Giovannoni (2017), [10] Kajander and Ciftcioglu (1998), [11] Fedotova et al. (2012).

2.2. Microbial adaptations

In the natural environments, microorganisms use an arsenal of mechanisms to cope with, and adapt to, constantly changing physio-chemical conditions, through changes in their gene expression profile, physiology and morphology (Schulz and Jørgensen, 2001; Chien et al., 2012). Here we highlight various survival strategies in prokaryotes, knowledge of which may stimulate future discoveries pertaining to small-sized organisms.

2.2.1. Extremely small size

In general, microorganisms do not fit into one standard model of size or shape (morphology) due to the impact environmental stressors (Young, 2006; Chien et al., 2012; Cesar et al., 2015; Lever et al., 2016). The efficiency of nutrients' uptake is dependent on organism size and the number of transporter systems on its surface (Button et al., 1998). Hence, in the case of cell size reduction, the surface area-to-volume ratio tends to increase (Fig. 2.2). This, however, does not imply that the percentage of genes encoding membrane-bound proteins in genomes is higher in organisms with a larger surface area-to-volume ratio (Stevens and Arkin, 2000) (Fig. 2.2).

Under conditions of starvation and energy limitations, microorganisms can drastically decrease in size, alter cellular morphology and motility to increase survivability (Torrella and Morita, 1981; Cesar and Huang, 2017; Lever et al., 2015). For example, in low organic phosphate conditions, *Caulobacter* spp. increase their surface area to volume ratio by growing a prosthecae, stalk-like protrusions, in order to enhance organic phosphate uptake (Wagner et al., 2006; Lever et al., 2015). Another example is the species *Sphingomonas alaskensis*, which also undergoes morphological changes in response to the fluctuations in nutrients availability. In its natural pelagic environment its body size is quite small (diameter 0.2-0.5 μm; length 0.5-3 μm) yet when grown on nutrient rich trypticase soy agar medium it increases in both diameter and length (diameter 0.8; length 2-3 μm) (Vancanneyt et al., 2001; Lever et al., 2015).

2.2.2. Lifestyle: free-living vs symbionts

Nano-sized organisms are thought to contain genomes coding for a very limited number of functions and pathways, which is a characteristic commonly associated with symbionts, however, nano-sized organisms do also exist in a free-living state. Generally, symbionts do not have the means for their existence without relying on essential metabolites provided by the host. However, these organisms do thrive probably due to their highly specialized and unique

functions which allows the host to be more competitive (McCutcheon and Moran, 2011). For instance, TM7 ("*Ca*. Saccharibacteria") bacteria isolated from the human oral mucosa can effectively conceal its host, *Actinomyces odontolyticus* subsp. *actinosynbacter* XH001, from the human immune system response (He et al, 2015; further discussion in the section " TM7 bacteria or '*Candidiatus* Saccharibacteria').

2.2.3. Oligotrophy and Copiotrophy

Oligotrophs also known as K-strategists, are organisms that prefer low-nutrient environments (Schut et al., 1997; Panikov, 2005; Torsvik and Øvreås, 2008). One of the most well-characterized oligotrophic environments is the open ocean, which encompasses 90% of the biosphere (i.e. the sum of all the ecosystems) (Schut et al., 1997; Hansell et al., 2009). In this environment, many essential nutrients are only present in very low concentrations: iron at 0.2-1.38 nmol kg^{-1}, nitrate at 1.04 µmol kg^{-1}, phosphate at 0.074 µmol kg^{-1}, silicate at 3.2 µmol kg^{-1}, dissolved inorganic carbon at 11 µmol kg^{-1}, and dissolved organic carbon at 40-80 µmol kg^{-1} (Johnson et al., 1997; Roshan and DeVries, 2017; Sauzède et al., 2017; Tagliabue et al., 2017), which makes it difficult to mimic such conditions and obtain a detectable growth of these microorganisms *in vitro*. At such low concentrations of nutrients microorganisms lower their metabolic rates and become less capable of forming aggregates (i.e. colonies), as seen in many pelagic organisms, such as SUP05 group bacteria and in "*Ca*. Pelagibacter ubique" (see references below in the sections 'SUP05 group' and "*Ca*. Pelagibacter ubique"). Overall, oligotrophs are characterized by small cell sizes, which are more advantageous in low nutrients conditions. The correlation between oligotrophy and diminutive size appears almost ubiquitously (Giovannoni et al., 2014), however, few studies have detected ultra-small-sized microorganisms in high-nutrient systems, such as eutrophic aquifers or the human oral cavity (Luef et al., 2015; He et al., 2015).

Copiotrophs or R-strategists, are active, fast-growing with larger cell body sizes, usually motile organisms well-suited to nutrient-rich environments; they represent the majority of bacteria and archaea cultured up to date (Dang and Lovell, 2016; Giovannoni, 2017). Despite being easy to culture, copitotrophs appear as rarer taxa in natural environments. They take advantage of sporadic high nutrients concentrations which in turn may transiently cause a rapid population growth (Vergin et al., 2013; Dang and Lovell, 2016). It is thought that copiotrophs are not nano-sized organisms as an increased surface area-to-volume ratio is not necessarily advantageous in nutrient-rich environments (Martínez-Cano et al., 2015). However, copiotrophic bacteria also tend to reduce their sizes as a response to starvation conditions in an attempt to increase their surface area-to-volume ratio, as in the case of *S. aureus* (40% reduction in size) and *P. syringae* (50% reduction in size) (Watson et al., 1998; Monier and Lindow, 2003).

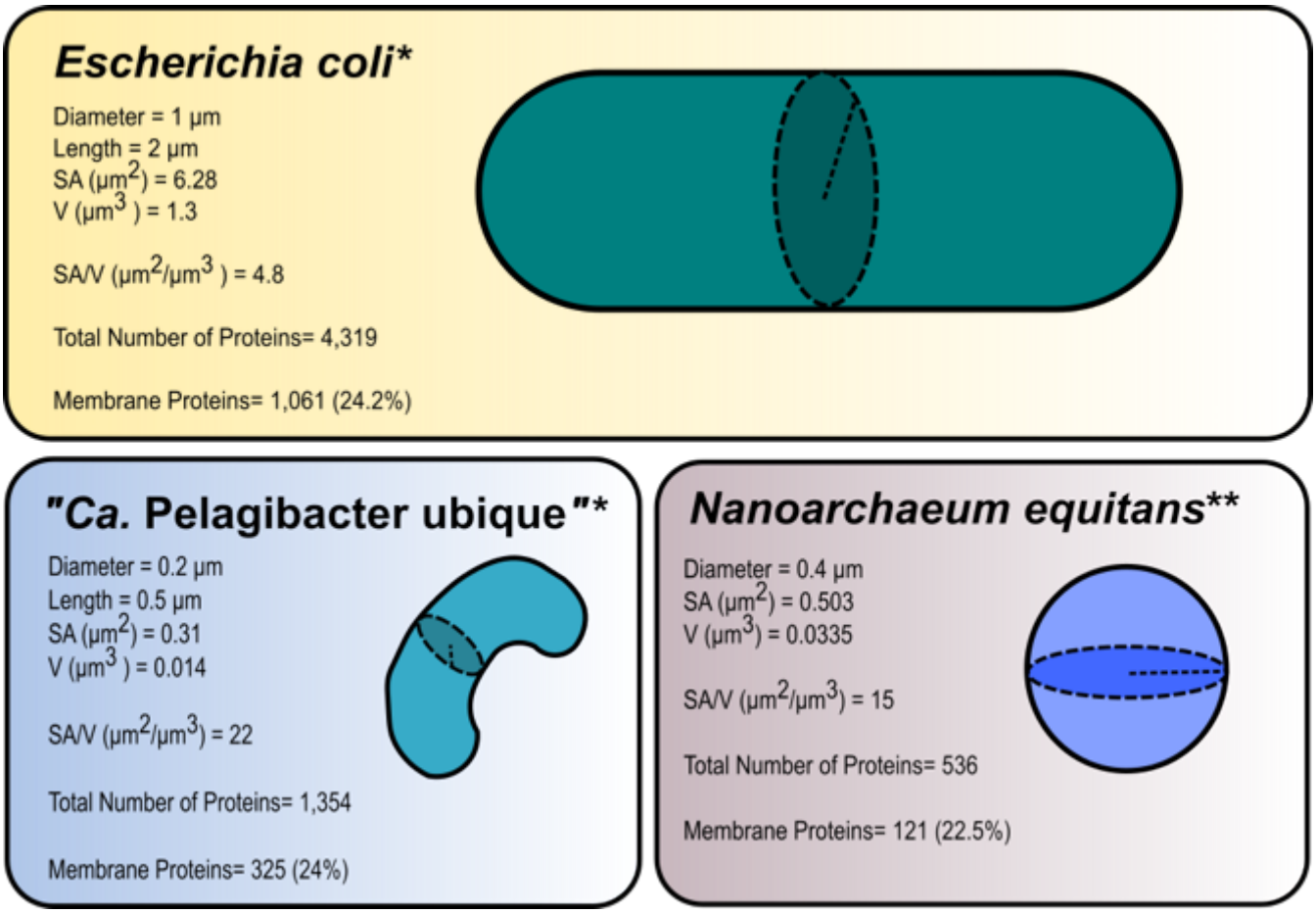


Figure 2.2- Surface area (SA) and volume (V) ratios in three selected species of different sizes: Escherichia coli, "*Candidatus* Pelagibacter ubique", and Nanoarchaeum equitans. The microorganism with the smallest dimensions ("Ca. P. ubique") had the largest ratio at 22. The habitat of "Ca. P. ubique" is the open ocean (oligotrophic environment) and hence its high SA/V ratio is advantageous to living in low nutrient conditions. The total protein numbers in encoded by genomes of E. coli (NCBI Reference Sequence: NC_000913.3), "Ca. P. ubique" (GenBank: CP000084.1), and N. equitans (GenBank: AE017199.1) are given and related with the proteins with membrane-spanning domains. For prediction of transmembrane helices in proteins, above genomes were analysed using TMMHMM 2.0 Server at http://www.cbs.dtu.dk/services/TMHMM/ (Krogh et al., 2001; Möller et al., 2001). Notes: * Dimensions and calculations of surface area and volume were obtained from Young (2006). ** The diameter was obtained from Huber et al. (2002), the equations for the surface area ($SA= 4\pi r^2$, where r is the radius) and volume ($V= \frac{4}{3}\pi r^3$, where r is the radius) of a sphere.

2.3. Characterization

Due to the constraints in accurately mimicking environmental settings *in vitro,* the cultivation of small organisms is often problematic and represents a main bottleneck in the process of their phenotypic characterization. In order to predict functional traits of nano-sized microorganisms as a part of the microbial community, culture-independent techniques are currently employed as primary approaches, as stand-alone or combinations of approaches: metagenome sequencing, flow-cytometry and fluorescence microscopy. Below is a brief overview of some culture-independent techniques and the challenges that arise when attempting to isolate nano-sized microorganisms.

2.3.1. Metagenomics

As indicated above, metagenomics has played a central role in attempts to detect small-sized and filterable organisms and elucidate their functions. In turn, the isolation and characterization of nano-sized organisms has yielded, and to some extent, validated new genomic data (Huber et al., 2002; Giovannoni et al., 2005). In many of the large-scale metagenomics studies, the significant proportion of assembled genomes exhibited small sizes (Rappé et al., 2002; Venter et al., 2004). In particular, an in-depth investigation of the SAR11 clade led to the discovery of "*Ca*. Pelagibacter ubique*",* a ubiquitous and predominant marine bacterium (Giovannoni, 2017; Zhao et al., 2017) . Also, microbial communities in the deep biosphere proved to be more diverse than previously anticipated, with a plethora of miniature cells with small genomes (Wu et al., 2015). Finally, hypersaline lakes, a good model for extreme habitats, were found to contain filterable cells, about 0.6 μm in diameter, that were termed "*Ca.* Nanohaloarchaeota" (Narasingarao et al., 2012). This study was in large facilitated by a more targeted sample preparation (filtration) procedure and *de novo* sequencing approach. However, we must note that small genomes and the ability to pass via 0.1, 0.22, and 0.45 μm pore-size filters are not necessarily the evidence of small sizes of microorganisms (i.e. filterable

microorganisms), for instance, the symbiont "*Ca.* Tremblya princeps*"* has an extremely reduced genome of 0.13 Mbp, yet, examination by microscopy showed its length to be ca. 2.3 μm (McCutcheon and Moran, 2011).

2.3.2. Flow cytometry and FACS cell sorting

The further culture-independent techniques, flow cytometry (Gasol and Morán, 1999; Miteva and Brenchley, 2005; Wang et al., 2007; Neuenschwander et al., 2015) and fluorescence in situ hybridization (FISH) (Glaubitz et al., 2013; Neuenschwander et al., 2015; Munson-McGee et al., 2015;) have been widely used to study microbial populations in their natural environments. In combination with fluorescence probes targeting SSU rRNA or immunolabelling cellular proteins, this approach allows quantification of a certain taxonomic group of microorganisms (Neuenschwander et al., 2015). Combining FISH/CARD-FISH (Fluorescence In Situ Hybridization/Catalyzed Reporter Deposition-Fluorescence In Situ Hybridization) and flow cytometry (also known as 2C-FISH) allowed for sorting and obtaining relatively pure populations of microorganisms, as it was the case of LD12 clade of ultramicrobacteria from freshwater. These ultramicrobacteria were known to be very difficult to isolate and characterize due to their small genomes and hence limited metabolic repertoires, cell sorting was therefore the crucial starting point for their subsequent genomic studies (Salcher et al., 2013; Neuenschwander et al., 2015). Although improvements in individual techniques were achieved in this study, the methodology of sample preparation is still tedious and time-consuming with relatively limited yields of cells (Neuenschwander et al., 2015). Whatever the case, the applications of cell sorting have been successful in resolving a number of "single-cell-genomes" (Ishoey et al., 2008; Probst et al., 2018).

2.3.3. Isolation of nano-sized microorganisms

Although isolation is an essential step in characterizing organisms, it is often overlooked and traditional approaches to culture them frequently prove unsuccessful. Many of the studies presented in this review employed filtering through 0.1-1.2 μm pore size filters to facilitate enrichment and isolation (Table 2.1). The exception to the filtration methodology was *Nanoarchaeum equitans*, which was co-cultured with the host, *Ignicoccus hospitalis,* and then separated out via centrifugation (Huber et al., 2002; Waters et al., 2003). Conversely, while the target microorganisms may be small enough to pass through the membrane, certain larger organisms can squeeze through pores, due to a lack of rigidity of their cells. Another example of organisms squeezing through small-sized pores are archaea of families *Ferroplasmaceae* (0.2-3 μm in diameter in average) and *Thermoplasmataceae* (0.5-3 μm in length and 0.2-0.5 μm thick), that can easily pass through a <0.45 μm pore filter due to the lack of a rigid cellular envelope (Golyshina, 2014; Nagy et al., 2016).

In previous studies, along with 'small-sized-organisms', many other microorganisms have been co-isolated (Venter et al., 2004; Tringe et al., 2005; Garza and Dutilh, 2015). An extra level of authentication is therefore necessary to reliably confirm the existence and metabolic function of these organisms, e.g. through an improvement in isolation and culturing techniques. Small cell size is the only certainty related to nano-sized organisms that belong to a range of taxa and do not share a common metabolism. For their characterization, a prior genomic analysis of the source community is critical. This would allow the targeting e.g. organism-specific surface proteins to enable FACS- or immunoprecipitation-based techniques targeted organisms of interest.

2.4. Nano-sized and filterable microorganisms

Though the different characterization techniques as mentioned above, the story of ultra-small microorganisms and our understanding of their ecosystem functioning is rapidly evolving. Here, some of the major milestones are outlined in regards to successful isolation and characterization of a variety of nano-sized organisms. Further, we have summarized the data on various microorganisms covered in this section in Table 2.1 and Figure 2.3.

Table 2.1- An overview of small-sized and filterable organisms denoting average cell size, average genome size, environment, separation technique (filter pore sizes), cultivability, affiliation to a confirmed species, lifestyle (free living or host-dependent), and corresponding references. NA denotes information not available within the respective source. Some studies showed that the results were inconclusive meaning that there were conflicting conclusions in the literature. *Colonies were slow-growing, taking up to a few months to become visible. **Proposed Candidatus status. †Parasitic ultramicrobacteria discussed in Duda et al. (2012) review.

Small--sized organism(s)	***Environment***	***Average genome size***	***Average/range cell size***	***Free-living?***	***Filter(s) pore size used***	***Cultured?***	***Validly published Species***	***Reference***
'Ca. Pelagibacter ubique'	Open ocean	1.3 Mbp	0.01 μm^3 (volume)	Yes	0.2 µm	Yes	Yes	Rappè, et al. (2002), Giovannoni et al. (2005), Carini et al (2012), Zhao et al. (2017), Giovannoni (2017)
Nanoarchaeum equitans	Submarine hot vent	0.5 Mbp	0.4 µm (diameter)	No	None	Yes	Yes	Huber et al. (2002) Waters, et al. (2003), Jahn et al (2008)
Ultrasmall Microorganisms	120,000 year old Greenland ice core	NA	<0.10 μm^3 (volume)	NA	0.4 µm, 0.2 µm, and 0.1 µm	Yes*	No	Miteva and Brenchley (2005)
ARMAN cells	Acid mine drainage biofilm	1 Mbp	0.03 μm^3 (volume)?	Inconclusive	0.45 µm	No	No	Comolli et al. (2009), Comolli and Banfield, (2014), Baker, et al. (2010), Baker, et al. (2006)
'*Ca.* Nanobsidianus. stetteri'	Obsidian Pool, Yellowstone National Park	0.651 Mbp	NA	No	0.4 µm	No	No	Podar, et al (2013), Munson-McGee, et al. (2015)
Oral TM7 'Ca. Saccharibacteria'	Human oral cavity	0.705 Mbp	200-300 nm (diameter)	No	0.22 µm	Yes	No	He, et al. (2015)

"*Ca.* 'Nanopusillus acidilobi'	Cistern Spring, Yellowstone National Park	0.605 Mbp	100-300 nm (diameter)	No	0.1 µm	Yes	No**	Wurch, et al. (2016)
WWE3/OP11/OD1 groundwater ultra-small bacteria	Anoxic aquifer	0.878 Mbp (WWE3) 0.694 Mbp (OD1) 0.820 to 1.050 Mbp (OP11)	0.009 μm^3 (volume)	No	1.2 µm, 0.2 µm, and 0.1 µm	No	No	Luef, et al. (2015), Wrighton, et al. (2012), Kantor, et al. (2013)
'Nanobacterium sanguineum'	Human and bovine blood	NA	50 nm (diameter)	NA	0.1 µm	Inconclusive	No	Kajander and Ciftcioglu (1998), Kajander and Ciftcioglu (2003), Cisar et al (2000),
Fossil remains	Meteorite ALH84001	NA	10-200 nm (length)	NA	NA	NA	No	McKay et al. (1996), McKay et al. (2001)
SUP05 Bacteria	Pelagic redox zones	1.164 Mbp to 1.53 Mbp	0.01-0.09 μm^3 (volume)	Yes	0.2 µm	No	No	Rogge et al. (2017), Glaubitz et al. (2013), Shah et al. (2017)
Filterable forms	Lake Motykino and Lake Dubrovskoe (Peatland bog)	NA	0.3-0.5 µm (rod diameter)	NA	0.22 µm	No	No	Fedotova et al. (2012)
Aurantimicrobium minutum Str. KNCT	River water	1.62 Mbp	0.04-0.05 μm^3 (volume)	Yes	0.22 µm	Yes	Yes	Nakai et al. (2016)
Curvibacter sp. Str. PAE-UM	River sediment	3.28 Mbp	<0.05 μm^3 (volume)	Yes	NA	Yes	Yes	Ma et al. (2016)
Free-living Ultramicroscopic bacteria	Natural biotopes (i.e., permafrost, oil slime, soil, lake silt, thermal swamp moss, Xenopus laevis, skin	1.5-2.4 Mbp	0.02-1.3 μm^3 (volume)	Yes	NA	No	No	Suzina et al. (2015)
Bdellovibrio spp.†	NA	3.78 Mbp	0.13 μm^3 (volume)	No	NA	Yes	Yes	Duda et al. (2012)
Micavibrio admiranndus†	NA	NA	0.05 μm^3 (volume)	No	NA	Yes	Yes	Duda et al. (2012)
Vampirovibrio chlorellavorus†	Reservoir water	NA	0.3-0.6 µm (diameter)	No	NA	Yes	Yes	Duda et al. (2012)
Kaistia adipata, str. NF1, NF3†	Soil and lake sediment	2.4 Mbp	0.1-0.5 μm^3 (volume)	No	0.22 µm	Yes	Yes	Duda et al. (2012)
Chryseobacterium solincola, str. NF4, NF5†	Soil and lake sediment	1.7 Mbp	<0.1 μm^3	No	0.22 µm	Yes	Yes	Duda et al. (2012)

2.4.1. Rise of the very small

Although ultramicrobacteria have been known for a long time (Oppenheimer, 1952), the subject laid dormant for a number of years. This was in part due to the limitations in microbiological techniques, and the lack of knowledge of their physiology and metabolism. That changed when McKay et al. (1996) first claimed their existence in Martian rocks. Not only did this imply that life may exist on exoplanets, but it also challenged the ideas on lower limit of size of a lifeform (McKay et al., 1996; Gibson et al., 2001). It was suggested that the smallest free living organism must be in the spherical diameter range of 250-300 nm to properly contain the 250-300 proteins essential to life (including the ribosomal proteins), although it was also suggested that, theoretically, a primitive organism can be as small as 50 nm (Kajander and Ciftcioglu, 1998). This was similar to an earlier study by Mushegian and Koonin (1996) who hypothesized that the minimal number of genes required for life ranges between ca. 250-450, however, there was no consensus on the number of ribosomal proteins that were actually needed. Importantly, it was never established in the McKay et al. (1996) study whether these nano-scale objects were free-living organisms, nor was it confirmed that these objects were living at all.

2.4.2 Nanoarchaeum equitans

Huber et al. (2002) found that a new archaeal species, *Ignicoccus hospitalis,* isolated from hot submarine vents, had in its culture a companion of a small cell size. The new phylum *Nanoarchaeota* and corresponding species *Nanoarchaeum equitans* were described as the first nano-sized archaea. The genome analysis revealed that it contained a chromosome of only 0.5 Mbp (Huber et al., 2002), while electron and fluorescence microscopy suggested that the cells of *N. equitans* were ca. 400 nm in diameter and were attached to the cell surface of its host, *I. hospitalis.* Further, it was shown that *N. equitans* was incapable of growing without its host, which in contrary neither benefited or was impaired by *N. equitans* (Huber et al., 2002; Jahn et

al., 2008). The inability of *N. equitans* to survive without its host is reflected in its small streamlined genome, which was a result of massive gene losses (Huber et al., 2002) including those for key biosynthetic pathways for vitamins, cofactors and amino acids (Torrella and Morita, 1981; Mushegian and Koonin, 1996; McCutcheon and Moran, 2011).

2.4.3 "ARMAN" cells

"ARMAN" (Archaeal Richmond Mine Acidophilic Nanoorganism) were first detected through *de novo* shotgun sequencing of aqueous sample obtained from an acid mine drainage (AMD) system and not through standard PCR-based surveys (Baker et al., 2006). Subsequent cryo-TEM analysis revealed an accumulation of filterable cells that were 0.03 μm^3 in volume with clearly defined cell walls (Comolli et al., 2009). "ARMAN" cells were initially considered free-living, possibly slow-growing, organisms possessing some intracellular tubular structures (Comolli et al., 2009), however, later on, their ability to free-living lifestyle was questioned (Comolli and Banfield, 2014).

According to the metagenome analysis with almost fully assembled "ARMAN" genomes of ca. 1 Mbp in size and proteomics, these organisms contain a rather unique set of genes with 45% of the genes failing to match to a known biological function, while 63% of the proteins identified could not be assigned to known archaeal protein families (Baker et al., 2010). Due to the small sizes of their genomes, it was assumed that "ARMAN" cells are certainly dependent on other community members, being either symbionts or commensals (Baker et al., 2010).

Cultivation of an "ARMAN"-related organism, '*Ca.* Mancarchaeum acidiphilum' Mia14 revealed that it was dependent on its host, euryarchaeon *Cuniculiplasma divulgatum* (Golyshina et al., 2017). As in the above examples, Mia14 underwent streamlining of its genome (0.95 Mbp) due to the massive gene loss. Similarly, it exhibits significant voids in its biosynthesis of amino acids, CoA, NAD and NADP, vitamins and heme. Additionally, its

central metabolism lacks glycolysis and gluconeogenesis, pentose phosphate pathway and tricarboxylic acid cycle (Golyshina et al., 2017). Interestingly, Mia14 cell sizes were only marginally smaller than *Cuniculiplasma* cells, which were 0.1 to 2 μm in size (Golyshina et al., 2016).

2.4.4. Other Archaea

"Candidatus Nanobsidianus stetteri" Nst1, a member of phylum *Nanoarchaeota* was first reported after the single-cell isolation alongside its host from the order *Sulfolobales* (phylum *Crenarchaeota*) by Podar et al. (2013). Unlike *N. equitans*, which is associated with a single host species, *I. hospitalis*, "*Ca.* N. stetteri" can use a multitude of *Sulfolobales* species as hosts. Its genome was ca. 20% larger than that of *N. equitans* and possessed a complete gluconeogenesis pathway (Podar et al., 2013; Munson-McGee et al., 2015). The genome analysis also indicated that "*Ca*. N. stetteri" genome coded for cellular functions previously not associated with the *Nanoarchaeota* taxon; the study concluded that these archaea share a common ancestor with *N. equitans* (Podar et al., 2013; Munson-McGee et al., 2015). Another study (Munson-McGee et al., 2015) has partially resolved two further single-cell genomes of "Nanobsidianus"-related archaea from Yellowstone hot springs and suggested their close relatedness with "*Ca.* N. stetteri" Nst1, but pointed at their association with archaea of "Acidicryptum spp." of *Sulfolobales*. "*Ca.* Nanopusillus acidilobi" is another success story, where this small-sized, reduced-genome archaeon was co-cultured with its host, *Acidilobus* sp. A7 by Wurch et al. (2016). "*Ca.* Nanopusillus acidilobi" is a thermophilic ectosymbiont, much like *N. equitans* and "*Ca*. Nanobsidianus stetteri". This particular species is only marginally smaller in body size than *N. equitans* (approximately 100-300 nm in diameter), both share approximately 80% SSU rRNA gene sequence identity (and 97-98% with *'Ca*. Nanobsidianus stetteri'), and exhibit much of the same functions as judged from genomic data (Wurch et al., 2016). "*Ca.* Nanopusillus acidilobi" genome possesses no genes related to respiration, ATP

synthesis and cannot produce its own amino acids, lipids, nucleic acids, and co-factors. Genomic data suggests that, like in its relative, "*Ca.* N. stetteri", glycogen may serve as a storage compound and facilitate its short-term energetic independence from the host (Wurch et al., 2016). A high density of "*Ca.* Nanopusillus acidilobi" on the surface of its host *Acidilobus* sp. 7A, deficiency of its genome in genes for central metabolic, biosynthetic and energy-generating pathways suggest a commensal or ectoparasitic lifestyle of these nanoarchaea (Wurch et al., 2016). Expression of flagellar proteins reported in proteomic data further suggests that "*Ca.* Nanopusillus acidilobi" has the ability to migrate from one host to another (Wurch et al., 2016).

2.4.5. "Ca. Pelagibacter ubique"

While the existence of oceanic ultramicrobacteria has been well documented, obtaining them in a pure culture remained difficult. Earlier studies (Rappé et al., 2002; Morris et al., 2002) revealed a very abundant clade of *Alphaproteobacteria*, SAR11, which makes up to 25% of plankton in the open ocean and is represented by small-sized, simple-metabolism bacteria (Giovannoni, 2017) . Initially found in pelagic water sampled from the Sargasso sea, these bacteria termed "*Ca*. Pelagibacter ubique*"* had genomes of approximately 1.3 Mbp and are considered to be one of the smallest free living cell (Giovannoni, 2017; Zhao et al., 2017). Their genomes contained the necessary gene sets for producing all 20 amino acids as well as other essential biosynthetic pathways (Giovannoni et al., 2005; Carini et al., 2012). Subsequent studies indicated that "*Ca*. P. ubique*"* required an unconventional medium, which was composed of methionine, glycine, pyruvate, and artificial seawater (Carini et al., 2012).

It was also found that "*Ca*. P. ubique*"* had a rather unique metabolism because of its ability to use glycolate instead of glycine at low glycine concentrations. Glycolate can be used in glycine biosynthesis through glyoxylate amination, with the glycine consequently being used for serine biosynthesis (Carini et al., 2012; Tripp, 2013). The glycolate to serine pathways are

regulated by two glycine riboswitches, the first of which controlling the glyoxylate to glycine biosynthesis and the second regulating the glycine to serine biosynthesis. At low glycine concentrations, the first riboswitch is turned on to produce more glycine (Tripp, 2013). When there are ample amounts of glycine in the cell, the first riboswitch turns off the glycine biosynthesis and the second riboswitch induces the conversion of glycine to serine. The ability to use glycolate instead of glycine to further create serine may be an evolutionary response to relative excesses of glycolate formed by phytoplankton in carbon limited conditions (Carini et al., 2012). As a free-living organism, "*Ca*. P. ubique*"* has the ability to adapt to changing conditions fairly well despite having a streamlined genome. It also challenged the previous assumption that small genome sizes were restricted to symbiotic organisms (Huber et al., 2002; Giovannoni, 2017).

2.4.6. SUP05 group

Oxygen-depleted zone in pelagic systems with dissolved oxygen concentrations below 60 μmol kg^{-1} present a unique challenge to organisms moving through the transition zone from high to low nutrient availability (Glaubitz et al., 2013; Rogge et al., 2017). According to cell counts from flow cytometry, SUP05 bacteria are a common bacterioplankton component in depleted oxygen zones (Glaubitz et al., 2013; Rogge et al., 2017). As chemolithoautotrophic organisms, they metabolize sulfur compounds and play a key role in the carbon, sulfur and nitrogen cycles to facilitate life in the redoxclines across the globe (Glaubitz et al., 2013; Rogge et al., 2017; Shah et al., 2017). They have the ability to carry out denitrification and uptake carbon dioxide in pelagic low oxygen zones, which is supported by genomic predictions, radioisotopic data and cultivation attempts (Glaubitz et al., 2013; Rogge et al., 2017; Shah et al., 2017). Cultivation attempts of one of the members of the SUP05 group, "*Candidatus* Thioglobus autotrophicus", revealed the utilization of ammonium under anaerobic conditions and nitrite production (Shah et al., 2017). Studies on

the SUP05 group have suggested cellular volumes ranging within 0.01-0.09 μm^3 and a genome of 1.164-1.53 Mbp, which indicates that these bacteria have undergone streamlining in their evolutionary past, much like "*Ca*. P. ubique*"* (Rogge et al., 2017; Shah et al., 2017).

2.4.7. Filterable forms in peatland bogs

Despite the abundance of organic carbon in aquatic subsystems of peatland bogs, its mineralization is very slow due to the elevated concentrations of phenolic compounds causing acidification (pH 4.4-4.8), enzyme inhibition and nitrogen limitation (Fedotova et al., 2012). This is the case for sphagnum peatland bogs in northern Russia, that contain a high number of filterable bacteria and archaea, $1.69 \pm 0.53 \times 10^4$ and $3.16 \pm 0.43 \times 10^4$ cells/mL, correspondingly (Fedotova et al., 2012). Phylogenetic analysis of 16S rRNA genes shows they were derived from several phyla (Fedotova et al., 2012). One-third of the archaeal sequences had a high identity (94-99%) with representatives of the orders *Methanobacteriales* and *Methanosarcinales,* while the rest exhibited a distant relatedness (71-74% sequence identity) to cultured methanogens and collectively belonged to the LDS (Lake Dagow sediment) cluster (Glissmann et al., 2004). All detected bacterial species had high SSU rRNA gene sequence identities (94-99%) to the *Betaproteobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, and *Actinobacteria,* which confirms that small size is an adaptation to low nutrient conditions common across the broad range of higher taxa. The study also attempted to culture filterable microorganisms on solid media: from the total microscopic cell count numbers, only a fraction of approx. 0.5-1.2% did form colonies represented by bacterial genera *Mesorhizobium, Bradyrhizobium, Sphingomonas* and *Agrobacterium*. A major discrepancy between the SSU rRNA amplicon libraries sequences of microbial communities in those freshwater samples and the taxonomy of cultured bacteria was also observed (Fedotova et al., 2012).

2.4.8. Ultra-small bacteria from Greenland ice

Glacial ice presents a rather unique challenge to many microbial species due to its sub-zero temperatures and oligotrophic conditions and is considered a freshwater-like habitat for microorganisms (Hodson et al., 2008). It has been previously noted that a number of ultrasmall organisms have been detected in several ice cores (Miteva, 2008). A plethora of bacteria in 120,000 year-old Greenland ice, which, after melting the ice cores, passed through filters with pore sizes of 0.4, 0.2 and even 0.1 μm was detected (Miteva and Brenchley, 2005). Scanning electron microscopy and flow cytometry confirmed that the filtration methodology was effective at removing larger cells residing in the melted ice water. The authors also stated that a considerable amount of fungal colonies were also present, although these were not discussed in further detail (Miteva and Brenchley, 2005), however, one can assume those were derived from filterable fungal spores. It is not clear if all >1,200 cultured bacteria were ultra-small, as there was evidence of larger organisms (e.g. spores of fungi and of *Firmicutes*), which possibly were cultured due to the non-uniform sizes of filter pores, over-pressurizing filtration units or non-rigid cell envelops of microorganisms that allowed them passing through filters (Wang et al., 2007, 2008). Whatever the case, the study of Miteva and Brenchley (2005) clearly demonstrated the viability in and cultivability of very small microorganisms with experimentally measured average volumes ranging between 0.043-0.1μm^3 from, a polar ice environment.

2.4.9. WWE3, OD11 and OP1 candidate phyla of ultra-small bacteria from groundwater

Much of the bacterial species discussed so far have been identified in oligotrophic environments, however, ultra-small organisms are not exclusive to these habitats. The WWE3-OD11-OP1 candidate phyla of groundwater bacteria were found in an eutrophic environment (Luef et al., 2015). Although these bacteria have not been cultivated, ultra-small cells have been successfully imaged challenging previous ideas on possible habitats of these organisms.

Luef et al. (2015) described the cellular structures present within ultra-small-sized-organisms: using cryo-TEM images they identified pili, cell walls, cellular division and the presence of viruses. The study investigated the freshwater collected from an anoxic, organic carbon rich groundwater located several meters below the surface. Until that point, small-sized microorganisms were thought to be either associated with oligotrophic conditions or microbial communities with a reduced diversity, e.g. AMD. Importantly, it appears that small size can also be beneficial in other environments. The study was unable to successfully perform CARD-FISH on the proposed ultra-small cells (Luef et al., 2015) and therefore could not confirm that small cells seen were indeed of the candidate phylum that they reported on.

Metagenomic analyses by Wrighton et al. (2012) and Kantor et al. (2013) have revealed that WWE3, OP1, OD11, TM7, and SR1 candidate phyla of bacteria possessed small genomes, lacked genes for several essential metabolic processes and contained genes of both archaeal and bacterial origin. The genomic predictions inferred that WWE3, OP1, and OD11 candidate phyla are capable of growing in organic carbon-rich environments (Wrighton et al., 2012; Luef et al. 2015; Kantor et al. 2013). The RuBisCO (type II/III ribulose-1, 5-biphosphate carboxylase-oxygenase), which was predicted in these groundwater ultrasmall bacteria, is not likely to be involved into the classical CBB (Calvin-Benson-Bassham) pathway, but into the CO_2 fixation linked with the AMP (adenosine monophosphate) recycling for ultimate ATP (adenosine triphosphate) production, similarly to the type III archaeal RuBisCo (Kantor et al., 2013; Wrighton et al. 2012). The occurrence of this pathway suggests that these organisms are not restricted to oligotrophic environments, but can survive with higher levels of available nutrients.

2.4.10. TM7 bacteria or "Candidatus Saccharibacteria"

Recent studies have shown that nano-sized organisms can also be a component of the human microbiome. A member of the bacterial candidate phylum TM7 ("*Ca*. Saccharibacteria") was cultivated and co-isolated with *Actinomyces odontolyticus* subsp. *actinosynbacter* strain XH001 by He et al. (2015). Having spherical cells of 200-300 nm in diameter and a genome of 0.705 Mbp, this bacterium of phylotype TM7 (strain TM7x) is associated with human oral microflora and was found to have a rather unique lifestyle. Like many of others discussed here, it is dependent on its basibiont, the host of the epibiont, an organism that resides on the surface of the host, *Actinomyces odontolyticus* subsp. *actinosynbacter* XH001. Under normal conditions, TM7x is an obligate epibiont, but during starvation it changes its lifestyle to parasitic, which eventually kills its own host and which is not usual for oral microorganisms (He et al., 2015; McLean et al., 2016). Additionally, TM7x lacks the ability to produce its own amino acids which further suggests its dependence on *A. odontolyticus* subsp. *actinosynbacter* XH001 (He et al., 2015). Its relationship with the host is thought to exacerbate oral mucosal diseases by concealing host immune responses by inhibiting *A. odontolyticus* XH001-induced TNF-α mRNA expression in macrophages (He et al., 2015). However, not all *Candidate* phylum TM7 members reside in the oral mucosa like TM7x: for example, RAAC3 with a small (0.845 Mbp) genome was originally found in a sediment obtained from an acetate-stimulated aquifer (Kantor et al., 2013). Another representative of TM7 group, "*Candidatus* Saccharimonas aalborgensis", with the genome of 1.0 Mbp was obtained from the activated sludge bioreactor (Albertsen et al., 2013; He et al., 2015). It remains unclear why TM7x has a more streamlined genome than the other phylotypes, a possible explanation of this adaptation is its specific human microbiome habitat and its complete dependency on its actinomycete host.

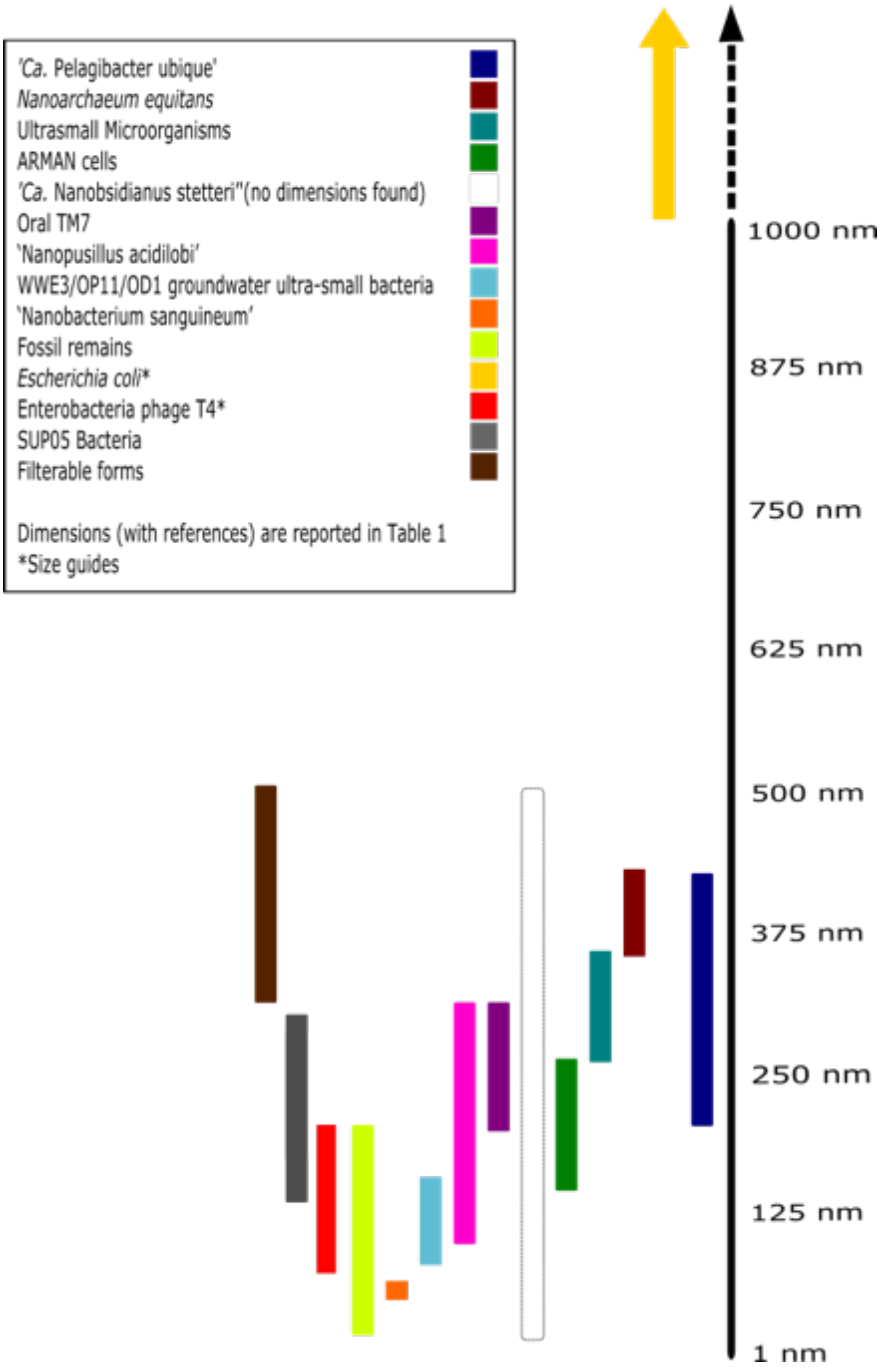


Figure 2.3- Size comparison of nano-sized organisms. Each of the colored lines represents relative range of sizes (in one dimension) of each individual. References and numerical ranges for individuals can be found in Table 2.1. If size was reported with volume, the organism was assumed to be spherical and then obtained the radius with the equation, $V=\frac{4}{3}\pi r^3$, where r is the radius. *References for size guides: Escherichia coli (approximately 1 µm x 2 µm) and phage T4 (approximately 90 nm x 200 nm) (Leiman et al., 2003). Note: Ca. Nanobsidianus stetteri' has no available information concerning cellular dimensions.

2.5. Selective pressures for small size

An important conclusion that can be made from the aforementioned studies on small-size microorganisms is that their sizes and distribution are a direct consequence of nutrient availability. As mentioned previously, increasing the surface area-to-volume ratio, which is an attribute of smaller cells, provides microorganisms with the ability to take up nutrients more efficiently (Giovannoni et al., 2014). Both symbiotic and free-living organisms seem to have benefited from this change. The results from existing studies suggest that in environments with high nutrient concentrations, a nano-sized organism will likely be a symbiont (or epibiont) with a decreased cell size being a result of limited metabolic capabilities with complete metabolic dependence on a host (Martínez-Cano et al., 2015). *Nanoarchaeum equitans* is a good example of this, as hydrothermal vents are relatively nutrient-rich, but these archaea are completely dependent on *Ignicoccous hospitalis* (Giannone et al., 2014). As nutrients become less available, the more likely the small-sized organism will be free-living because an increased surface-area-to-volume ratio is incredibly advantageous under such conditions (Martínez-Cano et al., 2015). The species "*Ca*. Pelagibacter ubique" is a good illustration of this scenario. Residing in the nutrient-depleted open ocean, it needs to produce its own essential amino acids, vitamins, etc. to survive (Carini et al., 2012). This raises the question, as to why this typical adaptation (small size and limited metabolic capabilities) does also exist in relatively stable nutrient-rich habitats. One possibility is that there may be selective pressures coming from predatory species, especially in aquatic systems (Pernthaler et al., 2001; Simon et al., 2002; Pernthaler, 2017). In the study of Pernthaler et al. (2001), the presence of the protozoan, *Ochromonas* sp*.,* resulted in an increasing population of members of *Actinobacteria* cluster Ac1. When an alternate protozoan predator, *Cyclidium glaucoma*, was introduced, no increase in population densities of Ac1 bacteria was observed (Pernthaler et al., 2001). Apparently, *Ochromonas* sp. prefers preys that are 0.8 to 4 μm in size, while *C. glaucoma* prefers those

smaller than 0.8 μm. Since the Ac1 are smaller than 0.8μm, the presence of only *Ochromonas* sp. allowed them to proliferate (Pernthaler et al., 2001). It was later found that some isolates of Ac1 were in fact ultramicro-sized (less than 0.1 μm^3 volume) and this small size prevented them from predation by *Ochromonas* sp. strain DS (Hahn et al., 2003). Hence, large populations of small organisms may also be a response to, or the result of, protozoan grazing (Salcher, 2014).

Another driver of selection of particular organisms in the environment are viruses and phages. Phages are host-specific and in most cases infect highly populous and dense bacterial subpopulations, which allows for less competitive (e.g. slow-growing) cells to proliferate (Winter et al., 2010; Salcher, 2014). Lysis of infected cells releases nutrients into the environment and makes them available to other community members allowing for overall microbial population growth (Weinbauer, 2004; Salcher, 2014). Viruses, similarly to predators, act as population control by culling overpopulated microorganisms ("killing the winner") while providing nutrients in the form of lysed cells to other species in the community (Weinbauer, 2004; Winter et al., 2010; Salcher, 2014).

2.6. Functional role of small-sized organisms

As documented here, small-sized organisms are not characterized by any specific type of metabolism or taxonomic affiliation. Therefore, we assume that their functional role is not restricted and may highly vary depending on the environment and actual physio-chemical conditions. Aquatic systems are incredibly complex, as fluctuations between high and low nutrient availability are common. In marine systems, the addition of nutrients e.g. in the form of nitrogen-rich fertilizers from agricultural runoffs, can greatly change the once oligotrophic environment into a copitrophic one, leading to harmful large scale phytoplanktonic blooms (Beman et al., 2005). Depending on concentrations of nutrients, populations of free-living small-celled microorganisms can either be enriched in R-strategists, or in K-strategists playing

distinct roles in the community. K-strategists, e.g. SUP05 clade and "*Ca*. P. ubique", are heavily involved with carbon and nitrogen cycling in oligotrophic areas (such as the open ocean and oxygen-depleted zones) (Giovannoni, 2017; Rogge et al., 2017). They are slow-growing and are widely dispersed, and rarely form colonies (Roshan and DeVries, 2017; Dang and Lovell, 2016; Giovannoni, 2017). R-strategists, e.g. Marine *Roseobacter* Clade (MRC) members and *Bacteroidetes,* are widely distributed and typically reside in nutrient-rich systems, e.g. in coastal systems (Dang and Lovell, 2016). These free-living organisms under favorable conditions grow quickly and may form large densely packed colonies and biofilms (Dang and Lovell, 2016). MRC bacteria can produce auxins and vitamins that are beneficial for algae (Dang and Lovell, 2016), whereas catabolically versatile *Bacteriodetes* play key roles in degrading high molecular weight dissolved organic matter (DOM) and biopolymers (Dang and Lovell, 2016).

In vertebrate systems, the role of these organisms appears variable. As seen in the case of TM7x, it may be beneficial or harmful to the host. *Actinomyces* strain XH001 normally elicits an immune response but TM7x modulates this response by either suppressing TNF-α gene expression in macrophages or "masking" it from macrophage detection altogether. However, under extended starvation conditions, TM7x can turn parasite, which leads to the host's demise (He et al., 2015).

Much of the literature discussed in this review has focused on a few species, however, the concerted effect of the entire ultra-small-sized microbial community in ecosystem functioning remains unknown. As discussed earlier, filtration through <0.45 μm pore size filters, is a common method to isolate small cells from aqueous samples. Interestingly, ultrafiltration was considered a method of choice to preserve freshwater samples during their storage and prior the hydrochemistry analysis (Brailsford et al., 2017). 0.22 μm pore size filters were considered as a safe tool for sterilization and for effective removal of microorganisms.

However, a recent study, which monitored the depletion of ^{14}C-glucose, ^{14}C-amino acid mixture, and ^{33}P-orthophosphate in filtered and unfiltered freshwater samples showed significant activity and utilization of substrates by organisms capable of passing this barrier (Brailsford et al., 2017). The previous studies clearly support this claim, as a number of the species were able to pass through ultrafiltration membranes (e.g. Wang et al., 2008). The great abundance of small-sized organisms in aqueous environments may also be attributed to selective pressures of predator-prey-viral interactions (Salcher, 2014). As discussed, protists feed on bacterioplankton and select prey based on cell size (Pernthaler et al., 2001; Salcher et al., 2013; Pernthaler, 2017). Conversely, viruses select for high-density preys and promote generation of DOM from lysed cells (Salcher, 2014), which can then be utilized by nano-sized microorganisms.

Nutrient cycling by ultra-small-sized organisms is not restricted to aquatic environments. A number of studies have shown an active population of ultramicrobacteria within a wide range of soil types (Soina et al., 2012; Lysak et al., 2013; Dobrovol'skaya et al., 2015). It was previously thought that soil pores <1 μm would be inaccessible to cells, leading to physical protection of organic carbon in soil. However, the potential of small-sized organisms to occupy this void space alongside their functional significance in soil remain unknown.

2.7. Conclusions and outlook

Discovery of small cells in the environment has reshaped our understanding of the microbial world and life on this planet. Using culture-independent tools first insights into the functionality of these organisms and a precise definition of the minimal sizes of living forms have been gained. Hence, it is reasonable to think that small-sized organisms may play a significant role in many environments. Many studies performed to date, however, have not considered the functionality of these organisms. Future studies should therefore shift their focus

to understanding their physiology and function. As more ecosystems are explored and as techniques are improved, the possibility of finding small-sized organisms is increasing. Culture- independent analysis will remain a critical tool for modelling and predicting functionalities and abundance of these organisms, however, the functional analysis of their activities remains essential to validate genome-based predictions.

2.8. Acknowledgements and Contributions

The review was conceived by all the authors. Lydia-Ann Ghuneim searched the literature, synthesized the data and wrote the manuscript. David Jones, Peter Golyshin, and Olga Golyshina provided significant revisions to the manuscript including data interpretation and wrote parts of the manuscript. All authors read the final manuscript. This work was carried out under the DOMAINE project, which is funded by the UK Natural Environment Research Council (NERC) (large grant NE/K010689/1). Olga Golyshina and Peter Golyshin acknowledge the support of the Centre of Environmental Biotechnology Project part-funded by the European Regional Development Fund (ERDF) through the Welsh Government and thank for its support the European Union Horizon 2020 Research and Innovation Program [Blue Growth: Unlocking the Potential of Seas and Oceans, grant agreement no. 634486, Project acronym 'INMARE'].

2.9. References

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., Nielsen, P. H., et al. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533. doi:10.1038/nbt.2579.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer - Reil, L. A., and Thingstad, F. (1983). The ecological role of water - column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi:10.3354/meps010257.

Baker, B. J., and Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* 44, 139–152. doi:10.1016/S0168-6496(03)00028-X.

Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic , ultrasmall , uncultivated Archaea. *Proc. Natl. Acad. Sci.* 107, 8806–11. doi:10.1073/pnas.0914470107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.0914470107.

Baker, B. J., Tyson, G. W., Webb, R. I., Flanagan, J., Hugenholtz, P., Allen, E. E., et al. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science* 314, 1933–1935. doi:10.1126/science.1132690.

Beman, J. M., Arrigo, K. R., and Matson, P. A. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434, 211–214. doi:10.1038/nature03370.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Button, D. K., Robertson, B. R., Lepp, P. W., and Schmidt, T. M. (1998). A small, dilute-cytoplasm, high-affinity, novel bacterium isolated by extinction culture and having kinetic constants compatible with growth at ambient concentrations of dissolved nutrients in seawater. *Appl. Environ. Microbiol.* 64, 4467–4476.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* 108, 4516–4522. doi:10.1073/pnas.1000080107.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S. J. (2012). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J.* 7, 592–602. doi:10.1038/ismej.2012.122.

Cesar, S., and Huang, K. C. (2017). Thinking big: the tunability of bacterial cell size. *FEMS Microbiol. Rev.* 41, 672–678. doi:10.1093/femsre/fux026.

Chien, A. C., Hill, N. S., and Levin, P. A. (2012). Cell size control in bacteria. *Curr. Biol.* 22, R340–R349. doi:10.1016/j.cub.2012.02.032.

Comolli, L. R., Baker, B. J., Downing, K. H., Siegerist, C. E., and Banfield, J. F. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *Isme J* 3, 159–167. doi:10.1038/ismej.2008.99.

Comolli, L. R., and Banfield, J. F. (2014). Inter-species interconnections in acid mine drainage microbial communities. *Front. Microbiol.* 5, 1–8. doi:10.3389/fmicb.2014.00367.

Dang, H., and Lovell, C. R. (2016). Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* 80, 91–138. doi:10.1128/MMBR.00037-15.Address.

Dobrovol'skaya, T. G., Zvyagintsev, D. G., Chernov, I. Y., Golovchenko, A. V., Zenova, G. M., Lysak, L. V., et al. (2015). The role of microorganisms in the ecological functions of soils. *Eurasian Soil Sci.* 48, 959–967. doi:10.1134/S1064229315090033.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Fedotova, A. V., Belova, S. E., Kulichevskaya, I. S., and Dedysh, S. N. (2012). Molecular identification of filterable bacteria and archaea in the water of acidic lakes of northern Russia. *Microbiology* 81, 281–287. doi:10.1134/S002626171203006X.

Fenchel, T. (1982). Ecology of Heterotrophic Microflagellates. I. Some Important Forms and Their Functional Morphology. *Mar. Ecol. Prog. Ser.* 8, 211–223. doi:10.3354/meps008211.

Garza, D. R., and Dutilh, B. E. (2015). From cultured to uncultured genome sequences: Metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* 72, 4287–4308. doi:10.1007/s00018-015-2004-1.

Gasol, J. M., and Morán, X. A. G. (1999). Effects of filtration on bacterial activity and picoplankton community structure as assessed by flow cytometry. *Aquat. Microb. Ecol.* 16, 251–264. doi:10.3354/ame016251.

Giannone, R. J., Wurch, L. L., Heimerl, T., Martin, S., Yang, Z., Huber, H., et al. (2014). Life on the edge: functional genomic response of Ignicoccus hospitalis to the presence of Nanoarchaeum equitans. *ISME J.* 9, 101–14. doi:10.1038/ismej.2014.112.

Gibson, E. K., McKay, D. S., Thomas-Keprta, K. L., Wentworth, S. J., Westall, F., Steele, A., et al. (2001). Life on Mars: Evaluation of the evidence within Martian meteorites ALH84001, Nakhla, and Shergotty. *Precambrian Res.* 106, 15–34. doi:10.1016/S0301-

9268(00)00122-4.

Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann. Rev. Mar. Sci.* 9, 231–255. doi:10.1146/annurev-marine-010814-015934.

Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1–13. doi:10.1038/ismej.2014.60.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science (80-.).* 309, 1242–1245. doi:10.1126/science.1114057.

Glaubitz, S., Kießlich, K., Meeske, C., Labrenz, M., and Jürgens, K. (2013). SUP05 Dominates the gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central baltic and black seas. *Appl. Environ. Microbiol.* 79, 2767–2776. doi:10.1128/AEM.03777-12.

Glissmann, K., Chin, K. J., Casper, P., and Conrad, R. (2004). Methanogenic pathway and archaeal community structure in the sediment of eutrophic Lake Dagow: Effect of temperature. *Microb. Ecol.* 48, 389–399. doi:10.1007/s00248-003-2027-2.

Golyshina, O. V. (2014). The Family Ferroplasmaceae. *Prokaryotes Other Major Lineages Bact. Archaea*, 29–34.

Golyshina, O. V., Lünsdorf, H., Kublanov, I. V., Goldenstein, N. I., Hinrichs, K. U., and Golyshin, P. N. (2016). The novel extremely acidophilic, cell-wall-deficient archaeon Cuniculiplasma divulgatum gen. Nov., sp. nov. represents a new family, Cuniculiplasmataceae fam. nov., of the order Thermoplasmatales. *Int. J. Syst. Evol. Microbiol.* 66, 332–340. doi:10.1099/ijsem.0.000725.

Golyshina, O. V., Toshchakov, S. V., Makarova, K. S., Gavrilov, S. N., Korzhenkov, A. A., La Cono, V., et al. (2017). “ARMAN” archaea depend on association with euryarchaeal

host in culture and in situ. *Nat. Commun.* 8, 1–11. doi:10.1038/s41467-017-00104-7.

Hahn, M. W., Lünsdorf, H., Wu, Q., Höfle, M. G., Boenigk, J., Stadler, P., et al. (2003). Isolation of Novel Ultramicrobacteria Classified as Actinobacteria from Five Freshwater Habitats in Europe and Asia Isolation of Novel Ultramicrobacteria Classified as Actinobacteria from Five Freshwater Habitats in Europe and Asia. *Appl. Environ. Microbiol.* 69, 1442–1451. doi:10.1128/AEM.69.3.1442.

Hansell, D. A., Carlson, C. A., Repeta, D. J., and Schlitzer, R. (2009). Dissolved organic matter in the ocean. *Oceanography* 22, 202–211. doi:10.1038/ncomms8422.

He, X., McLean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S.-Y., et al. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci. U. S. A.* 112, 244–9. doi:10.1073/pnas.1419038112.

Hodson, A., Anesio, A. M., Tranter, M., Fountain, A., Osborn, M., Priscu, J., et al. (2008). Glacial Ecosystems. *Ecol. Monogr.* 78, 41–67. doi:10.1002/ecm.1283.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67.

Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M., and Lasken, R. S. (2008). Genomic sequencing of single microbial cells from environmental samples. *Curr. Opin. Microbiol.* 11, 198–204. doi:10.1016/j.mib.2008.05.006.

Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., et al. (2008). Nanoarchaeum equitans and Ignicoccus hospitalis: New insights into a unique, intimate association of two archaea. *J. Bacteriol.* 190, 1743–1750. doi:10.1128/JB.01731-07.

Johnson, K. S., Gordon, R. M., and Coale, K. H. (1997). What controls dissolved iron concentrations in the world ocean? *Mar. Chem.* 57, 137–161. doi:10.1016/S0304-4203(97)00043-1.

Kajander, E. O., and Ciftcioglu, N. (1998). Nanobacteria: an alternative mechanism for pathogenic intra- and extracellular calcification and stone formation. *Proc.Natl.Acad.Sci.USA* 95, 8274–8279. doi:10.1073/pnas.95.14.8274.

Kantor, R. S., Wrighton, K. C., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., et al. (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *MBio* 4, 1–11. doi:10.1128/mBio.00708-13.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi:10.1006/jmbi.2000.4315.

Kuczynski, J., Liu, Z., Lozupone, C., McDonald, D., Fierer, N., and Knight, R. (2010). Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat. Methods* 7, 813–819. doi:10.1038/nmeth.1499.

Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2003). Structure and morphogenesis of bacteriophage T4. *Cell. Mol. Life Sci.* 60, 2356–2370. doi:10.1007/s00018-003-3072-1.

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., et al. (2015). Life under extreme energy limitation: A synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* 39, 688–728. doi:10.1093/femsre/fuv020.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi:10.1038/ncomms7372.

Lysak, L. V, Kadulin, M. S., Konova, I. A., Lapygina, E. V, Ivanov, A. V, and Zvyagintsev, D. G. (2013). Population number, viability, and taxonomic composition of the bacterial nanoforms in iron-manganic concretions. *Eurasian Soil Sci.* 46, 668–675. doi:10.1134/S1064229313060069.

Ma, D., Hao, Z., Sun, R., Bartlam, M., and Wang, Y. (2016). Genome Sequence of a Typical Ultramicrobacterium, Curvibacter Strain PAE-UM, Capable of Phthalate Ester Degradation. *Genome Announc.* 4, 1–2. doi:10.1128/genomeA.01510-15.Copyright.

Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 6. doi:10.3389/fmicb.2014.00742.

McCutcheon, J. P., and Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi:10.1038/nrmicro2670.

McKay, D. S., Gibson, E. K., Thomas-Keprta, K. L., Vali, H., Romanek, C. S., Clemett, S. J., et al. (1996). Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. *Science* 273, 924–930. doi:10.1126/science.273.5277.924.

Miteva, V. (2008). "Bacteria in snow and glacier ice," in *Psychrophiles: From Biodiversity to Biotechnology*, 31–50. doi:10.1007/978-3-540-74335-4_3.

Miteva, V. I., and Brenchley, J. E. (2005). Detection and Isolation of Ultrasmall Microorganisms from a. *Appl. Environ. Microbiol.* 71, 7806–7818. doi:10.1128/AEM.71.12.7806.

Möller, S., Croning, M. D. R., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17, 646–653. doi:10.1093/bioinformatics/17.7.646.

Monier, J.-M., and Lindow, S. E. (2003). Pseudomonas syringae Responds to the

Environment on Leaves by Cell Size Reduction. *Phytopathology* 93, 1209–1216. doi:10.1094/PHYTO.2003.93.10.1209.

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806–810. doi:10.1038/nature01240.

Munson-McGee, J. H., Field, E. K., Bateson, M., Rooney, C., Stepanauskas, R., and Young, M. J. (2015). Nanoarchaeota, their Sulfolobales Host, and Nanoarchaeota virus distribution across Yellowstone National Park hot springs. *Appl. Environ. Microbiol.* 81, 7860–7868. doi:10.1128/AEM.01539-15.

Mushegian, A. R., and Koonin, E. V (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273. doi:10.1073/pnas.93.19.10268.

Nagy, I., Knispel, R. W., Kofler, C., Orsini, M., Boicu, M., Varga, S., et al. (2016). Lipoprotein-like particles in a prokaryote: Quinone droplets of Thermoplasma acidophilum. *FEMS Microbiol. Lett.* 363, 1–7. doi:10.1093/femsle/fnw169.

Nakai, R., Fujisawa, T., Nakamura, Y., Nishide, H., Uchiyama, I., Baba, T., et al. (2016). Complete Genome Sequence of *Aurantimicrobium minutum* Type Strain KNC ^{T} , a Planktonic Ultramicrobacterium Isolated from River Water. *Genome Announc.* 4, e00616-16. doi:10.1128/genomeA.00616-16.

Narasingarao, P., Podell, S., Ugalde, J. A., Brochier-Armanet, C., Emerson, J. B., Brocks, J. J., et al. (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* 6, 81–93. doi:10.1038/ismej.2011.78.

