Analysis of dynamics and interactions within planktonic microbial communities in Lake Constance by Next-Generation Sequencing

Doctoral thesis for obtaining the

academic degree Doctor of Natural Sciences

(Dr. rer. nat.)

submitted by

Corentin Fournier

at the

Faculty of Science

Department of Biology

Konstanz, 2023

Date of oral examination: 27.10.2023

1. Referent: Prof. Dr. David Schleheck

2. Referentin: Prof. Dr. Laura Epp

3. Referent: Prof. Dr. Peter kroth

“To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of”

Ronald Fisher

“Papers are increasingly not reproducible [….]. Getting a splashy result passed the reviewers is what makes your career, not having a reputation for reproducibility and detail-oriented work.”

C. Titus Brown

“*First step in solving any problem is recognizing there is one*”

Will McAvoy

“Not all those who wander are lost”

Bilbo Baggins

Acknowledgements

First of all, I would like to express my deep gratitude to Prof. Dr. David Schleheck and Prof. Dr. Peter Kroth for accepting me as a PhD student even though I was completely naïve in the ecosystems and the techniques. I will never forget a cigarette break during the application process when David told me he was going to support my application because he thought I could bring an interesting perspective to this work. It allowed me to discover a whole new world and to develop skills and new ways of thinking.

Thanks to Jun.-Prof. Dr. Laura Epp. Your knowledge of sequencing, bioinformatics and biostatistics, coupled with great motivation and always good humour, pushed my understanding of these fields, how to use them and, most importantly, how to present and write them.

Thanks to Alfred Sulger, Angelika Seifried, Pia Mahler and Josef Halder for taking care of the routine field trip. These little adventures every two weeks on the lake were each time exciting: once hot and sunny, other in the middle of winter, surrounded by ice or full of VTK students looking for a slight bit of heat.

Thanks to Julia Schmidt for continuing the work beyond my PhD: field trip, DNA extraction and DNA library preparation so my work can be included in larger studies. Thanks too for being patient with me while I was teaching you the techniques and to go along with my obsession for sterility.

Thanks to Dr. Paolo Franchini who have been my entry point into the fascinating world of bioinformatics which I will never leave!

To all my students: Alexander Fiedler, Annika Hepp, Diana Kostenko, Stefanie Dörr and Lea-Sophie Pfau. Thanks for helping me and going along with my work overexcitement that could results in some “slightly” long days of work.

Prof. Dr. David Schleheck thank you again. Your happiness, curiosity and excitation in discovering new things (like coding an Arduino program from scratch!) have been a driving force for me to explore, learn and do the best during my PhD.

Franziska Klotz, Franziska, Franzi, where do I start as there is so much to say. You have always been here, through laugh and cry, supportive and genuine besties :P. Our relation started by moving furnitures through windows and have continued as an

Acknowledgements

amazing adventure comprised of lonely long nights of filtration, diving, party, building snowman (with legitimate genitalia), party again, discussion in our fortress of solitude called office, covid birthday/party, your PhD defence, etc etc. I would not change anything!

Adrian and Enzo, the full package! What else to say that it has been an honour to work and live alongside you all these years. All these parties in Rosmarin, Kolbenfresser and wherever we could do were pivotal in the life of so many PhD students. From the tchou tchou to DJ Adrien. You were always full of energy and happy, ready to help when lost in the middle of the night and in need of a bed. Enzo and our bouldering session by whatever weather (yes even when it was way too cold!) were a blast!

Joana, with Anna, Anja and Benjamin, you were the old wise PhD students here to teach us, the newcomers. What a lively lab it was and the best way of making the starting of our PhD easy and smooth. I still have some grudge, like for example (and chosen randomly) a certain session of waxing Joana, I didn´t forget!

Or course I want to thanks Eva Riehle. I remember when we started the cooperation, which turned into a magnificent study (first paper as first author for both of us woohoo !!!). What an adventure it was! You are an amazing person and your kindness, motivation and desire for adventure set an example for everyone.

Obviously, everyone affiliated to AG Schleheck and AG Kroth. Thank you for being here during parties, discussion or during my daily, very needed, social walk on M9. Our bike trips, hikes, social gathering are fond and precious memories.

Thanks to all RTG members. It has been a marvellous adventure with all of you.

A special thanks to Tina Romer who has been the backbone of the RTG and each and every one of the students. The German bureaucracy has been much easier to handle thanks to your knowledge, always good mood and willingness to help us.

Thank you of course to everyone that I had the chance to meet, I don´t forget you!

Finally, I want to thanks my family: Edith, Ambre and Clair. You have supported me in every way: through my emotional roller coaster, teaching me valuable lessons, or through my passion for science that helped connecting together.

Table of Contents

Table of Contents 1

Summary 5

Zusammenfassung 7

List of Figures 9

List of Tables 11

Abbreviations 12

General introduction 14

The freshwater microbial plankton 14

Threat to the freshwater microbial plankton 15

Technologies to disentangle microbial plankton communities 15

Sequencing revolution 17

Bioinformatics: change of approach in biology 18

Biostatistics in microbial ecology 19

Lake Constance 20

Objectives 22

Chapter I Are toxin-producing *Planktothrix* spp. an emerging species in Lake Constance? 23

Abstract 25

Introduction 26

Results 28

Bloom of red-pigmented phytoplankton at water depths below the chlorophyll-a maximum in Lake Constance 28

Phycoerythrin-rich *Synechococcus* phylotypes dominated the DRM cyanobacterial community in 2019 and 2020, while *Planktothrix* spp. were detectable at low abundance 29

Synechococcus rubescens and *Cyanobium gracile* clusters in 2019 and 2020 33

Dynamics of *Synechococcus* ASVs in 2019 and 2020 35

Dynamics of *Planktothrix* and *Microcystis* ASVs, the abundance of microcystin biosynthesis genes and the concentration of microcystins in samples taken during 2019 and 2020. 37

Retrospective evaluation of depth profiles for the Lake-*Überlingen* routine sampling site 39

Discussion 41

Conclusions 44

Table of Contents

Materials and Methods 46
Sample collection 46
DNA extraction and PCR 46
RT-PCR/TNA 47
Sanger sequencing 48
Amplification and Illumina sequencing 48
Bioinformatics pipeline 48
Phylogenetic analysis 49
Statistical and network analysis 50
Toxin extraction and analysis 51
Evaluation of bbe Moldaenke FluoroProbe data 52

Chapter II Description of a 'plankton filtration bias' in sequencing-based community analysis and of an Arduino microcontroller-based flowmeter device that can help to resolve it 54
Abstract 55
Introduction 56
Materials and Methods 59
Study sites and sampling campaign 59
Filtration using a self-constructed Arduino-based monitoring device (flowmeter) 59
DNA extraction, PCR amplification and Illumina-amplicon sequencing 62
Bioinformatic pipeline 63
Biostatistics 64
Results 67
Impact of different filtration volumes on the bacterioplankton community composition observed by 16S rRNA-gene fragment amplicon sequencing 67
Alpha diversity 69
Beta diversity 69
Individual taxonomic groups affected by filtration volume 71
Normalization of the plankton filtration bias using our flowmeter device 72
Alpha and beta diversities 74
Individual taxonomic groups affected. 76
Discussion 78

Chapter III Nano- and pico-plankton succession in Upper and Lower Lake Constance followed by 18S and 16S rDNA amplicon sequencing: Same seasonal dynamics but different communities 83
Abstract 85
Introduction 86

Table of Contents

Materials and Methods 88
Study sites, sampling campaign and filtration procedure 88
Molecular work: DNA extraction, library preparation and sequencing 89
Bioinformatic pipeline 91
Biostatistics 92
Results 94
Plankton diversity dynamics 96
Taxonomic community composition 101
Unique ASVs in either ULC or LLC 102
ASVs detected for both ULC and LLC with different relative abundances 107
Discussion 112
Microbial plankton dynamics 112
ULC unique microbial plankton community 113
LLC unique microbial plankton community 116
Differentially abundant community 118
Conclusions 121

Chapter IV Seasonal dynamics of the microbial plankton community is determined through both environmental factors and biotic interactions 123
Abstract 124
Introduction 125
Materials and methods 127
Study sites, sampling campaign and filtration procedure 127
Molecular work: DNA extraction, library preparation and sequencing 128
Bioinformatic pipeline 130
Biostatistics 130
Results 133
General overview of the data and analyses 133
JSDM and Bayesian framework in microbial ecology 134
Dynamics of the of community composition across the year 135
Co-occurence networks of ASVs 138
Discussion 144
JSDM and Bayesian framework in microbial ecology 144
Temporal environmental factors shaping the microbial plankton community 144
ASV co-occurrence 147
Conclusions 150

Table of Contents

General discussion 152
Raising our knowledge of the sequencing technology 152
Bias introduced by biological variability 152
Bias introduced by methodological limitations 153
The dependence to bioinformatic and its problematic 154
A good analysis starts by a good data preparation 156
The nature of NGS datasets? 157
Limits of statistics and a way to overcome them? 158
Bayesian statistics: the future? 159
Concept of resilience and stability in microbial ecology 161

Conclusions 163

Future prospects 164

Record of achievement 165
Chapter I 165
Chapter II 165
Chapter III 165
Chapter IV 165

Supplementary Files 166
Chapter I: Are toxin-producing *Planktothrix* spp. an emerging species in Lake Constance? 166
Chapter II: Description of a “plankton filtration bias” and of an Arduino microcontroller-based flowmeter device that can help to resolve it 176
Chapter III: Nano- and pico-plankton succession in Upper and Lower Lake Constance followed by 18S and 16S rDNA amplicon sequencing: Same seasonal dynamics but different communities ... 182
Chapter IV: Lake Constance microbial plankton dynamic is heavily impacted by the vertical mixing and ASVs associations: a Joint species distribution modelling approach 200

References 217

Summary

The advent of Next-Generation Sequencing (NGS) is helping to reveal much of the vast, yet hidden biodiversity of microbial communities in various habitats on Earth. Using DNA extracted from environmental samples, researchers now have access to phylogenetic and functional information, and thus to composition, dynamics and abiotic and biotic interactions of the microorganisms that make up these communities. The amount of information accumulated is enormous and allows worldwide comparisons. However, the accurate and optimal processing of such large NGS datasets is still an intensive field in development.

Throughout this thesis, DNA sampling and NGS has been established and applied to the description of the microbial plankton community of the epilimnion of Lake Constance, while pushing scientific boundaries to optimise processing and accurately analysing these datasets.

Planktothrix rubescens, a red-pigmented toxin-producing cyanobacterial species has regularly formed blooms in Alpine and pre-Alpine lakes. Lake Constance, however, has remained largely unaffected by blooms of this toxic cyanobacterium. The observation of red biomass on plankton filters collected from summer onwards in 2019 and 2020 raised the suspicion whether it was prominently present also in Upper Lake Constance (*Überlinger See*). DNA extraction and NGS sequencing of cyanobacteria-specific 16S-rDNA amplicons as well as Sanger sequencing of the *mcy* gene for toxin production, confirmed the presence of *P. rubescens* in Lake Constance, though at very low abundance. Two types of microcystin toxins were also detectable, though at extremely low concentrations. Indeed, *Synechococcus* spp. were responsible for the high abundance of red-pigmented biomass, and our phylogenetic analysis linked them to Lake Constance isolates already described in 2003.

Plankton filtration for collecting biomass for DNA extraction is traditionally performed with a fixed sample volume, resulting in different amounts of biomass and thus DNA depending on the plankton density of the water samples. Further, this introduces a so-called ‘filtration bias’ into the NGS community analysis. A new filtration approach was developed, using flowrate as a proxy for biomass loading and clogging of the filters. The new method successfully collected equivalent amounts of DNA

regardless of plankton density, and the NGS results showed no ‘filtration bias’ when tested. This filtration method was used to collect plankton biomass for DNA extraction in the following study.

The composition of the nanoplankton (here, organisms with 180 – 5 µm in diameter) and picoplankon (5 – 0.2 µm in diameter) of the epilimnion of Upper and Lower Lake Constance was analysed every two weeks from March 2018 to March 2019. Therefore, DNA was extracted from plankton samples taken of the water column between 0 – 20 m depth (integrated sample) and 18S- and 16S-rDNA amplicons were sequenced by NGS. While the temporal dynamics of the microbial community showed a similar seasonal pattern for both sampling sites, the phylogenetic analysis revealed specific taxa that were more abundant or even uniquely present in either Upper or Lower Lake Constance, which could be attributed to specific environmental factors in these different parts of the lake. For example, in Upper Lake Constance, the vertical mixing in winter brought a hypolimnion community into the epilimnion, while for Lower Lake Constance, a hypoxic event during summer and autumn allowed also for anaerobic microorganisms to thrive. The presence of several taxa capable of mixotrophy or predation at either the one or the other sampling site, also indicated specific predator-prey interactions in the different parts of the lake.

The frequentist statistics used in biology have limitations that can affect the accuracy of the results and, by extension, the full analysis. The main problem is the lack of consideration of the statistics in the experimental design. Joint Species Distribution Models (JSDM) and Bayesian inference, which are widely used in (non-microbial) ecology, are less sensitive to such limitations. My aim was to apply this approach to the temporal dynamics of the microbial plankton community in ULC described above, with emphasis on the effects of winter vertical mixing and biotic interactions. Vertical mixing was confirmed to affect a high and diverse proportion of the plankton community, with for example Actinobacteria being highly sensitive. A large network of biotic interactions was observed, confirming known interactions such as the mycoloop or the association of Proteobacteria and Bacteroidota with specific phytoplankton species. The models also revealed additional, predicted interactions that require further experimental analysis. Hence, the models showed conclusive results, indicating that JSDM is applicable also to the NGS-based analysis of microbial communities

Zusammenfassung

Next-Generation Sequencing (NGS) Methoden ermöglichen, die riesige, bisher verborgene Vielfalt mikrobieller Gemeinschaften in den verschiedenen Lebensräumen der Erde aufzudecken. Die Verwendung von DNA aus Umweltproben gibt Zugang zu phylogenetischer und funktioneller Information und damit zur Zusammensetzung, Dynamik und Interaktion der Mikroorganismen in solchen Lebensgemeinschaften. Die Menge der gesammelten Informationen ist enorm und ermöglicht weitläufige Vergleiche. Die optimale Analyse solcher NGS-Datensätze steht jedoch immer noch unter kontinuierlicher Entwicklung.

Im Rahmen dieser Arbeit wurden erstmals DNA-Probenahmen etabliert und NGS-Methoden genutzt zur Beschreibung der mikrobiellen Planktongemeinschaft des Epilimnions des Bodensees, wobei auch die Optimierung der Analysen dieser Datensätze im Fokus stand.

Planktothrix rubescens, ein rot-pigmentiertes Toxin-produzierendes Cyanobakterium, bildet in alpinen und voralpinen Seen regelmäßig Blüten. Der Bodensee ist jedoch weitgehend verschont geblieben. Rote Biomasse auf Planktonfiltern im Sommer und Herbst 2019 und 2020 ließ den Verdacht aufkommen, dass *P. rubescens* auch im Bodensee (*Überlinger See*) in hoher Abundanz vorkommt. DNA-Extraktion und NGS-Sequenzierung von Cyanobakterien-spezifischen 16S-rDNA PCR-Produkten sowie Sanger-Sequenzierung des *mcy*-Gens für die Toxinproduktion bestätigten das Vorhandensein von *P. rubescens* im Bodensee, wenn auch in sehr geringer Abundanz. Microcystin-Toxine waren ebenfalls nachweisbar, wenn auch in extrem niedriger Konzentration. Tatsächlich waren rot-pigmentierte *Synechococcus* spp. abundant, die bereits 2003 beschriebenen wurden.

Planktonfiltration zur DNA-Extraktion wird traditionell mit einem festen Probenvolumen durchgeführt, was je nach Planktondichte der Wasserproben zu unterschiedlichen Mengen an Biomasse und damit DNA führt. Außerdem führt dies zu einem so genannten "Filtrations-Bias" bei der Analyse der Gemeinschaft anhand NGS. Es wurde eine neue Filtrationsmethode entwickelt, bei der die Durchflussrate als Proxi für die Biomassebeladung der Filter verwendet wird. Mit der neuen Methode konnten in etwa gleiche DNA-Mengen unabhängig von der Planktondichte gesammelt werden,

und die NGS-Ergebnisse zeigten keinen "Filtrations-Bias". Die Methode wurde in der folgenden Studie der saisonalen Plankton-Sukzession des Bodensees angewendet.

Die Zusammensetzung des Nanoplanktons (hier Organismen mit 180 - 5 µm Durchmesser) und Picoplanktons (5 - 0,2 µm Durchmesser) im Epilimnion des Ober- und Untersees wurde alle zwei Wochen von März 2018 bis März 2019 analysiert. Dazu wurde DNA aus Planktonproben der Wassersäule zwischen 0 - 20 m Tiefe (integrierte Probe) extrahiert und 18S- und 16S-rDNA PCR-Produkte mittels NGS sequenziert. Während die zeitliche Dynamik der Zusammensetzung des Planktons für beide Seeteile ein ähnliches saisonales Muster zeigte, ergab die phylogenetische Analyse spezifische Taxa und Organismengruppen, die entweder im Ober- oder *Untersee* häufiger oder sogar ausschließlich vorhanden waren und die auf spezifische Umweltfaktoren in diesen Seeteilen zurückgeführt werden können. Im *Obersee* beispielsweise brachte die Winterdurchmischung eine Hypolimnion-Gemeinschaft ins Epilimnion, während eine Hypoxia im *Untersee* ab Sommer auch anaerobe Mikroorganismen auftreten ließ. Unterschiedliche Taxa, die zur Mixotrophie oder Prädation fähig sind, deuten auf spezifische Räuber-Beute-Interaktionen in den verschiedenen Seeteilen hin.

Die in der Biologie häufig verwendete frequentistische Statistik weist Einschränkungen auf, die sich negativ auf die Ergebnisse und damit auf die gesamte Analyse auswirken können. Das Hauptproblem ist eine mangelnde Berücksichtigung der Statistik bei der Versuchsplanung. Gemeinsame Artenverteilungsmodelle (JSDM) und Bayes'sche Statistik, die in der Ökologie weit verbreitet sind, sind weniger anfällig für solche Einschränkungen. Mein Ziel war es, diesen Ansatz auf die zeitliche Dynamik der mikrobiellen Planktongemeinschaft im *Obersee* anzuwenden, wobei die Schwerpunkte auf den Auswirkungen der Winterdurchmischung und den biotischen Interaktionen lag. Es wurde bestätigt, dass die vertikale Durchmischung einen großen und vielfältigen Anteil der Planktongemeinschaft beeinflusst. Es wurde ein komplexes Netz biotischer Interaktionen beobachtet, das bekannte Interaktionen wie den Mykoloop oder die Interaktion von Proteobakterien und Bacteroidota mit Phytoplankton bestätigt. Die Modelle zeigten auch zusätzliche, vorgesagte Interaktionen auf, die weitere experimentelle Analysen erfordern. Die Modelle ergaben also schlüssige Ergebnisse, die bestätigen, dass JSDM auch für NGS-basierte Analysen mikrobieller Gemeinschaften anwendbar ist.

List of Figures

Figure 1: Sequence Read Archive (SRA) data growth. SRA is an online database release in 2009 that stores sequence data from NGS technology. 18

Figure I.1. Representative depth profile recorded with a Moldaenke FluoroProbe multi-channel fluorometer indicating a high abundance of red-pigmented biomass at a water depth below the chlorophyll-a maximum in Lake Constance on July 1st, 2019. 29

Figure I.2. Krona plot of the DRM cyanobacterial community composition as determined by 16S rRNA gene-amplicon sequencing across the sampling periods in 2019 and 2020. 30

Figure I.3. Illustration of the phylogenetic relationship of the *Synechococcus* spp. taxa observed in 2019 and 2020..... 32

Figure I.4. Dynamics of relative abundance changes for the *Synechococcus* ASVs observed during the sampling campaigns 2019 and 2020..... 34

Figure I.5. Relative abundance of the *Planktothrix* and *Microcystis* ASVs observed 2019 and 2020 and of microcystin concentrations as determined for samples taken in 2019. 36

Figure I.6. Depth-profiles recorded with the FluoroProbe across the sampling campaigns 2019 and 2020 in comparison to the year 2016 in which *P. rubescens* blooms were reported for Lake Constance..... 40

Figure II.1. Schematic of the filtration setup and flowmeter device used in this study. 61

Figure II.2 Decrease of flowrate over time during plankton filtrations due to increased biomass loading of the filters, as recorded by the Arduino flowmeter, and DNA yields after extraction of the corresponding picoplankton filters. 66

Figure II.3. Alpha and beta diversities of the bacterioplankton communities observed for the 0.2-µm filters in dependence on the filtration volume. 68

Figure II.4. Representative taxa that displayed significant changes in relative abundances in dependence on the filtration volumes used. 70

Figure II.5 DNA yield after extraction of the picoplankton filters..... 73

Figure II.6. Alpha and beta diversities of undiluted and diluted bacterioplankton collected on 0.2-µm filters by threshold filtration in comparison to fixed-volume filtration. 75

List of Figures

Figure II.7. Specific taxa that displayed different relative abundances in dependence on the filtration method used..... 77

Figure II.8. Secchi depths, filtration volumes recorded for threshold filtrations, and DNA yields obtained from a seasonal sampling of Upper and Lower Lake Constance plankton from March 2018 – March 2019. 81

Figure III.1: Shannon diversity calculated per sampling site and across the sampling dates. 97

Figure III.2: Principal Component Analysis of nano- and picoplankton biodiversity across the seasonal sampling of the different parts of the lake. 99

Figure III.3: Venn diagrams illustrating the number of shared and unique ASVs in the Upper and Lower Lake Constance plankton across all sampling dates. 102

Figure III.4: Relative abundances of the unique ASVs detected in either ULC or LLC across the sampling dates. 103

Figure III.5: Relative abundance temporal variation of the 18S-NP unique ASVs affiliated to Myzozoa. 106

Figure III.6: Forest plots showing of the effect size of the differentially abundant 18S-NP ASVs grouped by season. 108

Figure III.7: Forest plots showing of the effect size of the differentially abundant 18S-PP ASVs grouped by season..... 110

Figure III.8: Forest plots showing of the effect size of the differentially abundant 16S-PP ASVs grouped by season..... 111

Figure IV.1: Variance partitioning of bacterioplankton and microeukaryote community composition of the probit and the lognormal models, ordered by descending random factor (time) proportion..... 135

Figure IV.2: Interaction between ASVs and the fixed factors sequencing depth and vertical mixing in the probit occurrence model and lognormal abundance model. 137

Figure IV.3: ASV associations based on the temporal random effect of the probit and the lognormal models..... 140

Figure IV.4: ASV associations with correlation above 0.9 built on the temporal random effect of the probit (A) and the lognormal (B) FULL models..... 141

Figure 2: Relationship between relative abundance and differences between group of ASVs. The relationship was measured using aldex2 and the analysis was based on effect size and confidence interval or q-value..... 159

List of Tables

Table I.1. Primers used in this study**. 47**

Table III.1: Sequence and targeted region of the primers used for the amplification of the 16S and 18S rRNA gene for barcoded amplicon sequencing library preparation. **..... 91**

Table III.2: Overview of samples and sampling sites, at individual steps of the bioinformatics processing **..... 95**

Table III.3: Results of the Mantel tests and Procrutes analysis testing the correlation of the temporal patterns of plankton biodiversity change in ULC compared to LLC.**..... 100**

Table III.4: Results of PermANOVA and ANOSIM tests comparing the biodiversity differences of the three plankton communities between ULC and LLC**. 100**

Table IV.1: Sequence and targeted region of the primers used for the amplification of the 16S and 18S rRNA gene for barcoded amplicon sequencing library preparation. **..... 129**

Table IV.2: Variance proportion explained by each model factors **..... 134**

Table IV.3: number of ASV associations with a posterior probability support of 0.95. **..... 139**

Abbreviations

18S-NP: eukaryotic nanoplankton dataset

18S-PP: eukaryotic picoplankton dataset

16S-PP: prokaryotic picoplankton dataset

ANOSIM: analysis of similarity

AOA: ammonium-oxidizing archaea

AOB: ammonium-oxidizing bacteria

ASV: amplicon sequence variant

CTAB: cetyltrimethylammonium bromide

DDBJ: DNA Data Bank of Japan

DNA: deoxyribonucleic acid

DRM: deep red-pigment maxima

ENA: European Nucleotide Archive

FP: fluorometer profile

HMSC: Hierarchical Modelling of Species Communities

JGI: Joint Genome Institute

JSDM: Joint Species Distribution Model

HTS: High Throughput Sequencing

LLC: Lower Lake Constance

MC: microcystin

MC-LR: microcystin-LR

MC-YR: microcystin-YR

MCMC: Monte Carlo Markov chains

NaCl: sodium chloride

Abbreviations

NCBI: National Center for Biotechnology Information

NGS: Next-Generation Sequencing

NHST: null hypothesis significance testing

NOB: nitrite-oxidizing bacteria

OTU: operational taxonomic units

PCA: principal component analysis

PCoA: principal coordinate analysis

PCR: polymerase chain reaction

PermANOVA: permutational analysis of variances

Rcf: relative centrifugal force

rRNA: ribosomal ribonucleic acid

ser/thr: serine/threonine

SDS: sodium dodecyl sulfate

SRA: Sequence Read Archive

ULC: Upper Lake Constance

UPLC-MS/MS: Ultra performance liquid chromatography – tandem mass spectrometer

General introduction

The freshwater microbial plankton

Freshwater microbial plankton is a very broad term that includes a large diversity of microorganisms that can be divided into bacterioplankton and microeukaryotes (Figueiras et al. 2020; Cabrerizo et al. 2022). Members of the bacterioplankton are prokaryotic cells belonging to bacteria and archaea, mostly showing a heterotrophic lifestyle, with the exception of some autotrophic groups dominated by cyanobacteria (Newton et al. 2011; Callieri et al. 2016; Salmaso et al. 2018). The microeukaryotes are a much broader assemblage of microorganisms related to plants, fungi or animals, making it neither monophyletic nor a uniform functional group (Urry et al. 2017; del Campo et al. 2020). For example, this phylogenetic diversity has led to a diversity of feeding behaviours with autotrophic, heterotrophic or mixotrophic microeukaryotes, such as autotrophic organisms such as diatoms or Chlotophyta, predatory Ciliophora and saprophytic/parasitic fungi, , or mixotrophic dinoflagellates or Cryptophyta (Anderson 1988; Pålsson and Granéli 2003; Hammer and Pitchford 2006; Banerji et al. 2018; Frenken et al. 2020). Although these organisms are extremely small, they are central to the functioning of freshwater ecosystems. The autotrophic microbial community, known as phytoplankton, captures carbon dioxide and inorganic matter to produce biomass and release dioxygen into the ecosystem (Jones 1998). This activity is the first trophic level as it keeps the ecosystem stable and able to support oxygen-breathing organisms and produce organic matter used by heterotrophic microorganisms but also the zooplankton as a food source (Ducklow et al. 1986). The mineralization of organic matter into inorganic constituents and the recycling of nutrients for renewed growth of the phytoplankton is catalysed by heterotrophic microorganisms. Furthermore, the heterotrophic microorganisms themselves serve as food for e.g. ciliates, thereby linking their biomass directly to higher trophic levels, a process known as the microbial loop (Cole et al. 1988; Jetten 2008; Fenchel 2008). Finally, at the community level, the taxonomic composition and total biomass of the microbial plankton community is highly dynamic in space and time due to environmental changes and biotic interactions (Alavi 2004; Heil et al. 2005; Munir et al. 2013; Frenken et al. 2016; Woodhouse et al. 2016; Woodhouse et al. 2018).

Threat to the freshwater microbial plankton

Lake ecosystems have been described as sentinels of environmental change (Adrian et al. 2009), and the same observation can be made for the freshwater microbial plankton community. Many large and long-term effects on the community as a result of environmental change have already been observed. For example, the resurgence of the cyanobacterium *Planktothrix rubescens*, which forms toxic blooms in Alpine and pre-Alpine lakes, correlates with changes in the trophic state of these lakes (Jacquet et al. 2005; Garneau et al. 2015). Acidification of water, another major environmental change resulting from increased atmospheric carbon dioxide, has been correlated with lower phytoplankton (Geelen and Leuven 1986) and bacteria richness (Griffiths et al. 2000). The so called “dead zone”, increasing in occurrence in aquatic ecosystems, are area of oxygen depletion resulting from a combination of environmental factors and anthropogenic activities (Dodds 2006; Rocke et al. 2013). Recurring seasonal hypoxic event happen naturally in freshwater lakes like Lake Erie (Conroy et al. 2011) or Lake Constance (Chapter III). However, some hypotheses that anthropogenic activities: direct with nutrient runoff and indirect with climate change, could intensify the occurrence and length of such event (Conroy et al. 2011).

Hence, studying the resistance and resilience capacity of microbial communities is an intense field of work, sometimes showing contrasting results depending on the community or ecosystem studied (Shade et al. 2012a; Shade 2023). Nevertheless, the high diversity, plasticity and functional redundancy of the microbial plankton community tend to indicate a certain level of resistance and resilience to disturbance (Shade et al. 2011; Shade et al. 2012b; Mandakovic et al. 2018; Avila-Jimenez et al. 2020). However, anthropogenic activities and their associated environmental changes will continue to pose a threat to such ecosystems. More studies, both correlational and experimental, must be promoted to further increase our knowledge of microbial plankton communities, firstly to satisfy scientific curiosity and secondly but probably more importantly, to increase our predictive capacity for such complex ecosystems.

Technologies to disentangle microbial plankton communities

The study of microbial plankton communities has always been limited by their small size. Microscopic counting techniques have primarily been used to distinguish

individual species and the composition of phytoplankton communities, with species assignment based on the morphology of taxa (Sommer 1986; Gaedeke and Sommer 1986; Rodriguez et al. 2002; Chalar 2009; Su et al. 2016). This is time consuming and requires expertise to correctly differentiate the species. However, morphological variability of the species leads to inequalities in taxonomic affiliation and the detection capability of rare species is low: Next-Generation Sequencing (NGS) analysis has 50 to 83% higher richness (Naselli-Flores 2014; Xiao et al. 2014). Furthermore, the presence of human variability - in respect to the persons doing the counting - must be taken into account, as only 43% of taxonomic affiliation consensus was observed when comparing phytoplankton counting experts (Culverhouse et al. 2003).

For the next smaller size class than the phytoplankton, the lack of morphological variation between bacteria and archaea cells and thus different taxa and species makes the use of microscopic differentiation to study their community structure impossible. Methods for analysing bacterioplankton communities can be divided into two categories: culture-dependent and culture-independent techniques. Culture-dependent techniques are based on the isolation, e.g., of bacterial colonies on agar plates, which are used for molecular analysis such as Sanger sequencing. This method, applied to large and diverse communities, is time and energy consuming as isolation of pure bacterial colonies is a long and tedious process. Furthermore, cultivation is a bottleneck as it is estimated that only 0.1 to 1% of environmental bacterial species can be considered to be cultivable under laboratory conditions (Steen et al. 2019). In respect to freshwater lakes, some important taxonomic groups are also difficult to cultivate, such as the freshwater Actinobacteria acI lineage, which was isolated for the first time in 2019, despite being an highly abundant and thus important bacterioplankton lineage in lakes (Kim et al. 2019a; Kim et al. 2021); the same applies for the SAR11-sister clade LD12 in freshwater lakes (Salcher et al. 2011b). Therefore, culture-dependent approaches underestimate community richness and bias community composition (Hugenholtz et al. 1998; Vaz-Moreira et al. 2011). Culture-independent techniques overcome the limitation of cultivation while providing a more accurate representation of complex microbial community structure. The first type of analysis was the community fingerprinting techniques, which employ a unique characteristics of a universal component or section of a biomolecule (molecular markers) to differentiate taxa and provide a profile of the community (Zwisler et al.

2003; Hewson and Fuhrman 2004; Schwalbach et al. 2005; Kan et al. 2006). However, the size or composition of the targeted biomolecules may be so similar between individual microbial taxa that they are indistinguishable. Therefore, these techniques usually do not have enough resolution to accurately represent a microbial community in its natural diversity, and the phylogenetic information derived from these analyses is often insufficient to make an unequivocal taxonomic assignment (Popa et al. 2009; Portillo et al. 2011). The development of DNA-based taxonomic markers, for example using the 16S and 18S rRNA gene sequences, and of NGS technologies has allowed a huge step forward in analysing microbial plankton communities, as detailed in the following.

Sequencing revolution

The first generation of DNA sequencing, developed by Sanger in 1977, revolutionised molecular biology by providing access to genomic information (Sanger et al. 1977) and this technology has evolved rapidly since then. It took almost 30 years to develop the second generation, also called High Throughput Sequencing (HTS) or Next-Generation Sequencing (NGS) (Margulies et al. 2005), while only five to ten years separate the second and third generations, also called long-read sequencing (Eid et al. 2009; van Dijk et al. 2018). The development of these technologies has opened up new ways of studying the microbial community. Sanger sequencing is able to process only an individual DNA sequence of a single species, meaning that if the DNA came from a mixed culture, the sequencing results cannot be interpreted.

One of the major advances of NGS is the ability to process millions of different DNA sequences in parallel in a single run (Voelkerding et al. 2009). As with fingerprinting techniques, NGS allows culture-independent analysis of communities, while overcoming the limitations of the fingerprinting approach. Indeed, by using the sequence information of an amplified DNA fragment rather than comparing its size or composition (without knowing it), coupled with the ability to process millions of DNA sequences in parallel, the resolution of NGS overcame the limited representation of highly diverse microbial communities and the low phylogenetic information of fingerprinting approaches (Karczewski et al. 2017).

While the Human Genome Project took 13 years, 20 institutions and three billion dollars to sequence the entire human genome, it currently takes three days and

costs US$ 1000 (Li and Chen 2014). This decrease in price and time, combined with the diversification of sequencing technologies adapted to a wide range of applications, has led to the democratisation of sequencing technologies (Hugoni et al. 2013b; Hugoni et al. 2017; Debroas et al. 2017; Salmaso et al. 2018; Cabello-Yeves et al. 2020; Salmaso et al. 2020; Wu et al. 2020; Klotz et al. 2022). Genomic and metagenomic approaches have become so commonplace that we are now producing information on a massive scale (Figure 1) This new form of information production has propelled biology into the -omics or big data era, bringing exciting opportunities but also challenges to overcome (Li and Chen 2014; Gauthier et al. 2019). One of the challenges is how to handle such large amounts of information.

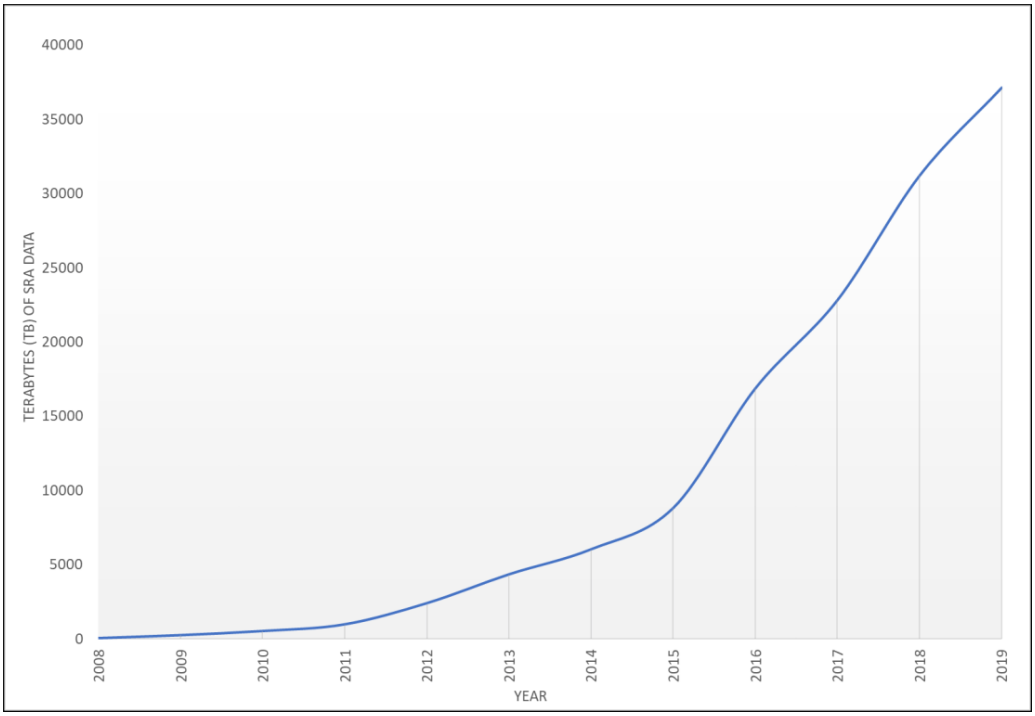

Figure 1: Sequence Read Archive (SRA) data growth. SRA is an online database release in 2009 that stores sequence data from NGS technology. Since its opening, the dataset has grown exponentially, reaching 73 petabytes in 2022.

Bioinformatics: change of approach in biology

Bioinformatics existed before the advent of the omics era. The first recorded bioinformatics software was COMPROTEIN, created in 1962 to study primary protein structure (Dayhoff and Ledley 1962) and the first DNA analysis software was developed in 1979 to analyse Sanger sequence reads (Staden 1979). Looking at the history of bioinformatics, molecular biology and bioinformatics have developed in parallel. With the development of more sensitive and diverse molecular biology

techniques, software was developed to analyse these new types of data, while computational power continuously increased (Gauthier et al. 2019). Bioinformatics was strongly influenced by the free software philosophy, which started in 1985, and participated in its dissemination at an early stage (Williams 2011). With the democratisation of the personal computer, informatics became accessible to more people and software was copied and shared for free on demand, and this mindset still follows today, with most bioinformatics software being open source and freely accessible (Rice et al. 2000; Bolyen et al. 2019; R core team 2021). The same applies to online data repositories such as ENA, NCBI or DDBJ, where sequence data is stored and accessible to anyone. The mass production of information that followed the second generation of sequencing has required a greater reliance on bioinformatics than ever before. Between 1990 and 2017, approximately 23000 pieces of software have been developed that are accessible to researchers without computer skills (Clément et al. 2018). This democratisation of bioinformatics tools, combined with advances in molecular biology techniques and the volume of information generated, has led to a major shift in biology. While the laboratory used to be the nerve centre of any type of study, whether experimental or environmental, (bio)informatics approaches have also taken a dominant place, placing the computer at the heart of biology. Until now, bioinformatics and biology have been considered as two separate "fields of study". However, the ubiquitous place that computers have taken in biology leads some researchers to hypothesise that in the near future, the term bioinformatics will be obsolete, as the division between bioinformatics and biology will no longer exist (Brown 2014; Gauthier et al. 2019).

Biostatistics in microbial ecology

Similar to the growth of bioinformatics, biostatistics has become a necessity in microbial ecology. Initially, there was a translation of statistics used in community ecology, hence the existence of equivalent terminology such as alpha, beta and gamma diversity, and the use of diversity indices such as Bray-Curtis distances or the Shannon-Wiener diversity index (Bray and Curtis 1957; Spellerberg and Fedor 2003). However, the nature of the data differs between observational ecological data, which provide quantitative data, and sequence-based microbial ecology data, which provide qualitative or compositional data (Aitchison 2003). This difference in nature is important when analysing the data, and previously named ecological indices, as well

as microbial ecology indices such as Unifrac (Lozupone et al. 2011), do not take this into account (Gloor et al. 2017). Unfortunately, there is still a long way to go, as biostatistics are still often used unthinkingly (Gloor et al. 2016b; Carr et al. 2019; Jia et al. 2022), and an increase in the recognition of statistics in microbial ecology is necessary.

The development started a few years ago with the recognition of the importance of the nature of the dataset, which allowed a first development of the statistics used in microbial ecology, while still maintaining the thought of the process present in ecological studies (Aitchison 2003; Fernandes et al. 2013; Gloor et al. 2016b; Gloor et al. 2017). However, the lack of consideration of statistics in experimental design is still an obstacle to the correct use of statistics(Halsey et al. 2015; Pollard et al. 2019). In ecology, joint species distribution models (JSDMs) are used to find statistical relationships between the occurrence or abundance of species and their biotic or abiotic environment (Ovaskainen and Abrego 2020). Such models have proven to be powerful analytical tools (Clark et al. 2014; Ovaskainen and Abrego 2020; Zhang et al. 2020; Poggiato et al. 2021; Elo et al. 2021) , but have been limited in microbial ecology because such models are too computationally intensive for sequence-based datasets. As they are based on Bayesian inference rather than frequentist statistics, their sensitivity to the initial experimental design is less important (van de Schoot et al. 2014; van de Schoot et al. 2015). Slowly but surely, such models have started to appear in microbial ecology, thanks to some precursors and the increase in computational power (Minard et al. 2019; Odriozola et al. 2021). These models may provide an alternative to frequentist statistics for microbial ecology studies in the near future. I hope that we will soon reach a stage where bioinformatics and biostatistics in microbial ecology reach a certain “maturity”, where researchers start to better understand and use statistics to make better and more accurate scientific contributions.

Lake Constance

Lake Constance is a pre-Alpine oligotrophic and monomictic lake located at the border between Germany, Switzerland and Austria and is ecologically and economically central to the region due to its geographical location between three countries and its size, being the second largest European pre-alpine lake. Its landscape is dominated by agriculture, as it is the largest grape growing area in

Germany, while at the same time acting as a drinking water reservoir for 5 million people. It is also one of the most important wetland habitats in Central Europe in terms of biodiversity, with more than 300 bird species and 36 fish species, including an endemic fish species thought to be extinct but rediscovered in the 2010s (Petri 2006; Doenz and Seehausen 2020). Like most pre-alpine lakes, Lake Constance underwent eutrophication in the 1970s due to agricultural intensification and lack of wastewater treatment, and is now re-oligotrophic thanks to intensive efforts.