Neuenschwander, S. M., Salcher, M. M., and Pernthaler, J. (2015). Fluorescence in situ

hybridization and sequential catalyzed reporter deposition (2C-FISH) for the flow cytometric sorting of freshwater ultramicrobacteria. *Front. Microbiol.* 6, 1–8. doi:10.3389/fmicb.2015.00247.

Oppenheimer, C. H. (1952). The membrane filter in marine microbiology. *J. Bacteriol.* 64, 783–786.

Panikov, N. S. (2005). Contribution of Nanosized Bacteria to the total biomass and activity of a soil microbial community. *Adv. Appl. Microbiol.* 57, 245–296.

Pernthaler, J. (2017). Competition and niche separation of pelagic bacteria in freshwater habitats. *Environ. Microbiol.* 19, 2133–2150. doi:10.1111/1462-2920.13742.

Pernthaler, J., Posch, T., Šimek, K., Vrba, J., Pernthaler, A., Glöckner, F. O., et al. (2001). Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. *Appl. Environ. Microbiol.* 67, 2145–2155. doi:10.1128/AEM.67.5.2145.

Podar, M., Makarova, K. S., Graham, D. E., Wolf, Y. I., Koonin, E. V, and Reysenbach, A.-L. (2013). Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct* 8, 9. doi:10.1186/1745-6150-8-9.

Probst, A. J., Ladd, B., Jarett, J. K., Geller-Mcgrath, D. E., Sieber, C. M. K., Emerson, J. B., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3, 328–336. doi:10.1038/s41564-017-0098-y.

Rappé, M. S., Connon, S. A., Vergin, K. L., and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplanton clade. *Nature* 418, 630–633. doi:10.1038/nature00917.

Rogge, A., Vogts, A., Voss, M., Jürgens, K., Jost, G., and Labrenz, M. (2017). Success of chemolithoautotrophic SUP05 and Sulfurimonas GD17 cells in pelagic Baltic Sea redox zones is facilitated by their lifestyles as K- and r-strategists. *Environ. Microbiol.* 19, 2495–2506. doi:10.1111/1462-2920.13783.

Roshan, S., and DeVries, T. (2017). Efficient dissolved organic carbon production and export in the oligotrophic ocean. *Nat. Commun.* 8, 2036. doi:10.1038/s41467-017-02227-3.

Salcher, M. M. (2014). Same same but different: Ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi:10.4081/jlimnol.2014.813.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907. doi:10.1038/ismej.2012.162.

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., et al. (2017). Estimates of Water-Column Nutrient Concentrations and Carbonate System Parameters in the Global Ocean: A Novel Approach Based on Neural Networks. *Front. Mar. Sci.* 4, 1–17. doi:10.3389/fmars.2017.00128.

Schut, F., Jansen, M., Pedro Gomes, T. M., Gottschal, J. C., Harder, W., and Prins, R. A. (1995). Substrate uptake and utilization by a marine ultramicrobacterium. *Microbiology* 141, 351–361. doi:10.1099/13500872-141-2-351.

Schut, F., Prins, R. A., and Gottschal, J. C. (1997). Oligotrophy and pelagic marine bacteria: facts and fiction. *Aquat. Microb. Ecol.* 12, 177–202. Available at: papers3://publication/uuid/C7BE943C-B654-458E-95B1-736B0B6BC849.

Shah, V., Chang, B. X., and Morris, R. M. (2017). Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *ISME J.* 11, 263–271. doi:10.1038/ismej.2016.87.

Sieburth, J. M. N., Smetacek, V., and Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol. Oceanogr.* 23, 1256–1263. doi:10.4319/lo.1978.23.6.1256.

Simon, M., Grossart, H. P., Schweitzer, B., and Ploug, H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat. Microb. Ecol.* 28, 175–211. doi:10.3354/ame028175.

Soina, V. S., Lysak, L. V, Konova, I. A., Lapygina, E. V, and Zvyagintsev, D. G. (2012). Study of ultramicrobacteria ({Nanoforms}) in soils and subsoil deposits by electron microscopy. *Eurasian Soil Sci.* 45, 1048–1056. doi:10.1134/S1064229312110087.

Stevens, T., and Arkin, I. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes. *Proteins* 39, 417–420.

Suzina, N. E., Esikova, T. Z., Oleinikov, R. R., Gafarov, A. B., Shorokhova, A. P., Polivtseva, V. N., et al. (2015). Comparative Characteristics of Free-Living Ultramicroscopical Bacteria Obtained from Natural Biotopes. *Microbiology* 51, 159–168.

Tagliabue, A., Bowie, A. R., Boyd, P. W., Buck, K. N., Johnson, K. S., and Saito, M. A. (2017). The integral role of iron in ocean biogeochemistry. *Nature* 543, 51–59. doi:10.1038/nature21058.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

Torrella, F., and Morita, R. Y. (1981). Microcultural study of bacterial size changes and microcolony and ultramicrocolony formation by heterotropic bacteria in seawater. *Appl. Environ. Microbiol.* 41, 518–527.

Torsvik, V., and Øvreås, L. (2008). Microbial Diversity , Life Strategies , and Adaptation to Life in Extreme Soils. *Microbiol. Extrem. Soils*, 15–43. doi:10.1007/978-3-540-74231-9.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science (80-.).* 308, 554–557. doi:10.1126/science.1107851.

Tripp, H. J. (2013). The unique metabolism of SAR11 aquatic bacteria. *J. Microbiol.* 51, 147–153. doi:10.1007/s12275-013-2671-2.

Vancanneyt, M., Schut, F., Snauwaert, C., Goris, J., Swings, J., and Gottschal, J. C. (2001). Sphingomonas alaskensis sp. now., a dominant bacterium from a marine oligotrophic environment. *Int. J. Syst. Evol. Microbiol.* 51, 73–80. doi:10.1099/00207713-51-1-73.

Velimirov, B. (2001). Nanobacteria, Ultramicrobacteria and Starvation Forms: A Search for the Smallest Metabolizing Bacterium. *Microbes Environ.* 16, 67–77. doi:10.1264/jsme2.2001.67.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi:10.1126/science.1093857.

Vergin, K. L., Done, B., Carlson, C. A., and Giovannoni, S. J. (2013). Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *Aquat. Microb. Ecol.* 71, 1–13. doi:10.3354/ame01661.

Wagner, J. K., Setayeshgar, S., Sharon, L. A., Reilly, J. P., and Brun, Y. V (2006). A nutrient uptake role for bacterial cell envelope extensions. *Proc. Nat. Acad. Sci. U. S. A.* 103, 11772–11777. doi:10.1073/pnas.0602047103.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent

enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi:10.1021/es0707198.

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* 42, 6749–6754. doi:10.1021/es800720n.

Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of Nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism. *PNAS* 100, 12984–12988.

Watson, S. P., Clements, M. O., and Foster, S. J. (1998). Characterization of the starvation-survival response of Staphylococcus aureus. *J. Bacteriol.* 180, 1750–1758.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181. doi:10.1016/j.femsre.2003.08.001.

Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. (2010). Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. *Microbiol. Mol. Biol. Rev.* 74, 42–57. doi:10.1128/MMBR.00034-09.

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla (supplement materials). *Science (80-.).* 337, 1661–5. doi:10.1126/science.1224041.

Wu, X., Holmfeldt, K., Hubalek, V., Lundin, D., Åström, M., Bertilsson, S., et al. (2015). Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. *ISME J.*, 1–12. doi:10.1038/ismej.2015.185.

Wurch, L., Giannone, R. J., Belisle, B. S., Swift, C., Utturkar, S., Hettich, R. L., et al. (2016). Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat. Commun.* 7, 1–10. doi:10.1038/ncomms12115.

Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., Mcintosh, J. R., and Nicastro, D. (2017). Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium "Candidatus Pelagibacter ubique. *Appl. Environ. Microbiol.* 83, 1–14.

Chapter 3

Assessment of methodologies of DNA extraction and whole genome amplification (WGA) from freshwater and seawater samples for consecutive genomic analysis

Abstract

Metagenomics requires large quantities of viable DNA. In aquatic systems this often means that 100 to 1000L of water is collected for a single study. An alternate solution is multiple displacement amplification (MDA) via whole genome amplification (WGA) to amplify pre-existing material from low aqueous volumes. Hence, in this study, we focused on the applicability of MDA for river and seawater samples in volumes of 50 mL or less. Using *Oleispira antarctica RB-8* as a control we showed that MDA is able to amplify $1.6x10^{-5}$ ng/μL of DNA accurately with minimal mutations. We also compared environmental samples obtained from a freshwater source (the Conwy River) and the saltwater source (the Menai Strait) to examine the efficacy of two ultrafiltration methods. The first is using a single 0.22 μm syringe filter and the second is a double filter method (0.2 μm followed by a 0.05 μm pore size filter). We examined the bacterial communities by assembling a 16S rRNA clonal library. There was no discernible difference in community composition of the freshwater of seawater samples with either ultrafiltration method.

Keywords: Whole genome amplification (WGA), multiple displacement amplification (MDA), tangential flow ultrafiltration, freshwater, saltwater, bacterial 16S rRNA clonal library

3.1. Introduction

Metagenomics offers valuable insight into various ecosystems; however, collecting enough high-quality DNA can be challenging. Such challenges; i.e. weather conditions, geography, accessibility to a site, and low-biomass environments; can all can lead to the recovery of poor-quality DNA. The cumulative logistical problems associated with general sampling techniques are significant. Hence, leading to low quantity and, often, poor quality of DNA material from these samples for downstream genomic analysis (limited sample volumes, kit contamination, environmental stressors, etc.) questions result viability and any conclusions reached.

There are two potential ways to circumvent the issue. The first is to collect large volumes of environmental sample (100 to 1000 L in aquatic systems) and ultrafilter (< 0.45 μm pore size) the sample to acquire viable DNA (Venter et al., 2004; Tringe et al., 2005). Cells and DNA collected on the filter can then be further processed downstream for subsequent sequencing (Venter et al., 2004; Tringe et al., 2005). The issue is that there are many microorganisms that can easily pass through ultrafilters. Filterable microorganisms have the unique ability to pass through a 0.45 μm filter unharmed (Velimirov, 2001; Panikov, 2005; Wang et al., 2008; Duda et al., 2012). Therefore, this technique does not fully represent the community present.

The second is to use whole genome amplification (WGA) through either PCR (DOP-PCR and PEP) or non-PCR methods (MDA) to amplify DNA (Spits et al., 2006; Arneson et al., 2008; Binga et al., 2008; Blagodatskikh et al., 2017). Although both the PCR based tecniques implement *Taq*-DNA polymerase, the do different in primers and annealing temeraptures (Spits et al., 2006; Arneson et al., 2008; Blagodatskikh et al., 2017). Degenerate oligonucleotide PCR (DOP-PCR) uses semi-degenerate oligonucleotides and high annealing temperature (Spits et al., 2006; Arneson et al., 2008; Blagodatskikh et al., 2017). Whereas, primer extension preamplification (PEP) utilizes random primers and a low PCR annealing temperature (Spits et al., 2006; Arneson et al., 2008; Blagodatskikh et al., 2017). Even Yet *Taq* DNA polymerase is limited in that (1) at least 3 kb fragment sizes are needed to have successful amplification, (2) more errors are produced in the sequence, (3) amplification bias due to specific pimers being needed, and (4) overall coverage is lacking (Spits et al., 2006; Arneson et al., 2008; Blagodatskikh et al., 2017).

MDA on the other hand is a non-PCR amplification technique that utilizes a high processivity enzyme, such as phi29 DNA polymerase, and random hexameric primers to amplify minutiae amounts (less that 0.3 ng) of DNA (Spits et al., 2006; Binga et al., 2008).

MDA can be especially useful to examine minority populations, the uncultured majority, small genome populations, and viruses (Spits et al., 2006; Binga et al., 2008). Although MDA has the tendency to produce chimeras, it is still a popular choice due to the ability to amplify 1 ng of DNA template (Spits et al., 2006; Binga et al., 2008). The study of filterable microorganisms can benefit from MDA technology and ultrafiltration. As previously described, their characterization is rather broad yet they may play a crucial role in recycling nutrients. For instance, "*Ca*. Pelagibacter ubique", a very small free-living bacteria capable of producing 20 essential amino acids in oceanic pelagic environments (Carini et al., 2012; Tripp, 2013).

This fact alone raises the question of what are these filterable microorganisms and what role do they play *in situ*. The aim of the trial was therefore to assess the applicability of MDA for river and seawater samples from as low as <50 mL aqueous sample. The goal was to circumnavigate much of the problems associated with many previous metagenomics studies and gain the ability to detect these minority species.

3.2. Methods

3.2.1. Site description

One freshwater and one marine site were used in the study. The freshwater sample was obtained from the Conwy catchment, which is located in North Wales. Its main drainage is in the form of the river Conwy (Afon Conwy). Mean annual rainfall ranges from 500-3500 mm and the mean annual air temperature ranges from 5-15 °C with an average annual temperature of 10 °C (Emmett et al., 2016). The river itself is 55 km long; starting from Llyn Conwy (450 m above sea level) and drains the Migneint, a large peatland bog that is a major store of carbon. Three tributaries (Machno, Lledr, and Llugwy), originating from the eastern side of the Snowdonia mountain range, join the main river further downstream before reaching the tidal limit (20 km inland) (Emmett et al., 2016). Average concentrations at the tidal limit are as

follows: nitrite 0.2-2.8 mg/L, ammonium <0.03-0.04 mg/L, phosphate <0.02-0.05 mg/L, and dissolved organic carbon 1.5-10 mg/L (Emmett et al., 2016). The pH ranges from 5.7 to 7.2. The site used in this study (code: NM29) is located at Cwm Llanerch and is associated with the main Afon Conwy (53° 6' 24.7068" N, 3° 47' 28.7556" W). The site is located approximately 4 km from the tidal limit (Simpson et al., 2001).

The marine sample was obtained from the Menai Strait. It is a body of salt water ca. 25 km long that separates the isle of Anglesey from mainland Wales: the tidal excursion on spring tides is over 14 km (Kratzer et al., 2003). The width of the strait ranges from 200 m to 2000 m, reaching a maximum depth of 18 m (Kratzer et al., 2003). Currents from the Irish Sea flow from both ends of the strait because water flow is dominated by tidal forcing and the tidal ranges, at spring tide, is 6 m (Kratzer et al., 2003). Dissolved organic carbon concentration ranges from 0.5-6.8 mg/L (Morris and Foster, 1971).

3.2.2. Sample collection and preparation

10 L of seawater was collected near the surface of the Menai Strait and 10 L of river water was collected from the NM29 site in the Conwy River catchment in October 2015. Samples were given a Sample ID based on the on source of the sample and filtration method (Table 3.1 and Fig. 3.1). The SA and RFM1 samples were passed through a 0.2 µm FiberFlo® hollow fibre capsule filter followed by a 0.05 µm FiberFlo® hollow fibre capsule filter (Minntech Corporation; Minneapolis, MN, USA) (Figs. 3.1, 3.2 and Table 3.1). These samples were concentrated to a volume of 50 mL via centrifugation at 65,202 *g* for 10 minutes. Centrifuged, unfiltered samples (20 mL) were passed through a 0.22 µm Sterivex® PVDF (hydrophilic polyvinylidene fluoride) filter (Millipore Corporation, Billerica, MA, USA) and labelled as SFM3 and RFM3 (Table 3.1 and Fig. 3.1). Aliquots of 1 mL of each sample were taken and the cells were fixed with 100 µL formaldehyde and 50 µl of 50:50 (v/v) phosphate buffer solution (PBS; pH 7.4)/ethanol solution.

Table 3.1- Sample ID table. Described here are the freshwater and sea water samples used in this analysis. The source, filtration method, and WGA signal (presence/absence of signal on an electrophororated 0.8% agarose gel.) *The filtration methods are outlined in Fig. 3.1.

Sample ID	*Source*	*Filtration method**	*Signal after WGA*
SA	Menai Strait	Method 1	(+)
RFM1	Conwy River	Method 1	(+)
RFM2	Conwy River	Method 2	(+)
RFM3	Conwy River	Method 3	(+)
SFM3	Menai Strait	Method 3	(+)

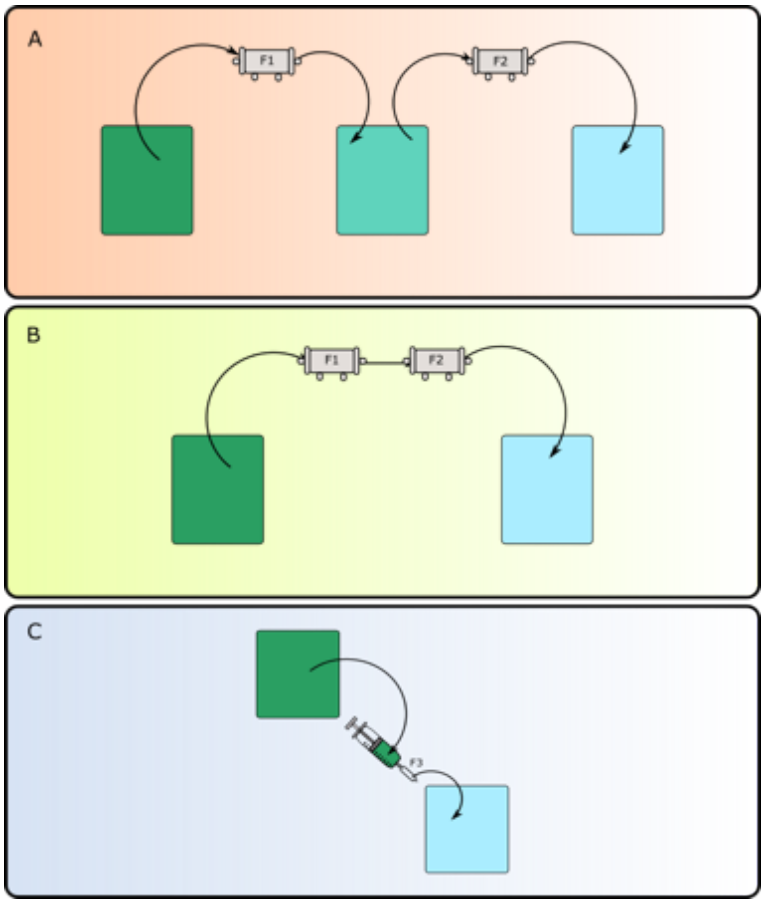


Figure 3.1- Schematic of filtering methods. F1 is the 0.2 µm FiberFlo® filter, F2 is the 0.05 µm FiberFlo® filter, and F3 is the 0.22 µm Sterivex™ filter. In method 1 (A) 10L of collected aqueous sample is first passed through F1, collected and then passed through F2. In method 2 (B) 10L of collected aqueous sample is passed through both F1 and F2 in series. We should note that method 1 and 2 produce the same level of filtering and thus only used samples from method 1. Finally, in method 3 (C) 20 mL of aqueous sample is passed through F3 via syringe.

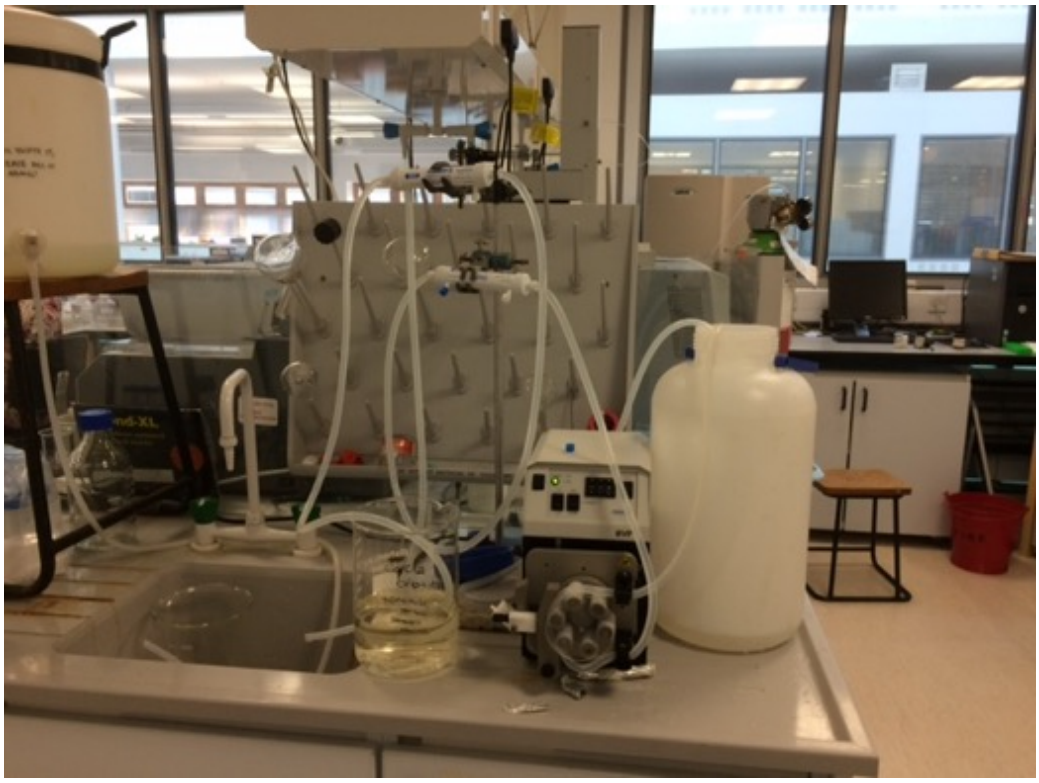


Figure 3.2- Experimental setup of the double filtration method. Using a periplasmic pump, 10 L of aqueous samples were first passed through a 0.2 µm filter and then a 0.05 µm FiberFlo® hollow fibre capsule filter (Minntech Corporation; Minneapolis, MN, USA). Final product was collected in a beaker for further processing and eventual whole gene amplification.

3.2.3. Quality control preparation

The species, *Oleispira antarctica RB-8*, is an obligate hydrocarbonoclastic bacterium (4.4 Mbp genome) which is found in the deep ocean (Kube et al., 2013). Purified DNA from this organism was used for the methodology assessment. The initial concentration was determined using the Quan-it dsDNA Assay kit® (LifeTechnologies, Eugene, OR, USA). Serial dilutions were taken from 1:10 to $1:10^{11}$ using PCR grade water (Sigma Aldrich Co, St Louis, MO, USA) and stored at -20 ^{0}C (Table 3.2).

Table 3.2- Table of the serial dilution concentrations for the quality control test. Serial dilutions were done on a DNA sample of the species, O. antarctica RB-8. *Samples (after WGA) were selected based on presence/absence of signal on an electrophororated 0.8% agarose gel. The estimated DNA concentration was determined using the Quan-it Kit®.

Sample (DNA from O. antarctica RB-8)	***Estimated/Calculated DNA concentration before WGA (ng/µL)***	***Signal after WGA***
Undiluted*	16	(+)
1:10	0.16	(+)
1:100	$1.6x10^{-2}$	(+)
1:1000 (10^3)*	$1.6x10^{-3}$	(+)
1:10,000 (10^4)*	$1.6x10^{-4}$	(+)
1:100,000 (10^5)*	$1.6x10^{-5}$	(+)
1:1,000,000 (10^6)	$1.6x10^{-6}$	(-)
1:10,000,000 (10^7)	$1.6x10^{-7}$	(-)
$1:10^8$	$1.6x10^{-8}$	(-)
$1:10^9$	$1.6x10^{-9}$	(-)
$1:10^{10}$	$1.6x10^{-10}$	(-)
$1:10^{11}$	$1.6x10^{-11}$	(-)

3.2.4. Extraction and generating clonal libraries

First WGA was achieved via MDA using the GEillustra Ready-to-go Genomi Phi V3 DNA amplification kit® (GE Healthcare UK Ltd, Little Chalfont, Buckinghamshire, UK). Successful WGA products were visualised in Bio-Rad ChemiDoc XR after tris-borate 0.8% agarose gel electrophoresis in SybrSafe (ThermoFisher) (Tables 3.1 and 3.2). Positive signals, as indicated in Tables 3.1 and 3.2, were then selected for further bacterial 16S rRNA gene amplification via PCR. Hence the WGA products were amplified using the universal primers 27F (5'-AGAGTTTGGATCMTGCTCAG-3') and 1492R (5'-TACCTTGTTACGACTT-3') with MyTaq™ DNA polymerase (Bioline Reagents Ltd, UK). QIAquick gel extraction kit® (Qiagen GmbH, Hilden, Germany) was used to isolate and further concentrate the DNA. The PCR product was then ligated into vector TOPO 2.1 and then transformed into TOP10F' *E. coli* electro competent cells (Invitrogen™, Life Technologies Corporation, Carlsbad, CA,

USA) using the TOPO® TA cloning® Kit (Invitrogen™, Life Technologies Corporation, Carlsbad, CA, USA) as per the manufacturer's protocol.

The cells were grown on LB media with kanamycin (50 µg/mL) and Xgal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside) (40 mg/mL). White colonies selected and inserts were extracted bi-directionally using M13 vector specific primers M13 Forward (5′-GTAAAACGACGGCCAG-3′) and M13 Reverse (5′-CAGGAAACAGCTATGAC-3′). All PCR reactions were performed in a Bio-Rad Tetrad 2 Peltier thermal cycler with the following program: denaturation step at 95°C for 2 minutes; followed by 30 cycles, each including denaturation 94°C for 1 minute, annealing (50°C for 1 minute), and DNA synthesis (72°C for 2 minutes). Final extension at 72°C for 10 minutes followed by cooling to 4°C. PCR products were visualised in Bio-Rad ChemiDoc XR after tris-borate 0.8% agarose gel electrophoresis in SybrSafe (ThermoFisher) and sent to Macrogen Inc, Seoul, Korea for Sanger sequencing.

3.2.5. 16S rRNA gene sequence analysis

The resultant sequences were screened and trimmed in BioEdit (Hall, 1999) and predicted relatives were determined using the BLAST 2.2 program (www.ncbi.nim.nih.gov/blast/). For 16S phylogenetic analysis, the sequences were aligned with ClustalW (Thompson et al., 1994) using 1,000 bootstraps within the BioEdit software. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). A Disparity Index test, (Kumar and Gadagkar 2001) was tabulated in MEGA5 (Tamura, et al. 2011) to illustrate similarity. A Monte Carlo test (500 replicates) was used to estimate the P-values (Kumar and Gadagkar 2001) and all positions containing gaps and missing data were eliminated. Three phylogenetic tree (Figs. 3.4, 3.5, and 3.7) calculations were made using the Neighbour-joining method (Saitou and Nei, 1987) with a discrete gamma distribution (5 categories) and 1,000 bootstraps. The evolutionary distances were computed using the Tajima-Nei method (Tajima and Nei, 1984). One phylogenetic tree calculation (Fig. 3.6) was made using the Maximum

likelihood method with a discrete gamma distribution (5 categories) and 1,000 bootstraps (Saitou and Nei, 1987). was All trees were generated using MEGA5 software (Tamura et al., 2011).

3.3. Results

3.3.1. 'Oleispira antarctica RB-8' quality control

To determine the efficacy and sensitivity of the WGA, *O. antarctica* RB-8 was initially assessed to predict the lower limits of detection to include the examination of minority populations (i.e., small size in terms of cellular dimensions and genome). Other factors such as changes within the sequence (i.e., level of mutation) as well as the integrity of the DNA, while maintaining the same identity were additionally ascertained to determine whether our WGA methodology had any effect upon identification. As shown in Table 3.3, WGA was indicated to be effective down to 1.6×10^{-5} ng/µL and further phylogenetic analysis indicated high degree of accuracy, as mutations were minimal (Fig. 3.3). One clone retrieved from the 1.6×10^{-5} ng/µL, 5H3_27F, showed the lowest degree of similarity to *O. antarctica* RB-8 (Table 3). The homogeneity test suggests that there is a high degree of similarity between all the clones with the exception of 2 samples, 4D3_27F and 4B3_27F, that had a concentration of 1.6×10^{-4} ng/µL (Table 3.3, Fig. 3.3).

	1 25
***Oleispira antarctica strain RB-8* partial sequence**	G C G G A C G G G T G A G T A A T G C T T A G G A
UD_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
UB_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
UA2_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
UA_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
5H3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
5F3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
5E3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
4D3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
4C3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
4B3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
4A3_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
3D2_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
3C2_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
3B2_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
3A2_27F	G C G G A C G G G T G A G T A A T G C T T A G G A
	26 50
***Oleispira antarctica strain RB-8* partial sequence**	A T C T A C C G A G T A G T G G G G G A T A G C C
UD_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
UB_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
UA2_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
UA_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
5H3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
5F3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
5E3_27F	A C C T A C C G A G T A G T G G G G G A T A G C C
4D3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
4C3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
4B3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
4A3_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
3D2_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
3C2_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
3B2_27F	A T C T A C C G A G T A G T G G G G G A T A G C C
3A2_27F	A T C T A C C G A G T A G T G G G G G A T A G C C

	51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	76		
***Oleispira antarctica strain RB-8* partial sequence**	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
UD_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
UB_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
UA2_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
UA_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
5H3_27F	A	T	T	G	G	A	A	C	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
5F3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
5E3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
4D3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
4C3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
4B3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
4A3_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
3D2_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
3C2_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
3B2_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
3A2_27F	A	T	T	G	G	A	A	A	C	G	A	T	G	A	T	T	A	A	T	A	C	C	G	C	A		
	77	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	102
***Oleispira antarctica strain RB-8* partial sequence**	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
UD_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
UB_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
UA2_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
UA_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
5H3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
5F3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
5E3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
4D3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	A	G	A	A	A	G	C	A	G	G	G	G	A	C
4C3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
4B3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
4A3_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
3D2_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
3C2_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
3B2_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C
3A2_27F	T	A	T	A	T	C	C	T	A	C	G	G	G	G	G	A	A	A	G	C	A	G	G	G	G	A	C

	103	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	128
Oleispira antarctica strain RB-8 **partial sequence**	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
UD_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
UB_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
UA2_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
UA_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
5H3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
5F3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
5E3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
4D3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
4C3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
4B3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
4A3_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
3D2_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
3C2_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
3B2_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
3A2_27F	C	T	T	C	G	G	G	C	C	T	T	G	C	G	C	T	A	T	T	C	G	A	T	G	A
	129	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	152
Oleispira antarctica strain RB-8 **partial sequence**	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
UD_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
UB_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
UA2_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
UA_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
5H3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
5F3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
5E3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
4D3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
4C3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
4B3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
4A3_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
3D2_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
3C2_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
3B2_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T
3A2_27F	G	C	C	T	G	A	G	T	G	A	G	A	T	T	A	G	C	T	A	G	T	T	G	G	T

	153	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	178		
Oleispira antarctica strain RB-8 **partial sequence**	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
UD_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
UB_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
UA2_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
UA_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
5H3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
5F3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
5E3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
4D3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
4C3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
4B3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
4A3_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
3D2_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
3C2_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
3B2_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
3A2_27F	G		G	G	G	T	A	A	A	G	G	C	C	T	A	C		C	A	A	G	G	C	G	A	C	G	
	179	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	204	
Oleispira antarctica strain RB-8 **partial sequence**	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
UD_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
UB_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
UA2_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
UA_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
5H3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
5F3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
5E3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
4D3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
4C3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
4B3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
4A3_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
3D2_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
3C2_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
3B2_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C
3A2_27F	A	T	C	T	C	T	A	G	C	T	G	G	T	C	T		G	A	G	A	G	G	A	T	G	A	T	C

	205	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	230
***Oleispira antarctica strain RB-8* partial sequence**	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
UD_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
UB_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
UA2_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
UA_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
5H3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
5F3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
5E3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
4D3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
4C3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
4B3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
4A3_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
3D2_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
3C2_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
3B2_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
3A2_27F	A		G	C	C	A	C	A	C	T	G	G	G	A	C	T		G	A	G	A	C	A	C	G	G	C
	231	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	254
***Oleispira antarctica strain RB-8* partial sequence**	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
UD_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
UB_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
UA2_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
UA_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
5H3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
5F3_27F	C	C	A	G	G	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
5E3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
4D3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
4C3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
4B3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
4A3_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
3D2_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
3C2_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
3B2_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T
3A2_27F	C	C	A	G	A	C	T	C	C	T	A	C	G	G	G		A	G	G	C	A	G	C	A	G	T

	255	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	280		
Oleispira antarctica strain RB-8 **partial sequence**	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
UD_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
UB_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
UA2_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
UA_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
5H3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	G
5F3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
5E3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
4D3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
4C3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
4B3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
4A3_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
3D2_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
3C2_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
3B2_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
3A2_27F	G		G	G	G	A	A	T	A	T	T	G	C	A	C	A		A	T	G	G	A	C	G	A	A	A
	281	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	306
Oleispira antarctica strain RB-8 **partial sequence**	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
UD_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
UB_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
UA2_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
UA_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
5H3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
5F3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
5E3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
4D3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
4C3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
4B3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
4A3_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
3D2_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
3C2_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
3B2_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G
3A2_27F	G	T	C	T	G	A	T	G	C	A	G	C	C	A	T		G	C	C	G	C	G	T	G	T	G	T	G

	307	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	332
Oleispira antarctica strain RB-8 **partial sequence**	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
UD_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
UB_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
UA2_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
UA_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
5H3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
5F3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
5E3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
4D3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
4C3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
4B3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
4A3_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
3D2_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
3C2_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
3B2_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
3A2_27F	A	A	G	A	A	G	G	C	C	T	T	C	G	G	G	T	T	G	T	A	A	A	G	C	A
	333	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	356
Oleispira antarctica strain RB-8 **partial sequence**	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
UD_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
UB_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
UA2_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
UA_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
5H3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
5F3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
5E3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
4D3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
4C3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
4B3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
4A3_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
3D2_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
3C2_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
3B2_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T
3A2_27F	C	T	T	T	C	A	G	C	G	A	G	G	A	G	G	A	A	A	G	G	T	C	A	G	T

	357	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	382		
Oleispira antarctica strain RB-8 **partial sequence**	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
UD_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
UB_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
UA2_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
UA_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
5H3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
5F3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
5E3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
4D3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
4C3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
4B3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
4A3_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
3D2_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
3C2_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
3B2_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
3A2_27F	A	A	G	T	A	A	T	A	T	T	T	G	C	T	G	G	C	T	G	T	G	A	C	G	T		
	383	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	408
Oleispira antarctica strain RB-8 **partial sequence**	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
UD_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
UB_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
UA2_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
UA_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
5H3_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
5F3_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
5E3_27F	T	A	C	T	C	G	C	A	G	A	A	A	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
4D3_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
4C3_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
4B3_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
4A3_27F	T	A	C	T	C	G	C	A	G	A	G	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
3D2_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
3C2_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
3B2_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T
3A2_27F	T	A	C	T	C	G	C	A	G	A	A	G	A	A	G	C	A	C	C	G	G	C	T	A	A	T	T

	409	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	434
***Oleispira antarctica strain RB-8* partial sequence**	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
UD_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
UB_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
UA2_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
UA_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
5H3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
5F3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
5E3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
4D3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
4C3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
4B3_27F	T	A	G	T	G	C	C	A	G	C	A	A	C	C	G	C	G	G	T	A	A	T	A	C	T
4A3_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
3D2_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
3C2_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
3B2_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
3A2_27F	T	A	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	T
	435	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	458
***Oleispira antarctica strain RB-8* partial sequence**	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
UD_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
UB_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
UA2_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
UA_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
5H3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
5F3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
5E3_27F	A	A	A	G	G	T	G	C	G	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
4D3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
4C3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
4B3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
4A3_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
3D2_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
3C2_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
3B2_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T
3A2_27F	A	A	A	G	G	T	G	C	A	A	G	C	G	T	T	A	A	T	C	G	G	A	A	T	T

	459	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	484		
***Oleispira antarctica strain RB-8* partial sequence**	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
UD_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
UB_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
UA2_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
UA_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
5H3_27F	A	C	T	G	G	G	C	G	T	A	A	G	G	C	G	C	G	C	G	T	A	G	G	T	G			
5F3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
5E3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
4D3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
4C3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
4B3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
4A3_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
3D2_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
3C2_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
3B2_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
3A2_27F	A	C	T	G	G	G	C	G	T	A	A	A	G	C	G	C	G	C	G	T	A	G	G	T	G			
	485	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	510	
***Oleispira antarctica strain RB-8* partial sequence**	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
UD_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
UB_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
UA2_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
UA_27F	G	C	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
5H3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
5F3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	A	A	T	G	T	G	A	A	A	G	C	C	C	A	
5E3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
4D3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
4C3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
4B3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
4A3_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
3D2_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
3C2_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
3B2_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	
3A2_27F	G	T	T	T	G	T	T	A	A	G	T	T	G	G	A	T	G	T	G	A	A	A	G	C	C	C	A	

	511	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	536
***Oleispira antarctica strain RB-8* partial sequence**	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
UD_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
UB_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
UA2_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
UA_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
5H3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
5F3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
5E3_27F	G		G	G	C	T	C	A	A	T	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
4D3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
4C3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
4B3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
4A3_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
3D2_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
3C2_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
3B2_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
3A2_27F	G		G	G	C	T	C	A	A	C	C	T	T	G	G	A		A	C	T	G	C	A	T	T	C	A
	537	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	560
***Oleispira antarctica strain RB-8* partial sequence**	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
UD_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
UB_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
UA2_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
UA_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
5H3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
5F3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
5E3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
4D3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
4C3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
4B3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
4A3_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
3D2_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
3C2_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
3B2_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G
3A2_27F	A		A	A	C	T	G	A	C	T	C	A	C	T	A	G		A	G	T	A	C	G	A	G	A	G

	561	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	586		
Oleispira antarctica strain RB-8 **partial sequence**	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
UD_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
UB_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
UA2_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
UA_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
5H3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
5F3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
5E3_27F	A		G	G	C	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
4D3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
4C3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
4B3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
4A3_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
3D2_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
3C2_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
3B2_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
3A2_27F	A		G	G	T	T	A	G	T	G	G	A	A	T	T	T		C	C	T	G	T	G	T	A	G	C		
	587	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	611
Oleispira antarctica strain RB-8 **partial sequence**	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
UD_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
UB_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
UA2_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
UA_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
5H3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
5F3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
5E3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
4D3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
4C3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
4B3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
4A3_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
3D2_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
3C2_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
3B2_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A
3A2_27F	G		G	T	G	A	A	A	T	G	C	G	T	A	G	A		G	A	T	G	G	G	A	A	G	G		A	A

	612	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	637
Oleispira antarctica strain RB-8 **partial sequence**	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
UD_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
UB_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
UA2_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
UA_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
5H3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
5F3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
5E3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
4D3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
4C3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
4B3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
4A3_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
3D2_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
3C2_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
3B2_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
3A2_27F	C	A	C	C	A	G	T	G	G	C	G	A	A	G	G	C	G	A	C	T	G	A	C	T	G
	638	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	661
Oleispira antarctica strain RB-8 **partial sequence**	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
UD_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
UB_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
UA2_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
UA_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
5H3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
5F3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	G	G	G	T	G	C	G	A
5E3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
4D3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
4C3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
4B3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
4A3_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
3D2_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
3C2_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	C	G	T	G	C	G	A
3B2_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A
3A2_27F	G	C	T	C	G	A	T	A	C	T	G	A	C	A	C	T	G	A	G	G	T	G	C	G	A

	662	.	.	.	.	.	.	.	.	.	.	.	.	.	677
***Oleispira antarctica strain RB-8* partial sequence**	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
UD_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
UB_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
UA2_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
UA_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
5H3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
5F3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
5E3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
4D3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
4C3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
4B3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
4A3_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
3D2_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
3C2_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
3B2_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A
3A2_27F	A	A	G	C	G	T	G	G	G	G	A	G	C	A	A

Figure 3.3- *Oleispira antarctica* RB-8 clonal library sequences by nucleotide. 16 ng/µL initial concentration and *O. antarctica* RB-8 reference (black); concentrations of: $1.6x10^{-5}$ ng/µL (red), $1.6x10^{-4}$ ng/µL (blue), and $1.6x10^{-3}$ ng/µL (green). 5H3_27F showed the lowest degree of similarity to O. antarctica RB-8. Each nucleotide, G (pink), A(yellow), T (dark green), and C (brown) are also indicated.

Table 3.3- A disparity index test. The P-value estimates of the disparity index per site are shown for each sequence pair below the diagonal. These values were calculated via Monte-Carlo test (500 replicates). P-values smaller than 0.05 are considered significant (marked with yellow highlights). 16 ng/µL initial concentration and O. antarctica RB-8 reference (black); concentrations of: 1.6 x 10^{-5} ng/µL (red), 1. 6 x10^{-4} ng/µL (blue), and 1.6x10^{-3} ng/µL (green).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	*Oleispira antarctica* strain RB-8 partial sequence		0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	UD_27F	1.00		0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	UB_27F	1.00	1.00		0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	UA2_27F	1.00	1.00	1.00		0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	UA_27F	1.00	1.00	1.00	1.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	5H3_27F	0.06	0.09	0.11	0.10	0.21		0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01
7	5F3_27F	1.00	1.00	1.00	1.00	1.00	1.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	5E3_27F	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	4D3_27F	1.00	1.00	1.00	1.00	1.00	0.01	1.00	1.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	4B3_27F	1.00	1.00	1.00	1.00	1.00	0.01	1.00	1.00	1.00		0.00	0.00	0.00	0.00	0.00	0.00
11	4C3_27F	1.00	1.00	1.00	1.00	1.00	0.06	1.00	1.00	1.00	1.00		0.00	0.00	0.00	0.00	0.00
12	4A3_27F	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.09	1.00		0.00	0.00	0.00	0.00
13	3D2_27F	1.00	1.00	1.00	1.00	1.00	0.10	1.00	1.00	1.00	1.00	1.00	1.00		0.00	0.00	0.00
14	3C2_27F	1.00	1.00	1.00	1.00	1.00	0.09	1.00	1.00	1.00	1.00	1.00	0.44	1.00		0.00	0.00
15	3B2_27F	1.00	1.00	1.00	1.00	1.00	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		0.00
16	3A2_27F	1.00	1.00	1.00	1.00	1.00	0.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

3.3.2. Freshwater and seawater samples

As observed in Figure 3.4 and 3.5, both freshwater samples RFM1 and RFM3 had low levels of biodiversity and there was no marked difference between the filtration methods. The sequences detected resided largely in the *Burkholderariales* order, specifically related to *Ralstonia pickettii* and *Variovorax paradoxus*. Additionally, *Cutibacterium acnes* (formally *Propionibacterium acnes*) was also observed in RFM3 samples.

The seawater sample passed through a 0.2 µm and then a 0.05 µm filter (sample SA) showed the highest level of diversity (Fig. 3.6), whereas the same seawater passed through a 0.22 µm Sterivex® filter (sample SFM3) showed a significantly decreased level of diversity (Fig. 3.7). The classes *Gammaproteobacteria*, *Alphaproteobacteria*, *Cytophagia*, and *Flavobacteria* were detected in the SA sample. SFM3 samples contained sequences of microorganisms normally found in freshwater systems of the *Burkholderiales* order and *Actinobacteria* phylum.

Table 3.4- Description of detected species. Range of sizes, morphology, and source sample are listed here, with one noted exception, *Spirulina major*.

Species Name	***Sample***	***Morphology***	***Dimension 1 (diameter) (µm)***	***Dimension 2 (length) (µm)***	***Reference***
Cutibacterium acnes	RFM3	Rod or branched	0.2-1.5	1.0-5.0	(Perry and Lambert, 2006; Jones and Goodfellow, 2012)
Ralstonia picketti	RFM1, RFM3, SFM3	slightly curved rod	0.5-0.8	1.2-3.0	(Bergey, 2005; Ryan et al., 2007)
Variovorax paradoxus	RFM1, RFM3, SFM3	slightly curved rod	0.5-0.6	1.2-3.0	(Bergey, 2005; Satola et al., 2013)
Arcicella aurantiaca	RFM3	Vibrioid (curved rod)	0.5-0.6	1.2-3.0	(Sheu et al., 2010)
Porticoccus hydrocarbonclasticus	SA	rod	1.0-2.0	0.5-0.6	(Gutierrez et al., 2012)
Marinobacterium litorale	SA	rod	0.5-0.8	1.2-2.0	(Kim et al., 2007)
Neptunomonas acidovorans	SA	rod	0.6-0.8	1.2-1.6	(Yang et al., 2014)
Amylibacter marinus	SA	rod	0.5-0.9	1.0-2.2	(Teramoto and Nishijima, 2014)
Spirulina major	SA	filamentous	1.5 (average)	100 (average)	(Olenina et al., 2006)
Phaeocystidibacter luteus	SA	rod	0.2-0.4	1.0-1.5	(Zhou et al., 2013)
Geojedonia litorea	SA	rod, coccoid, or filamentous	0.2-0.4	0.4-10	(Park et al., 2013)
Formosa arctica	SA	Rod	0.3-0.7	1.5-2.1	(Kwon et al., 2014)
Sphingomonas leidyi	SFM3	rod	0.3-0.8	0.7-1.9	(Chen et al., 2012)

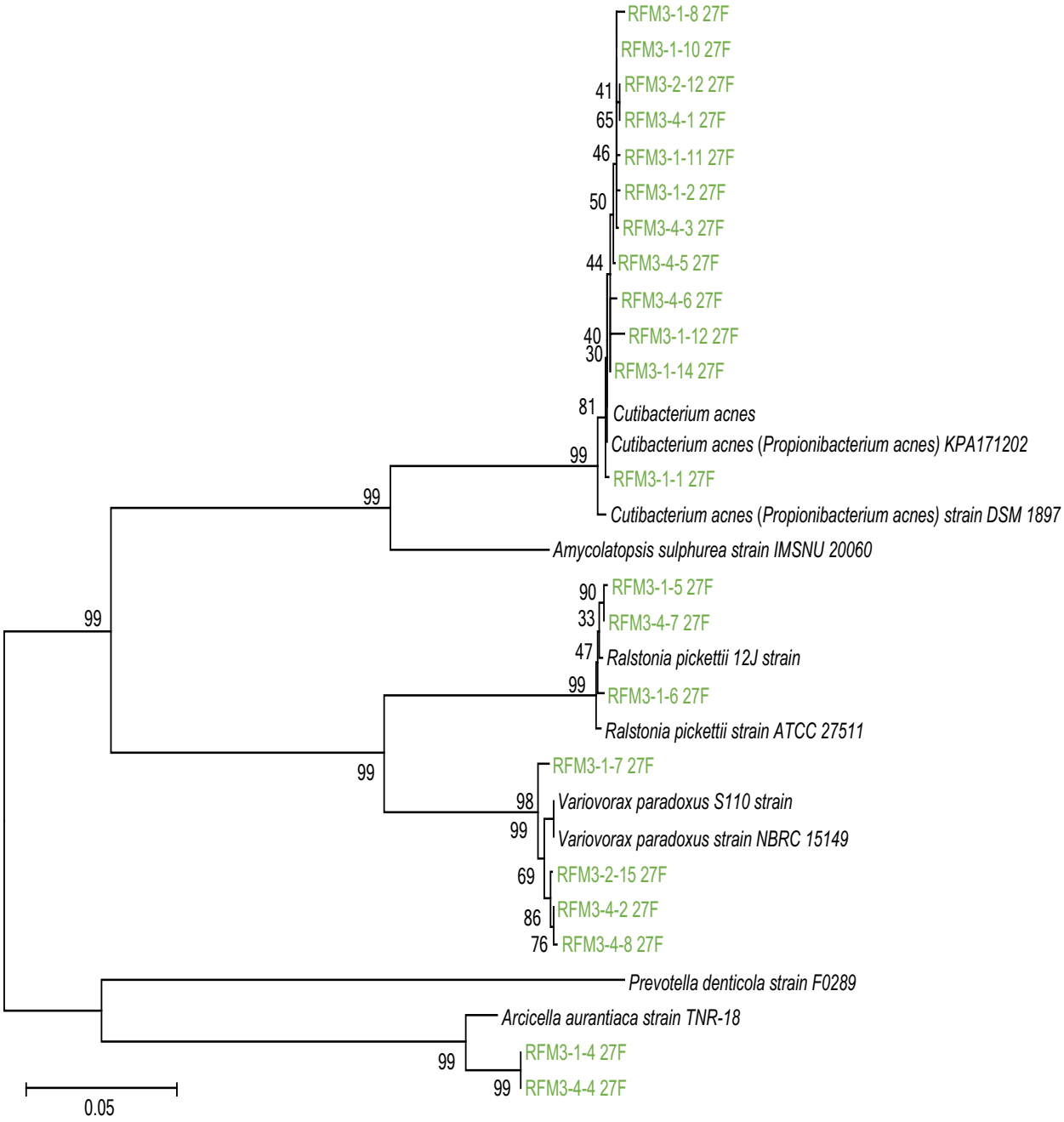


Figure 3.4- Unrooted neighbour-joining phylogenetic tree of River water samples from Cwm Llanerch passed through a 0.22µm filter. The green individuals correspond to the selected clones. Bar represents five substitutions per 100 nucleotides.

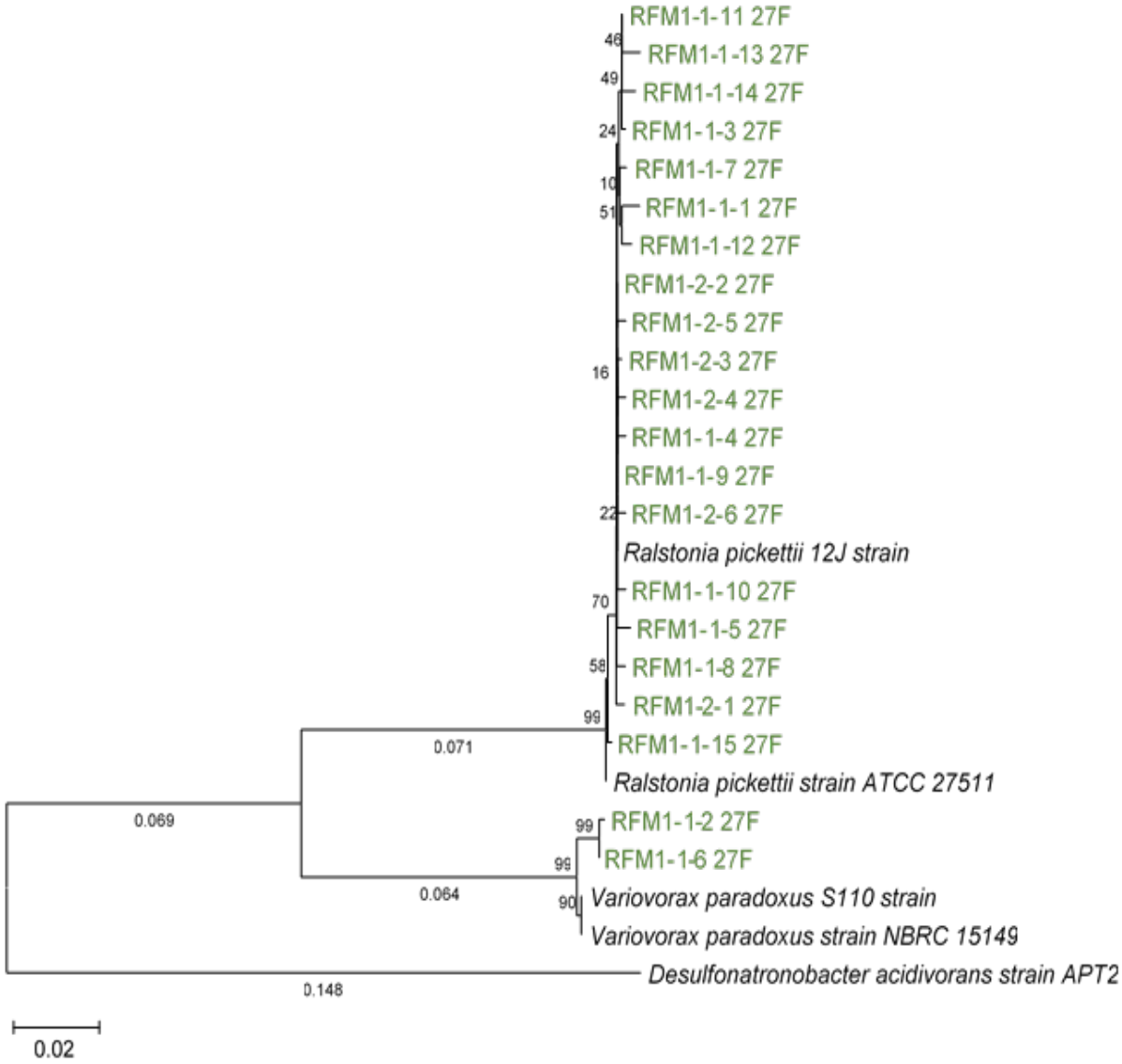


Figure 3.5- Unrooted neighbour-joining phylogenetic tree river water samples from Cwm Llanerch filtered through a 0.2 µm then 0.05 µm filter. The green individuals correspond to the selected clones. Bar represents two substitutions per 100 nucleotides.

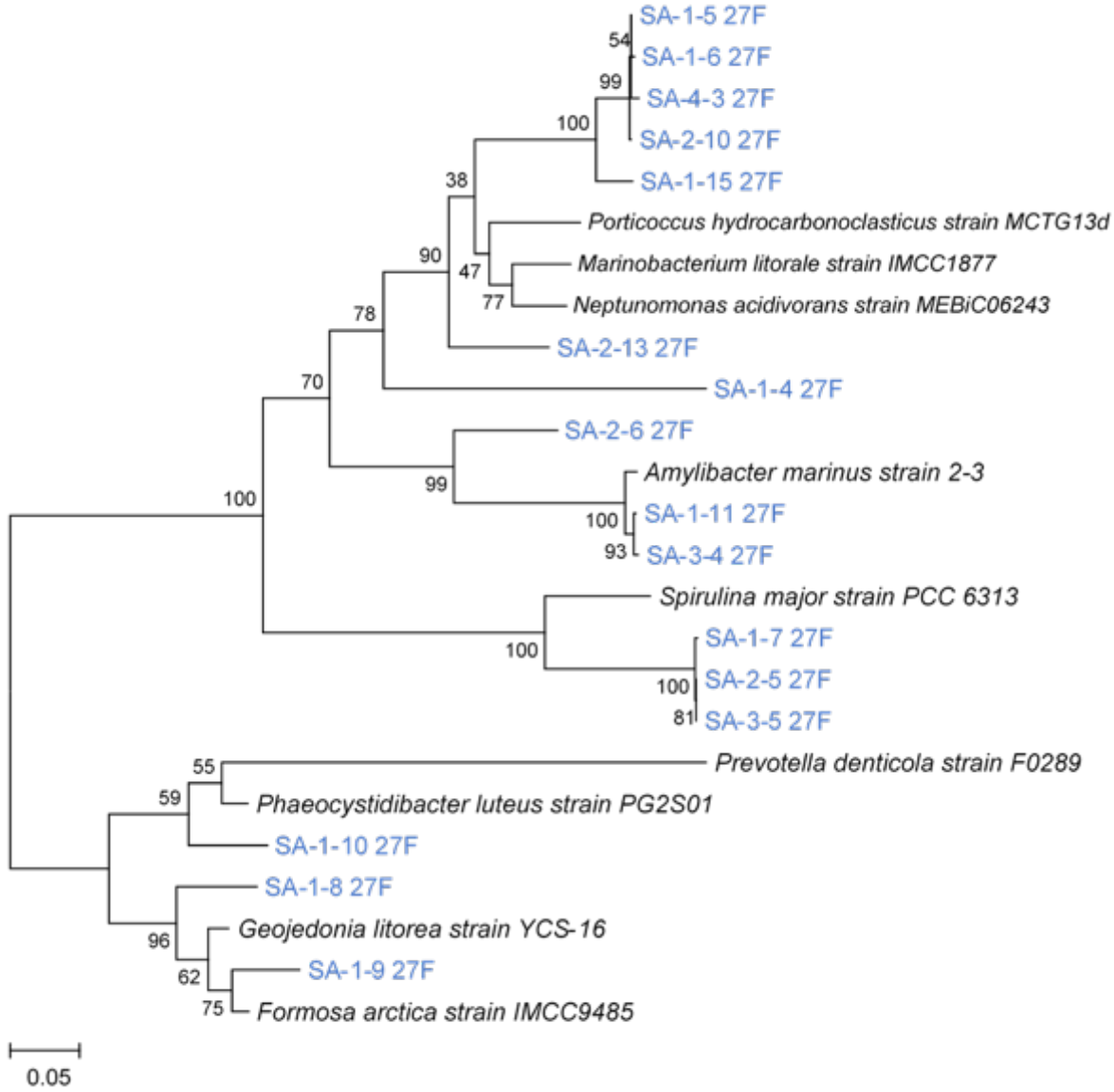


Figure 3.6- Unrooted phylogenetic tree based on 16S rRNA gene for seawater filtered through a 0.2 µm then 0.05 µm filter. The blue individuals correspond to the selected clones. Bar represents five substitutions per 100 nucleotides.

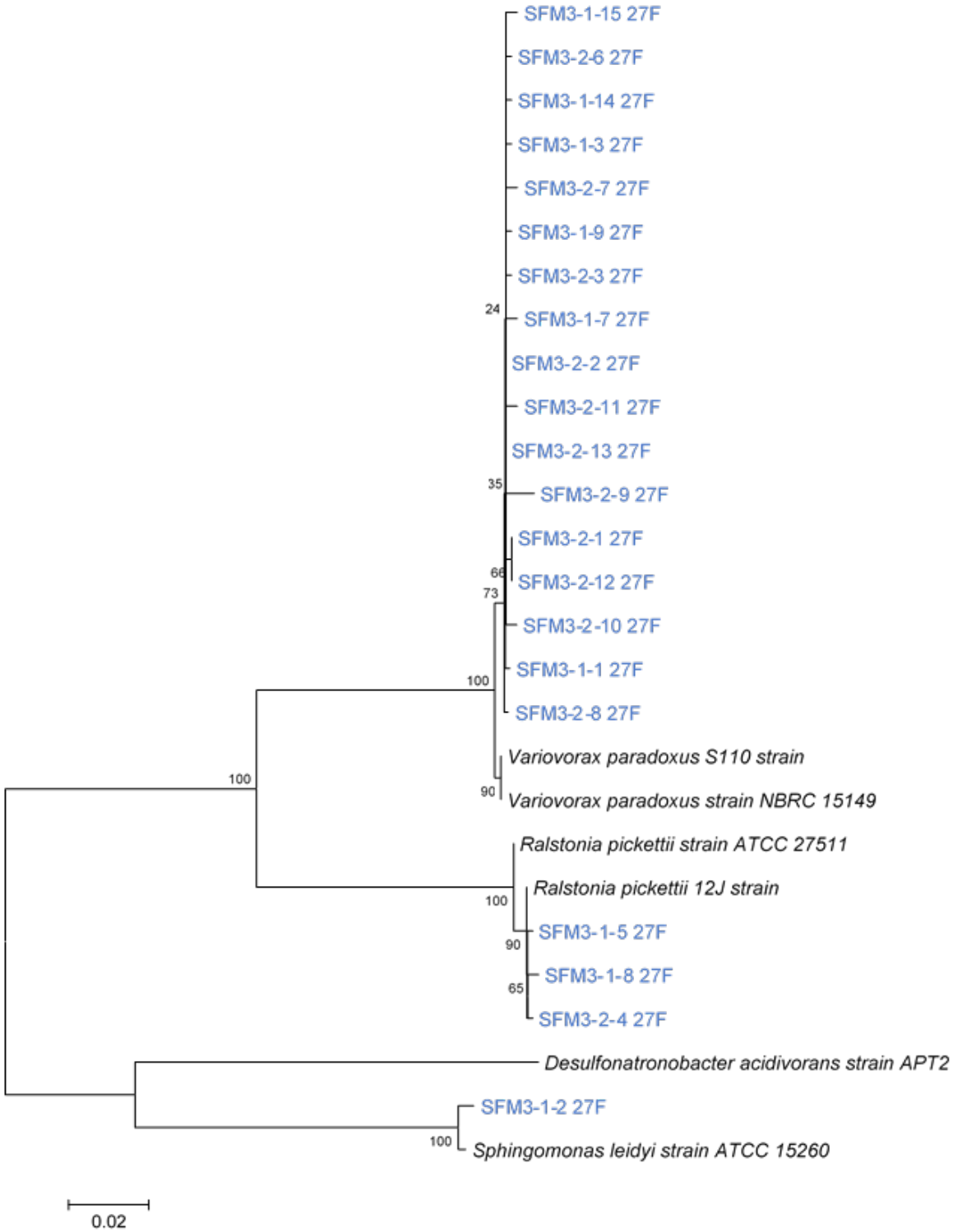


Figure 3.7- Unrooted neighbour-joining phylogenetic tree forriver water samples from Cwm Llanerch filtered through a 0.2 µm then 0.05 µm filter. The blue individuals correspond to the selected clones. Bar represents two substitutions per 100 nucleotides.

3.4. Discussion

3.4.1. Evaluation of the method used for microbial identification

Using *Oleispira antarctica RB-8* (Kube et al., 2013) as a model organism, our WGA methodology was shown to have high sensitivity to low DNA concentrations (10^{-5} ng/μL range). Further, the amplification of low concentrations of DNA showed negligible point mutations between the replicates and various DNA concentrations (Fig. 3.3). Phylogenetic analysis of the amplified control bacteria, after filtration processes, were identified as the correct species (*i.e.*, *O. antarctica* RB-8) via BLAST (Fig. 3.3 and Table 3.3). These observations align with previous investigations using MDA in a microbial ecology setting (Abuelncia et al., 2006; Biddle et al., 2008; Binga et al., 2008). For instance, using MDA, 99.2% genome coverage of *Escherichia coli* cells has been observed (Abuelncia et al., 2006). In fact, MDA is most effective at template quantities > 1 ng, where it is estimated that over 93% of genes in the amplified version are also found in the original template (Binga et al., 2008). However, at 0.1 or 0.01 ng of template, this decreases to 64% and 51%, respectively (Binga et al., 2008). This analysis shows that MDA is a robust approach for amplifying DNA and therefore it is likely that the species identified in the freshwater as well as seawater sample would be the precise representation of the microorganisms present within the samples.

3.4.2. Evaluation of microbial species in the different types of aqueous systems

Ralstonia pickettii and *Variovorax paradoxus* like sequences were detected in freshwater samples RFM1 and RFM3. They are known for their bioremediation abilities (Ryan et al., 2007; Pérez-Pantoja et al., 2012; Satola et al., 2013). These abilities include hydrocarbon degradation (Ryan et al., 2007), volatile organic compounds degradation (VOCs) (Ryan et al., 2007), polymer degradation (Satola et al., 2013), and also the catabolism of aromatic (Pérez-Pantoja et al., 2012) and sulfur compounds (Satola et al., 2013). They are typically found in soil, suggesting that runoff or soil erosion by the point of entry to the river. Significant levels

of *Cutibacterium acnes*, a common pathogenic species that typically resides on human skin (Perry and Lambert, 2006; Mollerup et al., 2016), however, were observed which may be attributed to possible sample contamination or sewage discharge. More likely the latter due to the fact that all negative controls (PCR grade water) showed no contamination.

In comparison, diverse ranges of bacteria were present in saltwater samples. The classes *Gammaproteobacteria* (Cho and Giovannoni, 2004; Kim et al., 2007; Williams et al., 2010; Gutierrez et al., 2012), *Alphaproteobacteria* (Gupta and Mok, 2007; Chen et al., 2012; Teramoto and Nishijima, 2014), *Cytophagia* (Kirchman et al., 2002; Sheu et al., 2010), and *Flavobacteria* (Kirchman et al., 2002; Zwart et al., 2002; Park et al., 2013; Kwon et al., 2014) were detected in the SA sample. All these are bacteria that are commonly found in marine environments. The function of each of these varies widely. For instance, *Cytophagia* and *Flavobacteria* are one of the primary consumers of dissolved organic matter (Kirchman et al., 2002). In addition, SFM3 samples contained microorganisms normally found in freshwater systems, which included the *Burkholderiales* order and *Actinobacteria* phylum. The difference in biodiversity may be due to the filter pore size (i.e. decreasing the filter pore size allowed us to see different minority populations not normally found through conventional methods).

In both aquatic environments, the identified bacteria present are significantly larger than the filter pore size (0.05 μm), although it should be noted that all identified species, except for *Cutibacterium acnes* (Perry and Lambert, 2006), are Gram-negative bacteria, which have been previously observed to have the ability to squeeze through small pores (Wang et al., 2007). The presence of these bacteria, larger than the filter pore size, may be attributed to the bacterial morphology and cell rigidity (Wang et al., 2007, 2008) and under certain environmental conditions, the cell may fall out of the range normally seen in laboratory conditions (Fedotova et al., 2012). In addition, the natural variations of size of the pores and the material of the filter (Wang et al., 2008) may also be a contributing factor. Another possibility is that cell fragments

(i.e eDNA) are passing through the filters instead of whole cells. Addition of microscopy evaluation may be able to aid in discerning whole cells from fragments. It should be noted that the DNA sequences were read in the forward direction only to roughly achieve a sense of which species are present and whether single or double filtration methodology affects the outcome of identification.

3.5. Conclusions

This study validated the use of the WGA methodology and its suitability to obtain readable DNA sequences from less than 50 mL of aqueous sample with a high level of accuracy. In reference to the effectiveness of filtration methodology, both methods are equally effective as there is no discernible difference in the observed communities between the single and double filter method. Therefore, we conclude that the single filter method represents a suitable in-field pre-processing method when identifying filterable organisms in aqueous solutions.

3.6. Acknowledgements and contributions

Lydia-Ann Ghuneim conducted the experiments, analysis, data interpretation and writing the manuscript. Francesca Brailsford assisted with freshwater sample collection. Peter Golyshin assisted with seawater collection. David Jones, Peter Golyshin, and Olga Golyshina provided significant revisions to the manuscript as well as insight into data interpretation. This work was carried out under the DOMAINE project, which is funded by the UK Natural Environment Research Council (NERC) (large grant NE/K010689/1). OG and PG acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government and thank for its support the European Union Horizon 2020 Research and Innovation Program [Blue Growth: Unlocking the Potential of Seas and Oceans, grant agreement no. 634486, Project acronym 'INMARE'].

3.7. References

Abuelncia, C. B., Wyborski, D. L., Garcia, J. A., Podar, M., Chen, W., Chang, S. H., et al. (2006). Environmental Whole-Genome Amplification To Access Microbial Populations in Contaminated Sediments. *Appl. Environ. Microbiol.* 72, 3291–3301. doi:10.1128/aem.72.5.3291-3301.2006.

Arneson, N., Hughes, S., Houlston, R., and Done, S. (2008). Whole-genome amplification by improved primer extension preamplification PCR (I-PEP-PCR). *Cold Spring Harb. Protoc.* 3, 1–5. doi:10.1101/pdb.prot4921.

Bergey, D. H. (2005). *Bergey's Manual of Systematic Bacteriology - Vol 2*. doi:10.1007/978-0-387-68572-4.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., and House, C. H. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci.* 105, 10583–10588. doi:10.1073/pnas.0709942105.

Binga, E. K., Lasken, R. S., and Neufeld, J. D. (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J.* 2, 233–241. doi:10.1038/ismej.2008.10.

Blagodatskikh, K. A., Kramarov, V. M., Barsova, E. V., Garkovenko, A. V., Shcherbo, D. S., Shelenkov, A. A., et al. (2017). Improved DOP-PCR (iDOP-PCR): A robust and simple WGA method for efficient amplification of low copy number genomic DNA. *PLoS One* 12, 1–12. doi:10.1371/journal.pone.0184507.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S. J. (2012). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J.* 7, 592–602. doi:10.1038/ismej.2012.122.

Chen, H., Jogler, M., Rohde, M., Klenk, H. P., Busse, H. J., Tindall, B. J., et al. (2012). Reclassification and emended description of Caulobacter leidyi as Sphingomonas leidyi comb. nov., and emendation of the genus Sphingomonas. *Int. J. Syst. Evol. Microbiol.* 62, 2835–2843. doi:10.1099/ijs.0.039636-0.

Cho, J. C., and Giovannoni, S. J. (2004). Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. *Appl. Environ. Microbiol.* 70, 432–440. doi:10.1128/AEM.70.1.432-440.2004.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Fedotova, A. V., Belova, S. E., Kulichevskaya, I. S., and Dedysh, S. N. (2012). Molecular identification of filterable bacteria and archaea in the water of acidic lakes of northern Russia. *Microbiology* 81, 281–287. doi:10.1134/S002626171203006X.

Gupta, R. S., and Mok, A. (2007). Phylogenomics and signature proteins for the alpha Proteobacteria and its main groups. *Bmc Microbiol.* 7, 106. doi:Artn 106\rDoi 10.1186/1471-2180-7-106.

Gutierrez, T., Nichols, P. D., Whitman, W. B., and Aitken, M. D. (2012). Porticoccus hydrocarbonoclasticus sp. nov., an aromatic hydrocarbon-degrading bacterium identified in laboratory cultures of marine phytoplankton. *Appl. Environ. Microbiol.* 78, 628–637. doi:10.1128/AEM.06398-11.

Hall, T. T. A. T. (1999). BioEdit: a user-friendly biological sequence alignment editor and

analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98. doi:citeulike-article-id:691774.

Jones, A., and Goodfellow, M. (2012). "Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria," in *Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria, Volume 5, Parts 1-2*, 437–464. doi:10.1007/978-0-387-86233-4.

Kim, H., Choo, Y. J., Song, J., Lee, J. S., Lee, K. C., and Cho, J. C. (2007). Marinobacterium litorale sp. nov. in the order Oceanospirillales. *Int. J. Syst. Evol. Microbiol.* 57, 1659–1662. doi:10.1099/ijs.0.64892-0.

Kirchman, D. L., Hugenholtz, P., Goebel, B. M., Pace, N. R., Karner, M. B., DeLong, E. F., et al. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* 39, 91–100. doi:10.1111/j.1574-6941.2002.tb00910.x.

Kratzer, S., Buchan, S., and Bowers, D. G. (2003). Testing long-term trends in turbidity in the Menai Strait, North Wales. *Estuar. Coast. Shelf Sci.* 56, 221–226. doi:10.1016/S0272-7714(02)00159-2.

Kube, M., Chernikova, T. N., Al-Ramahi, Y., Beloqui, A., Lopez-Cortez, N., Guazzaroni, M.-E., et al. (2013). Genome sequence and functional genomic analysis of the oil-degrading bacterium Oleispira antarctica. *Nat. Commun.* 4, 2156. doi:10.1038/ncomms3156.

Kwon, T., Baek, K., Lee, K., Kang, I., and Cho, J. C. (2014). Formosa arctica sp. nov., isolated from Arctic seawater. *Int. J. Syst. Evol. Microbiol.* 64, 78–82. doi:10.1099/ijs.0.056465-0.

Mollerup, S., Friis-Nielsen, J., Vinner, L., Hansen, T. A., Richter, S. R., Fridholm, H., et al. (2016). Propionibacterium acnes: Disease-causing agent or common contaminant?

detection in diverse patient samples by next- generation sequencing. *J. Clin. Microbiol.* 54, 980–987. doi:10.1128/JCM.02723-15.

Morris, A. W., and Foster, P. (1971). The Seasonal Variation of Dissolved Organic Carbon in the Inshore Waters of the Menai Strait in Relation to Primary Production. *Limnol. Oceanogr.* 16, 987–989.

Olenina, I., Hajdu, S., Edler, L., Wasmund, N., Busch, S., Göbel, J., et al. (2006). Biovolumes and size-classes of phytoplankton in the Baltic Sea. *HELCOM Balt.Sea Environ. Proc.* 106, 144 pp.

Panikov, N. S. (2005). Contribution of Nanosized Bacteria to the total biomass and activity of a soil microbial community. *Adv. Appl. Microbiol.* 57, 245–296.

Park, S., Lee, J. S., Lee, K. C., and Yoon, J. H. (2013). Geojedonia litorea gen. nov., sp. nov., a member of the family Flavobacteriaceae isolated from coastal seawater. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* 103, 1255–1262. doi:10.1007/s10482-013-9904-6.

Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M., Seeger, M., Pieper, D. H., et al. (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. *Environ. Microbiol.* 14, 1091–1117. doi:10.1111/j.1462-2920.2011.02613.x.

Perry, A. L., and Lambert, P. A. (2006). Propionibacterium acnes. *Lett. Appl. Microbiol.* 42, 185–188. doi:10.1111/j.1472-765X.2006.01866.x.

Ryan, M. P., Pembroke, J. T., and Adley, C. C. (2007). Ralstonia pickettii in environmental biotechnology: Potential and applications. *J. Appl. Microbiol.* 103, 754–764. doi:10.1111/j.1365-2672.2007.03361.x.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for

reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–25. doi:citeulike-article-id:93683.

Satola, B., Wübbeler, J. H., and Steinbüchel, A. (2013). Metabolic characteristics of the species Variovorax paradoxus. *Appl. Microbiol. Biotechnol.* 97, 541–560. doi:10.1007/s00253-012-4585-z.

Sheu, S. Y., Yang, C. S., Chen, M. H., Arun, A. B., Young, C. C., and Chen, W. M. (2010). Arcicella aurantiaca sp. nov., isolated from stream water. *Int. J. Syst. Evol. Microbiol.* 60, 2979–2983. doi:10.1099/ijs.0.021527-0.

Simpson, J. H., Vennell, R., and Souza, A. J. (2001). The salt fluxes in a tidally-energetic estuary. *Estuar. Coast. Shelf Sci.* 52, 131–142. doi:10.1006/ecss.2000.0733.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., et al. (2006). Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.* 1, 1965–1970. doi:10.1038/nprot.2006.326.

Tajima, F., and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1, 269–285. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6599968.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi:10.1093/molbev/msr121.

Teramoto, M., and Nishijima, M. (2014). Amylibacter marinus gen. nov., sp. nov., isolated from surface seawater. *Int. J. Syst. Evol. Microbiol.* 64, 4016–4020. doi:10.1099/ijs.0.065847-0.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi:10.1093/nar/22.22.4673.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science (80-.).* 308, 554–557. doi:10.1126/science.1107851.

Tripp, H. J. (2013). The unique metabolism of SAR11 aquatic bacteria. *J. Microbiol.* 51, 147–153. doi:10.1007/s12275-013-2671-2.

Velimirov, B. (2001). Nanobacteria, Ultramicrobacteria and Starvation Forms: A Search for the Smallest Metabolizing Bacterium. *Microbes Environ.* 16, 67–77. doi:10.1264/jsme2.2001.67.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi:10.1126/science.1093857.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi:10.1021/es0707198.

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* 42, 6749–6754. doi:10.1021/es800720n.

Williams, K. P., Gillespie, J. J., Sobral, B. W. S., Nordberg, E. K., Snyder, E. E., Shallom, J. M., et al. (2010). Phylogeny of gammaproteobacteria. *J. Bacteriol.* 192, 2305–2314.

doi:10.1128/JB.01480-09.

Yang, S.-H., Seo, H.-S., Lee, J.-H., Kim, S.-J., and Kyoung Kwon, K. (2014). Neptunomonas acidivorans sp. nov., isolated from sediment, and emended description of the genus Neptunomonas. *Int. J. Syst. Evol. Microbiol.* 64, 3650–3654. doi:10.1099/ijs.0.064253-0.

Zhou, Y., Su, J., Lai, Q., Li, X., Yang, X., Dong, P., et al. (2013). Phaeocystidibacter luteus gen. nov., sp. nov., a member of the family Cryomorphaceae isolated from the marine alga Phaeocystis globosa, and emended description of Owenweeksia hongkongensis. *Int. J. Syst. Evol. Microbiol.* 63, 1143–1148. doi:10.1099/ijs.0.030254-0.

Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., and Han, S. K. (2002). Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* 28, 141–155. doi:10.3354/ame028141.

Chapter 4

Prokaryotic diversity of filtered (0.22 μm) and unfiltered fractions residing in the Conwy River

Abstract

Microorganisms in lotic systems are vital to regulating and maintaining ecosystem function. In this study, we examined the microbial community residing in the Afon Conwy (Conwy River) using metagenomics. We compared microbial fractions in river water samples passed through 0.22 μm pore size filters alongside those in unfiltered samples using 16S rRNA gene amplicon clone library Sanger-sequencing, 16S rRNA gene amplicon sequencing using Oxford Nanopore™ platform and shotgun metagenome sequencing (Illumina® HiSeq). We examined the presence of clusters of orthologous groups of proteins (COGs) in both fractions. Finally, we compared these COG distributions, via cluster analysis, to 14 metagenomes (from an environmental source) and 20 genomes (of various genome sizes) obtained from the IMG/MER database. The dominate phyla in both fractions and across all sequencing methods were *Proteobacteria*, *Bacteriodetes*, and *Firmicutes*. Sanger sequencing contained a high number of sequences affiliated with *Chloroflexi* and *Actinobacteria*. Shotgun sequencing and Nanopore™ showed similar distributions where (1) the dominate phylum of both fractions was *Proteobacteria*, (2) the filtered fraction had a consistently higher percentage of sequences associated with *Firmicutes* than the unfiltered fraction, and (3) the unfiltered fraction had a higher percentage of *Bacteriodetes* than the filtered fraction. The 16S rRNA sequences retrieved from shotgun and Sanger sequencing showed a degree of similarity to many uncultured representatives and candidate phyla in the filtered fraction. There are more COGs in the filtered fraction that under the functional categories of replication, recombination and repair; and cell wall/membrane/envelope biogenesis than the unfiltered fraction. Cluster analysis in the form of a PCA, K-clustering, and hierarchal clustering show that the distribution of COGs in the filtered fraction is unique when compared to the other metagenomes and the unfiltered fraction. Clustering metagenomes

against single genomes revealed that the filtered fraction COG distribution is closely related to COG distribution of organisms with limited/streamlined genomes.

Keywords: lotic systems, filterable microorganisms, metagenomics, 16S rRNA gene amplicon sequencing, shotgun metagenome sequencing, nanopore, clusters of orthologous groups of proteins (COGs)

4.1. Introduction

It is estimated that 38,600 km^3 of lotic freshwater carries 0.9 Pg of carbon per year to marine systems (Sigee, 2005; Cole et al., 2007; Battin et al., 2008), effectively making lotic systems integral to global carbon cycling. The carbon in these systems is largely derived from allochthonous sources (i.e. from an external source in the form of leaf detritus, etc), which is further processed and degraded to dissolved organic matter (DOC) by heterotrophic microorganisms (Meyer, 1994). Therefore, knowledge of the composition of lotic communities is key to gaining a better understanding of the carbon cycle. However, studying them can be challenging due to the large number of uncultivable microorganisms (i.e. 'microbial dark matter') ubiquitous to earth's microbiome.

The advent of culture-independent methodologies allowed for further study of 'microbial dark matter', in the form of metagenomics. Metagenomics, also known as community genomics, is the study of microbial genomic material obtained directly from an environmental source (Handelsman et al., 1998). Stemming off from the parent field of genomics, it was first used to create libraries to identify new microbial species in a sample (Desai et al., 2012) . These techniques are used to recognise sequences of the microbial members and to predict their function within a particular environment Metagenomics requires the collection of genetic material to compare to a database of pre-existing genomes to map the organism's taxonomic relationship (Baker and Banfield, 2003; Wrighton et al., 2012; Brown et al., 2015). Two distinct methodologies have emerged: 'sequence-centred' (also

termed ‘genome-gazing’) metagenomics, relying on the bioinformatics and sequencing data analysis and ‘activity-based’ metagenomics, which relies on the functional screens of metagenomic libraries for enzymatic activities or metabolites, mostly, for biotechnological applications (Handelsman, 2004).

The focus of our study is lotic system of the Conwy River in North Wales (UK). There has been continuous human occupation in the Conwy catchment for hundreds of years involving activities ranging from mining to agriculture (Emmett et al., 2016). Because of this, the river, catchment, and estuary have been subject of numerous studies, ranging from spatial patterns (Emmett et al., 2016), virome investigations (Adriaenssens et al., 2018; Farkas et al., 2018), trace metal prevalence (Zhou et al., 2003), and lipid biomarkers (Mudge and Norris, 1997). Yet there are no studies exploring the general microbial ecology. However, recent investigations have suggested that this is important to understanding freshwater biogeochemical cycling. The uptake of ^{14}C-labelled low molecular weight DOC in freshwater samples of different fractions has revealed significant activity in water passed through a 0.22 μm membrane filter (Brailsford et al., 2017). It was previously though that using ultrafiltration on aqueous sample would render them sterile, but as evident from this study and many others, there is a significant number of microorganisms that can pass through these filters with little to no resistance, i.e. filterable microorganisms (Hahn et al., 2004; Wang et al., 2007, 2008; Brailsford et al., 2017). The identity and ecological role that these organisms play is varied depending on the ecosystem. In freshwater, they can participate in DOC cycling as well as be prey for larger protists (Duda et al., 2012; Salcher, 2014; Dang and Lovell, 2016).

Therefore, the purpose of this study was to examine the microbial content of the 0.22 μm filtered fraction and to compare it to the bulk (unfiltered) fraction of the Conwy River in terms of community makeup and predicted functionalities using culture-independent

methods, i.e. metagenomics. We used Sanger sequencing of 16S rRNA gene amplicon libraries, Oxford Nanopore™ 16S rRNA gene amplicon sequencing, as well as metagenome shotgun sequencing to examine overall community composition and clusters of orthologous groups (COGs) present.

COGs are a compilation of homologous genes that is made up of both orthologues and paralogues (Tatusov et al., 2000). The COG database (https://www.ncbi.nlm.nih.gov/COG/) categorizes proteins domains based by their inherent structures. By examining the COGs present, we hope to determine overall community potential functionality, relationships between other community members, and to compare these to other genomes and metagenomes contained in public databases.

4.2. Methods

4.2.1. Description of the sites

The freshwater samples were obtained from the Conwy catchment, which is located in North Wales. Its main drainage is in the form of the river Conwy (Afon Conwy). Mean annual rainfall ranges from 500-3500 mm and the mean annual air temperature ranges from 5-15 °C with an average annual temperature of 10 °C (Emmett et al., 2016). The river itself is 55 km long; starting from Llyn Conwy (450 m above sea level) and drains the Migneint, a large peatland bog that is a major store of carbon. Three tributaries (Machno, Lledr, and Llugwy), originating from the eastern side of the Snowdonia mountain range, join the main river further downstream before reaching the tidal limit (20 km inland) (Emmett et al., 2016). Average concentrations at the tidal limit are as follows: nitrite 0.2-2.8 mg/L, ammonium <0.03-0.04 mg/L, phosphate <0.02-0.05 mg/L, and dissolved organic carbon 1.5-10 mg/L (Emmett et al., 2016). The pH ranges from 5.7 to 7.2. The primary site used in this study (code: NM29) is located at Cwm Llanerch and is associated with the main Conwy River (53°

6' 24.7068" N, 3° 47' 28.7556" W). The site is located approximately 4 km from the tidal limit (Simpson et al., 2001).

4.2.2. Collection, preparation, and whole genome amplification (WGA)

All water samples were collected from Cwm Llanerch in December 2015. We collected five samples of 50 mL volume (150 mL each) of unfiltered samples (UF) and of filtered samples (F) (50 mL of sample passed through a 0.22 µm Sterivex® PVDF (hydrophilic polyvinylidene fluoride) filter (Millipore Corporation, Billerica, MA, USA). Unfiltered samples were centrifuged at 39,443 *g* for 15 minutes. Filtered samples were centrifuged for 20 min at 28,919 *g*. Supernatants from both treatments were removed. The remaining pellet was then washed (x3) with PBS (phosphate buffer solution) (pH 7.4). Whole genome amplification (WGA) was achieved using the GE Illustra Ready-to-go Genomi Phi V3 DNA amplification kit® (GE Healthcare UK Limited, Little Chalfont, Buckinghamshire, UK) as per the manufacturer's protocol. Concentration of DNA in WGA was estimated using Quan-it dsDNA Assay kit® (Invitrogen™, Life Technologies Corporation, Carlsbad, CA, USA) as per the manufacturer's protocol. Successful WGA products were visualised in Bio-Rad ChemiDoc XR after tris-borate 0.8% agarose gel electrophoresis in SybrSafe (ThermoFisher).

4.2.3. Generating 16S rRNA Clonal Libraries

16S rRNA was amplified from the WGA product using the universal primers 27F (5'-AGAGTTTGGATCMTGCTCAG-3') and 1492R (5'-TACCTTGTTACGACTT-3') with MyTaq™ DNA polymerase (Bioline Reagents Ltd, UK). A QIAquick gel extraction kit® (Qiagen) was used to isolate the PCR product. The PCR product was then ligated into vector TOPO 2.1 and then transformed into TOP10F' *E. coli* electro competent cells (Invitrogen™, Life Technologies Corporation, Carlsbad, CA, USA) using the TOPO ® TA cloning® Kit (Invitrogen™, Life Technologies Corporation, Carlsbad, CA, USA) as per the manufacturer's

protocol. The cells were grown on LB media with kanamycin (50 µg/mL) and Xgal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside) (40 mg/mL). White colonies were selected and inserts were PCR-amplified using vector specific primers M13F (5´-GTAAAACGACGGCCAG-3´) and M13R (5´-CAGGAAACAGCTATGAC-3´). All PCR reactions were performed in a Bio-Rad Tetrad 2 Peltier thermal cycler with the following program: denaturation step at 95°C for 2 minutes; followed by 30 cycles, each including denaturation at 94°C for 1 minute, annealing (50°C for 1 minute), and DNA synthesis (72°C for 2 minutes). Final extension at 72°C for 10 minutes followed by cooling to 4°C. PCR products were visualised in Bio-Rad ChemiDoc XR after tris-borate 0.8% agarose gel electrophoresis in SybrSafe (ThermoFisher) and sent to Macrogen Inc, Seoul, Korea for Sanger sequencing.

The resultant sequences were screened and trimmed in BioEdit (Hall, 1999). For phylogenetic analysis of 16S rRNA genes, contigs were first assembled utilizing the PRABI-Doua CAP3 assembly program (http://doua.prabi.fr/software/cap3) Then full multiple alignment was done using ClustalW (Thompson et al., 1994) using 1,000 bootstraps. Bellerophan software (http://comp-bio.anu.edu.au/bellerophon/bellerophon.pl) was used to detect chimeric sequences. The Neighbour joining unrooted phylogenetic tree was created based on a maximum likelihood best fit model using the Kimura-2 parameter (Saitou and Nei, 1987) with uniform distribution and 1,000 pseudo-replicates (Felsenstein, 1985) in MEGA7 (Kumar et al., 2016). Taxonomically closest relatives were determined using the BLAST 2.2 program using both the 16S bacteria-archaea and nucleotide collection (nr/nt) (www.ncbi.nim.nih.gov/blast/).

4.2.4. Shotgun sequencing and phylogenetic analysis

80 µL of ammonium acetate, 320 µL of absolute ethanol, and 30 µL of water were added to 20 µL of WGA product. Then product was washed with 300 µL of 70% ethanol then

air dried for 10 minutes to remove excess ethanol. Next, 300 µL of distilled water was added and then incubated at 50 °C for 10 min. The concentration of DNA was estimated using Quan-it dsDNA Assay kit® (Invitrogen™, Life Technologies Corporation, Carlsbad, CA, USA) as per the manufacturer's protocol. The final DNA concentration of both the unfiltered and filtered fractions were 44.1 ng/l and 110.3 ng/l respectively. DNA was then shotgun-sequenced using Illumina HiSeq1500 platform, assembled and annotated at Fidelity Systems, Inc (Gaithersburg, MD). Data were submitted and organized in the IMG/MER database (Markowitz et al., 2012). Resultant 16S rRNA gene sequences from shotgun-sequencing longer than 350 bp were selected for further phylogenetic analysis. The sequences were trimmed in BioEdit (Hall, 1999) and full multiple alignment was done using ClustalW (Thompson et al., 1994) with 1,000 bootstraps. Neighbour joining unrooted phylogenetic tree was created based on maximum likelihood best fit model using the Kimura-2 parameter method (Kimura, 1980) with a gamma distribution (shape parameter=0.63) and 1,000 pseudoreplicate bootstraps (Felsenstein, 1985) in MEGA7 (Kumar et al., 2016). Taxonomically closest relatives were determined using the BLAST 2.2 program using both the 16S bacteria-archaea and nucleotide collection (nr/nt) (www.ncbi.nim.nih.gov/blast/).

4.2.5. Amplicon sequencing (Oxford Nanopore)

The bacterial 16S rRNA genes were amplified from the WGA product with the universal primers 27F (5'-AGAGTTTGGATCMTGCTCAG-3') and 1492R (5'-TACCTTGTTACGACTT-3') using the 16S barcoding kit® (Oxford Nanopore Technologies, UK). The MinION Flow Cell® (Oxford Nanopore Technologies, UK) was primed using the Flow Cell Priming kit® (Oxford Nanopore Technologies, UK) as per the manufacturer's protocol. Finally, samples were loaded onto a MinION Flow Cell® (Oxford Nanopore Technologies, UK) and sequenced using MinION® device (Oxford Nanopore technologies, UK). Data processing and base-calling was done via the 16S workflow in the EPI2ME

platform. Taxonomically closest relatives to the base-called sequences were determined via the BLAST 2.2 program using 16S bacteria-archaea and nucleotide collection (nr/nt) (www.ncbi.nim.nih.gov/blast/).

4.2.6. Statistical analysis and heatmap creation

We used the R programming language (R Core Team, 2017) for statistical analysis and heatmap generation. The following packages were used: ggplot2 (Wickham, 2009), grid (R Core Team, 2017), gridExtra(Auguie, 2016), plyr (Wickham, 2011), dplyr (Wickham and Francois, 2016), tidyr (Wickham, 2017), reshape2 (Wickham, 2007), readr (Wickham et al., 2017b), tibble (Wickham et al., 2017a). Rarefaction curve was generated using the vegan package (Oksanen et al., 2018).

4.3. Results

4.3.1. 16S rRNA gene amplicon libraries sequencing

In this analysis, we examined the content of filtered freshwater fraction via sequencing 16S rRNA gene amplicon library. We chose site NM29 because it is the body of the main Conwy River and is 4 km upstream from the tidal limit.

The principal phyla across all fractions were *Chlorofexi* (31.6%), *Cyanobacteria* (15.9%), *Proteobacteria* (15.8%), *Actinobacteria* (10.5%), and *Plantomycetes* (5.3%) (Table 4.1). According to Table 3, there was a much higher percentage of *Chloroflexi* members detected in the filtered samples than unfiltered. Approximately 40% of the species in the filtered samples were of the *Chloroflexi* phylum. Conversely, a much higher percentage of *Proteobacteria* were detected in the unfiltered samples.

A majority (89%) of the sequences from both filtered and unfiltered samples have a low percent identity to known species, between 70%-89% (Table 4.1). As clearly illustrated in the phylogenetic tree (Fig 4.1), there is a large cluster that contains sequences that have less than 77% percent identity to the query sequence. The majority are from the filtered

sample. BLAST analysis of the cluster showed the closest relations to the following known phyla: *Chloroflexi*, *Cyanobacteria*, *Proteobacteria*, *Actinobacteria*, and *Plantomycetes*. Of all the species identified, only three have documented dimensions allowing them to pass through 0.22 µm-pores, *Dehalococcoides mccartyi, Peloniea submarnia,* and *Campylobacter upsaliensis* (Table 4.1).

The sequences were more closely related (i.e over 90% percent identity) to many known culturable species (Fig. 4.1 and Tables 4.1-4.4). There was only one sequence with an identity of less than 90%, which was clone 21B. Five clones have percent identities above 90%, with TM7 ('*Candidatius* Saccharibacteria') and '*Ca.* Yanofskybacteria'; which have been previously been predicted to be potential nanoorganisms (Brown et al., 2015; He et al., 2015) (Table 4.2). Only two clones, 2B and 2D, from this set were found in the unknown cluster (Fig 4.1).

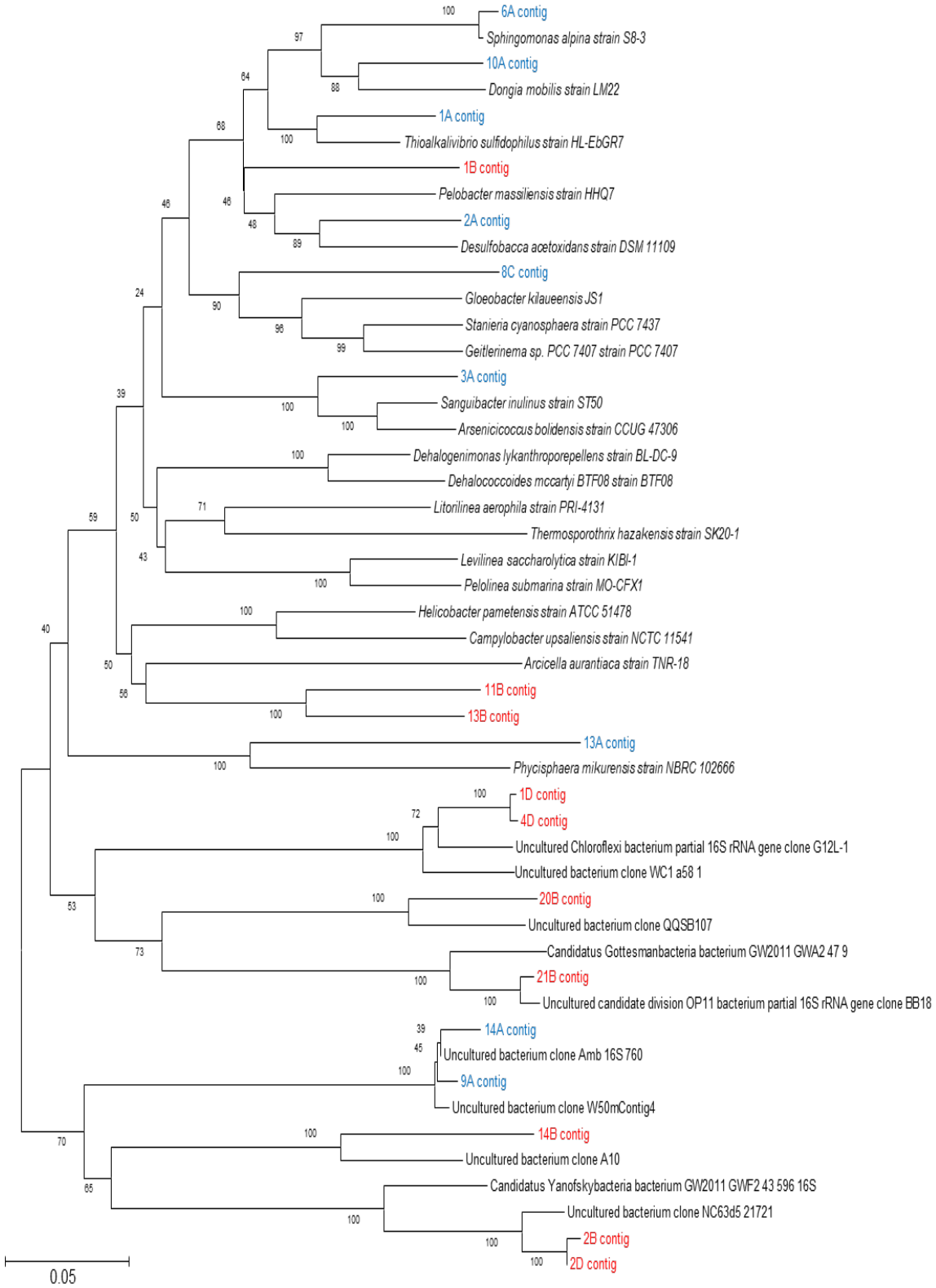


Figure 4.1- Unrooted neighbour joining tree of NM29 site (Cwn llanerch, main Afon Conwy). Clones from the unfiltered sample are in blue and those from the filtered sample (0.22 µm sterivex filter) are in red. Bar represents five substitutions per 100 nucleotides.

Table 4.1- Summary of taxa retrieved from 16S rRNA BLAST search. The BLAST results are listed along with percent identities and associated sequence. Phylum, morphology, cellular dimensions, and referenced are also indicated here. NA designates information was unavailable.

Sequence name	*Residing fraction*	*16S rRNA BLAST match species*	*Phylum*	*Percent Identity*	*Morphology*	*Dimension 1 (diameter) (µm)*	*Dimension 2 (length) (µm)*	*Reference*
10A contig	Unfiltered	*Dongia mobilis*	*Proteobacteria*	88%	Slightly curved to straight rod	0.3-0.5	0.6-1.0	(Liu et al., 2010)
6A contig	Unfiltered	*Sphingomonas alpina*	*Proteobacteria*	98%	Short rod	0.6-0.8	1.1-1.4	(Margesin et al., 2012)
1A contig	Unfiltered	*Thioalkalivibrio sulfidophilus*	*Proteobacteria*	90%	Slightly curved rod	0.4	3.0-8.0	(Muyzer et al., 2011; Sorokin et al., 2012)
2A contig	Unfiltered	*Desulfobacca acetoxidans*	*Proteobacteria*	84%	Oval	1.3	1.9-2.2	(Oude Elferink et al., 1999)
1B contig	Filtered	*Pelobacter massiliensis*	*Proteobacteria*	85%	Thin rod	0.4-0.5	0.7-1.5	(Schnell et al., 1991)
3A contig	Filtered	*Sanguibacter inulinus*	*Actinobacteria*	89%	Irregular rods	NA	NA	(Pascual et al., 1996)
21B contig	Filtered	*Arsenicicococcus bolidensis*	*Actinobacteria*	75%	Coccoid	NA	NA	(Collins et al., 2004; Routh et al., 2007)
1D contig	Filtered	*Stanieria cyanosphaera*	*Cyanobacteria*	75%	Coccoid	Baeocyte: 1.0-2.0 Vegetative: 30	NA	(Bergey, 2005)
11B contig	Filtered	*Geitlerinema sp PCC 7406*	*Cyanobacteria*	78%	Filamentous	1-4	NA	(Anagnostidis, K. and Komárek, 1985; Bergey, 2005)
8C contig	Unfiltered	*Gloebacter kilaueensis*	*Cyanobacteria*	84%	Rod	~1.5	~3.5	(Saw et al., 2013)
13B contig	Filtered	*Thermosphorothrix hazakensis*	*Chloroflexi*	81%	Filamentous, "fig like" spores	Spores: 1.0-2.0	Spores: 0.7-1.0	(Yabe et al., 2010)
9A contig	Unfiltered	*Dehalogenimonas lykanthroporepellens*	*Chloroflexi*	76%	Irregular coccoid	0.3-0.6	NA	(Moe et al., 2009)
20B contig	Filtered	*Dehalococcoides mccartyi*	*Chloroflexi*	74%	Disk	0.1-0.2	0.3-1.0	(Löffler et al., 2013; Pöritz et al., 2013)
14A contig	Unfiltered	*Litorilinea aerophila*	*Chloroflexi*	77%	Filamentous	0.45-0.55	100	(Kale et al., 2013)
2B contig	Filtered	*Levilinea saccharolytica*	*Chloroflexi*	74%	Filamentous	0.4-0.5	100	(Yamada et al., 2006)
4D contig	Filtered	*Pelolinea submarina*	*Chloroflexi*	76%	Filamentous	0.130-0.150	10	(Imachi et al., 2014)
13A contig	Unfiltered	*Phycisphaera mikurensis*	*Plantomycetes*	80%	Coccoid	0.5-1.5	NA	(Fukunaga et al., 2009)
2D contig	Filtered	*Helicobacter pametensis*	*Proteobacteria*	75%	Rod	0.4	1.5	(Dewhirst et al., 1994)
14B contig	Filtered	*Campylobacter upsaliensis*	*Proteobacteria*	74%	Curved or spiral rods	0.2-0.5	0.5-8	(Goossens et al., 1990)

Table 4.2- Summary of taxa retrieved from nr/nt BLAST search. Top BLAST matches (nr/nt) from the amplicon library sequence analysis. Listed are sequence identities (%), environmental location, NCBI accession number, and reference. Information and references were retrieved from the NCBI database**.**

Sequence name	*Residing fraction*	*Nr/nt BLAST match*	*Accession number (NCBI)*	*Percent Identity*	*Isolation location*	*References*
1A contig	Unfiltered	Uncultured bacterium clone SC142	EU735630	95.4%	oil contaminated soil in Jidong Oilfield, China	(Liu et al., 2009)
2A contig	Unfiltered	Uncultured bacterium clone p7o13ok	FJ478616	96.0%	undisturbed tall grass prairie, top 5 cm, OK, USA	(Youssef et al., 2009)
3A contig	Unfiltered	Uncultured actinobacterium clone LT1D11	EU117831	99.7%	lake epilimnion, WI , USA	(Newton et al., 2007)
6A contig	Unfiltered	Sphingomonas sp. 3B3	EU337119	99.3%	Fermented milk, South Africa	(Bauer et al., 2009)
8C contig	Unfiltered	Uncultured bacterium clone KWK12F.82	JN656845	98.4%	hypolimnion (hypoxic bottom) water from permafrost thaw pond KWK12, Nunavik, Canada	(Rossi et al., 2013)
9A contig	Unfiltered	Uncultured bacterium clone BSB0101-06	JN397714	95.8%	River bank of Jing-Mei river, Taipei, Taiwan	(Yueqing, 2013)
10A contig	Unfiltered	Uncultured bacterium clone TG-20	JQ769530	98%	biological soil crust of copper mine tailings wastelands, Tongling, China	(Zhan and Sun, 2012)
13A contig	Unfiltered	Uncultured bacterium clone DSM-R32	GU172204	97.1%	Eucalyptus plantation soil, China	NA
14A contig	Unfiltered	Uncultured bacterium clone Amb_16S_760	EF018382	94.8%	trembling aspen rhizosphere under ambient CO2 conditions,	(Lesaulnier et al., 2008)
1B contig	Filtered	Uncultured bacterium clone 51B1-013	LC081024	96.1%	Inside part of ferromanganese nodule, Japan	(Shiraishi et al., 2016)
1D contig and 4D contig	Filtered	Uncultured Chloroflexi bacterium clone G12L-1	HE614853	97.0%	Arsenic and gold mine, Lower Silesia, Zloty Stok, Poland	(Tomczyk-Zak et al., 2013)
2B contig and 2D contig	Filtered	*Candidatus* Yanofskybacteria bacterium GW2011_GWF2_43_596	KX123608	94.8%	Aquifer sediment, Rifle, CO, USA	(Brown et al., 2015)
11B contig	Filtered	Uncultured *Candidatus* Saccharibacteria bacterium clone BMr13	KT759369	98.6%	Plant roots, China	(Jin et al., 2014)
13B contig	Filtered	Uncultured TM7 bacterium clone QEDN3BD04	CU925854	94.1%	mesophilic anaerobic digester which treats municipal wastewater sludge, Evry, France	(Rivière et al., 2009)
14B contig	Filtered	Uncultured bacterium clone A10	JQ323111	94.7%	industrial circulating cooling water system, China	NA
20B contig	Filtered	Uncultured bacterium clone QQSB107	JF17826	95.6%	Coalbed, Eerduosi Basin, China	(Tang et al., 2012)
21B contig	Filtered	Uncultured bacterium gene clone RH1001	AB510990	88.5%	Soil, Japan	(Tabei and Ueno, 2010)

4.3.2. Shotgun metagenome sequencing

4.3.2.1. Overview of the Conwy River community

It is clear from the cladogram of all taxa detected in both fractions (Fig. 4.2), that the Afon Conwy is a microbiologically diverse system. There was a total of 386,586 and 308,756 sequences in the filtered and unfiltered fractions, respectively. In terms of actual assembled bases 114,345,546 in the unfiltered fraction and 227,308,057 within the filtered fraction. For protein coding genes, there are 623,535 genes in the filtered fraction versus 369,868 genes in the unfiltered fraction, all of which had a product name associated with them. The number of scaffolds in the filtered fraction are 386,586 and for the unfiltered fraction it is 308,756.

The majority of protein encoding genes, 91.9% in the filtered and 60.1% in the unfiltered fractions, did not belong to any COG category (Table 4.3). RNA genes only made up 0.45% in the filtered and 0.34% in the unfiltered fractions (Table 4.3). According to Figure 3, the most prevalent group in both filtered and unfiltered fractions are *Proteobacteria* at 5.40% and 18.6% respectively. Only two groups that had a greater percentage in the filtered vs unfiltered was the dsDNA viruses, no RNA stage, (0.15% in the unfiltered fraction vs 0.95% in the filtered fraction) and the unclassified viruses (0.02% in the unfiltered fraction vs 0.07% in the filtered fraction) (Fig. 4.3).

There were only a few of 16S rRNA gene sequences that were identifiable (Fig. 4.4). Only nine sequences from both filtered and unfiltered fractions were identified via BLAST matches (Tables 4.4 and 4.5) with relatively low percent identities (< 96 %) to known species. Figure 4.4, Table 4.4 and 4.7 show that the sequence Ga0136236 10202521 is an entirely separate clade from the rest of the sequences with its closest relative, uncultured candidate division OD1 bacterium (accession number: JN54016), with a percent identity of 82.4% (Table 4.5).

Table 4.3- General statistics from de novo sequencing analysis of the filtered and unfiltered fraction. Data were coalesced and organised by the IMG/MER database (Markowitz et al., 2012). The breakdown of protein coding and RNA genes are listed here in terms of number of assembled genes and percent of the total of protein coding genes and RNA genes combined. Genes that were affiliated with databases are provided here. NA denotes no information available.

Unfiltered fraction			***Filtered fraction***		
Genes	***Number of genes Assembled***	***Percent of Assembled***	***Genes***	***Number of genes Assembled***	***Percent of Assembled***
RNA genes	1,248	0.34%	*RNA genes*	4,698	0.75%
rRNA genes	183	0.05%	rRNA genes	29	0.00%
5S rRNA	19	0.01%	5S rRNA	4	0.00%
16S rRNA	53	0.01%	16S rRNA	9	0.00%
18S rRNA	6	0.00%	18S rRNA	NA	NA
23S rRNA	97	0.03%	23S rRNA	16	0.00%
28S rRNA	8	0.00%	28S rRNA	NA	NA
tRNA genes	1,065	0.29%	tRNA genes	4,669	0.75%
Protein coding genes	368,620	99.66%	*Protein coding genes*	618,837	99.25%
with Product Name	369,868	100.00%	with Product Name	623,535	100.00%
with COG	147,500	39.88%	with COG	50,353	8.08%
with Pfam	126,879	34.30%	with Pfam	59,498	9.54%
with KO	120,025	32.45%	with KO	26,848	4.31%
with Enzyme	75,479	20.41%	with Enzyme	17,510	2.81%
with MetaCyc	48,037	12.99%	with MetaCyc	8,391	1.35%
with KEGG	76,064	20.57%	with KEGG	15,786	2.53%

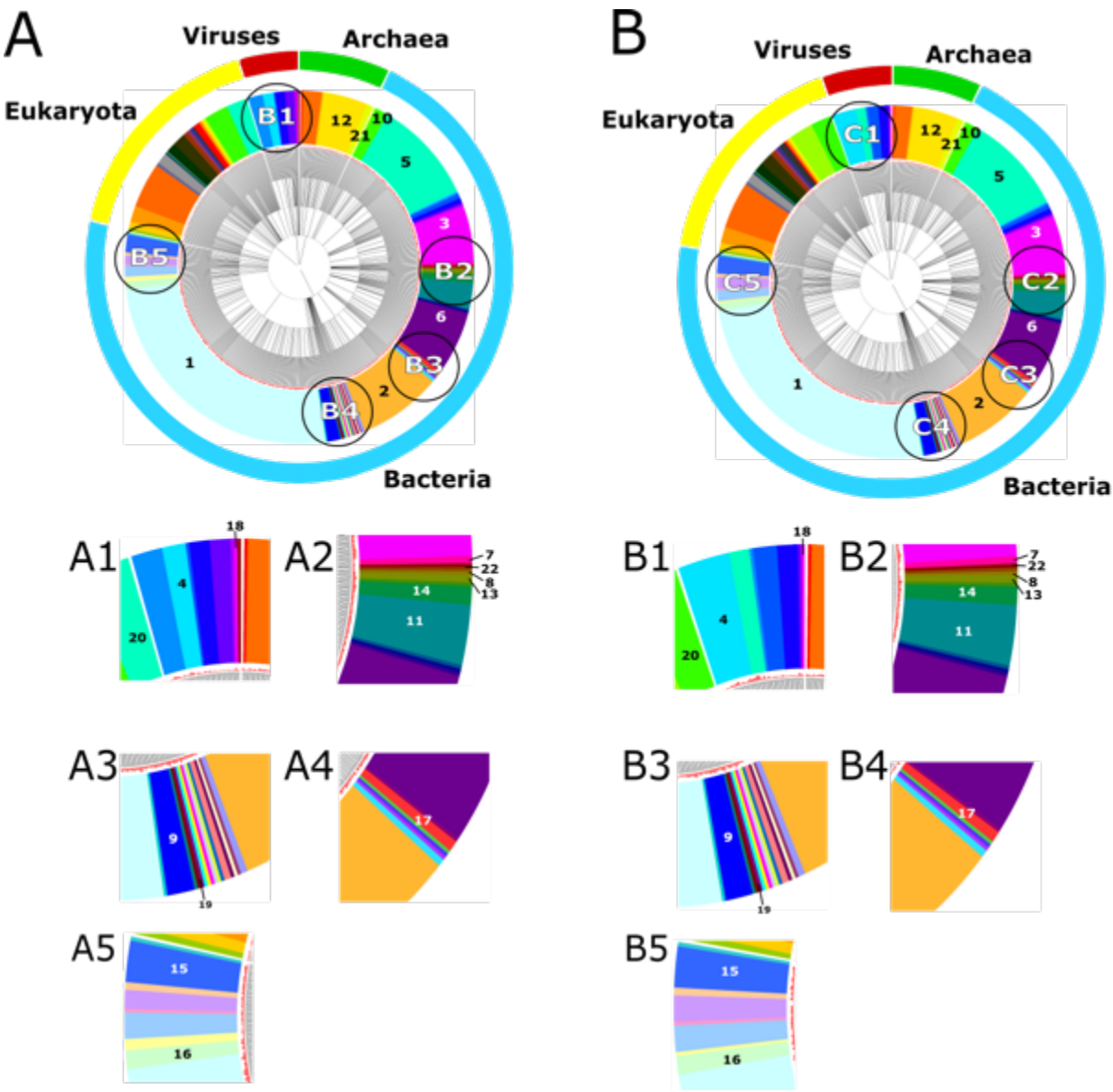


Figure 4.2- Cladogram radial trees of genetic data obtained from de novo sequencing of filtered and unfiltered samples. Radial trees with BLAST matches over 30% were generated for both the (A) unfiltered fraction and (B) filtered fraction (generated October 7th 2019). The outer ring denotes domains: Eukaryota (yellow), Viruses (red), Archaea (green), and Bacteria (blue). The inner ring colours represents the phyla and taxonomic rank at the terminus end of the branches is family. The following numbers correspond to specific phyla: 1-*Proteobacteria*, 2-*Firmicutes*,3-*Bacteriodetes*, 4-dsDNA viruses (no RNA stage), 5-*Actinobacteria*,6-*Cyanobacteria*,7-*Candidatus* Microgenomates,8-*Candidatus* Parcubacteria, 9-*Planctomycetes*, 10-*Acidobacteria*, 11-*Chloroflexi,* 12-*Euryarchaeota,*13-*Candidatus* Saccharibacteria, 14-*Chlamydiae*, 15-*Verrucomicrobia*, 16-*Spirochaetes*, 17-*Deinococcus-Thermus*, 18-Unclassified virus, 19-*Nitrospirae*, 20-Unclassified Eukaryota, 21-*Thaumarchaeota*, and 22-*Candidatus* Omnitrophica. Subsequent subcategories highlight sections of the unfiltered cladogram (A1-A5) and the filtered cladogram (B1-B5).

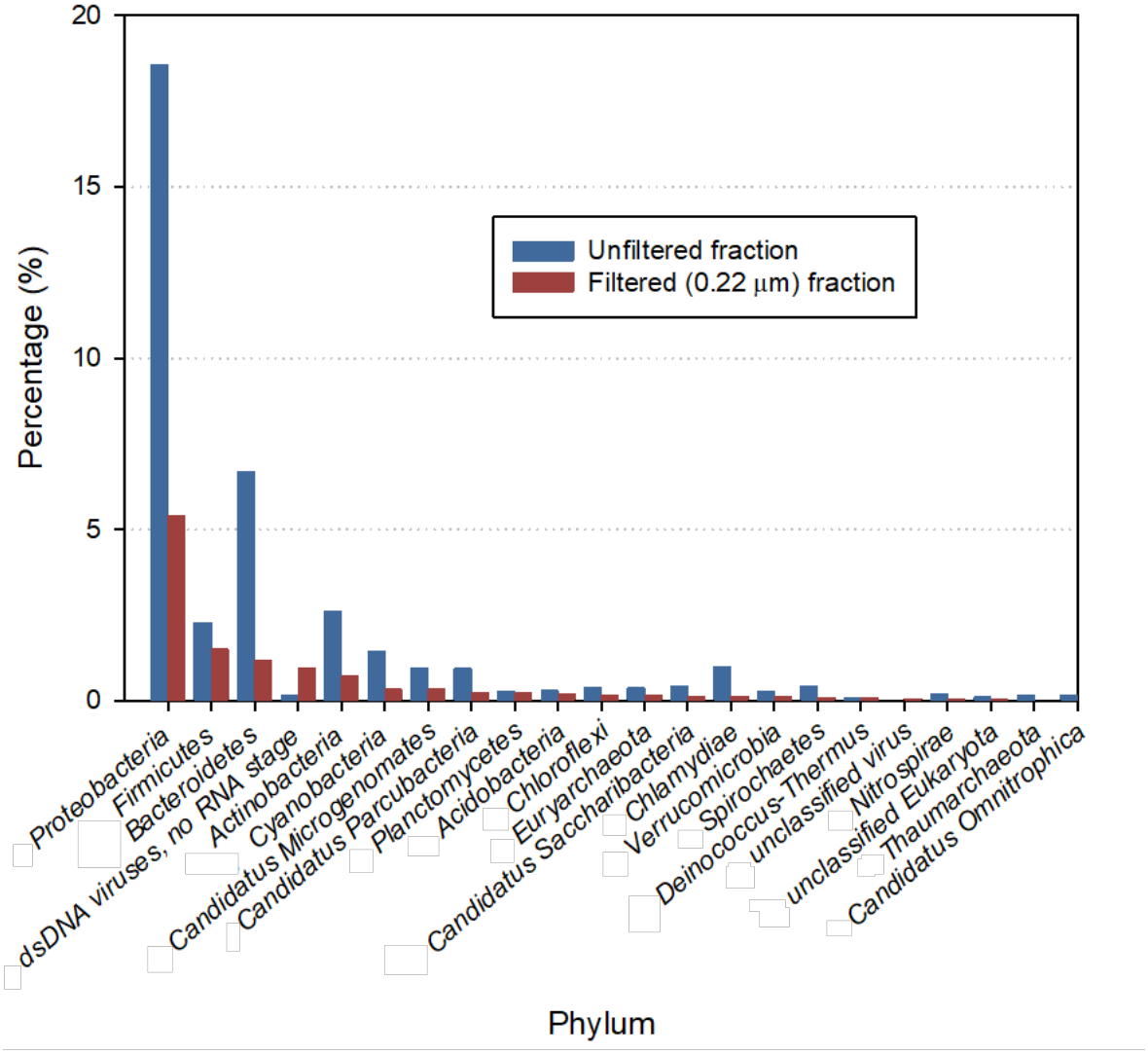


Figure 4.3- Distribution of the most common phyla obtained from de novo sequencing. The most common phyla are listed here in both filtered (red) and unfiltered (blue) fractions.

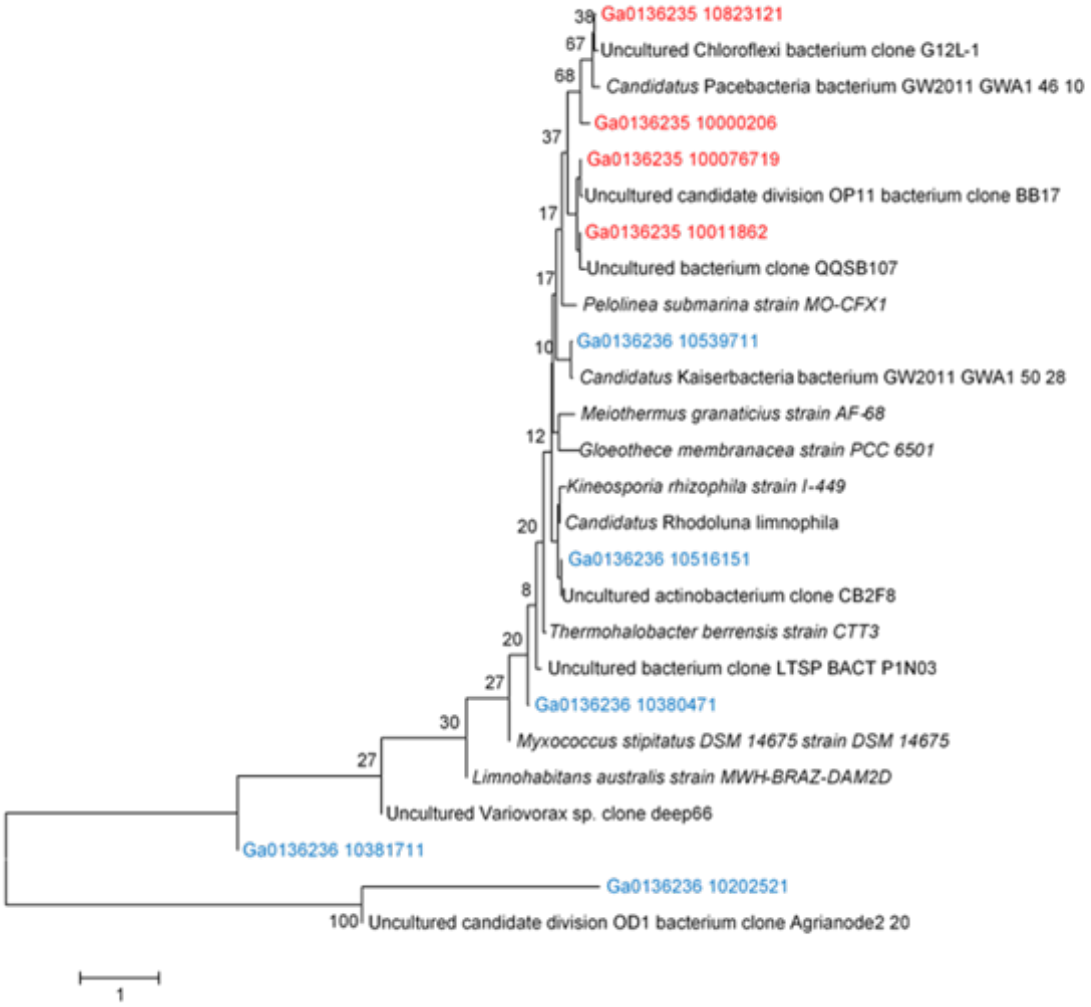


Figure 4.4- Unrooted neighbour joining tree of 16S rRNA sequences obtained from shotgun sequencing. Sequences from the unfiltered sample are in blue and those from the filtered sample (0.22 µm sterivex filter) are in red. Bar represents 100 substitutions per 100 nucleotides (complete substitution of all sites).

Table 4.4- Summary of taxa retrieved from 16S rRNA BLAST search. BLAST matches of 16S rRNA gene sequences obtained from shotgun sequencing data. Here we selected the highest percent matches to the clones. The sequence name, residing fraction, estimated copies, length of the gene (bp), BLAST match species, phylum, percent identity, dimensions and reference are provided here. Information and references were taken from the NCBI database. NA indicates information is not available.

Sequence Name	*Residing fraction*	*Estimated number of copies*	*Length of gene (bp)*	*16S rRNA BLAST match species*	*Phylum*	*Percent Identity*	*Dimension 1 (diameter) (μm)*	*Dimension 2 (length) (μm)*	*Reference*
Ga0136235 10823121	Filtered	2	642	*Pelolinea submarina strain MU-CFX1*	*Chloroflexi*	77%	0.130-0.150	10	(Imachi et al., 2014)
Ga0136235 10011862	Filtered	3	1479	*Thermohalobacter berrensis strain CTT3*	*Firmicutes*	75%	0.5	3.0-8.0	(Cayol et al., 2000)
Ga0136236 10202521	Unfiltered	2	757	*Gloeothece membranacea strain PCC 6501*	*Cyanobacteria*	80%	NA	NA	(Ohki et al., 2008)
Ga0136236 10539711	Unfiltered	1	360	*Meiothermus granaticius strain AF-68*	*Deinococcus-Thermus*	76%	0.8-1.0	1.5-10	(Albuquerque et al., 2010)
Ga0136235 10000206	Filtered	18	1516	*Kineosporia rhizophila strain I-449*	*Actinobacteria*	76%	Spores 1-2	NA	(Kudo et al., 1998)
Ga0136235 100076719	Filtered	4	1635	*Candidatus Rhoduluna limnophila*	*Actinobacteria*	78%	NA	NA	(Hahn, 2009)
Ga0136236 10380471	Unfiltered	3	448	*Myxococcus stipitatus DMS 14675*	*Proteobacteria*	83%	NA	NA	(Huntley et al., 2013)
Ga0136236 10381711	Unfiltered	2	495	*Limnohabitans australis strain MWH-BRAZ-DMA2D*	*Proteobacteria*	96%	0.4-0.5	1.0-1.7	(Hahn et al., 2010)

Table 4.5- Summary of taxa retrieved from nr/nt BLAST search. BLAST matches from the sequences obtained via IMG/MER analysis. Here we selected the highest percent matches to the clones. Listed are sequence name, residing fraction, estimated copies, length of the gene (bp), associated identities, environmental location, NCBI accession number, and references. Information and references were obtained from the NCBI database.

Sequence Name	*Residing fraction*	*Estimated number of copies*	*Length of gene (bp)*	*Nr/nt BLAST match name*	*Accession number (NCBI)*	*Percent Identity*	*Isolation location*	*References*
Ga0136236 10516151	Unfiltered	3	369	Uncultured actinobacterium clone CB2F8	HQ532245	99.7%	Integrated lake epilimnion from Crystal Bog Lake, WI, USA	(Newton et al., 2011)
Ga0136236 10380471	Unfiltered	3	448	Uncultured bacterium clone LTSP_BACT_P1N03	FJ550938	97.3%	Forest soil from the long-term soil productivity (LTSP) site Skulow Lake, Canada	(Hartmann et al., 2009)
Ga0136236 10381711	Unfiltered	1	360	Uncultured *Variovorax sp*. clone deep66	JQ684483	98.8%	Ground water, Ireland	(Akinbami, 2013)
Ga0136236 10202521	Unfiltered	2	757	Uncultured candidate division OD1 bacterium clone Agrianode2_20 16S	JN540162	82.4%	Soil from an agricultural site, Harvard Forest, MA, USA	(Dunaj et al., 2012)
Ga0136236 10539711	Unfiltered	1	360	*Candidatus* Kaiserbacteria bacterium GW2011_GWA1_50_28	KX123506	90.7%	Aquifer sediment, Rifle, CO, USA	(Brown et al., 2015)
Ga0136235 10000206	Filtered	18	1516	*Candidatus* Pacebacteria bacterium GW2011_GWA1_46_10	KX123533	91.3%	Aquifer sediment, Rifle, CO, USA	(Brown et al., 2015)
Ga0136235 10823121	Filtered	2	642	Uncultured *Chloroflexi* bacterium clone G12L-1	HE614853	97.4%	Arsenic and gold mine, Zloty Stok, Lower Silesia, Poland	(Tomczyk-Zak et al., 2013)
Ga0136235 100076719	Filtered	4	1635	Uncultured candidate division OP11	AM167966	94.2%	Mineral Spring, Wettinquelle, Bad Brambach,Germany	(Wagner et al., 2007)
Ga0136235 10011862	Filtered	3	1479	Uncultured bacterium clone QQSB107	JF417826	93.7%	Coalbed, Eerduosi Basin, China	(Tang et al., 2012)

4.3.2.2. Clusters of Orthologous Groups of proteins (COGs)

The number of protein encoding genes designated to a COG in the filtered fraction was 50,353 (8.08% of all assembled protein encoding genes) whereas the unfiltered fraction was 147,500 (39.9% of assembled protein encoding genes) (Table 4.3). There is a higher diversity of COGs within the unfiltered fraction than the filtered fraction (Figs. 4.5 and 4.6). The majority of the COGs in the filtered fraction fell into the functional category of replication, recombination, and repair (20.5%) followed closely by cell wall/membrane/envelope biogenesis (12.9%) (Fig 4.5). The majority of COGs from the unfiltered fraction fell into the functional categories of: translation, ribosomal structure and biogenesis (8.96%), cell wall/membrane/envelope biogenesis (8.79%), amino acid transport and metabolism (8.49%), and finally general function (8.18%) (Fig 4.5). The overrepresentation seen in the filtered fraction was narrowed down to the pathway of basal replication machinery (Figs. 4.7- 4.8).