Being such central economically and ecologically, Lake Constance has been and is studied extensively. Since its foundation in 1919, the Limnological Institute of Constance has host numerous researchers that have observed, recorded, analysed and accumulated data about every aspect of Lake Constance limnology: physical, chemical and biological (Lauterborn 1925; Stabel 1986; Grossart and Simon 1993; Schmidt and Conrad 1993; Straile and Geller 1998; Sabel et al. 2020; Straile 2021; Ogorelec et al. 2022). Yet, amidst this accumulation of information, the microbial plankton community of Lake Constance, and particularly its prokaryote component, remains a black box. Microeukaryotes were and are well studied, but remain analysed through microscopic approach (Sommer 1985; Sommer et al. 1993; Sommer et al. 2012). To the best of my knowledge, the only record of bacterioplankton community composition analyses of the bacterioplankton was in 2003 and performed using fingerprinting techniques (Zwisler et al. 2003). The NGS approach has yet to be introduced in Lake Constance to study its microbial plankton community.

Objectives

One major aim of this thesis was the development and optimization of a new nano- and picoplankton DNA-sampling procedure as part of the plankton routine sampling on the *Robert Lauterborn* research ship of the Limnological Institute of University of Konstanz. This included the development of an alternative filtration method that normalizes for a 'filtration bias' impacting the microbial plankton community analysis. Furthermore, it included the development and optimization of an appropriate DNA extraction method.

The samples collected every two weeks at two different sites (Upper and Lower Lake Constance) were conserved as DNA repositories of the seasonal succession of the microbial plankton community, and used for an NGS approach, employing PCR and 16S and 18S rRNA gene-fragment amplicon sequencing, which had to be developed as a second major aim of this thesis.

As a third major aim, a bioinformatics and biostatistics pipeline, necessary to process and analyse large NGS dataset, had to be established. This field of microbial ecology is in intense expansion with numerous new approaches developed or existing ecological models being transposed to microbial ecology; the necessity of applying newly developed methods to NGS datasets was central in the work presented in this thesis.

The final aim of this thesis was a first and fundamental, 16S and 18S rRNA gene-fragment based analysis of the temporal and spatial dynamics of the microbial plankton communities in Lake Constance.

Chapter I
Are toxin-producing *Planktothrix* spp. an emerging species in Lake Constance?

Corentin Fournier[2*] & Eva Riehle[1*], Daniel R. Dietrich[1#] and David Schleheck[2,3#]

[1] Human and Environmental Toxicology, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

[2] Microbial Ecology and Limnic Microbiology, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

[3] Limnological Institute, University of Konstanz, D-78457 Konstanz, Germany

* shared 1st authorship

shared senior authorship, correspondence: daniel.dietrich@uni-konstanz.de (DRD), david.schleheck@uni-konstanz.de (DS)

Keywords: Planktothrix; Synechococcus; Microcystins; temperate lakes

Key Contribution: Deep-water, red-pigmented cyanobacterial blooms in Lake Constance were characterized by phylogenetic community sequencing during summer 2019 and 2020, demonstrating the presence of low abundances of *Planktothrix* spp. and a predominance of picoplanktic phycoerythrin-rich *Synechococcus* spp., as well as very low concentrations of cyanobacterial toxins on individual sampling days. In response to climate change, changes of the physiochemical and biological parameters of the lake may in future support the establishment of toxic *Planktothrix* spp. blooms and/or mass developments of potentially toxic picocyanobacteria.

This chapter was published in Toxins in 2021
https://doi.org/10.3390/toxins13090666

Abstract

Recurring blooms of filamentous, red-pigmented and toxin-producing cyanobacteria *Planktothrix rubescens* have been reported in numerous deep and stratified pre-alpine lakes, with exception of Lake Constance. In a 2019 and 2020 Lake Constance field-campaign, we collected samples from a distinct red-pigmented biomass maximum below the chlorophyll-a maximum, determined via fluorescence-probe measurements at depths between 18-20 m. Here, we report the characterization of these deep red-pigment maxima (DRM) as cyanobacterial blooms. Using 16S rRNA gene-amplicon sequencing we found evidence that the blooms indeed were contributed by *Planktothrix* spp., but that phycoerythrin-rich *Synechococcus* taxa constituted most of the biomass (>96% relative read abundance) of the cyanobacterial DRM community. With UPLC-MS/MS, we also detected toxic microcystins (MCs) in the DRM at individual sampling days at concentrations of ≤ 1.5 ng/L. We next re-evaluated fluorescence-probe measurements collected over the past decade and found that in summer DRM have been present in Lake Constance at least since 2009. Our study highlights the need of a continuous monitoring program targeting also the cyanobacterial DRM in Lake Constance and of future studies on the competition of the different cyanobacterial taxa in relation to climate-change driven physiochemical and biological parameters of the lake.

Introduction

The formation of cyanobacterial blooms is a complex interplay of regional and biological variables and blooms have been reported worldwide with an increasing frequency (Huisman et al. 2018; Mantzouki et al. 2018). Albeit it has been shown that climate change and eutrophication are major contributors to bloom formation, the mechanisms with which nutrients and temperature interact to amplify blooms is still not fully understood (Mantzouki et al. 2018). The filamentous cyanobacterial genus *Planktothrix* occurs preferentially in pre-alpine and alpine lakes in temperate regions and has been responsible for many blooms in the past, for example in Lake Zurich (Walsby and Schanz 2002; Van den Wyngaert et al. 2011), Lake Bourget (Jacquet et al. 2005), Lake Garda (Salmaso 2010) and Mondsee (Kurmayer et al. 2004), where its mass occurrence can evolve to becoming a major factor within the food web (Posch et al. 2012). The success of *Planktothrix* spp. is attributed to its adaptability, for example its ability to regulate buoyancy and the use of phycobilins in addition to chlorophyll a (Salcher et al. 2011a; Kurmayer et al. 2016) as chromophores. By virtue of the phycobilins, phycocyanin and phycoerythrin, *Planktothrix* spp. can absorb light in large parts of the light spectrum, specifically blue and green light, thus conferring its red appearance. Consequently, the free-floating *Planktothrix* spp. usually develop blooms at depths of 9 – 15 m (Walsby et al. 2004; Kurmayer et al. 2016). Bloom depths within a stratified lake can be influenced by internal waves and can impact the growth of *Planktothrix* spp. by changing light availability (Cuypers et al. 2011). Similar to many other cyanobacterial species, *Planktothrix* spp. mass occurrences are supported by warmer water temperature, mediating more stable stratification of lakes in summer (Paerl and Huisman 2008; Posch et al. 2012). The genus *Planktothrix* is currently distinguished in nine species including red and green phenotypes, the most prominent representatives being *P. rubescens* and *P. agardhii*, respectively, which mostly occur in freshwater ecosystems of temperate regions (Komárek and Anagnostidis 2005). Like many other cyanobacteria, *Planktothrix* spp. are capable of producing microcystins (MCs), known to be toxic to humans as well as other mammalian and non-mammalian species (Dietrich and Hoeger 2005; Ernst et al. 2009; Chorus and Welker 2021). Hence, cyanobacterial blooms are often associated with the detection of increased intra- and extracellular toxin(s) (Buratti et al. 2017). Despite a plethora of efforts, neither the trigger for toxin production nor the factors resulting in the

development of toxic cyanobacterial blooms have been elucidated. Irrespective of the latter, the potential adverse impact of toxic cyanobacterial blooms on human health, society, economy and ecology highlights the importance of an improved understanding of cyanobacterial bloom formation with/without concomitant toxin production (Berry et al. 2017a).

Microcystins (MCs) share one common monocyclic structure with a molecular weight of approx. 1 kDa, composed of seven amino acids, amongst which are three *D*-amino acids, a *N*-methyldehydroalanine, two *L*-amino acids at the hypervariable positions 2 and 4 and the unique amino acid ADDA (Meriluoto et al.). Variations in amino acid composition and modifications such as methylations make up for a huge structural diversity, of which at least 279 different congeners have been reported (Bouaïcha et al. 2019). MC synthesis is encoded by 9-10 genes constituted as one *mcy* gene cluster (Tillett et al. 2000). Toxicity of MCs is induced by covalent binding and inhibition of ser/thr protein phosphatases and the concomitant hyperphosphorylation of cellular proteins (MacKintosh et al. 1990; Eriksson et al. 1990; MacKintosh et al. 1995), whereby the toxicodynamically relevant biological availability of MCs highly depends on the route of exposure and the respective MC congener (Feurstein et al. 2009; Fischer et al. 2010; Feurstein et al. 2010). MC concentrations <35 µg/L were reported in *Planktothrix* spp. blooms (Yéprémian et al. 2007; Catherine et al. 2008). When considering the current WHO guideline value of 1 µg MC/ L drinking water, which translates to approx. 10 µg/L raw water, largely depending on the type of water treatment used (Hoeger et al. 2002; Hoeger et al. 2004), it becomes obvious that toxin producing *Planktothrix* spp. blooms must be taken as a serious threat to freshwater systems serving as drinking water resources.

Unlike other pre-alpine lakes, Lake Constance, a drinking water reservoir for more than four million inhabitants, has so far not seen prominent and recurring, well-documented blooms of *Planktothrix* spp. Although *P. rubescens* has the capability to dominate entire lake ecosystems even at low nutrient concentrations (e.g. Lac de Bourget (Jacquet et al. 2005)), analyses of Lake Constance samples revealed only low abundances of *Planktothrix* spp. to date (Posch et al. 2012; Limnologischer Zustand des Bodensees). The first prominent appearance of *Planktothrix rubescens* in Lake Constance was documented in 2016, when *P. rubescens* filaments were

observed at various sampling sites of Lake Constance (SK 2016; Limnologischer Zustand des Bodensees).

No molecular phylogenetic studies have been conducted to date to evaluate the composition of the red-pigmented cyanobacterial community in Lake Constance in water depths typical for blooms of *Planktothrix* spp., i.e. below the chlorophyll-a maximum. In a field-campaign in the *Überlingen* embayment of Lake Constance (47.7571°N 9.1273°E), we observed reddish-colored plankton filters for samples taken at 18 m water depth in summer 2019. Subsequently, water samples were taken at the depths of maximal red-pigment (phycoerythrin) concentration, as determined by a fluorescence probe, on each of the fortnightly sampling days between June to October 2019, and July to September 2020. The composition of the cyanobacterial community at this deep-water red-pigment maximum (DRM) was assessed by Illumina PCR-amplicon sequencing using cyanobacteria-specific primers while UPLC-MS/MS analyses were applied to determine toxin concentrations. Potentially toxin-producing species were identified and quantified via quantitative PCR using *Planktothrix*- and microcystin-biosynthesis-specific (*mcy*) primers and Sanger sequencing.

Results

Bloom of red-pigmented phytoplankton at water depths below the chlorophyll-a maximum in Lake Constance

Multi-wavelength fluorometer profiles (FP) using a Moldaenke FluoroProbe were taken along the water column and evaluated right on the ship to determine the depths of the maxima of green pigment (chlorophyll-a maximum; predominantly diatoms and green algae) and red pigment concentration (deep-water red-pigment maximum, DRM). An example of a depth profile for all recorded fluorescence channels as well as of the interpretation of the abundance of different algae classes, as calculated by the Moldaenke FluoroProbe, is depicted in Figure I.1. An increased abundance of red-pigmented cyanobacteria, such as phycoerythrin-rich *Planktothrix* spp. or *Synechococcus* spp. was suggested by an elevated fluorescence intensity at 570 nm (and 525 nm) excitation wavelength (Figure I.1). According to the distinction of algae-classes via the Moldaenke FluoroProbe used, these phycoerythrin-rich algae were roughly attributed to represent ‘cryptophyta’.

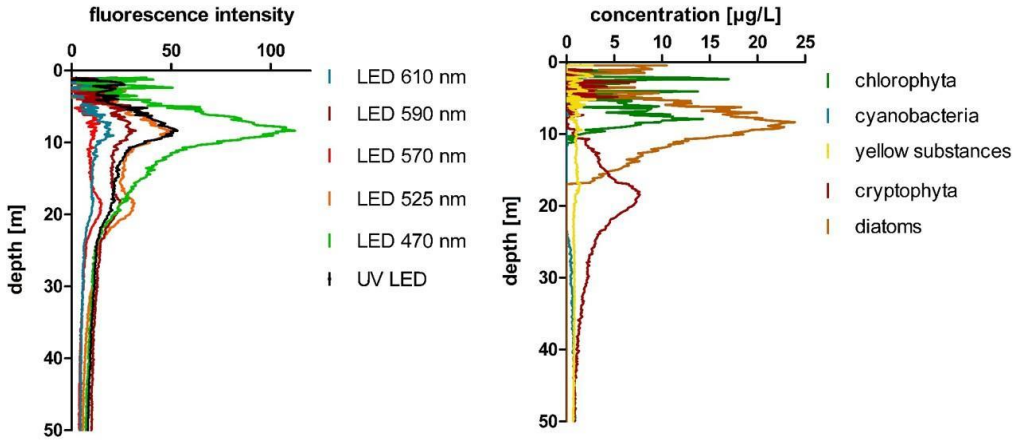

Figure I.1. Representative depth profile recorded with a Moldaenke FluoroProbe multi-channel fluorometer indicating a high abundance of red-pigmented biomass at a water depth below the chlorophyll-a maximum in Lake Constance on July 1st, 2019. A: Absolute fluorescence intensities recorded at the different excitation wavelengths. B: Abundance of the different algae classes as attributed by FluoroProbe. Red-pigmented cyanobacteria can be attributed to 'cryptophyta'. In this example (taken on July 1st, 2019), the chlorophyll-a maximum was determined between 8 - 10 m water depth (Figure I.1A, 470 nm) and a second maximum, indicating a red-pigmented biomass, at approx. 18 - 20 m water depth (Figure I.1A, 570 nm; 1B, cryptophyta).

Phycoerythrin-rich Synechococcus phylotypes dominated the DRM cyanobacterial community in 2019 and 2020, while Planktothrix spp. were detectable at low abundance

In order to characterize the cyanobacterial community composition presumably present at the DRM, plankton biomass at the DRM was collected on Whatman GF6 glass fiber filters. The filters appeared reddish, compared to yellow-green filters obtained from the chlorophyll-a maximum sample (see supplementary file, Figure SI.1). Total DNA was extracted from the filters and a fragment of the 16S rRNA gene was amplified using the cyanobacteria-specific primers CYA359F and CYA784R (Nübel et al. 1997). Illumina sequencing with 300 bp paired-end reads was employed. These primers amplified cyanobacteria phylotypes, which allowed for collecting phylogenetic information at a finer resolution, of also low abundant cyanobacteria at the DRM. Taxonomic affiliation was done using two different reference databases, SILVA 138 and Greengenes. Cyanobacteria taxonomy was consistent between SILVA and Greengenes, with the exception of the *Synechococcus* genus in Greengenes (replaced by *Cyanobium*_PCC-6307 in SILVA; *Cyanobium*_PCC-6307 is a heterotypic synonym of *Synechococcus* sp PCC-6307).

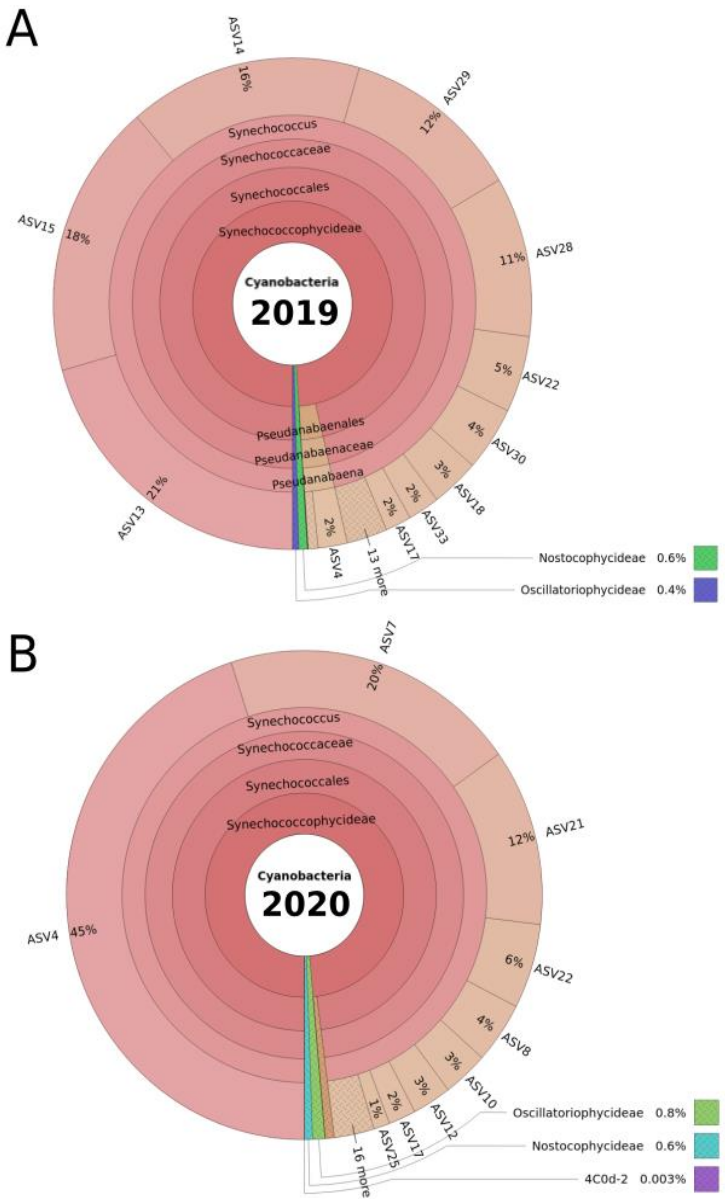

Figure I.2. Krona plot of the DRM cyanobacterial community composition as determined by 16S rRNA gene-amplicon sequencing across the sampling periods in 2019 (A) and 2020 (B). The community analysis was done by filtration of DRM water samples, total DNA extraction of the filters, and PCR amplicon sequencing of the cyanobacteria-specific 16S rRNA gene region V3-V4 (380 bp length). For this Krona plot and as an overview, the results shown are from all samples combined per year. Taxonomic affiliation was carried out using the Greengenes reference database and all taxonomic ranks are represented in the plot. Amplicon sequence variants (ASVs) as output of the Dada2 software package used (see Materials and Methods) distinguish also sequence variations by a single nucleotide, giving ASVs a higher resolution than the operational taxonomic units (OTUs) typically used. Therefore, ASVs were used as the deepest taxonomic rank in our study. Total relative abundance was calculated by dividing the number of reads affiliated to an ASV in a sample by the total number of reads in the sample.

Subsequent to bioinformatics processing and the removal (filtering) of the extremely low abundant phylotypes (i.e., phylotypes represented by less than four read in at least 20% of all samples; see Material and Methods), 35 amplicon sequence variants (ASVs) were detected in 2019, and 37 ASVs in 2020 (Figure I.2). Each ASV was affiliated at the level of either genus or order, as the last common taxonomic rank between SILVA and Greengenes databases (note that species rank could not be affiliated by the amplicon sequencing we used).

For both the 2019 and 2020 sampling campaigns, the genus *Synechococcus* clearly dominated the cyanobacterial DRM community, as examined by amplicon sequencing, occupying 96% (SILVA) or 98% (Greengenes) of the total relative read abundance and representing 65.7% (SILVA) and 67.6% (Greengenes) of the detected ASVs in the community (see supplementary file, Table SI.1). For 2019, five *Synechococcus* ASVs represented 78% of the total relative abundance, with ASV13 being the most abundant with 21% total relative abundance (Figure I.2A). For 2020, only three ASVs affiliated to *Synechococcus* contributed 77% of the total relative read abundance, with ASV4 contributing almost half (45%) throughout the year (Figure I.2B).

Although *Synechococcus* dominated the cyanobacterial community at the DRM in 2019 and 2020, each of the two ASVs affiliated to *Planktothrix* could be detected in both years. *Planktothrix* (Oscillatoriaphycideae) with 0.4% (2019) and 0.9% (2020) of the total relative abundance (Figure I.2A and 2B), represented only low abundant taxa together with Nostocophycea. In addition, two *Microcystis* spp. ASVs were detected in 2019, at 0.06% of the total relative read abundance. Only one *Microcystis* spp. ASV was detected in 2020, with a very low total relative abundance of 0.008%. Although the respective relative abundances of *Planktothrix* and *Microcystis* species are low, the strong bioinformatic filtration confirms the biological significance of these amplicon sequencing results. The ASVs of most abundant *Synechococcus* spp. as well as the *Planktothrix* spp. and *Microcystis* spp. ASVs were used for further analyses.

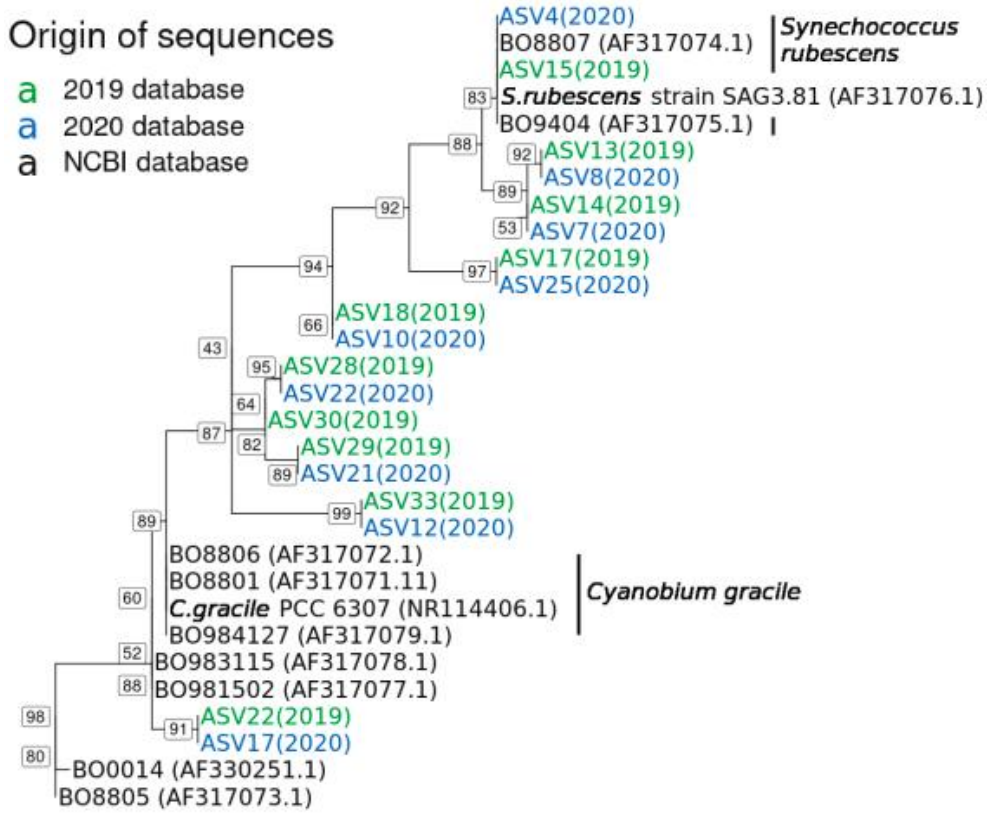

Figure I.3. Illustration of the phylogenetic relationship of the Synechococcus spp. taxa observed in 2019 and 2020. The phylogeny is based on the cyanobacteria-specific V3-V4 (380 bp) 16S rRNA gene region. Colors correspond to the origin of the sequences, with green for ASVs observed in 2019 and blue for ASVs observed in 2020 (sampling year also in brackets). Sequences from *Synechococcus spp.* taxa isolated previously from Lake Constance (Ernst et al. 2003) and sequences of *Synechococcus rubescens* strain SAG3.81 and *Cyanobium gracile* PCC 6307 from NCBI were used as references (indicated in black). IQ-TREE was used for phylogenetic inference using maximum likelihood. The model of nucleotide substitutions used was TPM2+F+I (Minh et al. 2020), determined as best-fit model by ModelFinder (Kalyaanamoorthy et al. 2017) based on the Bayesian information criterion scores and weights. Numbers at the internal nodes represent the percentage support of this specific node in a 1000-bootstraps testing. The tree was rooted using the sequences of the cyanobacterium *Anacystis nidulans* PCC 6301 as an outgroup (not shown). The reference sequences are identified by their accession number in brackets.

Synechococcus rubescens and Cyanobium gracile clusters in 2019 and 2020

We examined the phylogenetic relationship of the main *Synechococcus* ASVs detected in 2019 and 2020 with reference sequences of (*i*) all cultivated phycoerythrin-rich *Synechococcus* spp. of Lake Constance, as established by Ernst and colleagues in 2003 (Ernst et al. 2003), and (*ii*) with NCBI reference sequences of *Synechococcus rubescens* and phycoerythrin-rich *Cyanobium gracile* (Figure I.3). The relationship was established using the appropriate sequence fragments, representing the PCR amplicon of 380 bp, and the program IQ-TREE (Minh et al. 2020), when doing a phylogenetic inference using maximum likelihood coupled with ModelFinder to determine the best fitted nucleotide substitution model (Kalyaanamoorthy et al. 2017). Although the target sequence is shorter than the full 16S rRNA gene sequences established by Ernst *et* al., 2003, the phylogenetic relationship between the reference sequences remained the same, thereby confirming our analyses. The ASVs from this study were always grouped in pairs, with one ASV from 2019 (Figure I.3, green font) and another from 2020 (Figure I.3, blue font), as a reflection of re-occurring *Synechococcus* phylotypes across the two years, further confirming our analysis.

Overall, the ASVs grouped into two main clusters, either more closely related to the *S. rubescens* or to the *C. gracile* reference sequence (Figure I.3). The top-three most abundant ASVs for 2019 (ASV13, ASV14 and ASV15; see Figure I.2 and below) and the top-two most abundant for 2020 (ASV4 and ASV7; Figure I.2), were more closely related to the *Synechococcus rubescens* NCBI reference sequence. For this group, also two reference sequences of Lake Constance, phycoerythrin-rich *Synechococcus* isolates (Ernst et al. 2003), were available (Figure I.3; BO8807 and BO9404), while for the *C. gracile*-group, reference sequences of seven Lake Constance isolates were available (Figure I.3).

Megablast results were analogous to the phylogenetic tree shown, with the same 2019 and 2020 ASVs being affiliated closer to either *Synechococcus rubescens* or *Cyanobium gracile* (Table SI.2a and SI.2b), with the exception of ASV18 of 2019 and ASV10 of 2020, which showed identical percentage identities and E-values in megablast to both *S. rubescens* and *C. gracile* (Table SI.2a and SI.2b) and, hence, grouped in between both clusters (Figure I.3).

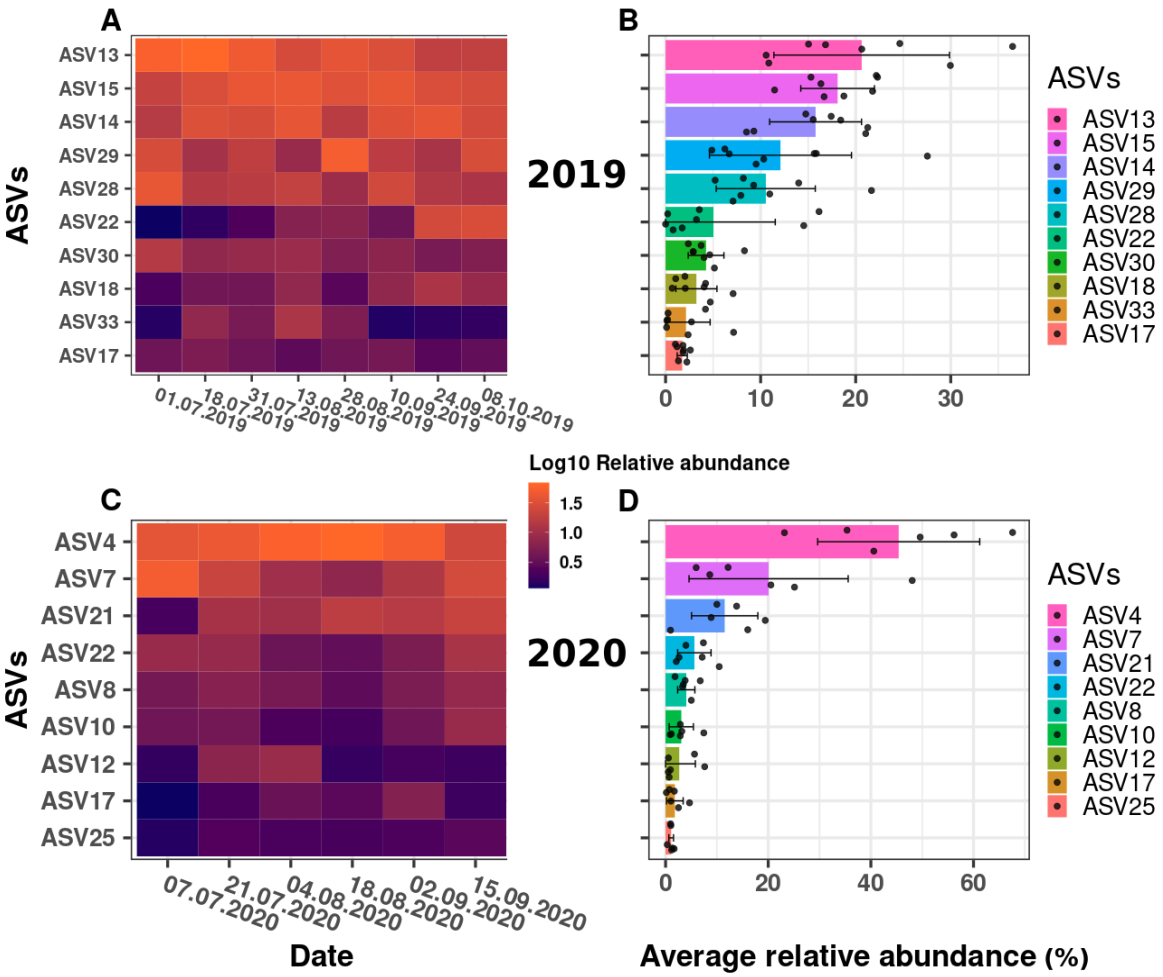

Figure I.4. Dynamics of relative abundance changes for the Synechococcus ASVs observed during the sampling campaigns 2019 and 2020. Shown are heatmaps across all sampling dates (A, C) and the ASV's average relative abundances (B, D) for each year (bars) with each individual data point indicated (grey dots). For the y-axes of heatmaps (A, C) and average relative abundances (B, D), the different ASVs were ordered from the more abundant (top) to the least abundant (bottom). The heatmap data was log10 transformed from the read count matrix, and to avoid introducing infinity for zero read counts, we added one artificial read to every cell prior to log10 transformation (log10[x + 1]). Color schemes vary between dark blue for low log10 value to bright orange for high log10 value; a higher log10 value means a higher relative abundance. Please note that the y-axes for B and D do not have the same scale. The average relative abundance across all samples of each year was calculated as for the Krona plot (see Figure I.2). Standard deviation was also calculated and represented for each ASV; if the graphical representation of the standard deviation was below 0, the minimal error-bar value was set to 0.

Dynamics of Synechococcus ASVs in 2019 and 2020

We examined the change in relative abundance over time of the *Synechococcus* ASVs detected. Figure I.4 illustrates the dynamics of the *Synechococcus* ASVs as a heatmap using relative abundance data after log10 transformation (Figure I.4A and 4C) and the data distribution (Figure I.4B and 4D) as the mean and standard deviation of relative abundances in a combined bar and jitter plot, where each dot represents the relative abundance of the taxa at a specific date (relative abundance values (%) per ASV across sampling dates are shown in Table SI.4a and SI.4b).

Overall, the observed changes in the phylogenetic structure across the sampling dates suggested a high degree of successional change within the *Synechococcus* spp. community at the DRM. Some ASVs varied from being almost undetectable in the beginning to a relative-abundance maximum later in the year, or showed the opposite trend, being abundant at the beginning or in the middle of the sampling campaigns, while other ASVs showed a comparatively stable (and low) relative abundance across the sampling campaigns. Indeed, while 10 ASVs were noted for 2019, only 9 ASVs were registered for 2020 (Figure I.4A and 4C). Thus, different ASVs appeared to dominate and therefore provide for a successional change in the two field campaigns. For example, ASV13 as the most prominent ASV in 2019 (Figure I.4C), had its maximum in July (up to 36%, Figure I.4A) and decreased thereafter in its relative abundance (to about 10%) at the end of September, while the two second most abundant ASVs in 2019, ASV 14 and 15 (Figure I.4C), had their maxima in mid-August and later in the year (see Figure I.4A and Table SI.4a). Similarly, for 2020, ASV7 represented almost 50% of the total relative abundance in early July but decreased to about 6% by mid-August, while the most abundant taxon in the 2020 sampling campaign, ASV4 (Figure I.4D), had its peak end of August (68%) and decreases to about 23% by the end of the campaign (Figure I.4B); further, the third most abundant ASV21 in 2020 represented only about 1% of the total relative abundance at the beginning of the sampling campaign, but almost 20% at its end (see (Figure I.4C, Table SI.4b).

The statistical relevance of the difference in *Synechococcus* ASV relative abundance was confirmed by a Kruskal-Wallis analysis of variance, with *p*-value far below the threshold of 0.05 (*p*-value of $3.95e^{-10}$ and $1.27e^{-5}$ for 2019 and 2020,

respectively). In an attempt to group taxa according to their relative abundance differences, a post-hoc Conover-Iman test was performed with a Benjamini Yekutieli *p*-value adjustment method for False Discovery Rate control (Table SI.3).

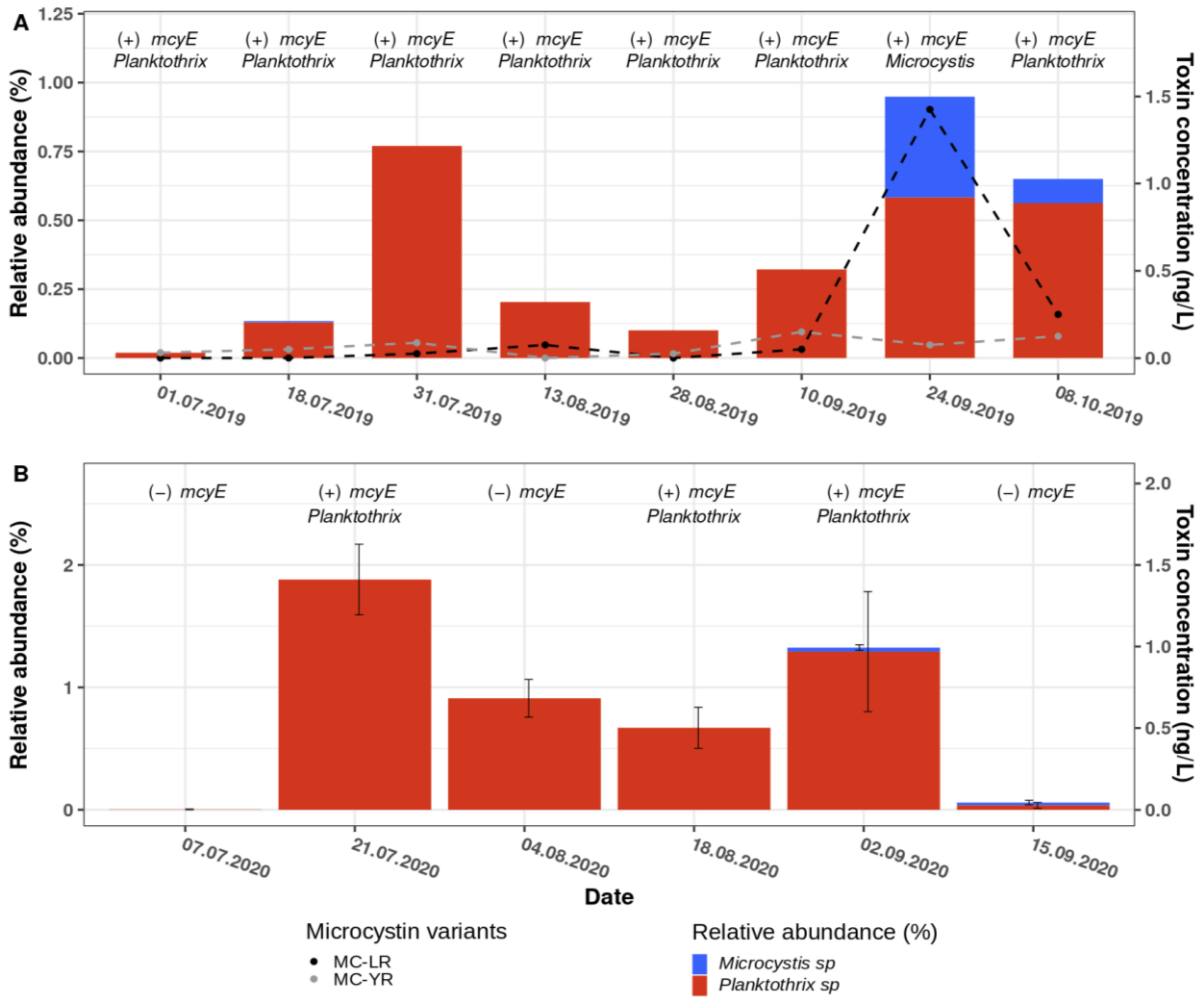

Figure I.5. Relative abundance of the Planktothrix and Microcystis ASVs observed 2019 and 2020 and of microcystin concentrations as determined for samples taken in 2019. The bar plots represent the total relative abundance observed for the two *Planktothrix* ASVs (red bars) detected in 2019 (A) and 2020 (B) and for the two *Microcystis* ASVs (blue bars) detected in 2019 and of the single *Microcystis* ASV detected in 2020. The x-axes represent the different sampling dates and the right y-axis the relative abundance (%). The error bars in B represent the standard deviation of the relative abundance as calculated from biological triplicates (n = 3); no error-bars are represented for 2019, as only one sample was collected per sampling date. The left y-axis represents the toxin concentration (ng/L) determined for independently collected DRM filters, for which the microcystin variant concentrations are indicated by dashed lines in black (MC-LR) and grey (MC-YR) (see main text). The text on top of each bars indicates the results of PCR amplification of the microcystin synthesis gene *mcyE* (+, detected; -, not detected) and of the phylogenetic affiliation of the *mcyE* consensus sequence to either *Planktothrix* or *Microcystis*, as established by Sanger sequencing of the PCR amplicons (see main text).

For 2019, the results suggested two groups. First, a group comprising ASVs that comparatively stably dominated the cyanobacterial community at the DRM throughout July and October (together 63% to 87% of the total relative abundance), i.e., ASV13, 14, 15, 28 and 29. A second group statistically differed from the first group, i.e., ASV17, 18, 22, 30 and 33, for which abundance differences appeared to be larger and/or occurred within shorter time intervals: For example, ASV22 had its high relative abundance of 15% only at two sampling dates (in September and October, Figure I.4A and Figure SI.4a),as also discussed above. For 2020, the Conover-Iman test did not significantly separate the ASVs into two groups as for 2019 (Table SI.3), even though visually (Figure I.4B); ASV4 dominated the cyanobacterial community with an average relative abundance of 45% throughout 2020.

Dynamics of Planktothrix and Microcystis ASVs, the abundance of microcystin biosynthesis genes and the concentration of microcystins in samples taken during 2019 and 2020.

In both the 2019 and the 2020 amplicon-sequencing datasets, we detected two ASVs affiliated to the *Planktothrix* genus. The relative abundance of *Planktothrix* ASVs in 2019 increased from July to late September, with a peak on July 31st, where 0.75% of the total cyanobacterial community were *Planktothrix* ASVs (Figure I.5A). Likewise, in 2020 *Planktothrix* ASVs were abundant from the end of July to the end of September with a maximum of approx. 2% relative abundance on July 21st (Figure I.5B). While one of the *Planktothrix* ASVs was affiliated to *Planktothrix rubescens*, a toxin-producing *Planktothrix* species, amplicon sequencing of the cyanobacteria-specific 16S rRNA gene fragment of the other ASVs did not allow for the taxonomical distinction at the species level, i.e. between the mostly non-toxin producing *P. agardhii* and toxin-producing *P. rubescens*. For example, a Megablast alignment of the *Planktothrix* ASVs suggested *Planktothrix agardhii* and *Planktothrix rubescens* as top hits with identical query coverage and E-values, while the percentage identity varied between 99.70% and 100% for *P. agardhii* and 99.38% to 99.74% for *P. rubescens*.

Beyond the *Planktothrix* ASVs, two ASVs affiliated to the genus *Microcystis* were detected in 2019, however only one *Microcystis* ASV was detected in 2020. As both genera, *Planktothrix* and *Microcystis*, are renowned for their toxin producing species, more in-depth analyses were carried out regarding the toxin producing potential of the species found in the *Überlingen* embayment.

To confirm the detection of the genus *Planktothrix* in Lake Constance at very low levels, especially when compared to *Synechococcus*, quantitative PCR (qPCR) was performed to estimate the abundance of toxin-producing *Planktothrix* genotypes. Specifically, we used *Planktothrix*-specific primer pairs for the 16S rDNA gene and the *mcyBA1* gene, as described by Ostermaier & Kurmayer, 2009 (Ostermaier and Kurmayer 2009). *Planktothrix mcyBA1* encodes for the first adenylation domain in non-ribosomal peptide synthase (NRPS) gene cluster and is present only in species capable of toxin production. Briefly, we created a standard curve using a dilution series of *Planktothrix* DNA (0.00001-100% *Planktothrix* DNA diluted in *Microcystis* DNA) and then calculated the relative abundance of *Planktothrix* DNA in our samples using a linear regression. The calculated relative abundance of *Planktothrix* ranged between 0.1 and 0.6% in 2019 and 0.01% in 2020 (Figure SI.2). Although the calculated relative abundance by qPCR differs from the relative abundances found with amplicon sequencing (Figure SI.2 and Figure I.5), the trend aligned well between both methods. Statistical analyses of the difference between 16S-rRNA gene and *mcyBA1* amplifications (ANOVA, see Ostermaier & Kurmayer, 2009) showed no significant difference in abundance with respect to toxin-producing or non-toxin producing *Planktothrix* genotypes, suggesting that there is only one genotype of *Planktothrix* present in 2019 and 2020. To confirm the presence of toxin-producing cyanobacterial species in Lake Constance, we used universal *mcyE*-specific PCR primers (HEPF/R, (Jungblut and Neilan 2006)). Being a member of the MC production gene cluster, *mcyE* is partly responsible for the synthesis of the ADDA chain in microcystins as well as the incorporation and synthesis of *D*-Glu (Christiansen et al. 2003). For the 2019 sampling campaign, the PCR yielded amplicons for every sample tested, suggesting the presence of potential microcystin producers throughout the year. Subsequent Sanger sequencing of the PCR products and analysis of the consensus sequences with Megablast attributed toxin-producing capabilities to *Planktothrix* species (Figure I.5A), except for the sample taken on September 24th, 2019, where the *mcyE* consensus sequence had the highest alignment scores with *Microcystis*, thus matching the date with the highest relative abundance of the *Microcystis*-affiliated ASV. For 2020, *mcyE* amplicons were observed less consistently than in 2019, but as seen in 2019, Sanger sequencing of these amplicons attributed the toxin-producing capabilities to *Planktothrix* spp. (Figure I.5B).

Collectively, we provide evidence that toxin-producing *Planktothrix* spp. (and/or *Microcystis* spp.) are present in the *Überlingen* embayment of Lake Constance. However, our data did not allow to conclude whether or not toxin production took place during the sampling campaigns as many species can carry the gene cluster without actively producing the toxins (Beversdorf et al. 2015). Consequently, we analysed the biomass samples collected independently on filters from the DRM for microcystins using UPLC-MS/MS (Figure I.5A and 5B). A peak in microcystin concentration was detected at the end of September 2019, where in total approximately 1.5 ng/L intracellular microcystins were found. The microcystin variants present were MC-LR (leucine and arginine in hypervariable region) and MC-YR (leucine and tyrosine in hypervariable region), with MC-LR being almost solely responsible for the peak in toxin concentration in September 2019 (Figure I.5A). Strikingly, for the samples collected during 2020, no toxins were detected under the conditions we used.

Retrospective evaluation of depth profiles for the Lake-Überlingen routine sampling site

During our sampling campaigns 2019 and 2020 (and the ongoing 2021 campaign), prominent DRM were observed from June onwards, especially after long and stable good-weather periods. This is illustrated best on July 1st, 2019, when we observed a first, prominent DRM at 18.4 m depth at the routine sampling site '*Wallhausen'* in the *Überlingen* embayment of Lake Constance. This DRM showed up after the weather presented with a stable window of approx. two weeks with predominant sunshine and no precipitation, low wind and elevated temperatures, as depicted in Figure SI.4, Figure SI.6 and Figure SI.7.

Three-dimensional (3D) plots of the FluoroProbe depth profiles for 'cryptophyta' content as proxy in the water column from 0-40 m depth (Figure I.1) across the sampling campaigns 2019 and 2020 are depicted in Figure I.6A. Furthermore, we retrospectively evaluated the 'cryptophyta' depth profiles as collected in the past years from the routine sampling site and transformed these into 3D plots using MATLAB (for years 2009 – 2018, see Figure SI.5). During the past twelve years, DRM have occurred multiple times (Figure SI.5, plots from 2011 and 2015 – 2020). Specifically, in 2016, we observed a prominent DRM in the summer, with maximum 'cryptophyta' concentrations of 6 µg/L on September 6th and September 20th (Figure I.6B). Corresponding to the DRM, a high abundance of *Planktothrix rubescens* has been

reported at various sampling sites of Lake Constance in 2016 (SK 2016; Limnologischer Zustand des Bodensees).

Overall, prominent DRM can be observed from July – October, with maximum 'cryptophyta' content of about 2 – 4 µg/L (as estimated based on the FluoroProbe calibration) and at water depths ranging between 10 and 20 m (Figure I.6A, Figure SI.5). The observed variation of the DRM depths likely follows the lake's internal waves, as previously found in Lake Ammer in Germany and Lac de Bourget in France (Cuypers et al. 2011; Hingsamer et al. 2014). Although we could speculate that these tree peaks represent high abundances of *P. rubescens* in the *Überlingen* embayment, the absence of appropriate samples allowing for DNA or toxin analyses available from that time preclude any corroboration.

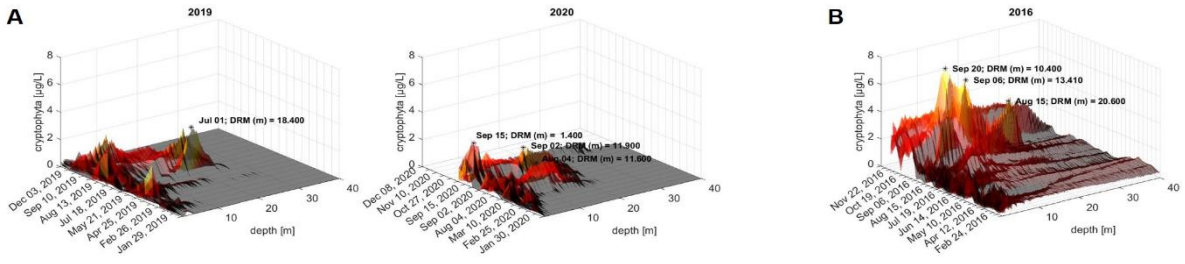

Figure I.6. Depth-profiles recorded with the FluoroProbe across the sampling campaigns 2019 and 2020 (A) in comparison to the year 2016 in which P. rubescens blooms were reported for Lake Constance (B). Depicted are the FluoroProbe profiles for 'cryptophyta' abundance (cf. Figure I.1) (expressed in µg chlorophyll-a per liter) recorded as proxy of red-pigment abundance in the water column from 0-40 m depth at the routine sampling site '*Wallhausen'* in the *Überlingen* embayment of Lake Constance. Prominent DRM are highlighted with date and water depths. Coordinates of the study site: 47.7571°N 9.1273°E; for an illustration see Figure SI.3. In 2016, blooms of *P. rubescens* were reported for Lake Constance in September-October at various sampling sites, i.e. of the German, Austrian and Swiss sections of Lake Constance (SK 2016; Limnologischer Zustand des Bodensees). For 2016, FluoroProbe data is available for the *Überlingen* embayment of Lake Constance (B).

Discussion

The characterization of recurring blooms of red-pigmented cyanobacteria in the *Überlingen* embayment of Lake Constance at water depths of 15 – 20 m by amplicon sequencing demonstrated the presence of *Synechococcus*, *Planktothrix* and *Microcystis*. Briefly, the relative abundance data suggested that *Synechococcus* taxa predominated the community (96 – 98%) at these water depths, while *Planktothrix* and *Microcystis* taxa were detectable only at very low abundances. For example, in 2020 up to 45% of the total relative abundance was represented by a single *Synechococcus* ASV (Figure I.2B and Figure I.4CD; ASV4). Moreover, the observed changes in relative abundance of *Synechococcus* ASVs across the sampling dates suggest a high degree of successional change within the *Synechococcus* spp. community at the DRM (Figure I.4).

Isolated phycoerythrin-rich *Synechococcus* species from Lake Constance have been investigated in the past, and similar to *Planktothrix* spp., they can express a large variety of phycobilins, thereby exploiting diverse light conditions (Ernst et al. 2003; Six et al. 2007). Interestingly, the highest abundant *Synechococcus* taxa found in our studies in 2019 and 2020 are closely related to either *S. rubescens* or *C. gracile*, forming two main clusters (Figure I.3, Table SI.2a and SI.2b). For the 2019 and 2020 sequences, a discrimination between *S. rubescens* and *C. gracile* is difficult (Figure I.3), as the 380 bp PCR amplicon used in this study is cyanobacteria specific and allows to affiliate taxa with high confidence only up to the genus rank. However, each *Synechococcus* taxon detected in 2019 is paired with a taxon detected in 2020, and their sequence alignment showed a 100% identity with no gap, suggesting recurring phylotypes/ecotypes at least across the two years. The close association of our sequences with those from the early 2000s (Ernst et al. 2003) suggests a stable deep-water cyanobacterial community in Lake Constance. The latter interpretation is supported by the earlier finding that different lineages of *Synechococcus* spp. can adapt to and thrive in specific ecological niches (Salazar et al. 2020).

Predominance of *Synechococcus* species over other cyanobacterial genera was reported for experimental co-cultures of *Synechococcus* and *Microcystis* strains as well as varying phosphate and nitrogen concentrations (Tan et al. 2019b; Tan et al. 2019a). Indeed, at low nutrient concentrations, which could also mirror the currently

oligotrophic conditions of Lake Constance (Limnologischer Zustand des Bodensees), *Synechococcus* outcompeted *Microcystis* in growth rate and final biomass (Tan et al. 2019b; Tan et al. 2019a). The predominance of *Synechococcus* may be explained its seemingly higher affinity for orthophosphates, more efficient nutrient uptake due to a larger surface-to-volume ratio as well as by potential competitive inhibition via quorum sensing/quenching molecules between the two strains (Verity et al. 1992; Santhakumari et al. 2016; Tan et al. 2019b; Tan et al. 2019a). The latter laboratory findings were also corroborated in natural habitats, at least in their trend (Ruber et al. 2018).

Despite the seemingly stable predominance of *Synechococcus* in the deep cyanobacterial community observed in Lake Constance the past two decades (Ernst et al. 2003), *de novo* occurrence of *Planktothrix* spp. in 2016 and reconfirmation of this earlier finding with our samples of 2019 and 2020 could suggest that *Planktothrix* spp. is in the process of establishing a stable presence in the *Überlingen* embayment of Lake Constance. The latter is of critical importance as mass occurrences of toxin producing cyanobacteria at water depths >20 m could become a threat to the water intake for the Sipplingen water treatment plant (https://www.bodensee-wasserversorgung.de) serving > 4 million people with drinking water. Indeed, we detected a low abundance of possibly MC-producing *Planktothrix* spp. by amplicon sequencing as well as by genus *Planktothrix*-specific qPCR for 2019 and 2020 (Figure I.5 and Figure SI.2). Furthermore, presence of the microcystin-biosynthesis gene cluster *mcy* was detected by PCR amplification of *mcyE* (Figure I.5). Subsequent Sanger sequencing of these amplicons identified *Planktothrix* spp. to be the main contributor to this gene sequence in our samples, except for September 24th, 2019, when the amplified *mcyE* consensus sequence showed highest alignment scores to *Microcystis* spp. In *Planktothrix* spp., the *mcy* gene cluster can be inactivated by various mutations, e.g. insertions or deletions, and thus non-toxic strains can develop (Christiansen et al. 2006). Non-toxic strains are less successful in competition than their toxic relatives, and thus *Planktothrix* blooms are usually dominated by toxic strains (Ostermaier and Kurmayer 2009). Corresponding to the presence of *mcyE* in 2019, low amounts of the microcystins MC-LR and MC-YR were detected by UPLC-MS/MS. In 2020, although *mcyE* amplicons were detectable for some of our samples, concentrations of microcystins were below the detection levels of the UPLC-MS/MS

used (LOD of UPLC-MS/MS method: 0.5 ng/mL (Altaner et al. 2019), therefore the resulting LOD in our water samples: 62 pg/L lake water). Despite the low abundance of *Planktothrix* spp. and low concentrations of detected toxins, the question does arise whether the latter is a sign of a fundamental change in water body dynamics in the *Überlingen* embayment of Lake Constance, e.g. resulting from global warming. Indeed, greater and prolonged stratification of water bodies in conjunction with lowered nutrient levels would promote deep-water euphotic ecosystem encompassing low-light specialized species such as the picocyanobacterial *Synechoccocus* spp. and *Planktothrix* spp. (Paerl and Huisman 2008; Posch et al. 2012), as was reported for other pre-alpine lakes e.g. Lake Zurich (Van den Wyngaert et al. 2011), Mondsee (Kurmayer et al. 2004) or Lac de Bourget (Jacquet et al. 2005).

As *Planktothrix* spp. are stimulated by increased temperature (Paerl and Huisman 2008), the continuous shift toward higher overall temperatures could favor the perseverant establishment of toxin producing *Planktothrix* spp. to the disadvantage of today's *Synechococcus*-dominated DRM ecosystems. Considering that *Synechococcus* spp. currently are markedly outcompeting other cyanobacterial species occurring at these water depths may suggest that allelopathic compounds from *Synechococcus* spp. can have an adverse effect on co-occurring species. Indeed, such effects have been observed for freshwater *Synechococcus* spp., which were able to impact the growth of other freshwater cyanobacteria or green algae (Kovács et al. 2018; Bubak et al. 2020). Adverse effects caused by allelopathic compounds from marine *Synechococcus* spp. were also observed on various marine invertebrates (Costa et al. 2015) as well as other bacterial species (Santhakumari et al. 2016). In conclusion, the described effects suggest a widespread production of allelopathic compounds by *Synechococcus* species that can even influence bacteria, plants and invertebrates.

In most cases, co-occurrence of species occupying the same or a similar ecological niche leads to dominance of one species, largely depending on the individual species competitive advantage e.g. nitrogen fixation, uptake of inorganic phosphorus, regulation of buoyancy, allelopathic compounds, etc. as also shown by suggesting that if *Synechococcus* outcompetes *Planktothrix* in its ecological niche, amongst other factors allelopathic compounds such as toxins may be used to compete against the respective other species. Indeed, allelopathic activity is one of the major

competitive strategies of freshwater *Synechococcus* against coexisting phytoplankton species (Bubak et al. 2020), further supporting the potential role of picocyanobacterial exudates in the competition with other cyanobacteria, such as *Planktothrix*. Counter-intuitively, some freshwater *Synechococcus* spp. possess a positive allelopathic activity towards *Microcystis* spp. and no effect on *Phormidium* spp., whereby *Phormidium* spp. like *Planktothrix* spp. is part of *Oscillatoriaceae* (Bubak et al. 2020), suggesting that *Synechococcus* spp. could even promote growth of toxin producing species. Such complex interplay of species competition implied for Lake Constance emphasizes the urgent need of further studies regarding co-occurrence and dominance of *Synechococcus* spp. relative to *Planktothrix* and *Microcystis* spp. in this lake.

The geographical setting of the *Überlingen* embayment of Lake Constance with its minor wind influence (Figure SI.6 and SI.7) is considered an additional factor that could favor continuous development of deep-water *Planktothrix* spp. populations. Indeed, light winds and convective mixing are highly important in the seasonal cycling of *P. rubescens* communities within a strongly stratified medium-sized lake (Fernández Castro et al. 2021). In consequence, this would mean that toxin producing *Planktothrix* spp. possibly could establish themselves as the predominant cyanobacteria species at deeper water levels and represent a serious concern for the quality of the *Überlingen* embayment of Lake Constance as a drinking water resource in the not-too-distant future.

Conclusions

Our study characterized the deep-water red-pigment maxima (DRM) in the *Überlingen* embayment of Lake Constance in 2019 and 2020 as being dominated by phycoerythrin-rich picocyanobacteria, namely *Synechococcus rubescens* and *Cyanobium gracile.* Unlike other pre-alpine lakes, the DRM in Lake Constance is not dominated by the phycoerythrin-rich, filamentous and often toxin-producing *Planktothrix rubescens*. Indeed, alignment with sequences from the Ernst *et al*. study (2003) demonstrated high sequence similarity and suggested that the same species have been dominant in 15 – 20 m depth in Lake Constance for the past 20 years. However, we could confirm the reports from 2016 that *Planktothrix* spp. does occur in Lake Constance in the years 2019 and 2020, albeit yet at very low relative

abundances. Nevertheless, microcystin concentrations of up to 1.5 ng/L were detected by UPLC-MS/MS in 2019, which appeared to be produced by *Planktothrix* and/or *Microcystis* spp. Hence, at present, Lake Constance seems to have a rather stable deep-water cyanobacterial ecosystem predominated by *Synechococcus* spp., although the geographical setting as well as continued climate warming could favor the development and a steady predominance of toxin producing *Planktothrix* spp. This highlights the importance of a future monitoring program for Lake Constance, with emphasis on sequencing-based cyanobacterial community studies, as well as the importance of competition studies of the different cyanobacterial taxa in relation to physiochemical and biological parameters of the lake, particularly in respect to the ongoing climate change. Monitoring programs and hypothesis driven competition studies may in return provide the required database to predict future deep-water mass occurrences of toxin producing cyanobacteria and thus help to secure Lake Constance as the drinking water resource for millions in the future

Materials and Methods

Sample collection

Samples were taken every two weeks from 20. June to 08. October in 2019 and 07. July to 15. September in 2020. The initial sampling in 2019 contained only one replicate/day, while everything was sampled in triplicates during 2020. A bbe Moldaenke FluoroProbe (FP) was used to determine the deep red maximum (DRM). Samples were collected at that depth every other week in Upper Lake Constance (47.7571°N 9.1273°E). Water was filtered through a 200 µm filter to exclude zooplankton and larger particles. Two liters of water were then filtered on one Whatman GF6 glass fiber filter while applying 2 bars pressure for collection of bacterial biomass. Filters were stored at -20°C until analysis. This sampling method was applied to DNA extractions 2019 and toxin extractions in 2019 and 2020. In 2020, DNA samples were collected on 0.2 µm polycarbonate filters.

DNA extraction and PCR

DNA from 2019 samples was extracted using the ZYMO Research Fecal/Soil/Microbe Microprep kit following the manufacturer's instructions. DNA of 2020 samples was extracted using a phenol/chloroform/isoamyl alcohol protocol adapted from Rusch et al., 2007 (Rusch et al. 2007) and the JGI protocol (William et al. 2012). Standard PCR was performed with Taq polymerase using 2X Taq MasterMix (NEB) and 30 cycles. The microcystin synthesis gene *mcyE* was amplified with the HEPF/R primer set (Jungblut and Neilan 2006) and DNA from cultured *Microcystis aeruginosa* strain 78 was used as a positive control for presence of *mcyE*. Planktothrix-specific primer pairs PcPI+/- (PC-IGS) and peamso+/- (*mcyA*) (Kurmayer et al. 2004) were used for additional PCR reactions, with *Microcystis aeruginosa* strain 78 as a negative control and *Planktothrix rubescens* strain 101 as a positive control.

Table I.1. Primers used in this study.

name of forward and reverse primers	sequence (5'-3')	target	reference
PcPI+	TGCTGTCGCCTAATTTTTCA	*cpcB-cpcA* PC-IGS	Kurmayer et al., 2004
PcPI-	CCACTGATCAGGCTGTCAGA		
peamso+	ATCAAACAGATGTACTGACAGGT	*mcyA*	Kurmayer et al., 2004
peamso-	AGGCCAGACTATCCCGTT		
HEPF	TTTGGGGTTAACTTTTTTGGGCATAGTC	*mcyE*	Jungblut & Neilan, 2006
HEPR	AATTCTTGAGGCTGTAAATCGGGTTT		
16S rDNA PTX fw	ATCCAAGTCTGCTGTTAAAGA	16S rDNA (only *Planktothrix* spp.)	Ostermaier & Kurmayer, 2009
16S rDNA PTX rv	CTCTGCCCCTACTACACTCTAG		
16S rDNA PTX TaqMan[1]	AAAGGCAGTGGAAACTGGAAG		
mcyBA1 PTX fw	ATTGCCGTTATCTCAAGCGAG	*mcyBA1* (only *Planktothrix* spp.)	Ostermaier & Kurmayer, 2009
mcyBA1 PTX rv	TGCTGAAAAAACTGCTGCATTAA		
mcyBA1 PTX TaqMan[1]	TTTTTGTGGAGGTGAAGCTCTTTCCTCTGA		
CYA359F	GGGGAATYTTCCGCAATGGG	16S rRNA (Cyanobacteria)	Nübel et al., 1997
CYA784R	ACTACWGGGGTATCTAATCCC		

[1] TaqMan probes contain 5' FAM (6-carboxyfluorescein, fluorescent reporter dye) and 3' TAMRA (6-carboxy-tetramethylrhodamine, fluorescent quencher dye).

RT-PCR/TNA

Quantitative Taq Nuclease Assays (TNA or TaqMan PCR) were performed with primers specific to *Planktothrix* spp. (Ostermaier and Kurmayer 2009). We quantified both *Planktothrix*-specific 16S rDNA and *mcyBA1*, which encodes the first adenylation domain of *mcyB* and is indicative of all genotypes containing the *mcy* gene cluster. Both probes contained 5’ FAM as a fluorescent marker and 3’ TAMRA as a quencher dye. Each reaction contained 50 ng template DNA, 200 nM of primers and probes and

KAPA probe fast MasterMix. Amplification and quantification were carried out as triplicates in a Bio-Rad CFX96 cycler with the following protocol: 10 min at 95°C to activate the hot start polymerase were followed by 50 cycles of 15 s at 95°C, 60 s at 51.5°C for 16S and 57°C for *mcyBA1* and 60 s at 68°C and a final elongation for 5 min at 68°C. The standard curve contained *Planktothrix* DNA mixed with *Microcystis* DNA in different ratios (0.00001-100% *Planktothrix* DNA diluted in *Microcystis* DNA). Data analysis using linear regression from the standard curve was performed with BioRad CFX manager, Microsoft Excel and GraphPad Prism 5.

Sanger sequencing

Sanger sequencing was performed with amplicons produced by the HEPF/R primer pair to identify the main microcystin producer in the samples. PCR amplicons were purified using QIAquick PCR cleanup kit and sent to Eurofins Genomics for analysis. The identity of the obtained sequences was determined using Nucleotide BLAST – megablast (Altschul et al. 1990).

Amplification and Illumina sequencing

Amplification of the V3-V5 hypervariable regions and the cyanobacterial specific V3-V4 hypervariable regions of the 16S rRNA gene was performed with 0.02 U/ µl of Phusion High Fidelity DNA polymerase, 1X Phusion HF Buffer, 200 µM of dNTPs (New England Biolabs, USA) and 0.5 µM of each primer. Primers targeting the V3-V5 hypervariable regions were 357F and 926R (Schuurman et al. 2004; Walters et al. 2016). Cyanobacteria specific primers were CYA359F and CYA784R (Nübel et al. 1997). Each PCR comprised an initial denaturation step of 3 min at 98°C, then 30 cycles of denaturation for 45 s at 98°C, annealing for 20 s at 62.4°C (V3-V5) or 60°C (cyanobacteria specific) and extension for 8 s at 72°C, then a final extension step for 5 min at 72°C. Extracted DNA was added at a final concentration of 0.12 ng/µl. No purification step was performed, the PCR products were directly sent to Eurofins Genomics for sequencing using Illumina MiSeq 2*300 bp with the Microbiome Profiling -Indexing only- package. The data presented in this study are accessible on NCBI under the bioproject number PRJNA727470.

Bioinformatics pipeline

The analysis was carried out using the already merged dataset provided by Eurofins Genomics. The expected fragment sizes were 569 bp for the V3-V5 amplicon

and 425 bp for the V3- V4 cyanobacterial specific amplicon. Reads were first trimmed using Trimmomatic (Bolger et al. 2014), removing all reads with a phred quality below 3 for the start and the end of the reads, below an average quality of 10 on a window of 3 base within the reads and below a size of 500 bp for the V3-V5 amplicons and 380 bp for the V3- V4 amplicons. FastQC was used for quality control of the reads before and after trimming (Simon 2010). The following steps were performed using QIIME2 2019.10 (Bolyen et al. 2019). Filtration of chimeras by consensus method, denoising and dereplication of the quality reads were performed using the denoise and dereplicate single-end sequences (Dada2, denoise-single) and a reads learn of 2.000,000 reads for the training error model (Callahan et al. 2016). Quality trimming was already performed using Trimmomatic so no trimming step was performed here. Taxonomic affiliation was achieved using the classify-consensus vsearch program and the databases SILVA_138 and Greengenes with a percentage identity of 80%, 90%, 97% and 100%. Taxonomic results were merged and the highest percentage identity taxonomic affiliation was kept. Reference databases were previously trained using the feature-classifier extract reads script with the sequences of the primers used for the amplification of the hypervariable regions. After training, the databases only contained the part of interest of the 16S rRNA gene for taxonomic assignation, V3-V5 hypervariable regions for the general 16S rRNA gene amplification and V3-V4 hypervariable regions for the cyanobacteria specific primers. Taxonomy was mostly consistent between SILVA and Greengenes databases.

Phylogenetic analysis

A megablast (highly similar sequence) was performed on the sequences affiliated to the genus *Planktothrix, Microcystis* and the most abundant *Synechococcus* using the 16S rRNA sequences reference database on NCBI. The top ten hit sequences and description table were extracted. *Synechococcus* sequences from Lake Constance analysed and sequenced by Anneliese Ernst (Ernst et al. 2003) were downloaded from the NCBI database. Two other 16S rRNA gene sequences, belonging to *Synechococcus rubescens* and *C*yanobium *gracile*, were also collected from NCBI and used as reference for the phylogenetic analysis, as previously done by Anneliese Ernst. These collected dataset was then compared to the most abundant *Synechococcus* sequences of 2019-2020. Sequences were merged with our sequences into a fasta file and aligned using SeaView (Gouy et al. 2010). IQ-TREE

2.1.2 (Minh et al. 2020) was used to calculate the phylogenetic tree by maximum likelihood using 1000 bootstrap parameters. The model of nucleotide substitution used was TPM2+I+F (Minh et al. 2020), determined as best-fit model by ModelFinder (Kalyaanamoorthy et al. 2017) based on the Bayesian information criterion scores and weights. The tree was visualized on R using the package ggtree (Yu 2020).

Statistical and network analysis

Statistical analysis was performed with R software (R core team 2021) using the package Phyloseq (McMurdie and Holmes 2013), vegan (Oksanen et al. 2015) and visualized with the package ggplot2 (Wickham 2016). Features represented by less than 3 reads in below 20% of the samples were discarded. Chloroplast affiliated features were removed from both datasets, cyanobacterial affiliated taxa were also removed from the cyanobacterial specific dataset only. No rarefaction has been applied on the dataset (McMurdie and Holmes 2014). After filtration, the lowest number of reads in a sample was 19,175 for the cyanobacteria specific dataset and 24,329 reads for the general 16S rDNA dataset in 2019. The 2020 dataset consisted only of the cyanobacteria specific amplicon and the lowest number of reads was 24,341. This number of reads allowed us to have confidence in catching all the richness present in our dataset as the rarefaction curves showed that we were in the stationary phase of the curve. Read count matrix was transformed in relative abundance by dividing the reads affiliated to one taxa by the total number of reads in the sample (x/sum(x)) and a logarithmic transformation was also applied. To avoid the presence of 0 in the matrix, one artificial read was added to each cell of the data (Log10(x+1)). Alpha diversity was analysed using the Observed richness, Pielou's evenness index, Shannon diversity index and Inverted Simpson diversity index. Community composition was observed using barplot, jitter plot and krona plot using the relative abundance data and heatmap with the Log10 transformed data. Statistical analyses were performed using non parametrical statistical test, sequencing data does not fulfill parametric test requirements (e.g. data normality and homoscedasticity and independence of observation) using a p-value and False discovery rate threshold of 0.05. Kruskal-Wallis one-way analysis of variance was performed on the relative abundance sub-data of the main Synechococcus taxa to test if at least one taxa dominate the community. The hypothesis that the relative abundance distribution of the tested taxa are equal is the null hypothesis (H0) and the alternative hypothesis

(H1) is that the relative abundance distribution of the tested taxa are not equal. If the Kruskal-Wallis test rejected H0, a post hoc test using the Conover-Iman squared rank test with a Benjamini Yekutieli p-value adjustment procedure was performed to determine which taxa or group of taxa dominate the community. Benjamini Yekutieli procedure was chosen as p-value adjustment because of the dependency between taxa's relative abundance observation. Barplot of the taxa of interest (*Planktothrix spp.*, *Microcystis spp., Synechococcales*) and correlation observation of the main node from the network analysis were produced. Network analysis was made using the network tool inference CoNet (v. 1.1.1 beta) (Faust and Raes 2016). A read count matrix, each column representing a sample and each row representing a single ASV, and a metadata featuring the Microcystin toxin concentration obtained by UPLC-MS/MS were loaded. No filtration was applied as the low abundant taxa were filtered from the matrix beforehand. Data was normalized by relative abundance calculations. Pair-wise associations were calculated using the Pearson, Spearman and Bray Curtis measures. The maximal number of positive and negative edges was set at 200 for each calculated network. Networks were calculated with 1000 row-permutation and bootstrap. P-value obtained for each measures and edges were merged using the Brown's method (Brown 1975) followed by a Benjamini-Hochberg false discovery rate correction. Edges supported by at least 2 of the three measures and a q-value below 0.05 were kept. Network visualization was performed using Cytoscape 3.8.2 (Shannon et al. 2003).

Toxin extraction and analysis

Toxins were extracted from filters using a methanolic extraction method and subsequently analysed via UPLC-MS/MS (Weisbrod et al. 2020). Briefly, 3 mL 50% (v/v) aqueous methanol was added to each filter and soaked for 30 min at room temperature. After vigorous mixing (vortex) for 10 min, samples were sonicated in an ultrasonic bath for 15 min before a centrifugation at maximum speed for 10 min. The supernatant was collected in a separate tube and the above steps excluding the initial 30 min soaking time were repeated twice with the remaining pellet. The pooled supernatant was dried overnight using a speed-vacuum system (Univapo 100H) and resuspended in 250 µL 50% aqueous methanol. Toxin samples were stored in glass vials at -20°C until analysis.

Concentrations of different MCs were measured via UPLC-MS/MS with an internal standard containing deuterated MC-LR and MC-LF (D_5-MC-LF and D_7-MC-LR) (Shannon et al. 2003; Weisbrod et al. 2020). Samples were ionized using electrospray ionization, a capillary voltage of 3 kV as well as a nebulizer pressure of 7.0 bar and analysed with a UPLC-MS/MS system (Acquity H-class liquid chromatograph with Acquity BEH C18 column with a corresponding guard column (kept at 40°C) and a Waters XEVO TQ-S mass spectrometer). Solvents A and B were composed of 10% and 90% acetonitrile, respectively and 100 mM CH_2O_2 and 6 mM NH_3. Inital conditions were 25% B, held for 30s, then 45% B within 30s, 60% B within 180s and 99% B within 12s, which was held again for 30s. Prior to application of the next sample, column was re-equilibrated to 25% B over 78s and held for 60s. Injection volume was 5 µL. As described in Altaner et al., 2019, simultaneous analysis of MC congeners was achieved using five analysis windows (Altaner et al. 2019; Weisbrod et al. 2020).

Evaluation of bbe Moldaenke FluoroProbe data

For data acquisition, the Moldaenke FluoroProbe (FP) was slowly lowered from the water surface to 100 m depth, then slowly pulled up again while collecting data both ways. FP data was preprocessed and later plotted in 3D plots using MATLAB R2020b (MATLAB 2020). Full spectra were chopped below 1 m depth and above 100 m depth and subsequently sorted. Data were interpolated to 0.1 m steps and 570 nm fluorescence measurements and cryptophyta content (in µg/L) data were extracted. Plotting was achieved using the MATLAB standard functions *surf* (3D) and *plot* (2D), peaks were analysed using *max*.

Author Contributions: Conceptualization, E.R., C.F., D.R.D. and D.S.; methodology, E.R. and C.F.; formal analysis, E.R. and C.F; data curation, E.R., C.F. and J.S.; writing—original draft preparation, E.R. and C.F; writing—review and editing, D.R.D. and D.S.; project administration, D.R.D. and D.S.; funding acquisition, D.R.D. and D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DFG Research Training Group R3 - Responses to biotic and abiotic Changes, Resilience and Reversibility of Lake Ecosystems (GRK 2272) and MTU Friedrichshafen GmbH (MTU Umweltstiftung).

Data Availability Statement: The Illumina sequence data presented in this study is accessible at NCBI under the bioproject number PRJNA727470.

Acknowledgments: We would like to highlight the excellent support received from Julia Schmidt, Microbial Ecology and Limnic Microbiology, with regard to DNA extraction from samples she helped organize from the *Überlinger* embayment. The assistance with UPLC-MS/MS by Sarah Krassnig and Aswin Mangerich the help in the laboratory by Leon Walther is thankfully acknowledged0. We would also like to thank Beatrix Rosenberg for access to long-term data and Alfred Sulger, Angelika Seifried, Pia Mahler and Josef Halder at the Limnological Institute of University of Konstanz for managing the ship cruises and Anneliese Ernst for helpful discussions. Further, we would like to thank the MTU Umweltstiftung and all members of the RTG-R3, especially Tina Romer and Frank Peeters, for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Chapter II
Description of a 'plankton filtration bias' in sequencing-based community analysis and of an Arduino microcontroller-based flowmeter device that can help to resolve it

Corentin Fournier[1*], Alexander Fiedler[3], Maximilian Weidele[2], Harald Kautz[2] and David Schleheck[1*]

[1] Microbial Ecology and Limnic Microbiology, Limnological Institute, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

[2] Scientific Engineering Service, University of Konstanz, D-78457 Konstanz, Germany

[3] Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Gløshaugen, Norway

*Shared corresponding authorship:
corentin.fournier@uni-konstanz.de, david.schleheck@uni-konstanz.de

Manuscript in preparation for submission

Target Journal: PLOS ONE

Keywords: Arduino, plankton filtration bias, microbial plankton community, flowrate, NGS

Abstract

Sequencing-based diversity studies of aquatic picoplankton (‘bacterioplankton’) communities using size-class filtration, DNA extraction, PCR and sequencing of phylogenetic markers, require a robust methodological pipeline. Biases have been demonstrated essentially at all levels, DNA extraction, primer choice and PCR. Even different filtration volumes of the same water sample and, thus, different biomass loading of the filters, can distort the sequencing results. We attempted to normalize biomass loading during filtrations, independent of plankton density in water samples, and designed a microcontroller-based flowmeter device for monitoring the decrease of initial flowrate as a proxy for increasing biomass loading and clogging of the filters. Test filtrations using freshwater plankton (Lake Constance) were done, DNA was extracted and an amplicon of the 16S rDNA was sequenced. The impact of filter loading was examined both at the community and at the taxonomic level. First, we confirmed that different filtration volumes used for the same water sample affects the outcome of the sequencing results. Significant differences were visible in alpha and beta diversities and across all taxonomic ranks. Taxa most affected by filtration volume were typical freshwater Actinobacteria and Bacteroidetes, increasing up to 38% and decreasing up to 29% in relative abundance, respectively. Second, a lake water sample was filtered undiluted and three-fold diluted, and each filtration was stopped at the same flowrate-threshold, i.e., when the flowrate was reduced to 50% of the initial flowrate. The three-fold diluted sample required three-fold filtration volumes, while differences for both samples across all taxonomic ranks were not statistically significant. This work confirms a volume/biomass-dependent bacterioplankton filtration bias for sequencing-based community analysis and provides an improved procedure for controlling the biomass loading during filtrations and recovery of equivalent amounts of total DNA from samples with different plankton densities, for example, across depth profiles and along seasonal cycles of temperate lakes.

Introduction

The advent of 'omics' technologies has enabled microbial ecology research to go a huge step forward in the disentanglement of the complexity of the highly diverse and dynamic environmental microbial communities on this planet. This also applies to the plankton in pelagic marine and freshwater ecosystems, particularly the microscopic, so-called nanoplankton as a representation of the, predominantly, phytoplankton (herein, the organismal group with 180 µm – 5 µm in diameter) as well as picoplankton communities as a representation of the, predominantly, single-cell bacterioplankton (herein, the organismal group with 5 µm – 0.2 µm in diameter), which each on their own and by their interactions are playing key roles for functioning and stability of these ecosystems (Pierrou 1976; Jones 1998; Jetten 2008; Madsen 2011). Total DNA extraction and PCR amplicon sequencing of phylogenetic markers (e.g., fragments of 16S and 18S rRNA genes or other markers) enables the analysis of the nano- and picoplankton community in respect to composition and relative abundance and the identification of the active members within these communities, while metagenomic sequencing allows for an evaluation of the functional repertoire encoded within these communities (Hugoni et al. 2013b; Hugoni et al. 2017; Chafee et al. 2018); metatranscriptomics and proteomics allow for a detection of the traits that are actually expressed under any particular environmental condition (Delhomme et al. 2015; Grassl et al. 2016). However, the high sensitivity of these modern techniques and the long methodological pipelines involved, are accompanied by increased sensitivity to biases that may be introduced by any of the methodological steps, e.g., through DNA extraction, primer choice and PCR (Krakat et al. 2017), and DNA template dilution for the PCR (Wu et al., 2010), as well as library preparation for sequencing (Berry et al. 2011; van Dijk et al. 2014; Krakat et al. 2017).

For analyzing nano- and picoplankton communities *via* omics methods, another key step is the sampling process. Plankton cellular biomass needs to be isolated from the water column in sufficient quantity, e.g., for extraction of total nucleic acids. Filtration is the principal process most commonly used for collecting nano- and picoplankton biomass. Usually, a pre-filtration is applied to remove zooplankton and other larger particles using a pore-size filter in the range 100 – 200 µm (Filker et al. 2016; Llorens-Marès et al. 2016), and the sample is then passed through either one, or a series of different, smaller pore-size filters, in order to collect the nano- and/or

picoplankton size-class contingents. An intermediate pore-size filter, for example 5 µm pore diameter, can be considered to collect the nanoplankton, i.e., predominantly the eukaryotic (protist) phytoplankton, but also particle-associated and filamentous bacteria, archaea or fungi (collectively, nanoplankton) (Díez et al. 2001; Fuchsman et al. 2012; Allen et al. 2012; Bradford et al. 2013; Hugoni et al. 2013b; Hugoni et al. 2013a), while a smaller pore-size filter, usually in the range 0.1 - 0.22 µm diameter, can be considered to collect single-cell prokaryotes and picoeukaryotes (collectively, picoplankton) (Díez et al. 2001; Eiler et al. 2012; Hugoni et al. 2013b; Hugoni et al. 2013a; Baltar et al. 2016). Further, a variety of filter materials is existing (e.g., glass fiber, polycarbonate, polyethersulfone) and their efficiency in collecting plankton biomass has been subject of discussions (Taguchi and Laws 1988; Knefelkamp et al. 2007). Finally, the filtration volume is also variable, for example, in between studies (from 300 ml up to 25 L) (Eiler and Bertilsson 2004; Llorens-Marès et al. 2016; Hugoni et al. 2017), within the same study (Eiler and Bertilsson 2004), or in between different types of omics-analyses done on the same water sample (Frias-Lopez et al. 2008; Hunt et al. 2013).

It is easy to rationalize that using different pore-size filters and/or filter materials will result in different final outcomes of DNA-based diversity analyses (Lee et al. 1995). But can also the filtration volume influence to outcome of the results under a given filtration setting? And if so, in which proportion? Indeed, Padilla *et al*., 2015 analysed the picoplankton community structure for an identical water sample (seawater) by 16S rDNA amplicon sequencing in dependence on the filtration volume, and their results showed significant differences in the relative abundances of bacterial community members (Padilla et al. 2015).

We considered whether the cause of such a volume-dependent ‘plankton filtration bias’ might indeed be a build-up of a so-called ‘filtration cake’: during the filtration process, the biomass collected on the filters might increasingly act as an additional filter, thereby retaining microorganisms on the filter that would normally belong into the filtrate and, thus, onto the smaller pore-size filter, if a serial size-class filtration is applied.

We also considered whether one way of normalizing for such a sample volume/ plankton density-dependent filtration bias would be to collect each the same amount

of total biomass on each filter, independent of the plankton density of the water samples. Therefore, we build and tested an Arduino-microprocessor based flowmeter device in order to monitor the filter loading, as the decrease of initial, maximal flow rate during the filtration, and in order to stop the filtrations each at a same minimal-flowrate threshold and, hence, potentially at about the same biomass loading (Figure II.1). We used samples from Lake Constance, when sequentially filtering the nano- and picoplankton size classes, first, from 180 µm to 5.0 µm (nanoplankton) and, second, from 5.0 µm to 0.2 µm (picoplankton). Then, the single-cell bacterial community structures represented on the 0.2-µm pore size filter were examined, using 16S rRNA gene amplicon sequencing. Our first objective was to confirm the volume-dependent filtration bias as revealed by Padilla *et al*., 2015 (see above). Second, we tested another Lake Constance water sample at two different plankton densities, i.e., the original, undiluted sample and a three-fold diluted sub-sample, when using our flowmeter to stop the filtration each at the same flowrate-threshold. Hence, the second experiment was conducted to evaluate whether the flowmeter device may be helpful in normalizing a density/volume-dependent filtration bias.

Materials and Methods

Study sites and sampling campaign

Lake Constance is a deep (maximal depth, 251 m) oligotrophic pre-Alpine lake situated in the northern part of the Alps (47°35´N 9°28´E). The lake is composed of three water bodies called *Obersee* (Upper Lake*), Untersee* (Lower Lake) and *Seerhein* (Lake Rhine); the Lake Rhine is connecting the Upper Lake to the Lower Lake. In this study, samples were taken at the routine sampling site *Wallhausen* situated in the *Überlinger See* (47.7571°N 9.1273°E), a fjord-like northwestern arm of the main basin of Upper Lake. Integrated samples of the epilimnion (0 – 20 m) were collected by boat using an integrating water sampler (model IWS II, Hydro-Bios, Germany). The samples used for the first experiment were collected on 08. November 2017 and the samples for the second experiment on 14. May 2019. The water samples were stored in stainless steel barrels (19 L soda-kegs).

Filtration using a self-constructed Arduino-based monitoring device (flowmeter)

Filtrations were carried out directly on the boat. The barrel containing the water sample was connected to a pressurized air tank and overpressure of 2 bar was applied. The barrel was connected *via* valve and PVC tubing to a series of three in-line filter holders (Swinnex 47 mm filter holder; Millipore), carrying (*i*) a 180-µm filter (hydrophilic nylon net, 47 mm diameter; Millipore) to remove zooplankton and larger particles, and (*ii*) a 5.0-µm and (*iii*) a 0.2-µm polycarbonate membrane filter (Isopore, 47 mm diameter; Millipore) to collect the nanoplankton and picoplankton, respectively.