Hence, we further examined the genes associated with basal replication machinery and DNA polymerase III subunits (both of which are part of the replication, recombination, and repair functional category) and found that replicative DNA helicase and DNA polymerase I 3'-5' exonuclease and polymerase domains were significantly higher, in terms of gene count in the filtered fraction relative to the unfiltered fraction (Fig. 4.9).

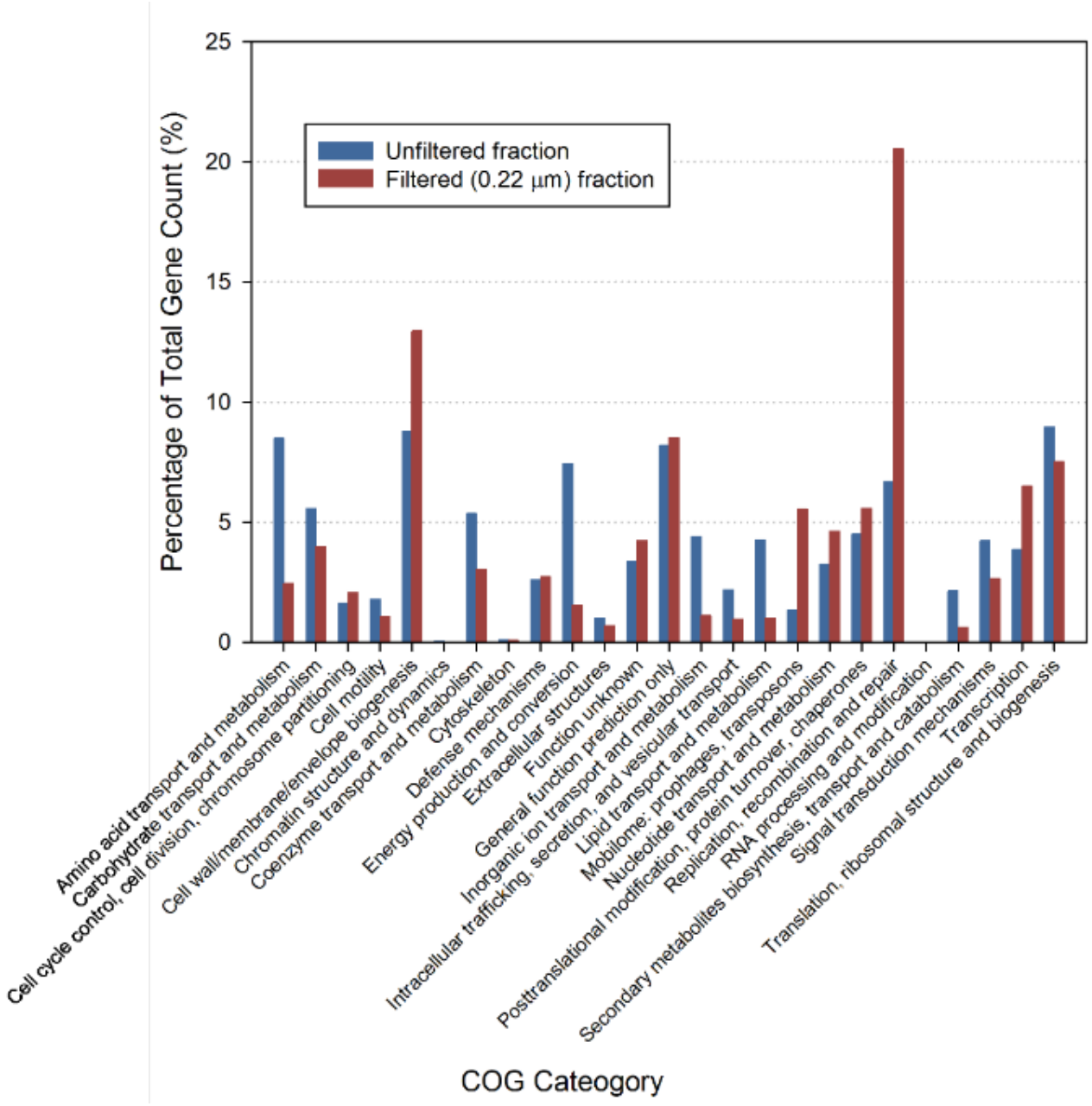


Figure 4.5- Distribution of COG-coding genes as per functional category. The percentage of total gene count that falls under a COG category is compared between the filtered (red) and unfiltered (blue) fraction.

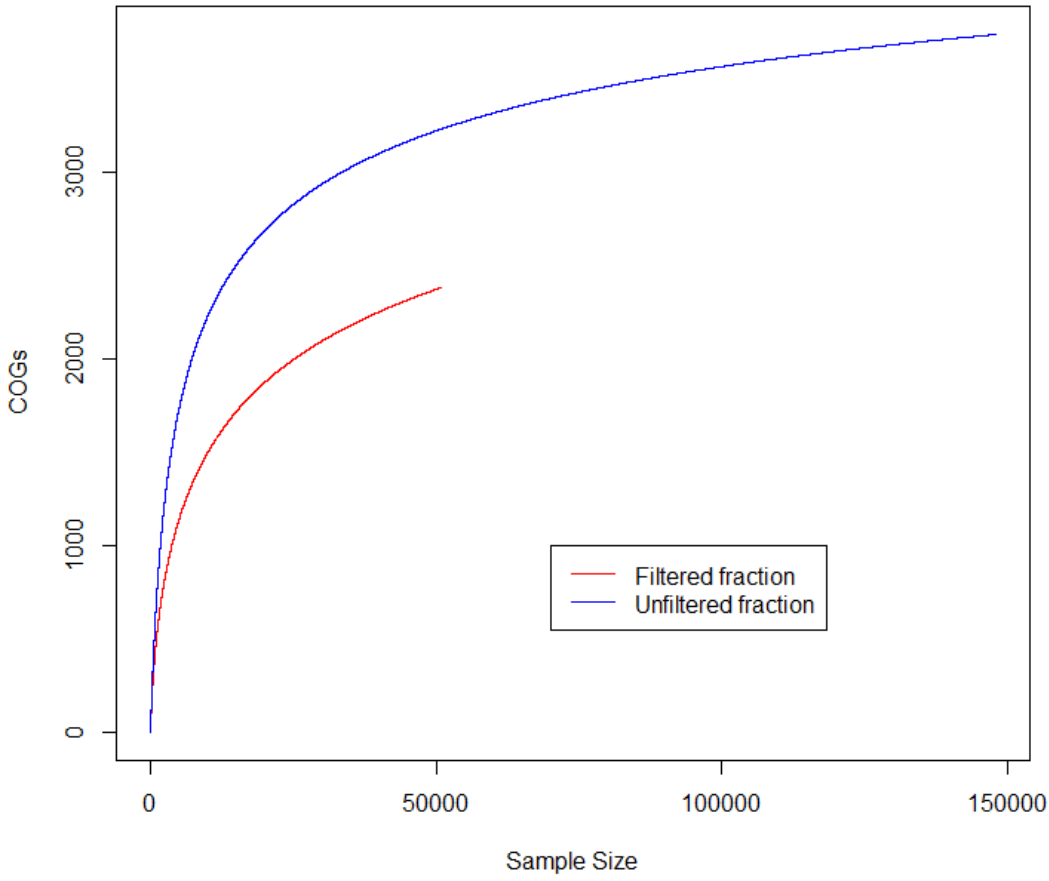
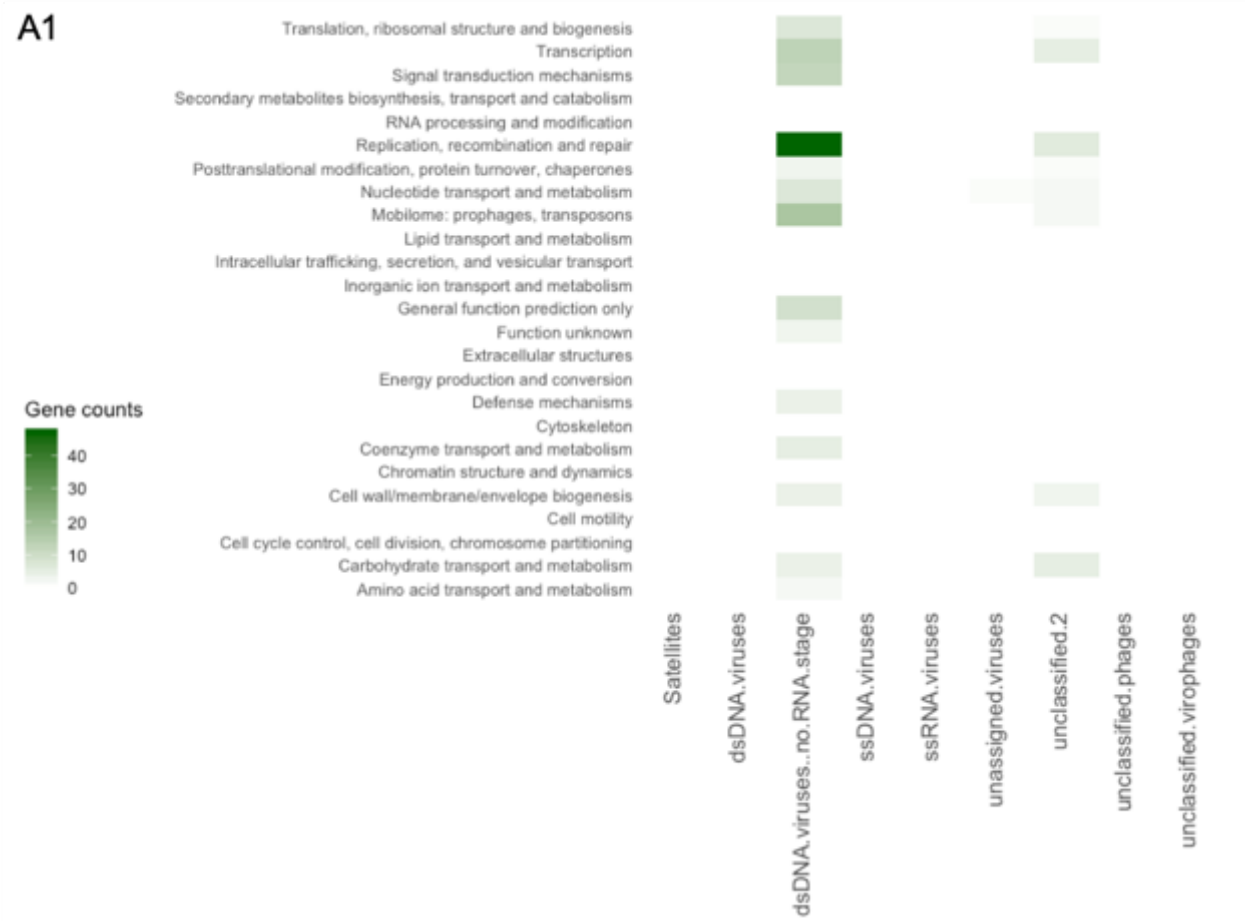
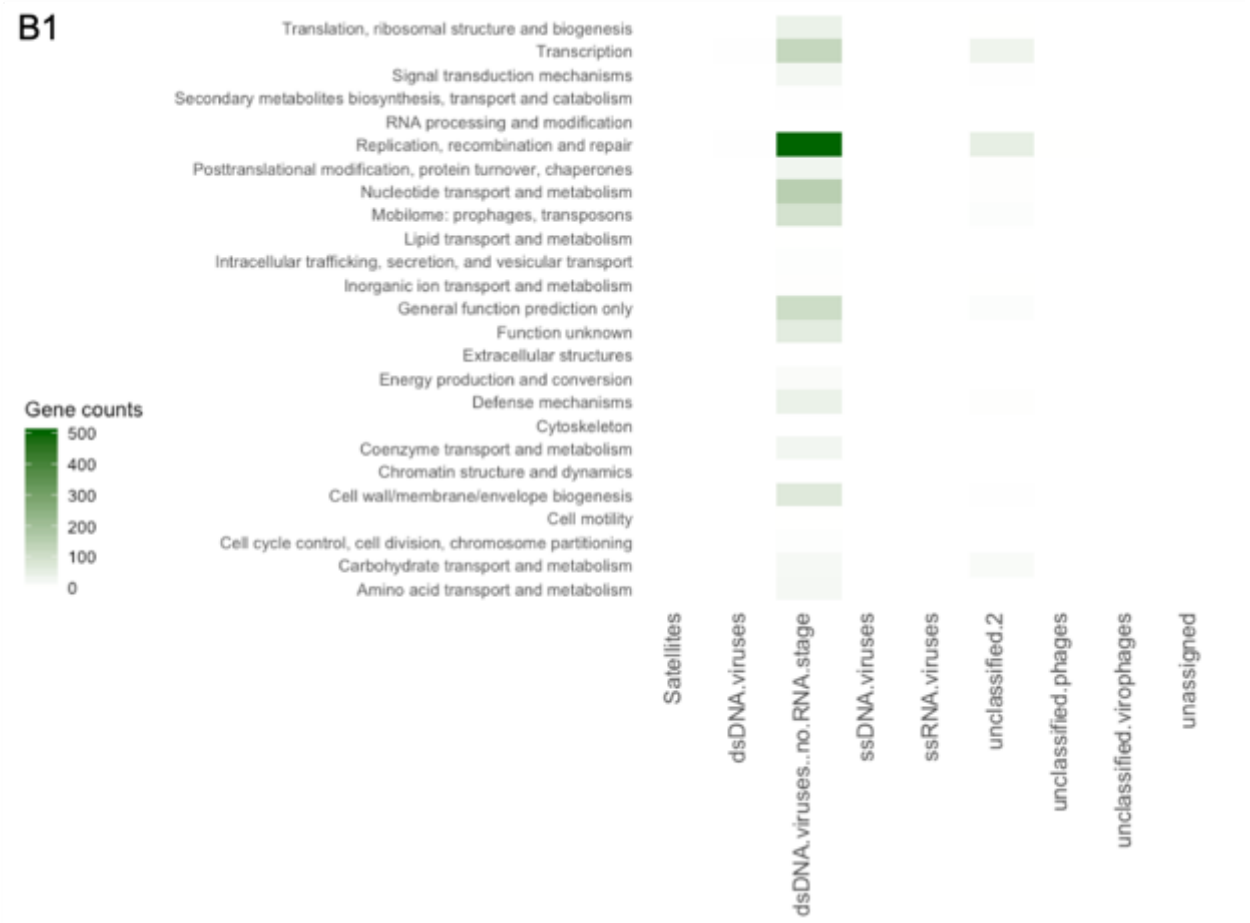
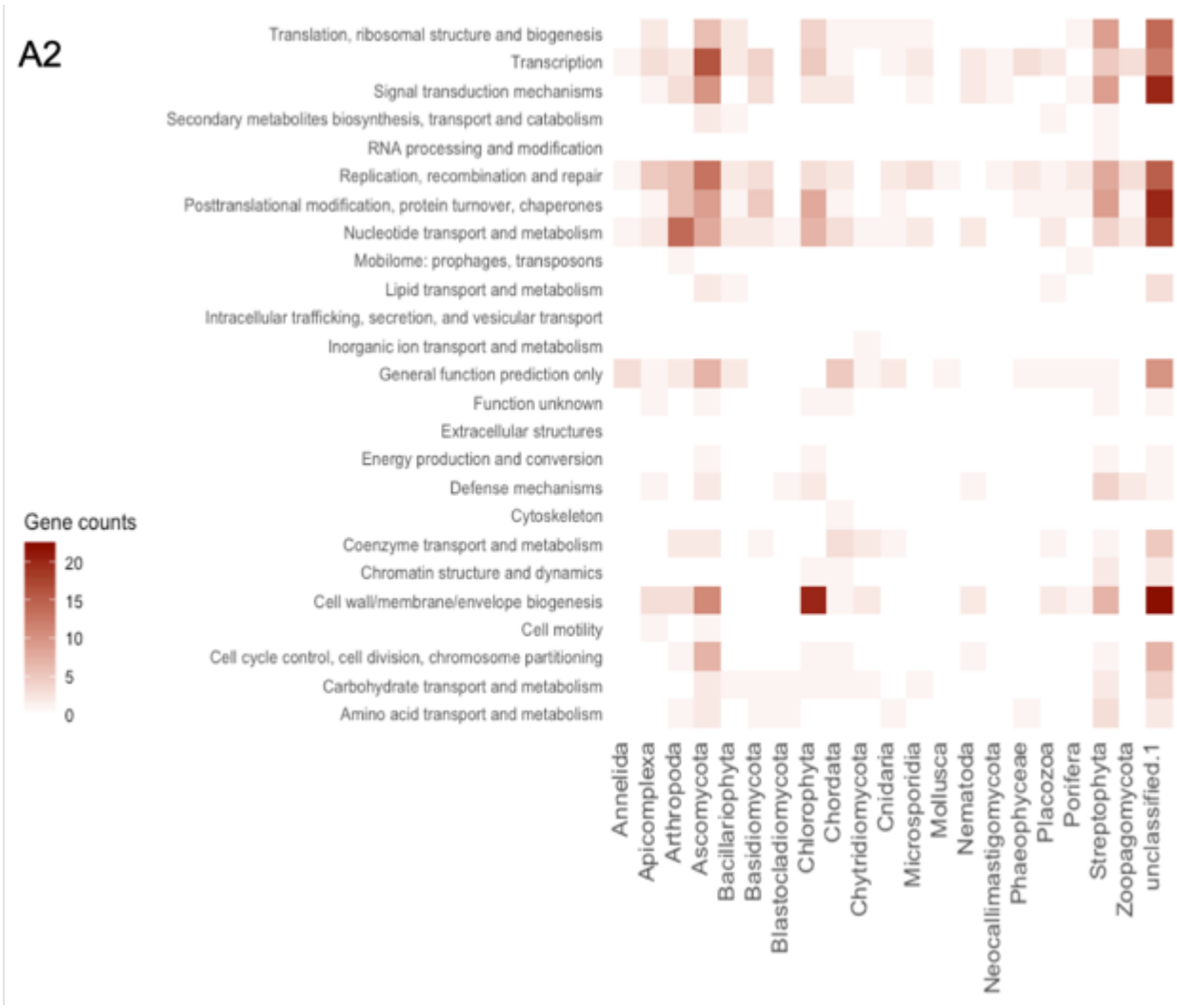
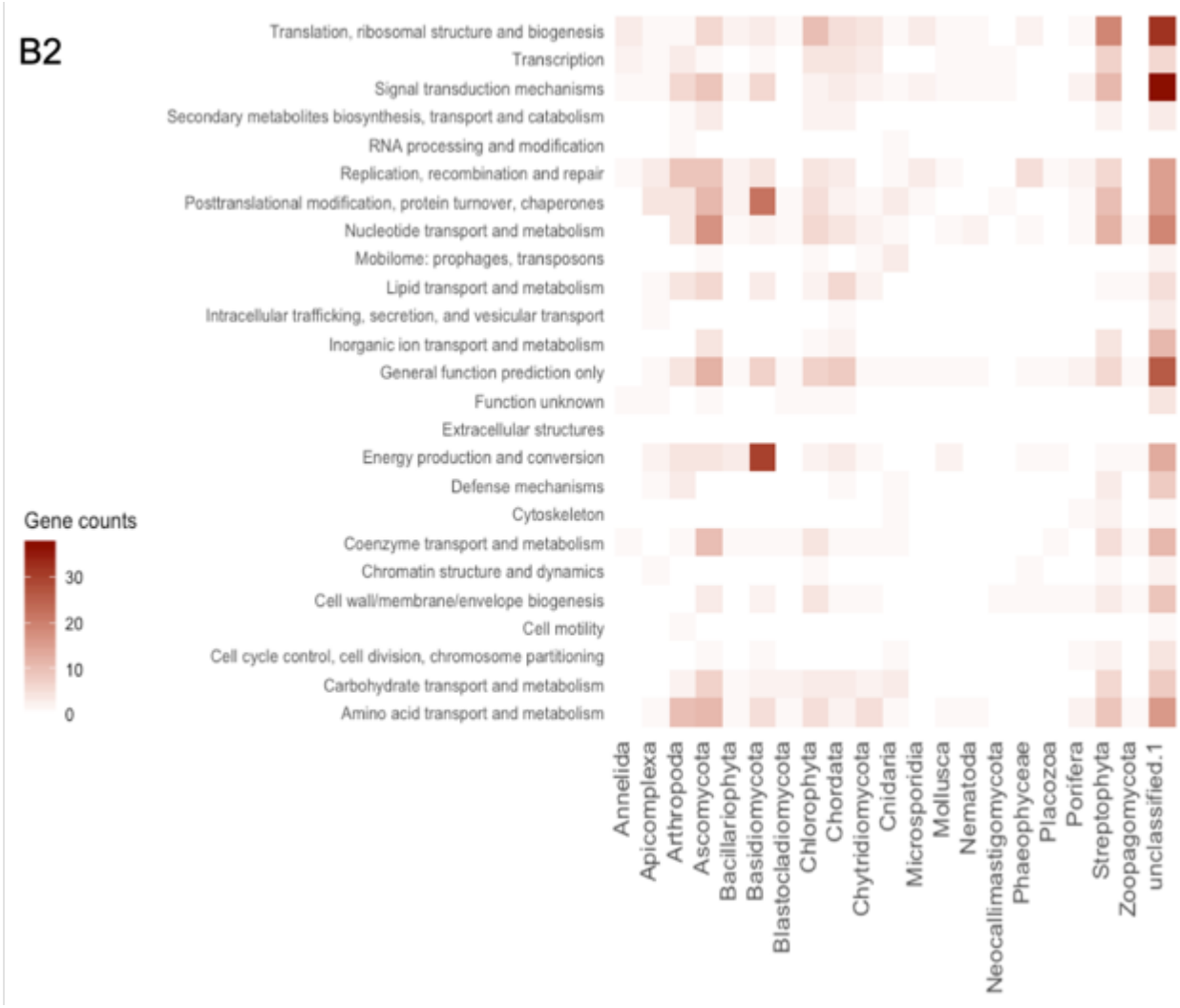


Figure 4.6- Rarefaction curve of the filtered and unfiltered fraction based on COGs. Measures richness of the COGs present in both the filtered (red) and unfiltered (blue) fraction.

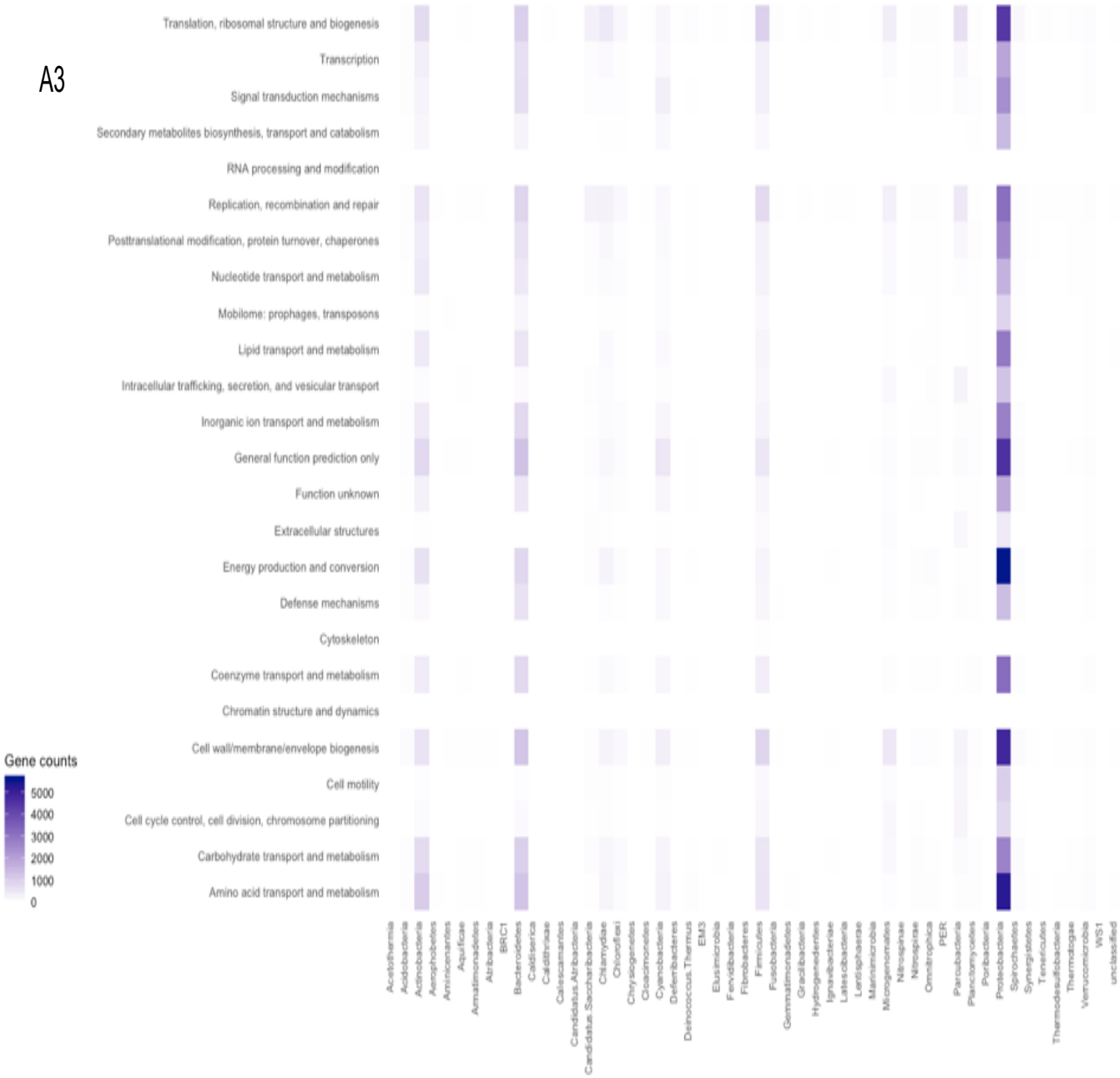

A1
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
40
30
20
10
0
Satellites
dsDNA.viruses
dsDNA.viruses..no.RNA.stage
ssDNA.viruses
ssRNA.viruses
unassigned.viruses
unclassified.2
unclassified.phages
unclassified.virophages


B1
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
500
400
300
200
100
0
Satellites
dsDNA.viruses
dsDNA.viruses..no.RNA.stage
ssDNA.viruses
ssRNA.viruses
unclassified.2
unclassified.phages
unclassified.virophages
unassigned

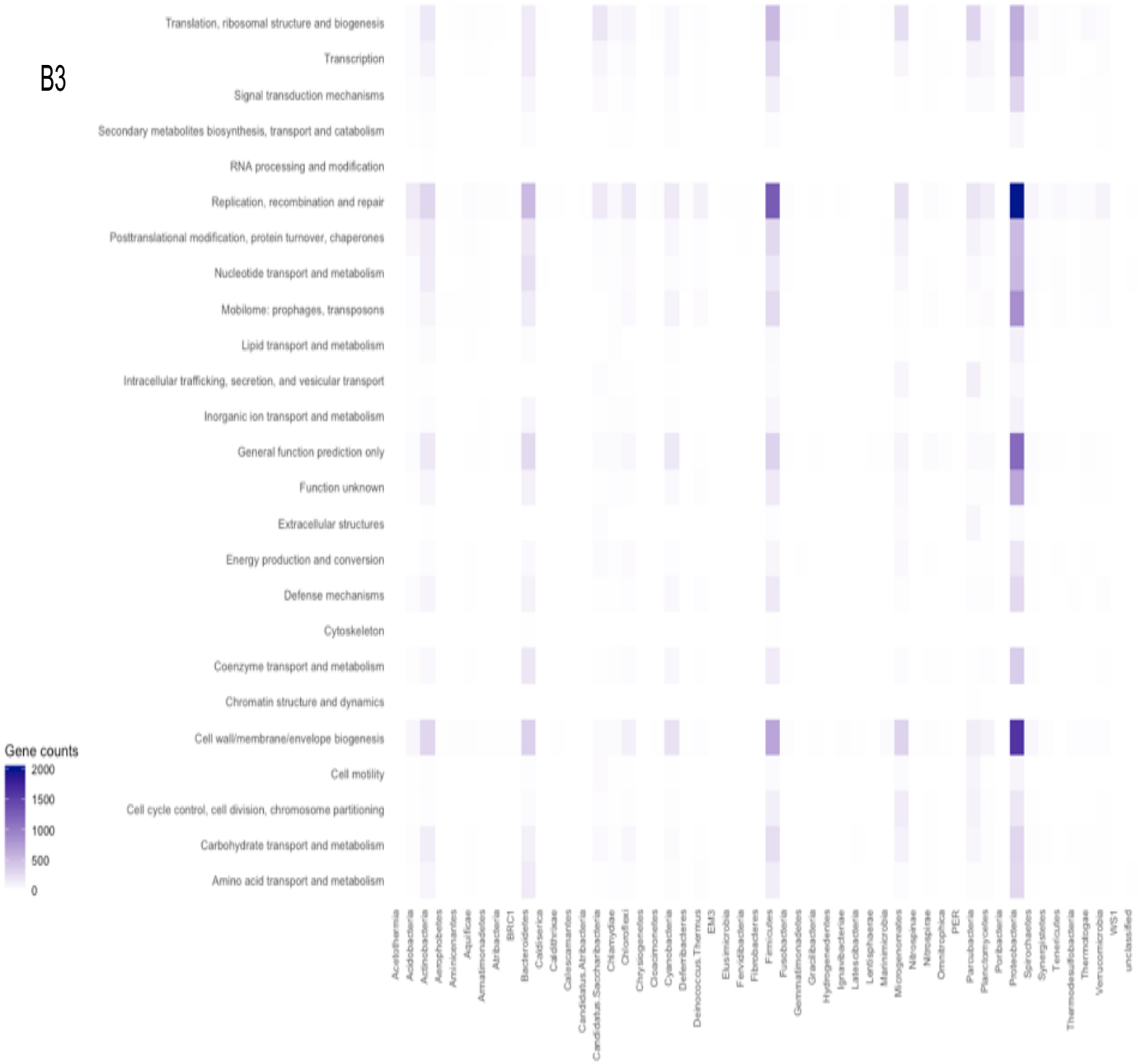

A2
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
20
15
10
5
0
Annelida
Apicomplexa
Arthropoda
Ascomycota
Bacillariophyta
Basidiomycota
Blastocladiomycota
Chlorophyta
Chordata
Chytridiomycota
Cnidaria
Microsporidia
Mollusca
Nematoda
Neocallimastigomycota
Phaeophyceae
Placozoa
Porifera
Streptophyta
Zoopagomycota
unclassified.1


B2
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
30
20
10
0
Annelida
Apicomplexa
Arthropoda
Ascomycota
Bacillariophyta
Basidiomycota
Blastocladiomycota
Chlorophyta
Chordata
Chytridiomycota
Cnidaria
Microsporidia
Mollusca
Nematoda
Neocallimastigomycota
Phaeophyceae
Placozoa
Porifera
Streptophyta
Zoopagomycota
unclassified.1

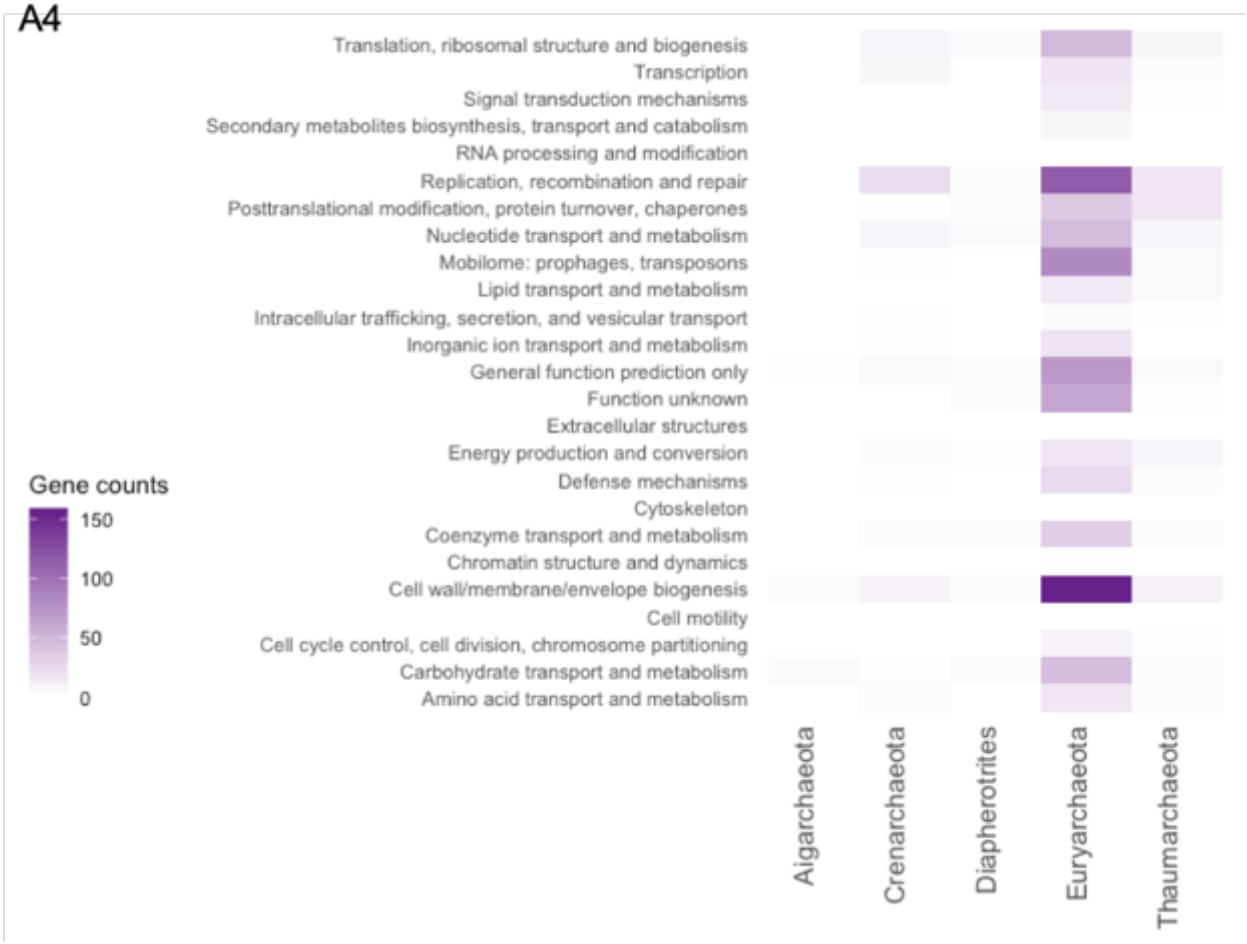
A3


Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
5000
4000
3000
2000
1000
0
Acetothermia
Acidobacteria
Actinobacteria
Aerophobetes
Aminicenantes
Aquificae
Armatimonadetes
Atribacteria
BRC1
Bacteroidetes
Calditrichae
Calescamantes
Candidatus Atribacteria
Candidatus Saccharibacteria
Chlamydiae
Chloroflexi
Chrysiogenetes
Cloacimonetes
Cyanobacteria
Deferribacteres
Deinococcus Thermus
EM3
Elusimicrobia
Fervidibacteria
Fibrobacteres
Firmicutes
Fusobacteria
Gemmatimonadetes
Gracilibacteria
Hydrogenedentes
Ignavibacteriae
Latescibacteria
Lentisphaerae
Marinimicrobia
Microgenomates
Nitrospinae
Nitrospirae
Omnitrophica
PER
Parcubacteria
Planctomycetes
Poribacteria
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Thermodesulfobacteria
Thermotogae
Verrucomicrobia
WS1
unclassified

B3


Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
2000
1500
1000
500
0
Acetothermia
Acidobacteria
Actinobacteria
Aerophobetes
Aminicenantes
Aquificae
Armatimonadetes
Atribacteria
BRC1
Bacteroidetes
Caldiserica
Calditrichaeae
Calescamantes
Candidatus_Atribacteria
Candidatus_Saccharibacteria
Chlamydiae
Chloroflexi
Chrysiogenetes
Cloacimonetes
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
EM3
Elusimicrobia
Fervidibacteria
Fibrobacteres
Firmicutes
Fusobacteria
Gemmatimonadetes
Gracilibacteria
Hydrogenedentes
Ignavibacteriae
Latescibacteria
Lentisphaerae
Marinimicrobia
Microgenomates
Nitrospinae
Nitrospirae
Omnitrophica
PER
Parcubacteria
Planctomycetes
Poribacteria
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Thermodesulfobacteria
Thermotogae
Verrucomicrobia
WS1
unclassified

A4


Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
RNA processing and modification
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Mobilome: prophages, transposons
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
General function prediction only
Function unknown
Extracellular structures
Energy production and conversion
Defense mechanisms
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Gene counts
150
100
50
0
Aigarchaeota
Crenarchaeota
Diapherotrites
Euryarchaeota
Thaumarchaeota

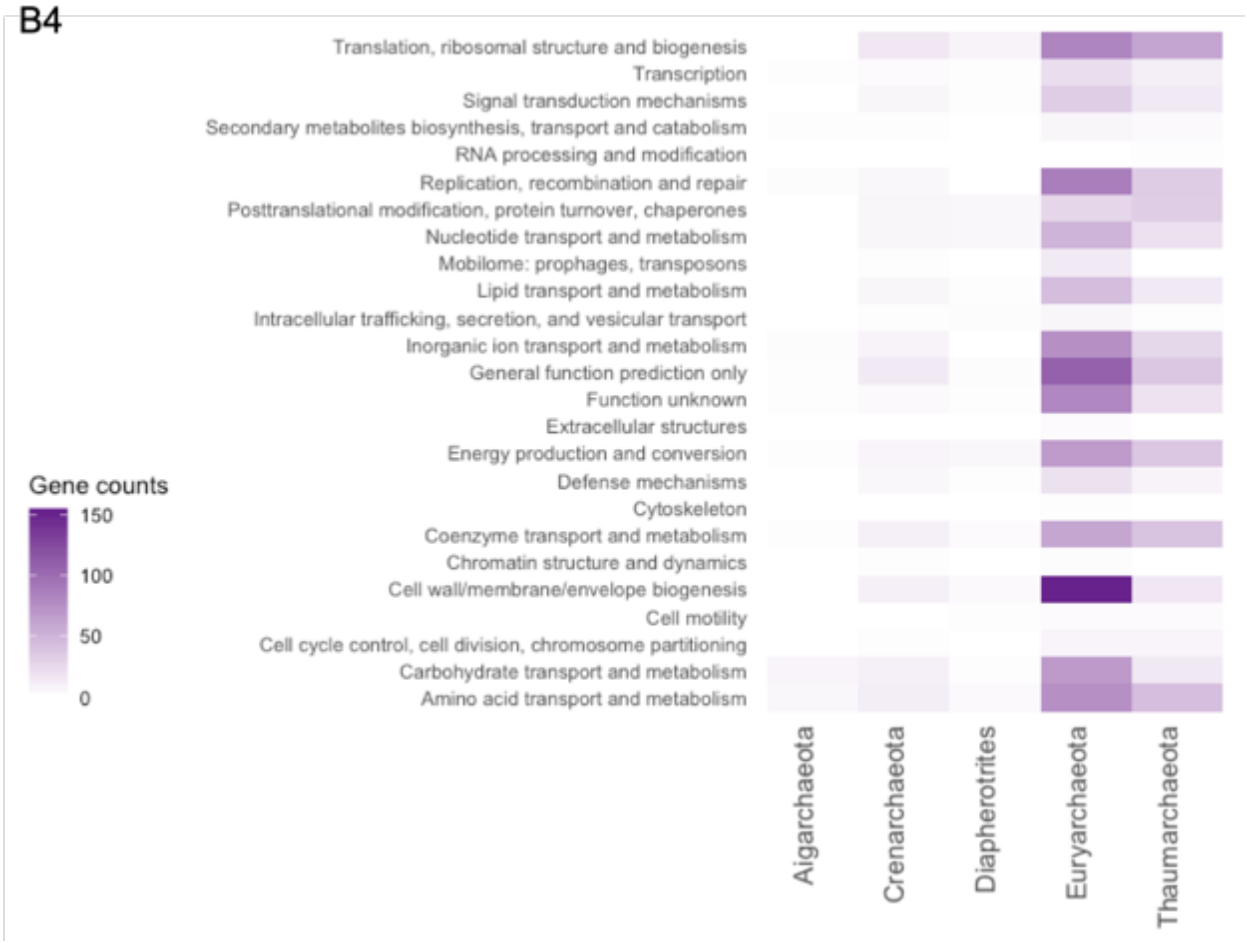
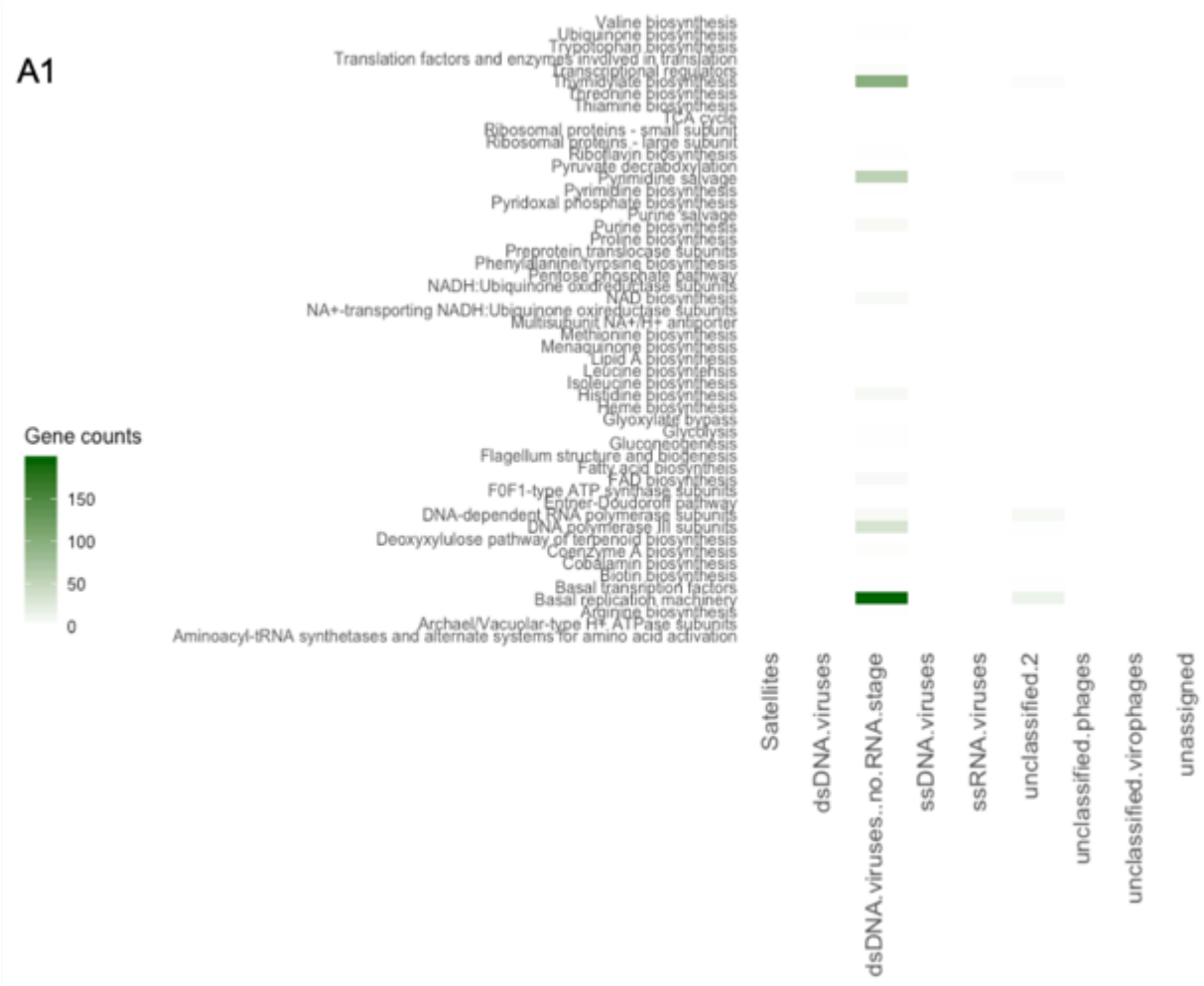
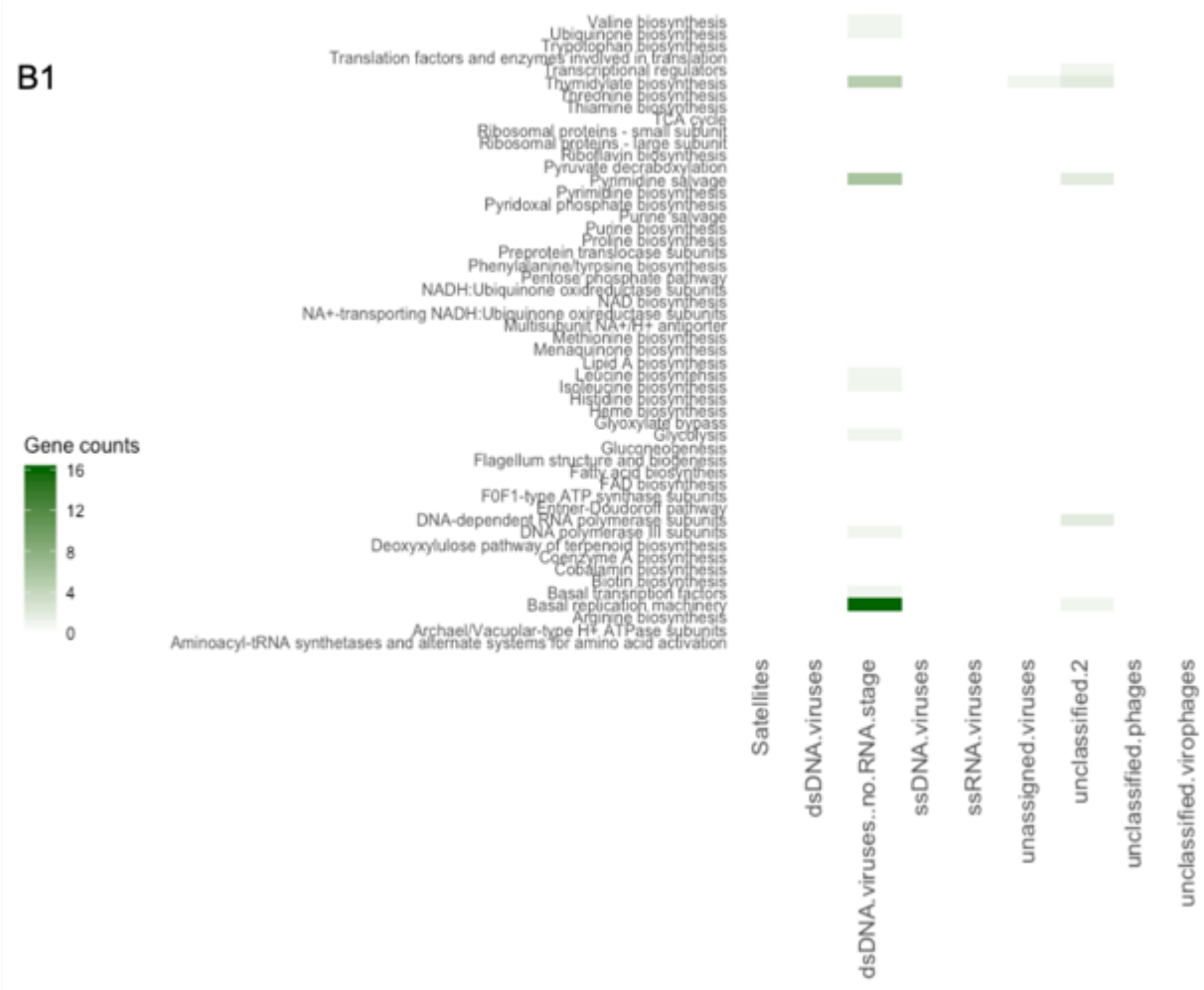


Figure 4.7- Functional categories of COG-coding genes in (A) filtered fraction and (B) unfiltered fraction of Conwy River water samples. Colour intensity is correlated to gene counts plotted microbial and viral phyla and the graphs are divided into the domains. (1) Viruses are green, (2) Eukaryotes are red, (3) Bacteria are blue, and (4) Archaea are purple.

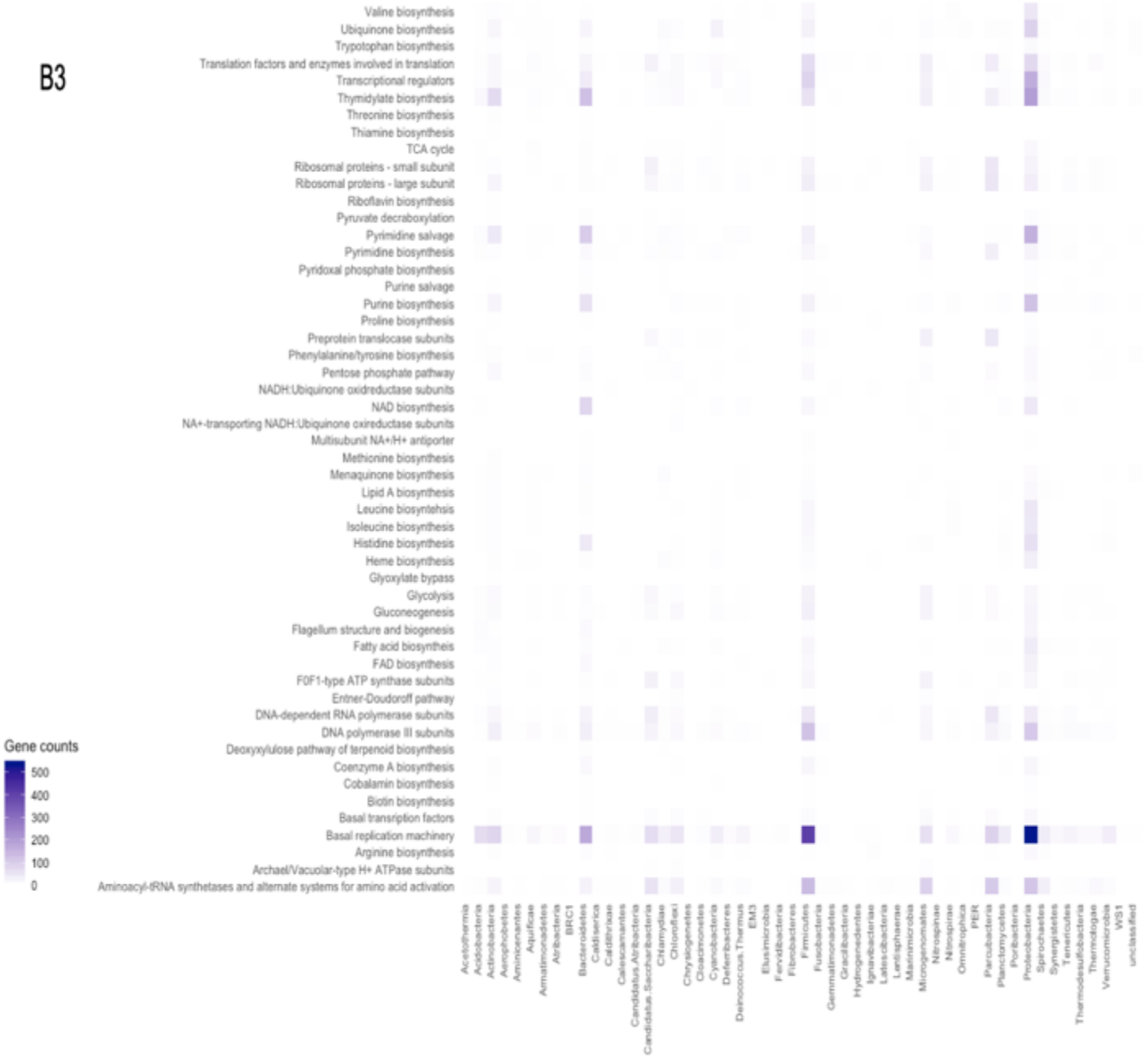

A1
Gene counts
150
100
50
0
Valine biosynthesis
Ubiquinone biosynthesis
Tryptophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decarboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrpsine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidoreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosynthesis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosynthesis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transcription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Satellites
dsDNA.viruses
dsDNA.viruses..no.RNA.stage
ssDNA.viruses
ssRNA.viruses
unclassified.2
unclassified.phages
unclassified.virophages
unassigned

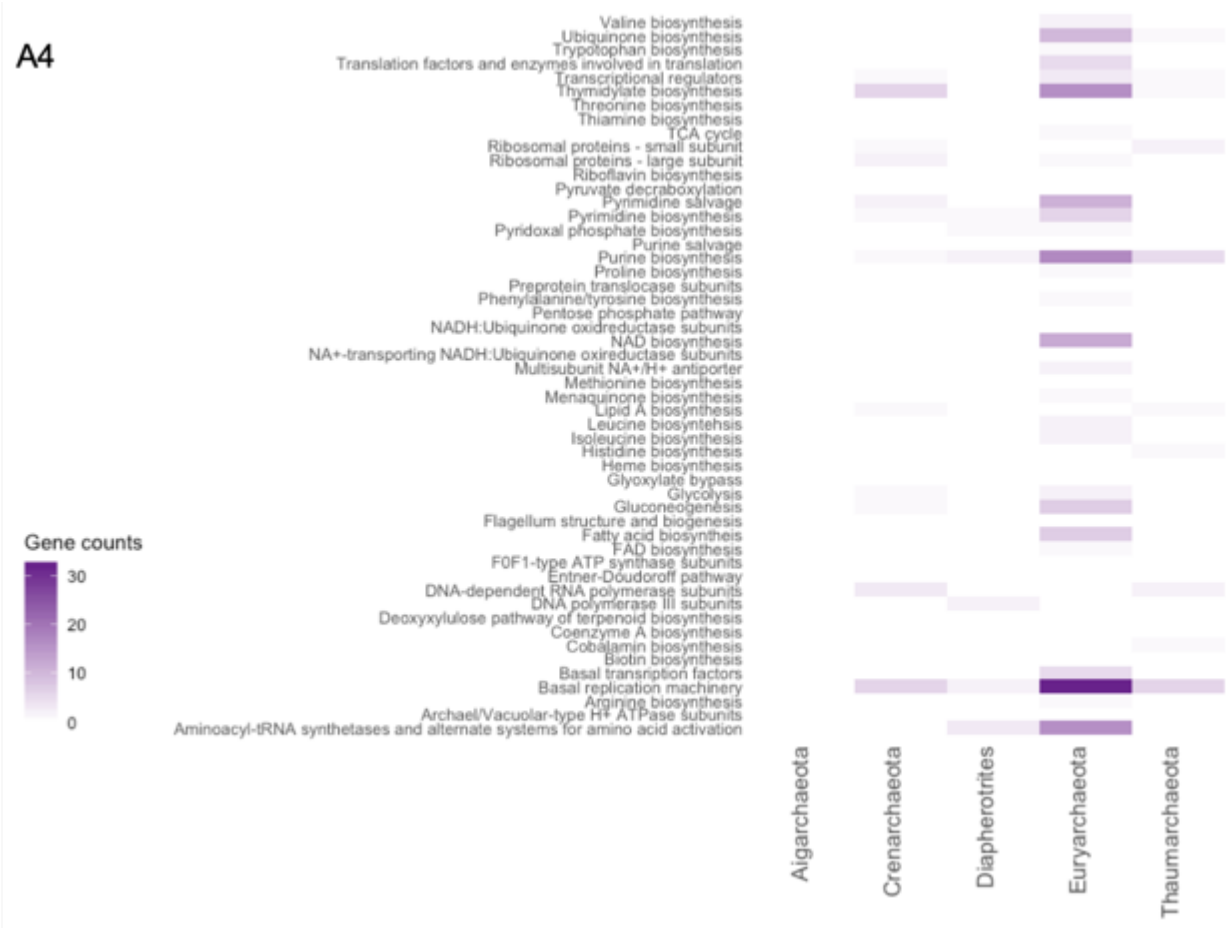

B1
Gene counts
16
12
8
4
0
Valine biosynthesis
Ubiquinone biosynthesis
Tryptophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decarboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrpsine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosynthesis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosynthesis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transcription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Satellites
dsDNA.viruses
dsDNA.viruses..no.RNA.stage
ssDNA.viruses
ssRNA.viruses
unassigned.viruses
unclassified.2
unclassified.phages
unclassified.virophages


A2
Valine biosynthesis
Ubiquinone biosynthesis
Tryptophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decraboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrosine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosynthesis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosynthesis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Gene counts
10.0
7.5
5.0
2.5
0.0
Annelida
Apicomplexa
Arthropoda
Ascomycota
Bacillariophyta
Basidiomycota
Blastocladiomycota
Chlorophyta
Chordata
Chytridiomycota
Cnidaria
Microsporidia
Mollusca
Nematoda
Neocallimastigomycota
Phaeophyceae
Placozoa
Porifera
Streptophyta
Zoopagomycota
unclassified.1


B2
Valine biosynthesis
Ubiquinone biosynthesis
Tryptophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decraboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrosine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosynthesis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosynthesis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Gene counts
25
20
15
10
5
0
Annelida
Apicomplexa
Arthropoda
Ascomycota
Bacillariophyta
Basidiomycota
Blastocladiomycota
Chlorophyta
Chordata
Chytridiomycota
Cnidaria
Microsporidia
Mollusca
Nematoda
Neocallimastigomycota
Phaeophyceae
Placozoa
Porifera
Streptophyta
Zoopagomycota
unclassified.1


A3
Valine biosynthesis
Ubiquinone biosynthesis
Trypotophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decraboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrosine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosyntehsis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosyntheis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Gene counts
1250
1000
750
500
250
0
Acetothermia
Acidobacteria
Actinobacteria
Aerophobetes
Aminicenantes
Aquificae
Armatimonadetes
Atribacteria
BRC1
Bacteroidetes
Caldiserica
Calditrichae
Calescamantes
Candidatus_Atribacteria
Candidatus_Saccharibacteria
Chlamydiae
Chloroflexi
Chrysiogenetes
Cloacimonetes
Cyanobacteria
Deferribacteres
Deinococcus_Thermus
EM3
Elusimicrobia
Fervidibacteria
Fibrobacteres
Firmicutes
Fusobacteria
Gemmatimonadetes
Gracilibacteria
Hydrogenedentes
Ignavibacteriae
Latescibacteria
Lentisphaerae
Marinimicrobia
Microgenomates
Nitrospinae
Nitrospirae
Omnitrophica
PER
Parcubacteria
Planctomycetes
Poribacteria
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Thermodesulfobacteria
Thermotogae
Verrucomicrobia
WS1
unclassified


B3
Valine biosynthesis
Ubiquinone biosynthesis
Trypotophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decraboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrosine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosyntehsis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosyntheis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Gene counts
500
400
300
200
100
0
Acetothermia
Acidobacteria
Actinobacteria
Aerophobetes
Aminicenantes
Aquificae
Armatimonadetes
Atribacteria
BRC1
Bacteroidetes
Caldiserica
Calditerrae
Calescamantes
Candidatus Atribacteria
Candidatus Saccharibacteria
Chlamydiae
Chloroflexi
Chrysiogenetes
Cloacimonetes
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
EM3
Elusimicrobia
Fervidibacteria
Fibrobacteres
Firmicutes
Fusobacteria
Gemmatimonadetes
Gracilibacteria
Hydrogenedentes
Ignavibacteriae
Latescibacteria
Lentisphaerae
Marinimicrobia
Microgenomates
Nitrospinae
Nitrospirae
Omnitrophica
PER
Parcubacteria
Planctomycetes
Poribacteria
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Thermodesulfobacteria
Thermotogae
Verrucomicrobia
WS1
unclassified


A4
Valine biosynthesis
Ubiquinone biosynthesis
Trypotophan biosynthesis
Translation factors and enzymes involved in translation
Transcriptional regulators
Thymidylate biosynthesis
Threonine biosynthesis
Thiamine biosynthesis
TCA cycle
Ribosomal proteins - small subunit
Ribosomal proteins - large subunit
Riboflavin biosynthesis
Pyruvate decraboxylation
Pyrimidine salvage
Pyrimidine biosynthesis
Pyridoxal phosphate biosynthesis
Purine salvage
Purine biosynthesis
Proline biosynthesis
Preprotein translocase subunits
Phenylalanine/tyrosine biosynthesis
Pentose phosphate pathway
NADH:Ubiquinone oxidreductase subunits
NAD biosynthesis
NA+-transporting NADH:Ubiquinone oxireductase subunits
Multisubunit NA+/H+ antiporter
Methionine biosynthesis
Menaquinone biosynthesis
Lipid A biosynthesis
Leucine biosyntehsis
Isoleucine biosynthesis
Histidine biosynthesis
Heme biosynthesis
Glyoxylate bypass
Glycolysis
Gluconeogenesis
Flagellum structure and biogenesis
Fatty acid biosyntheis
FAD biosynthesis
F0F1-type ATP synthase subunits
Entner-Doudoroff pathway
DNA-dependent RNA polymerase subunits
DNA polymerase III subunits
Deoxyxylulose pathway of terpenoid biosynthesis
Coenzyme A biosynthesis
Cobalamin biosynthesis
Biotin biosynthesis
Basal transription factors
Basal replication machinery
Arginine biosynthesis
Archael/Vacuolar-type H+ ATPase subunits
Aminoacyl-tRNA synthetases and alternate systems for amino acid activation
Gene counts
30
20
10
0
Aigarchaeota
Crenarchaeota
Diapherotrites
Euryarchaeota
Thaumarchaeota

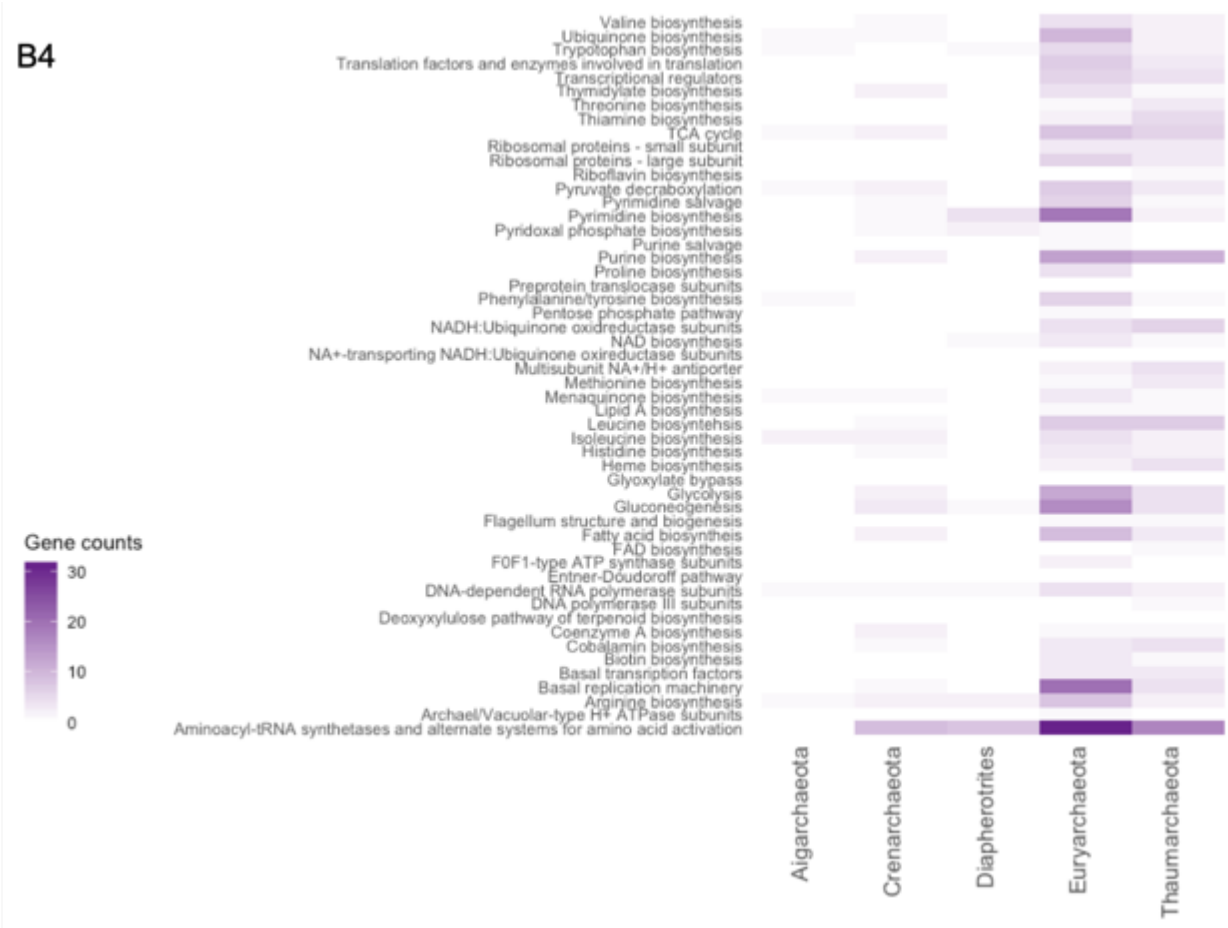


Figure 4.8- Associated pathways of COG-coding genes in (A) filtered fraction and (B) unfiltered fraction of Conwy River water samples. Colour intensity is correlated to gene counts. The x-axis are phyla and the graphs are divided into the domains. (1) Viruses are green, (2) Eukaryotes are red, (3) Bacteria are blue, and (4) Archaea are purple.