At the outlet of the last filter holder, a mini-flowmeter (model FCH-m-PP –3.0 LPM 82202739; B.I.O-TECH, Vilshofen, Germany) was connected, from which pulses were counted and flowrates calculated (in ml/min) by a programmed Arduino I/O device; the calibration factor had to be set directly in the Arduino program. A detailed description of the most basic, low-cost, open-source version of this Arduino flowmeter device, and its electronic components, a circuit diagram, and the software used, is available at https://github.com/Uni-Konstanz-WWE. This flowmeter device comprised also an LCD screen that displayed, (*i*) the maximal flowrate (F_m) recorded right at the start of a new filtration (i.e., directly after the valve had been fully opened), (*ii*) the current flowrate during the filtration (F_c), and (*iii*) the flowrate threshold (F_t) at which the filtration has to be stopped; F_t is calculated by the Arduino from F_m using a

threshold factor (e.g., 50% of F_m), and this threshold factor had to be set directly in the Arduino program. When the flowrate had decreased and F_t was reached, the Arduino gave both an optical signal (LED) and acoustic signal (piezo beeper), in order to alert the operator to stop the filtration by closing the valve. Optionally, the Arduino was streaming the monitored flowrates to a computer which stored the data in an Excel sheet with graphical display, in order to follow the flowrate on screen at any time during the filtration procedure (Figure II.1). Such a basic, low-cost flowmeter device, which can be operated in the field also by battery and without USB connection to a computer, was used for the experiments described in this paper. A much further developed version of the device comprises six flowmeter channels, a casing with battery, a number pad for adjusting flowrate thresholds and other parameters, and additional I/O ports for data storage on USB stick and for controlling electrically-actuated valves for automation of the filtration procedure (Figure II.1); it is currently used for our routine sampling of Lake Constance nano- and picoplankton.

For the first experiment, we aimed at confirming the observations made by Padilla *et al*., 2015 (see also Introduction), but without prior definition of the volumes to be filtered. Instead, we programmed the flowmeter to alarm us for stopping the filtrations when either 66%, 50%, 25% or 10% of the initial, maximal flowrate (F_m) had been reached, using each the same water sample from Lake Constance (see above); for each threshold, filtrations were done in triplicates (n = 3). For the second experiment, we simulated a change in plankton density, to evaluate whether the flowmeter can indeed be helpful in a normalization of the volume/density-based filtration bias. First, a water sample from Lake Constance was filtered using the flowmeter at a flowrate-threshold of 50% F_m; the filtration volumes were recorded (590 ml +/- 64 ml). Then, the lake water was diluted 1:3 with autoclaved distilled water and a new series of filtrations (n = 4) was started, using the flowmeter flowrate-threshold of 50%. Another series of filtrations (n = 4) was done volume-controlled, i.e., until the same volume was reached as for the filtration of the undiluted samples (590 ml). Hence, this gave three sets of samples: undiluted (UD), diluted flowrate threshold filtration (TF1/3) and diluted fixed volume-filtration (VF1/3).

When the filtrations had been stopped, the filter holders were opened and each filter membrane was carefully curled up using two forceps, and transferred into a 15 ml conical tube (Eppendorf, Hamburg, Germany) so that the biomass-containing side

of the filter was pointing to the inside of the tube. Then, the biomass was immersed each with 3 ml of lysis buffer (50 mM Tris-HCl buffer pH 8.0, 50 mM EDTA, 50 mM EGTA) and the tube stored at -20 °C.

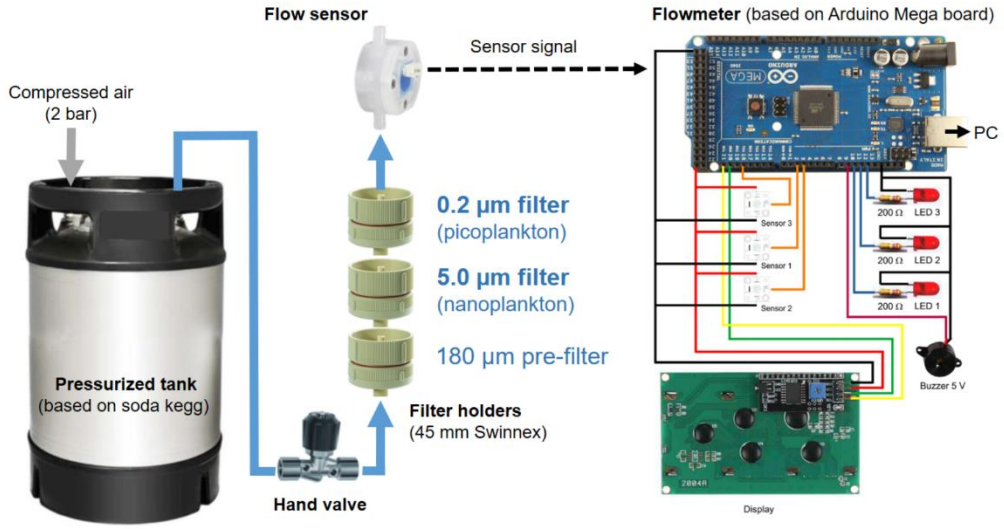

Figure II.1. Schematic of the filtration setup and flowmeter device used in this study. We used overpressure filtration for plankton size-class filtration with three filters in series, (*i*) a 180-µm nylon net pre-filter to remove zooplankton and larger particles, and (*ii*) a 5.0-µm and (*iii*) a 0.2-µm pore-size polycarbonate membrane filter, in order to collect the nanoplankton and picoplankton, respectively, for DNA extraction. The flowrate of the filtrate was monitored continuously during the filtration, by using a flow sensor and a flowmeter device based on an Arduino board. The Arduino was programmed to record the initial, maximal flowrate (F_m) when a filtration had been started with new filters, and to monitor the continuous decrease of flowrate caused by increasing biomass loading and clogging of the filters. Once a pre-defined (programmed) minimal flowrate threshold (F_t) has been reached, the Arduino was configured to give an optic (LED) and acoustic (beeper) alarm to the operator, so that the filtration could be stopped manually using a valve, for example, each at $F_t = 50\% F_m$ across all water samples filtered with variable plankton densities. The flowrate data was also streamed to a PC for recording (Figure II.2A). Note that the illustration is not to scale. Details on components and operation of the filtration monitoring setup can be found in the Material and methods section, and details on the Arduino configuration and the Arduino code at Github (https://github.com/Uni-Konstanz-WWE).

DNA extraction, PCR amplification and Illumina-amplicon sequencing

Our study focused on the single-cell bacterioplankton community and therefore; DNA extraction were applied on the 0.2 µm polycarbonate membrane filters. The DNA extraction protocol used was adapted from (Rusch et al. 2007) and the JGI bacterial DNA extraction protocol (William et al. 2012). After thawing of the samples at room temperature, 200 µl of 0.1- and 1-mm diameter zirconium beads each (Carl Roth, Karlsruhe, Germany) were added to each tube. Then, the tubes were treated in an ultrasonic water bath (Sonorex super RK 510, Bandelin, Germany) for 1 min, followed by 15 min of vortexing at full speed in a horizontal tube holder. In the next step, 150 µl of freshly prepared lysozyme solution (final concentration 2.5 mg/ml) was added, and the tubes incubated for 1 h at 37 °C by horizontal shaking at 1,400 rpm (Thermomixer comfort, Eppendorf, Hamburg, Germany); then 315 µl of 10% sodium dodecyl sulfate (SDS) (final concentration, 1%) and 31.5 µl of freshly prepared proteinase K solution (final concentration 500 µg/ml) were added, followed by an incubation the tubes for 1 h at 55°C in a water bath. After one hour of incubation, proteinase K was added again at the same concentration and the solution incubated again for 1 h at 55°C. Finally, 236 µl of 5 M NaCl solution was added, the solution vortexed, followed by addition of 236 µl of CTAB/NaCl-solution (10% CTAB, 0.7 M NaCl; preheated at 65°C); the solution was again vortexed, and then incubated for 10 min at 65°C. The DNA was purified by phenol-chloroform extraction. Therefore, 1 vol. of phenol/chloroform/isoamyl alcohol (25:24:1 vol %) (Carl Roth, Karlsruhe, Germany) was added, the suspension mixed by vortexing, and the tubes centrifuged for phase separation at 13,000 rcf for 20 min at 4°C. After transfer of the supernatant into a new 15 ml tube, 1 vol. of chloroform/isoamyl alcohol (24:1 vol %) was added, the suspension mixed by vortexing, and the tube centrifuged for phase separation at 13,000 rcf for 20 min at 4°C. The supernatant was transferred into a new 15 ml tube. 0.5 µl of glycogen (Thermo Fisher Scientific, USA) and 0.7 vol. of isopropanol were added to the tube. The solution was mixed well and incubated for 15 min at -20°C. The precipitated DNA was collected by centrifugation at 15,000 rcf for 25 min and 4°C. Isopropanol was removed and the DNA pellet was washed using 500 µl of ice-cold 70% ethanol. After a final centrifugation of 5 min, the supernatant was removed and the DNA dried at air for 5 min. Then, 50 µl of PCR-grade water was added and the DNA dissolved. Finally, the DNA concentration was measured using a Nanodrop

2000c spectrophotometer (Thermo Fisher Scientific, USA) and the quality of the DNA evaluated by agarose gel electrophoresis.

Amplification of the V3-V5 hypervariable regions (Huttenhower et al. 2012; Boers et al. 2015) of the 16S rRNA gene was performed with 0.02 U/ µl of Phusion High Fidelity DNA polymerase, 1X Phusion HF Buffer and 200 µM of dNTPs (New England Biolabs, USA). DNA template was added at a final concentration of 0.12 ng/µl. The primers pair used was 357F and 926R with universal adapter, required for the second PCR, added to their 5' end. The primer sequences, with universal adapter, are 5´–TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-CTCCTACGGGAGGCAGCAG– 3´ for 357F and 5´–GTCTCGTGGGCTC GGAGATGTGTATAAGAGACAG-CCGYCAATTYMTTTRAGTTT– 3´ for 926R (Schuurman et al. 2004; Parada et al. 2016; Walters et al. 2016), at a final concentration of 0.5 µM each. The following PCR program was used on a T100 Thermal cycler (Bio-Rad, USA): first denaturation for 3 min at 98°C; 30 cycles of denaturation for 45 s at 98°C; annealing for 20 s at 62.4°C, and extension for 8 s at 72°C; final extension for 5 min at 72°C. The PCR products were sent to Eurofins GATC Biotech for amplicon sequencing, using the Illlumina MiSeq 2*300 bp with the NGSelect Amplicons 2nd PCR package. The reads were merged by Eurofins. The expected V3 – V5 amplicon size was 569 bp.

Bioinformatics pipeline

The sequence libraries were trimmed using Trimmomatic (Bolger et al. 2014), removing all reads below 500 bp, with a phred quality below 3 for the start and the end of the reads and below an average quality of 10 on a window of 3 base within the reads. FastQC was used to check the quality of the reads before and after the trimming (Simon 2010); before to test the trimming parameters, and thereafter to verify that the reads had the necessary quality for the downstream analysis. The following bioinformatics steps were done using QIIME2 2018.11 (Bolyen et al. 2019). Denoising and dereplication of the reads was performed using the denoise and dereplicate single-end sequences (Dada2 denoise-single) with a chimera filtration done using the consensus method (Callahan et al. 2016). This program is classifying sequences as ASV (amplicon sequence variant) that distinguish sequence variation by a single nucleotide difference (Callahan et al. 2016). Phylogenetic tree construction was carried by creating a sequence alignment and removing phylogenetically

uninformative alignment using MAFFT. Taxonomic affiliation was done using the classify-consensus vsearch program and the TaxAss pipeline (Rohwer et al. 2018), which used both a general database, SILVA_132, and a freshwater ecosystem-specific database, FreshTrain (Newton et al. 2011; Quast et al. 2013; Rohwer et al. 2018). Before the taxonomic classification, the dataset was split into two groups using Blastn: sequences with a low percent identity to ecosystem-specific reference sequences, and sequences with high percent identity to ecosystem-specific reference sequences. The group containing the low percent identity sequences was affiliated using the general SILVA_132 database and the group with the high percent identity sequences was affiliated using the Freshtrain database. The two groups were then recombined and used for downstream analysis.

Biostatistics

All statistical analyses were performed with R software (R core team 2021) using the package Phyloseq (McMurdie and Holmes 2013), Vegan (Oksanen et al. 2015) and EdgeR (Robinson et al. 2010). Graphical display was done using ggplot2 (Wickham 2016).ASVs representing more than 9 reads in at least one replicate were kept for downstream analysis. Chloroplast affiliated ASVs were also removed from the dataset. After removal of the low quality, chimera and spurious sequences (false positive sequences) during the bioinformatic treatment and the filtration of the lowest abundant reads, the lowest number of reads in a sample was 18,696 for the first experiment and 17,938 reads for the second experiment. These minimal numbers of reads allowed to have confidence in recovering all the taxa richness present in our samples as indicated by the rarefaction curves (see supplementary file Figure SII.1). No rarefaction has been applied on the dataset (McMurdie and Holmes 2014), the data was normalized by relative abundance in percentage by dividing the number of reads affiliated to an ASV by the total read number in the sample and multiply by 100. Alpha diversity was measured with the richness using the Observed OTU, evenness with the Pielou index and both Shannon-Wiener and Simpson diversity index (Simpson 1949; Pielou 1966; Spellerberg and Fedor 2003). The comparison between samples using Beta diversity was done using Weighted Unifrac distance based on the phylogenetic tree build previously under QIIME2 (Lozupone et al. 2011) and visualized using Principal Coordinates Analysis (PCoA). For both experiments, Permutational multivariate analysis of variance (PermANOVA) with 999 permutations on the Unifrac

distance matrices was done to test the significance of the community composition difference between groups (Anderson 2017). The samples of the first experiment were grouped by volume of filtered water and the second experiment compared the difference two time independently: UD *versus* TF1/3 and UD *versus* VF1/3. The comparison of relative abundance of taxa between the different conditions was done with the package EdgeR (Robinson et al. 2010). The analysis was performed on the raw reads datasets normalized using the Relative Log Expression method (Anders and Huber 2010). P-value adjustment was done using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). The tested taxonomic ranks were ASV, family/lineage, order, class and phylum. Species and genus taxonomic ranks were not included because 16S rRNA gene fragment (V3 – V5) did not yielding enough taxonomic depth to be confident in these ranks. Only two conditions were comparable, so the two lower and the two larger filtration volumes were merged for the first experiment as the samples of these two groups clustered together on the PCoA (Figure II.3E). For the second experiment, like for the community composition differences test using PermANOVA, the condition VF1/3 and TF1/3 were each tested independents against UD. A False Discovery Rate (FDR) threshold of 0.05 was used to consider a result significant or not

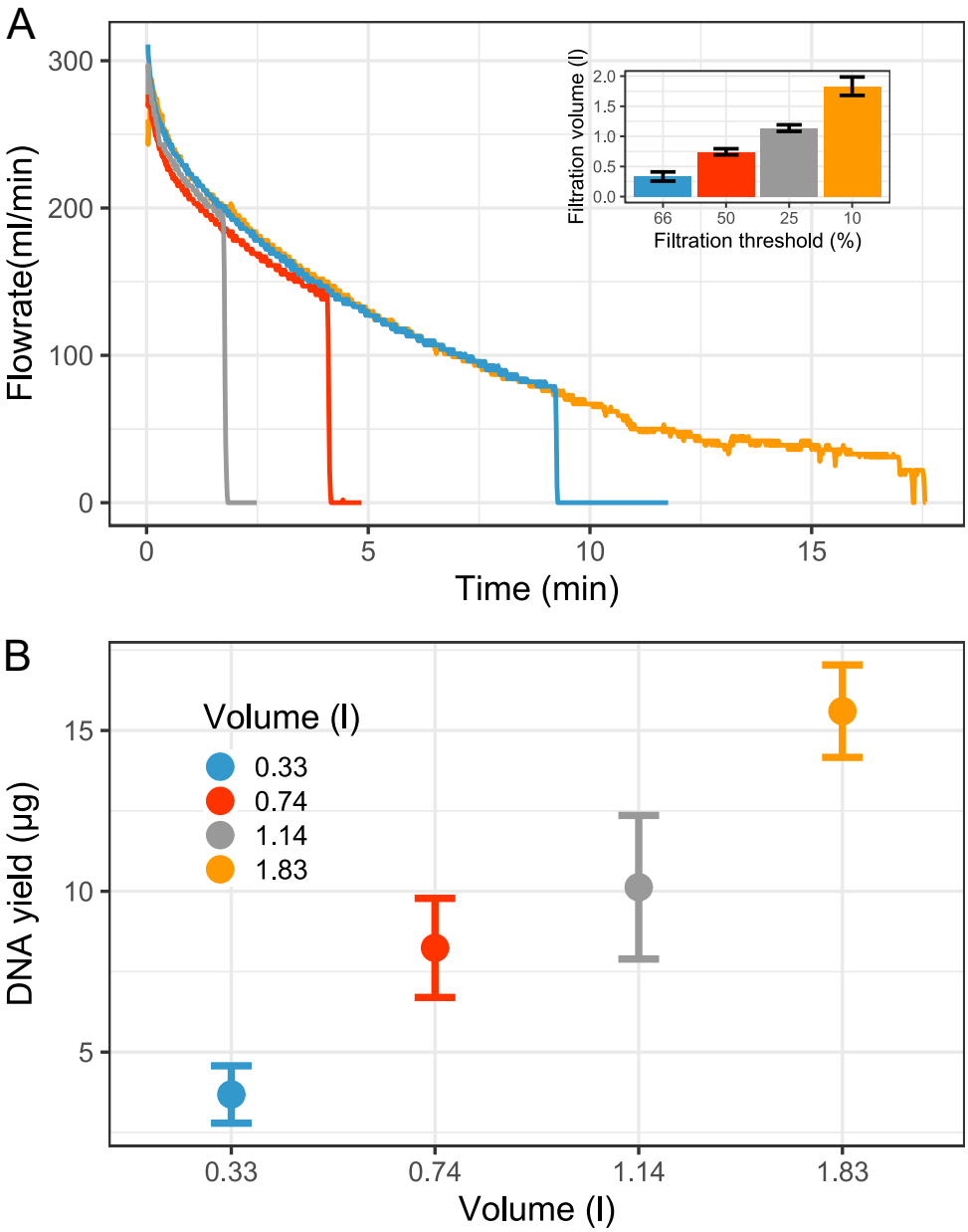

Figure II.2 Decrease of flowrate over time during plankton filtrations due to increased biomass loading of the filters, as recorded by the Arduino flowmeter (A), and (B) DNA yields after extraction of the corresponding picoplankton filters. (**A**) The filtrations were stopped at a threshold of 66% (blue), 50% (red), 25% (grey) and 10% (yellow) of the initial, maximal flow rate (F_m) at the start of the filtration. Each one replicate per triplicate threshold-filtration is shown. (**B**) Average DNA yield (n = 3) in 50 µl extract volume obtained from each set of filters; with standard deviation.

Results

Impact of different filtration volumes on the bacterioplankton community composition observed by 16S rRNA-gene fragment amplicon sequencing

For the first experiment with our flowmeter device (Figure II.1), we used a plankton sample taken from Lake Constance (08. November 2017; 0-20 m integrated sample, sampling site *Wallhausen*). We collected nano- and picoplankton samples by serial filtration through a 180-µm pre-filter onto 5-µm pore size (nanoplankton) and 0.2-µm pore size (picoplankton) polycarbonate membranes. The advantage of polycarbonate was that the material completely dissolved in the phenol-chloroform extraction step during DNA preparation (see Materials and Methods). The filtrations were stopped at four different flowrate-thresholds (F_t), as monitored with the flowmeter, at F_t = 66%, 50%, 25% and 10% of the initial, maximal flow rate (F_m, 100%). Each threshold filtration was done in triplicate (n = 3). The decrease of flowrate over time is shown in Figure II-2A. The total filtration volumes recorded after the filtrations were 0.33 ± 0.07 L, 0.74 ± 0.05 L, 1.14 ± 0.06 L and 1.83 ± 0.15 L, respectively, as illustrated as in Figure II.2A. The total DNA yields from the picoplankton filters were 3.7 ± 0.9 µg, 8.2 ± 1.5 µg, 10.12 ± 2.2 µg and 15.6 ± 1.4 µg, respectively (Figure II.2B). The linear regression (adjusted R-squared = 0.89) confirmed the relation between filtration volume and DNA yield. Notably, the obtained yields also suggested that smaller relative amounts of total DNA per liter of water sample were recovered from the filters when the filter had been loaded with higher amounts of biomass (e.g., 23% less total DNA per liter of water from the 1.83-L vs. the 0.33-L filtration).

We focused on a possible ‘filtration bias’ particularly for the single-cell bacterioplankton community composition, hence, on the picoplankton samples on the 0.2-µm filters (Padilla et al. 2015). Therefore, amplicon sequencing (Illumina technology) of a 16S rRNA gene fragment (V3-V5 hypervariable region) was applied to evaluated possible differences in the observed bacterioplankton community composition in dependence on the filtration volume. Note that the sequencing data obtained of one of the replicates of the 50% flowrate threshold had to be removed because it was an outlier.

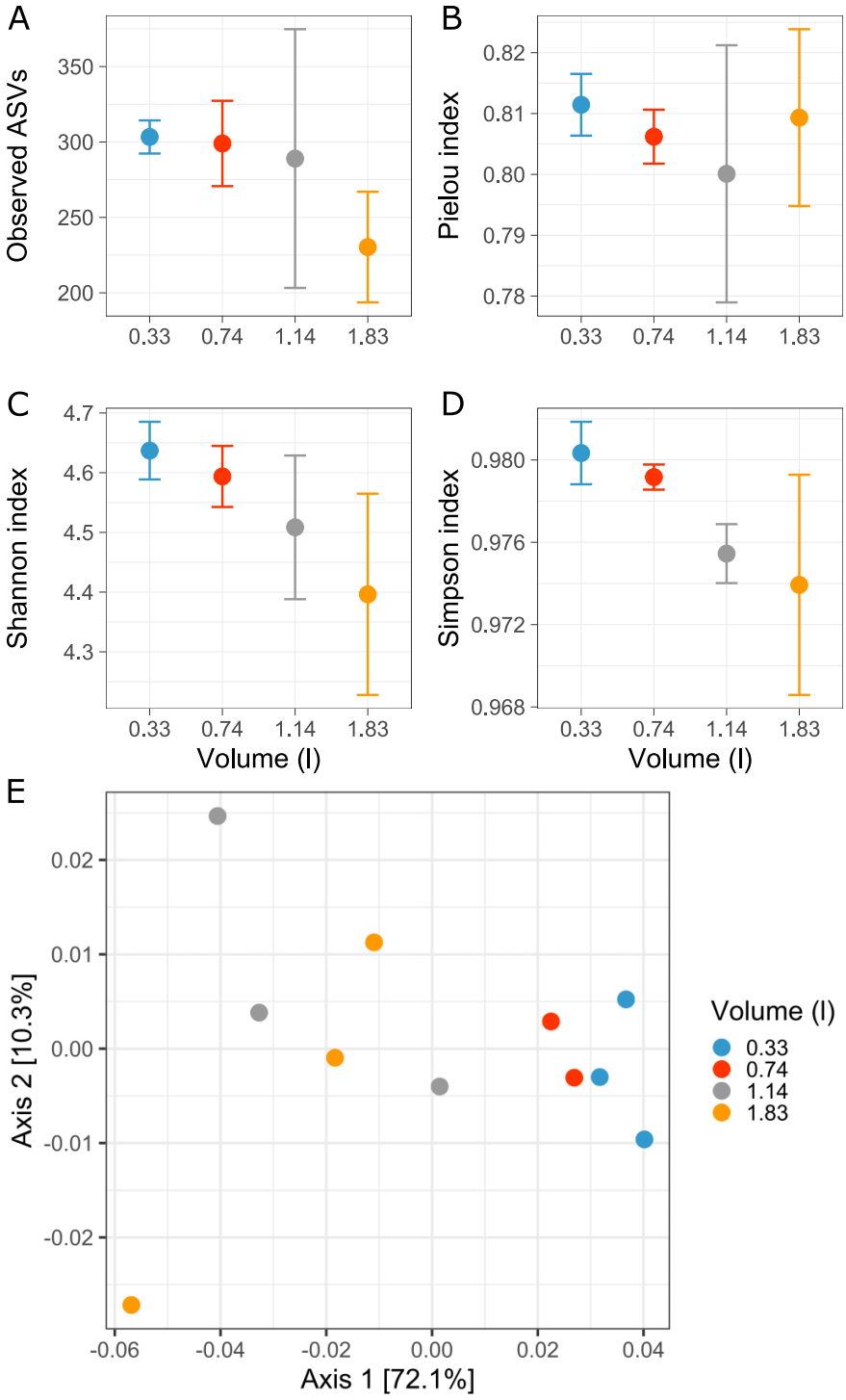

Figure II.3. Alpha and beta diversities of the bacterioplankton communities observed for the 0.2-µm filters in dependence on the filtration volume. For the picoplankton filters (0.2-µm pore size) loaded with variable flowrate thresholds (Figure II.2A), the bacterioplankton community composition was analysed by 16S rRNA-gene fragment amplicon sequencing. Alpha diversity is represented by richness of the observed ASVs (**A**) and by Pielou´s community evenness (**B**), Shannon-Wiener community diversity (**C**) and Simpson community diversity (**D**) indices. Dots represent the average values and the error bar represent the standard deviation of the replicates for each condition (n = 3). (**E**) Principal Coordinate Analysis (PCoA) of the Weighted Unifrac metric calculated from the bacterioplankton community composition

Alpha diversity

960 bacterial ASVs were present in the processed dataset after removal of ASVs represented by less than nine reads in at least one replicate. The community diversity was analysed within each sample (alpha diversity), using the Shannon-Wiener and Simpson indexes. In addition, richness was expressed as the number of observed ASVs and the evenness using the Pielou´s index. Interestingly, increased filtration volume coincided with decreased Shannon and Simpson diversity indices (Figure II.3CD). Richness was also decreasing with increased filtration volume, with an average ASV number of 303 for the lowest volume (66% F_m) and 230 ASVs for the highest volume (10% F_m) (Figure II.3A). Thus, in average 24% of the bacterial taxa detected for the lowest filtration volume were not detectable for the largest filtration volume. The evenness was relatively stable, apart from an increase of variability for increased filtration volumes (Figure II.3B). The decrease of diversity observed with the Shannon and Simpson index can be explained by the decreasing richness. This decrease of richness was due to a loss of low-abundant taxa (see explanation below) with increasing filtration volume.

Beta diversity

The community structure was analysed in respect to diversity between samples (beta diversity) and statistically tested for the null hypothesis (H_0), that the filtration volume had no influence on the observable community composition, and the alternative hypothesis (H_a), that the filtration volume had an effect on the bacterial community composition. The analyses were done using the Weighted Unifrac distance metric, visualized by Principal Coordinate analysis (PCoA), and the statistical significance was tested using a PermANOVA. The PCoA plot (Figure II.3E) illustrates a clear shift of the community compositions in dependence on filtration volume. Each replicate of a filtration threshold clustered together, while the higher-volume conditions showed a higher variability between replicates. The first axis of the PCoA represented 72.1% of the overall variability and the second axis 10.3% (Figure II.3E and Figure SII.2A), which suggested that one variable was the major driver of the observed shift of the community composition, i.e., most likely the variable parameter of our experiment, the filtration volume (biomass loading). This difference of variance may also reflect that an equivalent observable difference between replicates indicate a stronger difference of diversity on the x-axis than the y-axis. PermANOVA results

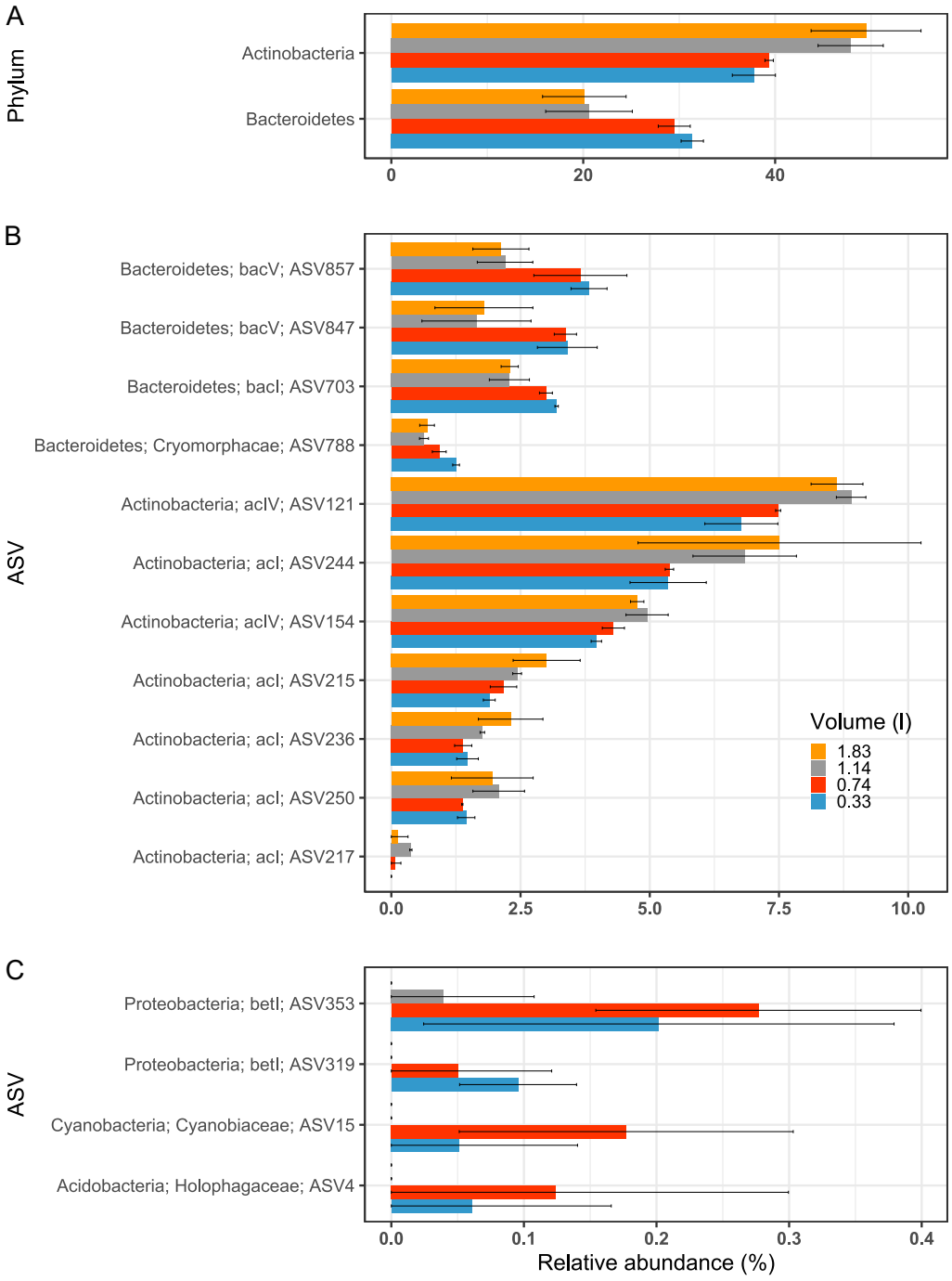

Figure II.4. Representative taxa that displayed significant changes in relative abundances in dependence on the filtration volumes used. Shown are (**A**) phyla-taxonomic rank, (**B**) and (**C**) as ASV with family and phyla affiliation. The relative abundance is expressed as percent (%) reads for a specific taxon relative to all reads observed. Error bars represent the standard deviation (n = 3). Note the different x scales.

showed a p-value of 0.002 and R^2 of 0.57. The R^2 indicated that the filtration volume explained 57% of the community composition variability which, coupled with the p-value below 0.05, allowed to reject H_0. Under these conditions, we can conclude that the filtration volume indeed had an impact on the observed microbial community composition.

Individual taxonomic groups affected by filtration volume

Next, we examined which taxonomic groups were specifically affected by the filtration volume, using EdgeR to compare the relative abundance changes between conditions. Only two conditions can be compared so the data of the two lowest (0.33 and 0.74 L) and the two highest (1.14 and 1.83 L) filtration volumes were grouped together as their replicates clustered closely on the PCoA, indicating similar community composition (Figure II.3E). All taxonomic ranks were analysed with the exception of genus and species which were rarely assigned in our dataset. The results showed that the community composition was affected at all taxonomic levels by the filtration volume, in that two phyla, six classes, eight orders, 14 families and 15 ASVs showed a significant difference in relative abundance, as illustrated in Figure II.4AB and Figure SII.3A-C. The number of affected taxa is minor compared to the total dataset, as they represented 11.8%, 17.1%, 9.8%, 9.0% and 1.6% of the total phylum, class, order, family/lineage and ASV, but represented a majority of the relative abundance as these few taxa represented between 34.0% (ASV) to 78.8% (class) of the total abundance.

Most changed in relative abundance in dependence on filtration volume, were taxa affiliated to Actinobacteria and Bacteroidetes, together representing 69% of all affected taxa. The differences followed opposing trends, with Actinobacteria increasing with increased filtration volume (up to 38% increase at the phylum rank) and Bacteroidetes decreasing with increased filtration volume (up to 29% decrease for the phylum rank) (Figure II.4A). The percentage of relative abundance change varied from a maximal increase of 67% for the family/lineage Sporichthyaceae (Figure SII.3A) to a maximal decrease of 50% for ASV847 affiliated to the lineage bacV (Figure II.4A, Table SII.1). Four others ASVs not affiliated with either Actinobacteria or Bacteroidetes were only detected for the lowest filtration volumes (0.33 and 0.74 L) (Figure II.4C). These four ASVs were detected by EdgeR because their initial relative abundance was high enough in the low filtration volumes, yet we can expect that more

ASVs may display the same pattern. This decrease of total number of ASV detected in the highest volumes could explain why the richness was decreasing with increased filtration volume (cf. Figure II.3A).

Normalization of the plankton filtration bias using our flowmeter device

With the next experiment, we tested whether our flowmeter device may be helpful to normalize the filtration bias described above. We tested if the threshold filtration could uniform the biomass loading of the filters for the bacterioplankton community, as represented on the 0.2-µm filters. One of the replicate of condition VF1/3 (see below) had to be removed because it was an outlier, leaving three replicate for this condition. A Lake Constance sample (14. May 2019; 0-20 m depth, *Wallhausen*) was filtered at its given plankton density (termed 'undiluted condition', UD) and after the sample had been diluted three-fold (1:3), each with threshold filtrations (TF) using the flowmeter device and stopping the filtrations again at Ft = 50% Fm. The total filtration volumes recorded for the undiluted sample were 0.57 ± 0.06 L (n = 8) and for the 1:3-diluted sample (termed TF1/3) 1.71 ± 0.04 L (n = 4), hence, the 3-fold volume. As a third filtration condition (termed 'fixed-volume filtration' VF1/3), we filtered the three-fold diluted sample with the same volume as the undiluted sample (0.57 L), hence without the flowmeter device and as control of the filtration bias (n = 3). The DNA yields obtained for the UD and TF1/3 conditions were relatively similar with 3.5 ± 0.4 µg and 3.9 ± 0.5 µg, while the VF1/3 condition yield was about 3-fold lower with 1.2 ± 0.2 µg, as expected (Figure II.5).

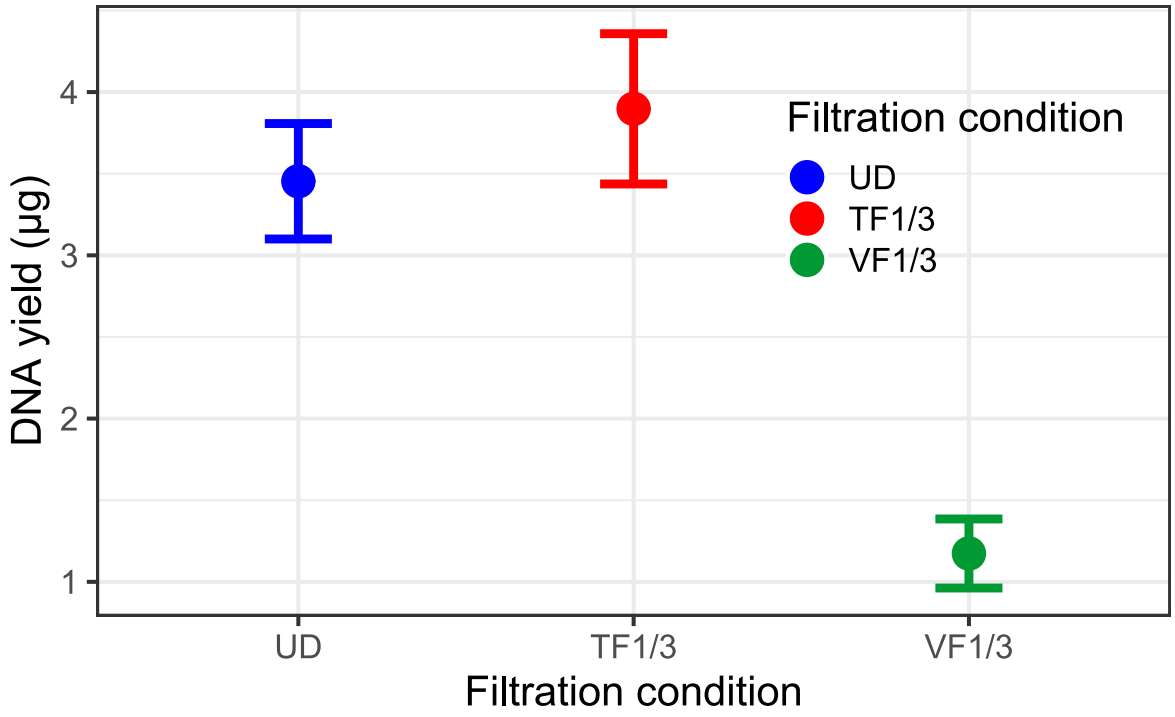

Figure II.5 DNA yield after extraction of the picoplankton filters. (**A**) UD corresponds to the undiluted water condition, TF1/3 to the filtration of the 3-fold diluted water sample using threshold filtration with the flowmeter device, and VF1/3 to the filtration of the 3-fold diluted water using the same filtration volume as for the undiluted condition.

Alpha and beta diversities

The average number of ASVs in the processed data was 257 ± 27, 271 ± 26 and 278 ± 15 for the conditions UD, TF1/3 and VF1/3 respectively, hence the richness in the diluted samples seemed to be slightly higher, although all three standard deviations overlap (Figure II.6A). The same trend was visible for the evenness (Pielou), for which the standard deviation (SD) for VF1/3 was much smaller (Figure II.6B). The Shannon-Wiener and Simpson values of the UD condition were (cf. SDs) smaller than the VF1/3 condition, while the TF1/3 diversity values were in between the two conditions (SDs overlapping with both UD and VF1/3 conditions) (Figure II.6CD). For the beta diversity, the first axis of the PCoA represented 69.3% of the variability and the second axis 8.4% (Figure II.6E and Figure SII.2B). Hence, one factor was the major driver of the observed variability and like in the first experiment. An equivalent difference between replicates indicates a stronger difference of diversity on the x-axis than the y-axis. Replicates of each condition clustered together, with the VF1/3 filtrations more clearly separated from the undiluted filtrations, and the TF1/3 filtrations in between (Figure II.6E).

The goal of this experiment was to examine whether the filtration bias may be normalized when filtrations are done using the flowrate device (undiluted *vs.* TF1/3) in comparison to fixed-volume filtration (undiluted *vs.* VF1/3). Hence, the statistical hypotheses for the analysis of the fixed-volume condition were, (H_0) plankton density does not affect the observed diversity (centroid of the compared group are equivalent), and (H_a) plankton density does affect the observed diversity (centroid of the compared group are not equivalent). The PermANOVA output was a p-value of 0.027 and a R^2 of 0.66, thus, we can reject H_0 and conclude that the plankton density (filter loading) impacted the diversity, when using fixed-volume filtration. For the threshold filtration, which may have no impact (H_0) or may have an impact (H_a) on the diversity in the diluted sample, the p-value was 0.114 and the R^2 0.20, and therefore, we cannot reject H_0. Hence, at the community level, a difference in plankton density did not statistically affect the observed community when using our flowrate device, suggesting that threshold filtrations may compensate for the filtration bias.

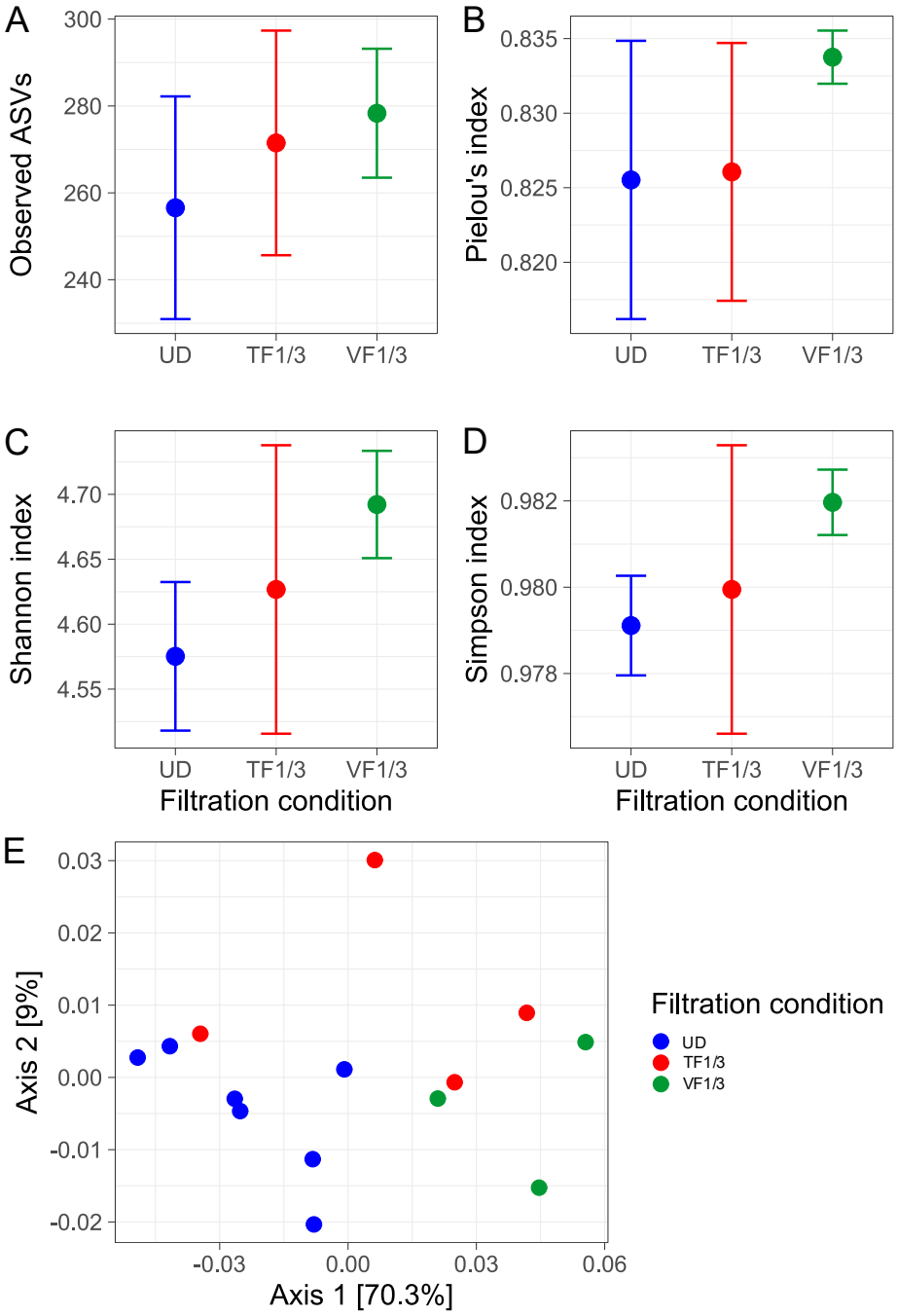

Figure II.6. Alpha and beta diversities of undiluted and diluted bacterioplankton collected on 0.2-µm filters by threshold filtration in comparison to fixed-volume filtration. A Lake Constance water sample was filtered undiluted (UD) and 3-fold diluted (threshold filtration, TF1/3) using each the flowmeter device at 50% flow-rate threshold. For control, the 3-fold diluted sample was also filtered with the same volume as the undiluted sample (fixed-volume filtration, VF1/3). Alpha diversity is represented by (**A**) richness represented by the observed ASVs, and (**B**) community evenness using Pielou´s, (**C**) Shannon-Wiener community diversity and (**D**) Simpson community diversity indices. Dots represent the average values and the error bar represent the standard deviation of the replicates for each condition. (**E**) Principal Coordinate Analysis (PCoA) of the Weighted Unifrac metric calculated from the bacterioplankton community composition.

Individual taxonomic groups affected.

EdgeR was used to compare the relative abundance change of taxa between the conditions (genus and species excluded). For the undiluted *vs*. VF1/3 condition and the confirmation of the filtration bias, all taxonomic ranks except order showed taxa with abundance significantly changed, i.e., three phyla, four classes, four families/lineages and two ASVs (Figure II.7 and Figure SII.4). Like in the first experiment, the number of affected taxa is minor compared to the total number of taxa as they represented 18.8%, 12.9%, 2.5% and 0.4% of the total taxa in their respective taxonomic rank. The relative abundance of the impacted taxa is less dominating, representing between 2.4% (ASV) to 51.6% (class) of the total relative abundance. As with the first experiment, most taxa belonged to the phyla Actinobacteria and Bacteroidetes and these two phyla showed a decrease of 27% and an increase of 30% of relative abundance respectively, when the plankton had been diluted and filtered by fixed-volume filtration (Figure II.7B). Decrease in relative abundance in relation to plankton dilution was up to 34% for the lineage acI (Figure SII.4 and Table SII.2). In addition, significant increase of also phylum Planktomycetes (28%) and its belonging class Phycisphaerae (40%) (Figure II.7A) was detected for the VF1/3 condition in this experiment.

The undiluted *vs*. TF1/3 condition showed no taxa, whichever taxonomic rank, that were changed at statistically significant level. This is illustrated in Figure II.7 and Figure SII.4, where the standard deviations (SDs) for the TF1/3 condition are always overlapping with these of the undiluted condition, hence, showing a difference not strong enough to be significant, while the SDs for the VF1/3 condition illustrate significant differences (filtration bias).

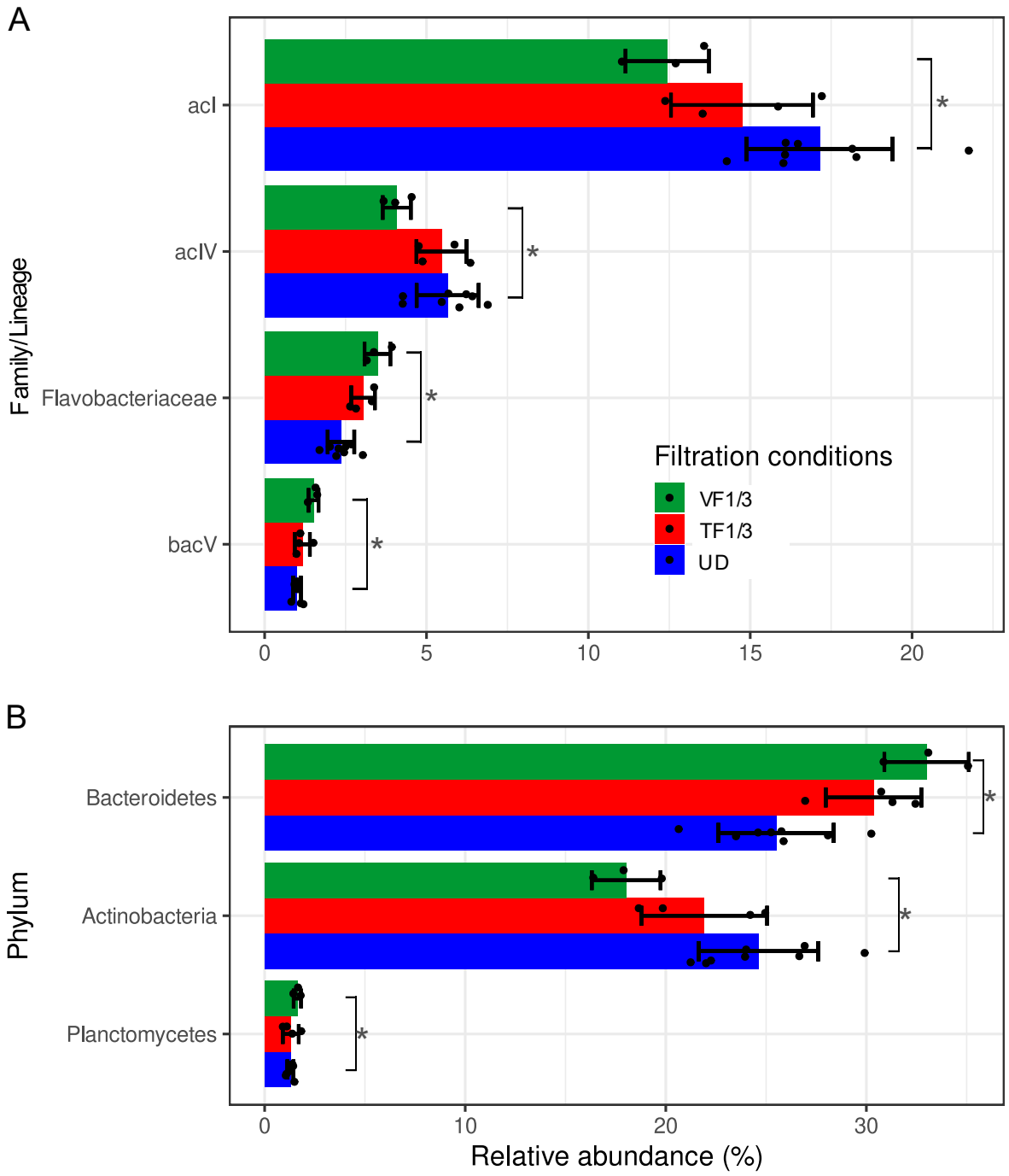

Figure II.7. Specific taxa that displayed different relative abundances in dependence on the filtration method used. Lake Constance water was filtered undiluted and 1:3 diluted (threshold filtration, TF1/3) using the flowmeter device, and the 1:3-diluted sample was also filtered with the same volume as the undiluted sample (fixed-volume filtration, VF1/3). Shown are examples on the (A) family and (B) phyla level. Error bars represent the standard deviation (n = 3). Statistical significance is flagged with one star (*). Note the variation of y axis scales.

Discussion

The first experiment using the flowmeter device confirmed and expanded on the description of a pico-bacterioplankton filtration bias as reported previously (Padilla et al. 2015; Torres-Beltrán et al. 2019). Less bacterial taxa could be detected for the largest sample volume (Figure II.3A), which may appear counterintuitive, as more water was filtered, so more bacterioplankton biomass was collected, which should lead to a better detection of also the lowest-abundant taxa.

As described in the Introduction, we considered whether the cause of such a volume-dependent filtration bias may be a build-up of a filtration cake during serial filtration, which is increasingly acting as an additional, smaller-pore size filter, thereby increasingly retaining microorganisms on this filter that would normally belong into the filtrate and onto the next smaller pore-size filter. Through the implied biomass dependence, samples with different plankton densities may produce a similar bias if each the same volume of water is filtered, as confirmed with the second experiment (Figure II.6; UD *vs*. VF1/3 conditions). This effect is particularly relevant for plankton diversity studies in lakes of temperate regions, where the plankton density can change drastically in between winter mixing and the vegetative season, for example for deep Lake Constance, which is the subject of our own plankton diversity studies. Lake Constance currently displays Secchi depths in winter of down to 21 m (February 2022, *Überlinger* See) and up to 2.5 m during spring bloom and summer.

Members of the phyla Actinobacteria and Bacteroidetes were among the taxa most affected by increased filtration volume. Typical freshwater Bacteroidetes decreased in relative abundance for the 0.2-µm filters (Figure II.4), as well as similarly affected family Pedosphaeracae, and phyla Verrucomicrobia, Holophagaceae and Acidobacteria (Figure S II.3), which are known to be filamentous or particle-associated bacteria (Allgaier and Grossart 2006; Eloe et al. 2011). They may increasingly be captured by the 5-µm filter and therefore be detected at lower relative abundance on the 0.2-µm filters with increasing filtration volume. *Vice versa*, a major part of the typical freshwater Actinobacteria are contributed by very small, ‘ultramicrobacteria’ (Hahn et al. 2003; Newton et al. 2011; Kim et al. 2019b), and an increased biomass cakes on the filters may have retained these cells more efficiently on the 0.2-µm filter with higher volumes.

The 16S rDNA diversity analyses by amplicon sequencing as performed in this study, can also be subject of a ‘DNA template dilution bias’ for the PCR (Wu et al. 2010). In this study it was shown that the number of taxa detected was decreasing with higher template dilution (i.e., decreased probability for low abundant taxa-DNA to be amplified), but the relative abundance of detected ASVs was not affected, therefore the global community structure was not changed when comparing the different dilution condition (Wu et al. 2010). In our study, the different DNA yields recovered from the different filtration volumes (Figure II.2B) required higher dilution of the DNA for the highest filtration volume and, thus, this additional bias may also be a factor that contributed to the observed decrease of richness through loss of some of the lowest-abundant taxa. The lowest filtration volume showed 93 ASVs more than the highest, and these bacterial taxa belong to the low abundant bacteria with an average relative abundance of 0.029%. In the first experiment, four ASVs may illustrate this ‘DNA template dilution bias’, as they were detectable for the low filtration volume and not anymore for the high filtration volume (Figure II.4C). Although the design of our experiments does not allow to disentangle also the ‘DNA template dilution bias’, it should not have affected the results and conclusion of our analyses. Indeed, the ‘DNA template dilution bias’ affect the capability of detecting ASVs or not (Wu et al. 2010), while the filtration bias as described in this study altered the ASVs relative abundance. Hence, DNA template dilution bias and filtration bias likely affect each a different component of the community analysis output.

In the second experiment, no statistically significant difference was observed between the undiluted and flowrate-threshold filtration condition (Figure II.6). The 50%-threshold filtration allowed for the collection of equivalent amounts of DNA regardless of the three-fold dilution of the original water sample (Figure II.5), and thus, a different dilution of the DNA for PCR amplification was not necessary, which likely eliminated also a DNA template dilution bias at least between the UD and TF1/3 samples (Wu et al. 2010). However, the number of taxa for the VF1/3 condition was slightly higher (Figure II.6A), and this increase may indeed be due to DNA template dilution bias, as the UD and TF1/3 DNA samples had to be diluted three-fold relative to the VF1/3 DNA sample for PCR. Hence it seems that the compensation of the filtration bias by collecting equivalent amount of biomass, resulting in similar dilution for the PCR, may also compensate for the DNA template dilution bias.

Hence, the two experiments confirmed a volume/biomass-dependent bacterioplankton filtration bias for community analysis and showed that threshold filtrations can improve the recovery of equivalent amounts of total DNA from samples with different plankton densities, thereby reducing significant biases for community analysis.

We considered whether this method of threshold filtration may be useful for sampling across depth profiles and along seasonal cycles of temperate lakes, and we therefore tested our filtration setup (Figure II.1) (at 50% threshold) in a yearly sampling campaign for Lake Constance, when collecting every two weeks an integrated 0 – 20 m depth sample as a representation of the photic zones in two different lake parts, Upper Lake (*Überlinger See*) and Lower Lake Constance (*Zeller See*). In Figure II.8, the Secchi depth as a representation of plankton density is shown, and the recorded filtration volumes until the threshold was reached, and the DNA yields obtained from extraction of the 5-µm and 0.2-µm pore size filters are shown. Increased Secchi depths correlated well with increased filtration volumes in order to reach the 50%-threshold. For example, a sample taken at >16 m Secchi depth needed almost 2 L water, and at 5 m Secchi depth around 0.4 – 0.7 L, while similar amounts of DNA were recovered from the filters at both high and low Secchi depths, particularly throughout the seasonal cycle of the *Zeller See* (Figure II.8CD). For *Überlinger See*, the DNA yield from threshold filtration tended to be highest at low Secchi depth in winter and during the spring bloom (Figure II.7AB), while in summer and beginning of fall, the DNA yields were unexpectedly low. We believe this may have resulted from calcite precipitation, typical for Upper Lake Constance in summer (Stabel 1986; Küchler-Krischun and Kleiner 1990; Pulvermüller et al. 1995), in that calcite particles most likely have led to a faster clogging of the filters and thus to a lower ratio of cellular biomass collected until the 50%-threshold was reached, and to lower amounts of DNA extracted. Hence, it appeared that the effect of threshold filtration for normalizing DNA yields may itself be distorted by calcite (or other inorganic particles) contributing to the clogging of the filters.

The filtration bias can be expected to not only affect DNA-based community analyses. Transcriptomic and proteomic analyses also require collection of plankton biomass by filtration, making these methods pipelines also susceptible. This may particularly be relevant for multi-omics analyses, in which the different plankton

samples for each pipeline are collected by different filtration methods and/or volumes. For example, if a study would examine the plankton community composition and its activity through comparing genomic and transcriptomic data, each generated from biomass collected by each a different filtration method, this could lead to potential over- or underestimation of taxa representation and/or their activity. The relevance of the filtration bias can be expanded also to studies of different lakes or mesocosms experiments in which plankton densities vary strongly. Hence, we expect that using flowrate-threshold filtration to normalize the biomass loading on each filter can increase the accuracy of also such analyses.

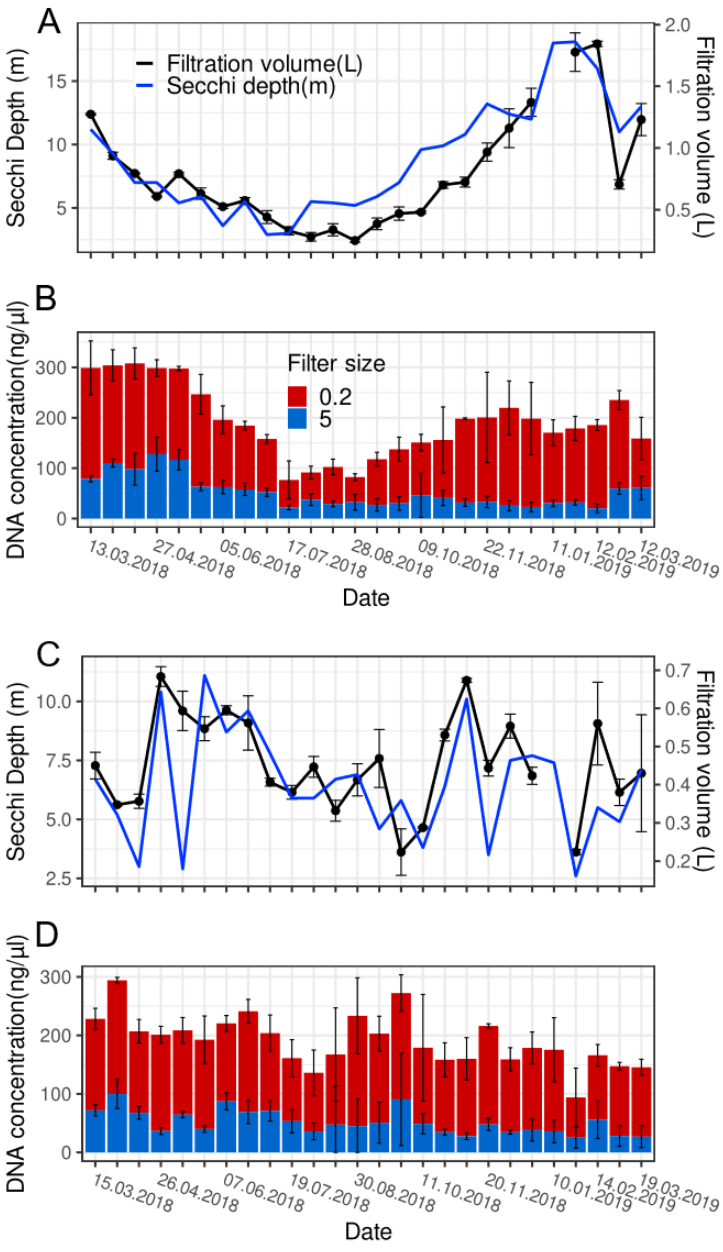

Figure II.8. Secchi depths, filtration volumes recorded for threshold filtrations, and DNA yields obtained from a seasonal sampling of Upper and Lower Lake Constance plankton from March 2018 – March 2019. The sampling was done at Upper Lake Constance, *Überlinger See* (A, B), and Lower Lake Constance, *Zeller See* (C, D). Integrated samples from 0-20 m water depth were collected and filtered using the flowmeter device set to a threshold of 50% Fm. The filtration volumes and DNA yields are shown as average of triplicates per sampling date; error bars represent standard deviation. Note that the DNA concentrations from extracting both size-class filters, the 5.0-µm (red) and 0.2-µm (blue) pore-size filters, are shown.

Funding: This research was funded by DFG Research Training Group R3 - Responses to biotic and abiotic Changes, Resilience and Reversibility of Lake Ecosystems (GRK 2272).

Data Availability Statement: The Illumina sequence data presented in this study will be available online upon publication.

Acknowledgments: We would like to highlight the excellent support received from Alfred Sulger, Julia Schmidt, Angelika Seifried, Pia Mahler and Josef Halder at the Limnological Institute of University of Konstanz for managing the ship cruises. Further, we would like to thank all members of the RTG-R3, especially Tina Romer, and members of the AG Schleheck for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Chapter III
Nano- and pico-plankton succession in Upper and Lower Lake Constance followed by 18S and 16S rDNA amplicon sequencing: Same seasonal dynamics but different communities

Corentin Fournier[1*] and David Schleheck[1*]

[1] Microbial Ecology and Limnic Microbiology, Limnological Institute, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

*Shared corresponding authorship:
corentin.fournier@uni-konstanz.de, david.schleheck@uni-konstanz.de

Manuscript in preparation for submission

Keywords: NGS, Lake Constance, microbial plankton, seasonal dynamics, geographical difference, community composition

Abstract

Lake Constance is a pre-Alpine, warm-monomictic, oligotrophic lake situated at the southern end of Germany composed of two main water bodies, deep Upper Lake Constance (ULC) and the shallow, less oligotrophic Lower Lake Constance (LLC). This lake has been studied in respect to its seasonal plankton succession and was a test system for the development of the Plankton Ecology Group (PEG). Until today, no DNA-sample repository and sequencing-based study exist of the seasonal succession of the nano- and picoplankton in Lake Constance. Hence, as part of this study, a DNA-sampling procedure has been set up and added to the routine sampling campaign of the Limnological Institute of University of Konstanz, for future study of the nano- and picoplankton composition and its seasonal succession using Next-Generation Sequencing. Every two weeks, samples from the top-20 m of the water column (approx. photic zone) were filtered into two size classes, 180 – 5 µm diameter (nanoplankton, NP) and 5 – 0.2 µm diameter (picoplankton, PP). With the present communication, the differences in composition and dynamics between ULC and LLC are described for March 2018 to March 2019, based on rDNA amplicon sequencing: the NP samples were analysed by 18S rDNA (18S-NP dataset), and the PP samples by 18S (18S-PP dataset) and 16S rDNA (16S-PP dataset) sequencing. While the general pattern of temporal diversity dynamic was similar for all three datasets, marked differences in composition could be observed. In average, 30% of the ASVs were unique to ULC or LLC and ASVs with different relative abundances between ULC and LLC could be detected at every seasons. Relative abundances of the unique as well as differentially abundant ASVs showed very specific patterns linked to the ecological conditions of ULC and LLC. For LLC, the development of hypoxic deep water in late summer/fall, allowed for the development of hypoxic/anoxic adapted taxa. The vertical mixing of winter allowed the development of taxa adapted to the deep oxygenated hypolimnion in ULC. This study and the ongoing DNA-repository being established, are important references for a future detection of changes of the Lake Constance nano- and picoplankton community given the ongoing effects of climate change.

Introduction

Lake Constance is a pre-Alpine, warm-monomictic, oligotrophic lake situated at the southern end of Germany composed of two main water bodies, termed Upper Lake Constance (ULC) and Lower Lake Constance (LLC), which are connected via the Lake Rhine. It is acting as drinking water reservoir for more than 5 million people (Petri 2006) and is a biodiversity reservoir for numerous species such as birds and fish, but also of a highly diverse and dynamic microbial plankton community. The lake underwent eutrophication in the 1970s-1980s followed by a re-oligotrophication since the 1990s (Straile and Geller 1998; Limnologischer Zustand des Bodensees). Lake Constance has been, and is being extensively studied using environmental long term and experimental data (Sommer et al. 1993; Straile and Geller 1998; Peeters et al. 2007; Kunzmann et al. 2019; Sabel et al. 2020; Ogorelec et al. 2022). It also served as one of the lakes for the development of the Plankton Ecology Group (PEG) model. The PEG model, created by the International Society of Limnology (SIL) Plankton Ecology working group, is a verbal model used as standard template to describe patterns and driving factors of seasonal phytoplankton and zooplankton succession in oligotrophic and eutrophic lakes (Sommer et al. 1986; Sommer et al. 2012).

The first recognition of the ecological importance of the phytoplankton as contingent of the microbial plankton, was made by J.D Hooker in 1847 from observation during an Antarctic cruise while the importance of the bacterioplankton in freshwater environment was recognized in 1940s by Lindeman (Lindeman 1991). Phototrophic microbial plankton produces biomass and dioxygen from carbon dioxide maintaining the aquatic environment stable and capable of hosting organisms relying on oxygen for respiration, and organic matter used as nutrient sources by heterotrophic bacteria (Vollenweider et al. 1974; Ducklow et al. 1986; Jones 1998; Lewis et al. 2020). Biogeochemical cycling of phosphorus, sulfur and nitrogen also rely heavily on microorganisms (Pierrou 1976; Jetten 2008; Wu et al. 2019). Through the heterotrophic microbial plankton such as bacteria, nutrients are recycled back into the trophic network and made available for higher trophic organisms (Cole et al. 1988), In addition, bacteria are being consumed by protozoa, and finally that the protozoa could then enter the food chain formed by larger creatures, as part of the so-called microbial loop (Azam et al. 1983; Fenchel 2008). Phytoplankton and bacterioplankton are also the prey of grazing and parasitic microorganisms, such as Ciliophora and parasitic

fungi, the last representing an alternative to the microbial loop called the mycoloop (Anderson 1988; Kagami et al. 2014; Agha et al. 2016). Ecosystem-wise, a high microbial plankton diversity is an indicator of a stable and resilient environment. The functional redundancy within a highly diverse community can act as buffer to undergo a disturbance while retaining its major functions (Eisenhauer et al. 2012; Li et al. 2012).

The species identification and composition of the phytoplankton community is traditionally being studied using microscopy counting techniques with a taxonomic affiliation based on morphological attributes (Sommer 1986; Gaedeke and Sommer 1986; Rodriguez et al. 2002; Chalar 2009; Su et al. 2016). Microscopic identification and counting however, is not capable of revealing the full phytoplankton community diversity (Xiao et al. 2014), as the taxonomic affiliation can be conflicted due to morphological variability (Naselli-Flores 2014) and this method is highly dependent on the experience of the individual person that is counting. But microscopic species identification cannot be applied to study the composition of the heterotrophic pico/bacterioplankton. Fingerprinting techniques such as denaturing gradient gel electrophoresis have allowed for a first insight into the species composition of the bacterioplankton community of Lake Constance (Zwisler et al. 2003; Hewson and Fuhrman 2004; Schwalbach et al. 2005; Kan et al. 2006).

The development of Next-Generation Sequencing (NGS) opened a new door to the study of the smallest size classes of the plankton. Amplicon sequencing targeting the phylogenetic marker 16S rDNA allows for a much higher resolution in the determination of the bacterioplankton diversity as well as a more accurate taxonomy (Newton et al. 2007; Salmaso et al. 2018; Cabello-Yeves et al. 2020), while 18S rDNA amplicon sequencing allows for a deeper view into the micro-eukaryote community (Debroas et al. 2017; Salmaso et al. 2020; Wu et al. 2020), besides that also other phylogenetic marker genes can be applied. Further, the democratization of these techniques allowed for a much more intensive study of freshwater communities leading to an accumulation of information precious to unravel how freshwater environments are functioning and impacted by disturbances. We are in an era of biodiversity loss and climate change that impact biological events like the trophic network, modifying the nutrient transport between organisms and altering the overall ecosystem dynamics (Gronchi et al. 2022). It is therefore becoming urgent to create

plankton DNA-sample repositories (DNA banks) for Lake Constance, serving as today's reference for comparison to the changes coming in future.

We started routine sampling of the total DNA of the nano- and picoplankton community of Lake Constance, every two weeks, and established a DNA bank since 2018 of these two plankton size classes. Integrated water sampling from 0 – 20 m was chosen as approximate representation of the photic zone (epilimnion) that represents the highest primary production. The results presented in this communication are focusing on the differences in the temporal composition and dynamics of the microbial plankton communities in two water bodies that represent marked differences in environmental conditions at Lake Constance, i.e., deep (up to 252 m depth), fully-yearly aerobic, oligotrophic ULC (*Wallhausen*-sampling site; *Überlinger See*) *versus* a shallow and less oligotrophic LLC part (*Zeller See*, *Radolfzell*-site) (20 m depth) with an oxycline and anaerobic zone at the end of the vegetative season.

Materials and Methods

Study sites, sampling campaign and filtration procedure

Lake Constance is a deep (maximal depth of 252 m) oligotrophic pre-Alpine lake composed of three water bodies called *Obersee* (Upper Lake Constance or ULC), *Untersee* (Lower Lake Constance or LLC) and *Seerhein* (Lake Rhine). Samples were taken at the routine sampling site *Wallhausen* in the *Überlinger See* (47.7571°N 9.1273°E), a north-western arm of ULC with a maximal depth of 140 m, and in *Zeller See* (47.716557 °N 8.991767 °E), a shallow north-western part of LLC with a maximal depth of about 20 m. We sampled the approximate photic zone (epilimnion), which represent the 0 - 20 m depth surface layer. Integrated samples were collected on a boat using an integrating water sampler (model IWS II, Hydro-Bios, Germany). The sampling campaign represented in this study started in March 2018 and ended in March 2019. Samples were taken every two weeks and are stored in stainless steel barrels (19 L soda-keggs). Table SIII.1 summarizes the sampling dates and their grouping by seasons. Four biological replicates were collected each time (n = 4).

Filtrations were done directly on the boat, in order to reduce alteration of the samples. A pressure of 2 bars was applied in the barrels using a pressurized air tank. The barrels were connected to a series of filter holder (Swinnex 47 mm filter holder; Millipore) using PVC tubing. The pore-sizes of the filters were 180 µm (hydrophilic

nylon net, 47 mm diameter; Millipore) to remove zooplankton and larger particles, and 5.0 µm to collect the nanoplankton and 0.2 µm to collect the picoplankton, each using polycarbonate membrane filters (Isopore, 47 mm diameter; Millipore). The Swinnex filter holders with the filters assembled, were autoclaved when being wrapped in aluminum foil and the transported onto the boat. Filtration was done following the protocol developed in our laboratory (Chapter II), when using the flowrate at the outlet of the filters as a proxy of the biomass loading of the filters and the degree of filter clogging. This protocol was developed in response to a study (Padilla et al. 2015) showing that there is a bias introduced by the filtration volume and the microbial plankton density on analysis of microbial plankton by -omics techniques (Chapter II). Each filter was transferred by using sterile forceps and stored when being rolled into 15 ml Falcon tubes, with the filter side harboring the collected biomass facing to the inside of the tube. The tubes were stored on dry ice on the boat and in the laboratory at -20°C until the DNA was extracted. Only the 5.0 and 0.2 µm filters were used for the downstream analysis.

Molecular work: DNA extraction, library preparation and sequencing

The DNA extraction protocol used was adapted from Rusch et al. 2007 (Rusch et al. 2007) and from the JGI protocol (version 3 by William S., Helene Feil and A. Copeland). It started with thawing of the filters by adding 3 ml of lysis buffer, composed of EDTA, EGTA and Tris-HLC [pH8.0] each at a final concentration of 50 mM, then 1 min of ultrasonication (Sonorex super RK 510, Bandelin, Germany) was applied, followed by adding 0.1- and 1-mm diameter zirconium beads (approximately 200 µg of each) and 15 min of vortexing at highest speed; the tubes were held vertically during vortexing. Lysozyme (2.5 mg/ml) was added and the tubes incubated for 1 h at 37°C in a shaking incubator. Sodium dodecyl sulfate (SDS) at a final concentration of 1% and proteinase K (500 µg/ml) were added, followed by 2 h of incubation at 55°C. After one hour of incubation, proteinase K was added again at the same concentration. A final step was done consisting of 10 min incubation at 65°C with 0.07 Volume of 5M NaCl and 0.07 vol. of CTAB/NaCl (preheated at 65°C). DNA was then purified by adding 1 vol. of phenol/chloroform/isoamyl alcohol followed by 20 min of centrifugation at 13,000 rcf and 4°C. After recovering the supernatant in a new 15 ml Falcon tube, 1 vol. of chloroform/isoamyl alcohol was added followed by 20 min of centrifugation at 13,000 rcf and 4°C. The supernatant was recovered in a new 15 ml Falcon tube. 0.5

µl of glycogen (Thermo Fisher Scientific, US) and 0.7 vol. of isopropanol were added to the tube, followed by incubation for 15 min at -20°C. Another centrifugation was done for 25 min at 15,000 rcf and 4°C. The isopropanol was decanted and the DNA pellet was washed using 500 µl ice-cold 70% ethanol. After a final centrifugation of 5 min, the ethanol was removed and the pellet was dried at air for approx. 5 min. Then, 50 µl of PCR-grade water was added to each tube. Finally, the DNA concentration was measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and the DNA quality (resp. HMW) examined after agarose gel electrophoresis.

The 180 – 5 µm size class nanoplankton (NP) was studied using 18S rRNA gene-fragment amplification, targeting the eukaryotic nanoplankton community (18S-NP). The 5 – 0.2 µm size class picoplankton (PP) was studied using both 16S and 18S rRNA gene-fragment amplification, targeting respectively the eukaryotic picoplankton (18S-PP) and the free living bacterioplankton (16S-PP). A two-step PCR approach was used for the DNA library preparation with Illumina dual indexing. The primer pairs of the first PCR amplified the DNA fragment of interest and included a universal 5´-sequence tail. Amplification of the V3-V5 hypervariable region of the 16S rRNA gene (569 bp) was performed using the 357F and 926R primers, respectively (Schuurman et al. 2004) , (Walters et al. 2016). Amplification of the V4 hypervariable region of the 18S rRNA gene (378 bp) was done using the 573F and 951R primers (Mangot et al. 2013). Table III.1 includes the sequences of the universal 5´-tail and the rDNA primes sequences. PCR was performed with 0.02 U/ µl of Phusion High Fidelity DNA polymerase, 1X Phusion HF buffer and 200 µM of dNTPs (New England Biolabs, USA). Extracted DNA was added at a final concentration of 0.12 ng/µl. A final concentration of 0.5 µM of each primer was used. The number of PCR cycles was 10. It comprised the initial denaturation step of 3 min at 98°C, the denaturation of 45 s at 98°C, primers annealing step of 20 s at 62.4°C, extension step of 8 s at 72°C and the final extension step of 5 min at 72°C. The reactions were done using a T100 Thermal cycler (Bio-Rad, USA). PCR products of the first PCR were used as template for the second PCR with Illumina adaptors. The sets of primers pairs used included an individual index (barcode sequence) specific to each sample and the Illumina adapters. The same conditions were used as with the first PCR, but the number of cycles was 20. PCR products of both steps were each purified following the Agencourt AMPure XP protocol, with a 0.9 and 0.8 vol. beads and a final elution volume of 11 µl and 14 µl for the first and second

purification, respectively. PCR products were pooled in equivalent quantities and send to Eurofins GATC Biotech for sequencing using Illlumina MiSeq 2*300 bp with the Microbiome Profiling Indexing only-package. Each amplification, 16S and 18S DNA amplicon, was sequenced independently two times.

Table III.1: Sequence and targeted region of the primers used for the amplification of the 16S and 18S rRNA gene for barcoded amplicon sequencing library preparation.

Primer name	Sequences (5´-3´)	Targeted region	Reference
Universal 5´ forward	TCGTCGGCAGCGTCAGATGTG TATAAGAGACAG		Amplicon deep sequencing
Universal 5´ reverse	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG		
357F	CTCCTACGGGAGGCAGCAG	16S rDNA V3-V5	Schuurman *et al*., 2004
926R	CCGYCAATTYMTTTRAGTTT		Walters *et et*., 2016
573F	CGCGGTAATTCCAGCTCCA	18S rDNA V4	Mangot *et al*., 2013
951R	TTGGYRAATGCTTTCGC		

Bioinformatics pipeline

The sequencing data provided by Eurofins GATC Biotech was paired end. Quality of the sequences was checked using FastQC (Simon 2010). The poor quality of the 16S rDNA reverse reads made the output of the merging of the paired end reads not reliable. Therefore, the analysis of the 16S rDNA dataset was done using only the forward reads. Reads were trimmed using Trimmomatic (Bolger et al. 2014) discarding reads with a phred score below 3 at the start and the end of the reads, an average quality of 11 over a window of 3 bases within the reads. The 18S rDNA reads were also cropped at a maximum size of 285 to remove base with lowest quality on the merging area. The 18S rDNA paired end reads were merged using NGmerge (Gaspar 2018) with parameters of a minimum overlap of 30 bases and an acceptable error rate of maximum 10%. Reads quality was checked between each step previously described. Denoising and taxonomic affiliation were then performed on QIIME2 2020.11. Filtration of chimeras using the consensus method, denoising and dereplication of the quality reads were done using the denoise and dereplicate single-end sequences (dada2, denoise-single) and a read learn of 2,000,000 reads for the

training error model (Callahan *et al*., 2016). For the 16S rDNA database, taxonomic affiliation was done using the classify-consensus vsearch program and the TaxAss pipeline (Rohwer et al. 2018), which used both a general database, SILVA_138, and a freshwater ecosystem-specific database, FreshTrain (Newton et al. 2011). Four thresholds of identity were used, 80%, 90%, 97% and 100%. Outputs were then compared to check for consistency and to keep the deepest taxonomic affiliation while having a minimum of unassigned sequences. The 18S rDNA taxonomic affiliation was performed using Blast+ with the top 5 hits Megablast on the SSU_eukaryote_rRNA reference database. After checking the consistency of the top-5 hits, one was kept and its Unique Subject Taxonomy ID (staxids) was linked to the full lineage using a custom reference database: the reference database was created from the NCBI archive new_taxdump using NCBItax2lin (available via https://github.com/zyxue/ncbitax2lin). A parallel taxonomic affiliation was performed using the classify-consensus vsearch program and the reference database PR2 (Protist Ribosomal database) (Guillou et al. 2013). Both taxonomies were compared to verify the consistency of the taxonomy. Phylogenetic trees were constructed using Fasttree 2.1 under the General Time Reversible (GTR) evolution model (Price et al. 2010) and rooted with an outgroup on R (R core team, 2018) using the function root of the ape package (Paradis et al. 2004).

Biostatistics

All statistical analyses were performed with R software (R core team, 2018) using the package Phyloseq (McMurdie and Holmes 2013), vegan (Oksanen et al. 2015) and aldex2 (Fernandes et al. 2013; Gloor et al. 2016b). The rendering of the graphs was done using the package ggplot2 (Wickham, 2016). Before filtering the lowest abundant ASV, the dataset was split into the different conditions and sub-conditions (see Table III.2 in the Results section of this Chapter). ASV represented with less than 0.01% in relative abundance in less than 5% of the samples were discarded. ASV affiliated to mitochondria and chloroplast for the 16S rDNA datasets and ASV affiliated to Metazoa for the 18S rDNA datasets were also discarded. No rarefaction to a minimum number of reads for all samples was applied on the datasets (McMurdie and Holmes 2014). Samples with a number of reads below 10,000 were discarded. This number was considered as the minimal number of reads to cover the overall richness observable from the sequencing.

Shannon-Wiener and Simpson diversity indices were used for the observation of the alpha diversity, with the Observed ASVs and Pielou index as complement for the richness and evenness (Simpson 1949; Pielou 1966; Spellerberg and Fedor 2003). The beta diversity was visualized by Principal Component Analysis (PCA) based on the PhilR distance calculation (phylogenetic isometric log-ratio transform) (Silverman et al. 2021). 95% confidence interval assuming a multivariate t-distribution were calculated. Correlation between distance matrices (ULC *versus* LLC) was performed using a Procrustes analysis and Mantel test (Nekola and White 1999; Peres-Neto and Jackson 2001). Permutational multivariate analysis of variance (PermANOVA) with 999 permutations and Analysis of similarity (ANOSIM) were used to test the difference in the community composition. Homogeneity of variance between group was analysed using betadisper, which calculates the average distance to the group median. PermANOVA, ANOSIM and betadisper tests were all performed on the PhilR distance matrix. P-value threshold for significance was set to 0.05. The pseudo-F value and R2 of the PermANOVA can be interpreted as the effect size and the percentage of variance explained by the tested groups, respectively. The ANOSIM statistic F-value, which ranges from 0 to 1, represents the similarity of the groups tested, with a high value representing a high degree of dissimilarity. To test which taxa are more represented either in the ULC or LLC microbial community, aldex2 was used (ANOVA-Like differential expression tool for high throughput sequencing data). No p-value threshold was used to test the significance. Instead, we considered that a taxon was more represented in one of the conditions when the effect size was above 1.4 with a 95% confidence interval that did not overlap with 0. Phylum, class, order, family and ASV taxonomic ranks were tested. To generate the centered log ratio transformed values, 1000 Monte-Carlo instances using the geometric mean abundance of all ASVs were calculated. Our samples were independent so the paired.test function was set to FALSE for the calculation of the effect size and confidence interval.

Results

Our datasets comprise the two sampling sites Upper Lake Constance (ULC) and Lower Lake Constance (LLC), sampled every two weeks each from March 2018 to March 2019, thus, is comprising a total of 26 sampling dates per site. Each sampling date per site was represented by four biological replicates (n = 4), as water samples were collected independently with the integrating water sampler from 0 – 20 m water depth, into four separate canisters for filtration. The nano- and picoplankton biomass was collected separately into two size classes, the 180 – 5 µm (nanoplankton) and 5 – 0.2 µm (picoplankton) size class. Three replicates were used for the downstream analysis. The nanoplankton (NP) samples were used for sequencing of only the 18S rDNA fragment, as a representation of the eukaryote nanoplankton community (18S-NP dataset). The picoplankton (PP) samples were used for amplicon sequencing of both 16S and 18S rDNA fragments, resulting in two separate datasets, as representations of the eukaryotic picoplankton (18S-PP dataset) and the free-living (single cell) prokaryote picoplankton community (16S-PP dataset). Two independent Illumina paired-end sequencing runs were performed on each sample in order to examine the variability introduced by the sequencing step and to increase the number of reads per sample by pooling them. The results did not show divergence of samples between the sequencing runs; hence they could be merged confidently (see supplementary file Figure SIII.1 to Figure SIII.3). In total, before merging the forward and reverse reads, 1872 files were retrieved from the sequencing provider.