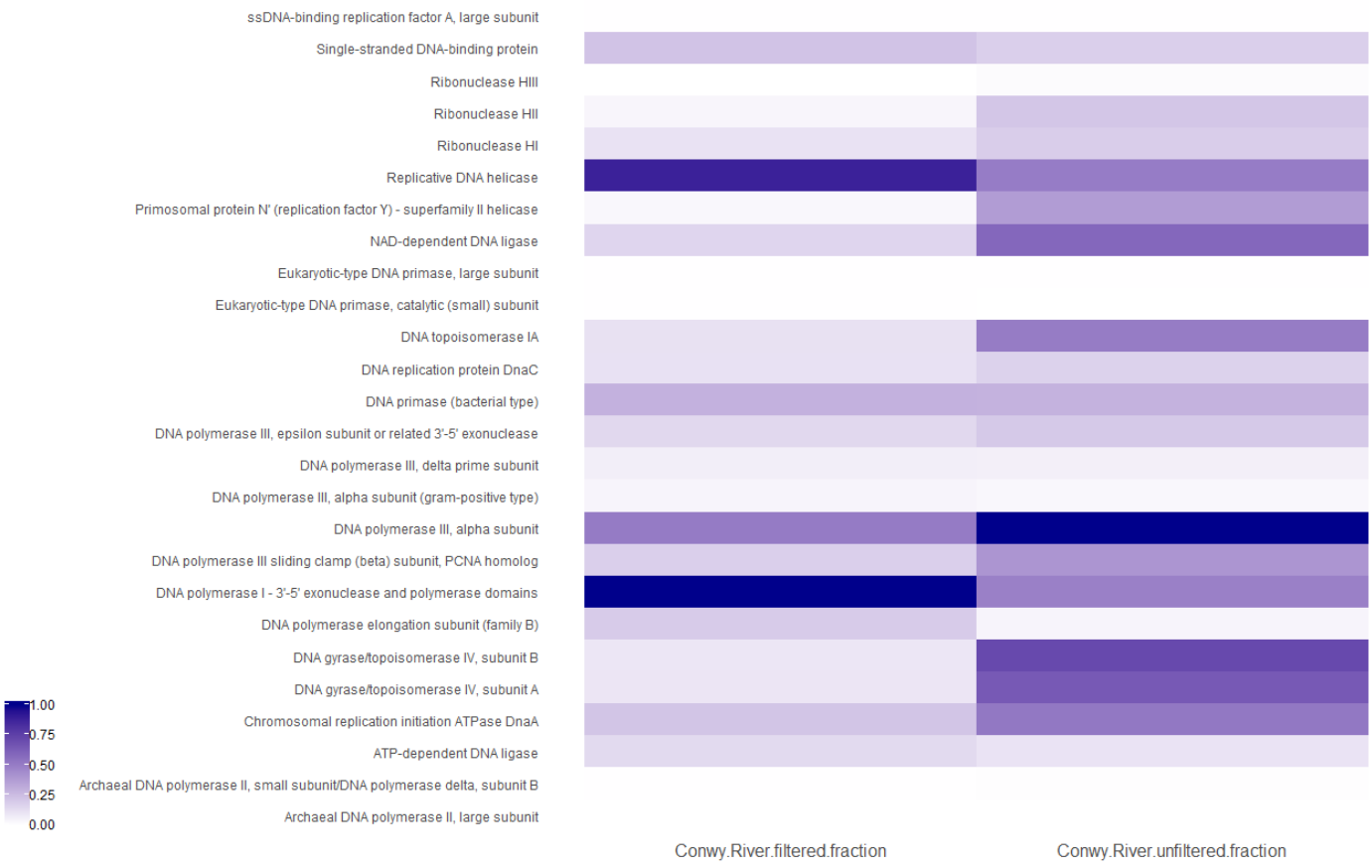


Figure 4.9- Individual COGs associated with the pathway of basal replication machinery in both fractions. The gene counts were scaled (normalized) to a value between 0 and 1.

4.3.2.3. Distribution of COG functional categories

We chose 20 prokaryotic species of various genome sizes (Table 4.6) for comparative analysis to determine essential functionalities by utilizing a scaled (values between 0 and 1) gene count of proteins that fall under broad COG categories (Fig. 4.10). We also compared 14 similar ecosystems with the filtered and unfiltered fractions (Table 4.6). The filtered water from the Afon Conwy has a unique distribution when compared to the other systems (Fig. 4.10). The principle component analysis (PCA) of functional COG categories confirms this as well, showing that in terms of COG functional categories present, the filtered fraction has unique distribution of COGs when compared to the other metagenomes across various environments (Fig. 4.11 and Table 4.6). K clustering showed that the filtered Afon Conwy microbial content is unlike all metagenomes analysed here (Fig. 4.12). Further hierarchical clustering confirmed this and the dendrogram presented showed that the filtered fractions share similar COG categories present within small genome-sized microorganisms than with those with larger genomes and other metagenomes (Fig. 4.12).

Table 4.6- List of metagenome and genomes for clustering used for comparative analysis. Genome size, lifestyle and domain are indicated for genomes used, whereas location and environment are listed for metagenomes. Metagenomes and genomes were obtained from the IMG/MER database.

Number	*Name*	*Type*	*Genome Size (Mbp)*	*Free Living?*	*Domain*	*Location*	*Environment*
G1	*Sinorhizobium meliloti 2011*	Genome	6.7	Yes	Bacteria	NA	NA
G2	*Agrobacterium tumefaciens C58*	Genome	5.7	Yes	Bacteria	NA	NA
G3	*Pseudomonas putida F1*	Genome	5.6	Yes	Bacteria	NA	NA
G4	*Escherichia coli K12 DH1 ATCC 33849*	Genome	4.6	Yes	Bacteria	NA	NA
G5	*Methanosarcina acetivorans C2A*	Genome	5.75	Yes	Archaea	NA	NA
G6	*Haloterrigena turkmenica VKM B 1734 DSM 5511*	Genome	5.4	Yes	Archaea	NA	NA
G7	*Sphingomonas aestuarii DSM 19475*	Genome	3.0	Yes	Bacteria	NA	NA
G8	*Haemophilus influenzae 2019*	Genome	2.0	No	Bacteria	NA	NA
G9	*Metallosphaera cuprina Ar 4*	Genome	1.8	Yes	Archaea	NA	NA
G10	*Ferroplasma acidarmanus Fer1*	Genome	1.8	Yes	Archaea	NA	NA
G11	*Thermoplasma acidophilum DSM 1728*	Genome	1.6	Yes	Archaea	NA	NA
G12	*Candidatus Pelagibacter ubique SAR11 HTCC1062*	Genome	1.3	Yes	Bacteria	NA	NA
G13	*Methanothermus fervidus V24S DSM 2088*	Genome	1.2	Yes	Archaea	NA	NA
G14	*Ignicoccus hospitalis KIN4 I DSM 18386*	Genome	1.3	Yes	Archaea	NA	NA
G15	*Rickettsia prowazekii BuV67 CWPP*	Genome	1.1	Yes	Bacteria	NA	NA
G16	*Wigglesworthia glossinidia*	Genome	0.7	No	Bacteria	NA	NA
G17	*Buchnera aphidicola APS*	Genome	0.66	No	Bacteria	NA	NA
G18	*Mycoplasma genitalium G37*	Genome	0.58	Yes	Bacteria	NA	NA
G19	*Nanoarchaeum equitans Kin4 M*	Genome	0.49	No	Archaea	NA	NA
G20	*Candidatus* Tremblaya princeps PCIT	Genome	0.14	No	Bacteria	NA	NA
M1	Freshwater microbial communities	Metagenome	NA	NA	NA	Columbia River, USA	Freshwater
M2	Lotic microbial communities from Mississippi River from River Site 1	Metagenome	NA	NA	NA	Mississippi River, MN, USA	Freshwater
M3	Lotic microbial communities from Mississippi River from River Site 7	Metagenome	NA	NA	NA	Mississippi River, MN, USA	Freshwater
M4	Sediment microbial communities from subsurface aquifer	Metagenome	NA	NA	NA	Rifle, CO, USA	Freshwater
M5	Isolated Sinkhole microbial communities	Metagenome	NA	NA	NA	Lake Huron, MI, USA	Freshwater
M6	Marine microbial community	Metagenome	NA	NA	NA	La Parguera, Puerto Rico	Marine
M7	Pelagic marine microbial communities	Metagenome	NA	NA	NA	North Sea	Marine

M8	Hypoxic sulfidic aquatic microbial communities from Monarch Geyser	Metagenome	NA	NA	NA	Yellowstone National Park, USA	Freshwater
M9	Fountain water microbial communities from Alpena County Library	Metagenome	NA	NA	NA	Alpena County, MI, USA	Freshwater
M10	Groundwater microbial communities from the Olkiluoto Island deep subsurface site	Metagenome	NA	NA	NA	Finland	Freshwater
M11	Groundwater microbial communities from the Aspo Hard Rock Laboratory HRL deep subsurface site	Metagenome	NA	NA	NA	Sweden	Freshwater
M12	Groundwater microbial communities from subsurface biofilms in sulfidic aquifier	Metagenome	NA	NA	NA	Frasassi Gorge, Italy	Freshwater
M13	Acid mine drainage microbial communities from abandoned Hg mine	Metagenome	NA	NA	NA	Los Rueldos, Spain	Acid Mine
M14	Bog forest soil microbial communities	Metagenome	NA	NA	NA	Calvert Island, British Columbia, Canada	Soil
M15	Afon Conwy community unfiltered fraction	Metagenome	NA	NA	NA	North Wales, UK	Freshwater
M16	Afon Conwy community filtered fraction (0.2 μm filter)	Metagenome	NA	NA	NA	North Wales, UK	Freshwater

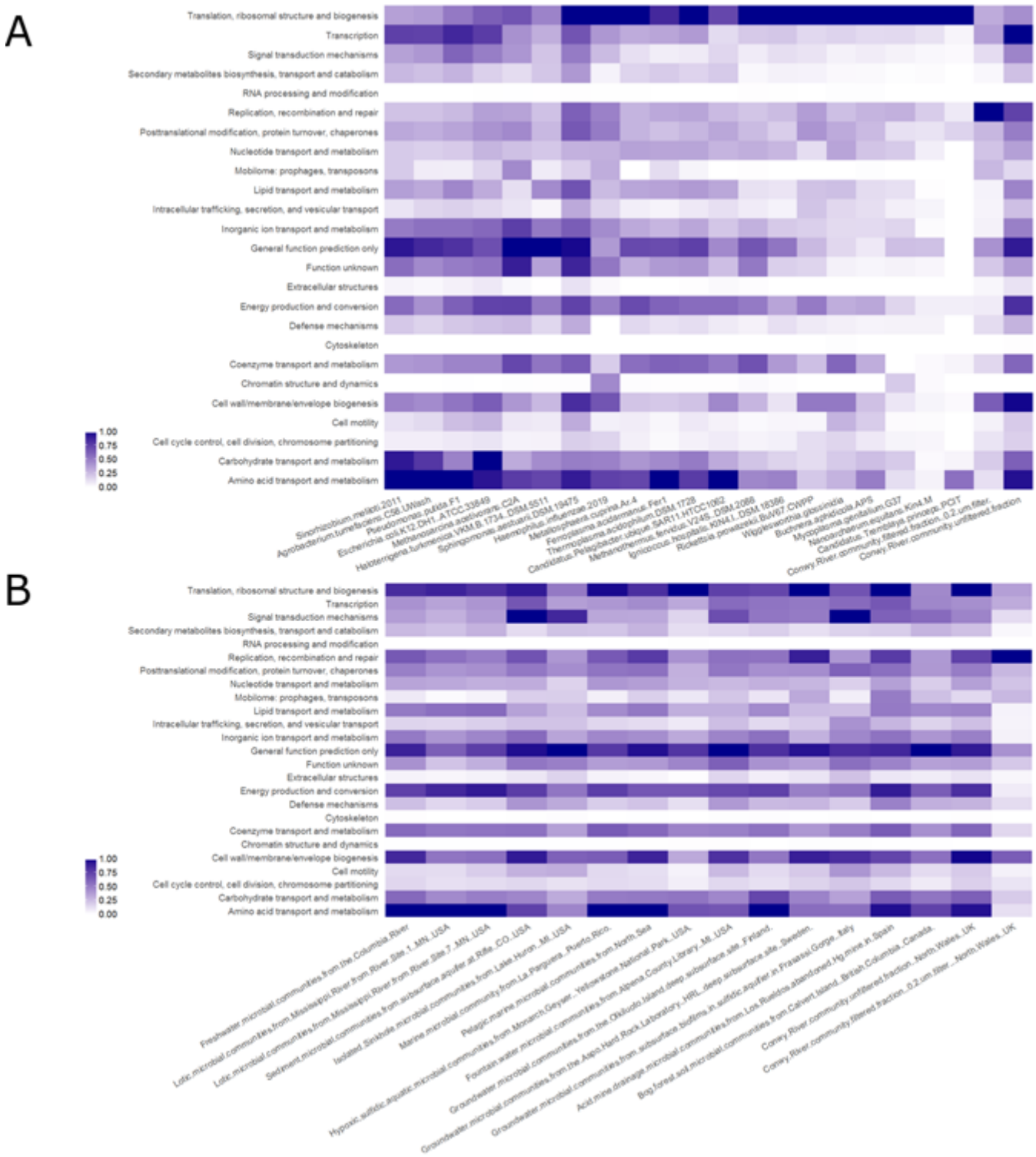


Figure 4.10- Distribution heat map of COG functions comparing Conwy River fraction metagenomes to (A) prokaryotic genomes and (B) other metagenomes (environments). The gene counts were scaled (normalized) to a value between 0 and 1. For (A), organisms are arranged left to right in order of decreasing genome size. Genome sizes ranged from 6.7 Mbp (Sinorhizobium meliloti) to 0.14 Mbp (Ca. Trembleya princeps). Conwy river water fractions are on the far right side. See table 8 for further information regarding metqagenomes and genomes.

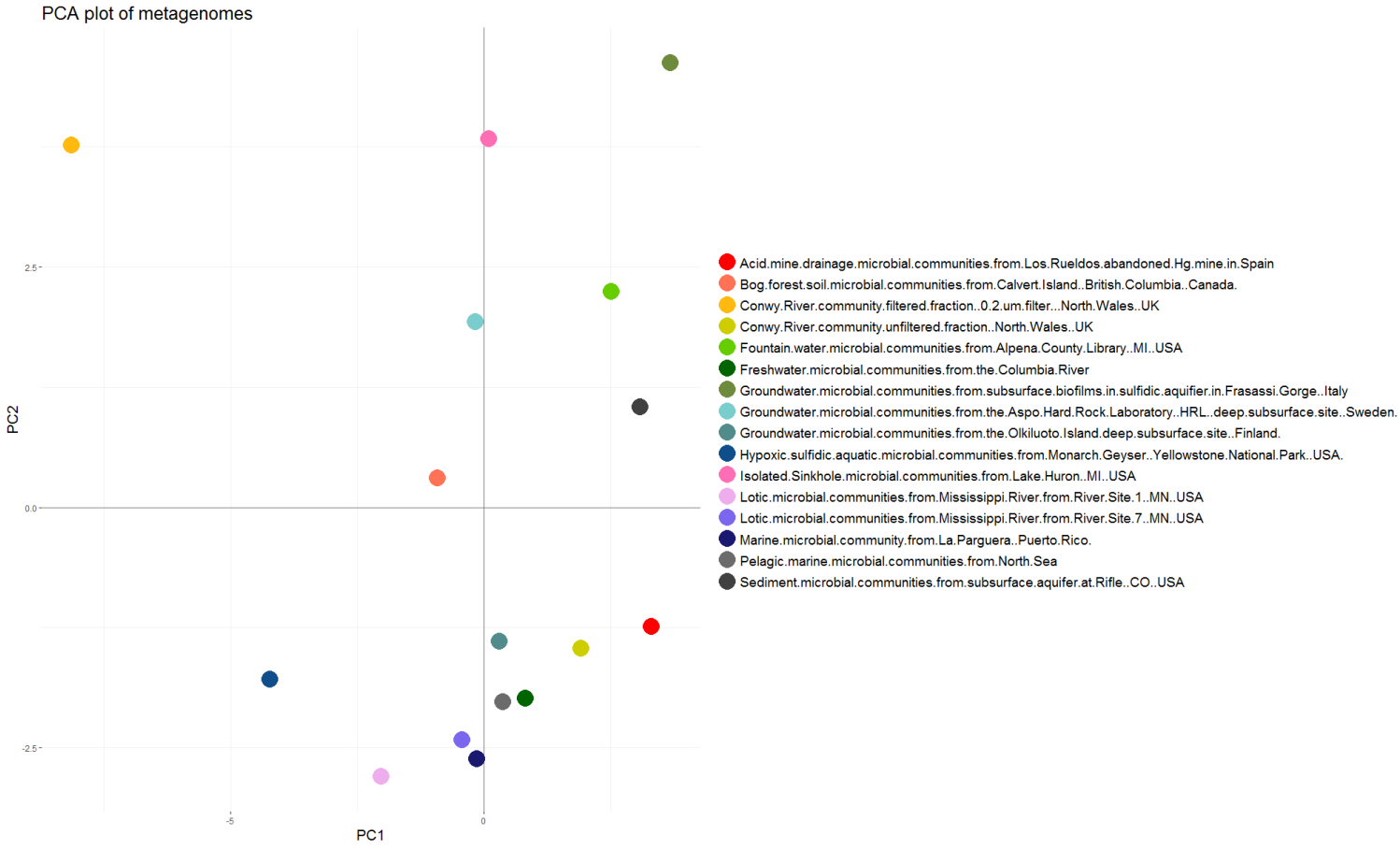


Figure 4.11- PCA of metagenomes. Calculations were made using the normalized gene counts based on COG functions.

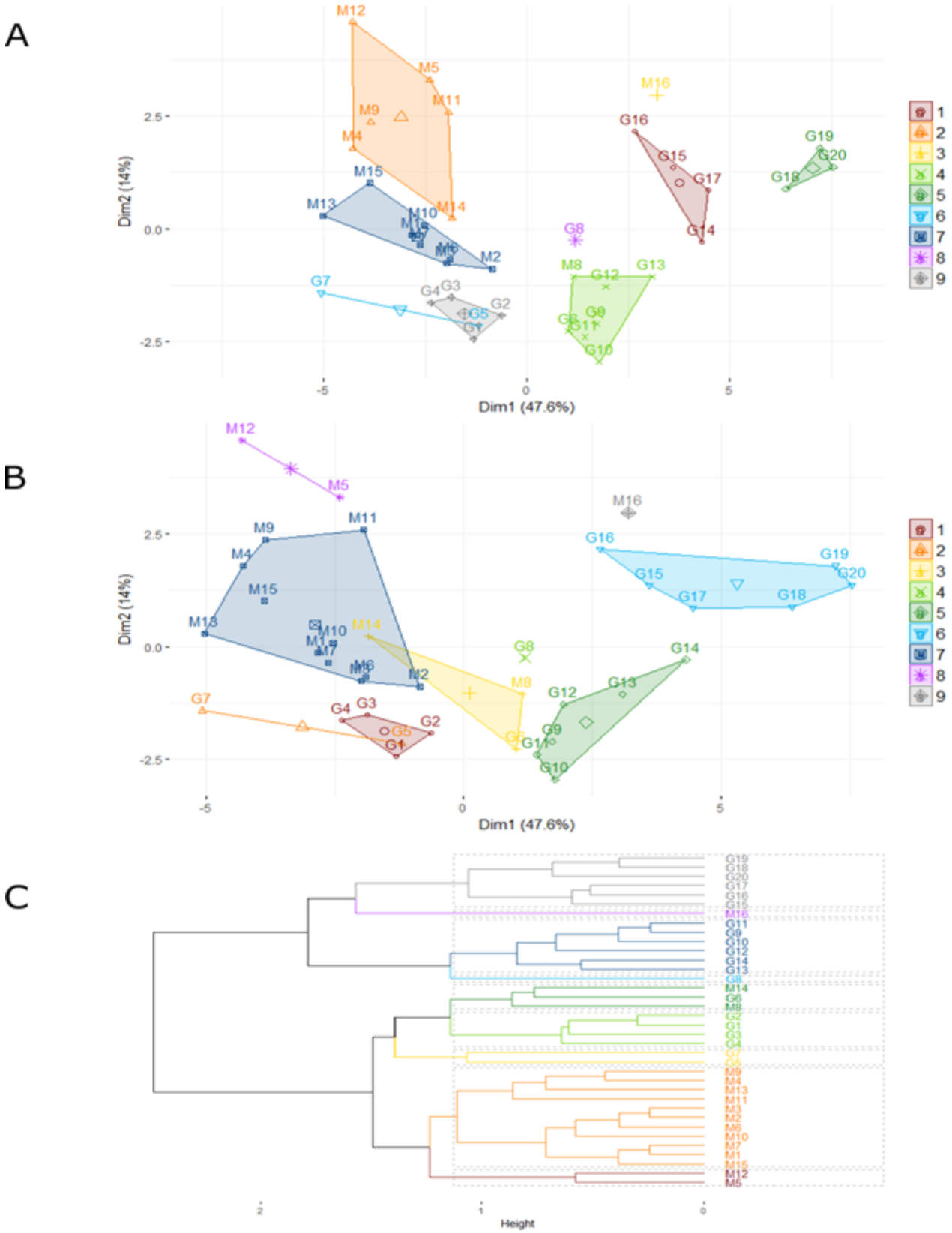


Figure 4.12- Clustering of single celled genomes and metagenomes. (A) K cluster plot, (B) hierarchal cluster plot and (C) hierarchal dendrogram are shown here. Calculations were made using the normalized gene counts based on COG functions. It was predicted that the optimal number of clusters was 9.

4.3.3. Amplicon sequencing (Oxford Nanopore™)

Similar to the shotgun sequencing results, there were more reads in the filtered versus the unfiltered section (Table 4.7). In addition, there is a higher percentage of classified reads in the unfiltered sample versus the filtered sample (Table 4.8). Average alignment accuracies, i.e. the degree of similarity the resultant sequences share with the sequences within the NCBI database, of both samples remained about the same (83-84%). *Proteobacteria* was the dominating phylum in both samples (Fig. 4.13 and Table 4.8). For the most part, the distribution of phyla in the filtered fraction mirrored the unfiltered fraction, except for *Firmicutes* (7.6% versus 2.1%) and *Spirochaetes* (9% versus 0.23%) (Fig. 4.13 and Table 4.8). At the genus level *Polynucleobacter*, *Pseudomonas* and *Legionella* are the three most prevalent genera in the unfiltered fraction, whereas, *Ralstonia*, *Variovorax*, *Spirochaeta,* and *Geobacter* are the most prevalent in the filtered fraction (Fig. 4.13 and Table 4.9). Of the 34 genera examined, only six are present in both fractions (*Ralstonia, Geobacter, Cupriavidus, Bdellovibrio, Rhodoferax, and Polarmonas*). The remaining 28 are found in one fraction and not the other (Fig. 4.13 and Table 4.9). According to Table 4.9, the genus with the highest alignment accuracy was *Roseovarius* in the filtered fraction (92.9%) and the lowest was *Oxobacter*, also in the filtered fraction (77.8%).

Table 4.7-Summary of amplicon sequencing. Total yield (Mbp), total number of reads, average sequence length (bp), quality score, number of classified and unclassified reads, and average alignment accuracy are listed here.

Sample	*Total yield (Mbp)*	*Total Reads*	*Average sequence length (bp)*	*Average Quality Score*	*Reads Classified*	*Reads Unclassified*	*Average Alignment Accuracy*
Filtered fraction	5.2	4,313	1,199	9.49	564 (13%)	3,749 (87%)	84%
Unfiltered fraction	5.0	3,475	1,446	9.58	2,632 (76%)	843 (24%)	83%

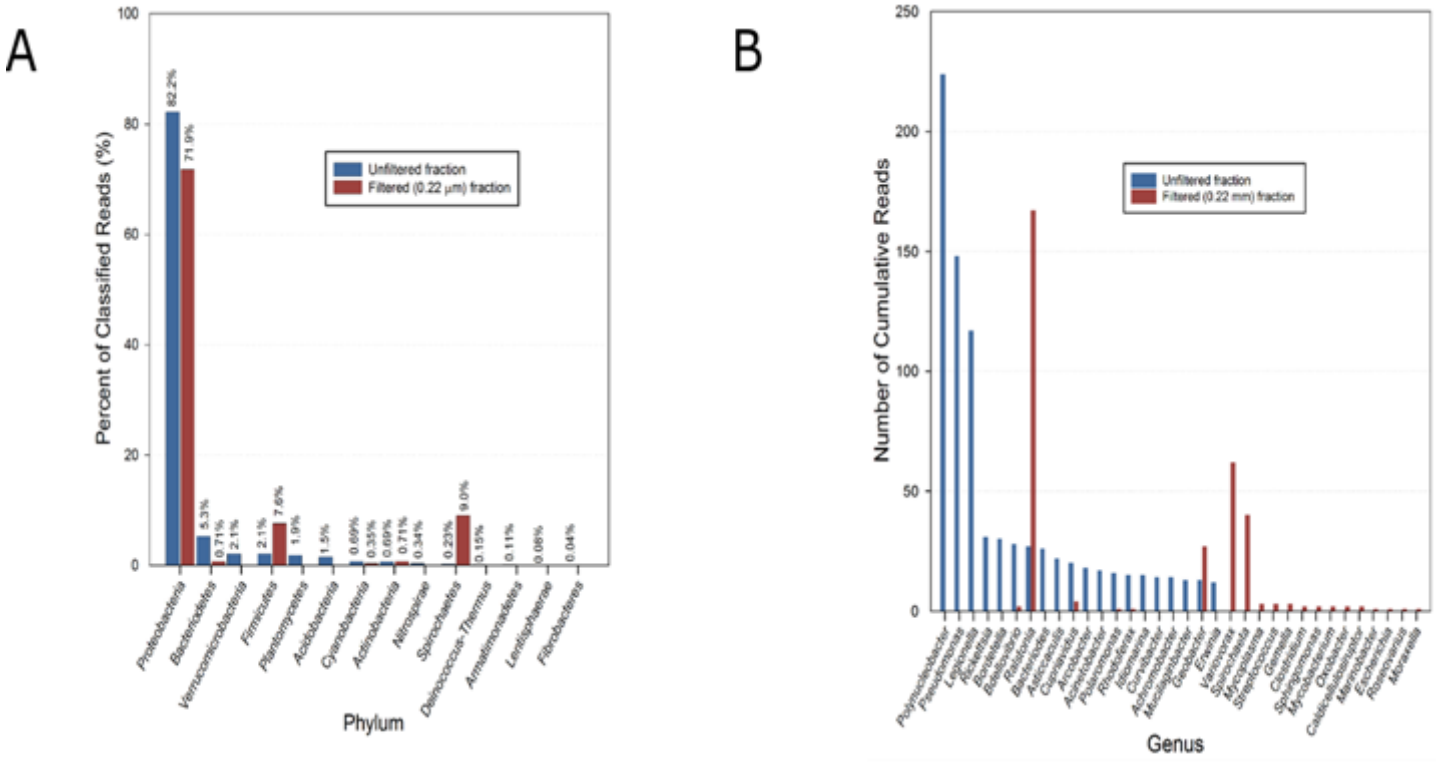


Figure 4.13-(A) Bacterial phyla distributions and (B) cumulative classified reads of the top 20 genera present in the Afon Conwy. More information regarding number of cumulative reads and average alignment accuracy can be found in Tables 4.8 and 4.9.

Table 4.8-Phyla identified from the sample NM29 classified reads in unfiltered and filtered fractions. The number and percentage of classified reads are listed as well as corresponding alignment accuracy.

	Unfiltered Fraction		***Filtered Fraction***	
Phylum	***Classified cumulative reads***	***Average Alignment Accuracy (%)***	***Classified cumulative reads***	***Average Alignment Accuracy (%)***
Proteobacteria	2,155 (82.2%)	83.3	405 (71.9%)	85.4
Bacteriodetes	138 (5.3%)	83.9	4 (0.71%)	84.8
Verrucomicrobacteria	55 (2.1%)	81.3		
Firmicutes	54 (2.1%)	81.9	43 (7.6%)	80.1
Plantomycetes	50 (1.9%)	80.9		
Acidobacteria	39 (1.5%)	81.5		
Cyanobacteria	18 (0.69%)	81.2	2 (0.35%)	81.4
Actinobacteria	17 (0.69%)	81.0	4 (0.71%)	81.8
Nitrospirae	9 (0.34%)	82.3		
Spirochaetes	6 (0.23%)	85.4	51 (9.0%)	84.7
Deinococcus-Thermus	4 (0.15%)	80.7		
Armatimonadetes	3 (0.11%)	84.7		
Lentisphaerae	2 (0.08%)	82.8		
Fibrobacteres	1 (0.04%)	80.9		

Table 4.9- Top 20 genera obtained from the site NM 29 classified reads based on cumulative reads. Genus, Phylum, cumulative reads per taxa, and average alignment accuracy (%) for both fractions are listed

Unfiltered Fraction				***Filtered Fraction***			
Genus	***Phylum***	***Cumulative Reads***	***Average Alignment Accuracy (%)***	***Genus***	***Phylum***	***Cumulative Reads***	***Average Alignment Accuracy (%)***
Polynucleobacter	*Proteobacteria*	224	86.4	*Ralstonia*	*Proteobacteria*	167	87.2
Pseudomonas	*Proteobacteria*	148	85.6	*Variovorax*	*Proteobacteria*	62	87.1
Legionella	*Proteobacteria*	117	83.0	*Spirochaeta*	*Spirochaetes*	40	85.3
Rickettsia	*Proteobacteria*	31	82.1	*Geobacter*	*Proteobacteria*	27	81.0
Bordetella	*Proteobacteria*	30	85.7	*Cupriavidus*	*Proteobacteria*	4	83.1
Bdellovibrio	*Proteobacteria*	28	83	*Mycoplasma*	*Tenericutes*	3	79.1
Ralstonia	*Proteobacteria*	27	84.7	*Streptococcus*	*Firmicutes*	3	88.7
Bacteriodes	*Bacteriodetes*	26	85.8	*Gemella*	*Firmicutes*	3	86.7
Asticcaculis	*Proteobacteria*	22	85.3	*Bdellovibrio*	*Proteobacteria*	2	79.1
Cupriavidus	*Proteobacteria*	20	83.4	*Clostridium*	*Firmicutes*	2	78.1
Arcobacter	*Proteobacteria*	18	86.4	*Sphingomonas*	*Proteobacteria*	2	86.4
Acinetobacter	*Proteobacteria*	17	83.9	*Mycobacterium*	*Actinobacterium*	2	82.4
Polaromonas	*Proteobacteria*	16	84.4	*Oxobacter*	*Firmicutes*	2	77.8
Rhodoferax	*Proteobacteria*	15	85.3	*Caldicellulosiruptor*	*Firmicutes*	2	79.3
Idiomarina	*Proteobacteria*	15	86.6	*Marinobacter*	*Proteobacteria*	1	81.1
Curvibacter	*Proteobacteria*	14	84.8	*Escherichia*	*Proteobacteria*	1	86.8
Achromobacter	*Proteobacteria*	14	83.4	*Roseovarius*	*Proteobacteria*	1	92.9
Mucilaginbacter	*Bacteriodetes*	13	85.3	*Moraxella*	*Proteobacteria*	1	80.1
Geobacter	*Proteobacteria*	13	79.3	*Rhodoferax*	*Proteobacteria*	1	83.5
Erwinia	*Proteobacteria*	12	85.7	*Polaromonas*	*Proteobacteria*	1	86.4

4.4. Discussion

4.4.1. Taxonomic Community Composition

The 16S rRNA sequences retrieved from the *de novo* method, Sanger sequencing, and from Nanopore™ are related to uncultured microorganisms, i.e represent "microbial dark matter". More specifically, the 16S rRNA sequences of the filtered fraction are closely related to the following phyla, candidate division OP11 ('*Candidatus* Microgenomates'), candidate division OD1 ('*Ca.* Pacebacteria'), and '*Ca.* Kaiserbacteria' (Table 4.5). All of which are predicted to be ultramicro/ultra-small microbial species (Wrighton et al., 2012; Brown et al., 2015; He et al., 2015; Luef et al., 2015). As seen in Table 4.5, as well Figure 4.4, three clones of the five in the clonal library that were closely related to potential ultra-small bacteria were also found in the unfiltered fraction. In addition, in Figure 4.4 and Table 4.5, there is one sequence, Ga0136236 10202521, that was found in the unfiltered fraction even though it is again related to a potential ultra-small organism from the candidate phylum OD1 (accession number: JN540162). The abundance of these cells in the unfiltered fraction may indicate either one or both scenarios.

The BLAST matches from the Sanger sequencing revealed that there were a number of bacteria that either have the capability to degrade synthetic materials and hazardous wastes or have been isolated from hazardous/polluted anthropogenic sites (Table 4.1), namely *Dongia mobilis* (nitrate reducer), *Dehalogenimonas mccarti* (dehalogenates aliphatic and aromatic organohalides), *Thioalkalivibrio sulfidiphus* (sulphur oxidizer), *Arsenicicoccus bolidensis* (arsenic reducer), *Pelobacter massiliensis* (hydroxyhydroquinone degradation), *Thermosporothrix hazakensis*, *Levilinea saccharolytica*, and *Desulfobacca acetoxidans* (sulfate reducer) in the unfiltered fraction (Schnell et al., 1991; Oude Elferink et al., 1999; Yamada et al., 2006; Routh et al., 2007; Moe et al., 2009; Liu et al., 2010; Yabe et al., 2010; Sorokin et al., 2012; Pöritz et al., 2013). In addition to the filtered fraction, the sequences

were matched with microorganisms residing in two different environments, i.e. anthropogenic and non-anthropogenic. The first set are microorganisms obtained from natural environments (non-anthropogenic) such as: the soil/ rhizosphere (Lesaulnier et al., 2008; Youssef et al., 2009; Tabei and Ueno, 2010; Jin et al., 2014), freshwater (Newton et al., 2007; Rossi et al., 2013; Yueqing, 2013), and subterranean (Brown et al., 2015; Shiraishi et al., 2016). The second are microorganisms that originated from anthropogenic perturbed environments; such as mining operations (Liu et al., 2009; Tang et al., 2012; Zhan and Sun, 2012; Tomczyk-Zak et al., 2013), farming (Bauer et al., 2009), and wastewater treatment (Rivière et al., 2009). Although the organisms were first described in these specific environments (i.e nature vs man-made) they are in fact native prior to human interference.

In regards to amplicon sequencing via Nanopore™, the difference number of unassigned sequences versus assigned sequences is stark in both fractions (Table 4.7). Predictably, the majority of the sequences in the filtered fraction are unassigned (87%) comparared to the unfiltered fraction (13%) (Table 4.7). This is largely in part due to using the 16S rRNA genes of bacteria and archaea collection BLAST instead of nucleotide collection (nr/nt). In addition to this, the literature suggests that there is a relatively low level of accuracy, as noted in the average alignment accuracies (Tables 4.8- 4.9), especially with high G+C content sequences (Laver et al., 2015). This probably accounted for the high number of unclassified reads.

As for the reads that were classified, there was an overabundance of *Ralstonia*, *Variovorax*, *Spirochaeta*, and *Geobacter* genera related reads in the filtered fraction (Fig. 4.13 and Table 4.9). This is the result of two factors. Firstly, *Ralstonia* and *Variovorax* genera possess the ability to pass through 0.2 μm filters (Zwart et al., 2002; Ryan and Adley, 2014). As for *Geobacter* and *Spirochaeta*, although their filterability has not been extensively studied, it may be possible for them to pass too (Wang et al., 2007, 2008). Secondly, all the

microorganisms of these genera are native to both soil and freshwater systems (Hahn et al., 2004; Ryan et al., 2007; Bergey, 2010; Lovley et al., 2011; Satola et al., 2013). Soil erosion is a natural process and as a result a large number of soil microbiota were found.

In addition to the noted overabundance of the sequences affiliated with these two genera, there was a unique distribution at the phylum level in the filtered fraction. Namely *Firmicutes* and *Spirochaetes* were notable members (Fig. 4.13 and Table 4.8). As stated previously, *Spirochaeta* (which taxonomically fall under the *Spirochaetes* phylum), may be morphologically pliable. *Firmicutes* have a sporulation phase, therefore, we think that spores are more likely to be passing through the filter than metabolically active cells.

Overall the 16S rRNA amplicon sequencing data (both Sanger and Nanopore™), does resemble shotgun sequencing results on a phylum level in that (1) the dominant phylum is *Proteobacteria* in both fractions, (2) *Bacteriodetes* is the second most common phylum in the unfiltered fraction, and (3) *Firmicutes* is the second most common phylum in the filtered fraction.

In regards to the presence of archaea in the Conwy River, only the shotgun sequencing showed that there were *Euryarchaeota* (filtered fraction 0.17% and unfiltered fraction 0.38%) and *Thaumarchaeota* (filtered fraction 0.20% and unfiltered fraction 0.17%) in both fractions in low abundance (Fig. 4.5). This finding is not entirely surprising as it is estimated that less than 10% (up to 37% in extreme cases) of the microbial community in freshwater ecosystems are archaea (Wells et al., 2006; Bomberg et al., 2008; Herfort et al., 2009; Cavicchioli, 2011). It is thought that in non-extreme environments, archaea play a variety of roles in regards to general nutrient cycling (Wells et al., 2006; Bomberg et al., 2008; Herfort et al., 2009; Cavicchioli, 2011).

4.4.2. Overview of COGs present in the Conwy River

There were significant differences in terms of the distribution of COGs functional categories between the fractions. First, rarefaction of the COGs showed that there the filtered fraction is much less diverse than the unfiltered fraction (Fig. 4.6). The rarefaction curve plateaus around approximately 3,000 COGs, whereas the filtered fraction begins to plateau around 2,000 COGs (Fig. 4.6). This result is what we expected to find as the unfiltered community encompasses all individuals, including the filtered fraction.

Second, the filtered fraction had a higher distribution of COGs associated with the functional category of replication, recombination, and repair, more specifically the affiliated with the basal replication machinery pathway (Fig. 4.5, 4.7-4.9). Closer examination of specific domain(s) under the basal replication machinery pathway show that replicative DNA helicase and DNA polymerase I 3'-5' exonuclease and polymerase domains are overrepresented in the filtered fraction. Helicases and polymerases are fundamental to DNA replication, where helicases separate annealed nucleic acid strands via ATP hydrolysis and polymerase assembles and lengthens long chains of nucleic acid strands. Replicative DNA helicase, found in all cells, DNA polymerase I, first described in *E. coli* and is only found in prokaryotes, specifically removes the RNA primer and replaces it with corresponding DNA (Lehman et al., 1958; Patel et al., 2001; Tuteja and Tuteja, 2004). Their overall ubiquity may explain its apparent overabundance in the filtered fraction.

However, the overrepresentation of these domains may be tied to the presence of dsDNA viruses in the filtered fraction, as there are significantly more domains (especially basal replication machinery) associated with viral entities in the filtered fraction by almost an entire order of magnitude (Figs. 4.7-4.8). This is significant as some viruses are primarily comprised of DNA helicases, as in the case of bovine papilloma virus (Tuteja and Tuteja,

2004; Yang et al., 2006). Therefore the helicases present may not be from prokaryotic entities, but are actually viruses.

4.4.3. Comparing the distribution of COG functional categories in genomes and metagenomes

The relationship between the presence/absence of COGs to genome size of an organism was explored because it is commonly thought that as genome size, cellular complexity decreases (Konstantinidis and Tiedje, 2004; Giovannoni et al., 2014; Martínez-Cano et al., 2015). This complexity is often associated with cells of limited volumes, although there are numerous organisms that defy this assumption, such as "*Candidatus* Tremblaya princeps" (Martínez-Cano et al., 2015). Nonetheless, organisms that have limited functionality could also be an indicator of nutrient presence. Copiotrophic environments often harbour motile large bodied organisms that take advantage of the high concentration of nutrients (Ghuneim et al., 2018). Such capabilities could be considered "expensive" genomically as more proteins are needed to motor flagella, secondary metabolism, hence more base pairs needed to encode for said proteins. On the other hand, inhabitants of oligotrophic environments are often small (i.e. ultra-micro) with limited mobility, often encoding for bare essential functions (such as DNA upkeep and repair), hence are less "expensive" than copiotrophic counterparts (Ghuneim et al., 2018).

Our investigation showed that there was a clear relationship between absence/presence of specific COGs and an organism's genome size (Fig 4.10 and Table 4.6). As the genome size decreases, there was a high prevalence of COGs dedicated to essential function (i.e. transcription, ribosomal structure and biogenesis) whereas COGs dedicated to transcription and amino acid transport and metabolism were less prevalent (Fig. 4.10). The same trend was also observed in a previous study which 115 genomes were compared (Konstantinidis and Tiedje, 2004).

We then compared the composition of COG functional categories present in metagenomes (including the fractions of the Conwy) and genomes via cluster analysis. Clustering analysis revealed that there was little overlap in terms of COG functional groups of genomes and metagenomes (i.e. metagenomes cluster with other metagenomes while the same goes for genomes; Figs. 4.11-4.12). However, there was only one metagenome that did not follow this pattern, the Conwy River filtered fraction. When represented via PCA, K cluster, or hierarchal graphical cluster, the filtered fraction of the Conwy River is unique because it occupies its own cluster across all methods (Figs. 4.11-4.12). When we examined only the metagenomes, the filtered fraction of the Afon Conwy was unique in terms of COG composition when compared to all other environments in this analysis (Fig. 4.10). Upon closer inspection, the dendrogram of the hierarchal clustering shows that it shows the highest resemblance to small genome organisms (Fig. 4.12). This could be an indicator that there are small-bodied organisms residing in the filtered fraction. Yet, it may be unlikely as pelagic environments contain a plethora of small organisms but did not cluster with the filtered fraction (Figs. 4.12). Most likely, this along with the previous analysis, may be an indication that there are fragments (eDNA) and dormant forms of cells, and that only a select few of the cells present in this fraction are metabolically active.

4.5. Conclusion

There is a clear distinction between the content of filtered and unfiltered fractions in terms of the distribution of phyla and COGs present. The majority of microorganisms residing in the Conwy River are *Proteobacteria* followed by *Firmicutes* in the filtered fraction and *Bacteriodetes* in the unfiltered fraction. The filtered fraction of Conwy River water contains a high number of genes dedicated to COGs affiliated with replication, recombination and repair, more specifically COGs that fall under the COG pathway of basal replication and machinery (specifically COGs that are associated with DNA repair/replication

functions) when compared to the whole community and other metagenomes. Although there is a clear connection between genome size (i.e. complexity) and presence/absence of COGs, this could not be carried over to metagenomes. In conclusion, there is significant microbial diversity within the Afon Conwy that garners more attention.

4.6. Acknowledgments and contributions

Lydia-Ann Ghuneim conducted the experiments, analysis, data interpretation and writing the manuscript. Francesca Brailsford and Helen Glanville assisted with freshwater sample collection. Rafael Bargiela assisted in data interpretation and statistical analysis. Tatyana Chernikova carried out the Nanopore sequencing. Stephan Toshchakov carried out shotgun sequencing, assembly, and annotations. David Jones, Peter Golyshin, and Olga Golyshina provided significant revisions as well as insight into data interpretation. This work was carried out under the DOMAINE project, which is funded by the UK Natural Environment Research Council (NERC) (large grant NE/K010689/1). Olga Golyshina and Peter Golyshin acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government and thank for its support the European Union Horizon 2020 Research and Innovation Program [Blue Growth: Unlocking the Potential of Seas and Oceans, grant agreement no. 634486, Project acronym 'INMARE']. We declare no conflict of interest.

4.7. References

Adriaenssens, E. M., Farkas, K., Harrison, C., Jones, D., Allison, H. E., and Mccarthy, A. J. (2018). Viromic analysis of wastewater input 2 to a river catchment reveals a diverse 3 assemblage of RNA viruses. 3, 1–39. doi:10.1101/248203.

Akinbami, O. A. (2013). Use of multiple displacement amplification based approaches for detection and analysis of environmentally significant and contaminating bacteria in fresh water. Available at: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.602407.

Anagnostidis, K. and Komárek, J. (1985). Modern Approach to the Classification System of Cyanophytes. *Arch. Hydrobiol. Suppl. Algol. Stud.* 38, 291–302.

Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. Available at: https://cran.r-project.org/package=gridExtra.

Baker, B. J., and Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* 44, 139–152. doi:10.1016/S0168-6496(03)00028-X.

Battin, T. J., Kaplan, L. A., Findlay, S., Hokinson, C. S., Marti, E., Packman, A. I., et al. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* 1, 95–100. doi:10.1038/ngeo602.

Bauer, R., Bekker, J. P., Wyk, N. van, du Toit, C., Dicks, L. M. T., and Kossmann, J. (2009). Exopolysaccharide production by lactose-hydrolyzing bacteria isolated from traditionally fermented milk. *Int. J. Food Microbiol.* 131, 260–264. doi:10.1016/j.ijfoodmicro.2009.02.020.

Bergey, D. H. (2005). *Bergey's Manual of Systematic Bacteriology - Vol 2*. doi:10.1007/978-0-387-68572-4.

Bergey, D. H. (2010). *Bergey's Manual of Systematic Bacteriology - Vol 4: Bacteroidetes*. Springer doi:10.1007/978-0-387-68572-4.

Bomberg, M., Montonen, L., Münster, U., and Jurgens, G. (2008). Diversity and function of

archaea in freshwater habitats. *Curr. Trends Microbiol.* 4, 1–89. Available at: http://www.researchtrends.net/tia/abstract.asp?in=0&vn=4&tid=41&aid=2693&pub=2008&type=3.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208–211. doi:10.1038/nature14486.

Cavicchioli, R. (2011). Archaea - Timeline of the third domain. *Nat. Rev. Microbiol.* 9, 51–61. doi:10.1038/nrmicro2482.

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., et al. (2007). Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems* 10, 172–185. doi:10.1007/s10021-006-9013-8.

Collins, M. D., Routh, J., Saraswathy, A., Lawson, P. A., Schumann, P., Welinder-Olsson, C., et al. (2004). Arsenicicoccus bolidensis gen. nov., sp. nov., a novel actinomycete isolated from contaminated lake sediment. *Int. J. Syst. Evol. Microbiol.* 54, 605–608. doi:10.1099/ijs.0.02918-0.

Dang, H., and Lovell, C. R. (2016). Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* 80, 91–138. doi:10.1128/MMBR.00037-15.Address.

Desai, N., Antonopoulos, D., Gilbert, J. A., Glass, E. M., and Meyer, F. (2012). From genomics to metagenomics. *Curr. Opin. Biotechnol.* 23, 72–76. doi:10.1016/j.copbio.2011.12.017.

Dewhirst, F. E., Seymour, C., Fraser, G. J., Paster, B. J., and Fox, J. G. (1994). Phylogeny of Helicobacter isolates from bird and swine feces and description of Helicobacter pametensis sp. nov. *Int. J. Syst. Bacteriol.* 44, 553–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7520743.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Dunaj, S. J., Vallino, J. J., Hines, M. E., Gay, M., Kobyljanec, C., and Rooney-Varga, J. N. (2012). Relationships between soil organic matter, nutrients, bacterial community structure, and the performance of microbial fuel cells. *Environ. Sci. Technol.* 46, 1914–22. doi:10.1021/es2032532.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Farkas, K., Cooper, D. M., McDonald, J. E., Malham, S. K., de Rougemont, A., and Jones, D. L. (2018). Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. *Sci. Total Environ.* 634, 1174–1183. doi:10.1016/j.scitotenv.2018.04.038.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution (N. Y).* 39, 783–791. doi:Doi 10.2307/2408678.

Fukunaga, Y., Kurahashi, M., Sakiyama, Y., Ohuchi, M., Yokota, A., and Harayama, S. (2009). Phycisphaera mikurensis gen. nov., sp nov., isolated from a marine alga, and proposal of Phycisphaeraceae fam. nov., Phycisphaerales ord. nov and Phycisphaerae classis nov in the phylum Planctomycetes. *J. Gen. Appl. Microbiol.* 55, 267–275.

Ghuneim, L.-A. J., Jones, D. L., Golyshin, P. N., and Golyshina, O. V (2018). Nano-Sized

and Filterable Bacteria and Archaea: Biodiversity and Function. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.01971.

Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1–13. doi:10.1038/ismej.2014.60.

Goossens, H., Pot, B., Vlaes, L., Vandenborre, C., Vandenabbeele, R., Vannaelten, C., et al. (1990). Characterization and Description of Campylobacter-Upsaliensis Isolated from Human Feces. *J. Clin. Microbiol.* 28, 1039–1046.

Hahn, M. W., Stadler, P., Wu, Q. L., and Pöckl, M. (2004). The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *J. Microbiol. Methods* 57, 379–390. doi:10.1016/j.mimet.2004.02.004.

Hall, T. T. A. T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98. doi:citeulike-article-id:691774.

Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi:10.1128/MMBR.68.4.669-685.2004.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. doi:10.1016/S1074-5521(98)90108-9.

Hartmann, M., Lee, S., Hallam, S. J., and Mohn, W. W. (2009). Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. *Environ. Microbiol.* 11, 3045–3062. doi:10.1111/j.1462-2920.2009.02008.x.

He, X., McLean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S.-Y., et al. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and

epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci. U. S. A.* 112, 244–9. doi:10.1073/pnas.1419038112.

Herfort, L., Kim, J., Coolen, M., Abbas, B., Schouten, S., Herndl, G., et al. (2009). Diversity of Archaea and detection of crenarchaeotal amoA genes in the rivers Rhine and Têt. *Aquat. Microb. Ecol.* 55, 189–201. doi:10.3354/ame01294.

Imachi, H., Sakai, S., Lipp, J. S., Miyazaki, M., Saito, Y., Yamanaka, Y., et al. (2014). Pelolinea submarina gen. nov., sp. nov., an anaerobic, filamentous bacterium of the phylum Chloroflexi isolated from subseafloor sediment. *Int. J. Syst. Evol. Microbiol.* 64, 812–818. doi:10.1099/ijs.0.057547-0.

Jin, H., Yang, X. Y., Yan, Z. Q., Liu, Q., Li, X. Z., Chen, J. X., et al. (2014). Characterization of rhizosphere and endophytic bacterial communities from leaves, stems and roots of medicinal Stellera chamaejasme L. *Syst. Appl. Microbiol.* 37, 376–385. doi:10.1016/j.syapm.2014.05.001.

Kale, V., Björnsdóttir, S. H., Fridjónsson, Ó. H., Pétursdóttir, S. K., Ómarsdóttir, S., and Hreggvidsson, G. Ó. (2013). Litorilinea aerophila gen. nov., sp. nov., an aerobic member of the class Caldilineae, phylum Chloroflexi, isolated from an intertidal hot spring. *Int. J. Syst. Evol. Microbiol.* 63, 1149–1154. doi:10.1099/ijs.0.044115-0.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi:10.1007/bf01731581.

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3160–3165. doi:10.1073/pnas.0308653100.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, msw054.

doi:10.1093/molbev/msw054.

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi:10.1016/j.bdq.2015.02.001.

Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958). Enzymatic Synthesis of Deoxyribonucleic Acid. I . Preparation of substrates and partial purfication of an enzyme from Escherichia coli. *J. Biol. Chem.* 233, 163–170.

Lesaulnier, C., Papamichail, D., McCorkle, S., Ollivier, B., Skiena, S., Taghavi, S., et al. (2008). Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen. *Environ. Microbiol.* 10, 926–941. doi:10.1111/j.1462-2920.2007.01512.x.

Liu, R., Zhang, Y., Ding, R., Li, D., Gao, Y., and Yang, M. (2009). Comparison of archaeal and bacterial community structures in heavily oil-contaminated and pristine soils. *J. Biosci. Bioeng.* 108, 400–407. doi:10.1016/j.jbiosc.2009.05.010.

Liu, Y., Jin, J. H., Liu, Y. H., Zhou, Y. G., and Liu, Z. P. (2010). Dongia mobilis gen. nov., sp. nov., a new member of the family Rhodospirillaceae isolated from a sequencing batch reactor for treatment of malachite green effluent. *Int. J. Syst. Evol. Microbiol.* 60, 2780–2785. doi:10.1099/ijs.0.020347-0.

Löffler, F. E., Yan, J., Ritalahti, K. M., Adrain, L., Edwards, E. A., Konstantinidis, K. T., et al. (2013). Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov. and famil. *Int. J. Syst. Evol. Microbiol.* 63, 625–635.

Lovley, D. R., Ueki, T., Zhang, T., Malvankar, N. S., Shrestha, P. M., Flanagan, K. A., et al. (2011). Geobacter: The Microbe Electric's Physiology, Ecology, and Practical

Applications. *Adv. Microb. Physiol.* 59, 1–100. doi:10.1016/B978-0-12-387661-4.00004-5.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi:10.1038/ncomms7372.

Margesin, R., Zhang, D. C., and Busse, H. J. (2012). Sphingomonas alpina sp. nov., a psychrophilic bacterium isolated from alpine soil. *Int. J. Syst. Evol. Microbiol.* 62, 1558–1563. doi:10.1099/ijs.0.035964-0.

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40. doi:10.1093/nar/gkr1044.

Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 6. doi:10.3389/fmicb.2014.00742.

Meyer, J. L. (1994). The microbial loop in flowing waters. *Microb. Ecol.* 28, 195–199. doi:10.1007/BF00166808.

Moe, W. M., Yan, J., Nobre, M. F., da Costa, M. S., and Rainey, F. A. (2009). Dehalogenimonas lykanthroporepellens gen. nov., sp. nov., a reductively dehalogenating bacterium isolated from chlorinated solvent-contaminated groundwater. *Int. J. Syst. Evol. Microbiol.* 59, 2692–2697. doi:10.1099/ijs.0.011502-0.

Mudge, S. M., and Norris, C. E. (1997). Lipid biomarkers in the Conway Estuary (North Wales, UK): a comparison between fatty alcohols and sterols. *Mar. Chem.* 57, 61–84.

Muyzer, G., Sorokin, D. Y., Mavromatis, K., Lapidus, A., Clum, A., Ivanova, N., et al. (2011). Complete genome sequence of "Thioalkalivibrio sulfidophilus" HL-EbGr7. *Stand. Genomic Sci.* 4, 23–35. doi:10.4056/sigs.1483693.

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., and Bertilsson, S. (2011). *A guide to the natural history of freshwater lake bacteria.* doi:10.1128/MMBR.00028-10.

Newton, R. J., Jones, S. E., Helmus, M. R., and McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Appl. Environ. Microbiol.* 73, 7169–7176. doi:10.1128/AEM.00794-07.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). vegan: Community Ecology Package. Available at: https://cran.r-project.org/package=vegan.

Oude Elferink, S. J. W. H., Akkermans-van Vliet, W. ., Bogte, J. J., and Stams, A. J. M. (1999). Desulfobacca acetoxidans gen. nov., sp. nov., a novel acetate-degrading sulfate reducer islated from sulfidogenic granular sludge. *Int. J. Syst. Evol. Microbiol.* 49, 345–350.

Pascual, C., Collins, M. D., Grimont, P. A. D., Dominguez, L., and Fernandez-Garayzabal, J. F. (1996). Sanguibacter inulinus sp. nov. *Int. J. Syst. Bacteriol.* 46, 811–813.

Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L. A. (2001). Prokaryotic DNA polymerase I: Evolution, structure, and “base flipping” mechanism for nucleotide selection. *J. Mol. Biol.* 308, 823–837. doi:10.1006/jmbi.2001.4619.

Pöritz, M., Goris, T., Wubet, T., Tarkka, M. T., Buscot, F., Nijenhuis, I., et al. (2013). Genome sequences of two dehalogenation specialists - Dehalococcoides mccartyi strains BTF08 and DCMB5 enriched from the highly polluted Bitterfeld region. *FEMS Microbiol. Lett.* 343, 101–104. doi:10.1111/1574-6968.12160.

R Core Team (2017). R: A language and environment for statistical computing. Available at: https://www.r-project.org/.

Rivière, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., Weissenbach, J., et al. (2009). Towards the definition of a core of microorganisms involved in anaerobic

digestion of sludge. *ISME J.* 3, 700–14. doi:10.1038/ismej.2009.2.

Rossi, P. G., Laurion, I., and Lovejoy, C. (2013). Distribution and identity of bacteria in subarctic permafrost thaw ponds. *Aquat. Microb. Ecol.* 69, 231–245. doi:10.3354/ame01634.

Routh, J., Saraswathy, A., and Collins, M. D. (2007). Arsenicicoccus bolidensis a novel arsenic reducing actinomycete in contaminated sediments near the Adak mine (northern Sweden): Impact on water chemistry. *Sci. Total Environ.* 379, 216–225. doi:10.1016/j.scitotenv.2006.06.027.

Ryan, M. P., and Adley, C. C. (2014). Ralstonia spp.: Emerging global opportunistic pathogens. *Eur. J. Clin. Microbiol. Infect. Dis.* 33, 291–304. doi:10.1007/s10096-013-1975-9.

Ryan, M. P., Pembroke, J. T., and Adley, C. C. (2007). Ralstonia pickettii in environmental biotechnology: Potential and applications. *J. Appl. Microbiol.* 103, 754–764. doi:10.1111/j.1365-2672.2007.03361.x.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–25. doi:citeulike-article-id:93683.

Salcher, M. M. (2014). Same same but different: Ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi:10.4081/jlimnol.2014.813.

Satola, B., Wübbeler, J. H., and Steinbüchel, A. (2013). Metabolic characteristics of the species Variovorax paradoxus. *Appl. Microbiol. Biotechnol.* 97, 541–560. doi:10.1007/s00253-012-4585-z.

Saw, J. H. W., Schatz, M., Brown, M. V., Kunkel, D. D., Foster, J. S., Shick, H., et al. (2013). Cultivation and Complete Genome Sequencing of Gloeobacter kilaueensis sp. nov., from a Lava Cave in Kilauea Caldera, Hawai'i. *PLoS One* 8.

doi:10.1371/journal.pone.0076376.

Schnell, S., Brune, A., and Schink, B. (1991). Degradation of hydroxyhydroquinone by the strictly anaerobic fermenting bacterium Pelobacter massiliensis sp. nov. *Arch. Microbiol.* 155, 511–516. doi:10.1007/BF00244971.