After removing samples that were not fitting the minimal number of reads-threshold, no full triplicates from a single sampling date had to be removed, so our datasets did not comprise time frame gaps. We had to remove two replicates from the ULC 18S-NP dataset for the sampling date 10.04.2018, and also two from the LLC 16S-PP dataset for the date 20.11.2018. Hence, only a single replicate was left for these two sampling dates (e.g., no error bars in the alpha diversity graphs), but the downstream analysis was not affected by this single representation as the samples were clustered into larger groups, e.g., according to seasons. Unfortunately, the 16S rDNA Illumina reverse reads were not of sufficient in quality to allow for a merging with the forward reads, and we therefore had to perform the analysis with the 300 bp forward reads only. Although shorter, the amplicons still comprise enough phylogenetic information, by covering the complete V3-hypervariable region and a

large proportion of the V4-hypervariable region. Table III.2 summarizes the number of samples per plankton size class and sampling site, at each step of the bioinformatics process, and the final number of detected ASVs after the processing of the datasets.

Size class	Microbial plankton	Sampling site	Raw files paired end	Merge reads	Per sites	Concatenate	Filtration	ASVs number by site	ASVs number by microbial plankton
180 - 5 µm	18S-NP	ULC	624	312	156	78	73	1216	1708
		LLC			156	78	70	1378	
5 - 0.2 µm	18S-PP	ULC	624	312	156	78	78	1506	2195
		LLC			156	78	72	1769	
	16S-PP	ULC	624	312	156	78	78	1044	1513
		LLC			156	78	75	1158	

Table III.2: Overview of samples and sampling sites, at individual steps of the bioinformatics processing plus the number of detected ASVs per microplankton type within each sampling site after the data was processed. A total of 5416 ASVs could be distinguished across all samples and sampling sites. A representation of the individual sampling dates per site, can be found in the Supplemental information file, Table SIII.1. Abbreviations used: 18S-NP, eukaryotic nanoplankton dataset; 18S-PP, eukaryotic picoplankton dataset; 16S-PP, prokaryotic picoplankton dataset; ULC, Upper Lake Constance (*Wallhausen* sampling site); LLC, Lower Lake Constance (*Zeller See* sampling site); ASV, amplicon sequencing variant.

Plankton diversity dynamics

Community dynamics were analysed through the classical alpha and beta diversity analysis. The alpha diversity was observed using the Shannon and Simpson index complemented by the Observed ASVs for the richness, and the Pielou index for the evenness. The beta diversity was calculated using the PhilR method, taking into account the phylogenetic information as well as the compositional nature (relative abundance) of our data, and visualised with PCA. Statistical test of the difference of community composition between groups was done using PermANOVA and ANOSIM and the homogeneity of variance was examined with betadisper. Correlation between distance matrices calculated for the ULC and LLC data was tested with Procrustes analysis and Mantel test.

For the eukaryotic nano- (18S-NP) and picoplankton (18S-PP) datasets, the highest diversity was displayed for the warmer period from spring to summer compared to late autumn and winter, with the LLC diversity increasing slightly earlier in spring, for both the eukaryotic nano- and picoplankton (Figure III.1AB). The overall increase in diversity was due to a higher number of ASVs detected (Figure SIII.5AB), and the fact that the Simpson index (Figure SIII.4AB) increased less intensively than the Shannon index (Figure III.1AB), indicated that the additionally detected ASVs in spring and summer belong rather to the low abundant community members. However, the Shannon diversity for the 18S-PP dataset showed sudden decreases for specific sampling dates: two of these drops in diversity for ULC, one in summer (28.08.2018) and one in winter (26.02.2019), and one for LLC, also in winter (06.12.2019). These were due to decreases in evenness (Figure SIII.6B), with few ASVs strongly increasing in relative abundance compared to the previous sampling date, and becoming dominant, while being decreased again at the next sampling date.

For the prokaryotic picoplankton (16S-PP), high diversity was observed for the winter period for both ULC and LLC (Figure III.1C). These increases of diversity coincided with an increased evenness (Figure SIII.6C), which indicated that the diversity increase resulted from a more even distribution of the relative abundances of the detected ASVs. The main difference between the 16S-PP in ULC and LLC was that the mid-summer to late-autumn period showed the lowest 16S-PP diversity in ULC, while it was the highest in LLC. This difference was due to an increase in the number

of detected ASVs in LLC during this period, from about 400 detected ASVs to about 600 (Figure SIII.5C).

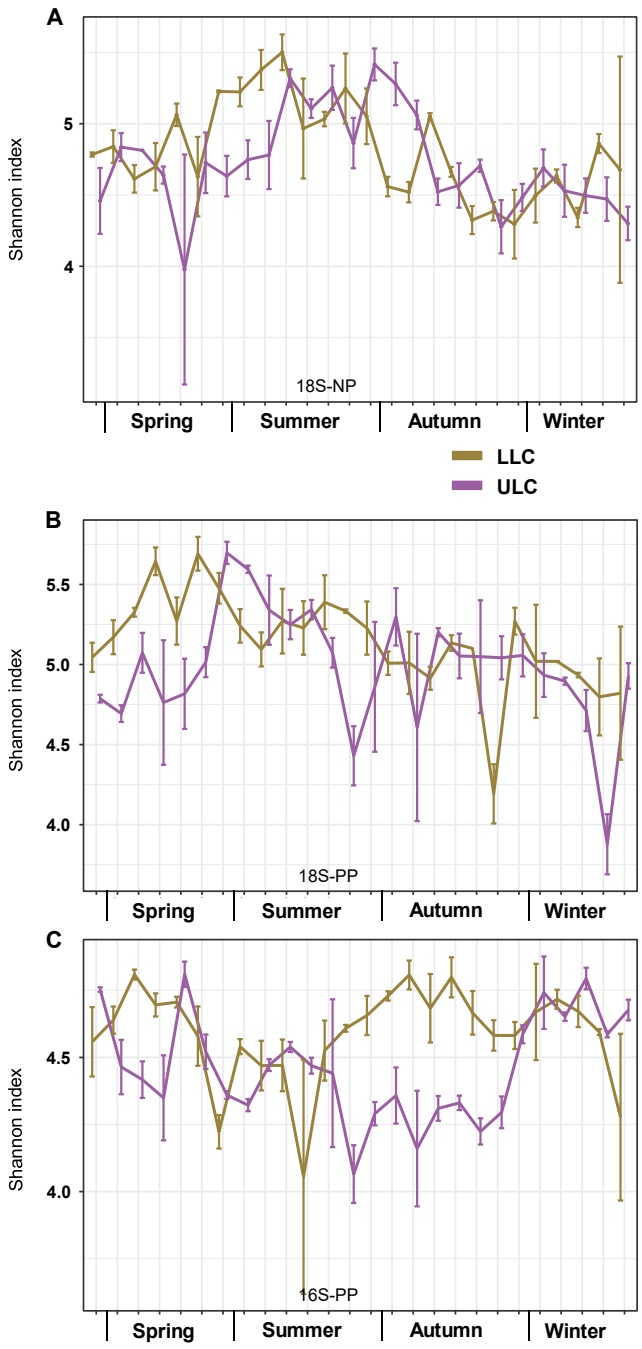

Figure III.1: Shannon diversity calculated per sampling site and across the sampling dates. The purple lines correspond to the Upper Lake Constance (ULC) and the gold lines correspond to the Lower Lake Constance (LLC) sampling site. The upper panel (A) shows the indices for the eukaryotic nanoplankton (18S-NP dataset), the central panel (B) for the eukaryotic picoplankton (18S-PP dataset) and the lower panel (C) for the prokaryotic picoplankton (16S-PP dataset). The x-axis represents the different seasons (for exact sampling dates, please see Table SIII.1). Note that the y-axes have different scale for each of the three graphs. Graphical representations of the Simpson indices are shown in the Supplemental information file (Figure SIII.1) as well as for the richness (Figure SIII.2) and evenness (Figure SIII.3).

The comparison of biodiversity (beta diversity) between samples was first calculated for each sampling site independently, in order to examine the seasonal dynamics of the 18S-NP, 18S-PP and 16S-PP communities in ULC and LLC. As expected, a cyclic temporal dynamic was observed for all three plankton communities of both ULC and LLC (Figure SIII.7A-F). The similarity of the three plankton community temporal dynamic between ULC and LLC was confirmed by Procrustes analysis of shape distribution and Mantel test of correlation between matrices. The Procrustes analysis used geometric analysis to compare the shape of the temporal dynamic between ULC and LLC while the Mantel test looked at the correlation between the ULC and LLC dissimilarity matrices. Both tests' results consisted of correlation values, p-values and sum of squares for the Procrustes. All p-values were below the chosen significance threshold of 0.05 and each correlation value ranged between medium to high depending on the test, coupled with a low sum of squares values for the Procrustes analysis. Table III.3 summarizes the results of the analysis and Figure SIII.8 visualizes the Procrustes analysis. The results suggested that all three plankton communities indeed had a similar temporal pattern of biodiversity in ULC and LLC.

However, even though the temporal dynamics of biodiversity was similar between ULC and LLC, this does to imply that the taxa constituting these communities were equivalent. Samples collected during the same time period (seasons) from ULC and LLC might not cluster together, indicating biodiversity differences. Therefore, distance matrices were calculated for each plankton type after the datasets for each ULC and LLC were pooled, as illustrated in Figure III.2. For the 18S-NP communities, the ULC and LLC samples clustered comparatively close together, i.e., showed centroids close together and overlapping ellipses of 0.95-confidence level (Figure III.2A). A somewhat similar pattern was observed for the 18S-PP biodiversity (Figure III.2B), again with centroids relatively close together and the confidence ellipse of LLC almost completely contained within the one for ULC. Nevertheless, part of the ULC samples, corresponding to autumn and winter, was clearly separated from that of the LLC, indicating a difference of biodiversity for autumn and winter (Figure III.2A). Note that also the axes on which these samples were spread were different, with the ULC samples more widely spread along the PC1-axis and the LLC samples along the PC2 axis, indicating potential differences in the main environmental factors shaping these communities. The strongest differences were observed for the biodiversity dynamics

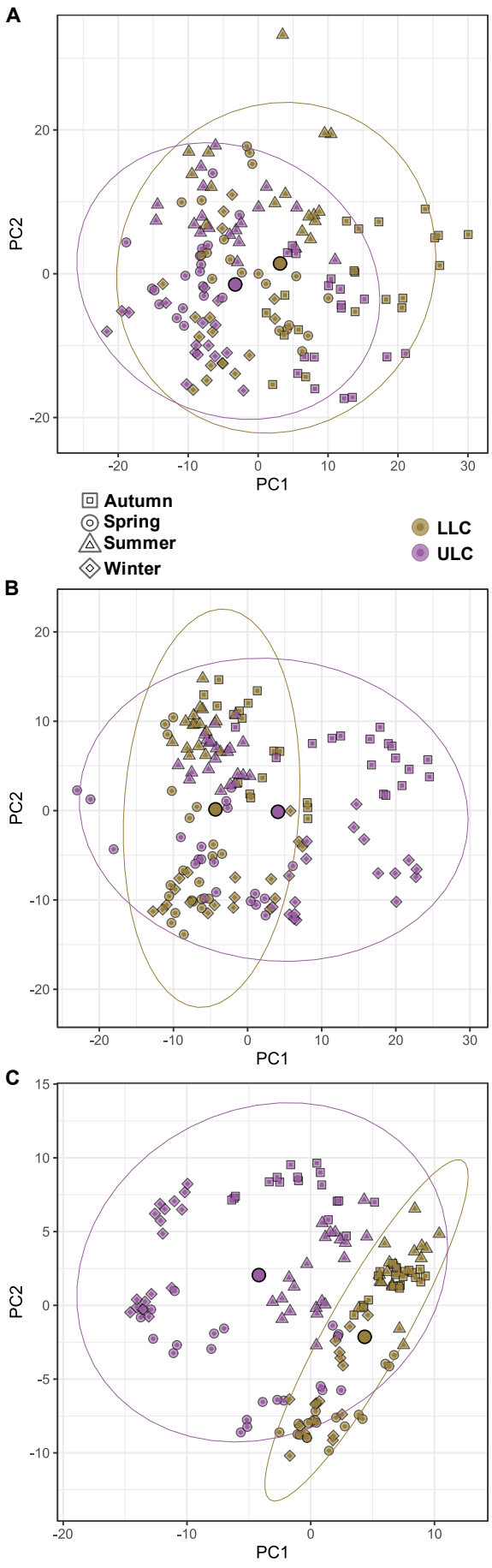

Figure III.2: Principal Component Analysis of nano- and picoplankton biodiversity across the seasonal sampling of the different parts of the lake. Dissimilarity matrices calculated based on the phylogenetic isometric log-ratio (PhilR)-transformed Euclidean distance matrix. Colours indicate the sampling site; gold, LLC; purple, ULC. Shape indicates the season, square (autumn), round (spring), triangle (summer) and diamond (winter). The large circle with black outline corresponds to the centroid per sampling site. Ellipses were calculated following a multivariate t-distribution with a confidence level of 0.95. The top panel (A) shows the data for the eukaryotic nanoplankton (18S-NP), with the ellipses pretty much overlapping, indicating similar biodiversity dynamics in ULC and LLC and across the seasons. The central panel (B) shows the data for the eukaryotic picoplankton (18S-PP), indicating a certain separation of biodiversity change, even though the ellipses still pretty much overlapped. The lower panel (C) shows the data for the prokaryotic picoplankton (16S-PP), indicating a clear separation of biodiversity change between ULC and LLC across the seasons.

of the prokaryotic picoplankton communities between ULC and LLC: the 16S-PP biodiversity displayed a clear difference between ULC and LLC, with clearly separated centroids and the 0.95 confidence ellipses not highly overlapping. Furthermore, both the eukaryotic and prokaryotic picoplankton of ULC showed a stronger biodiversity change than the LLC picoplankton communities, as samples for ULC were more widely spread along both axes than for LLC (cf. Figures III.2B and C). The PermANOVA and ANOSIM results are summarized in Table III.4. Based on the p-values, all three plankton community compositions showed significant differences based on the sampling sites. However, the 18S-NP and 18S-PP communities showed low pseudo-F ratio and the R^2 values for the PermANOVA as well as mild ANOSIM statistic F value (Table III.4). Contrary to the p-value, these results indicated relative similarity of biodiversity between the ULC and LLC 18S-NP and 18S-PP communities. These results are discussed further below.

Table III.3: Results of the Mantel tests and Procrutes analysis testing the correlation of the temporal patterns of plankton biodiversity change in ULC compared to LLC.

Size class	Microbial plankton	Procrustes analysis			Mantel test	
		Correlation	Sum of squares	p-value	Correlation	p-value
180 - 5 µm	18S-NP	0.79	0.38	<0.001	0.45	<0.001
5 - 0.2 µm	18S-PP	0.87	0.24	<0.001	0.67	<0.001
	16S-PP	0.84	0.29	<0.001	0.64	<0.001

Table III.4: Results of PermANOVA and ANOSIM tests comparing the biodiversity differences of the three plankton communities between ULC and LLC.

Size class	Microbial plankton	PermANOVA			ANOSIM	
		R^2	Pseudo-F ratio	p-value	Statistic R	p-value
180 - 5 µm	18S-NP	0.08	12.56	<0.001	0.17	<0.001
5 - 0.2 µm	18S-PP	0.09	14.83	<0.001	0.25	<0.001
	16S-PP	0.17	31.53	<0.001	0.50	<0.001

Taxonomic community composition

For the eukaryotic nano- and picoplankton (18S-NP and 18S-PP) and across all sampling sites and dates, Ciliophora, Myzozoa (composed of Dinoflagellates), Ochrophyta, Cercozoa, Cryptophyta and Chlorophyta were the major phyla detected. Additionally, Bacillariophyta (diatoms) was a major phylum in the 18S-NP dataset and Haptophyta in the 18S-PP dataset. These phyla together represented 93.9% and 85.1% of the NP and PP total 18S rDNA relative abundance (Figure SIII.9 and SIII.10). The average relative abundance of these major phyla in all three datasets, calculated across all sampling dates, but separated by sampling site and by size class, are summarized in Table SIII.2. Sequencing reads affiliated to Ciliophora were particularly dominant in the 18S-NP dataset, by up to 73.9% in relative abundance for individual sampling dates (Figure SIII.9), while for the 18S-PP dataset, these eight major phyla were more evenly shared (Figure SIII.10). At the family level, the 18S-NP dataset was dominated by Stephanodiscaceae, Strobilidiidae, Strombidiidae, Gymnodiniaceae, and Histiobalantiidae reads (together 43.7% of the total relative abundance) and the 18S-PP dataset by Chrysochromulinaceae, Geminigeraceae, Katablepharidaceae, Cryptomonadaceae and Colepidae reads (together 33.5% of the total relative abundance). For the prokaryotic picoplankton (16S-PP), the major phyla detected were Actinobacteria, Bacteroidota, Proteobacteria, Cyanobacteria, Planctomycetes, Chloroflexi, Verrucomicrobia and Acidobacteria, across all sampling dates and sites; these eight phyla represented 98.9% of the total 16S rDNA reads (Table SIII.2). At the family level, the 16S-PP dataset was dominated by the Actinobacteria acI and acIV lineages (26.3% and 14.2%, respectively), and together with the Cyanobiaceae family, Proteobacteria betI lineage and Bacteroidota bacI and bacV lineages, these represented about 65% of the total relative abundance across all sampling dates and sites (Figure SIII.11).

Differences in the plankton communities of ULC and LLC were examined. First, we examined whether there were unique taxa (ASVs) in either ULC or LLC, and second, whether ASVs detected for both parts of the lake were more or less abundant in ULC or LLC at particular time periods. Venn diagrams were drawn to find the unique plankton community members, and bar plot were created to illustrate their phylogenetic distribution. The aldex2 package was used to find ASVs with significantly different relative abundances in either ULC or LLC, by estimating the effect size

(median of the ratio of the between group difference) and the confidence interval for each ASV.

Unique ASVs in either ULC or LLC

Unique ASVs detected in either ULC or LLC across the 18S-NP, 18S-PP and 16S-PP datasets, represented 28.9% and 38.1% of total ASVs in ULC and LLC. Hence, a higher proportion of ASVs was unique to the LLC. Further, the 16S-PP dataset contained the highest proportion of unique ASVs (Figure III.3). Their temporal dynamics in relative abundances, relative to the total relative abundance change, is illustrated in Figure III.4, represented as stacked bar plots with the total numbers of unique ASVs added together at the phylum level. For ULC, the highest proportion of a unique community was detected during winter (Figure III.4A-C-E), with up to about 30% of the total community relative abundance for the 18S-PP dataset; another increase was visible for the summer. For LLC, the unique ASV abundance showed an opposite trend, with highest proportions for end of summer to mid-autumn across all three datasets (Figure III.4B-D-F). For the 18S-NP dataset, the unique ASV abundance showed also an increase for March 2018 and again for February-March 2019 (Figure III.4B).

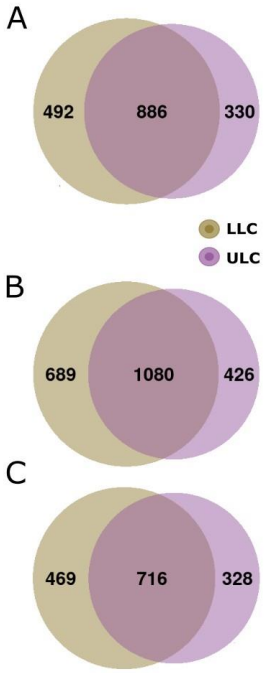

Figure III.3: Venn diagrams illustrating the number of shared and unique ASVs in the Upper and Lower Lake Constance plankton across all sampling dates. Purple corresponds to the Upper Lake (ULC) and gold to the Lower Lake Constance (LLC) sampling site. The top panel (**A**) represents the eukaryotic nanoplankton (18S-NP), and the central panel (**B**) the eukaryotic (18S-PP) and the lower panel (**C**) the prokaryotic (16S-PP) picoplankton.

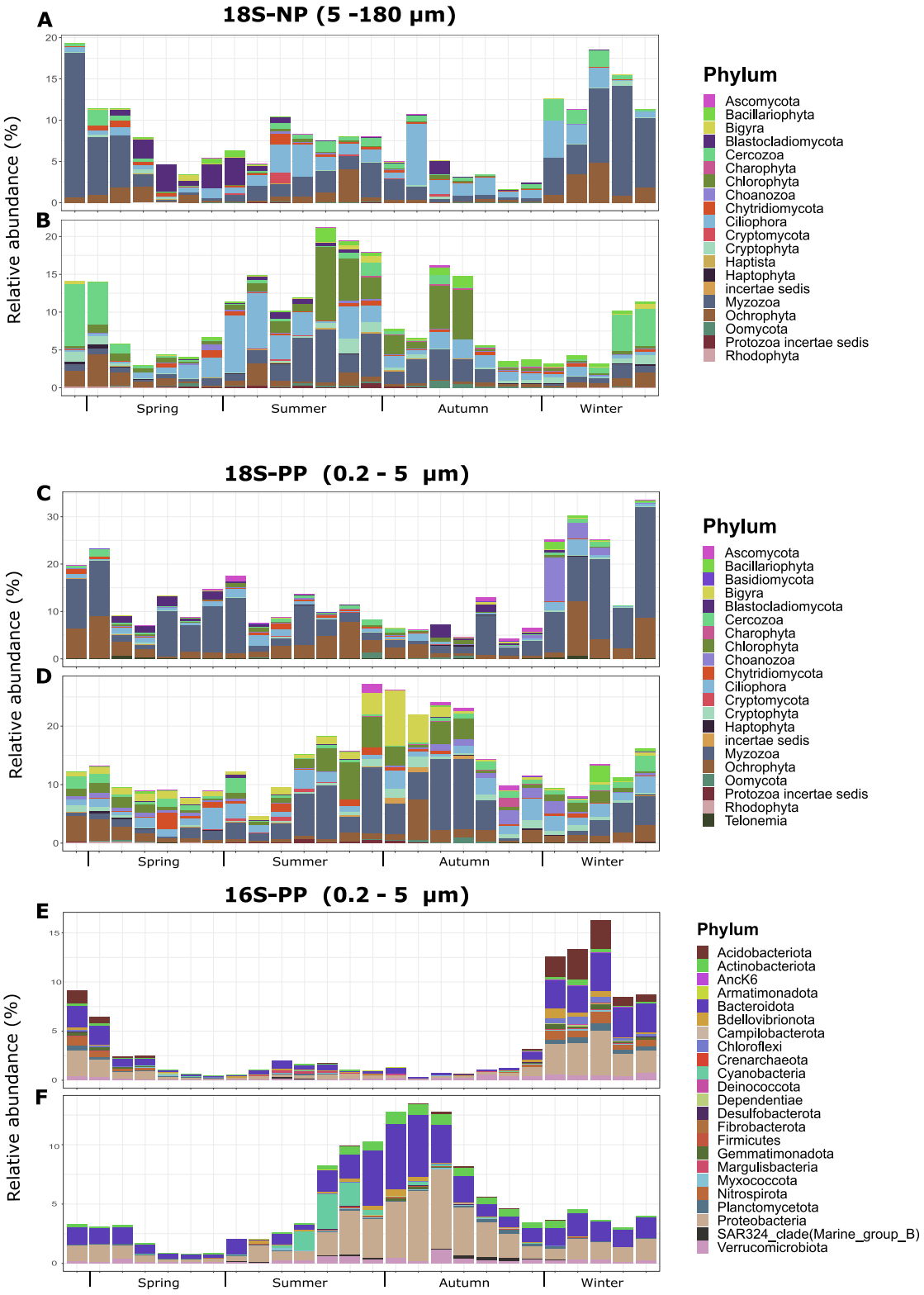

Figure III.4: Relative abundances of the unique ASVs detected in either ULC or LLC across the sampling dates. The stacked bar plots illustrate the percentages of unique ASV per phylum relative to the total number of ASVs (unique and shared) per lake parts ULC and LLC and across all sampling dates. The top panels represent the unique eukaryotic nanoplankton (18S-NP) in ULC (**A**) and LLC (**B**), the central panels the eukaryotic (18S-PP) picoplankton in ULC (**C**) and LLC (**D**) and the lower panels the prokaryotic (16S-PP) picoplankton in ULC (**E**) and LLC (**F**).

These unique ASVs were taxonomically very diverse, with affiliations across 20, 21, and 23 phyla of the 18S-NP, 18S-PP and 16S-PP datasets, respectively. In Table SIII.3, the main phyla of the unique ASVs are summarized with their relative abundance ranges. The richness in unique ASVs (on the phylum level) ranged from only three unique Acidobacteriota ASVs in the ULC 16S-PP dataset up to 206 unique Proteobacteria ASVs in the LLC 16S-PP dataset. For some phyla, a similar relative abundance in unique ASVs was observed for both ULC and LLC, such as for Ochrophyta (18S-NP) with 50 and 77 unique ASVs respectively, or Actinobacteria (16S-PP) with 22 and 35 unique ASVs. Other phyla showed clear site differentiation, such as Chlorophyta (18S-NP) with only 12 unique ASVs for ULC but 85 for LLC, or Bigyra (18S-PP) with only 11 unique ASVs for ULC but 93 for LLC (Table SIII.3).

The unique ASVs detected in the 18S-NP dataset for ULC were mainly affiliated to Myzozoa in winter, Ochrophyta in winter and mid-summer, Ciliophora from summer to mid-winter and Blastocladiomycota in spring (Figure III.4A and Table SIII.3). The unique 18S-NP ASVs detected in LLC affiliated mainly to Myzozoa, to Ciliophora during summer and autumn, and to Cercozoa specifically in March 2018 and 2019 (Figure III.4B and Table SIII.3). Further, Chlorophyta showed a high proportion of unique ASVs for LLC but not for ULC (Figure III.4B and Table SIII.3). The unique Myzoza, Ciliophora and Ochrophyta ASVs dominant in the 18S-NP datasets showed a clear separation between ULC and LLC for the deeper taxonomic rank: unique ASVs of the Myzozoa in ULC belonged in majority to the family Gymnodiniaceae. Gymnodiniaceae is also one of the ASVs major family of the unique LLC 18S-NP community (Figure III.5), but they differ in their genus affiliation as LLC ASVs affiliated to *Gyrodinium* while ULC ASVs affiliated to *Lepidodinium*. Although genus affiliation is often difficult to trust using short amplicons, the consistency observed here allowed to put some faith in this result. The unique Myzozoa community in LLC also showed a higher taxonomic diversity than the ULC one with the Protoperidiniaceae and Suessiaceae families also playing an important role (Figure III.5). Unique Ochrophyta ASVs (18S-NP) also showed a clear taxonomic separation between ULC and LLC. For ULC, they belonged to Chrysoamoebidaceae in summer and to Paraphysomonadaceae in winter. For LLC, the unique Ochrophyta community was taxonomically more diverse, being affiliated to Mallomonadaceae, Chrysophaeraceae, Chrysolepidomonadaceae and Paraphysomonadaceae (Figure SIII.12).

The unique ASVs detected in the 18S-PP dataset were mainly affiliated to Myzozoa, Ochrophyta and Ciliophora. A higher percentage of unique Bigyra and Chlorophyta were observed for LLC, while unique Blastocladiomycota and Choanozoa ASVs were more abundant in ULC (Figure III.4CD and Table SIII.3). One of the major family of Myzozoa in ULC and LLC was Gymnodiniaceae (Figure SIII.13), with the exact same genus separation: for LLC, these Gymnodiniaceae ASVs affiliated to *Gyrodinium* and for ULC to *Lepidodinium*. Apart from Gymnodiniaceae, further unique Myzozoa ASVs affiliated to the families Amphidiniaceae and Prorocentraceae for ULC and to Prorocentraceae and Goniodomataceae for LLC (Figure SIII.13). In addition, unique Bigyra ASVs showed a very specific presence in the LLC community. Their abundance peak was observed at beginning of autumn, constituted of only a few specific ASVs that all affiliated to the Pseudophyllomitidae family. A similar pattern was observed for ULC, though early in winter and with ASVs affiliated to Choanozoa and the two families Codonosigaceae and Filiasterea incerta sedis.

The unique ASVs detected for the 16S-PP in ULC affiliated mostly with the phyla Acidobacteria, Bacteroidota, Nitrospirota and Proteobacteria, while the unique 16S ASVs in LLC affiliated mostly with Actinobacteriota, Cyanobacteria, Bacteroidota and Proteobacteria (Figure III.4EF and Table SIII.3). Furthermore, Anck6 and Crenarchaeota (Thaumarchaeota) were present uniquely in ULC. Their differentiation at deeper taxonomic level was as described in the following. For ULC, the unique Proteobacteria ASVs belonged primarily to the Nitrosomonadaceae, Acetobacteraceae and Xanthobacteraceae families (Figure SIII.14). For LLC, the unique Proteobacteria ASVs were contributed mainly by the betI, Rhodocyclaceae, Methylomonadaceae and Methylophilaceae families (Figure SIII.14). The unique Bacteroidota ASVs as the second major phylum represented in the unique 16S-PP community, were for ULC affiliated to BSV26 and Chitinophagaceae (Figure SIII.15) and for LLC to bacI, bacII, Chitinophagaceae (*Terrimonas* genus) and Sphigobacteriaceae (Figure SIII.15). As observed for Bigyra and Choanozoa in the 18S-PP community, a sudden peak of relative abundance unique Cyanobiaceae family ASVs was observed in LLC at the end of summer. Thus, while unique ASVs affiliated to similar phyla for both parts of the lake, their family affiliation revealed a clear separation between ULC and LLC. The majority of the unique ASVs in the three datasets were mostly undetected at the exception of spike happening on specific short

period (Figure SIII.12 to SIII.15). The larger spike was observed in the 18S-NP dataset of ULC with a single Myzozoa ASVs reaching a maximum of 5.92% of the total community relative abundance while being below detection level or close to undetected most of the time (Figure III.5).

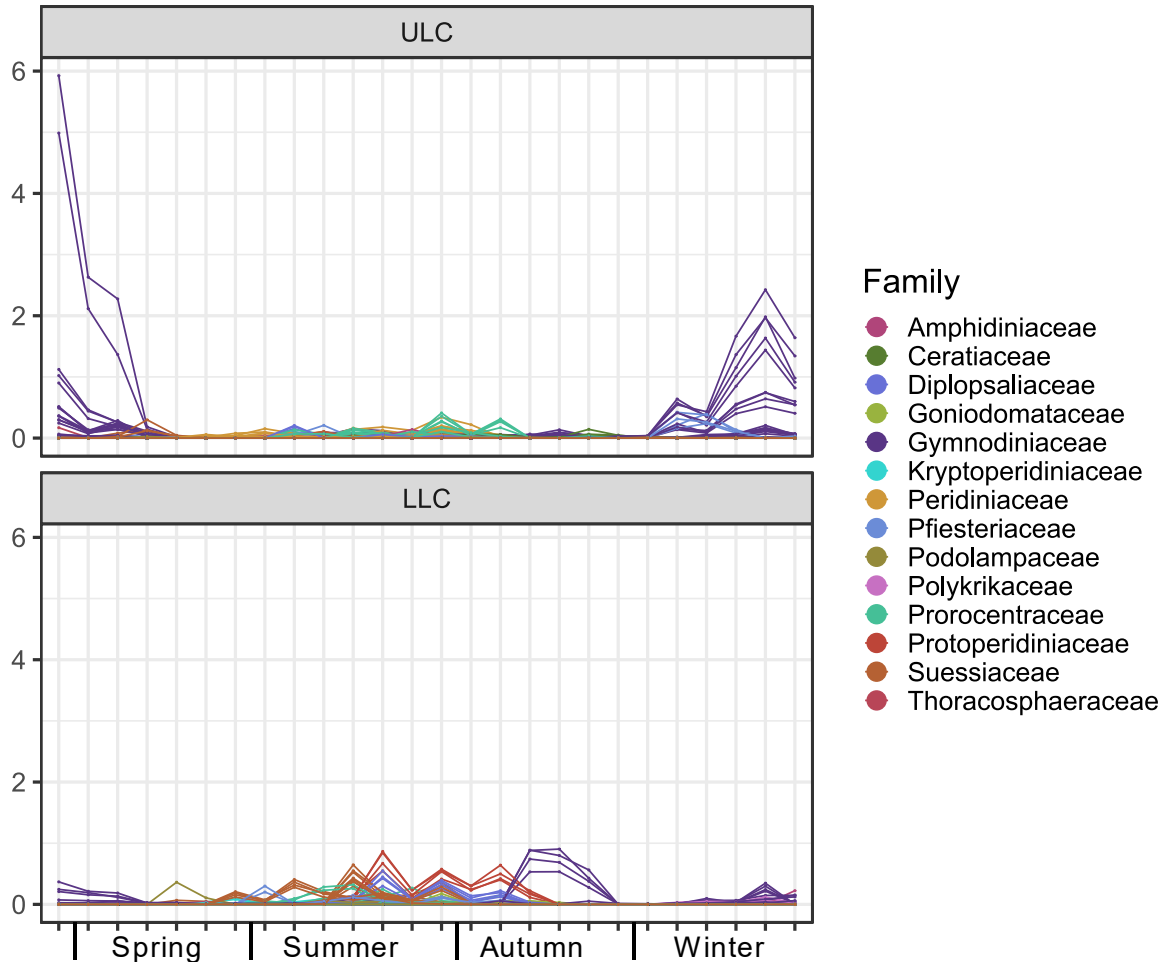

Figure III.5: Relative abundance temporal variation of the 18S-NP unique ASVs affiliated to Myzozoa. Each line represented a single ASV. ASVs are colored by family taxonomic affiliation. The top panel correspond to ULC and bottom panel correspond to LLC.

ASVs detected for both ULC and LLC with different relative abundances

An analysis for differentially abundant taxa using aldex2 was performed at the phylum, class, order, family and ASV taxonomic rank. Therefore, the dataset was divided into four different groups according to seasons. Table SIII.4 summarises the number of taxa that were indicated by aldex2 to be differentially abundant between ULC and LLC, each by season and by taxonomic rank. In total, 254 ASVs covering 50 families, 42 orders, 20 classes and 13 phyla were found to be differentially abundant across the four seasons and the three datasets. When compared to the total number of taxa in each taxonomic rank, they were in minority, representing 4.7, 10.9, 14.6, 14.0, 18.6% of the total number of ASV, family, order, class and phyla, respectively. Proportionally, the 16S-PP dataset represented 61.9% of the total differentially abundant taxa while 18S-NP and 18S-PP dataset represented only 28% and 10.1%. Figures III.6, III.7 and III.8 show forest plots illustrating the effect size of the differentially abundant taxa of the 18S-NP, 18S-PP and 16S-PP communities, respectively. A larger effect size indicates a more pronounced difference between ULC and LLC.

For the 18S-NP community, the total number of differentially abundant ASVs was quite evenly distributed over all four seasons. These were mostly affiliated to Myzozoa, Bacillariophyta and Ciliophora (Figure III.6). 37 of the 57 detected ASVs (64%) were more abundant in ULC. These taxa being more abundant in LLC were almost exclusively found in winter and spring, and belonged to the Gymnodiniaceae family of the Myzozoa phylum for winter and to the Stephanodiscaceae and *Chlamydomonadaceae* families of the phyla Bacillariophyta and Chlorophya, respectively, in spring (Figure III.6AB). For summer and autumn, a larger taxonomic diversity of differentially abundant ASVs was observed. ASVs affiliated to Myzozoa showed two opposite groups of four ASVs, one affiliated to Ceratiaceae and the other to Suessiaceae, which were more abundant in ULC and LLC, respectively. Results indicate that ASVs affiliated to the same family tended to behave similarly by season (Figure III.6).

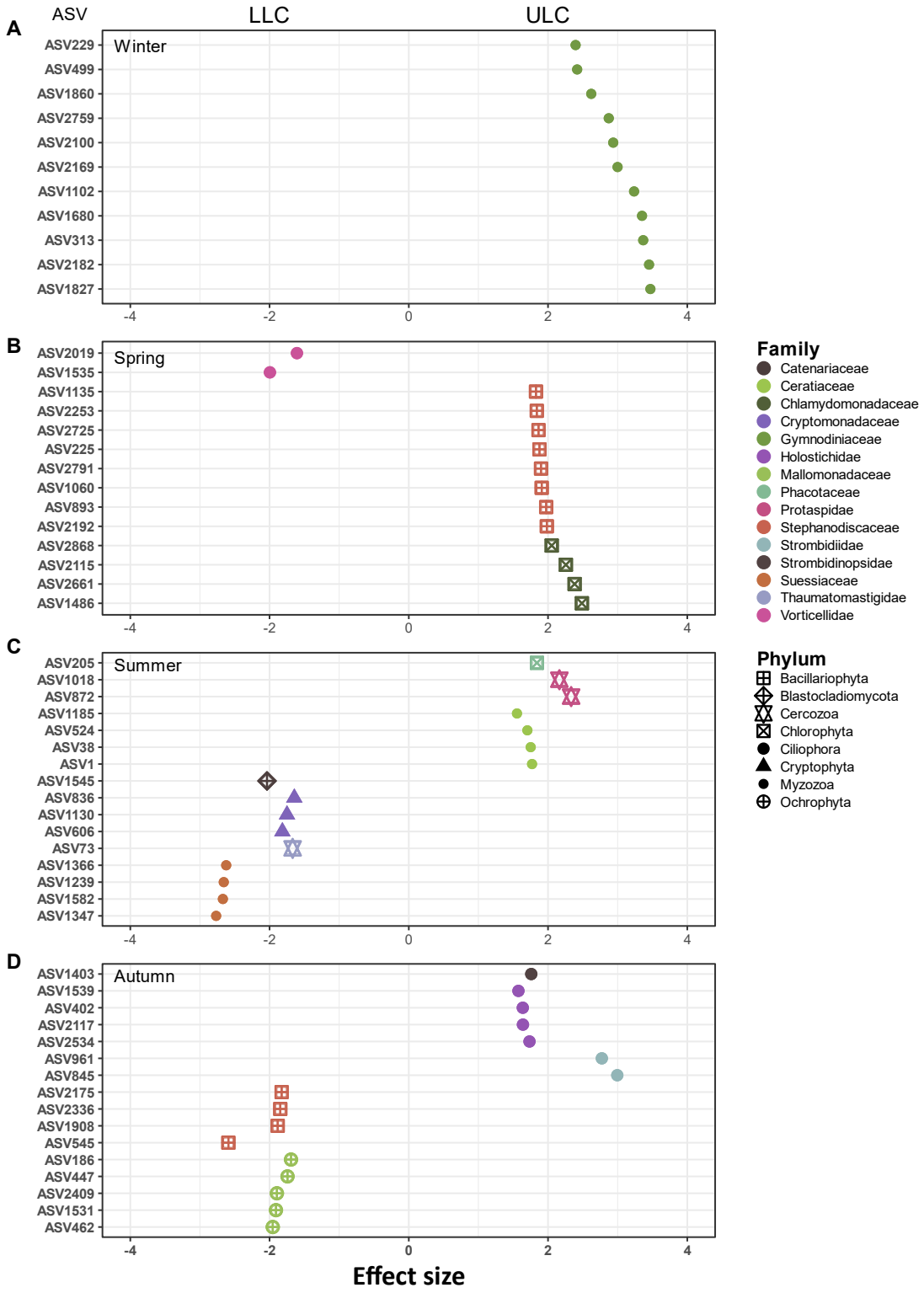

Figure III.6: Forest plots showing of the effect size of the differentially abundant 18S-NP ASVs grouped by season. A higher effect size indicates a larger difference of relative abundance of taxa between ULC and LLC, while a positive value indicates a higher relative abundance in ULC and a negative value a higher relative abundance in LLC. **(A)** winter, **(B)** spring, **(C)** summer, **(D)** autumn. The colour coding indicates the family taxonomic rank, and the symbols the phylum taxonomic rank.

For the 18S-PP community, differentially abundant ASVs were more evenly distributed over the seasons, at the exception of spring showing four ASVs with different abundance (Figure III.7). The 18S-PP dataset showed an opposite trend compared to the 18S-NP dataset, in that 34 of the its 45 ASVs (75%) were higher in relative abundance in LLC (Table SIII.4). Taxonomically, the differentially abundant 18S-PP ASVs were dominated by Cryptophyta (25 of the 45 ASVs) of the families Katablepharidaceae and Cryptomonadaceae (Figure III.7). Furthermore, ASVs affiliated to Gymnodiniaceae were found more abundant in ULC in winter. Like for the 18S-NP community, ASVs affiliated to the same family showed the same pattern of relative abundance differences. The exception was winter, where two clusters, with all ASVs affiliated to Katablepharidaceae, could be observed, one more abundant in ULC and the other in LLC (Figure III.7A). All these ASVs were taxonomically identical up to the species level.

For the 16S-PP community, the total number of differentially abundant ASVs showed a high seasonal variability, with winter accounting for 72.4% of the identified ASVs (Table SIII.4). As a much higher number of ASVs was found for the 16S-PP dataset compared to the other two datasets, the forest plots were made at the family level rather than ASV level, for a better visualisation (Figure III.8). Proteobacteria, Actinobacteria and Bacteroidota were the most affected phyla with 35, 40 and 41 ASVs from 15, six and six differentially abundant families, respectively. Spring and autumn showed ASVs of few families of Actinobacteria and Proteobacteria, mostly more abundant in LLC, such as the two gammaproteobacterial families Methylophilaceae and Methylomonadacae in autumn. Two clusters related to Proteobacteria were found for winter, with one being more abundant in ULC and one in LLC: the ULC cluster comprised Alphaproteobacteria of the families/lineages Rhodospirillaceae, Reyranellaceae, SAR11_clade1 and KF-JG30-B3, while the LLC cluster comprised Gammaproteobacteria of the families/lineages betI, betII, Comamonadaceae and Steroidobacteraceae (Figure III.8A). The single ASV affiliated to BSV26 in the unique 16S-PP ULC community was also detected here as it was abundant enough in ULC in winter (1.1% of relative abundance) to be detected by aldex2 (Figure III.7A).