Shiraishi, F., Mitsunobu, S., Suzuki, K., Hoshino, T., Morono, Y., and Inagaki, F. (2016). Dense microbial community on a ferromanganese nodule from the ultra-oligotrophic South Pacific Gyre: Implications for biogeochemical cycles. *Earth Planet. Sci. Lett.* 447, 10–20. doi:10.1016/j.epsl.2016.04.021.

Sigee, D. C. (2005). *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment*. Wiley Available at: https://books.google.co.uk/books?id=AJ8nIkoB1o0C.

Simpson, J. H., Vennell, R., and Souza, A. J. (2001). The salt fluxes in a tidally-energetic estuary. *Estuar. Coast. Shelf Sci.* 52, 131–142. doi:10.1006/ecss.2000.0733.

Sorokin, D. Y., Muntyan, M. S., Panteleeva, A. N., and Muyzer, G. (2012). Thioalkalivibrio sulfidiphilus sp. nov., a haloalkaliphilic, sulfur-oxidizing gammaproteobacterium from alkaline habitats. *Int. J. Syst. Evol. Microbiol.* 62, 1884–1889. doi:10.1099/ijs.0.034504-0.

Tabei, Y., and Ueno, K. (2010). Phylogenic analysis of bacteria passed through 0.45-microm-pore-size filters in the rhizosphere. *J. Gen. Appl. Microbiol.* 56, 129–36. doi:10.2323/jgam.56.129.

Tang, Y. Q., Ji, P., Lai, G. L., Chi, C. Q., Liu, Z. S., and Wu, X. L. (2012). Diverse microbial community from the coalbeds of the Ordos Basin, China. *Int. J. Coal Geol.* 90–91, 21–33. doi:10.1016/j.coal.2011.09.009.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*

28, 33–36. doi:10.1093/nar/28.1.33.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi:10.1093/nar/22.22.4673.

Tomczyk-Zak, K., Kaczanowski, S., Drewniak, Ł., Dmoch, Ł., Sklodowska, A., and Zielenkiewicz, U. (2013). Bacteria diversity and arsenic mobilization in rock biofilm from an ancient gold and arsenic mine. *Sci. Total Environ.* 461–462, 330–340. doi:10.1016/j.scitotenv.2013.04.087.

Tuteja, N., and Tuteja, R. (2004). Prokaryotic and eukaryotic DNA helicases: Essential molecular motor proteins for cellular machinery. *Eur. J. Biochem.* 271, 1835–1848. doi:10.1111/j.1432-1033.2004.04093.x.

Wagner, C., Mau, M., Schlömann, M., Heinicke, J., and Koch, U. (2007). Characterization of the bacterial flora in mineral waters in upstreaming fluids of deep igneous rock aquifers. *J. Geophys. Res. Biogeosciences* 112. doi:10.1029/2005JG000105.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi:10.1021/es0707198.

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* 42, 6749–6754. doi:10.1021/es800720n.

Wells, L. E., Cordray, M., Bowerman, S., Miller, L. A., Vincent, W. F., and Deming, J. W. (2006). Archaea in particle-rich waters of the Beaufort Shelf and Franklin Bay, Canadian Arctic: Clues to an allochthonous origin? *Limnol. Oceanogr.* 51, 47–59.

doi:10.4319/lo.2006.51.1.0047.

Wickham, H. (2007). Reshaping Data with the {reshape} Package. *J. Stat. Softw.* 21, 1–20. Available at: http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2009). *ggplot2: Elegant graphics for Data Analysis*. Springer-Verlag New York Available at: http://ggplot2.org.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* 40, 1–29. Available at: http://www.jstatsoft.org/v40/i01/.

Wickham, H. (2017). tidyr: Easily Tidy Data with "spread()" and "gather()" Functions. Available at: https://cran.r-project.org/package=tidyr.

Wickham, H., and Francois, R. (2016). dplyr: A Grammar of Data Manipulation. Available at: https://cran.r-project.org/package=dplyr.

Wickham, H., Francois, R., and Müller, K. (2017a). tibble: Simple Data Frames. Available at: https://cran.r-project.org/package=tibble.

Wickham, H., Hester, J., and Francois, R. (2017b). readr: Read Rectangular Text Data. Available at: https://cran.r-project.org/package=readr.

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla (supplement materials). *Science (80-.).* 337, 1661–5. doi:10.1126/science.1224041.

Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M., and Yokota, A. (2010). Thermosporothrix hazakensis gen. nov., sp. nov., isolated from compost, description of Thermosporotrichaceae fam. nov. within the class Ktedonobacteria Cavaletti et al. 2007 and emended description of the class Ktedonobacteria. *Int. J. Syst. Evol. Microbiol.* 60, 1794–1801. doi:10.1099/ijs.0.018069-0.

Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H., et al. (2006).

Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. nov., sp. nov., and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the. *Int. J. Syst. Evol. Microbiol.* 56, 1331–1340.

Yang, L., Mohr, I., Fouts, E., Lim, D. A., Nohaile, M., and Botchan, M. (2006). The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase. *Proc. Natl. Acad. Sci.* 90, 5086–5090. doi:10.1073/pnas.90.11.5086.

Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A., and Elshahed, M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl. Environ. Microbiol.* 75, 5227–5236. doi:10.1128/AEM.00592-09.

Yueqing, L. (2013). Microbial communities of spring pits in Jing-Mei River at the southeastern Taipei basin. *Master 's Degree Life Sci. ZTE Univ.*

Zhan, J., and Sun, Q. (2012). Diversity of free-living nitrogen-fixing microorganisms in the rhizosphere and non-rhizosphere of pioneer plants growing on wastelands of copper mine tailings. *Microbiol. Res.* 167, 157–165. doi:10.1016/j.micres.2011.05.006.

Zhou, J. L., Liu, Y. P., and Abrahams, P. W. (2003). Trace metal behaviour in the Conwy estuary, North Wales. *Chemosphere* 51, 429–440. doi:10.1016/S0045-6535(02)00853-6.

Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., and Han, S. K. (2002). Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* 28, 141–155. doi:10.3354/ame028141.

Chapter 5

Utilization of low molecular weight organic compounds by microbial communities residing in the Conwy River

Abstract

Filterable microorganisms have been shown to participate in carbon (C) cycling in freshwater systems. However, due to difficulties in their isolation and characterization, their functional role in freshwater ecosystems remains poorly understood. Previous investigations of dissolved organic C (DOC) cycling in the Conwy River provided preliminary evidence indicating the presence of metabolically active microorganisms within the 0.22 μm filtered water fraction. The purpose of this study was therefore to (1) determine taxonomic identity and community dynamics of bacterial and archaeal communities residing in ultra-filtered (0.22 μm pore size) and unfiltered freshwater obtained from the Conwy River (North Wales, UK), and (2) investigate their ability to cycle nutrients by examining their role in DOC (i.e. amino acids, sugars, and organic acids) utilization. Using ^{14}C-radioisotope tracking, 16S rRNA amplicon sequencing, shotgun sequencing, and targeted metabolomics, we uncovered an active microbial community capable of utilizing low molecular weight (LMW) compounds. The filtered fraction microorganisms began to utilize DOC after 74 h, overall consuming LMW DOC at a slower rate than the bulk (unfiltered) community. Microbial community structure was not affected by DOC addition, but changed with time. The dominant phyla of the filtered fraction included *Firmicutes*, *Proteobacteria*, *Bacteriodetes*, *Actinobacteria*, and *Spirochaetes*. In comparison, the unfiltered fraction was more diverse, comprising of *Proteobacteria*, *Bacteriodetes* and *Firmicutes,* as well as *Verrucomicrobia, Acidobacteria*, and *Actinobacteria*. Targeted metabolomics confirmed the existence of the lag phase in the filtered fraction within the first 74 h of the experiment. COGs connected to the utilization of DOC showed little variation over the course of the three weeks in both fractions as well as showing little effect of DOC addition. However, COGs associated with energy production increased in both fraction when substrate was added. Both shotgun sequencing and targeted metabolomics indicated that there was no preference in substrate type in either

fraction. We conclude that the filtered fraction (1) contains metabolically active cells that are capable of both uptake and mineralization of LMW DOC; (2) is comprised of a similar phyla found in the unfiltered fraction (with the exception of the overabundance of *Spirochaetes*); (3) is different in the distribution of these phyla; (4) has communities that are not influenced by LMW DOC addition; and (4) has COG functional categories associated with energy production are influenced by LMW DOC addition.

Keywords: dissolved organic carbon, low molecular weight compounds, filterable microorganisms, ultra-small microorganisms, 16S rRNA, lotic systems

5.1. Introduction

Filterable microorganisms refers to (1) nano-sized microorganisms, i.e. small-bodied microorganisms that have dimensions of 50-400 nm and a volume less than 0.1 μm^3, (2) larger cells that have the capability to squeeze through filters with pore sizes of less than 0.45 μm, and (3) small-cell variants of microorganisms with larger cell sizes (e.g. dormant or senescent forms) (Velimirov, 2001; Panikov, 2005; Duda et al., 2012; Ghuneim et al., 2018; Proctor et al., 2018). Our knowledge of the bacteria and archaea in the filterable fraction, however, remains poorly understood as most of these microorganisms have proven difficult to culture under laboratory conditions (Ghuneim et al., 2018). Nonetheless, filterable microorganisms appear ubiquitous throughout the biosphere and have been implicated in many geochemical processes ranging from sulfur reduction in pelagic systems, the consumption of photo-oxidation products of humic substances, and the production and consumption of dissolved organic matter (DOM) (Salcher, 2014; Dang and Lovell, 2016; Ghuneim et al., 2018).

DOM is defined as compounds that have the ability to pass through 0.45 μm filter and is the main source of organic nutrients in freshwater systems (Brailsford et al., 2017). The main source of DOM entering lotic systems is from external sources (i.e. agricultural runoff,

leaf detritus, wastewater discharges). The microorganisms responsible for using DOM in freshwater are largely heterotrophic benthic bacteria (Sigee, 2005). The over-proliferation of DOM can lead to algal blooms that decimate local wildlife populations, thus studying DOM is vital to ecosystem health (Beman et al., 2005). Yet examining DOM can be quite challenging, especially in lotic systems. One reason for this is the inherent ever-changing conditions of lotic systems (i.e. flow, weather events), while another is the constant anthropogenic inputs (agricultural runoff, pollutants) as a large proportion of towns and cities are located near bodies of water (Meyer, 1994; Sigee, 2005; Fenchel, 2008). Additionally, the levels of DOM within UK systems have been steadily increasing over the last 50 years due to intense anthropogenic activities (Ritson et al., 2014). As stated previously, this can lead to detrimental effects in the form of algal blooms by depleting local waters of oxygen and creating toxic metabolites that are harmful to local wildlife (Beman et al., 2005; Jones et al., 2016). DOM cycling can be observed using a combination of techniques including 16S rRNA amplicon sequencing, fluorescence in situ hybridization (FISH), stable isotope probing (SIP), stable isotope imaging (NanoSIMS), and radioisotope labelling (Roszak and Colwell, 1987; Findlay et al., 2003; Kirchman et al., 2004; Malmstrom et al., 2005; Rinta-Kanto et al., 2012; Landa et al., 2013).

A previous study on DOM cycling in the Conwy River showed the organisms passing through a 0.22 μm filter could remove ^{14}C-labelled glucose and amino acids and $^{33}PO_4^{3-}$ from solution (Brailsford et al., 2017). This is significant because it is normally thought that ultra-filtering aqueous samples would render the sample sterile (Brailsford et al., 2017; Ghuneim et al., 2018). It was concluded that there is a potentially robust community of filterable microorganisms which can actively participate in DOM cycling (Brailsford et al., 2017).

The purpose of this research was to compare microorganisms in 0.22 μm filtered aqueous samples (filterable microorganism community) versus those in the entire

community. Specifically, I wanted to assess: (1) whether the filterable microbial fraction could consume low molecular weight (LMW) C compounds, (2) whether it had a different taxonomic composition to the unfiltered fraction, and (3) whether the taxonomic composition of the filterable fraction would change in response to C addition and incubation time. I hypothesized that metabolically active cells are passing through the 0.22 µm filter that utilize LMW DOC and that these have a distinct taxonomic distribution and this community is sensitive to nM concentration changes in LMW DOC (i.e. the community distribution is altered and there is an increase in energy production).

5.2. Methodology

5.2.1. Description of catchment

The Conwy catchment is located in North Wales and its main drainage is the Conwy River (Afon Conwy). Mean annual rainfall ranges from 500-3500 mm and the mean annual air temperature ranges from 5-15 °C with an average annual temperature of 10 °C (Emmett et al., 2016). The river itself is 55 km long; starting from Llyn Conwy (450 m above sea level) and drains the Migneint, a large peatland bog that is a major store of carbon. Three tributaries (Machno, Lledr, and Llugwy), originating from the eastern side of the Snowdonia mountain range, join the main river further downstream before reaching the tidal limit (20 km inland) (Emmett et al., 2016). Average concentrations at the tidal limit are as follows: nitrite 0.2-2.8 mg/L, ammonium <0.03-0.04 mg/L, phosphate <0.02-0.05 mg/L, and dissolved organic carbon 1.5-10 mg/L (Emmett et al., 2016). The pH ranges from 5.7 to 7.2. The primary site used in this study (code: NM29) is located at Cwm Llanerch and is associated with the main Conwy River (53° 6' 24.7068" N, 3° 47' 28.7556" W). The site is located approximately 4 km from the tidal limit (Simpson et al., 2001).

5.2.2. Materials

Stock solutions of amino acid, sugar, and organic acid were generated. The amino acid stock was created by adding 600 µL of L-isomeric amino acid standard H (Thermo Fisher Scientific) to 100 mL of LC-MS water. Subsequent serial dilution achieved a final concentration, when added to each sample, of 15.05 nM. The sugar stock was created by dissolving 1.8 g of D-(+)-glucose, 1.8 g of D-(+)-fructose, and 3.42g of sucrose in 100 mL of LC-MS water along with serial dilutions to achieve a final concentration of glucose and fructose of 33.4 nM and sucrose of 31.6 nM per sample/replicate. Finally, the organic acid solution was created by adding 0.377 mL of formic acid, 1.34g of L-(-)-malic acid and 1.92 g of citric acid to 100 mL of LC-MS water. The final concentration of each organic acid per sample/replicate was 33.4 nM via serial dilution. All sugars, organic acids, and LC-MS water were purchased from Sigma-Aldrich with purities ≥ 99%.

5.2.3. Sample collection and substrate addition

Approximately 2 L of freshwater was collected from Cwm Llanerch on March 6, 2018. The collected samples were either unfiltered or filtered on-site through a 0.22 µm Sterivex® PVDF (hydrophilic polyvinylidene fluoride) filter (Millipore Corporation, Billerica, MA, USA). Samples were kept on ice and in the dark during transportation. The EC (electrical conductivity) and the pH was measured for both unfiltered (pH = 6.52, EC = 68 µS/cm) and filtered water (pH = 6.48, EC = 66 µS/cm) samples. Subsequent nutrients amendments via stock solutions were added dependent on whether the sample was being analyzed using non-radiolabeled assays or radiolabeled assays, *vide infra*. For the radiolabeled assays, 100 mL of sample (filtered, unfiltered, and blank) was placed into sterile 250 mL screw cap Erlenmeyer flasks. Subsequently, 1 ml of ^{14}C-substrate was added to the water, a 15 mL 1 M sodium hydroxide trap added and the flasks sealed (radiolabeled assay). An identical set of flasks was also set up for metabolite and metagenomic analysis but to

which no ^{14}C substrate was added (non-radiolabelled assays). All blank samples comprised autoclaved LC-MS grade water.

5.2.4. Radiolabeled assay (Depletion, CO_2 production, immobilized carbon, data collection and analysis)

^{14}C radioisotopes were used to determine: (i) the depletion of added substrate from solution, (ii) substrate-induced respiration (CO_2 production), and (iii) incorporation of C by the microorganisms into new biomass (C immobilization) in the filtered and unfiltered water samples. The filtered and unfiltered samples were spiked with one type of the radiolabeled ^{14}C radioisotope (3 different forms), separately, all with 0.1 kBq mL^{-1} activity: (1) radiolabeled amino acid stock solution: ^{14}C-amino acid mix (PerkinElmer, MA, USA Lot 3,590,279; 37,000 kBq/mL); (2) radiolabeled sugar stock solution: ^{14}C-glucose (PerkinElmer, MA, USA; Lot 3,632,475; 7,400 kBq/mL), ^{14}C-fructose (PerkinElmer, MA, USA; 37,000 kBq/mL), and ^{14}C-sucrose (PerkinElmer, MA, USA; 37,000 kBq/mL); or (3) radiolabeled organic acid stock solution: ^{14}C-citric acid (PerkinElmer, MA, USA; 1,850 kBq/mL), ^{14}C-formic acid (PerkinElmer, MA, USA; 34,906 kBq/mL) and ^{14}C-malic acid (PerkinElmer, MA, USA; 3,700 kBq/mL).

For substrate depletion, 500 µL of sample was aliquoted from the samples at 0, 1, 2, 4, 6, 22, 26, 49, 74, 141, 214, 333, and 506 h after substrate treatment and the samples placed into 1.5 mL Eppendorf tubes (Table S5.4). The samples were centrifuged at 20,817 *g* for 3 minutes. 250 µL of supernatant was removed and the remaining solution was discarded. 25 µL of 0.1 M HCl was added to the supernatant and allowed to incubate over 3 h to remove any dissolved CO_2 present. Finally, 4 mL of Optiphase HiSafe-3 scintillation cocktail (PerkinElmer) was added to the solution. For carbon dioxide production, 300 µL was taken from each 1 M NaOH trap at various times over the course of the experiment and then 4 mL of Optiphase HiSafe-3 scintillation cocktail (PerkinElmer) was added to the solution (Table

S5.4). ^{14}C in the solutions were taken using a Wallac 1404 liquid scintillation counter with automated quench correction (Wallac EG&G, Milton Keyes, UK). Biomass incorporation was calculated from the results obtained from CO_2 evolution and substrate utilization.

Mean and standard error at each time were calculated in R using the packages plyr (Wickham, 2011) and sciplot (Morales et al., 2017). Graphs were generated in R using the ggplot2 (Wickham, 2009) and gridExtra (Auguie, 2016) packages. Repeated measures ANOVA was performed on the ^{14}C data using SPSS Statistics 25 (IBM UK Ltd., Portsmouth, UK) to measure the effects of treatment over the three-week period. The Mauchly's test for sphericity was also performed. However, all values were non-significant ($p > 0.05$). The Greenhouse-Geisser estimate of sphericity was done to determine the proper correction value. If $\varepsilon < 0.75$, the Greenhouse-Geisser correction was applied and if $\varepsilon > 0.75$, the Huynh-Feldt correction was applied (Table S5.1). Post-hoc multiple pairwise testing was carried out using Tukey's post hoc multiple pairwise testing. The Games-Howell test was applied if the assumptions of the ANOVA test were not met.

5.2.5. Non-radiolabeled assays

5.2.5.1. 16S rRNA amplicon preparation, sequencing, and statistical analysis

Samples of river water (500 µL) were removed from the flasks at 0, 49, 141, 333, and 506 h and subsequently centrifuged (21,000 *g*, 10 min) removing the supernatant. The remaining pellet was then washed (x3) with PBS (phosphate buffer solution) (pH 7.4). For the preparation of Illumina-compatible libraries of V4 region of 16S rRNA gene, a dual-indexing primer system with heterogeneity spacer was used (Fadrosh et al., 2014). The rRNA-annealing parts of the primers corresponded to standard F515-R806 primers with slight modifications aimed to improve the coverage of environmental taxa (Table S5.2). All PCR reactions were performed in a Bio-Rad® thermocycler with the following program: 95 °C for 2 min for denaturation followed by 33 annealing cycles, 95 °C for 45 s, 50 °C for 1

min, 72 °C for 30 s, and finally 72 °C for 3 min. PCR products were checked using gel electrophoresis (1.8% agarose gel). A QIAquick gel extraction kit® (Qiagen) was used to purify PCR fragments from the agarose gel. A Qubit® dsDNA HS kit (Life Technology) with Qubit® Fluorometer was used to determine the concentration of DNA. Samples were then subsequently dried down via spin vacuum. The barcoded amplicons were sequenced with a MiSeq™ benchtop sequencer (Illumina Inc., San Diego, CA, USA) using paired-end 250 bp reads. All NGS reads were subjected to stringent quality filtering, parts of reads corresponding to 16S rRNA primers were removed using CLC Genomics Workbench 10.0 (Qiagen, Germany). After quality trimming, overlapping paired reads were merged with SeqPrep tool (https://github.com/jstjohn/SeqPrep).

Further processing, including demultiplexing, OTU generation and taxa assignment was performed with Qiime bioinformatics pipeline (Caporaso et al., 2010). Generation of OTUs was performed with open-reference algorithm (script pick_open_reference_otus.py). Taxa assignment was performed using 97% identity clustered sequences of Silva128 database (Yilmaz et al., 2014).

The R programming language (R Core Team, 2017) was used for statistical analysis and figure creation (Wickham, 2009). NMDS (using the Bray-Curtis calculation method) was used to examine beta diversity via phyloseq package (McMurdie et al., 2018). Rarefaction curves were created using the ranacapa package (Kandlikar, 2018). PERMANOVA via a mixed effect model was calculated from the relative abundances using the Bray-Curtis method with 999 permutations via the vegan package (Oksanen et al., 2018). Then subsequent stepwise model selection was utilized to determine which effects/mixed effects had the greatest influence on OTU absence/presence.

5.2.5.3. Shotgun sequencing

Samples of river water (500 µL) were removed from the flasks at the 0, 141, and 506 h and centrifuged (21,000 *g*, 10 min), removing the supernatant. Whole genome amplification (WGA) was achieved using the REPLI-g UltraFast Mini kit (Qiagen, Germany) as per the manufacturer's protocol. Ca.1 µg of DNA was sheared using a Bioruptor Pico sonicator (Diagenode) by sonicating at 4 °C with four cycles of 15 sec on and 90 sec off, to obtain fragments of 600-800 bp size.

Library preparation was performed using the NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs) according to the protocol provided by the manufacturer. Briefly, fragmented DNA was end repaired and ligated to the Illumina adaptor. Adaptor-ligated DNA was amplified with index primers provided in NEBNext Multiplex Oligos Set 1 and Set 2 (New England Biolabs). Size-selection and PCR clean up were performed by gel purification using the QIAEX II Gel extraction kit (Qiagen, Germany). Subsequently, the barcoded libraries were quantified using Qubit® dsDNA HS Assay Kit with the Qubit 4 Fluorometer. Samples were then pooled in equimolar amounts and the resulting pool was diluted to a final concentration of 4 nM.

The final pool was denatured and sequenced using the MiSeq Reagent Kit v3, 600 Cycles Sequencing kit on the MiSeq System (Illumina). Quality control was done using *fastqc* (Andrews, 2010) and adapter trimming using *cutadapt* (Martin, 2011), discarding those reads with average quality under 20 or shorter than 20 bps. Assembly was performed using MEGAHIT (Li et al., 2015). Gene prediction and annotation was done using PROKKA (Seemann, 2014). Additional annotation was added using emapper and DIAMOND (Buchfink et al., 2015) with the eggNOG database (Huerta-Cepas et al., 2016). Binning sample reads and assembled contigs were done using MaxBin (Wu et al., 2014). The resulting bins have been further classified with Kraken2 based on the NCBI RefSeq database

to get taxonomic classifications (Wood et al., 2019). Homoscedastic student's t-testing with two-tailed distribution was performed on the data (Table S5.4).

5.3. Results

5.3.1. Metabolic activity of ^{14}C-DOC

Overall, our blank (negative control) showed no signs of metabolic activity in comparison to the other treatments (Fig. 5.1). The unfiltered fraction showed the highest rate of consumption across the treatments, where the highest rate of change in concentration occurring after 74 h, and which then plateaued after 141 h. Within the filtered fraction, there was a clear lag phase across all substrates in the first 74 h (Fig. 5.1). Then from 74 to 141 h we observed a spike in metabolic activity across all substrate types until it slows at 214 h. Repeated measures ANOVA showed a significant difference ($p \leq 0.001$) between treatments and measurement time (Table S5.5). According to the F-values, there was a much larger effect due to treatment alone than the compounded effect of treatment and experiment duration (Table S5.5).

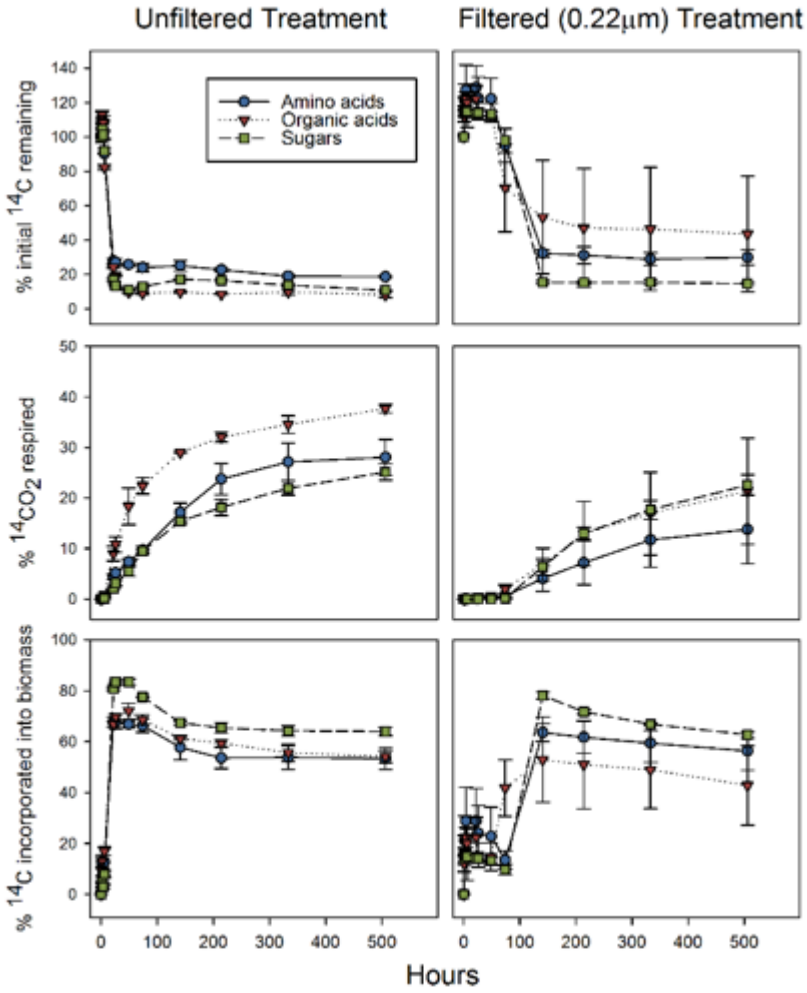


Figure 5.1- Utilization of three different ^{14}C-labelled substrates in 0.22 µm filtered and unfiltered river water. The upper panels show substrate depletion from the river water, the middle panels show the cumulative $^{14}CO_2$ production and the lower panels the amount of ^{14}C immobilized in the microbial biomass. Values represent means ± standard error (n = 3).

5.3.2. 16S rRNA amplicon sequencing

Over the course of three weeks, the prokaryotic (Bacteria and Archaea) community of both the fractions and nutrient amendment type (with and without addition) were examined using 16S rRNA amplicon sequencing analysis with short reads (approximately 250 base pairs). PCR was successful for 51 samples (including the replicates) which were subsequently analyzed by comparing the operational taxonomic units (OTUs) detected (Table S5.6). In total, there were 668,405 reads across all the samples and time points (306,244 in the filtered fraction, 362,161 in the unfiltered fraction) (Table S5.6).

Large differences in OTU composition and distribution were apparent between the filtering treatments and measurement times (Fig. 5.2). OTU of phyla that were prominent in the initial communities in the filtered fraction were "*Candidatus* Parcubacteria" (Candidate phylum OD1)*, Actinobacteria*, *Acidobacteria*, *Bacteriodetes*, *Firmicutes*, *Proteobacteria*, *Spirochetes*, and unassigned groups. As the experiment progressed, *Proteobacteria* became the dominant phyla as *Firmicutes*, "*Ca.* Parcubacteria", *Spirochetes*, *Cyanobacteria*, *Acidobacteria*, *Actinobacteria*, and unassigned groups decreased. In comparison, the unfiltered fraction was composed mainly of *Bacteriodetes*, *Verrucomicrobia*, *Acidobacteria*, *Actinobacteria*, *Firmicutes*, and *Proteobacteria* in the initial community (Fig. 5.2). *Firmicutes* decreased over the course of the three-week experiment while the proportion of *Actinobacteria*, *Armatimonadetes* and *Verrucomicrobia* increased. The appearance of minority phyla such as archaea of the *Euryarchaeota* and bacteria of the "*Ca.* Dependentiae" (Candidate phylum TM6) were dependent on whether substrate was added to the sample (Fig. 5.2).

The affiliation on the taxonomic levels of class and family was also undertaken. At the initial stages of the experiment, the filtered fraction was dominated by *Bacilli*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, *Spirochaetes*,

Alphaproteobacteria, *ZB2*, *Actinobacteriia*, and *Chloroplast* (Fig. 5.3 and Table S5.7). Upon further inspection at the family level, *Bacillaceae* (*Bacilli, Bacillales*), *Halomonadaceae* (*Oceanospirillales*, *Gammaproteobacteria*), and unassigned groups (i.e. unidentified taxa) were dominant (Figs. 5.3-5.4 and Table S5.7). On the other hand, the initial community of the unfiltered fraction was comprised predominantly of *Betaproteobacteria*, *Alphaproteobacteria*, *Bacteroidia*, *Clostridia*, *Flavobacteria*, *Gammaproteobacteria*, *Sphingobacteriia*, and *Bacilli*; with a fairly even distribution across a plethora families with the most dominant being *Oxalobacteraceae* (*Burkolderiales, Betaproteobacteria*), *Comamonadaceae* (*Burkolderiales, Betaproteobacteria*), *Flavobacteriaceae* (*Flavobacteriales*, *Flavobacteriia*), *Verrucomicrobiaceae* (*Verrucomicrobiales, Verrucomicrobiae*), and unassigned families (Figs. 5.3-5.4 and Table S5.7). During the later stages of the experiment (49-506 h), the filtered fraction was almost completely dominated by *Betaproteobacteria*, *Gammaproteobacteria*, and *Epsilonproteobacteria* (Fig. 5.3 and Table S5.7). More specifically, the most prevalent families were *Comamonadaceae*, *Campylobacteraceae* (*Campylobacterales, Epsilonproteobacteria*), *Oxalobacteraceae*, and *Pseudomonadaceae* (*Pseudomonadales, Gammaproteobacteria*) (Fig. 5.4 and Table S5.7). However, the microbial community in the unfiltered fraction remained relatively consistent in terms of distribution at both class and family levels, with the exception of *Bacteroidia* where they were only found in the first 49 h and *Verrucomicrobiaceae* which increased in the latter stages of the experiment (Figs. 5.3 and 5.4).

We also found that substrate addition did not alter OTU distribution greatly, however, there are some notable differences. Firstly, *Epsilonbacteria* in the filtered fraction appears primarily in the substrate amended samples (Fig. 5.3). Second, there are more *Alphaproteobacteria* in the filtered fraction without amendments than when substrate was present (Fig. 5.3). Third, *Actinobacteria* were more prevalent in the substrate amended

filtered fraction. Fourth, *Clostridia* were prevalent throughout the unfiltered fraction without substrate addition but not in the nutrient amended samples (Fig. 5.3). Finally, *Campylobacteraceae* in the filtered fraction are prevalent in nutrient amended sample (Fig. 5.4).

A PERMANOVA test was used to compare the effects of substrate addition, duration, and treatment on the relative abundance of OTUs present in the samples. Overall, the addition of substrate had no measureable effect on OTU abundance ($p > 0.05$), whereas treatment and duration have significant effects on the OTUs present ($p < 0.05$). After subsequent stepwise model selection, the greatest effect on the presence and absence of OTUs was a mixed effect of both treatment and experiment duration ($p < 0.05$). Examining alpha (rarefaction) and beta diversity by NMDS showed that there were measureable differences between the filtered fractions and experimental time points (Figs. 5.5 and 5.6). As shown in Figure 5.4, there was no noticeable effect of low molecular weight DOC addition on community composition. Based on the rarefaction curves (Fig. 5.6), there seems to be more diversity in the unfiltered fraction versus the filtered, and samples taken at the same time are clustered together. The initial communities and those measured at 49 h in both river water fractions were more diverse than those measured towards the end of the experiment (141-506 h) (Fig. 5.6). We note that within both fractions, diversity decreases over the three-week experimental period (Fig. 5.6).

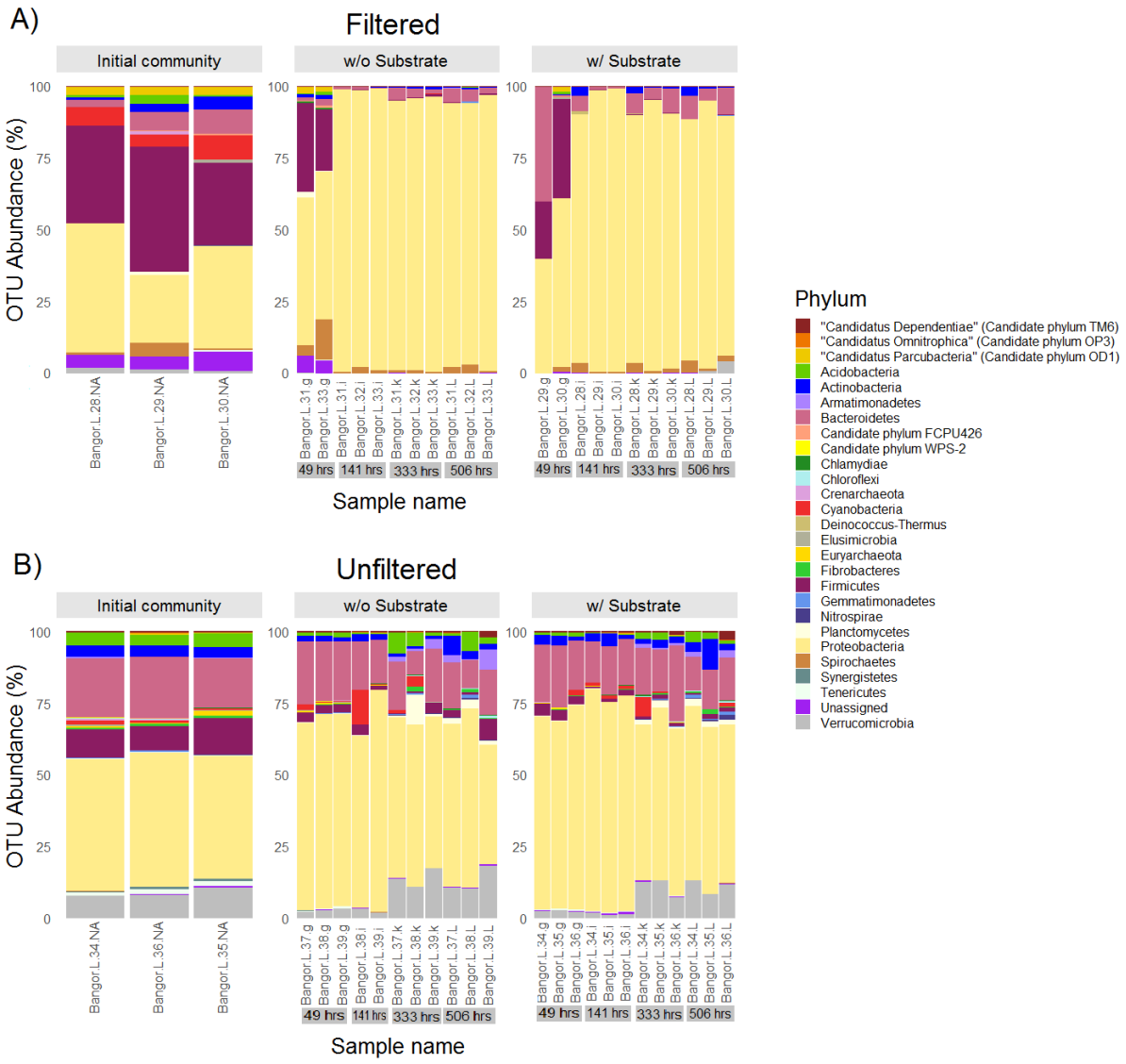


Figure 5.2- Phyla level taxa distribution of high abundance OTUs between treatments and substrate additions over the course of approximately three weeks (506 h). The top row is the (A) filtered treatment and the bottom is the (B) unfiltered treatment. The left column is the initial communities (n=3). The middle column shows the communities changing over the course of the three weeks w/o additional substrate versus the right column w/ additional substrates. Samples are duplicates or replicates dependent on nutrient amendment and time point (49 h, 141 h, 333 h, and 506 h) measured. Refer to Table S5.7 for further information regarding lower order taxa. Counts lower than 50 were removed.

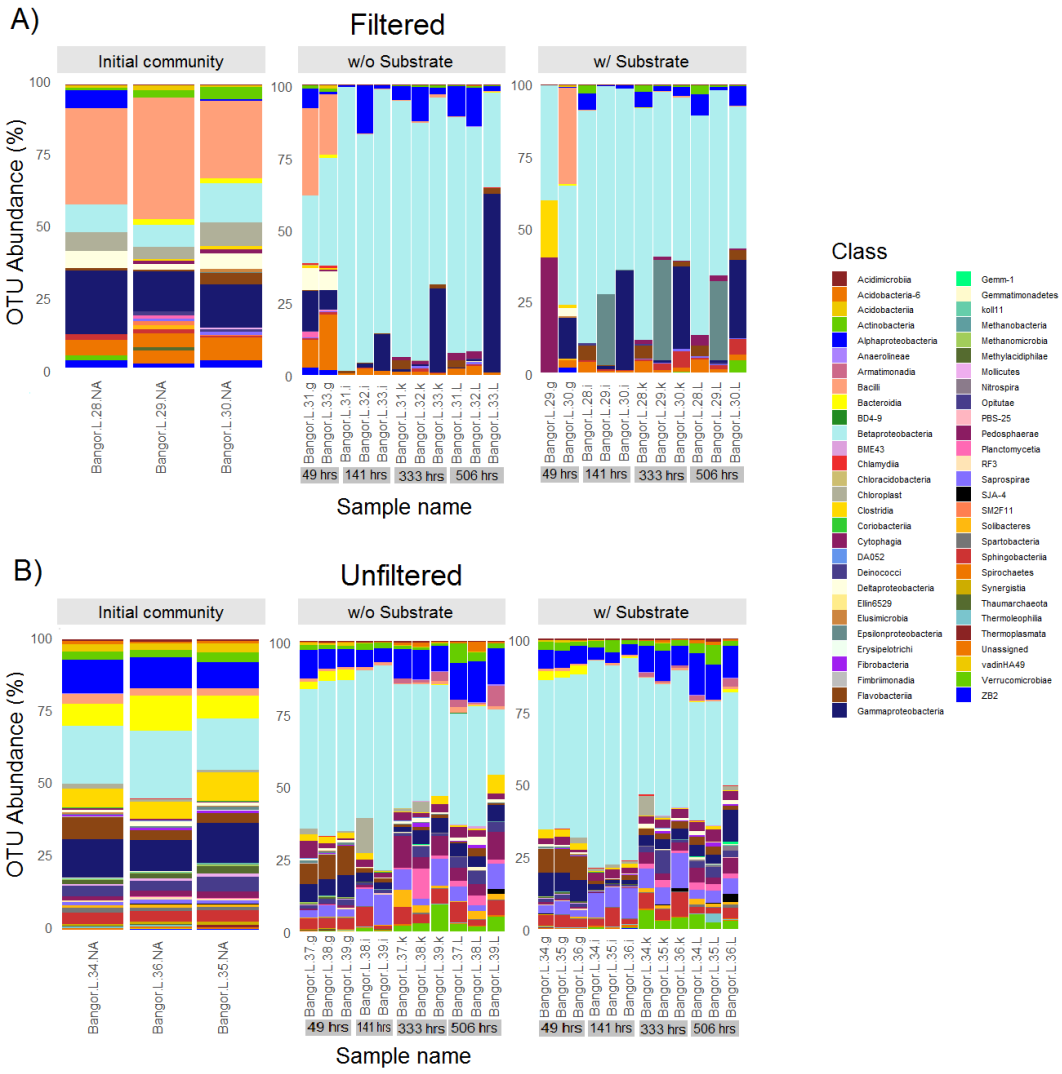


Figure 5.3- Class level taxa distribution of high abundance OTUs between treatments and substrate additions over the course of approximately three weeks (506 h). The top row is the (A) filtered treatment and the bottom is the (B) unfiltered treatment. The left column is the initial communities (n=3). The middle column shows the communities changing over the course of the three weeks w/o additional substrate versus the right column w/ additional substrates. Samples are duplicates or replicates dependent on nutrient amendment and time point (49 h, 141 h, 333 h, and 506 h) measured. Refer to figures 5.2, 5.4, and Table S5.7 for further information regarding high order taxa. Counts lower than 50 were removed.

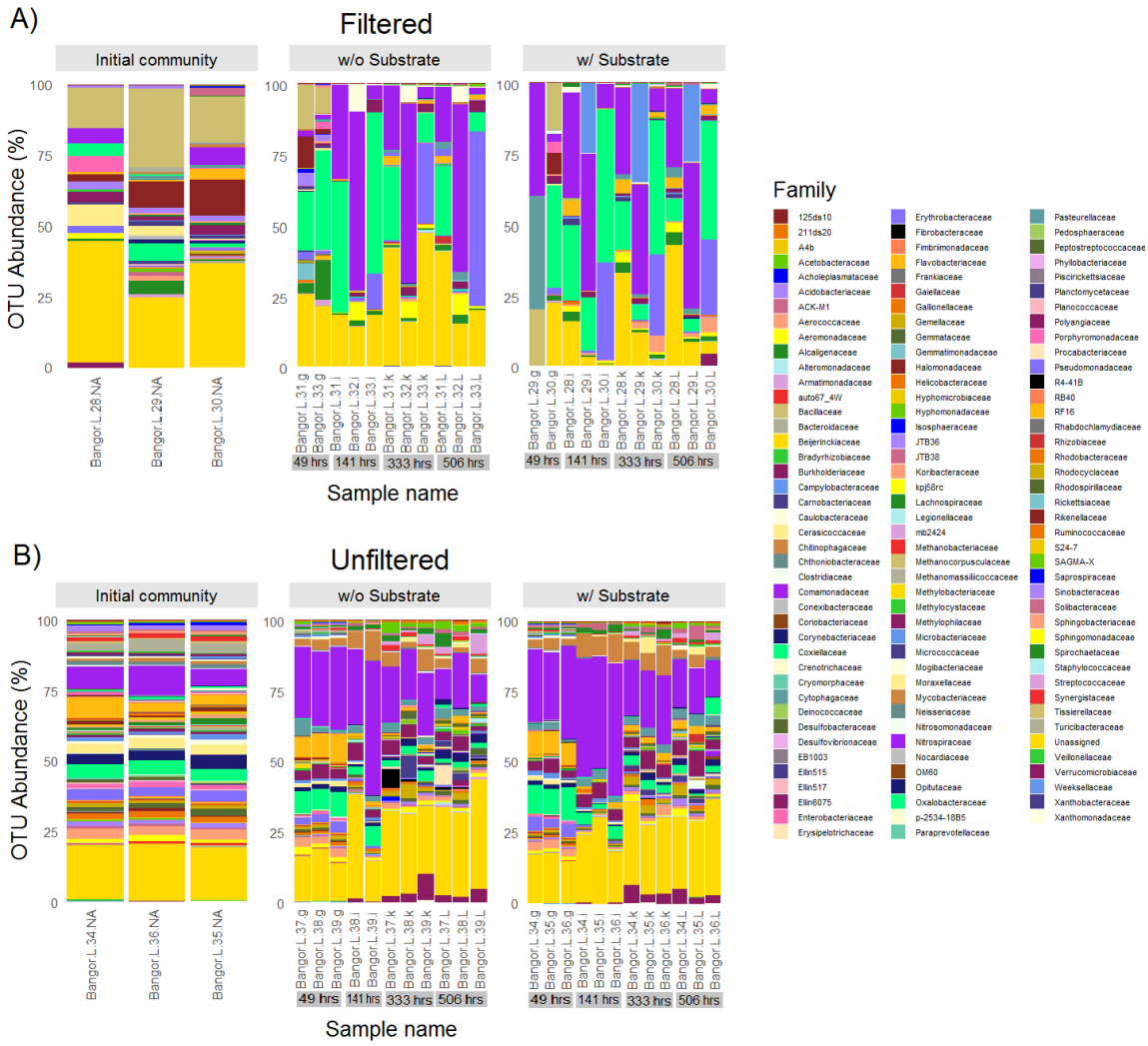


Figure 5.4- Family level taxa distribution of high abundance OTUs between treatments and substrate additions over the course of approximately three weeks (506 h). The top row is the (A) filtered treatment and the bottom is the (B) unfiltered treatment. The left column is the initial communities (n=3). The middle column shows the communities changing over the course of the three weeks w/o additional substrate versus the right column w/ additional substrates. Samples are duplicates or replicates dependent on nutrient amendment and time point (49 h, 141 h, 333 h, and 506 h) measured. Refer to Table S5.7 for further information regarding high order taxa. Counts lower than 50 were removed.

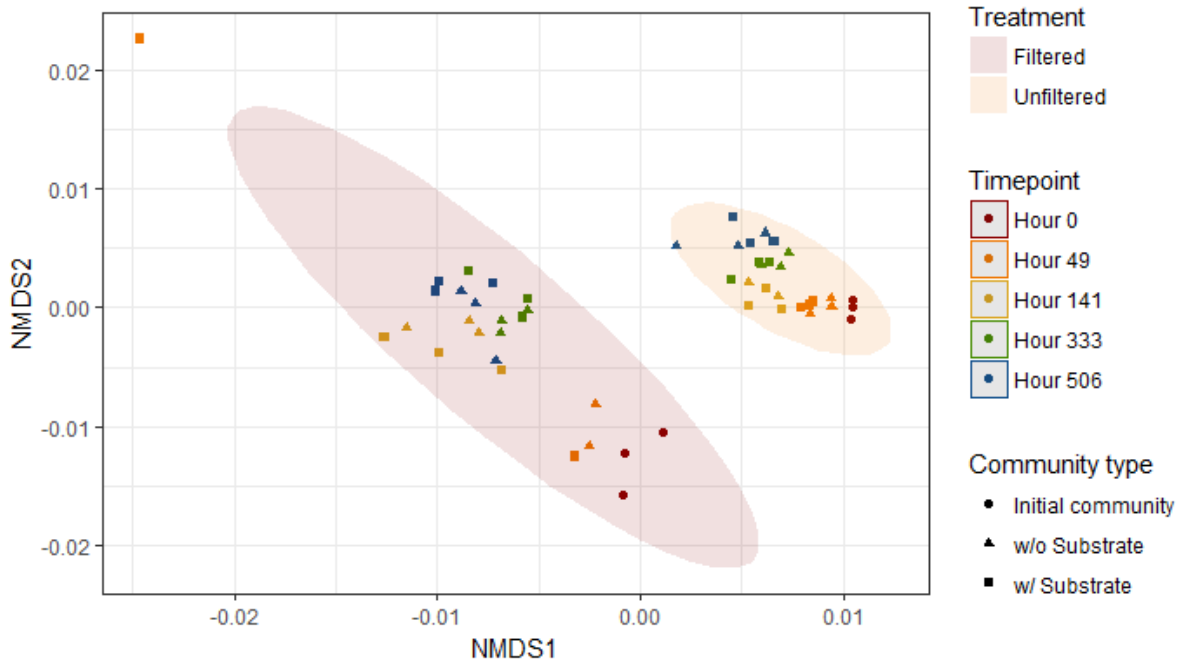


Figure 5.5- Beta diversity of both fractions at 49 h, 141 h, 333 h, and 506 h with and without substrate addition. NMDS using Bray Curtis calculation method (stress value = 0.094).

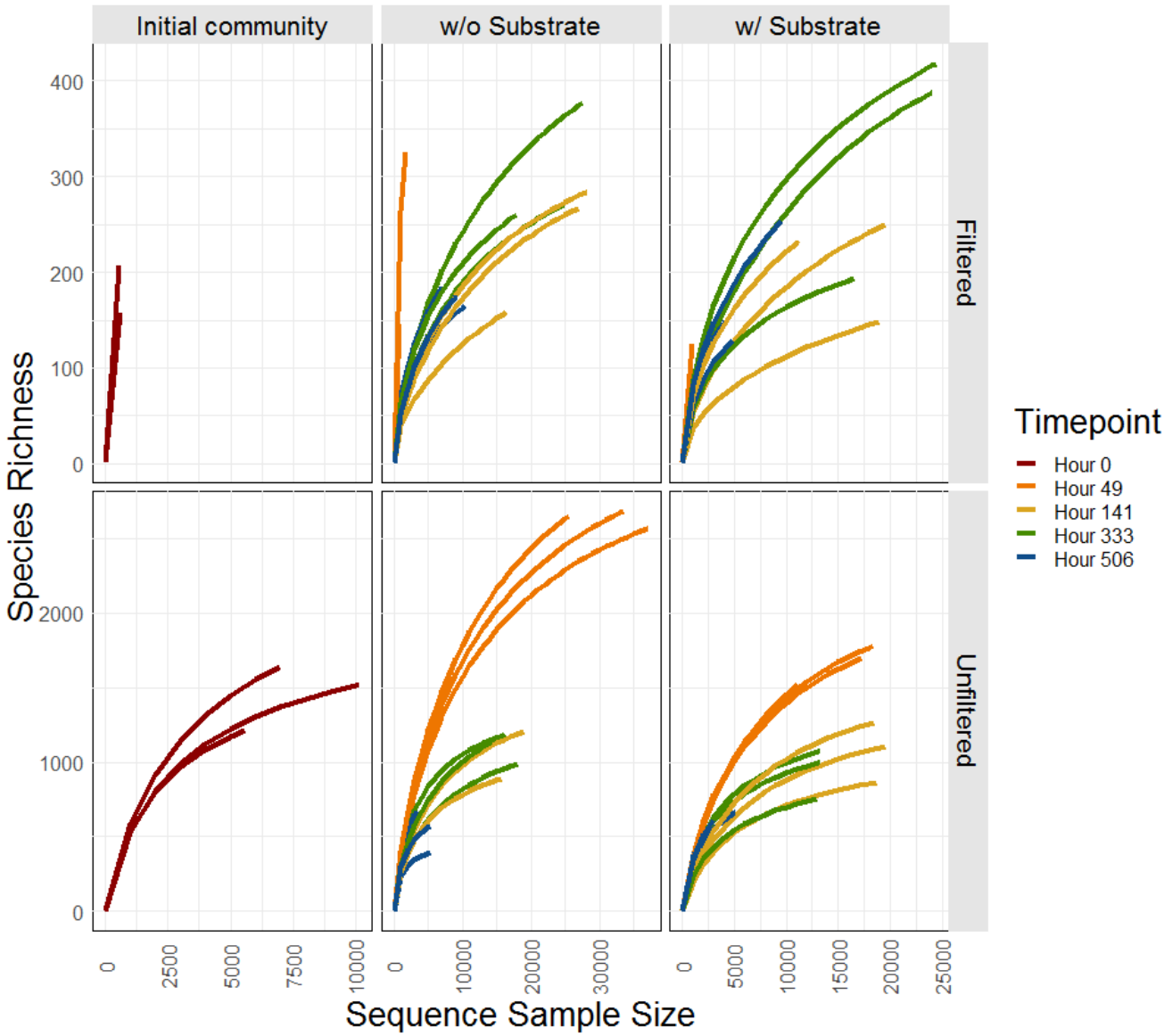


Figure 5.6- Rarefaction curves for all the samples and time points (49 h, 141 h, 333 h, and 506 h) illustrating the changes in diversity over the course of the three weeks (506 h). The top row is the filtered community and the bottom is the unfiltered community. The left column shows the initial community, the middle is the community from 49-506 h without substrate addition, and the right column is the community from 49-506 h with substrate addition.

5.3.3. GC-MS analysis of metabolites in river water

Sixteen out of the 24 metabolites from the three substrate types added to the river water were detected via GC-MS. This included the amino acids: alanine, aspartic acids, glycine, isoleucine, leucine, phenylalanine, proline, serine, threonine, tyrosine, and valine; the organic acids: citric acid and malic acid; and the sugars fructose, glucose, and sucrose. Our blank (negative control) showed no appreciable signs of metabolite loss during the experiment ($p > 0.05$). In the river water treatments, there was an overall decrease in metabolite concentration in the addition treatments in both fractions. T-tests between the 0 and 506 h time point indicated a significant difference in substrate concentration between the start and end of the incubation period ($p < 0.05$) (Table S5.3). The majority of the substrates were depleted within the first 49 h of the experiment in the unfiltered fraction (Fig. 5.7). In contrast, a lag phase was apparent in the filtered fraction across all substrate types, as consumption of all three substrate types did not occur until 49 h (Fig. 5.7).

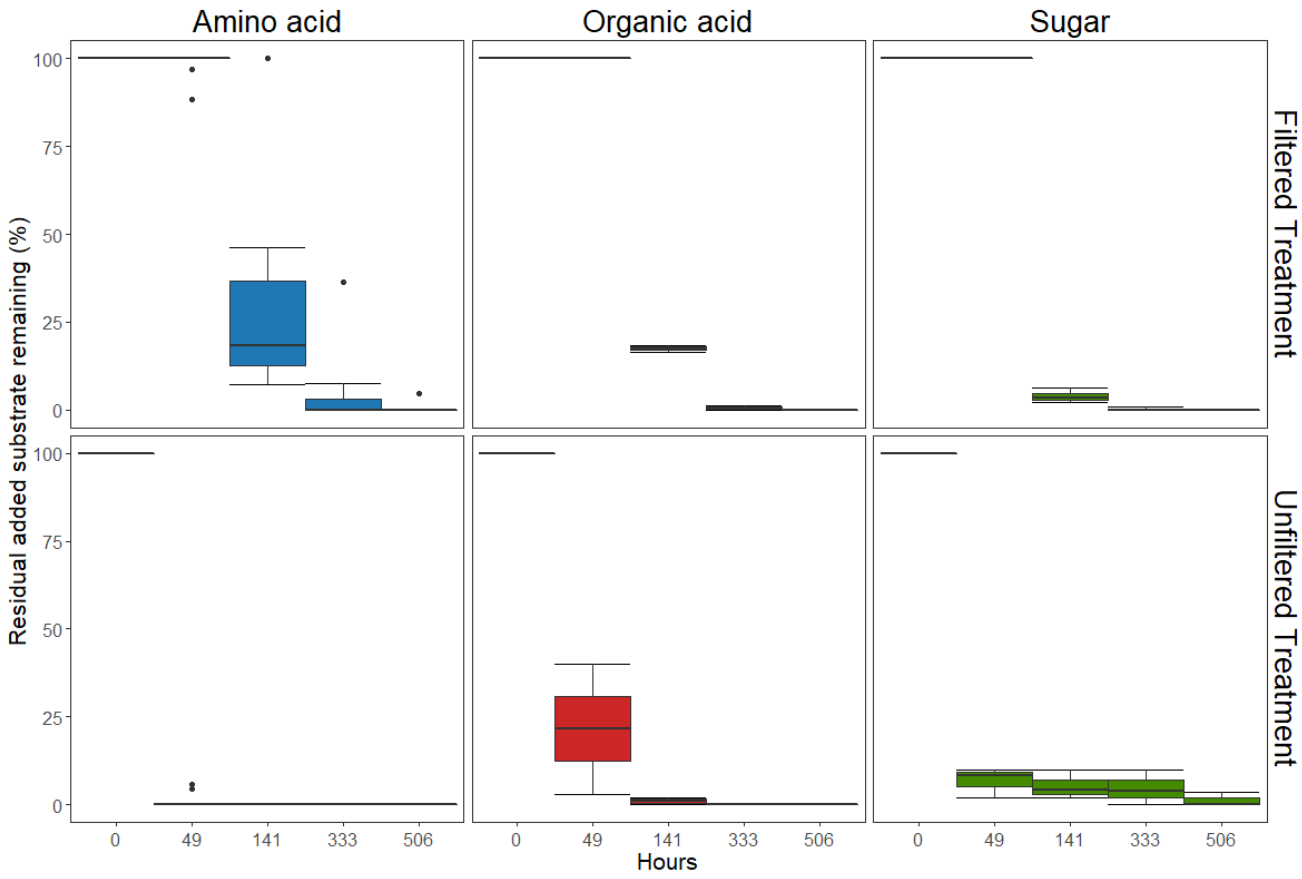


Figure 5.7- Boxplot of residual added substrate remaining (%). In total there are here 16 metabolites, which consist of 11 amino acids (left), 3 sugars (right), and 2 organic acids (middle) (n=3). Both fractions in the filtered (top) and unfiltered (bottom) were measured over the course of 506 h.

5.3.4. Shotgun sequencing: Assessment of COG categories of both fractions

The entire community of both fractions before and three weeks after substrate addition were examined using shotgun sequencing. Three time points selected (0, 141 and 506 h) were chosen to examine the evolution of the community in terms of protein encoding genes expressed/present. The number of contigs was variable ranging from 25,762 to 256,700 across all treatments, time points, and substrate amendments; with the unfiltered fraction containing overall more contigs than the filtered (Table S5.8). The number of contigs was not dependent on the duration of the experiment nor the addition of substrates (Table S5.8). Here we focus on the number of genes encoding for proteins that fall under the functional categories of clusters of orthologous groups (COGs). The number of individual COG assignments across all sample types ranged from 893 to 99,402 with a similar trend of the unfiltered fraction containing more COG assignments, yet no effect of substrate addition or duration was evident (Table S5.8). T-tests revealed that adding substrate had a greater effect on the distribution of functional categories within the filtered fraction than then unfiltered fraction ($p < 0.05$) (Table S5.4).

Examination of the COG assignments across both fractions and nutrient amendments over the course of the three weeks show some notable trends. Firstly, greatest number of COGs in both fractions across all time points and substrate addition fall under an unknown function (Fig. 5.8). Second, COGs that are affiliated with specific metabolic pathways (i.e. utilization of amino acids, carbohydrates, and lipids) did not vary in either fraction over the course of the three-week period regardless of substrate amendment (Figs. 5.8 and 5.9). Rather, the percentage of general COGs affiliated with energy production and conversion increased in both fractions with substrate addition (Figs 5.8 and 5.9). The other COGs in the filtered fraction were influenced by the presence of added substrate such as amino acid metabolism; coenzyme transport and metabolism; translation, ribosomal structure and

biogenesis; cell motility; inorganic ion transport; secondary metabolism; signal transduction; post-translational modification, protein turnover, and protein chaperones; and unknown functionalities (Fig.5.8 and Table S5.4). On the other hand, the COGs present in the unfiltered fraction were not influenced by the presence of nutrient amendment (Fig.5.8 and Table S5.4).

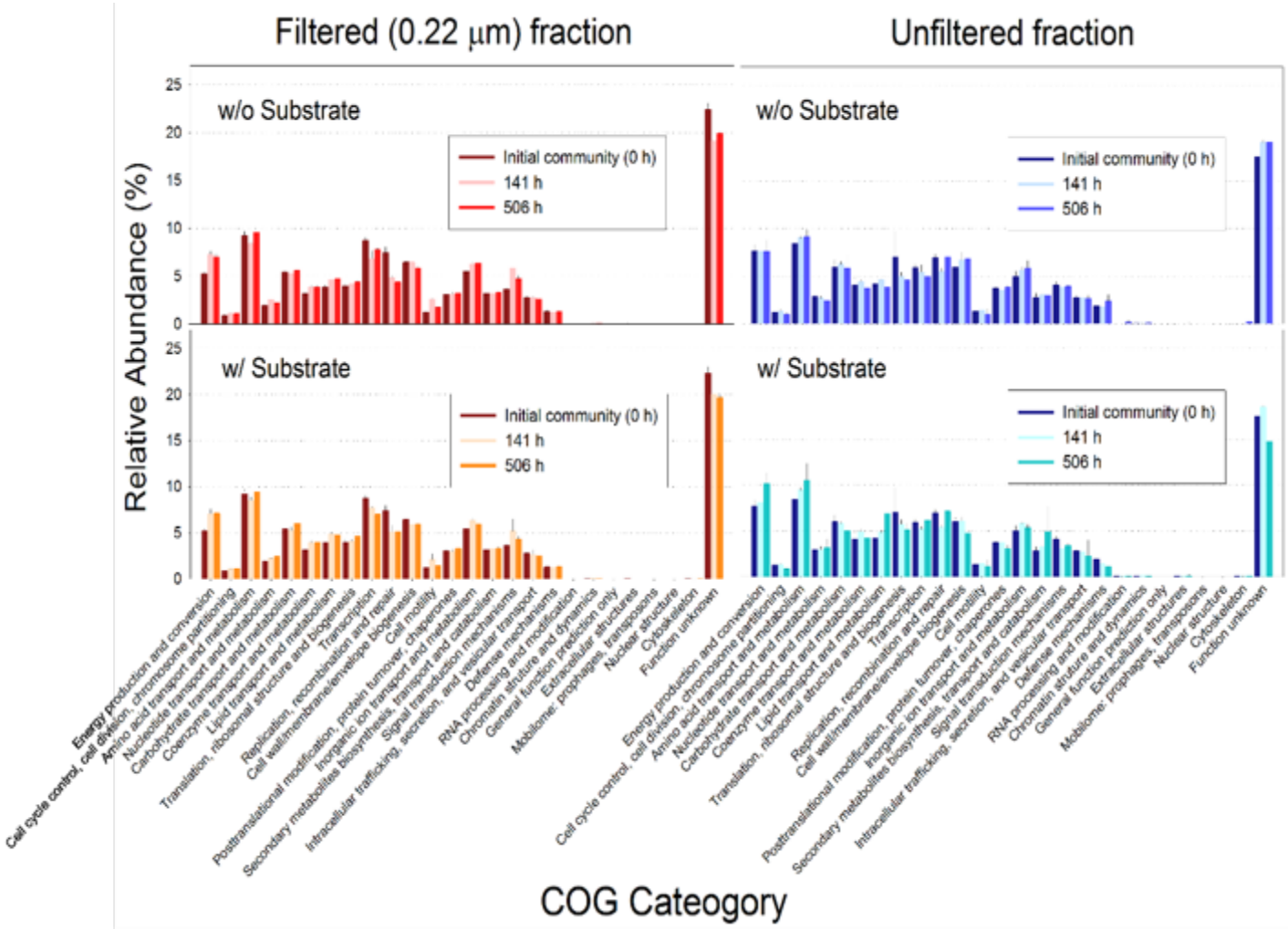


Figure 5.8- Distribution of COG-coding genes as per functional category. The relative abundance (%) of total gene count that falls under a COG category is compared between the filtered (warm colors) and unfiltered (cool colors) fraction over the course of 506 h with and without substrate amendments. Values represent means ± standard deviation (n = 2)

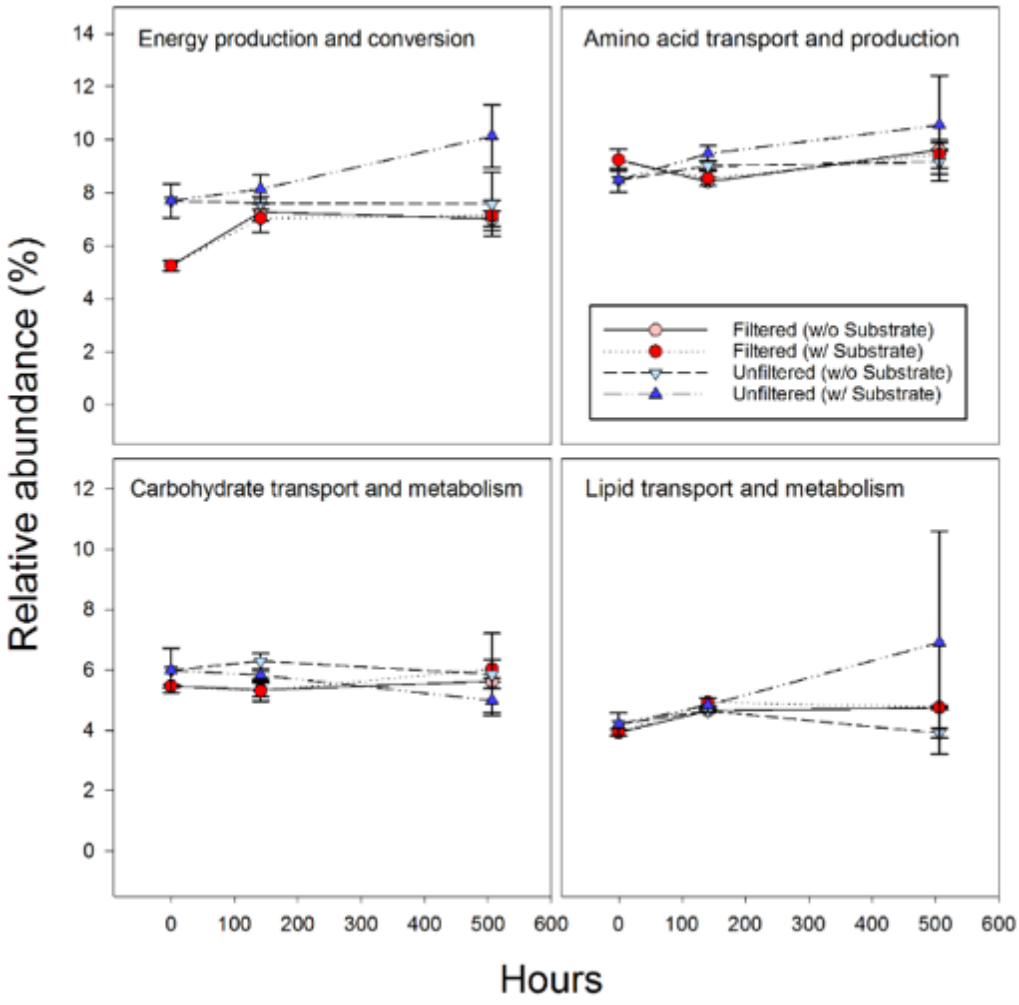


Figure 5.9- COG categories associated with DOM usage (energy production and conversion, amino acid transport and metabolism, carbohydrate transport and metabolism, and lipid transport and metabolism). The relative abundance (%) of total gene count that falls under a COG category is compared between the filtered (warm colors) and unfiltered (cool colors) fraction over the course of 506 h with and without substrate amendments. Values represent means ± standard deviation (n = 2).

5.4. Discussion

5.4.1. Taxonomy of microorganisms residing in the Conwy River

Firmicutes was one of the most abundant phyla in the initial phases of the experiment, especially in the filtered fraction (Figs. 5.2 and 5.3). It could be that the *Firmicutes* represented here are endospores (i.e. dormant forms) rather than metabolically active forms. Endospores are usually smaller (or the same size) as the original bacteria as in the case of *Bacillus subtilis* where the bacterium itself is 4-10 µm long and 0.25-1.0 µm in diameter but its endospore form is 0.89-1.53 µm long and 0.41 to 0.67 µm in diameter (Carrera et al., 2007; Yu et al., 2014). This fact, in conjunction with ^{14}C-labelled DOC measurements and the targeted metabolomics showing limited activity in the first 74 h, is strong evidence to support this (i.e. lag phase while spores germinate and become active). Another notable factor contributing to the noted decline in population in both fractions may stem from their anaerobic lifestyle. The constant exposure to oxygen, like that in the experiment, would be detrimental to their survival.

Spirochetes followed the same pattern as the *Firmicutes* in the filtered fraction. In the initial phase of the experiment, these bacteria were initially present but declined over time (Figs. 5.2 and 5.3). They are mostly motile free-living anaerobes and are widely distributed in a number of environments, from freshwater ponds to living within a range of eukaryotic hosts (Harwood and Canale-Parola, 1984). It should be noted, that the morphology of *Spirochaetes* is rather unique. Due to their inherent dimensions (3-500 µm long and 0.09 to 3 µm in diameter), it should not be able to fit through small pore sizes (Margulis et al., 1993) and should be found in large quantity in the unfiltered fraction. Yet their numbers were negligible in the unfiltered fraction but abundant in the filtered fraction. Evidence does show that *Spirochetes*, due to its morphology can easily squeeze through 0.22 µm filters relatively unharmed (Hahn, 2004; Wang et al., 2007, 2008). Similar to the *Firmicutes*, the constant

exposure to oxygen at regular intervals may be the reason for the noted decline in *Spirochetes*. This suggests that both these phyla present in the initial phases may not be metabolically active due to the sub-optimal growth conditions for these taxa.

Bacteriodetes and *Actinobacteria* were found across all the fractions, but they were much more prevalent in the unfiltered fraction (Figs. 5.2 and 5.3). However, their numbers in the filtered fraction declined sharply over the three week incubation. This may be due to the lack of high-molecular-weight polysaccharides in the samples, as they may prefer these substrates to the simple sugars, amino acids, and organic acids that were initially added (Thomas et al., 2011). It has been observed that large populations of *Bacteroidetes* in freshwater lakes resided in regions of the water column that have an accumulation of slowly degradable sinking macromolecules (Thomas et al., 2011). In addition, in marine systems, an influx of organic matter resulted in an increase of *Bacteroidetes* (Gooday, 2002; Thomas et al., 2011).

Actinobacteria and *Bacteroidetes* were present in low abundance in the filtered fraction. Two factors may be responsible for this. Firstly, some *Actinobacteria* are nano-sized and second there are inactive or senescent forms. In regards to the nano-sized organsims, it is known that in aquatic *Actinobacteria*, specifically Ac1, cell volumes can be less than 0.1 μm^3 (Jooste and Hugo, 1999; Pernthaler et al., 2001; Thomas et al., 2011; Ghai et al., 2013; Salcher et al., 2013). Although similar freshwater systems (oligotrophic prealpine lakes, etc) have Ac1 as a dominant taxa (Salcher et al., 2013), the Conwy river was dominated by *Microbacteriaceae*, *Micrococcaceae*, *Corynebacteriaceae* and ACK-M1. In addition, a prominent family in the Conwy river water was *Flavobacteriaceae*, which is the largest family in the phylum of *Bacteroidetes* with at least 90 genera (McBride, 2014). A notable quality to this family is the ability to utilize polysaccharides on the cell surface, i.e. can bind polysaccharides and transport oligomers via the outer membrane (McBride, 2014). The

morphology of individual species within this family, however, is highly variable. Usually, members are rod shaped with dimensions ranging from 0.3-0.6 μm in diameter and 1-10 μm long, and as they age may become spherical (Jooste and Hugo, 1999). Some, under specific growth conditions become filamentous and flexible (Jooste and Hugo, 1999). These dominant families of the *Bacteroidetes* found are common in freshwater systems, which also have a notable sporulation phase and or senescence (Jooste and Hugo, 1999; Hahn, 2004; McBride, 2014; Lewin et al., 2016; Chopyk et al., 2018). The lag phase suggests that the majority of these are senescent and or metabolically inactive.

"Candidatus Parcubacteria" (Candidate phylum OD1) was found to be exclusively in the initial phases (the first 49 h) of the filtered fraction, with their population declining afterwards. It has been postulated that they have reduced genomes (less than 1.5 Mbp), which can be an indication of a symbiotic lifestyle. For instance, lacking biosynthetic pathways like the synthesis of nucleic acids, vitamins and lipids, and mechanisms for DNA repair are commonplace for many symbiont species (Brown et al., 2015; Nelson and Stegen, 2015). Even with a streamlined genome, it is suggested that this group engage in sulfur cycling in aquatic sulfur-rich environments (Harris et al., 2004). A recent draft genome of "*Ca.* Parcubacteria" suggests that there is a potential to metabolize organic compounds (glucose, ribose, acetate) (Castelle et al., 2017). "*Ca.* Parcubacteria" were originally thought to be found in oligotrophic environments, yet recent research has shown these organisms also reside in anoxic ground water and river water (Harris et al., 2004; Luef et al., 2015; Proctor et al., 2018). The evidence collected here also confirms previous findings in that "*Ca.* Parcubacteria" are not limited to oligotrophic environments.

Acidobacteria and *Cyanobacteria* were notable members in the initial community of the filtered fraction and greatly decreased within that same fraction throughout the experiment. *Cyanobacteria* are ubiquitous throughout different environments and their

population decline may be attributed to limited exposure to light. *Acidobacteria* are another widely distributed taxa, which can be found in soils, hot springs, oceans, and in contaminated (uranium) areas (Quaiser et al., 2003; Barns et al., 2007; Kielak et al., 2016). They can effectively degrade a wide range of polysaccharides and can use nitrite as an nitrogen source (Kielak et al., 2016). *Acidobacteria* were not limited to the filtered fraction. This could explain why there were higher levels of nitrate present in the initial phases of the filtered sample, which was subsequently used up by other members of the population. The unfiltered fraction may have less microorganisms competing for nitrite.

Verrucomicrobia and *Armatimonadetes* were notable members of the unfiltered community, especially in the end stages of the experiment. It was initially thought that these bacteria, although ubiquitous, appeared in low frequencies across biomes (approximately less than 7% of bacterial sequences retrieved) (Janssen, 2006). Yet subsequent studies have revealed that they are more prevalent than previously thought. *Verrucomicriobia* were estimated to make up 35% of all bacterial sequences in grassland and in prairie soils (Bergmann et al., 2011). In freshwater lakes, *Verrucomicrobia* abundance ranges from 1.7-41.7% of all bacterial sequences (Chiang et al., 2018). *Verrucomicrobia* are sensitive to changes in DOM concentration. For instance, microbial communities derived from seawater were amended with micromolar concentrations (13 μM addition) of diatom-derived DOM; *Verrucomicriobia* (along with *Bacteriodetes* and *Alphaproteobacteria*) became more abundant in the DOM samples after the 15 day period (Landa et al., 2013). Yet our experiment showed no marked difference in the communities with added substrate.

The phyla *Armatimondetes*, formerly known as Candidate phylum OP10, are aerobic Gram-negative aerobic bacteria preferring oligotrophic environments and possess the ability to degrade polysaccharides (Tamaki et al., 2011; Lee et al., 2014). Although their preference for oligotrophy is well documented, *Armatimondetes* make up 37% of rRNA sequences

retrieved from temperate soils. This phylum has also been found from a plethora of other environments such as; human skin (17%), wastewater treatment facilities (12%), freshwater and sediments (7%), biofilms (6%), thermal springs (6%), plants and animal microbiomes (6%), marine sediments (3%), ice and snow (2%), mine tailings (1%), fossil fuel deposits (1%), and dust/atmosphere (1%) (Bond et al., 1995; Hugenholtz et al., 1998; Tamaki et al., 2011; Dunfield et al., 2012; Lee et al., 2014). There are only three isolates currently known: *Armatimonas rosea*, *Chthonomonas calidirosea* and *Fibriimonas ginsengisoli* (Lee et al., 2014). The presence of microorganisms of this phylum at the 506 h point only in the unfiltered fraction can signify the presence polysaccharides favourable for growth (Figs. 5.2 and 5.3). It should be noted that this phylum was not detected in the filtered fraction, as it is specultated that their cell body ranges between 0.5-5 µm in geothermal systems (Dunfield et al., 2012).

Finally, *Proteobacteria*, one of the most characterized bacterial phylum, was not exclusive to the unfiltered fraction (Figs. 5.2 and 5.3). We also note that not all currently known *Proteobacteria* are large cells. Most famously, "*Ca.* Pelagibacter ubique," one of the smallest free-living cells, falls under the phyla *Proteobacteria.* Members of this phylum overtook many of the sequences that were initially present in the filtered fraction after 141 h (Figs. 5.2 and 5.3). This, combined with Figures 5.1 and 5.6, is strong evidence to suggest that *Proteobacteria* are largely responsible for DOC utilization in the filtered fraction, not *Firmicutes* or *Spirochaetes*. Other aquatic systems also suggest that *Proteobacteria* were primarily responsible for LMW DOC usage. For instance when examining seawater, obtained from the Mediterranean and Baltic sea, *Gammaproteobacteria* thrived on allochthonous carbon sources (Gómez-Consarnau et al., 2012). Another study within the Mediterranean (coastal waters) suggested that *Alphaprotebacteria* were the most active in terms of glucose

and amino acid uptake while *Gammaproteobacteria* preferred amino acids (Alonso-Sáez and Gasol, 2007).

The change in the microbial community may be the result of "bottle effect", where the surface area to volume ratio of a growing chamber or vessel alters the composition of microorganisms (Hammes et al., 2010). There are contradicting reports when it comes to the bottle effect. One study suggested that for short term incubations (less than 5 days), the bottle effect is negligible hence microbial communities do not change (Fogg and Calvario-Martinez, 1989; Hammes et al., 2010). On the other hand, another study examined the picoplanktonic communities of oligotrophic marine water over a 24 h period and observed a shift from autotrophs to heterotrophs (Calvo-Díaz et al., 2011). Other accounts suggest that bottle size is a determinate factor (Fogg and Calvario-Martinez, 1989). The consensus is that bottle effect is something to account for, depending on the conditions (such as initial sample, ambient light, etc). For this experiment, the bottle effect may be the main driver for change especially in the filtered fraction because of the over proliferation of *Proteobacteria* and the decrease in *Firmicutes* and *Cyanobacteria* (Figs. 5.2-5.4).

5.4.2. Utilization of DOC and fractions of the Conwy River

Low molecular weight compounds (amino acids, organic acids and sugars) were chosen for this experiment because (1) they are the major constituents of DOM and (2) are a ubiquitous source of nutrients for many heterotrophic freshwater species. The river water in previous studies prior to addition of substrates showed a significant amount of DOC present (Emmett et al., 2016; Brailsford et al., 2017, 2019b). However, it is hard to determine what form that C is in because the definition of DOM also includes particulates that have the ability to pass through ultra-small filters (filter sizes <0.45 μm) (Brailsford et al., 2017). Therefore, specific substrates were used to determine C utilization by the intrinsic microbial community.

Both the filtered and unfiltered fractions possessed the ability to use low molecular weight compounds (Figs. 5.1, 5.7-5.9). The majority of the activity occurred in the first 10 d of the experiment. As the three weeks progressed, there was a clear distinction between the filtered and unfiltered fraction in terms of DOC usage. Most notably the lag phase only occurred in the filtered fraction and was independent of substrate/addition type (Figs. 5.1, 5.7-5.9).

In the initial investigation of Brailsford et al. (2017), DOC (in the form of glucose and amino acids) usage was dependent on location within the catchment. DOC usage in the 0.22 µm-filtered water of main Conwy river (at the sample site) closely resembles 0.22 µm filtered water the Migneint sub-catchment (land use type: upland peat bogs). In the Migneint sub-catchment, there was minimal metabolic activity 24-48 h into the experiment, whereas the other location the Hiraethyln sub-catchment (land use type: agriculture) showed immediate uptake of both glucose and amino acids (Brailsford et al., 2017). Further investigation comparing mesotrophic (i.e. moderate level of nutrient productivity) rivers to the oligotrophic (i.e. low level of nutrient productivity) main river, revealed that unfiltered oligotrophic water and sediment, and mesotrophic sediment (especially at higher µM concentrations) exhibited low levels of LMW DOC metabolic activity (µM concentrations) within the first 72 h (Brailsford et al., 2019a). The DOC metabolic activity of the filtered fraction within this current study resembled the behavior of the oligotrophic conditions observed in the previous study (Brailsford et al., 2019a). The lag phase is not unique to the Conwy catchment. This phenomenon has also been observed with marine systems (e.g. Mediterranean and Baltic Sea), where bacterial growth occurred within 1-3 days after µM additions of LMW DOC (Gómez-Consarnau et al., 2012) and also in deep subsoils from the Conwy catchment (de Sosa et al., 2018).

We should note that the marine system (especially pelagic areas) and major stretches of the Conwy river are classed as oligotophic (Alonso-Sáez and Gasol, 2007; Gómez-Consarnau et al., 2012). Organisms adapted to oligotrophic environments decrease in size to maximize surface area for nutrient consumption. However, when nutrients are introduced into the system, copiotrophs (i.e. microorganisms that prefer high nutrient, eutrophic environments) overproliferate and cells increase in size (Ghuneim et al., 2018). Therefore, the lag phase observed is probably the result of two factors. Firstly, there was a low abundance of microorganisms in the filtered fraction, due to removing larger microorganisms such as protists and other heterotrophic prokaryotes via filtration. Secondly, the microorganisms left over are metabollicaly inactive or possible oligotrophs, and once DOC is introduced causes the oligotrophs to decrease in number and copiotrophs to proliferate.

Neither the filtered fraction nor unfiltered fraction showed a preference of substrate type, hence suggesting that microorganisms in the Conwy River are generalists in regards to LMW DOC. This is evident as there was no discernible difference in rate of consumption in the radiolabeled experiments nor the COG assignments from shotgun sequences (Figs. 5.1, 5.8-5.9). In terms of COG assignments, both fractions have a similar distribution of COGs related to LMW DOC cycling (Fig. 5.9). As substrate addition also increased the percentage of COGs associated with functionalities related to energy production/conversion further supports this generalist view (Figs 5.8 and 5.9). The targeted metabolomics may also suggest that the filtered fraction prefers organic acids and sugars to amino acids, as there is less variability in the sugars and organic acids (Fig. 5.7). This could be due to two factors. First, much of the metabolic pathways for organic acids and sugars are ubiquitous as biochemical pathways such as glycolysis and the tricarboxylic acid cycle (TCA) are present across many heterotrophic species. Secondly, in order for amino acids to be metabolized they first need to be converted to sugars via oxidative deamination whereas sugars and organic acids are

instantly utilized via glycolysis and the TCA cycle. As for the unfiltered community, there is a notable preference for amino acids, indicating that many of the members posess the machinery to metabolize amino acids. The preference of one substrate over another and lag phase may be attributed to the overall larger population of metabolically active cells in the unfiltered fraction as illustrated in both the 16S rRNA amplicon sequencing and shotgun sequencing (Tables S5.6 and S5.8).

5.5. Conclusions

There is strong evidence of metabolically active forms residing in the 0.22 μm filtered river water fraction. Substrate utilization, CO_2 production, and biomass incorporation was observed in both radiolabeled and targeted metabolomics experiment within the filtered fraction. The addition of nano-molar concentrations of LMW DOC does not alter the structure of the microbial community in either fraction, which may be largely in part due to the "bottle effect". Previous research suggests that changes in community structure requires micro-molar concentrations (i.e. enough C to support growth) (Gómez-Consarnau et al., 2012; Landa et al., 2013). Bacterial and archaeal numbers change over time according to several different factors: (1) whether in the filtered or unfiltered fraction, (2) their cell life cycle/metabolic state, and (3) the metabolic preferences/optimal growth conditions. The most prominent phyla in both fractions were *Firmicutes, Proteobacteria, Bacteriodetes, Actinobacteria*, and *Acidobacteria*. The filtered fraction contained many more *Firmicutes* and *Spirochetes* that may not be metabolically active as these were quickly overtaken by *Proteobacteria*. It can be concluded that *Proteobacteria* are mainly responsible for the utilization of LMW DOC in the Conwy River within the filtered fraction and the community as a whole. Although the makeup of the members within the microbial communities were not greatly altered by the additional nutrients, COG functional category of energy production and conversion did, across both fractions. The percentage of energy production/conversion COGs

increased over the course of the three weeks in both fractions (i.e. gene expression was influenced by nM concentrations of LMW DOC).

5.6. Acknowledgements and contributions

Sample collection and filtration was done by Miles Marshall. Robert Brown assisted with setup and data collection in the initial phases of the experiment. Lydia-Ann J. Ghuneim co-conceived the research, conducted the experiments, analysis, data interpretation, and wrote the manuscript. Lydia-Ann J. Ghuneim, Tatyana N. Chernikova, and Evgenii A. Lunev prepared 16S rRNA amplicon sequencing libraries. Aleksei A. Korzhenkov performed 16S rRNA library sequencing. Stepan V. Toshchakov performed 16S rRNA data analysis including demultiplexing, OTU generation and taxonomy assignment. David Rojo and Coral Barbas did masspectrometry analysis of metabolites. Manuel Ferrer carried out the preparation of samples for target metabolomics analysis and metabolomics data analysis. Marco A. Distaso prepared libraries and ran Illumina sequencing. Rafael Bargiela processed sequencing data and contributed to data visualisation. David L. Jones, Peter N. Golyshin, and Olga V. Golyshina co-conceived the research plan and provided significant revisions to the manuscript as well as insight into data interpretation.

This work was carried out under the DOMAINE project, which is funded by the UK Natural Environment Research Council (NERC) (large grant NE/K010689/1). DLJ, OVG and PNG acknowledge the support of the Centre for Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. David L. Jones and Peter N. Golyshin thank Natural Environment Research Council (NERC) for funding the project ‘Plastic Vectors’ (NE/S004548/1). 16S rRNA sequencing and work of Aleksei A. Korzhenkov was supported by a grant from Ministry of Science and Higher Education of Russian Federation allocated to the Kurchatov Center for Genome Research (grant 075-15-2019-1659). Work of Stepan V. Toshchakov was supported by Ministry of

Science and Higher Education within the State assignment of FRC ‘Fundamentals of Biotechnology’ RAS.

5.7. References

Alonso-Sáez, L., and Gasol, J. M. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean coastal waters. *Appl. Environ. Microbiol.* 73, 3528–3535. doi:10.1128/AEM.02627-06.

Andrews, S. (2010). FastQC - A quality control tool for high throughput sequence data. *Babraham Bioinforma.* doi:citeulike-article-id:11583827.

Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. Available at: https://cran.r-project.org/package=gridExtra.

Barns, S. M., Cain, E. C., Sommerville, L., and Kuske, C. R. (2007). Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl. Environ. Microbiol.* 73, 3113–3116. doi:10.1128/AEM.02012-06.

Beman, J. M., Arrigo, K. R., and Matson, P. A. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434, 211–214. doi:10.1038/nature03370.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, G., Walters, W. A., et al. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol. Biochem.* 43, 1450–1455. doi:10.1016/j.soilbio.2011.03.012.The.

Bond, P. L., Hugenholtz, P., Keller, J., and Blackall, L. L. (1995). Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl. Environ. Microbiol.* 61, 1910–1916.

Brailsford, F. L., Glanville, H. C., Golyshin, P. N., Johnes, P. J., Yates, C. A., and Jones, D.

L. (2019a). Microbial uptake kinetics of dissolved organic carbon (DOC) compound groups from river water and sediments. *Sci. Rep.* 9, 1–11. doi:10.1038/s41598-019-47749-6.

Brailsford, F. L., Glanville, H. C., Golyshin, P. N., Marshall, M. R., Lloyd, C. E., Johnes, P. J., et al. (2019b). Nutrient enrichment induces a shift in dissolved organic carbon (DOC) metabolism in oligotrophic freshwater sediments. *Sci. Total Environ.* 690, 1131–1139. doi:10.1016/J.SCITOTENV.2019.07.054.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208–211. doi:10.1038/nature14486.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Calvo-Díaz, A., Ídaz-Pérez, L., Suárez, L. Á., Morán, X. A. G., Teira, E., and Marañón, E. (2011). Decrease in the autotrophic-to-heterotrophic biomass ratio of picoplankton in oligotrophic marine waters due to bottle enclosure. *Appl. Environ. Microbiol.* 77, 5739–5746. doi:10.1128/AEM.00066-11.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–6. doi:10.1038/nmeth.f.303.

Carrera, M., Zandomeni, R. O., Fitzgibbon, J., and Sagripanti, J. L. (2007). Difference

between the spore sizes of Bacillus anthracis and other Bacillus species. *J. Appl. Microbiol.* 102, 303–312. doi:10.1111/j.1365-2672.2006.03111.x.

Castelle, C. J., Brown, C. T., Thomas, B. C., Williams, K. H., and Banfield, J. F. (2017). Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. *Sci. Rep.* 7, 1–12. doi:10.1038/srep40101.

Chiang, E., Schmidt, M. L., Berry, M. A., Biddanda, B. A., Burtner, A., Johengen, T. H., et al. (2018). Verrucomicrobia are prevalent in north- temperate freshwater lakes and display class- level preferences between lake habitats. *PLoS One* 13, 1–20. doi:10.1371/journal.pone.0195112.

Chopyk, J., Allard, S., Nasko, D. J., Bui, A., Mongodin, E. F., and Sapkota, A. R. (2018). Agricultural Freshwater Pond Supports Diverse and Dynamic Bacterial and Viral Populations. *Front. Microbiol.* 9, 792. doi:10.3389/fmicb.2018.00792.

Dang, H., and Lovell, C. R. (2016). Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* 80, 91–138. doi:10.1128/MMBR.00037-15.Address.

de Sosa, L. L., Glanville, H. C., Marshall, M. R., Schnepf, A., Cooper, D. M., Hill, P. W., et al. (2018). Stoichiometric constraints on the microbial processing of carbon with soil depth along a riparian hillslope. *Biol. Fertil. Soils* 54, 949–963. doi:10.1007/s00374-018-1317-2.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Dunfield, P. F., Tamas, I., Lee, K. C., Morgan, X. C., McDonald, I. R., and Stott, M. B. (2012). Electing a candidate: A speculative history of the bacterial phylum OP10.

Environ. Microbiol. 14, 3069–3080. doi:10.1111/j.1462-2920.2012.02742.x.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., et al. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2, 6. doi:10.1186/2049-2618-2-6.

Fenchel, T. (2008). The microbial loop - 25 years later. *J. Exp. Mar. Bio. Ecol.* 366, 99–103. doi:10.1016/j.jembe.2008.07.013.

Findlay, S. E. G., Sinsabaugh, R. L., Sobczak, W. V., and Hoostal, M. (2003). Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter. *Limnol. Oceanogr.* 48, 1608–1617. doi:10.4319/lo.2003.48.4.1608.

Fogg, G. E., and Calvario-Martinez, O. (1989). Effects of bottle size in determinations of primary productivity by phytoplankton. *Hydrobiologia* 173, 89–94. doi:10.1007/BF00015518.

Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2013). Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* 3, 2471. doi:10.1038/srep02471.

Ghuneim, L.-A. J., Jones, D. L., Golyshin, P. N., and Golyshina, O. V (2018). Nano-Sized and Filterable Bacteria and Archaea: Biodiversity and Function. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.01971.

Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M., and Pinhassi, J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds.

Environ. Microbiol. 14, 2361–2378. doi:10.1111/j.1462-2920.2012.02804.x.

Gooday, A. J. (2002). Biological Responses to Seasonally Varying Fluxes of Organic Matter to the Ocean Floor: A Review. *J. Oceanogr.* 58, 305–332. doi:10.1023/A:1015865826379.

Hahn, M. W. (2004). Broad diversity of viable bacteria in "sterile" (0.2 μm) filtered water. *Res. Microbiol.* 155, 688–691. doi:10.1016/j.resmic.2004.05.003.

Hammes, F., Vital, M., and Egli, T. (2010). Critical evaluation of the volumetric "bottle effect" on microbial batch growth. *Appl. Environ. Microbiol.* 76, 1278–1281. doi:10.1128/AEM.01914-09.

Harris, J. K., Kelley, S. T., and Pace, N. R. (2004). New Perspective on Uncultured Bacterial Phylogenetic Division OP11. *Appl. Environ. Microbiol.* 70, 845–849. doi:10.1128/AEM.70.2.845-849.2004.

Harwood, C. S., and Canale-Parola, E. (1984). *Ecology of Spirochetes*. doi:10.1146/annurev.mi.38.100184.001113.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi:10.1093/nar/gkv1248.

Hugenholtz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R. (1998). Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring. 180, 366–376.

Janssen, P. H. (2006). Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes MINIREVIEWS Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Appl. Environ. Microbiol.* 72, 1719–1728.

doi:10.1128/AEM.72.3.1719.

Jones, T. G., Evans, C. D., Jones, D. L., Hill, P. W., and Freeman, C. (2016). Transformations in DOC along a source to sea continuum; impacts of photo-degradation, biological processes and mixing. *Aquat. Sci.* 78, 433–446. doi:10.1007/s00027-015-0461-0.

Jooste, P. J., and Hugo, C. J. (1999). The taxonomy, ecology and cultivation of bacterial genera belonging to the family Flavobacteriaceae. *Int. J. Food Microbiol.* 53, 81–94. doi:10.1016/S0168-1605(99)00162-2.

Kandlikar, G. (2018). ranacapa: Utility Functions For Downstream Anacapa Analyses in R.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of Acidobacteria: Moving beyond genes and genomes. *Front. Microbiol.* 7, 1–16. doi:10.3389/fmicb.2016.00744.

Kirchman, D. L., Dittel, A. I., Findlay, S. E. G., and Fischer, D. (2004). Changes in bacterial activity and community structure in response to dissolved organic matter in the Hudson River, New York. *Aquat. Microb. Ecol.* 35, 243–257. doi:10.3354/ame035243.

Landa, M., Cottrell, M. T., Kirchman, D. L., Blain, S., and Obernosterer, I. (2013). Changes in bacterial diversity in response to dissolved organic matter supply in a continuous culture experiment. *Aquat. Microb. Ecol.* 69, 157–168. doi:10.3354/ame01632.

Lee, K. C. Y., Dunfield, P. F., and Stott, M. B. (2014). "The Phylum Armatimonadetes," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 447–458. doi:10.1007/978-3-642-38954-2_388.

Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et

al. (2016). Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu. Rev. Microbiol.* 70, 235–54. doi:10.1146/annurev-micro-102215-095748.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi:10.1038/ncomms7372.

Malmstrom, R. R., Cottrell, M. T., Elifantz, H., Kirchman, D. L., Kiene, R. P., Yu, L., et al. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean Diversity and abunda. *Appl Env. Microbiol* 71, 2979–86. doi:10.1128/AEM.71.6.2979.

Margulis, L., Ashen, J. B., Sole, M., and Guerrero, R. (1993). Composite, large spirochetes from microbial mats: spirochete structure review. *Proc Natl Acad Sci U S A* 90, 6966–6970. doi:10.1073/pnas.90.15.6966.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12.

McBride, M. J. (2014). "The Family Flavobacteriaceae," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 643–676. doi:10.1007/978-3-642-38954-2_130.

McMurdie, P. J., Holmes, S., with contributions from Gregory Jordan, and Chamberlain, S. (2018). phyloseq: Handling and analysis of high-throughput microbiome census data.

Available at: http://dx.plos.org/10.1371/journal.pone.0061217.

Meyer, J. L. (1994). The microbial loop in flowing waters. *Microb. Ecol.* 28, 195–199. doi:10.1007/BF00166808.

Morales, M., R Core Team, Community, R. listerv, and Murdoch, D. (2017). sciplot: Scientific Graphing Functions for Factorial Designs. Available at: https://cran.r-project.org/package=sciplot.

Nelson, W. C., and Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* 6, 1–14. doi:10.3389/fmicb.2015.00713.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). vegan: Community Ecology Package. Available at: https://cran.r-project.org/package=vegan.

Panikov, N. S. (2005). Contribution of Nanosized Bacteria to the total biomass and activity of a soil microbial community. *Adv. Appl. Microbiol.* 57, 245–296.

Pernthaler, J., Posch, T., Šimek, K., Vrba, J., Pernthaler, A., Glöckner, F. O., et al. (2001). Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. *Appl. Environ. Microbiol.* 67, 2145–2155. doi:10.1128/AEM.67.5.2145.

Proctor, C. R., Besmer, M. D., Langenegger, T., Beck, K., Walser, J.-C., Ackermann, M., et al. (2018). Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J.* doi:10.1038/s41396-018-0070-8.

Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., Eck, J., et al. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. *Mol. Microbiol.* 50, 563–575.

doi:10.1046/j.1365-2958.2003.03707.x.

R Core Team (2017). R: A language and environment for statistical computing. Available at: https://www.r-project.org/.

Rinta-Kanto, J. M., Sun, S., Sharma, S., Kiene, R. P., and Moran, M. A. (2012). Bacterial community transcription patterns during a marine phytoplankton bloom. *Environ. Microbiol.* 14, 228–239. doi:10.1111/j.1462-2920.2011.02602.x.

Ritson, J. P., Graham, N. J. D., Templeton, M. R., Clark, J. M., Gough, R., and Freeman, C. (2014). The impact of climate change on the treatability of dissolved organic matter (DOM) in upland water supplies: A UK perspective. *Sci. Total Environ.* 473–474, 714–730. doi:10.1016/J.SCITOTENV.2013.12.095.

Roszak, D. B., and Colwell, R. R. (1987). Metabolic activity of bacterial cells enumerated by direct viable count. *Appl. Environ. Microbiol.* 53, 2889–2893. doi:PMC204217.

Salcher, M. M. (2014). Same same but different: Ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi:10.4081/jlimnol.2014.813.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907. doi:10.1038/ismej.2012.162.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153.

Sigee, D. C. (2005). *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment*. Wiley Available at: https://books.google.co.uk/books?id=AJ8nIkoB1o0C.

Simpson, J. H., Vennell, R., and Souza, A. J. (2001). The salt fluxes in a tidally-energetic estuary. *Estuar. Coast. Shelf Sci.* 52, 131–142. doi:10.1006/ecss.2000.0733.

Tamaki, H., Tanaka, Y., Matsuzawa, H., Muramatsu, M., Meng, X. Y., Hanada, S., et al. (2011). Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int. J. Syst. Evol. Microbiol.* 61, 1442–1447. doi:10.1099/ijs.0.025643-0.

Thomas, F., Hehemann, J.-H. H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut Bacteroidetes: The food connection. *Front. Microbiol.* 2, 1–16. doi:10.3389/fmicb.2011.00093.

Velimirov, B. (2001). Nanobacteria, Ultramicrobacteria and Starvation Forms: A Search for the Smallest Metabolizing Bacterium. *Microbes Environ.* 16, 67–77. doi:10.1264/jsme2.2001.67.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi:10.1021/es0707198.

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci. Technol.* 42, 6749–6754. doi:10.1021/es800720n.

Wickham, H. (2009). *ggplot2: Elegant graphics for Data Analysis*. Springer-Verlag New York Available at: http://ggplot2.org.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* 40, 1–29. Available at: http://www.jstatsoft.org/v40/i01/.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *bioRxiv*, 762302. doi:10.1101/762302.

Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin:

an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26. doi:10.1186/2049-2618-2-26.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42. doi:10.1093/nar/gkt1209.

Yu, A. C. S., Loo, J. F. C., Yu, S., Kong, S. K., and Chan, T. F. (2014). Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. *Appl. Microbiol. Biotechnol.* 98, 855–862. doi:10.1007/s00253-013-5377-9.

Alonso-Sáez, L., and Gasol, J. M. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean coastal waters. *Appl. Environ. Microbiol.* 73, 3528–3535. doi:10.1128/AEM.02627-06.

Andrews, S. (2010). FastQC - A quality control tool for high throughput sequence data. *Babraham Bioinforma.* doi:citeulike-article-id:11583827.

Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. Available at: https://cran.r-project.org/package=gridExtra.

Barns, S. M., Cain, E. C., Sommerville, L., and Kuske, C. R. (2007). Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl. Environ. Microbiol.* 73, 3113–3116. doi:10.1128/AEM.02012-06.

Beman, J. M., Arrigo, K. R., and Matson, P. A. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434, 211–214. doi:10.1038/nature03370.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, G., Walters, W. A., et al. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol. Biochem.* 43, 1450–1455. doi:10.1016/j.soilbio.2011.03.012.The.

Bond, P. L., Hugenholtz, P., Keller, J., and Blackall, L. L. (1995). Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl. Environ. Microbiol.* 61, 1910–1916.

Brailsford, F. L., Glanville, H. C., Golyshin, P. N., Johnes, P. J., Yates, C. A., and Jones, D. L. (2019a). Microbial uptake kinetics of dissolved organic carbon (DOC) compound groups from river water and sediments. *Sci. Rep.* 9, 1–11. doi:10.1038/s41598-019-47749-6.

Brailsford, F. L., Glanville, H. C., Golyshin, P. N., Marshall, M. R., Lloyd, C. E., Johnes, P. J., et al. (2019b). Nutrient enrichment induces a shift in dissolved organic carbon (DOC) metabolism in oligotrophic freshwater sediments. *Sci. Total Environ.* 690, 1131–1139. doi:10.1016/J.SCITOTENV.2019.07.054.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208–211. doi:10.1038/nature14486.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Calvo-Díaz, A., Ídaz-Pérez, L., Suárez, L. Á., Morán, X. A. G., Teira, E., and Marañón, E. (2011). Decrease in the autotrophic-to-heterotrophic biomass ratio of picoplankton in oligotrophic marine waters due to bottle enclosure. *Appl. Environ. Microbiol.* 77, 5739–5746. doi:10.1128/AEM.00066-11.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–6. doi:10.1038/nmeth.f.303.

Carrera, M., Zandomeni, R. O., Fitzgibbon, J., and Sagripanti, J. L. (2007). Difference between the spore sizes of Bacillus anthracis and other Bacillus species. *J. Appl. Microbiol.* 102, 303–312. doi:10.1111/j.1365-2672.2006.03111.x.

Castelle, C. J., Brown, C. T., Thomas, B. C., Williams, K. H., and Banfield, J. F. (2017). Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. *Sci. Rep.* 7, 1–12. doi:10.1038/srep40101.

Chiang, E., Schmidt, M. L., Berry, M. A., Biddanda, B. A., Burtner, A., Johengen, T. H., et al. (2018). Verrucomicrobia are prevalent in north- temperate freshwater lakes and display class- level preferences between lake habitats. *PLoS One* 13, 1–20. doi:10.1371/journal.pone.0195112.

Chopyk, J., Allard, S., Nasko, D. J., Bui, A., Mongodin, E. F., and Sapkota, A. R. (2018). Agricultural Freshwater Pond Supports Diverse and Dynamic Bacterial and Viral Populations. *Front. Microbiol.* 9, 792. doi:10.3389/fmicb.2018.00792.

Dang, H., and Lovell, C. R. (2016). Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* 80, 91–138. doi:10.1128/MMBR.00037-15.Address.

de Sosa, L. L., Glanville, H. C., Marshall, M. R., Schnepf, A., Cooper, D. M., Hill, P. W., et

al. (2018). Stoichiometric constraints on the microbial processing of carbon with soil depth along a riparian hillslope. *Biol. Fertil. Soils* 54, 949–963. doi:10.1007/s00374-018-1317-2.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Dunfield, P. F., Tamas, I., Lee, K. C., Morgan, X. C., McDonald, I. R., and Stott, M. B. (2012). Electing a candidate: A speculative history of the bacterial phylum OP10. *Environ. Microbiol.* 14, 3069–3080. doi:10.1111/j.1462-2920.2012.02742.x.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., et al. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2, 6. doi:10.1186/2049-2618-2-6.

Fenchel, T. (2008). The microbial loop - 25 years later. *J. Exp. Mar. Bio. Ecol.* 366, 99–103. doi:10.1016/j.jembe.2008.07.013.

Findlay, S. E. G., Sinsabaugh, R. L., Sobczak, W. V., and Hoostal, M. (2003). Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter. *Limnol. Oceanogr.* 48, 1608–1617. doi:10.4319/lo.2003.48.4.1608.

Fogg, G. E., and Calvario-Martinez, O. (1989). Effects of bottle size in determinations of primary productivity by phytoplankton. *Hydrobiologia* 173, 89–94. doi:10.1007/BF00015518.

Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2013). Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* 3, 2471. doi:10.1038/srep02471.

Ghuneim, L.-A. J., Jones, D. L., Golyshin, P. N., and Golyshina, O. V (2018). Nano-Sized and Filterable Bacteria and Archaea: Biodiversity and Function. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.01971.

Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M., and Pinhassi, J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ. Microbiol.* 14, 2361–2378. doi:10.1111/j.1462-2920.2012.02804.x.

Gooday, A. J. (2002). Biological Responses to Seasonally Varying Fluxes of Organic Matter to the Ocean Floor: A Review. *J. Oceanogr.* 58, 305–332. doi:10.1023/A:1015865826379.

Hahn, M. W. (2004). Broad diversity of viable bacteria in "sterile" (0.2 μm) filtered water. *Res. Microbiol.* 155, 688–691. doi:10.1016/j.resmic.2004.05.003.

Hammes, F., Vital, M., and Egli, T. (2010). Critical evaluation of the volumetric "bottle effect" on microbial batch growth. *Appl. Environ. Microbiol.* 76, 1278–1281. doi:10.1128/AEM.01914-09.

Harris, J. K., Kelley, S. T., and Pace, N. R. (2004). New Perspective on Uncultured Bacterial Phylogenetic Division OP11. *Appl. Environ. Microbiol.* 70, 845–849. doi:10.1128/AEM.70.2.845-849.2004.

Harwood, C. S., and Canale-Parola, E. (1984). *Ecology of Spirochetes*. doi:10.1146/annurev.mi.38.100184.001113.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional

annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi:10.1093/nar/gkv1248.

Hugenholtz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R. (1998). Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring. 180, 366–376.

Janssen, P. H. (2006). Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes MINIREVIEWS Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi:10.1128/AEM.72.3.1719.

Jones, T. G., Evans, C. D., Jones, D. L., Hill, P. W., and Freeman, C. (2016). Transformations in DOC along a source to sea continuum; impacts of photo-degradation, biological processes and mixing. *Aquat. Sci.* 78, 433–446. doi:10.1007/s00027-015-0461-0.

Jooste, P. J., and Hugo, C. J. (1999). The taxonomy, ecology and cultivation of bacterial genera belonging to the family Flavobacteriaceae. *Int. J. Food Microbiol.* 53, 81–94. doi:10.1016/S0168-1605(99)00162-2.

Kandlikar, G. (2018). ranacapa: Utility Functions For Downstream Anacapa Analyses in R.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of Acidobacteria: Moving beyond genes and genomes. *Front. Microbiol.* 7, 1–16. doi:10.3389/fmicb.2016.00744.

Kirchman, D. L., Dittel, A. I., Findlay, S. E. G., and Fischer, D. (2004). Changes in bacterial activity and community structure in response to dissolved organic matter in the Hudson River, New York. *Aquat. Microb. Ecol.* 35, 243–257. doi:10.3354/ame035243.

Landa, M., Cottrell, M. T., Kirchman, D. L., Blain, S., and Obernosterer, I. (2013). Changes

in bacterial diversity in response to dissolved organic matter supply in a continuous culture experiment. *Aquat. Microb. Ecol.* 69, 157–168. doi:10.3354/ame01632.

Lee, K. C. Y., Dunfield, P. F., and Stott, M. B. (2014). "The Phylum Armatimonadetes," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 447–458. doi:10.1007/978-3-642-38954-2_388.

Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et al. (2016). Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu. Rev. Microbiol.* 70, 235–54. doi:10.1146/annurev-micro-102215-095748.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi:10.1038/ncomms7372.

Malmstrom, R. R., Cottrell, M. T., Elifantz, H., Kirchman, D. L., Kiene, R. P., Yu, L., et al. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean Diversity and abunda. *Appl Env. Microbiol* 71, 2979–86. doi:10.1128/AEM.71.6.2979.

Margulis, L., Ashen, J. B., Sole, M., and Guerrero, R. (1993). Composite, large spirochetes from microbial mats: spirochete structure review. *Proc Natl Acad Sci U S A* 90, 6966–6970. doi:10.1073/pnas.90.15.6966.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12.

McBride, M. J. (2014). "The Family Flavobacteriaceae," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 643–676. doi:10.1007/978-3-642-38954-2_130.

McMurdie, P. J., Holmes, S., with contributions from Gregory Jordan, and Chamberlain, S. (2018). phyloseq: Handling and analysis of high-throughput microbiome census data. Available at: http://dx.plos.org/10.1371/journal.pone.0061217.

Meyer, J. L. (1994). The microbial loop in flowing waters. *Microb. Ecol.* 28, 195–199. doi:10.1007/BF00166808.

Morales, M., R Core Team, Community, R. listerv, and Murdoch, D. (2017). sciplot: Scientific Graphing Functions for Factorial Designs. Available at: https://cran.r-project.org/package=sciplot.

Nelson, W. C., and Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front. Microbiol.* 6, 1–14. doi:10.3389/fmicb.2015.00713.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). vegan: Community Ecology Package. Available at: https://cran.r-project.org/package=vegan.

Panikov, N. S. (2005). Contribution of Nanosized Bacteria to the total biomass and activity of a soil microbial community. *Adv. Appl. Microbiol.* 57, 245–296.

Pernthaler, J., Posch, T., Šimek, K., Vrba, J., Pernthaler, A., Glöckner, F. O., et al. (2001). Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in

mixed continuous culture. *Appl. Environ. Microbiol.* 67, 2145–2155. doi:10.1128/AEM.67.5.2145.

Proctor, C. R., Besmer, M. D., Langenegger, T., Beck, K., Walser, J.-C., Ackermann, M., et al. (2018). Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J.* doi:10.1038/s41396-018-0070-8.

Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., Eck, J., et al. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. *Mol. Microbiol.* 50, 563–575. doi:10.1046/j.1365-2958.2003.03707.x.

R Core Team (2017). R: A language and environment for statistical computing. Available at: https://www.r-project.org/.

Rinta-Kanto, J. M., Sun, S., Sharma, S., Kiene, R. P., and Moran, M. A. (2012). Bacterial community transcription patterns during a marine phytoplankton bloom. *Environ. Microbiol.* 14, 228–239. doi:10.1111/j.1462-2920.2011.02602.x.

Ritson, J. P., Graham, N. J. D., Templeton, M. R., Clark, J. M., Gough, R., and Freeman, C. (2014). The impact of climate change on the treatability of dissolved organic matter (DOM) in upland water supplies: A UK perspective. *Sci. Total Environ.* 473–474, 714–730. doi:10.1016/J.SCITOTENV.2013.12.095.

Roszak, D. B., and Colwell, R. R. (1987). Metabolic activity of bacterial cells enumerated by direct viable count. *Appl. Environ. Microbiol.* 53, 2889–2893. doi:PMC204217.

Salcher, M. M. (2014). Same same but different: Ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi:10.4081/jlimnol.2014.813.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907.

doi:10.1038/ismej.2012.162.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153.

Sigee, D. C. (2005). *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment*. Wiley Available at: https://books.google.co.uk/books?id=AJ8nIkoB1o0C.

Simpson, J. H., Vennell, R., and Souza, A. J. (2001). The salt fluxes in a tidally-energetic estuary. *Estuar. Coast. Shelf Sci.* 52, 131–142. doi:10.1006/ecss.2000.0733.

Tamaki, H., Tanaka, Y., Matsuzawa, H., Muramatsu, M., Meng, X. Y., Hanada, S., et al. (2011). Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *Int. J. Syst. Evol. Microbiol.* 61, 1442–1447. doi:10.1099/ijs.0.025643-0.

Thomas, F., Hehemann, J.-H. H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut Bacteroidetes: The food connection. *Front. Microbiol.* 2, 1–16. doi:10.3389/fmicb.2011.00093.

Velimirov, B. (2001). Nanobacteria, Ultramicrobacteria and Starvation Forms: A Search for the Smallest Metabolizing Bacterium. *Microbes Environ.* 16, 67–77. doi:10.1264/jsme2.2001.67.

Wang, Y., Hammes, F., Boon, N., and Egli, T. (2007). Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent enrichment of filterable bacterial communities. *Environ. Sci. Technol.* 41, 7080–7086. doi:10.1021/es0707198.

Wang, Y., Hammes, F., Düggelin, M., and Egli, T. (2008). Influence of size, shape, and flexibility on bacterial passage through micropore membrane filters. *Environ. Sci.*

Technol. 42, 6749–6754. doi:10.1021/es800720n.

Wickham, H. (2009). *ggplot2: Elegant graphics for Data Analysis*. Springer-Verlag New York Available at: http://ggplot2.org.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* 40, 1–29. Available at: http://www.jstatsoft.org/v40/i01/.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *bioRxiv*, 762302. doi:10.1101/762302.

Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26. doi:10.1186/2049-2618-2-26.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42. doi:10.1093/nar/gkt1209.

Yu, A. C. S., Loo, J. F. C., Yu, S., Kong, S. K., and Chan, T. F. (2014). Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. *Appl. Microbiol. Biotechnol.* 98, 855–862. doi:10.1007/s00253-013-5377-9.

Chapter 6

Discussion

6.1. Introduction

In this chapter, we further explore the experimental work presented in this thesis by summarizing the key findings, discussing them in relation to the objectives of the thesis, and identifying future areas of research.

6.2. Overall discussion

6.2.1. Taxonomic composition of the Conwy River

Overall, the Conwy River is a physically, chemically and biologically diverse lotic system. A schematic of our findings regarding the taxonomic compositions of the unfiltered and filtered fraction is presented in Figure 6.1. Taken together, the results from the taxonomic studies revealed that the microorganisms in the Conwy water are rather ubiquitous in freshwater systems, yet their distributions differed depending upon the size fraction examined (Fig. 6.1).

The results indicated that, regardless of fraction, *Proteobacteria* is the dominant phylum. *Proteobacteria*, as stated in the previous chapters, is the most characterized of all the bacterial phyla and are a major constituent of the global microbiome (Sunagawa et al., 2015; Thompson et al., 2017). However, the fractions significantly differ at the genera level. Nanopore™ sequencing suggested that *Polynucleobacter, Pseudomonas, Legionella, Rickettsia, Bordetella, Bdellovibrio,* and *Ralsontia* are the top genera in the unfiltered community. However, *Ralstonia*, *Variovorax,* and *Geobacter* are the major constituents of the filtered fraction. As discussed in Chapters 3 and 4, *Ralstonia* and *Variovorax* have been observed passing through 0.22 µm filters in medical studies. *Geobacter* is notable as it has the ability to oxidize organic compounds via the reduction of Fe (III) and Mn (IV) oxides and some spcies can oxidize aromatic hydrocarbons (Lovley et al., 2011).

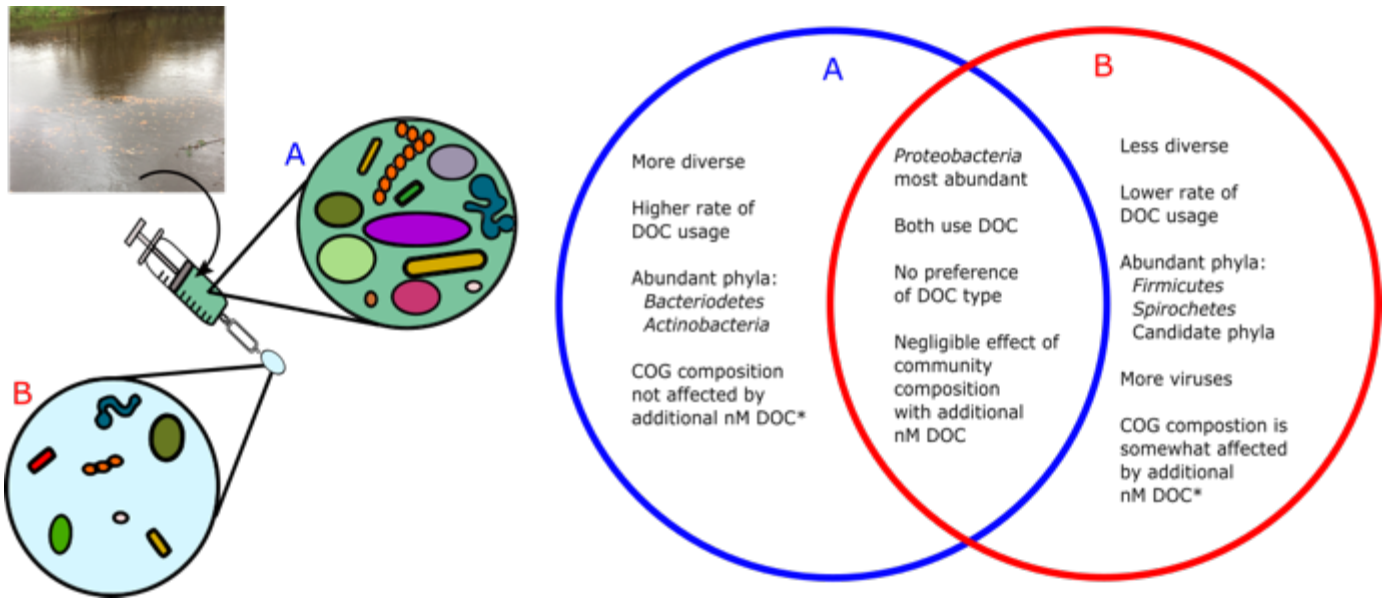


Figure 6.1- Summary of the fractions of the Conwy River. Firstly, freshwater samples were obtained from the Conwy River. The (A) unfiltered fraction is passed through a 0.22 µm syringe filter to obtain the (B) filtered fraction. Although there are similar groups found within both fractions, there is greater diversity within the (A) unfiltered fraction than the (B) filtered fraction. As larger majority populations are removed, minority populations can be detected. A Venn diagram summarizes the major findings of the thesis comparing the two fractions in terms of DOC usage, abundant phyla, and COGs. * Note: Further research is needed.

Interestingly, the meta-barcoding analysis provided a different insight. Of note is that we only examined approximately 250 bp section of the 16S rRNA genes so accurate taxanomic affiliation at the genera level remains subject to error. Again, there are differences in terms of distribution of OTUs within the *Proteobacteria* phylum that differ in the fractions. *Comamonadaceae* (unfiltered and filtered fractions), *Oxalobacteraceae* (filtered and unfiltered fraction), *Pseudomonadaceae* (filtered fraction), *Campylobacteraceae* (filtered fraction), and *Halomonadaceae* (filtered fraction) were the most prevalent families. These families are ubiquitous throughout Earth's microbiome and is difficult to pinpoint exact functionality due to this (Dewhirst et al., 1994; Williams et al., 2010; Baldani et al., 2014; de la Haba et al., 2014; Willems, 2014; Flynn et al., 2017).

The next most common phyla are *Bacteriodetes*, *Actinobacteria*, and *Firmicutes*, which are not only native to freshwater but also to the surrounding terrestrial system. As mentioned previously, the distribution of these three phyla differed depending on the size fraction examined. The filtered fraction showed a much higher proportion of *Firmicutes* than the *Bacteriodetes* and *Actinobacteria*. In fact, we can see in Chapters 3-6 that *Firmicutes* was the second most common phylum in this fraction. It is suspected that these are not actually metabolically active but are actually in a sporulation form. The evidence for this is from the lag phase observed in regards to DOC utilisation even though *Firmicutes* was the dominate phylum in this phase, and quickly dissipated as metabolic activity increased. While examining the most prominent taxa at the family level, *Bacillaceae* constituted the overwhelming majority. This family is well known for sporulation and species such as *Bacillus subtilis* are part of this taxa (Mandic-Mulec et al., 2015). The ubiquity of this family, especially in terrestrial systems, makes it rather difficult to identify their exact function other than a general nutrient cycling (Mandic-Mulec et al., 2015).

Bacteriodetes and *Actinobacteria* were predominately found in the unfiltered fraction. However, unlike in the case of *Firmicutes*, these are more likely to be metabolically active, especially in the unfiltered fraction, as seen in Chapter 5. *Actinobacteria* and *Bacteroidetes* are major players in C cycling as they can degrade high molecular weight biopolymers (e.g. chitin, cellulose, and hemicellulose) and other complex C molecules (e.g. hydrocarbons, nitroaromatics, sulfonated azo dyes, and pesticides) (Thomas et al., 2011; Lewin et al., 2016). A prominent family in the Conwy river water was *Flavobacteriaceae*, which is the largest family in the phylum of *Bacteroidetes* with at least 90 genera (McBride, 2014). A notable quality to this family is the ability to utilize polysaccharides on the cell surface, i.e. can bind polysaccharides and transport oligomers via the outer membrane (McBride, 2014).

As for individual sequences found in the filtered fraction belonging to *Bacteroidetes* and *Actinobacteria*, it is widely known that there is number of ultramicrobacteria (specifically *Actinobacteria*) that are found in planktonic communities (Pernthaler et al., 2001; Hahn et al., 2004; Hahn, 2009; Salcher et al., 2013; Lewin et al., 2016). For instance, Ac1 *Actinobacteria*, which are commonly characterised as being ultra-small, are the most abundant group in freshwater systems and can degrade chitin and chitin by-products from fungi, diatoms, and crustaceans (Pernthaler et al., 2001; Salcher et al., 2013; Lewin et al., 2016). Although Ac1 was not detected in the Conwy, there could be similar features for other groups of *Actinobacteria*, as the inherent DOC content of the freshwater flowing along the Conwy catchment fluctuates between oligotrophy and mesotrophy (approximately 1.5-10 mg/L) depending upon external conditions (agricultural run-off, precipitation) and location (upland vs lowland sites) (Emmett et al., 2016; Brailsford et al., 2017, 2019). Interestingly, Ac1 was only present in low abundance, unlike *Microbacteriaceae*, *Micrococcaceae*, *Corynebacteriaceae* and ACK-M1, which are common groups in freshwater systems (Lewin et al., 2016; Chopyk et al., 2018) .

In addition, we should also address *Chloroflexi* in the filtered fraction when examining 16S rRNA genes via Sanger sequencing. This was the only method to show that this phylum was a dominant component of the microbial community of the filtered fraction (and eve the general community). Even so, we must consider the level of similarity seen between the sequences identified to *Chloroflexi* species. These sequences were distantly related to many of the species in the phyla with the exception of two sequences, which did share a high degree of similarity of an uncultured *Chloroflexi* bacterium (Accession number HE614853). Therefore, although these 16S rRNA sequences were retrieved from bacteria, we know little else about this group. Even so, the degree of similarity to this particular phylum is something that could be further explored in future studies.

There was evidence of candidate phyla in the Conwy River that have been previously proposed to have intrinsically small cell sizes such as "*Candidatus* Parcubacteria" (Candidate phylum OD1/WWE3). Although evidence suggests some of these candidate phyla may have a small cell volume due to its minimized genome, these small genomes do not always correlate to small volume. The presence of this phylum in the filtered fraction may be due to another factor such as morphology and potential symbiosis with other microorganisms.

In regards to archaea, shotgun sequencing and 16S rRNA single amplicon showed that there were *Euryarchaeota* and *Thaumarchaeota* present in both fractions in low abundance. Archaea make up less than 10% of the microbial community in freshwater ecosystems and effectively utilise DOM (Wells et al., 2006; Bomberg et al., 2008; Herfort et al., 2009; Cavicchioli, 2011).

Finally, we address the viral entities in the Conwy. There was a notable difference between the filtered and unfiltered fraction as there was an overrepresentation of dsDNA viruses within the filtered fraction. Viruses, although considered non-living entities, are an important component to the greater microbial community as they can be drivers of selection. The majority of viruses are host-specific, which typically target dense populations of organisms and subsequent lysis releases typically unavailable nutrients to the wider community (Weinbauer, 2004; Winter et al., 2010; Salcher, 2014). This behaviour acts as both a population control ("killing the winner") and nutrient provider. In addition, they vary in size (20 nm to 500 nm) allowing them to easily pass through a 0.22 um filter (Leiman et al., 2003). This filterability along with 'killing the winner' scenario is the most likely explanation for higher population of viruses in the filterable community (Weinbauer, 2004; Winter et al., 2010; Salcher, 2014).

6.2.2. COGs of genomes vs COGs of metagenomes, is there a link?

The metagenomic investigations revealed that the Conwy River might contain several unidentified candidate phyla as discussed previously. These phyla are predicted to have small genomes (less than 1.5 Mbp) (Harris et al., 2004; Duda et al., 2012; Castelle et al., 2015; Luef et al., 2015; Proctor et al., 2018) . Organisms that have these streamlined genomes only contain essential functionalities, such as machinery for DNA repair and basic metabolic processes (such as glycolysis, etc) (Giovannoni et al., 2014). In light of this, we aimed to determine the distribution of organisms within a particular environment based on the available clusters of orthologous groups (COGs). We did this by comparing the distribution of COGs of both metagenomes and genomes via clustering analysis. The first step was to explore what COGs are actually in the Conwy River and if there was a difference in composition between the fractions to infer a distinct role that the filtered fraction plays in the overall community. The evidence suggested that the filtered fraction may have a distinctive role, as the distribution of COGs in this section did not resemble the unfiltered fraction. The unfiltered fraction contained more COGs that were affiliated with amino acid metabolism and general energy production, whereas the filtered fraction contained an overwhelming amount of COGs dedicated to replication, recombination and repair of DNA. This may be in large part due to the amount of viruses in the filtered fraction. As seen in Chapter 4, when examining the COGs assigned to specific taxonomic groups, there was an entire order of magnitude more dsDNA viruses in the filtered fraction than the unfiltered fraction and upon further examination of these groups there was an overrepresentation of DNA helicase and DNA polymerase I.