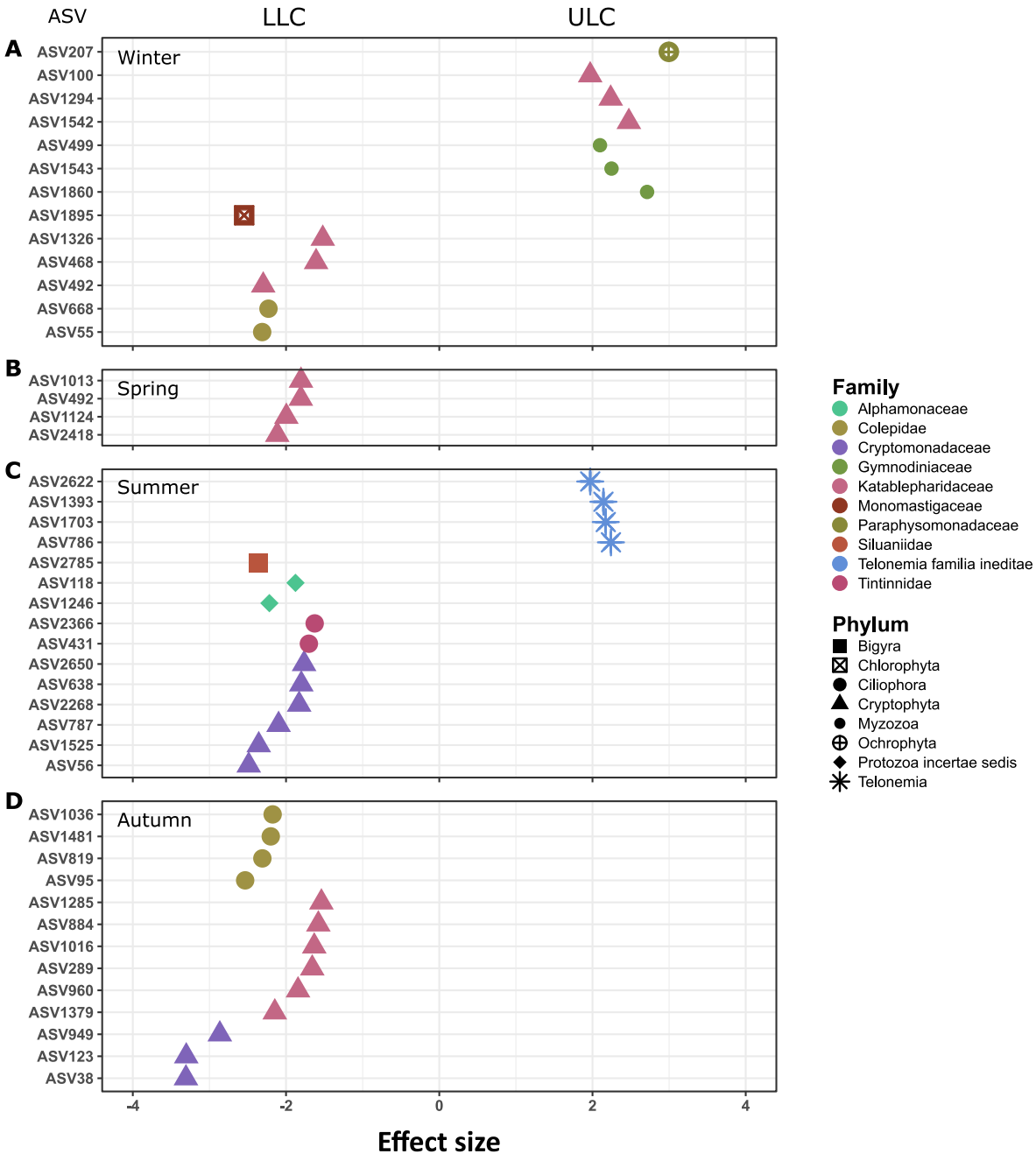

Figure III.7: Forest plots showing of the effect size of the differentially abundant 18S-PP ASVs grouped by season. A higher effect size indicates a larger difference of relative abundance of taxa between ULC and LLC, while a positive value indicates a higher relative abundance in ULC and a negative value a higher relative abundance in LLC. **(A)** winter, **(B)** spring, **(C)** summer, **(D)** autumn. The colour coding indicates the family taxonomic rank, and the symbols the phylum taxonomic rank.

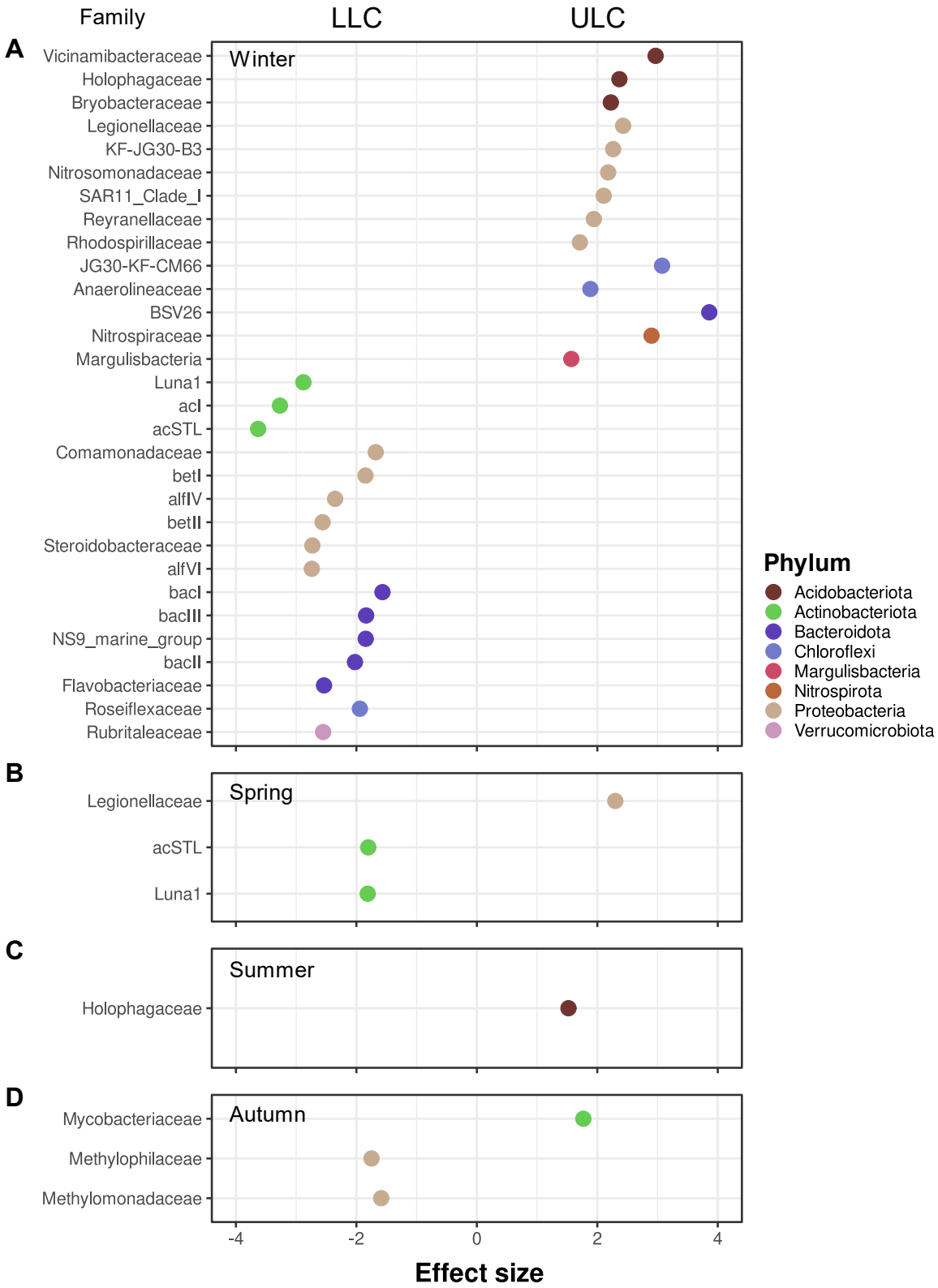

Figure III.8: Forest plots showing of the effect size of the differentially abundant 16S-PP ASVs grouped by season. A higher effect size indicates a larger difference of relative abundance of taxa between ULC and LLC, while a positive value indicates a higher relative abundance in ULC and a negative value a higher relative abundance in LLC. **(A)** winter, **(B)** spring, **(C)** summer, **(D)** autumn. The colour coding indicates the family taxonomic rank, and the symbols the phylum taxonomic rank.

Discussion

Lake Constance has been intensively studied in the past in respect to its seasonal plankton succession and was a test system for the development of the Plankton Ecology Group (PEG) model (Sommer et al. 2012). However, until today and to the best of our knowledge, no DNA-sample repository and sequencing-based study exist of the seasonal succession of the nano- and picoplankton in Lake Constance. Hence, as part of this study, a DNA-sampling procedure has been set up and added to the routine sampling campaign of the Limnological Institute of University of Konstanz, for future study of the nano- and picoplankton composition and its seasonal succession using Next-Generation Sequencing. Every two weeks, samples from the top-20 m of the water column (approx. photic zone) were filtered into two size classes, 180 – 5 µm diameter (nanoplankton, NP) and 5 – 0.2 µm diameter (picoplankton, PP). The DNA-repository is an important reference for a future detection of changes of the Lake Constance nano- and picoplankton community, given the ongoing effects of climate change, for example incomplete winter mixing due to warmer winters, and lower water inflow due to reduced snow melt in the Alpine region.

In the present study, we focused on a first, general 16S- and 18S-rDNA based ‘census’ of the taxa detectable in ULC and LLC nano- and picoplankton, across one seasonal succession. In total, at least 5416 different ASVs were distinguishable across all seasons and sampling sites, giving an unprecedented view of Lake Constance microbial plankton. However, the use of generalist 16S and 18S rDNA primers pair did not allow for a high resolution of the taxonomic rank or affiliation that can be trusted at deep taxonomic rank due to the low amount of information hold by the amplicon and the influence of the targeted hypervariable region on the phylogenetic affiliation. The use of primers pair specific to a group of organisms like cyanobacteria or diatoms (Ibrahim et al. 2021; Fournier & Riehle et al. 2021) could give a finer taxonomic resolution.

Microbial plankton dynamics

The nano- and picoplankton communities in ULC and LLC showed similar ‘circular’ temporal dynamics when displayed on PCA plots, which fits observations made for other freshwater lakes in temperate regions (Li et al. 2015; He et al. 2017; Salmaso et al. 2018; Salmaso et al. 2020).

When testing the difference in biodiversity between ULC and LLC, both PermANOVA and ANOSIM p-value were significant for the three communities, but the other values of the tests and the visualisation contradicted the p-value for the 18S-NP and 18S-PP communities (Figure III.2AB and Table III.4). PermANOVA and ANOSIM have been shown to confound location and dispersion effects resulting from heterogeneity of variance between groups, which is the case here, as the ULC and LLC groups presented heterogeneity of variance. This confounding effect is stronger in the case of unbalanced design where the number of samples between groups is not equal (Anderson and Walsh 2013; Anderson 2017), as in our study in which data for individual replicates had to be removed because they did not meet the quality criteria (Table III.2). Anderson *et al*., 2017 proposed the Behrens-Fisher modification of PermANOVA to make it more robust against unbalanced design (Anderson et al. 2017). Unfortunately, this modification has not yet been implemented in the adonis2() function of R. Thus, the results of the biodiversity comparison between the ULC and LLC 18S-NP and 18S-PP communities must be considered carefully due to this confounding effect, and neither a clear nor no difference can be concluded for these communities.

When season was included as an independent variable with the sampling site in the PermANOVA, the proportion of variance explained by season outweighed sampling site with R2 values around 0.3, indicating that seasonality is a much stronger explanation for the variance in community biodiversity than the different sampling sites.

ULC unique microbial plankton community

The relative abundance of the unique ULC 18S-NP and 18S-PP community members showed a peak during winter. Further, the unique 18S-NP and 18S-PP communities in ULC were represented by the same phyla and families. Myzozoa (dinoflagellate), with the mixotrophic genus *Lepidodinium* of the Gymnodiniaceae family (Liu et al. 2021), is the major taxa of the unique ULC community in winter. The free-living nanoflagellates belonging to Codonosigaceae and Filasterea of the Choanozoa, and the non-photosynthetic Paraphysomonadaceae of the Ochrophyta (Scoble and Cavalier-Smith 2014; Leadbeater 2015), represented the two other main unique taxa in winter in ULC. They all display a heterotrophic lifestyle and seem to favour grazing on bacteria, picocyanobacteria and detritus (Pettitt et al. 2002; Bec et al. 2010). Autotrophic activity and phototrophic phytoplankton density are negatively

correlated with the cold period (Straile et al. 2010) and species capable of mixotrophic behavior have been observed dominating the winter community (Millette et al. 2017). The relative abundance increase of these taxa during winter may be a result of feeding behavior switching from phototrophy to partial heterotrophy. They may exploit various organic and inorganic resources from their prey, complementing their nutrient demand (Stickney et al. 2000). The decrease of the phototrophic phytoplankton abundance during winter, while the more competitive mixotrophs and heterotrophs can keep a relatively stable abundance by diversifying their nutrition behaviour, would be reflected in the community composition of our relative-abundance datasets by an increase of the stable ones. Why these specific microorganisms are unique to ULC cannot be conclude here. However, their mixotrophic and heterotrophic nutrition behaviour could indicate a feeding specificity on ULC-specific microorganisms.

Another interesting feature of the unique ULC 18S-NP and 18S-PP communities was the relative abundance increase end of spring/beginning of summer, of Blastocladiomycota ASVs of the Catenariacae family. These organisms display a parasitic lifestyle and target algae, and have been shown as the major organisms responsible for the decline of an algae bloom in a drinking water reservoir (Porter et al. 2011; Zhang et al. 2018). The relative abundance increases of these unique Blastocladiomycota ASVs during the spring bloom decline could be the results of the fungi parasitic behaviour targeting specific algal of the ULC spring bloom.

The dynamics observed for the Shannon diversity of the ULC 16S-PP community, with a maximum in winter (Figure III.1C), has been observed also for other freshwater lakes (Tang et al. 2017; Ávila et al. 2017; Salmaso et al. 2018). Since evenness, but not richness, increased in winter, we can assume that most of the ASVs were detectable throughout the year as part of the low abundance community. The maximum of diversity coincides with the peak of relative abundance of the unique ULC 16S-PP community (Figure III.4E). Most of the phyla and families constituting the unique ULC 16S-PP community, namely Acidobacteria, Bacteroidota, Chloroflexi, Crenarchaeota, Nitrospirota and Proteobacteria, are known to be abundant in the oxygenated hypolimnion of deep stratified lakes (Urbach et al. 2001; Shade et al. 2008; Okazaki et al. 2013; Okazaki and Nakano 2016; Herber et al. 2020; Xing et al. 2020; Klotz et al. 2022). The abundance of ASVs of Sphingobacteriales and Chitinophagales (Bacteroidota), Rhizobiales (Proteobacteria), Vicinamibacterales

(Acidobacteria) and Nitrospirales (Nitrospirota) is positively correlated with nutrient loading and lower nutrient uptake capabilities (Pollet et al. 2011; Pollet et al. 2014; Jiang et al. 2019; Park et al. 2020; Pilgrim et al. 2022; Wu et al. 2022). The reduction of thermal stratification, which leads to mixing of the water column in winter, revived the nutrient-depleted epilimnion (Limnologischer Zustand des Bodensees). This nutrient uptake reduces selective pressure and allows the growth of bacteria that are non-competitive in nutrient acquisition (Logue et al. 2012). Nitrospora and Nitrosopumilaceae, belonging to the phyla Nitrospirota and Crenarchaeota (Thaumarchaeota) respectively, are known to be active in close association in order to perform nitrification. A single Nitrosopumilaceae ASV could be detected in our dataset, tending to confirm previous studies that studied the hypolimnion ammonia oxidizing community of Lake Constance by metagenomics and metatranscriptomics and found that the ammonia oxidizing archaea community was comprised of a single species affiliated to *Nitrosopumilus* (Thaumarchaeota) (Herber et al. 2020; Klotz et al. 2022).

Both phyla, Crenarchaeota and AncK6, were only detectable for ULC. Crenarchaeota are associated with nutrient poor lakes and deep oxygenated hypolimnion communities (Hugoni et al. 2013a; Callieri et al. 2016; Herber et al. 2020). Hence, the physico-chemical condition of LLC may not be suitable for the development of Crenarchaeota populations. AncK6 is mainly found in association with sponges and black coral (Jeong et al. 2013), and four sponges of the genus *Ephydatia* live in the northern part of the lake close to the sampling site, which may explain their presence only in ULC (Brümmer and Müller 2015).

Our results indicate that the depth of ULC coupled with the environmental changes brought by the winter, that is vertical mixing, is one of the main drivers of the unique microbial plankton community of ULC. As a high proportion of the ULC 18S-NP and 18S-PP community can exhibit mixotrophic and heterotrophic behaviour and interactions like predator – prey or parasitism, are likely to be the second driver for the development of the unique microbial plankton community in ULC relative to LLC.

LLC unique microbial plankton community

The increase of the Shannon diversity that was observed in mid-summer for the 16S-PP dataset of LLC, while that of ULC was at a minimum, was due to an increase in richness (Figure III.1C and Figure SIII.5C). This period coincides with the strongest oxygen gradient during the year, with the lower water column being hypoxic. Phytoplankton counts showed the highest density coupled with a high nitrate concentration just before the hypoxia event (Limnologischer Zustand des Bodensees). High phytoplankton activity due to an increase in nutrient concentration can be followed by oxygen depletion in the bottom water, as thermal stratification prevents reoxygenation. High nutrient concentrations such as total phosphorus, ammonium, silicate, manganese and iron were measured in the hypoxic layer (Limnologischer Zustand des Bodensees). These environmental parameters are consistent with observations in other freshwater environments, where low oxygen concentrations create chemically reducing conditions, resulting in accumulation of ammonia and other nutrients, and increased solubility of iron, manganese and other metals (Cleveland and Ayres 2004). Although not measured, we can expect high concentrations of hydrogen sulfide and methane, as observed under similar conditions in other lakes (Kosolapov et al. 2003; Biderre-Petit et al. 2011).

The maximum relative abundance of both LLC 18S-NP and 18S-PP unique communities overlapped with the hypoxia event (Figure III.4B-D). Among Myzozoa (dinoflagellate), the main families were the heterotrophic Protoperidiniaceae (Gribble and Anderson 2006), Suessiaceae which hosts species capable of mixotrophy, symbiosis and parasitism (Annenkova et al. 2011; Jang et al. 2020) and Gymnodiaceae with the heterotroph genus *Gyrodinium* capable of feeding on diatoms (Saito et al. 2006). Mesocosms experiments showed that the abundance of dinoflagellates, including Protoperidiniaceae and *Gyrodinium*, increased during anoxic events (Rocke et al. 2013; Hinode et al. 2019; Gauns et al. 2020), indicating the ability to thrive in low oxygen environments. A sudden increase in relative abundance with an equally rapid disappearance was observed for heterotrophic nanoflagellates of the phylum Bigyra, family Pseudophyllomitidae (Figure III.4D). Bigyra is a non-monophyletic group consisting of deep-branching environmental heterotrophic marine Stramenopiles (MAST) (Thakur et al. 2019). Pseudophyllomitidae is a known member of the MAST-6 group (Massana et al. 2009; Shiratori et al. 2017), comprising bacteria

and phytoplankton grazers with the ability to live in hypoxic environments (Behnke et al. 2006; Lin et al. 2022). Both Myzozoa and Bigyra taxa, which are unique to the LLC, showed potential mixotrophic and heterotrophic capabilities, consistent with previous observations that the proportion of bacterivorous and parasitic microbial plankton increases during seasonal hypoxia events (Santoferrara et al. 2022).

The third major phylum of the LLC 18S-NP and 18S-PP unique community was the autotrophic green algae Chlorophyta with the families Radiococcaceae, Sphaeropleaceae and Volvocaceae. A high abundance of Radiococcaceae has been positively correlated with nutrient loading, and the dominance of Chlorophyta in shallow lakes has been observed in the presence of high total phosphorus concentrations (Jensen et al. 1994; Yang et al. 2021). It is likely that ASVs belonging to these families were present in the oxic rather than the hypoxic layer. Although not as high as in the hypoxic bottom layer, the total phosphorus concentration in the oxic upper layer was also higher than the rest of the year and could have triggered their increase in abundance (Limnologischer Zustand des Bodensees). Ciliophora ASVs of the LLC unique 18S-NP community had their maximum abundance happening before the hypoxic event (Figure III.4B). It is likely that the high phytoplankton concentration observed during this period is the driver of this peak, as Ciliophora graze on bacteria and microeukaryotes (Posch et al. 2015; Pajdak-Stós et al. 2017). In addition to the hypoxic event, the presence of ASVs unique to LLC associated with mixotrophic and heterotrophic microorganisms also suggests a predator-prey specific interaction as observed in ULC.

The relative abundance peak of the 16S-PP unique community in LLC also coincided with the hypoxic event. The presence of ASVs affiliated to the anaerobic sulfate reducing bacteria Desulfobacterota, Sulfurimonadaceae and methylotrophs Methylomonadaceae and Methylophilaceae, only detectable during this period, tends to confirm the presence of hydrogen sulfide and methane in the hypoxic water (Beck et al. 2013; D'Angeli et al. 2019; Salcher et al. 2019; Flieder et al. 2021). Furthermore, abundant ASVs were detected for Rhodocyclaceae during this period, associated with ASVs for the anoxygenic sulfur-oxidising *Sulfuritalea* (Sperfeld et al. 2019), the obligate anaerobic Fe(III)-reducing *Ferribacterium*, and the hydrogen-metabolising *Dechloromonas* (Cummings et al. 1999; Achenbach et al.). Chitinophagaceae, bacI and certain betI species are described as facultative anaerobes, so the detection of

ASVs associated with them during the hypoxic event, along with the lineage, was not surprising (Hahn et al. 2010; Newton et al. 2011; Rosenberg 2014). Even if we can only infer their capabilities based on the taxonomy, the major ASVs of the unique 16S-PP community in LLC were all affiliated to known anaerobic or taxa containing species capable of anaerobic activities.

The hypoxia event observed in the bottom layer of LLC from end of summer to end of autumn is one of the main drivers of the unique microbial plankton community in LLC. Each of the three unique microbial plankton communities had its maximum relative abundance overlapping perfectly with the hypoxia event, and their taxonomic composition showed taxa adapted to environments with low or absent oxygen. The taxonomic composition of the unique 18S-NP and 18S-PP communities confirmed previous observations that microorganisms with mixotrophic and heterotrophic behaviour seem to be positively selected in low oxygen environments. As in ULC, the high proportion of microorganisms showing predatory behaviour also suggests predator-prey specificity in LLC. No quantitative analysis was carried out in this study, but observation indicate that oxygen depletion in water favours picoeukaryote organisms (Fenchel and Finlay 2008) so we could expect to see an increase of the 18S-PP community biomass in the hypoxic layer of LLC. The increase of richness of the 16S-PP community observed in LLC from end of summer to end of autumn (Figure SIII.5C) was the result of this hypoxic event allowing the development of two drastically different 16S-PP communities in the water column.

Differentially abundant community

ASVs of the 18S-NP community that were differentially abundant showed a diverse taxonomy. The presence of ASVs affiliated with Suessiaceae and Cryptomonadaceae, which were more abundant in LLC, may be related to their mixotrophic behavior and the presence of the hypoxic event (Bielewicz et al. 2011; Patidar et al. 2014; Jang et al. 2020). Suessiaceae ASVs were already detected in the unique 18S-NP community in LLC and had a maximum relative abundance overlapping the hypoxic event. ASVs more abundant in the ULC in winter all belonged to the and *Lepidodinium* genus of Gymnodiaceae. This has already been observed in the unique microbial plankton community of ULC and can be related to their mixotrophic behavior and their capability of increasing in abundance in winter (Conceição et al. 2021; Liu et al. 2021). The higher abundance of ASVs belonging to

Bacillariophyta in spring in ULC is consistent with phytoplankton counts showing a high abundance of Bacillariophyceae during the spring bloom in ULC (Limnologischer Zustand des Bodensees). The higher abundance of phototrophic ASVs belonging to Ochrophyta and Bacillariophyta in LLC in autumn could be due to this higher orthophosphate concentration measured in LLC compared to ULC (Limnologischer Zustand des Bodensees). ASVs belonging to Katablepharidaceae and Cryptomonadaceae of the Cryptophyta represented 56% of the differentially abundant ASVs in the 18S-PP community and were in majority more abundant in LLC. Both families showed heterotrophic or mixotrophic behavior (Novarino 2003; Grujcic et al. 2018) and low nutrient concentration seems to trigger the mixotrophic behavior for Cryptophyta (Bielewicz et al. 2011). The higher abundance of Cryptophyta in LLC in summer and autumn, when the hypoxic event occurs, may also be the result of Cryptophyta being able to dominate the heavy metal hypoxic environment (Patidar et al. 2014). Cyanobacteria, like *Microcystis*, *Planktothrix* and *Synechoccocus*, can produce allelopathic compounds that may inhibit the growth of other microbial plankton species (Śliwińska-Wilczewska et al. 2019; Bubak et al. 2020). The total relative abundance of cyanobacteria ASVs was approximately two-fold higher in ULC than in LLC (Figure SIII.10). It is tempting to speculate if this may have led to an allelopathic compound concentration that controlled the development of Cryptophyta species in ULC.

Among the differentially abundant taxa of the 16S-PP community, Proteobacteria ASVs were the most represented, with two clusters comprising six families, one in ULC and one in LLC, and each in winter. The families of the ULC cluster belonged mostly to Alphaproteobacteria, while the LLC cluster was composed of families belonging mostly to Gammaproteobacteria. A study comparing growth condition (Pinhassi and Berman 2003a) showed that Alphaproteobacteria performed better in oligotrophic conditions while Gammaproteobacteria outperformed them at higher nutrient concentrations. Nutrient regeneration in epilimnion during the winter vertical mixing is much stronger in LLC than ULC (Limnologischer Zustand des Bodensees). Hence, LLC conditions may therefore be more favorable to Gammaproteobacteria while ULC conditions favor Alphaproteobacteria species. Furthermore, the vertical mixing especially in ULC could also bring up organisms of the meta- and hypolimnion into our sample collected from the epilimnion (0-20m

integrated sample), such as Nitrosomonadaceae observed in the meta and hypolimnion of our ULC sampling site (Herber et al. 2020). Legionellaceae, more abundant in ULC, are protozoan parasites which may indicate a prey specificity coupled with favorable conditions in winter (Moffat and Tompkins 1992). The cluster of five Bacteroidota families more abundant in LLC during winter may be the result of the higher nutrient concentration of LLC during winter and the copiotroph behavior that some Bacteoroidota can display (Li et al. 2022). Interestingly, we found again Methylophylaceae and Methylomonadaceae ASVs with a higher abundance in LLC during autumn. The hypoxic event in LLC mentioned above, is most likely the reason for this difference, providing a more favorable environment for the development of species of these two families.

Analysis of the differentially abundant microbial plankton between ULC and LLC confirmed and expanded on the observation made with the unique ASV communities. A high number of differentially abundant ASVs of the 18S-NP and 18S-PP communities were affiliated with taxonomic groups known for their mixotrophic and heterotrophic lifestyle, which may indicate a central role of these phenotypes in the species selection between ULC and LLC. However, contrary to the temporal dynamics of the unique 18S-NP and 18S-PP communities, where winter and summer-to-autumn showed the most differences, the number of taxa with relative abundance differences between ULC and LLC was quite stable by season. The 16S-PP community showed a high number of differentially abundant taxa in winter, which is the period when the vertical mixing and nutrient regeneration is taking place. The difference in depth and nutrient regeneration seems to be the main driver of the relative abundance difference between ULC and LLC. However, in autumn, when the hypoxic event is taking place in LLC, this didn´t transcribe in a lot of differentially abundant ASVs.

Conclusions

While the diversity index presented an epilimnion temporal biodiversity dynamic similar between ULC and LLC, the taxonomic analysis presented an interesting pattern with non-neglectable proportion of the community specific to ULC and LLC or more abundant. The presence of a hypoxic layer in LLC from mid-summer to mid-autumn allows microaerophilic to anaerobic bacteria as well as mixotrophic and heterotrophic organisms to thrive. ULC is composed of a hypolimnion bacterioplankton community that increase in abundance during winter when the surface water layer is regenerated in nutrients by the water mixing. Similarly to LLC, the microbial plankton community specific to ULC comprised a lot of organisms capable of mixotrophy and heterotrophy. These two mechanisms seem to be at the centre of the microbial plankton capability to live in changing environment such as the hypoxic environment or winter that include the environmental changes of the vertical mixing. However, the presence of a high number of organisms potentially capable of mixotrophy and heterotrophy specific to ULC or LLC may also be the indicator of a predator-prey interaction. The 16S-PP community difference between ULC and LLC seems to be mainly determined by the environmental factor as we could observe the presence of hypoxic adapted bacteria in LLC, while oxygenated hypolimnion bacteria were more abundant in ULC. However, biotic interactions may also shape the 16S-PP community difference with the potential predator-prey link of the unique 18S-NP and 18S-PP community. A further study is in preparation using the ULC samples from this dataset to observe the temporal co-occurrence of microbial plankton ASVs to hopefully detect the presence of such a predator-prey system as well as any other mutualism or even parasitism interactions.

Funding: This research was funded by DFG Research Training Group R3 - Responses to biotic and abiotic Changes, Resilience and Reversibility of Lake Ecosystems (GRK 2272).

Data Availability Statement: The Illumina sequence data presented in this study will be available online upon publication.

Acknowledgments: We would like to highlight the excellent support received from Alfred Sulger, Julia Schmidt, Angelika Seifried, Pia Mahler and Josef Halder at the Limnological Institute of University of Konstanz for managing the ship cruises. Further, we would like to thank all members of the RTG-R3, especially Tina Romer, and members of the AG Schleheck for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Chapter IV
Seasonal dynamics of the microbial plankton community is determined through both environmental factors and biotic interactions

Corentin Fournier[1], David Schleheck[1] and Laura Epp[2]

[1] Microbial Ecology and Limnic Microbiology, Limnological Institute, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

[2] Environmental Genomics, Limnological Institute, University of Konstanz, D-78457 Konstanz, Germany

corentin.fournier@uni-konstanz.de, david.schleheck@uni-konstanz.de, laura.epp@uni-konstanz.de

Manuscript in preparation for submission

Target journal: Environmental Microbiology

Keyword: freshwater epilimnion, microbial plankton, interactions, HMSC, Bayesian inference

Abstract

Understanding the abiotic and biotic factors affecting lake freshwater microbial plankton community composition and dynamics is the first step to understanding its functioning and, by extent, how the ecosystem works. Our study focused on the temporal analysis of Lake Constance microbial plankton dynamic, the relation to environmental factors and the biotic dynamic of the community. Samples of the lake epilimnion were taken routinely for a year and the microbial plankton community was studied by amplicon sequencing with the temporal dynamics modelled using a joint species distribution modelling approach. The model variance was mostly explained by the temporal biotic interactions, but the vertical mixing still impacted approximately a third of the microbial plankton community. Except Actinobacteria that was almost entirely affected, the taxonomy of the affected ASVs was highly diverse. The ASVs association matrices revealed a huge network with 91% of the ASVs interacting. Diverse functional capability, motility and mixotrophic behaviour seemed to be two important components allowing to overcome the environmental changes brought by the vertical mixing. The phytoplankton community holds the central place in the community structure, most of the other members relying on it for their living. Alternative trophic pathway like the mycoloop also showed a prominent role.

Introduction

Lake and aquatic communities act as sentinels of environmental changes (Adrian et al. 2009; Aránguiz-Acuña et al. 2020), which have increased in the past century. Among these are lake eutrophication due to an intensification and generalization of chemical product use in agriculture and in detergents, leading to an accumulation of nutrients in the watershed and river to be transported to the lake (Straile and Geller 1998; Finger et al. 2013; Bernát et al. 2020). Alien species, introduced in the environment via human translocation or deliberate introduction, invading freshwater ecosystems and outcompeting native species is another current environmental change that leads to decrease of biodiversity (Bax et al. 2003; Strayer 2010; Baer et al. 2022). Anthropogenic activities are primarily responsible for these disturbances, and the intensity and occurrence have been increasing with time (Woolway et al. 2022). At the same time, human societies existentially depend on freshwater ecosystems. Understanding current dynamics of lake ecosystems and their functioning is thus of immediate importance to foresee responses and resilience to ongoing disturbances and safeguard our freshwater sources.

Species interactions play an important role in the functioning and the equilibrium of the ecosystems. Primary producers, i.e. phototrophic organisms, through their metabolisms, produce organic matter that is shed to the environment. These products act as substrate for heterotrophic microorganisms, which integrate the carbon in their biomass and are then eaten by bacterial grazers (Ducklow et al. 1986). This trophic pathway, called microbial loop, makes carbon available for higher trophic levels by an alternative pathway to the classical plankton food chain (Fenchel 2008). Likewise, biogeochemical cycles, central in the ecosystem's equilibrium and resilience, are primarily run by a multitude of microorganisms (Pierrou 1976; Jetten 2008; Madsen 2011), resulting in functional networks, which can be studied by focusing on functional rather than taxonomic diversity (Feist et al. 2009; Avila-Jimenez et al. 2020). Species associations not only play a role at the ecosystems level but also shape the structure of the microbial plankton community. For example, lake algal blooms are followed by an increase of saprophytic and parasitic microorganisms (Zhang et al. 2018), while allelopathic compounds produced by cyanobacteria can act as growth inhibitors of algae (Śliwińska-Wilczewska et al. 2019). Early theoretical studies have hypothesized

that complex networks are likely to resist more to disturbance (MacArthur 1955). Yet, mathematical models and recent theoretical work come to an opposite conclusion (Thébault and Fontaine 2010) and not enough environmental and experimental studies exist to support either point of view.

In network analysis, direct associations like mutualisms and commensalism, in which one or both parties gain, will lead to positive non-random co-occurrence. Direct associations including competition, predation or parasitism, in which the interaction is negative for one actor and positive for the other, will lead to negative non-random co-occurrence (Ovaskainen and Abrego 2020). Microorganisms responding similarly to change in the environment are likely to spatially and temporally co-exist due their overlap in ecological niches and present equivalent non-random patterns of co-occurrence (Coutinho et al. 2015). Unfortunately, deciphering interactions coming from ecological niche overlap to direct species to species relationships from a network built with environmental samples is challenging. Joint Species Distribution Modeling (JSDM) allows for calculation of species relation patterns based on environmental factors, followed by a latent variable approach for direct species associations based on the residual variance (Warton et al. 2015; Ovaskainen and Abrego 2020).

The Hierarchical Modelling of Species Communities (HMSC) is a correlative approach based on JSDM using environmental data. This approach is used in ecology for a large variety of datasets, from macroinvertebrates to bird communities (Tikhonov et al. 2020a; Hällfors et al. 2020; Chiu et al. 2020; Elo et al. 2021) to rare species prediction (Zhang et al. 2020). Within the field of microbial ecology, equivalent analyses have been recently implemented (Minard et al. 2019; Odriozola et al. 2021), but on datasets with a limited complexity. Here we use this approach to model the temporal co-occurrence patterns of microbial plankton in the epilimnion of Upper Lake Constance over the course of one year, from March 2018 to March 2019. More specifically, we 1) test, whether the existing Bayesian framework can be used on a large and complex dataset, such as that from Lake Constance, 2) investigate the temporal dynamics of microbial plankton in relation to environmental changes across the year, and 3) identify consistent associations of species that point to the existence of biotic networks. From the latter, we discuss hypotheses that can be tested in future experimental setups to obtain a more complete mechanistic understanding of the microbial plankton communities.

Materials and methods

Study sites, sampling campaign and filtration procedure

Lake Constance is a deep (maximal depth of 251 m) oligotrophic pre-Alpine Lake composed of three water bodies called *Obersee* (Upper Lake), *Untersee* (Lower Lake) and *Seerhein* (Rhine Lake). Samples were taken at a site in Lake *Überlingen* (47.7571°N 9.1273°E), a north-western arm of the main basin of the Upper Lake with a depth of 140 m. We sampled the approximate photic zone (epilimnion), by integrated sampling of the surface layer reaching from 0 - 20 m in depth. Samples were collected from a boat using an integrating water sampler (model IWS II, Hydro-Bios, Germany), from March 2018 to March 2019. Samples were taken every two weeks and stored in stainless steel barrels (19 L soda-keggs) prior to filtration. Four biological replicates were collected on each sampling occasion (n=4).

Samples were filtered directly on the boat, in order to reduce alteration of the samples. A pressure of 2 bars was applied to the barrels using a pressurized air tank. The barrels were connected to a series of filters´ holder (Swinnex 47 mm filter holder; Millipore) using PVC tubing. The pore - sizes of the filters were 180 µm (hydrophilic nylon net, 47 mm diameter; Millipore) to remove zooplankton and larger particles, 5.0 µm and a 0.2 µm polycarbonate membrane filter (Isopore, 47 mm diameter; Millipore) to collect the nanoplankton and picoplankton respectively. The Swinnex filter holders with the filters assembled, were autoclaved when being wrapped in aluminum foil and then transported onto the boat. Filtration was done following the protocol developed in our laboratory (Chapter II), when using the flowrate at the outlet of the filters as a proxy of the biomass loading of the filters and the degree of filter clogging. This protocol was developed in response to a study (Padilla et al. 2015) showing that there is a bias introduced by the filtration volume and the microbial plankton density on analysis of microbial plankton by -omics techniques (Chapter II). Each filter was transferred by using sterile forceps and rolled into 15 ml Falcon tubes for storage, with the filter side harboring the collected biomass facing to the inside of the tube. The tubes were stored on dry ice on the boat and in the laboratory at -20°C until the DNA was extracted. Only the 5.0 and 0.2 µm filters were used for downstream analysis.

Molecular work: DNA extraction, library preparation and sequencing

DNA extraction was performed on all the filters collected. The protocol was adapted from (Rusch et al. 2007) and the JGI protocol (version 3 by William S., Helene Feil and A. Copeland). 1 min of ultrasonication (Sonorex super RK 510, Bandelin, Germany) was applied, followed by 15 mins of vortex after adding 0.1- and 1-mm diameter zirconium beads. Lysozyme (2.5 mg/ml) was added and 1 hour of incubation at 37 °C in a shaking incubator was done. Sodium dodecyl sulfate (SDS) at a final concentration of 1% and proteinase K (500 µg/ml) were added followed by 2 hours of incubation at 55°C. After one hour of incubation, proteinase K was added again at the same concentration. A final step was done consisting of 10 mins incubation at 65°C with 0.07 Volume of 5M NaCl and 0.07 Volume of CTAB/NaCl (preheated at 65°C). DNA was then purified by adding 1 Volume of phenol/chloroform/isoamyl alcohol followed by 20 mins of centrifugation at 13000 rcf and 4°C. After recovering the supernatant in a new 15 ml falcon tube, 1 Volume of chloroform/isoamyl alcohol was added followed by 20 mins of centrifugation at 13000 rcf and 4°C, and the supernatant was recovered in a new 15 ml falcon tube. 0.5 µl of Glycogen and 0.7 Volume of isopropanol were added to the tube and incubated for 15 mins at -20°C. A new centrifugation was done during 25 mins at 15000 rcf and 4°C. Isopropanol was removed, and the DNA pellet was cleaned using 500 µl cold 70% ethanol. After a final centrifugation of 5 mins, ethanol was removed, and the pellet was dried. 50 µl of PCR-grade water was added in each tube. DNA concentration was measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA).

16S rRNA gene amplification was performed on the 0.2 µm DNA extract only and the 18S rRNA gene amplification was performed on both the 0.2 and the 5 µm DNA extract. A two-step PCR approach was used for the DNA library preparation with Illumina dual indexing. Primers of the first PCR contain the fragment of interest sequences as well as a universal 5´tail. Amplification of the V3-V5 hypervariable region of the 16S rRNA gene (569 bp) was performed using the 357F and 926R primers, respectively from (Schuurman et al. 2004) and (Walters et al. 2016). Amplification of the V4 hypervariable region of the 18S rRNA gene (378 bp) was done using the 573F and 951R primers (Mangot et al. 2013). Table IV.1 contains the sequences of the universal 5´tail and the primes sequences. PCR was performed with 0.02 U/ µl of Phusion High Fidelity DNA polymerase, 1X Phusion HF Buffer and 200

µM of dNTPs (New England Biolabs, USA). Extracted DNA was added at a final concentration of 0.12 ng/µl. A final concentration of 0.5 µM of each primer was used. The number of PCR cycles was 10. It comprised the initial denaturation step of 3 mins at 98°C, the denaturation of 45 seconds at 98°C, primers annealing step of 20 seconds at 62.4°C, extension step of 8 seconds at 72°C and the final extension step of 5 mins at 72°C. The reaction was done using a T100 Thermal cycler (Bio-Rad, USA). PCR products of the first PCR were used as template for the second PCR. Primers included a pair of indexes (barcode) specific to each sample and Illumina adaptors. Conditions of the PCR were equivalent to the first PCR. Number of cycles was increased to 20. PCR products of both steps were purified following the Agencourt AMPure XP protocol with a 0.9X and 0.8X volume beads and a final elution volume of 11 µl and 14 µl in the first and second purification respectively. PCR products were pooled in equivalent quantities and send to Eurofins GATC Biotech for sequencing using Illumina MiSeq 2*300 bp with the Microbiome Profiling -Indexing only- package. Two independents sequencing runs were performed for the 16S and 18S DNA library respectively.

Table IV.1: Sequence and targeted region of the primers used for the amplification of the 16S and 18S rRNA gene for barcoded amplicon sequencing library preparation.