Once a baseline was established, the next step was to determine if there was a trend that existed between the distribution of COGs and genome size. According to our study and previous investigations, there is a strong connection between the prevalence of certain COGs

and genome size (Konstantinidis and Tiedje, 2004). Microorganisms with smaller genomes have a higher percentage of COGs dedicated to the "essential functions" such as transcription, ribosomal structure and biogenesis and microorganisms with larger genomes had more varied capabilities (specifically amino acid transport and metabolism) (Konstantinidis and Tiedje, 2004).

Thirdly, we compared the Conwy River to other systems. That is, how do the COGs in the general and filterable communities of the Conwy water compare to metagenomes from other habitats, systems, etc. The majority of these ecosystems were freshwater (i.e. groundwater, lotic, and lentic waters) but we also examined marine, acid mine drainage, and soil systems. Although clustering suggested that the general community of the Conwy water was similar to other systems, the filterable community was far removed from all the other systems.

This finding led to the third and final step, which is comparison of COG compositions of genomes and metagenomes. As stated previously, the aim was to determine if the distribution of COGs seen at the metagenome level is representative of the type of microorganisms. According to our results, the metagenomes clustered with other metagenomes (and it goes for genomes as well). There isn't a correlation between COG distributions of metagenomes and genomes. Of note, the filtered fraction is unique across all these systems, as it was very different from all other systems. In addition, when examined via dendrogram, it was more closely related to small genome microorganisms than other metagenomes.

6.2.3. Usage of low molecular weight DOC by the filtered and unfiltered fractions

In previous studies and Chapter 6, the prokaryotic species residing in both fractions were shown to be capable of metabolizing low molecular weight compounds (amino acids, organic acids and sugars). The major difference between the fractions is the observed lag

phase, i.e. minimal metabolic activity in the first 74 h, in the filterable microbial component. That is, the microorganisms residing in the filtered fraction do not start metabolizing low molecular weight (LMW) DOC until after a period of 3 d.

The presence of the lag phase is not exclusive to this system, as previous investigations within the same catchment have shown a similar response. It was noted in Chapter 6 that the DOC usage in the 0.22 μm filtered water of main Conwy river (at the sample site) closely resembled 0.22 μm filtered water of the Migneint sub-catchment (land use type: upland peat bogs) (Brailsford et al., 2017). In addition, when comparing mesotrophic rivers within the catchment to the oligotrophic main river revealed minimal metabolic activity of LMW DOC (μM concentrations) within the first 72 h within unfiltered oligotrophic water and sediment and mesotrophic sediment, especially at higher μM concentrations (Brailsford et al., 2019a). The DOC metabolic activity of the filtered fraction was similar to oligotrophic conditions in Brailsford et al. (2019a). This further confirms that the lag phase is caused by both low abundance of microorganisms (removing larger microorganisms such as protists and other heterotrophic prokaryotes via filtration) and unfavourable conditions (i.e. oligotrophic/mesotrophic conditions and the removal of larger sediments via the act of filtration).

As for a preference in the metabolism of LMW substrates, neither the general microbial community nor the filtered fraction exhibited any strong preference. That is the rate of consumption across the substrate types were fairly consistent. This may be due to the functional redundancy seen in many prokaryotes as all contain a basic metabolism to utilise LMW DOC. Although the composition of the microbial community was not affected by LMW DOC, the COG functional categories were. The filtered fraction showed more variation across the three weeks than the unfiltered fraction.

6.3. Future research directions

Broadly, there are two approaches that could be taken to investigate microorganisms residing in the Conwy River: (1) whole microbial community level studies, or (2) studies in pure cultures (isolation attempts). Investigation at the community level treats each microorganism population as a singular entity and is ideally suited to investigating the relative role each microbial size fraction plays in DOM cycling. Techniques that could be included in further studies include the use of stable isotope tracer experiments (e.g. ^{13}C, ^{15}N, ^{18}O), flow cell cytometry (FACS sorting), un-targeted metabolomics, meta-transcriptomics, and proteomics. In addition, it would be desirable to use these approached to investigate DOM cycling under a wide range of environmental conditions, such as:

(1) Examination of other size fractions (e.g. 0.45 μm and 0.05 μm pore sizes).
(2) Changing the oxygen level via microcosm system (examining obligate anaerobes and aerobes, facultative anaerobes, microaerophiles, and aerotolerant organisms).
(3) Using larger MW DOM substrates (polypeptides, polysaccharides, humic substances etc).
(4) Utilizing common chemicals from an anthropogenic source (herbicides, pesticides, agrochemicals, medicines, etc).
(5) Investigation of other locations
    a. Within the catchment (upstream, downstream, small tributaries, near point source inputs, across the river-estuary transition zone).
    b. Other lotic systems with either different trophic conditions (copiotrophy) or similar nutrient levels (mesotrophy/oligotrophy).
    c. Other freshwater systems (lotic vs lentic vs ground water, etc).
    d. Other aquatic systems (marine vs freshwater).
    e. Other habitats (soil vs lotic, animal intestinal tract vs lotic etc).

(6) Investigating the effects of seasonality and extreme weather events (e.g. droughts, floods, heatwaves, and UV exposure).

(7) Investigation of the organisms which regulate ultra-small microbial populations in freshwaters (e.g. viral vs grazers).

(8) Investigate the relative abundance of ultra-small bacteria in the water column vs in sediments.

(9) Determine if ultra-small microorganisms reside in soil to ascertain if they are a major source entering freshwaters.

Another important aspect in community analysis is the implementation of ecological theory. Since many of the key principles in ecological theory arose from direct observation of macro natural systems, applying ecological theory can be challenging as phenomena such as horizontal gene transfer, rapid reproduction rates, and the lack of a delineation of species are not accounted for (Prosser et al., 2007; Hibbing et al., 2010). Hence further studies examining the entire microbial community are essential in altering and improving ecological theory as a whole.

The investigation of individual cells for eventual isolation would require an entirely different approach, as traditional culture dependent methodologies used in isolation are not ideal. Many microorganisms (i.e. 'microbial dark matter') cannot be cultured currently under typical *in vitro* conditions. Therefore, culture-independent methods should be the initial step before potential isolation. Techniques such as immunoprecipitation, microscopy (general, SEM, TEM, FISH, etc), and next generation sequencing (metagenomics, single-cell genomics, etc) can be implemented to expand our knowledge of the role these organisms play in freshwater ecosystems. These techniques can also answer questions concerning physiology (for physical separation from the general populous), genome composition (determining taxon,

constructing whole genomes, identifying potential targeting moieties) and growth conditions (nutrient preferences, ideal temperature, etc).

Even when a potential new organism is found, there is still the issue of taxonomic classification. There are two main issues in regards to taxonomic classification, the first being the lack of a clear species definition and second the discrepancies in naming convention. Firstly, species delineation via polyphasic microbial taxonomy must satisfy the following conditions: (1) DNA-DNA hybridization (DDH) values at least 70%, (2) at least 97% rRNA gene sequence similarity, (3) 2% of G+C span maximum, and (4) differentiation based on chemotaxonomic and phenotypic features, via techniques such as fatty acid methyl ester (FAME), polyamines, peptidoglycan type sphingolipids, and matrix-assisted laser desporption/ionization—time-of-flight mass spectrometer (Thompson et al., 2013, 2015). Although many species have been established this way, new sequencing technologies makes these conditions antiquated leading to broad species definitions not based on the definition of an evolutionary species. To combat this, the actions proposed include: (1) whole genome sequencing and check for similarity within species via multi-locus sequence analysis (MLSA), genome-to-genome distance (GGD), and average amino acid identity (AAI), (2) cross reference with Bergey's guide, (4) using a genome-to-phenotype approach rather than phenome to genome, (5) avoid techniques such as FAME, MALDI-TOF MS, etc, and finally (6) depositing sequences to open access and public databases (Thompson et al., 2015). Secondly, there is no consistency in naming convention across the databases, especially candidate phyla. For instance, candidate phyla WWE3 is WWE3 in NCBI and Greengenes databases whereas the SILVA database classifies this phylum as OD1 and "*Candidatus* Parcubacteria" (Pruesse et al., 2007; McDonald et al., 2012; Yilmaz et al., 2014). In response to this, a new database, the genome taxonomy database (GTDB) has been created (Parks et al., 2018). The goal of the GTDB is to standardise microbial taxonomy based on genome

phylogeny across the NCBI, Greengenes, and SILVA. The creation of the GTDB and the differentiation of species should alleviate the current issues that plague microbial taxonomy.

6.4. Final thoughts

According to the evidence provided in this thesis, we conclude that (1) there is a population of metabolically active cells that are capable of passing through a 0.22 μm filter residing in the Conwy River, and (2) that these cells participate in DOM (specifically DOC and DON) cycling. The specific taxa responsible for this DOM processing, however, remains elusive, as there is a high level of functional redundancy amongst many prokaryotes and the presence of 'microbial dark matter'. These findings give us a better understanding of the taxonomic microbial content of river water, the role that microorganisms including filterable microorganisms play in the behavior and fate of low molecular weight DOC within the Conwy River.

6.5. References

Baldani, J. I., Rouws, L., Cruz, L. M., Olivaries, F. L., Schmid, M., and Hartmann, A. (2014). *The Family Oxalobacteraceae*. doi:10.1007/978-3-642-30197-1.

Bomberg, M., Montonen, L., Münster, U., and Jurgens, G. (2008). Diversity and function of archaea in freshwater habitats. *Curr. Trends Microbiol.* 4, 1–89. Available at: http://www.researchtrends.net/tia/abstract.asp?in=0&vn=4&tid=41&aid=2693&pub=2008&type=3.

Brailsford, F. L., Glanville, H. C., Golyshin, P. N., Marshall, M. R., Lloyd, C. E., Johnes, P. J., et al. (2019). Nutrient enrichment induces a shift in dissolved organic carbon (DOC) metabolism in oligotrophic freshwater sediments. *Sci. Total Environ.* 690, 1131–1139. doi:10.1016/J.SCITOTENV.2019.07.054.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., et al. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Sci. Total Environ.* 598, 377–384. doi:10.1016/j.scitotenv.2017.04.049.

Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. doi:10.1016/j.cub.2015.01.014.

Cavicchioli, R. (2011). Archaea - Timeline of the third domain. *Nat. Rev. Microbiol.* 9, 51–61. doi:10.1038/nrmicro2482.

Chopyk, J., Allard, S., Nasko, D. J., Bui, A., Mongodin, E. F., and Sapkota, A. R. (2018). Agricultural Freshwater Pond Supports Diverse and Dynamic Bacterial and Viral Populations. *Front. Microbiol.* 9, 792. doi:10.3389/fmicb.2018.00792.

de la Haba, R. R., Arahal, D. R., Sánchez-Porro, C., and Ventosa, A. (2014). "The Family Halomonadaceae," in *The Prokaryotes: Gammaproteobacteria*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 325–360. doi:10.1007/978-3-642-38922-1_235.

Dewhirst, F. E., Seymour, C., Fraser, G. J., Paster, B. J., and Fox, J. G. (1994). Phylogeny of Helicobacter isolates from bird and swine feces and description of Helicobacter pametensis sp. nov. *Int. J. Syst. Bacteriol.* 44, 553–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7520743.

Duda, V. I., Suzina, N. E., Polivtseva, V. N., and Boronin, a. M. (2012). Ultramicrobacteria: Formation of the concept and contribution of ultramicrobacteria to biology. *Microbiology* 81, 379–390. doi:10.1134/S0026261712040054.

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., et al. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Sci. Total Environ.* 572, 1586–1600. doi:10.1016/j.scitotenv.2016.04.004.

Flynn, T. M., Koval, J. C., Greenwald, S. M., Owens, S. M., Kemner, K. M., and Antonopoulos, D. A. (2017). Parallelized, Aerobic, single carbon-source enrichments from different natural environments contain divergent microbial communities. *Front. Microbiol.* 8, 1–14. doi:10.3389/fmicb.2017.02321.

Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1–13. doi:10.1038/ismej.2014.60.

Hahn, M. W. (2009). Description of seven candidate species affiliated with the phylum Actinobacteria, representing planktonic freshwater bacteria. *Int. J. Syst. Evol. Microbiol.* 59, 112–117. doi:10.1099/ijs.0.001743-0.

Hahn, M. W., Stadler, P., Wu, Q. L., and Pöckl, M. (2004). The filtration-acclimatization

method for isolation of an important fraction of the not readily cultivable bacteria. *J. Microbiol. Methods* 57, 379–390. doi:10.1016/j.mimet.2004.02.004.

Harris, J. K., Kelley, S. T., and Pace, N. R. (2004). New Perspective on Uncultured Bacterial Phylogenetic Division OP11. *Appl. Environ. Microbiol.* 70, 845–849. doi:10.1128/AEM.70.2.845-849.2004.

Herfort, L., Kim, J., Coolen, M., Abbas, B., Schouten, S., Herndl, G., et al. (2009). Diversity of Archaea and detection of crenarchaeotal amoA genes in the rivers Rhine and Têt. *Aquat. Microb. Ecol.* 55, 189–201. doi:10.3354/ame01294.

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8, 15–25. doi:10.1038/nrmicro2259.

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3160–3165. doi:10.1073/pnas.0308653100.

Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2003). Structure and morphogenesis of bacteriophage T4. *Cell. Mol. Life Sci.* 60, 2356–2370. doi:10.1007/s00018-003-3072-1.

Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et al. (2016). Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu. Rev. Microbiol.* 70, 235–54. doi:10.1146/annurev-micro-102215-095748.

Lovley, D. R., Ueki, T., Zhang, T., Malvankar, N. S., Shrestha, P. M., Flanagan, K. A., et al. (2011). Geobacter: The Microbe Electric's Physiology, Ecology, and Practical Applications. *Adv. Microb. Physiol.* 59, 1–100. doi:10.1016/B978-0-12-387661-4.00004-5.

Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi:10.1038/ncomms7372.

Mandic-Mulec, I., Stefanic, P., and van Elsas, J. D. (2015). Ecology of Bacillaceae. *Microbiol. Spectr.* 3. doi:10.1128/microbiolspec.TBS-0017-2013.

McBride, M. J. (2014). "The Family Flavobacteriaceae," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 643–676. doi:10.1007/978-3-642-38954-2_130.

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi:10.1038/ismej.2011.139.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996. Available at: https://doi.org/10.1038/nbt.4229.

Pernthaler, J., Posch, T., Šimek, K., Vrba, J., Pernthaler, A., Glöckner, F. O., et al. (2001). Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. *Appl. Environ. Microbiol.* 67, 2145–2155. doi:10.1128/AEM.67.5.2145.

Proctor, C. R., Besmer, M. D., Langenegger, T., Beck, K., Walser, J.-C., Ackermann, M., et al. (2018). Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J.* doi:10.1038/s41396-018-0070-8.

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., et al. (2007). The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 384–392. doi:10.1038/nrmicro1643.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi:10.1093/nar/gkm864.

Salcher, M. M. (2014). Same same but different: Ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi:10.4081/jlimnol.2014.813.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907. doi:10.1038/ismej.2012.162.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science (80-.).* 348, 1261359. doi:10.1126/SCIENCE.1261359.

Thomas, F., Hehemann, J.-H. H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut Bacteroidetes: The food connection. *Front. Microbiol.* 2, 1–16. doi:10.3389/fmicb.2011.00093.

Thompson, C. C., Amaral, G. R., Campeão, M., Edwards, R. A., Polz, M. F., Dutilh, B. E., et al. (2015). Microbial taxonomy in the post-genomic era: Rebuilding from scratch? *Arch. Microbiol.* 197, 359–370. doi:10.1007/s00203-014-1071-2.

Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., and Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC Genomics* 14, 913. doi:10.1186/1471-2164-14-913.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181. doi:10.1016/j.femsre.2003.08.001.

Wells, L. E., Cordray, M., Bowerman, S., Miller, L. A., Vincent, W. F., and Deming, J. W. (2006). Archaea in particle-rich waters of the Beaufort Shelf and Franklin Bay, Canadian Arctic: Clues to an allochthonous origin? *Limnol. Oceanogr.* 51, 47–59. doi:10.4319/lo.2006.51.1.0047.

Willems, A. (2014). "The Family Comamonadaceae," in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 777–851. doi:10.1007/978-3-642-30197-1_238.

Williams, K. P., Gillespie, J. J., Sobral, B. W. S., Nordberg, E. K., Snyder, E. E., Shallom, J. M., et al. (2010). Phylogeny of gammaproteobacteria. *J. Bacteriol.* 192, 2305–2314. doi:10.1128/JB.01480-09.

Winter, C., Bouvier, T., Weinbauer, M. G., and Thingstad, T. F. (2010). Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. *Microbiol. Mol. Biol. Rev.* 74, 42–57. doi:10.1128/MMBR.00034-09.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42. doi:10.1093/nar/gkt1209.

Appendix I

Method development for long term ^{14}DOC isotopic experiments

Abstract

We attempted to examine how radiolabelled low molecular weight compounds (i.e. amino acids, sugars, and organic acids) within unfiltered river water and 0.22 μm-filtered water were metabolized by the various microbial populations over a three-week period. However, our efforts proved unsuccessful as metabolic activity was observed in the sterile control treatment after the initial 48-hour period in the organic acid treatment. Liquid chromatography–mass spectrometry (LC-MS) water was used instead of freshwater as it was thought to be chemically free and sterile. It was later discovered that the LC-MS water was only chemically free and not sterile. In addition, the setup itself allowed for outside sources for contamination, as the $^{14}CO_2$ trap was never replaced.

A1.1. Summary

The initial aims the experiments were to compare the changes in microbial communities residing in both the filter and unfiltered fraction over a three-week period by tracking the depletion and $^{14}CO_2$ respiration of added ^{14}C-labelled amino acids, organic acids, and sugars; and identifying taxonomic groups with 16S rRNA metabarcoding (metagenomics). The initial run of the experiment, lasting over a 48-hour period with no metabolomics component, showed that living cells are passing through the 0.22 μm filter. However, we were plagued by contamination issues and thus our focus shifted to remedy this. Here I document our attempts to determine potential sources of contamination and remedies to combat this prevalent issue. The methods used are described in Chapter 5.

As expected, the unfiltered fraction showed the highest rate of ^{14}C-substrate consumption across the treatments. However, there were signs of contamination within the organic acid controls, (blank treatments) very early on in the experiment (Fig. A1.1). Additionally, the large error bars observed across the majority of the samples may also be a product of variable microbial contamination within the flasks.

Once the source of contamination was identified, the experiment was repeated in November (also known as the November 2017 experiment). This time we ultra-filtered the substrates using a 0.22 μm sterile vacuum filter and we moved to a more sterile physical location. The experiment was halted after one week due to clear signs of contamination after 48 hours across all the blank treatments (Fig. A1.2). Similarly, to the previous attempt, this contamination was most apparent in the organic acid treatment. Unlike the previous experiment, however, the contamination was also apparent in the sugar and amino acid treatments as well (Fig. A1.2). We suspected that there might be a plethora of sources of contamination. Considering it was across all the substrates, our suspicions were leaning towards the LC-MS water itself. When the supplier was contacted, it was revealed that no sterilization was done to eliminate microbial contaminants from the water. Although the LC-MS water may be a source of contamination, another potential source was airborne during the sampling of the sodium hydroxide traps. As we were constantly perturbing the system and exposing the tube to the laboratory environment, there is a high chance of contamination. The final source of contamination could be from the radiolabelled substrates. The ^{14}C labelled chemicals are sterile at the point of sale, however, they are not maintained under sterile conditions after opening (they are just maintained at -20°C).

To determine this, a small-scale experiment was done to narrow down possible sources of contamination (also known as the contamination experiment). We examined the behaviour of ^{14}C labelled formic, malic and citric acid in reverse osmosis water (RO water). RO water has been previously used as a negative control (Brailsford et al., 2017) and showed no signs of contamination in these previous studies. The first 6 hours are “noisy” in comparison to the rest of the data, then after 6 hours the levels of organic acid stabilize (Fig. A1.3). This stabilization led us to conclude that the source is not from the isotopes and RO water is a good substitute for the blank treatments. However, where each stabilized is another

matter. Citric acid is the most extreme, where values are regularly over 100% (Fig. A1.3). This is clearly impossible to have more than what was added. Formic acid showed a dramatic drop in overall percent. This depletion is not characteristic of what was previously observed; as depletion continues to happen over 96 hours and then stabilizes. Hence, it can be concluded that the perceived utilization is not from a biological source. These anomalies in the data may stem from the inherent levels of radioactivity present in each of the radiolabelled substances or sorption of the substrates to the glass. In addition, it is likely that the isotope was not evenly distributed across all the samples at the beginning of the experiment. The highest volume added was citric acid and the lowest was formic acid. That correlates to high values associated with citric acid and low values with formic acid. Another important note is the matrix mixed with isotope contained the LC-MS water. If there were contamination, then there would be clear signs of depletion across all three substrate groups.

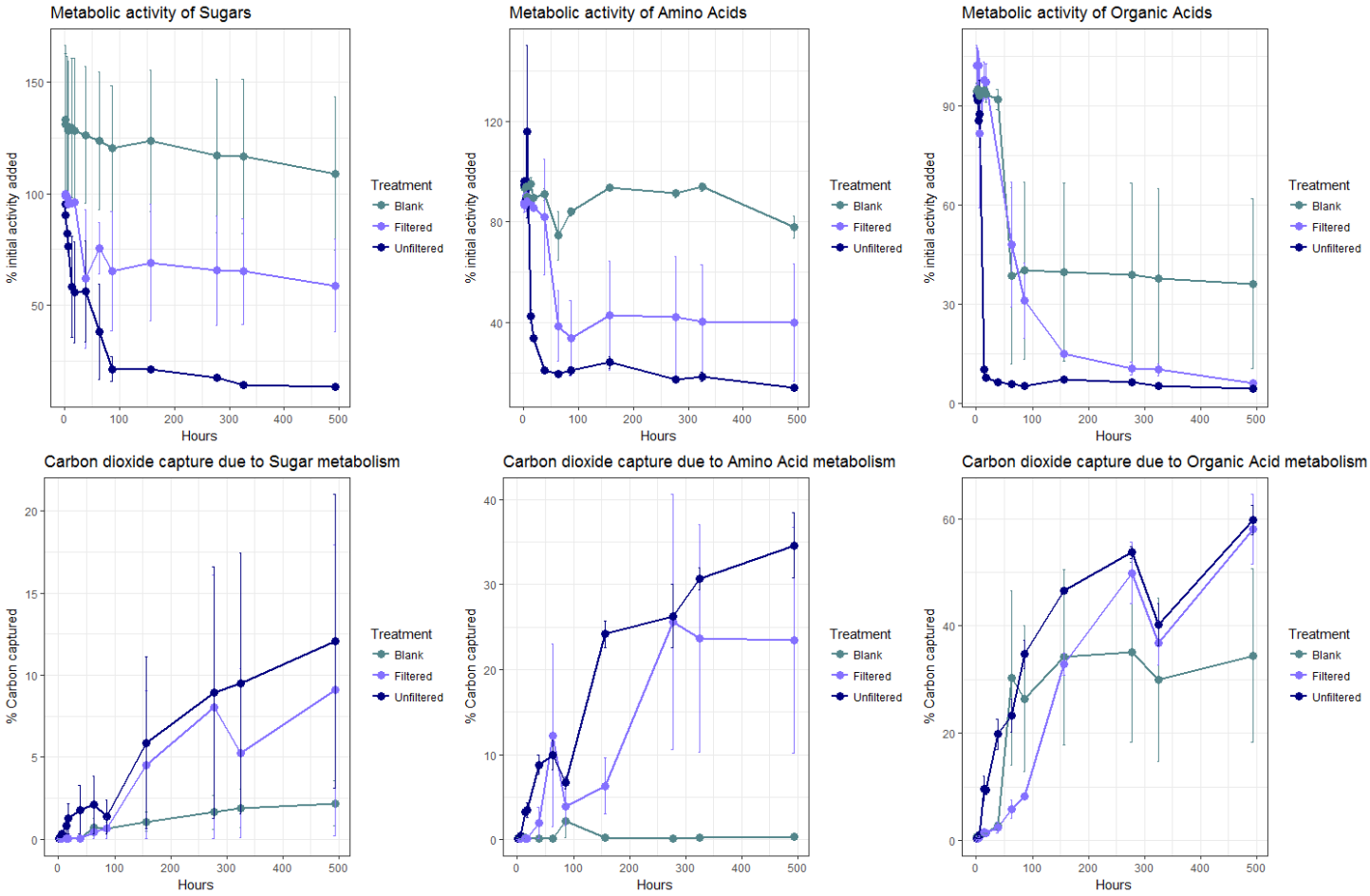


Figure A1.1- Depletion and CO_2 uptake measurements from September 2017 collection. These graphs indicate the activity of radioactivity *in* overall percent used or trapped. Values represent means ± standard error (n = 3).

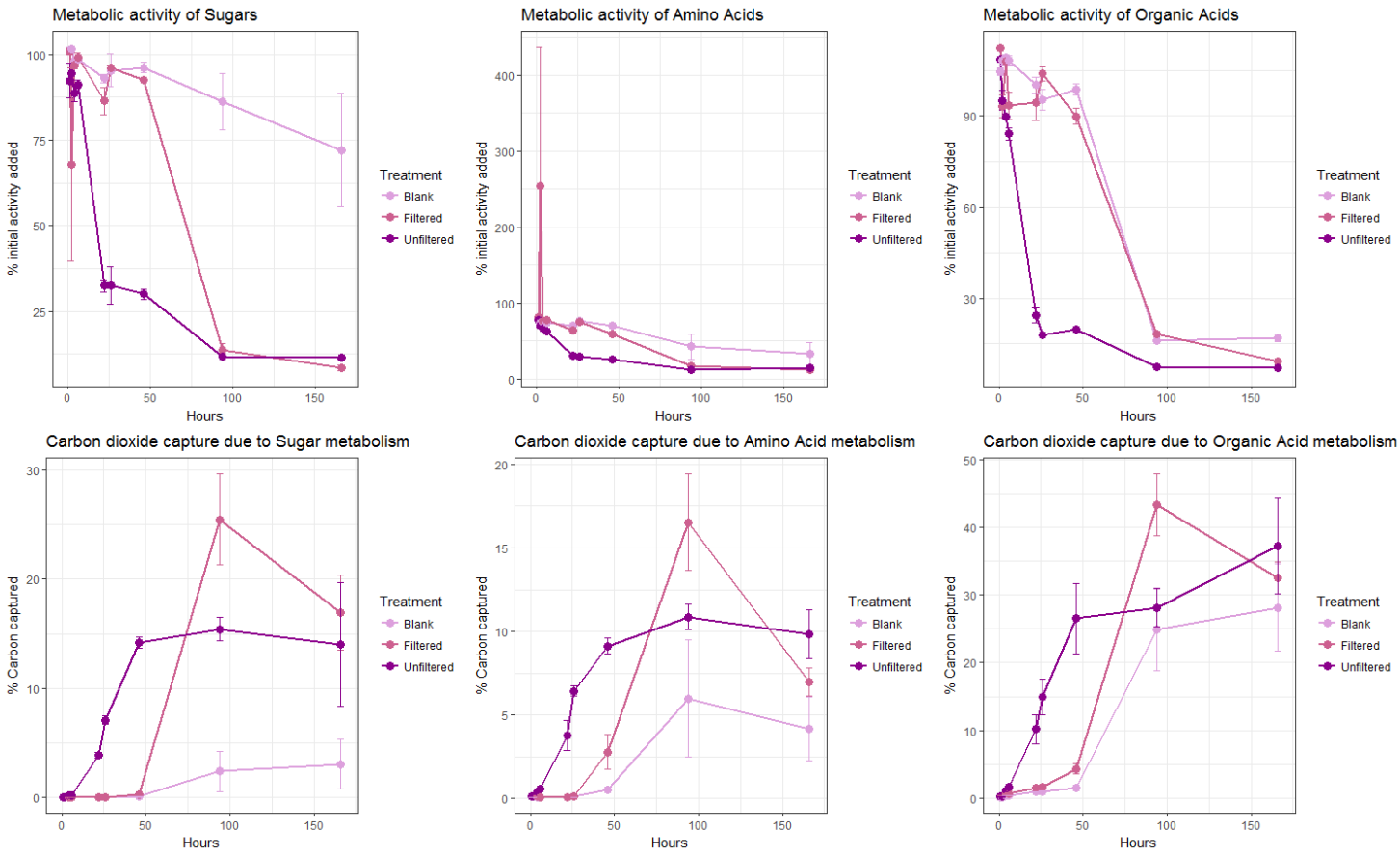


Figure A1.2-Depletion and CO_2 uptake measurements from November 2017 collection. These graphs indicate the activity of radioactivity in overall percent used or trapped. Values represent means ± standard error (n = 3).

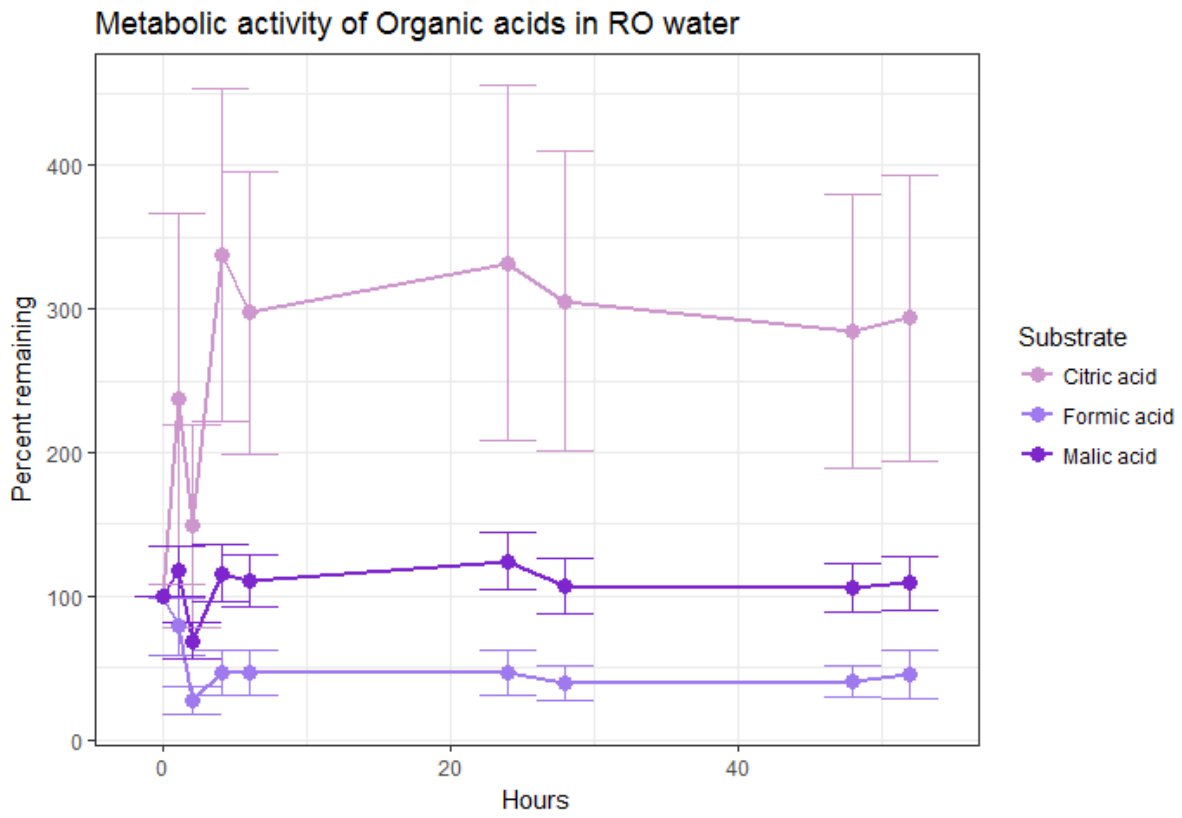


Figure A1.3- Depletion measurements from the contamination experiment. These graphs indicate the activity of radioactivity in overall percent remaining in the flask. Values represent means ± standard error (n = 3).

A1.2. Conclusion

The trends from this experiment match our previous observations over a longer time period. However, we cannot confirm whether or not perceived metabolic activity from the filtered fraction was from native microorganisms or from contaminants. It was initially suspected that the main source might be from the radiation laboratory, as this was not originally intended for microbial work. We changed the location to a more isolated environment as well as the use of ultra-filtered substrate solutions. However, this proved to be ineffective, as there was still levels of contamination after 48 hours. We suspect that the sodium hydroxide traps was the primary source. The contamination experiment demonstrated that the isotopes and the LC-MS water were not the source. Due to this, we decided not to explore the unlabelled samples, as these would not reflect the true nature of filterable species. Additionally, the resultant metagenomics and metabolomics would also show high levels of inaccuracy via the unidentified contaminant.

Despite this, the results presented indicate some differences between the filtered and unfiltered fractions. Yet, due to the obvious contamination in the controls, it is impossible to conclude that the metabolic activity was due, in fact, to the filterable microorganism population. Contamination is a serious issue and can come from many sources. Overall, these experiments help us identify areas of improvement and to help better streamline experimental procedures. The experiments following this were changed to include sterilization of the LC-MS water in the form of autoclaving and regular replacement of the NaOH traps with sterile containers and sterile sodium hydroxide solution.

Appendix II

Supplemental Material for Chapter 6

Table S5.1- Mauchly's test and Greenhouse-Geisser test for sphericity. Mauchy's W, p-value, and ε value are listed here. * Huyn-Feldt will need to be applied

				ε value	
Measurement	***Substrate***	***Mauchly's W***	***p-value***	***Greenhouse-Geisser***	***Huynh-Feldt***
^{14}C Substrate Depletion	^{14}C Amino acids	0.614	0.003	0.722	1.000
^{14}C Substrate Depletion	^{14}C Organic acids	0.301	<0.001	0.589	0.959
^{14}C Substrate Depletion	^{14}C Sugars	0.137	<0.001	0.537	0.800
$^{14}CO_2$ production	^{14}C Amino acids	0.773*	0.045	0.815	1.000
$^{14}CO_2$ production	^{14}C Organic acids	0.222	<0.001	0.562	0.895
$^{14}CO_2$ production	^{14}C Sugars	0.519	<0.001	0.675	1.000
^{14}C Biomass incorporation	^{14}C Amino acids	0.913*	0.336	0.920	1.000
^{14}C Biomass incorporation	^{14}C Organic acids	0.559	0.003	0.694	1.000
^{14}C Biomass incorporation	^{14}C Sugars	0.141	<0.001	0.538	0.802

Table S5.2- Forward and reverse primers used. Primers were produced by Eurofins Genomics.

Primer name	*Sequence (5' to 3')*
F515DI2	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TGCAGATCCAAC GTGBCAGCMGCCGCGGTAA
F515DI3	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CCATCACATAGG GTGBCAGCMGCCGCGGTAA
F515DI4	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTGGTATGGGAG T GTGBCAGCMGCCGCGGTAA
F515DI5	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ACTTTAAGGGTG T GTGBCAGCMGCCGCGGTAA
F515DI6	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GAGCAACATCCT T GTGBCAGCMGCCGCGGTAA
F515DI7	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TGTTGCGTTTCT GT GTGBCAGCMGCCGCGGTAA
F515DI8	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ATGTCCGACCAA GT GTGBCAGCMGCCGCGGTAA
F515DI9	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT AGGTACGCAATT GT GTGBCAGCMGCCGCGGTAA
F515DI10	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ACAGCCACCCAT CGA GTGBCAGCMGCCGCGGTAA
F515DI11	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TGTCTCGCAAGC CGA GTGBCAGCMGCCGCGGTAA
F515DI12	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GAGGAGTAAAGC CGA GTGBCAGCMGCCGCGGTAA
F515DI13	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTTACGTGGTTG ATGA GTGBCAGCMGCCGCGGTAA
F515DI14	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TACCGCCTCGGA ATGA GTGBCAGCMGCCGCGGTAA
F515DI15	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CGTAAGATGCCT ATGA GTGBCAGCMGCCGCGGTAA
F515DI16	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TACCGGCTTGCA TGCGA GTGBCAGCMGCCGCGGTAA
F515DI17	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ATCTAGTGGCAA TGCGA GTGBCAGCMGCCGCGGTAA
F515DI18	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CCAGGGACTTCT TGCGT GTGBCAGCMGCCGCGGTAA
F515DI19	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CACCTTACCTTA GAGTGG GTGBCAGCMGCCGCGGTAA
F515DI24	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GAGACTATATGC CCTGGAG GTGBCAGCMGCCGCGGTAA
R806DI2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT TGCAGATCCAAC GGACTACHVGGGTWTCTAAT
R806DI3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT CCATCACATAGG GGACTACHVGGGTWTCTAAT
R806DI4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT GTGGTATGGGAG A GGACTACHVGGGTWTCTAAT
R806DI5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT ACTTTAAGGGTG A GGACTACHVGGGTWTCTAAT
R806DI6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGCAACATCCT A GGACTACHVGGGTWTCTAAT

Table S5.3- Summary of the t-testing for COG distribution for targeted metabolomics compounds. Values were calculated from the relative abundance peak intensity of the GC system. Values for the initial 0 h and the final 506 h time point were used to check for utilization of each of the 16 compounds (n=3). P-values are reported where red indicates p <0.05, green is p < 0.01, and yellow is p <0.001.

	Filtered	***Unfiltered***
Compound	***p-value***	***p-value***
Alanine	0.01	0.075
Aspartic acid	0.007	0.244
Glycine	0.006	< 0.001
Isoleucine	0.001	0.008
Leucine	0.001	0.075
Phenylalanine	0.001	0.012
Proline	0.004	0.033
Serine	0.005	0.044
Threonine	0	0.046
Tyrosine	0.012	0.004
Valine	0.003	0.015
Fructose	< 0.001	0.12
Glucose	0.001	0.109
Sucrose	0	0.147
Citric acid	< 0.001	< 0.001
Malic acid	< 0.001	0.006

Table S5.4- Summary of the t-testing for COG distribution for both fractions and substrate addition. Results use to determine the effects of experiment duration, treatment (Filtered vs unfiltered), and substrate addition on the distribution of COG functional categories. P-values are reported where red indicates p <0.05, green is p < 0.01, and yellow is p <0.001. NA denotes no value.

	Filtered		***Unfiltered***	
COG Category	***w/ substrates***	***w/o substrates***	***w/ substrates***	***w/o substrates***
Energy production and conversion	0.416	0.179	0.864	0.253
Cell cycle control, cell division, chromosome partitioning	0.221	0.281	0.356	0.218
Amino acid transport and metabolism	0.047	0.019	0.125	0.912
Nucleotide transport and metabolism	0.182	0.017	0.153	0.394
Carbohydrate transport and metabolism	0.698	0.327	0.262	0.364
Coenzyme transport and metabolism	0.015	0.025	0.807	0.551
Lipid transport and metabolism	0.121	0.060	0.233	0.912
Translation, ribosomal structure and biogenesis	0.003	0.005	0.918	0.576
Transcription	0.012	0.009	0.412	0.398
Replication, recombination and repair	0.365	0.218	0.421	0.324
Cell wall/membrane/envelope biogenesis	0.020	0.061	0.823	0.155
Cell motility	0.029	0.018	0.869	0.958
Posttranslational modification, protein turnover, chaperones	0.011	0.056	0.041	0.408
Inorganic ion transport and metabolism	0.001	0.004	0.702	0.529
Secondary metabolites biosynthesis, transport and catabolism	0.010	0.103	0.135	0.443
Signal transduction mechanisms	0.036	0.004	0.589	0.313
Intracellular trafficking, secretion, and vesicular transport	0.463	0.062	0.388	0.546
Defence mechanisms	NA	NA	NA	NA
RNA processing and modification	0.115	0.054	0.117	0.598
Chromatin structure and dynamics	0.506	0.190	0.691	0.695
General function prediction only	0.599	0.997	0.058	0.403
Extracellular structures	0.125	0.122	0.481	0.305
Mobilome: prophages, transposons	NA	NA	NA	NA
Nuclear structure	NA	NA	NA	NA
Cytoskeleton	0.350	0.018	0.286	0.053
Function unknown	0.040	0.037	0.517	0.566

Table S5.5- Summary of the repeated measures ANOVA. Results use to determine the effects of filtering treatment and time on the use of 3 different ^{14}C-labelled substrates by river water microbial communities. F-values, p-values, and degrees of freedom (df), are reported based on correction from the Maulchy's test of sphericity (Table S5.1).

Measurement	*Substrate*	*Effect*	*Correction applied*	*df*	*F-value*	*p-value*
^{14}C Substrate Depletion	^{14}C Amino acids	Treatment	Greenhouse-Geisser	1.444	267.258	<0.001
^{14}C Substrate Depletion	^{14}C Amino acids	Treatment x Duration	Greenhouse-Geisser	17.322	13.085	<0.001
^{14}C Substrate Depletion	^{14}C Organic acids	Treatment	Greenhouse-Geisser	1.177	73.700	<0.001
^{14}C Substrate Depletion	^{14}C Organic acids	Treatment x Duration	Greenhouse-Geisser	14.126	4.381	0.001
^{14}C Substrate Depletion	^{14}C Sugars	Treatment	Greenhouse-Geisser	1.074	217.896	<0.001
^{14}C Substrate Depletion	^{14}C Sugars	Treatment x Duration	Greenhouse-Geisser	12.884	19.531	<0.001
$^{14}CO_2$ production	^{14}C Amino acids	Treatment	Huynh-Feldt	2.000	143.215	<0.001
$^{14}CO_2$ production	^{14}C Amino acids	Treatment x Duration	Huynh-Feldt	24.000	13.701	<0.001
$^{14}CO_2$ production	^{14}C Organic acids	Treatment	Greenhouse-Geisser	1.125	105.501	<0.001
$^{14}CO_2$ production	^{14}C Organic acids	Treatment x Duration	Greenhouse-Geisser	13.497	7.175	<0.001
$^{14}CO_2$ production	^{14}C Sugars	Treatment	Greenhouse-Geisser	1.351	326.369	<0.001
$^{14}CO_2$ production	^{14}C Sugars	Treatment x Duration	Greenhouse-Geisser	16.210	43.863	<0.001
^{14}C Biomass Incorporation	^{14}C Amino acids	Treatment	Greenhouse-Geisser	1.840	12.317	<0.001
^{14}C Biomass Incorporation	^{14}C Amino acids	Treatment x Duration	Greenhouse-Geisser	22.084	6.429	<0.001
^{14}C Biomass Incorporation	^{14}C Organic acids	Treatment	Huynh-Feldt	2.000	82.803	<0.001
^{14}C Biomass Incorporation	^{14}C Organic acids	Treatment x Duration	Huynh-Feldt	24.000	6.126	<0.001
^{14}C Biomass Incorporation	^{14}C Sugars	Treatment	Huynh-Feldt	1.604	140.036	<0.001
^{14}C Biomass Incorporation	^{14}C Sugars	Treatment x Duration	Huynh-Feldt	19.242	20.388	<0.001

Table S5.6- Overview of 16SrRNA amplicon sequencing samples. Samples are labelled via their number and letter designations. Signal refers to the presence of a band of a 1.8% agarose gel. The number of reads retrieved per sample are also listed. *Negative controls. **Samples that produced no signal and were not included in subsequent analysis. † False positive result. NA (not applicable) indicates inability to sequence samples, even though there was a positive signal. ††Inability to sequence the sample, even with a positive signal after primers.

Sample ID	*Hours after addition*	*Treatment*	*Substrate Addition*	*Signal*	*Meta-barcode ID*	*Number of Reads*
28	0	Filtered	+	+	Bangor.L.28.NA	241
29	0	Filtered	+	+	Bangor.L.29.NA	577
30	0	Filtered	+	+	Bangor.L.30.NA	560
34	0	Unfiltered	+	+	Bangor.L.34.NA	10,161
35	0	Unfiltered	+	+	Bangor.L.35.NA	5,558
36	0	Unfiltered	+	+	Bangor.L.36.NA	6,963
40*†	0	Blank	+	+	NA††	NA
41*	0	Blank	+	-		NA
42*	0	Blank	+	-		NA
28g**	49	Filtered	+	-		NA
29g	49	Filtered	+	+	Bangor.L.29.g	5
30g	49	Filtered	+	+	Bangor.L.30.g	909
31g	49	Filtered	-	+	Bangor.L.31.g	782
32g**	49	Filtered	-	-		NA
33g	49	Filtered	-	+	Bangor.L.33.g	1,755
34g	49	Unfiltered	+	+	Bangor.L.34.g	18,361
35g	49	Unfiltered	+	+	Bangor.L.35.g	17,286
36g	49	Unfiltered	+	+	Bangor.L.36.g	11,079
37g	49	Unfiltered	-	+	Bangor.L.37.g	37,037
38g	49	Unfiltered	-	+	Bangor.L.38.g	33,406
39g	49	Unfiltered	-	+	Bangor.L.39.g	25,473
28i	141	Filtered	+	+	Bangor.L.28.i	19,519
29i	141	Filtered	+	+	Bangor.L.29.i	18,969
30i	141	Filtered	+	+	Bangor.L.30.i	11,165
31i	141	Filtered	-	+	Bangor.L.31.i	16,445
32i	141	Filtered	-	+	Bangor.L.32.i	28,145
33i	141	Filtered	-	+	Bangor.L.33.i	26,989
34i	141	Unfiltered	+	+	Bangor.L.34.i	19,473
35i	141	Unfiltered	+	+	Bangor.L.35.i	18,626
36i	141	Unfiltered	+	+	Bangor.L.36.i	18,241
37i**	141	Unfiltered	-	-		NA
38i	141	Unfiltered	-	+	Bangor.L.38.i	15,668
39i	141	Unfiltered	-	+	Bangor.L.39.i	18,883
28k	333	Filtered	+	+	Bangor.L.28.k	24,025
29k	333	Filtered	+	+	Bangor.L.29.k	16,516
30k	333	Filtered	+	+	Bangor.L.30.k	24,453

31k	333	Filtered	-	+	Bangor.L.31.k		24,763
32k	333	Filtered	-	+	Bangor.L.32.k		27,545
33k	333	Filtered	-	+	Bangor.L.33.k		17,912
34k	333	Unfiltered	+	+	Bangor.L.34.k		13,228
35k	333	Unfiltered	+	+	Bangor.L.35.k		13,275
36k	333	Unfiltered	+	+	Bangor.L.36.k		12,981
37k	333	Unfiltered	-	+	Bangor.L.37.k		18,043
38k	333	Unfiltered	-	+	Bangor.L.38.k		16,158
39k	333	Unfiltered	-	+	Bangor.L.39.k		14,267
28L	506	Filtered	+	+	Bangor.L.28.L		4,096
29L	506	Filtered	+	+	Bangor.L.29.L		4,906
30L	506	Filtered	+	+	Bangor.L.30.L		9,605
31L	506	Filtered	-	+	Bangor.L.31.L		10,439
32L	506	Filtered	-	+	Bangor.L.32.L		6,882
33L	506	Filtered	-	+	Bangor.L.33.L		9,041
34L	506	Unfiltered	+	+	Bangor.L.34.L		1,380
35L	506	Unfiltered	+	+	Bangor.L.35.L		5,071
36L	506	Unfiltered	+	+	Bangor.L.36.L		2,927
37L	506	Unfiltered	-	+	NA††	NA	
38L	506	Unfiltered	-	+	Bangor.L.38.L		3,261
39L	506	Unfiltered	-	+	Bangor.L.39.L		5,355
40L*	506	Blank	+	-		NA	
41L*	506	Blank	+	-		NA	
42L*	506	Blank	+	-		NA	

Table S5.7- OTU taxonomy assignment in both fractions across all time points with and without addition. This table is the supplementary material for figures 5.2-5.4 indicating classification from phyla to family level. Note: unassigned indicates that there is currently no name associated with that taxa level.

Phylum	***Class***	***Order***	***Family***
"*Candidatus* Dependentiae" (Candidate phylum TM6)	SJA-4	Unassigned	Unassigned
"*Candidatus* Omnitrophica" (Candidate phylum OP3)	koll11	GIF10	kpj58rc
"*Candidatus* Omnitrophica" (Candidate phylum OP3)	BD4-9	Unassigned	Unassigned
"*Candidatus* Omnitrophica" (Candidate phylum OP3)	PBS-25	Unassigned	Unassigned
"*Candidatus* Parcubacteria" (Candidate phylum OD1)	SM2F11	Unassigned	Unassigned
"*Candidatus* Parcubacteria" (Candidate phylum OD1)	ZB2	Unassigned	Unassigned
Acidobacteria	*Acidobacteriia*	*Acidobacteriales*	*Acidobacteriaceae*
Acidobacteria	*Chloracidobacteria*	RB41	Ellin6075
Acidobacteria	*Acidobacteriia*	*Acidobacteriales*	*Koribacteraceae*
Acidobacteria	*Acidobacteria*-6 (Division 6)	iii1-15	mb2424
Acidobacteria	*Acidobacteria*-6 (Division 6)	iii1-15	RB40
Acidobacteria	*Solibacteres*	*Solibacterales*	*Solibacteraceae*
Acidobacteria	*Acidobacteria*-6 (Division 6)	CCU21	Unassigned
Acidobacteria	*Acidobacteria*-6 (Division 6)	iii1-15	Unassigned
Acidobacteria	DA052	Ellin6513	Unassigned
Acidobacteria	*Solibacteres*	*Solibacterales*	Unassigned
Actinobacteria	*Actinobacteria*	*Actinomycetales*	ACK-M1
Actinobacteria	*Thermoleophilia*	*Solirubrobacterales*	*Conexibacteraceae*
Actinobacteria	*Coriobacteriia*	*Coriobacteriales*	*Coriobacteriaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Frankiaceae*
Actinobacteria	*Thermoleophilia*	*Gaiellales*	*Gaiellaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Microbacteriaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Mycobacteriaceae*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*
Actinobacteria	*Acidimicrobiia*	*Acidimicrobiales*	Unassigned
Actinobacteria	*Actinobacteria*	*Actinomycetales*	Unassigned
Actinobacteria	*Thermoleophilia*	*Solirubrobacterales*	Unassigned
Armatimonadetes	*Armatimonadia*	*Armatimonadales*	*Armatimonadaceae*
Armatimonadetes	*Fimbriimonadia*	*Fimbriimonadales*	*Fimbriimonadaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	*Bacteroidaceae*
Bacteroidetes	*Saprospirae*	*Saprospirales*	*Chitinophagaceae*
Bacteroidetes	*Flavobacteriia*	*Flavobacteriales*	*Cryomorphaceae*
Bacteroidetes	*Cytophagia*	*Cytophagales*	*Cytophagaceae*
Bacteroidetes	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	p-2534-18B5
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	*Paraprevotellaceae*

Bacteroidetes	*Bacteroidia*	*Bacteroidales*	*Porphyromonadaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	RF16
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	*Rikenellaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	S24-7
Bacteroidetes	*Saprospirae*	*Saprospirales*	*Saprospiraceae*
Bacteroidetes	*Sphingobacteriia*	*Sphingobacteriales*	*Sphingobacteriaceae*
Bacteroidetes	*Bacteroidia*	*Bacteroidales*	Unassigned
Bacteroidetes	BME43	Unassigned	Unassigned
Bacteroidetes	*Saprospirae*	*Saprospirales*	Unassigned
Bacteroidetes	*Sphingobacteriia*	*Sphingobacteriales*	Unassigned
Bacteroidetes	*Flavobacteriia*	*Flavobacteriales*	*Weeksellaceae*
Candidate phylum FCPU426	Unassigned	Unassigned	Unassigned
Candidate phylum WPS-2	Unassigned	Unassigned	Unassigned
Chlamydiae	*Chlamydiia*	*Chlamydiales*	*Rhabdochlamydiaceae*
Chloroflexi	*Anaerolineae*	SBR1031	A4b
Chloroflexi	*Anaerolineae*	H39	Unassigned
Chloroflexi	*Anaerolineae*	WCHB1-50	Unassigned
Chloroflexi	Ellin6529	Unassigned	Unassigned
Crenarchaeota	*Thaumarchaeota*	*Cenarchaeales*	SAGMA-X
Cyanobacteria	*Chloroplast*	*Stramenopiles*	Unassigned
Cyanobacteria	*Chloroplast*	*Streptophyta*	Unassigned
Deinococcus-Thermus	*Deinococci*	*Deinococcales*	*Deinococcaceae*
Elusimicrobia	*Elusimicrobia*	*Elusimicrobiales*	Unassigned
Elusimicrobia	*Elusimicrobia*	FAC88	Unassigned
Euryarchaeota	*Methanobacteria*	*Methanobacteriales*	*Methanobacteriaceae*
Euryarchaeota	*Methanomicrobia*	*Methanomicrobiales*	*Methanocorpusculaceae*
Euryarchaeota	*Thermoplasmata*	E2	*Methanomassiliicoccaceae*
Fibrobacteres	*Fibrobacteria*	*Fibrobacterales*	*Fibrobacteraceae*
Fibrobacteres	*Fibrobacteria*	258ds10	Unassigned
Firmicutes	*Bacilli*	*Lactobacillales*	*Aerococcaceae*
Firmicutes	*Bacilli*	*Bacillales*	*Bacillaceae*
Firmicutes	*Bacilli*	*Lactobacillales*	*Carnobacteriaceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Clostridiaceae*
Firmicutes	*Erysipelotrichi*	*Erysipelotrichales*	*Erysipelotrichaceae*
Firmicutes	*Bacilli*	*Gemellales*	*Gemellaceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Lachnospiraceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Mogibacteriaceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Peptostreptococcaceae*
Firmicutes	*Bacilli*	*Bacillales*	*Planococcaceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Ruminococcaceae*
Firmicutes	*Bacilli*	*Bacillales*	*Staphylococcaceae*
Firmicutes	*Bacilli*	*Lactobacillales*	*Streptococcaceae*
Firmicutes	*Clostridia*	*Clostridiales*	*Tissierellaceae*
Firmicutes	*Bacilli*	*Turicibacterales*	*Turicibacteraceae*
Firmicutes	*Bacilli*	*Bacillales*	Unassigned
Firmicutes	*Clostridia*	*Clostridiales*	Unassigned
Firmicutes	*Clostridia*	*Clostridiales*	*Veillonellaceae*
Gemmatimonadetes	*Gemmatimonadetes*	*Gemmatimonadales*	*Gemmatimonadaceae*

Gemmatimonadetes	Gemm-1	Unassigned	Unassigned
Gemmatimonadetes	*Gemmatimonadetes*	KD8-87	Unassigned
Nitrospirae	*Nitrospira*	*Nitrospirales*	*Nitrospiraceae*
Planctomycetes	*Planctomycetia*	*Gemmatales*	*Gemmataceae*
Planctomycetes	*Planctomycetia*	*Gemmatales*	*Isosphaeraceae*
Planctomycetes	*Planctomycetia*	*Planctomycetales*	*Planctomycetaceae*
Planctomycetes	vadinHA49	DH61	Unassigned
Proteobacteria	*Gammaproteobacteria*	*Alteromonadales*	125ds10
Proteobacteria	*Gammaproteobacteria*	*Alteromonadales*	211ds20
Proteobacteria	*Alphaproteobacteria*	*Rhodospirillales*	*Acetobacteraceae*
Proteobacteria	*Gammaproteobacteria*	*Aeromonadales*	*Aeromonadaceae*
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	*Alcaligenaceae*
Proteobacteria	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Beijerinckiaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Bradyrhizobiaceae*
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	*Burkholderiaceae*
Proteobacteria	*Epsilonproteobacteria*	*Campylobacterales*	*Campylobacteraceae*
Proteobacteria	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	*Comamonadaceae*
Proteobacteria	*Gammaproteobacteria*	*Legionellales*	*Coxiellaceae*
Proteobacteria	*Gammaproteobacteria*	*Methylococcales*	*Crenotrichaceae*
Proteobacteria	*Deltaproteobacteria*	*Desulfobacterales*	*Desulfobacteraceae*
Proteobacteria	*Deltaproteobacteria*	*Desulfovibrionales*	*Desulfovibrionaceae*
Proteobacteria	*Betaproteobacteria*	A21b	EB1003
Proteobacteria	*Gammaproteobacteria*	*Enterobacteriales*	*Enterobacteriaceae*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Erythrobacteraceae*
Proteobacteria	*Betaproteobacteria*	*Gallionellales*	*Gallionellaceae*
Proteobacteria	*Gammaproteobacteria*	*Oceanospirillales*	*Halomonadaceae*
Proteobacteria	*Epsilonproteobacteria*	*Campylobacterales*	*Helicobacteraceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Hyphomicrobiaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Hyphomonadaceae*
Proteobacteria	*Deltaproteobacteria*	Sva0853	JTB36
Proteobacteria	*Deltaproteobacteria*	NB1-j	JTB38
Proteobacteria	*Gammaproteobacteria*	*Legionellales*	*Legionellaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Methylobacteriaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*
Proteobacteria	*Betaproteobacteria*	*Methylophilales*	*Methylophilaceae*
Proteobacteria	*Gammaproteobacteria*	*Pseudomonadales*	*Moraxellaceae*
Proteobacteria	*Betaproteobacteria*	*Neisseriales*	*Neisseriaceae*
Proteobacteria	*Betaproteobacteria*	*Nitrosomonadales*	*Nitrosomonadaceae*
Proteobacteria	*Gammaproteobacteria*	*Alteromonadales*	OM60
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	*Oxalobacteraceae*
Proteobacteria	*Gammaproteobacteria*	*Pasteurellales*	*Pasteurellaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Phyllobacteriaceae*
Proteobacteria	*Gammaproteobacteria*	*Thiotrichales*	*Piscirickettsiaceae*
Proteobacteria	*Deltaproteobacteria*	*Myxococcales*	*Polyangiaceae*
Proteobacteria	*Betaproteobacteria*	*Procabacteriales*	*Procabacteriaceae*
Proteobacteria	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*

Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*
Proteobacteria	*Betaproteobacteria*	*Rhodocyclales*	*Rhodocyclaceae*
Proteobacteria	*Alphaproteobacteria*	*Rhodospirillales*	*Rhodospirillaceae*
Proteobacteria	*Alphaproteobacteria*	*Rickettsiales*	*Rickettsiaceae*
Proteobacteria	*Gammaproteobacteria*	*Xanthomonadales*	*Sinobacteraceae*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Proteobacteria	*Alphaproteobacteria*	Ellin329	Unassigned
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	Unassigned
Proteobacteria	*Alphaproteobacteria*	*Rickettsiales*	Unassigned
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	Unassigned
Proteobacteria	*Alphaproteobacteria*	Unassigned	Unassigned
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	Unassigned
Proteobacteria	*Betaproteobacteria*	Ellin6067	Unassigned
Proteobacteria	*Betaproteobacteria*	IS-44	Unassigned
Proteobacteria	*Betaproteobacteria*	MND1	Unassigned
Proteobacteria	*Betaproteobacteria*	SC-I-84	Unassigned
Proteobacteria	*Betaproteobacteria*	Unassigned	Unassigned
Proteobacteria	*Deltaproteobacteria*	MIZ46	Unassigned
Proteobacteria	*Deltaproteobacteria*	*Myxococcales*	Unassigned
Proteobacteria	*Deltaproteobacteria*	*Spirobacillales*	Unassigned
Proteobacteria	*Deltaproteobacteria*	Unassigned	Unassigned
Proteobacteria	*Gammaproteobacteria*	*Chromatiales*	Unassigned
Proteobacteria	*Gammaproteobacteria*	*Legionellales*	Unassigned
Proteobacteria	*Gammaproteobacteria*	Unassigned	Unassigned
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Xanthobacteraceae*
Proteobacteria	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*
Spirochaetes	*Spirochaetes*	*Spirochaetales*	*Spirochaetaceae*
Synergistetes	*Synergistia*	*Synergistales*	*Synergistaceae*
Tenericutes	*Mollicutes*	*Acholeplasmatales*	*Acholeplasmataceae*
Tenericutes	RF3	ML615J-28	Unassigned
Unassigned	*Unassigned*	Unassigned	Unassigned
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	auto67_4W
Verrucomicrobia	*Opitutae*	*Cerasicoccales*	*Cerasicoccaceae*
Verrucomicrobia	*Opitutae*	Unassigned	*Cerasicoccaceae*
Verrucomicrobia	*Spartobacteria*	*Chthoniobacterales*	*Chthoniobacteraceae*
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	Ellin515
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	Ellin517
Verrucomicrobia	*Opitutae*	*Opitutales*	*Opitutaceae*
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	*Pedosphaeraceae*
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	R4-41B
Verrucomicrobia	*Methylacidiphilae*	*Methylacidiphilales*	Unassigned
Verrucomicrobia	*Pedosphaerae*	*Pedosphaerales*	Unassigned
Verrucomicrobia	*Verrucomicrobiae*	*Verrucomicrobiales*	*Verrucomicrobiaceae*

Table S5.8-Overview of shotgun sequencing across initial (0 h), 141 h, and 506 h in both filtered and unfiltered fractions with and without nutrient amendments. Sample ID, time point taken, treatment (filtered vs unfiltered) and substrate addition (NA, +/-). The number of contigs coding sequences (CDS), annotations, and COG assignments. NA indicates that this is the initial community at the beginning of the experiment.

Sample ID	*Time point*	*Treatment*	*Substrate addition*	*Contigs*	*CDS*	*Annotations*	*COG assignations*
29F	initial	Filtered	NA	55448	65137	25084	24228
28L	506	Filtered	+	82874	75794	26193	44711
31L	506	Filtered	-	56595	63879	44306	44877
28i	141	Filtered	+	46501	38146	18021	18296
31i	141	Filtered	-	30881	17982	7946	8100
U	initial	Unfiltered	NA	77690	31847	14813	14841
36L	506	Unfiltered	+	43038	24158	1191	893
37L	506	Unfiltered	-	47339	18946	6936	6828
35i	141	Unfiltered	+	71461	41921	21821	22401
38i	141	Unfiltered	-	144906	85698	43310	46063
28F	initial	Filtered	NA	74187	112132	38959	37031
29L	506	Filtered	+	107976	80606	44657	45096
32L	506	Filtered	-	82602	106232	77879	78659
29i	141	Filtered	+	76921	79720	53277	54579
32i	141	Filtered	-	68087	69763	45651	46643
34U	initial	Unfiltered	NA	240978	122695	60376	60269
35L	506	Unfiltered	+	25762	13018	2210	2142
39L	506	Unfiltered	-	80895	32200	6559	6559
36i	141	Unfiltered	+	256700	181140	97113	99402
37i	141	Unfiltered	-	182931	107126	59150	60379