Primer name	Sequences (5´-3´)	Targeted region	Reference
Universal 5´ forward	TCGTCGGCAGCGTCAGATGTG TATAAGAGACAG		Amplicon deep sequencing
Universal 5´ reverse	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAG		
357F	CTCCTACGGGAGGCAGCAG	16S rDNA V3-V5	Schuurman *et al*., 2004
926R	CCGYCAATTYMTTTRAGTTT		Walters *et et*., 2016
573F	CGCGGTAATTCCAGCTCCA	18S rDNA V4	Mangot *et al*., 2013
951R	TTGGYRAATGCTTTCGC		

Bioinformatics pipeline

The sequencing data provided by Eurofins GATC Biotech was paired end. Quality of the sequences was checked using FastQC. The analysis of the 16S rDNA dataset was done using only the forward reads. Reads were trimmed using Trimmomatic (Bolger et al. 2014), discarding reads with a Phred score below 3 at the start and the end of the reads, an average quality of 11 over a window of 3 bases within the reads. The 18S rDNA reads were also cropped at a maximum size of 285 to remove bases with lowest quality on the merging area. The 18S rDNA paired end reads were merged using NGmerge (Gaspar 2018) with parameters of a minimum overlap of 30 bases and an acceptable error rate of maximum 10%. Read quality was checked between each step previously described. Denoising and taxonomic affiliation were then performed on QIIME2 2020.11. Filtration of chimeras using the consensus method, denoising and dereplication of the quality reads were done using the denoise and dereplicate single-end sequences (Dada2, denoise-single) and a read learn of 2,000,000 reads for the training error model (Callahan *et al*., 2016). For the 16S rDNA database, taxonomic affiliation was determined using the classify-consensus vsearch program and the TaxAss pipeline (Rohwer et al. 2018), which used both a general database: SILVA_138 and a freshwater ecosystem-specific database: FreshTrain (Newton et al. 2011). 4 thresholds of identity were used: 80%,90%,97% and 100%. Outputs were then compared to check for consistency and to keep the deepest taxonomic affiliation while having a minimum of unassigned sequences. The 18S rDNA taxonomic affiliation was performed using the classify-consensus vsearch program and the reference database PR2 (Protist Ribosomal database) (Guillou et al. 2013).

Biostatistics

The dataset was analysed using the Hierarchical Modelling of Species Communities (HMSC) (Ovaskainen et al. 2017b). This model belongs to the Joint Species Distribution Modelling approach (Warton et al. 2015). HMSC constructs a hierarchical model based on the generalized linear model (GLM) pipeline and Bayesian inference (Ovaskainen et al. 2017a). A minimum threshold of 10000 reads per sample was used to include the samples in the analyses. This threshold was defined as the minimal number of reads to reach the plateau phase of the rarefaction curve, which indicates the minimal amount of reads to capture all ASVs sequenced.

The read numbers of 9 samples were below the threshold and they were therefore discarded from the analysis. The analyses were thus based on 69 samples from 26 sampling dates, each represented by two or three replicates. ASVs, whose relative abundance was below 0.5% in at least 20% of the samples were removed from the analysis.

Co-occurrence associations between microbial plankton ASVs were analysed by including a temporal random effect at the sampling date level (Ovaskainen et al. 2017b). These were modelled against fixed explanatory variables, consisting of environmental variables and the sequencing depth. As environmental variables, we included the seasons and whether the water column is mixed or thermally stratified (denoted as vertical mixing). The sequencing depth, as log transformed total number of reads per sample, was included to account for sequencing depth differences between samples. Two types of models were run: 1) FULL models, including all explanatory variables, i.e. the environmental variable and the sequencing depth, and 2) NULL models, including only the sequencing depth. ASV associations observed in the null models combine co-occurrence due to shared ecological niches as well as direct associations that are independent of the environmental variables. Associations observed in the full models are calculated after the effects attributed to the environmental variables are taken into account. The resulting associations are more likely to be direct interactions between species, or alternatively ASVs, i.e. associations that occur independently of the included environmental variables and are thus likely direct associations between species. A 0.95 threshold of posterior probability support is used for the environmental variable impact on ASVs as well as ASV associations.

As the dataset was zero inflated, even after the filtration, calculations were performed using a hurdle mode. This means that two separate models were run: 1) a probit model based on presence – absence of ASVs, or occurrence fitted to a binomial distribution and 2) a lognormal model for the relative abundance of ASVs conditional on presence. In effect, a total of four models were created, i.e. a FULL and a NULL model using both the probit and the lognormal model, under the assumption of a default prior distribution.

The explanatory and predictive powers of the models were calculated using area under the curve (AUC) for the probit model and the standard R^2 for the lognormal

models following the protocol of Odriozola et al., 2021 (Odriozola et al. 2021). Unconditional and conditional predictive power was measured using a 2-fold cross-validation approach across the sampling dates. A two-fold cross validation split the datasets in two independent sets, one was used as the training set while the other was used for testing and validation. Unconditional predictive power was tested by setting up the two independently of the ASV associations and only using the environmental variables. Conditional predictive power will consider the ASV associations in addition to the environmental variables. We examined how the predictive power can be improved when considering the link between Eukaryota and Bacteria, i.e. how well the model can predict bacteria when the Eukaryota community is known and vice versa.

Models were fitted using the R-package HMSC (Tikhonov et al. 2020a). For each model, the posterior distribution was sampled using 4 Markov Chain Monte Carlo (MCMC) chains. 150.000 iterations were calculated per chain, in which the 50.000 first iterations were discarded. To minimise autocorrelation, iterations were thinned to 100 giving a final 1000 posterior samples per chain and 4000 per models. MCMC convergence was measured by calculating the effective number of samples, measuring the autocorrelation of samples, and the potential scale reduction factor, measuring the consistency of the results between the model chains. Convergence was measured for the parameters accounting for the species niches (β), the species associations (Ω) and the residual covariance of species niches (V).

Results

General overview of the data and analyses

Our dataset covered a time period of a full year with sampling of biological triplicates every two weeks, resulting in 78 samples. After the filtration of the samples that did not contain enough reads, 69 samples remained. No full triplicates from a single sampling date were removed so our dataset did not have a time frame gap (see supplementary file Table SIV.1). Two replicates were removed for the sampling date 27.04.2018, making this date represented by a single replicate (Table SIV.1). The total number of raw reads, not discarded after the dada2 steps and remaining after filtering the low abundant ASVs is showed in Table SIV.2. The minimal number of reads in one sample was 28540, which was enough to detect the majority of the ASVs in the samples (Figure SIV.1). As two independent sequencing runs were performed on each sample, we explored the variability between sequencing runs before merging them into one dataset (Figure SIV.2 to SIV.4) The results did not show divergence of samples between the sequencing runs; hence they could be merged confidently. Sequencing data yielded a high number of detected ASVs, with a majority being low abundant and thus not holding enough information to give a satisfactory fit in the ASVs interaction models (Ovaskainen and Abrego 2020). Furthermore, as JSDM models demand high computing power, running the model with a dataset that is too large would require excessive computational power. Hence a very stringent filtration of the low abundant ASVs was applied. After filtration, 377 bacterioplankton and 670 microeukaryote ASVs, resulting in a total of 1047 ASVs, remained for the analysis. The 16S rDNA reverse reads were not of sufficient quality to allow for a correct merging with the forward reads. We therefore had to do the analysis with the 300 bp forward reads only. Although shorter, the amplicons still comprise enough taxonomic information, by covering the complete V3-hypervariable region and a large proportion of the V4-hypervariable region. The taxonomic database used for the microeukaryotes gave limited information for the deeper taxonomic ranks, hence we chose to stay at high taxonomic ranks, which delivered solid results.

JSDM and Bayesian framework in microbial ecology

The models had a greater explanatory power for the probit occurrence model (0.92) than the lognormal abundance models (0.76). The convergence of the beta and omega parameters, respectively measuring the species responses to environmental covariates and ASV-to-ASV associations were assessed using the sample size and potential scale reduction factor. The potential scale reduction factor results all showed values close to one, indicating that the four independent Monte Carlo Markov chains within each model showed similar results (Figure SIV.5). The sample size, which measured the autocorrelation level between the chain iterations, is expected to be close to 4000. The presence of autocorrelation resulted in a lower effective sample size that led to an increase of the uncertainty in the posterior probability support, decreasing the number of detected ASVS associations above a certain posterior probability support. The beta parameter results were close to the expected values, indicating low or no autocorrelation. The omega parameter presented an effective sample size around 2000, indicating the presence of autocorrelation (Figure SIV.6).

Table IV.2: Variance proportion explained by each model factors

	Bacterioplankton		**Microeukaryote**	
	Probit	**Lognormal**	**Probit**	**Lognormal**
Random factor (time)	0.68±0.15	0.82±0.08	0.68±0.17	0.84±0.08
Vertical mixing	0.13±0.07	0.09±0.05	0.11±0.07	0.07±0.05
Seasons	0.10±0.06	0.05±0.04	0.12±0.08	0.05±0.04
Sequencing depth	0.09±0.08	0.03±0.02	0.1±0.07	0.04±0.03

Dynamics of the of community composition across the year

The variance partitioning of community composition revealed that the date level random factor explained the majority of the variance, especially dominating in the lognormal models (Table IV.2 and Figure IV.1A-D). Both the vertical mixing and seasonality explained roughly the same proportion of variance (Table IV.2), and the presence of a large standard deviation indicated a strong variability among ASVs (Table IV.2 and Figure SIV.7). The sequencing depth variable explained the lowest variance proportion and was higher in the probit occurrence model than in the lognormal abundance model. Overall, the proportion of variance explained by the model variable showed similar patterns between bacterioplankton and microeukaryotes for both the probit occurrence and the lognormal abundance models (Figure IV.1A-D).

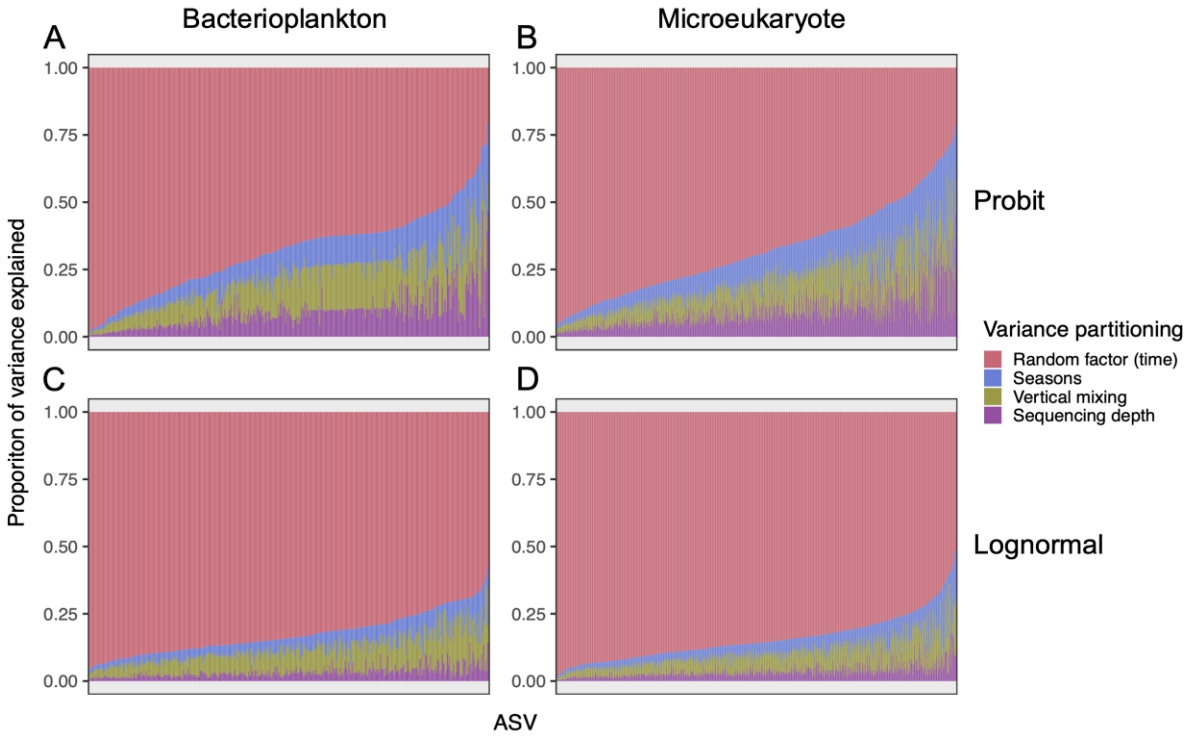

Figure IV.1: Variance partitioning of bacterioplankton (A & C) and microeukaryote (B & D) community composition of the probit (A & B) and the lognormal models (C & D), ordered by descending random factor (time) proportion. The bar plots represent the proportion of variance explained by each variable (y-axis) for each ASV (x-axis). Fixed factors comprise the sequencing depth, seasons and the vertical mixing. The random factor time determines ASV associations. ASVs are in descending order based on the proportion of variance explained by the random factor time in Figure IV.1 and vertical mixing descending order in Figure SIV.7.

Although the fixed factors seemed to play a minor role compared to the random factor time in the model variance explanation, a high number of ASVs presented a positive or negative interaction with the vertical mixing of the water column, either in their occurrence (Figure IV.2A) or their relative abundance (Figure IV.2B). Only interactions between ASVs and the fixed factor having a posterior probability support of at least 0.95 were kept for the analysis. The probit occurrence model showed 185 ASVs affected by the vertical mixing, 160 negatively and 25 positively (Figure IV.2A). The lognormal abundance model showed 237 affected ASVs, 228 negatively and 9 ASVs positively affected by the vertical mixing (Figure IV.2B). In total, 353 and 31 ASVs were respectively affected negatively and positively by the vertical mixing. They represented 36.7% of the total ASVs number, showing the large impact of the vertical mixing on the microbial plankton community.

All positively affected ASVS, in both models, were taxonomically affiliated to microeukaryotes. In the 160 negatively affected ASVs of the occurrence model, 103 were affiliated to microeukaryotes and 57 to bacteria. The 228 negatively affected ASVs of the abundance model were divided into 99 affiliated to microeukaryotes and 129 to bacteria. For a few exceptions, no clear cluster of taxonomically close affected ASVs could be observed (Figure IV.2AB). The major exception was Actinobacteria, which had 78% of its ASVs negatively impacted by the vertical mixing. Proteobacteria and Bacteroidota were the two main bacterioplankton phyla in terms of ASV richness, with 128 and 112 respectively (Table SIV.3). Of these 29% and 34%, respectively, were negatively impacted by the vertical mixing. Impacted ASVs of Proteobacteria were evenly distributed between Alpha- and Gammaproteobacteria. All ASVs of lineage alfVIII (Alphaproteobacteria) were negatively affected by the vertical mixing. The family Commamonadaceae and the lineage betI represented 60% of the impacted Gammaproteobacteria ASVs but only 26% and 31% of the Commamonadaceae and *betI* total richness. 74% of the impacted ASVs of Bacteroidea belong to the two orders Flavobacteriales and Chitinophagales, representing 29% and 46% of the total richness of Flavobacteriales and Chitinophagales.

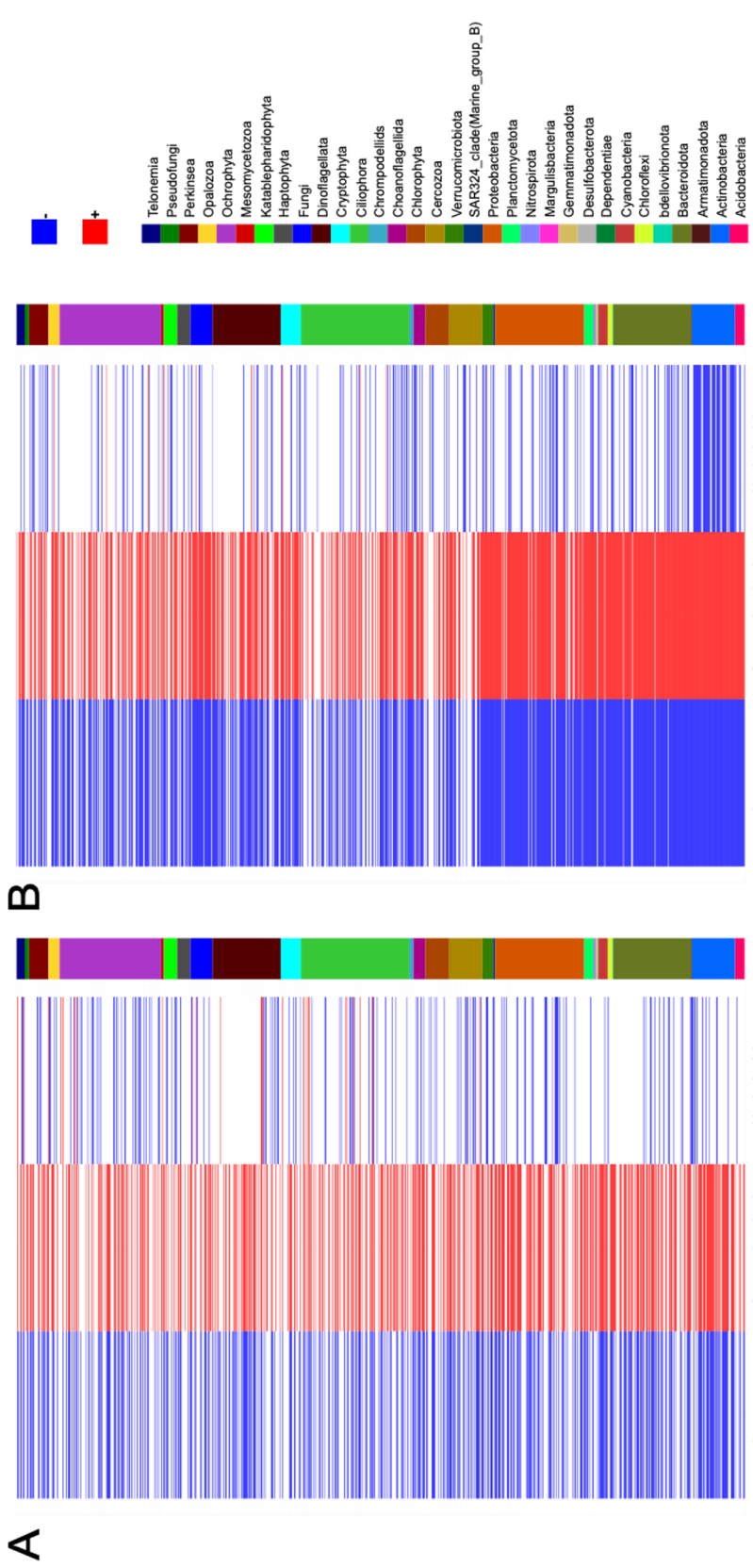

Figure IV.2: Interaction between ASVs and the fixed factors sequencing depth and vertical mixing in the probit occurrence model (A) and lognormal abundance model (B). The red color indicates a positive interaction of the ASVs with the environmental factors and the blue color indicate a negative interaction. Interaction with a posterior probability support above 0.95 were kept. x-axis represents the fixed factor "sequencing depth" and "vertical mixing". y-axis represents the ASVs ordered by taxonomic group. The colored barplot along the y axis represent the phyla to which the ASV is affiliated.

Microeukaryotes impacted by the vertical mixing were also taxonomically highly diverse, and presented differences in response to environmental change, with negative and positive associations. The main phyla Ciliophora, Ochrophyta, Dinoflagellata and Cercozoa represented respectively 25%, 27%, 16% and 35% of ASVs negatively impacted by the vertical mixing. Proportionally, Dinoflagellata had a low number of ASVs impacted negatively and most of them belong to the order Prorocentrales (50% of the order total richness). The majority of the ASVs interacting positively with the vertical mixing belong to Ciliophora (32% of the positive ASVs), Ochrophyta (19%) and Dinoflagellata (16%). Of the six Ochrophyta ASVs positively affected by the vertical mixing, four were Bacillariophyta (diatoms) and two Chrysophyceae. Chrysophyceae represented 67.5% of the negatively affected ASVs of Ochrophyta (29% of the class richness) while also having ASVs positively affected. Other taxonomic groups presented a high proportion of negatively affected ASVs by the vertical mixing like 50% of Perkinsea, 59% of Choanoflagellida and 75% of the fungi Chytridiomycota ASVs.

Even though the models´ variance was in majority explained by the random factor time, a high proportion of the community was impacted by the vertical mixing fixed factor. While most of the ASVs of Actinobacteria were negatively impacted, the models also revealed a taxonomically diverse microbial plankton community being impacted by the vertical mixing, with a large domination of negative interactions and only few taxonomically specific ASVs showing positive associations.

Co-occurrence networks of ASVs

Only associations with a probability support of 0.95 were kept, representing associations that are likely not by chance and therefore ecologically relevant, regardless of the strength of the associations (Odriozola et al. 2021). A positive association indicates ASV co-occurrence, while a negative association represents ASV co-exclusion. ASV associations derived from the NULL models (no environmental variables as fixed factor) and called raw associations, likely represent ecological niche overlap (positive associations) or exclusion (negative associations). ASV associations observed in the FULL models, including environmental variables, are called residual associations and likely represent direct interactions, independent of environmental variables, or are related to ecological niche overlap/exclusion from an environmental variable not included in the models.

The NULL and FULL models revealed a high number of ASVs associations (Figure IV.3A-D). A certain number of associations were present in both models, indicating ASV co-occurrences resulting from both an ecological niche overlap (NULL) and direct interactions or unknown ecological niches (FULL). Although the NULL models presented a total number of raw associations higher than the residual associations of the FULL models, a higher proportion of the residual associations had a strong correlation value, above 0.7, compared to the raw associations (Table IV.3). This indicates a denser network of highly connected ASVs in the FULL models (Figure IV.4). Furthermore, the lower number of associations found in the FULL models did not mean that associations only disappear from the NULL to the FULL models. A certain proportion of the associations were unique to the NULL models while other associations were detected only in the FULL models. This indicates ASV co-occurrences resulting only from direct interactions of the organisms (Figure IV.3A *versus* B and C *versus* D and Table IV.3). The proportion of negative and positive associations did not change substantially between models with on average of 43% of negative and 57% of positive associations (Table SIV.4).

Table IV.3: number of ASV associations with a posterior probability support of 0.95. The correlation strength indicates the number of associations with correlation values above the filtration threshold. The last row indicates the number of associations unique to each model.

Correlation strength	NULL		FULL	
	Probit	Lognormal	Probit	Lognormal
Total	72884	25271	55687	17256
Above 0.6	42216	2581	51636	3742
Above 0.7	16219	718	30672	1002
Above 0.8	3204	202	8854	267
Above 0.9	134	26	740	58
Unique	34932	15684	17720	7685

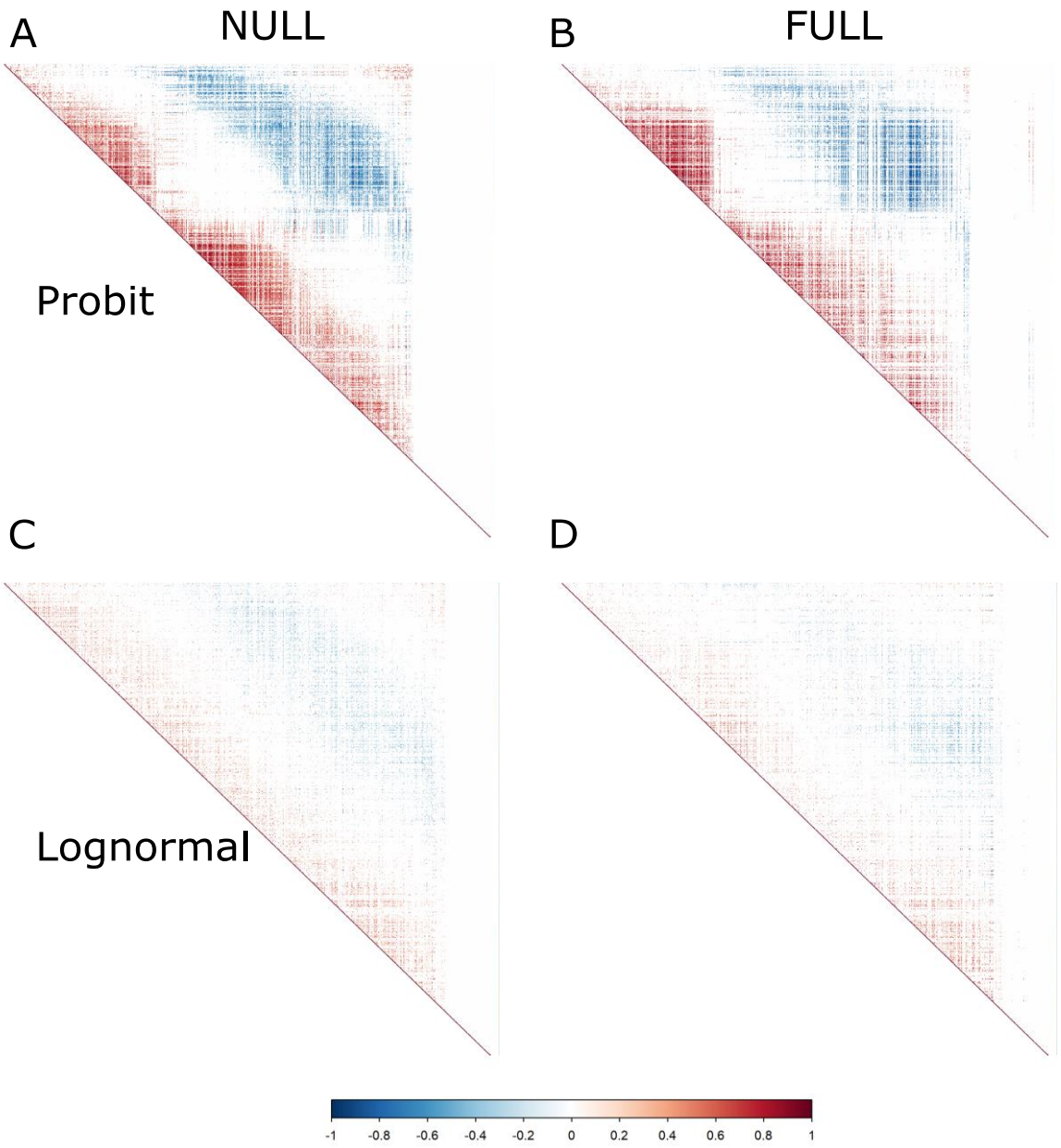

Figure IV.3: ASV associations based on the temporal random effect of the probit (A &B) and the lognormal models (C & D). A & C represent the raw associations of the NULL models with the sequencing depth as fixed factor. B & D represent the residual associations of the FULL models with all environmental variables as fixed factors: seasons, vertical mixing and sequencing depth. The red color indicates a positive interaction of the ASVs with the environmental factors and the blue color indicates a negative interaction. Interactions shown have a posterior probability support above 0.95. For a clear comparison between the NULL and FULL models, ASVs are ordered using the angular order (AOE) in the NULL models and this order is kept for the FULL model. All 1047 ASVS including bacterioplankton and microeukaryote are put in the same graphics.

Figure IV.4: ASV associations with correlation above 0.9 built on the temporal random effect of the probit (A) and the lognormal (B) FULL models. The red color indicates a positive interaction of the ASVs with the environmental factors and the blue color indicates a negative interaction. Interaction with a posterior probability support above 0.95. Taxonomy indicate ASV number, kingdom and phylum affiliation

The association matrices did not show a clear taxonomical cluster of co-occurrences or co-exclusion. All 32 phyla of our dataset presented ASV associations, even the lowest in abundance or richness (Table SIV.3). 84% and 91% of the ASVs, respectively in the probit occurrence and lognormal abundance models, presented at least one association. The phyla Ciliophora, Ochrophyta, Dinoflagellata, Proteobacteria and Bacteroidota dominated the associations dynamics by representing 59.2% of the total number of associations in the models (average calculated from the four models). These 5 phyla represented a high proportion of the community (40% of the relative abundance) and 61% of the total number of ASVs (Table SIV.3). Linear regression showed a direct relationship between the number of associations and phylum richness in each model (adjusted R-squared > 0.9), while no relationship could be established with the total relative abundance of the phylum.

An intricate network of associations including both co-occurrence and co-exclusion between phytoplankton phyla was visible in the FULL models. Phytoplankton referred to the photosynthetic microbial plankton community, represented in this dataset by the phyla Ochrophyta, Haptophyta, Cryptophyta, Chlorophyta, Dinoflagellata and Cyanobacteria. They represented 30.9% of the total ASV associations observed in the models and 32% of the total number of ASVs, while belonging to only 6 of the 32 phyla. Their positive interactions were about twice as numerous as their negative ones (Table SIV.6 and Table SIV.7). The only exception was the cyanobacteria, for which co-exclusion and co-occurrence were either equal in number or for which more cases of co-exclusion with other phytoplankton phyla were detected (Table SIV.6 and Table SIV.7). Within these taxa, the Dinophyceae, containing both photosynthetic and heterotrophic organisms, and the golden-brown algae Chrysophyceae were the most important phytoplankton classes, each accounting for 9% of the total number of associations in the FULL models.

Phytoplankton ASVs not only interacted with each other, but also with other taxonomic groups. 21% of the total associations of the 6 phytoplankton phyla were formed by Bacteroidota and Proteobacteria, making them the groups with most bacterioplankton-phytoplankton connections. Four ASVs of our dataset were assigned to the order Rhodobacterales, Four to Cytophagales, while 10 ASVs were assigned to the family Flavobacteriaceae and 18 to the lineage bacII. Table SIV.5 summarises the number of associations between these 4 bacterial taxa and phytoplankton ASVs. For

each of these taxa, the majority of bacterial ASVs were found to co-occur and co-exclude with one to several phytoplankton ASVs. Actinobacteria showed a high number of interactions with both other bacterioplankton and phytoplankton. Positive associations between Actinobacteria ASVs represented 29% of the total positive associations of Actinobacteria, making this phylum proportionally the most internally connected of the study. The FULL models showed close direct interaction of two Nitrospirota ASVs with four ASVs belonging to Nitrosomonadaceae family and a single methanotroph Methylococcales. Four Gammaproteobacteria ASVs belonging to the TRA3-20 (Burkholderiales) all showed a consistent positive association with the 2 Nitrospirota ASVs with correlation value between 0.7 and 0.93.

Fungi are important members of the microbial plankton community. In this study, 70% of the fungi were assigned to the Chytridiomycota. Interactions of Chytridiomycota in the FULL models were dominated by co-occurrences with ASVs belonging to the phytoplankton phyla Chlorophyta, Ochrophyta and Dinoflagellata (Table SIV.6 and Table SIV.7), representing 28.2% of the total direct interactions of Chytridiomycota. Co-occurrences were two to three times more detectable than co-exclusions (Table SIV.6 and Table SIV.7). 14.3% of the interactions of Chytridiomycota were with Ciliophora ASVs, largely consisting of co-occurrence. Finally, Proteobacteria and Bacteroidota ASVs also showed some association with Chytridiomycota (21.4% of interactions), with a more balanced co-occurrence and co-exclusion dynamic. Interestingly, the lognormal FULL models showed three Chytridiomycota ASVs that interacted with a specific group of Actinobacteria ASVs, almost exclusively belonging to lineages acI and acIV. Two showed only positive associations and one only negative. Such a consistency in the strength of the relationships clearly points to a rather intimate interaction between these ASVs.

Discussion

JSDM and Bayesian framework in microbial ecology

The first objective of this study was the implementation of JSDM models with Bayesian inference in microbial community analysis, more particularly for large sequencing data. The diagnosis of the MCMC convergence, made using the effective sample size and potential scale reduction factor, showed satisfactory results, indicating that the inference derived from the model can be trusted (Ovaskainen and Abrego 2020). Yet, the lower-than-expected effective sample size of the omega parameter, measuring the ASVs associations, showed the presence of autocorrelation (Figure S3). Satisfactory convergence of parameters is very hard to achieve for high dimensional data, such as DNA sequencing data (Rajaratnam and Sparks 2015). The presence of autocorrelation lead to an increase of the uncertainty in the posterior probability support. The uncertainty reduced the number of detected ASV associations but it did not invalidate the ASV associations having a posterior probability support above our threshold. Increasing the thinning of the chain is a possibility to improve parameters convergence but it also greatly increases the run time of the MCMC. The compositional data are not explicitly modelled with a multinomial distribution in the JSDM approach we used. Not taking into account the compositional nature of the data can result in an inflation of negative ASV associations (Gloor et al. 2017). We did not find an inflation of negative associations – indicating that the results are robust, confirming previously observed similar results (Odriozola et al. 2021). The development of methods for satisfactory convergence of high dimensional data and the integration of multinomial distribution for compositional data in JSDM and bayesian inference are a current active field of study (F. Dormann et al. 2007; Warton et al. 2015; Doser et al. 2022).

Temporal environmental factors shaping the microbial plankton community

The variance partitioning showed a very high proportion of variance explained by the random factor time, consistent with previous studies (Minard et al. 2019; Odriozola et al. 2021). Although the random factor time models covariation among ASVs which can´t be related to environmental variables (fixed factors), it also includes associations resulting from environmental variables that are not integrated into the fixed effect (Warton et al. 2015; Ovaskainen and Abrego 2020). Direct associations

could represent a lower amount of the variance if more environmental variables were included in the models. Higher sequencing depth increases the probability of detecting low abundance organisms while the relative abundance of detected taxa remains stable regardless of sequencing depth (Zaheer et al. 2018; Pereira-Marques et al. 2019). This explains the higher proportion of variance explained by the sequencing depth in the probit occurrence models. For ASVs whose variance proportion was best explained by seasons and vertical mixing, sequencing depth also often explained a high proportion of the variance (Figure IV.1AB and Figure SIV.1AB). This suggests that the most seasonal ASVs (in terms of season and vertical mixing) either belong to the low abundance community and require a high sequencing depth to be detected, or appear only during a specific period/season and remain below detection level for the rest of the year.

Water vertical mixing corresponds to several sudden changes in environmental parameters, including physical water turbulence, low irradiance, nutrient regeneration, reoxygenation and low water temperature (Diehl et al. 2002; Huisman et al. 2004; Barton et al. 2014; Limnologischer Zustand des Bodensees). The cold period also corresponds to low autotrophic and heterotrophic community activity. The high taxonomic diversity observed in the affected ASVs is probably also due to the diversity of metabolic and physiological capabilities. Species adapted to warm and oligotrophic environments, such as Actinobacteria, or bacterial species dependent on phytoplankton exudates for feeding, such as in the lineage betI (Proteobacteria), Flavobacteriales and Chitinophagales (Bacteroidota), are negatively associated with vertical mixing, as it alters conditions from their optimum (Šimek et al. 2006; Haukka et al. 2006; Zeder et al. 2009; Šimek et al. 2011; Woodhouse et al. 2016). Despite having different optimal nutrient concentrations, Alpha- and Gammaproteobacteria were equally affected by vertical mixing, indicating that nutrient concentration is not the only factor affecting these bacteria (Pinhassi and Berman 2003b). Only lineage alfVIII (Alphaproteobacteria) had all its ASVs affected. Little is known about this lineage and its classification is still under debate (Garcia et al. 2013; Rubbens et al. 2019), but like other Proteobacteria, it relies on phytoplankton exudates to thrive (Rubbens et al. 2019; Chun et al. 2020). Proteobacteria and Bacteroidota-associated organisms show high variation in gene content and metabolic capabilities (Zavarzin et al. 1991; Bradley and Pollard 2017). This diversity and species-dependent capabilities

may be the reason why we observed different responses to vertical mixing for taxonomically close ASVs. Bottom-up selection by the viral community could also be responsible for the negative effect of vertical mixing on microbial plankton. Indeed, the physical perturbation of vertical mixing can trigger both the viral lysis activity of bacteriophages in the ocean (Blottière et al. 2017; Winter et al. 2018). Such increased activity could also occur in a freshwater lake. No taxonomy was reported in the studies, but the host range of bacteriophages could lead to bottom-up pressure that differentially affects bacterioplankton species during vertical mixing.

Environmental pressure is likely to play a role in the negative effects of vertical mixing on microeukaryotes. Parasitic taxa such as Perkinsea and fungi Chytridiomycota showed a high proportion of negatively affected ASVs, which can be linked to the low microbial plankton biomass of this period. Prorocentrales, a spring bloom-forming phytoplankton that prefers warm waters (Heil et al. 2005), correspond to most of the negatively affected ASVs of Dinoflagellata. Motile and mixotrophic organisms seemed to cope with vertical mixing, probably by overcoming the disturbance and remaining in their optimum conditions or by diversifying their food input. Such observation were made for Dinoflagellata that are often found aggregating at depths corresponding to their optimum growth conditions using chemotaxis, phototaxis and geotaxis (Cullen and Macintyre 1998). Mixotrophic behaviour has been observed in Dinoflagellata and Ochrophyta (Olrik 1998; Gribble and Anderson 2006), and low irradiance has been shown to trigger mixotrophic behaviour in Dinoflagellata species, allowing them to dominate the winter community (Millette et al. 2017). Among the ASVs positively affected by vertical mixing, the majority belong to taxa capable of motility, such as Ciliophora and Dinoflagellata (Buskey and Stoecker 1989; Cullen and Macintyre 1998; Smayda 2010), and mixotrophic or heterotrophic behaviour taxa, such as Ciliophora, Dinoflagellata and Ochrophyta. Observations of the phytoplankton community have shown the presence of cell size selection during vertical mixing. Regeneration of nutrients allows larger phytoplankton species with low affinity for nutrients to outcompete smaller phytoplankton species, as observed for diatoms (Bacillariophyta) (Winder and Hunter 2008; Barton et al. 2014; Fraisse et al. 2015). Although we can’t infer size from taxonomy, cell size selection could play a role for ASVs to be positively or negatively affected within Bacillariophyta. Likewise, species-dependent response to environmental parameters (Munir et al. 2013), differences in

motility ability or feeding behaviour within the deeper taxonomic levels can be the reason behind the difference of response to the vertical mixing for closely related ASVs.

ASV co-occurrence

ASV associations observed in our model can be attributed to different aspects of the microbial plankton community, from associations representing large trophic networks including mutualisms, predation, parasitism to the smallest and more intimate associations of ASVs performing a specific step of a biogeochemical cycles. The first striking observation was the high interconnection of phytoplankton ASVs. Photosynthetic phytoplankton contribute to carbon fixation and produce organic matter, which is exuded from the cells and serves as a food source for heterotrophic microorganisms (Ducklow et al. 1986; Field et al. 1998; Pierson 2012; Thornton 2014). Phytoplankton, like those belonging to the classes Dinophyceae and Chrysophyceae, also exhibit alternative feeding behaviors, such as mixotrophy (Olrik 1998; Gribble and Anderson 2006) and heterotrophy (Grujcic et al. 2018). Some of the interactions observed in the FULL models result from these alternative feeding behaviours, as they have a wide prey range and are known to play an important role in microbial plankton community structure (Marshall and Laybourn-Parry 2002; Jeong et al. 2004; Flynn et al. 2013; Stoecker et al. 2017; Yoo et al. 2017; Jang et al. 2020; Liu et al. 2022). Mutualisms and symbioses have been extensively studied between phytoplankton and heterotrophic bacteria, and less among phytoplankton, but our model showed a highly connected phytoplankton community. One known association is the diazotroph-diatom interactions in which cyanobacteria fix dinitrogen to organic nitrogen available to diatoms, which produce photosynthate usable by the cyanobacteria (Zehr and Ward 2002). All our cyanobacterial ASVs belong to the Synechococcales, which are capable of nitrogen fixation (Spiller and Shanmugam 1987). However, our model showed low interactivity between the two groups, with only a few cyanobacteria co-occurring with Bacillariophyta and the majority co-excluding, indicating other type of interactions.

The bacterial community interacting with algae was highly dominated by Bacteroidota and Proteobacteria, which was expected as they are known to form close associations (Yang et al. 2009; Lachnit et al. 2011; Zhou et al. 2018; Woodhouse et al. 2018). Flavobacteria (Bacteroidota) are able to degrade high molecular weight components produced by phytoplankton, and the alphaproteobacterial *Roseobacter* (Rhodobacterales) form mutualistic interactions by feeding on algal exudate and

promoting algal growth by biosynthesising growth stimulants (Geng and Belas 2010; Seyedsayamdost et al. 2011; Buchan et al. 2014). Indeed, it is a two-way interaction and the phytoplankton also benefit from the associations. Certain bacteria belonging to the Bacteroidota have been observed to positively influence the phytoplankton community via phytohormones and algal growth regulation (Matsuo et al. 2005). However, co-exclusion has also been observed between phytoplankton and bacterioplankton ASVs, which usually interact positively. Algae senescence can trigger opportunistic pathogen behavior with the production of algicide compounds in the normally mutualistic bacteria *Roseobacter* (Seyedsayamdost et al. 2011). The presence of a proteobacterial strain in the environment can enhance the grazing activity of dinoflagellates on the Rhodobacterales genus Rhodomonas (Cryptophyta), creating a tripartite interaction (Alavi 2004).

The FULL models also uncovered interactions related to the nitrogen cycles. Such biogeochemical cycles rely on microorganisms to perform the different conversion. The nitrification, oxidation of ammonia to nitrite, happens in two steps, each requiring different actors: ammonium-oxidizing bacteria or archaea (AOB or AOA) and nitrite-oxidizing bacteria (NOB). In lake Constance, the only NOB ASV detected in the epilimnion all year long affiliates with Nitrospira, and such ASVs presented strong interactions with the AOB Nitrosomonadaceae and Methylococccales (AOB) (Llorens-Marès et al. 2020; Cai et al. 2022). The order Burhkolderiales, containing Nitrosominadaceae, also contains the family TRA3-20 that host ASVs with strong and consistent interactions with Nitrospirota. Burhkolderiales exhibit high nitrogen assimilation (conversion of ammonia or nitrate to organic nitrogen) in soil, and a close interaction with *Nitrospirota* assures an easy access to nitrate (Morrissey et al. 2018). The taxonomy of TRA3-20 is specific to the SILVA database, with only little information available except its first isolation from a freshwater environment (Kim et al. 2021). It is impossible to rule out the precise nature of the TRA3-20 – Nitrospirota connection, but the strength of the correlation coupled with the fact that all ASVs affiliated to TRA3-20 had such connection indicate very close relationships that could be related to the nitrogen cycles.

Fungi affiliated to Chytridiomycota are parasites of phytoplankton with a large spectrum of hosts (Leshem et al. 2016; Frenken et al. 2017; Frenken et al. 2017; McKindles et al. 2021). This feeding behavior makes them central in the microbial

plankton community and could explain their strong connection with other members of the microbial plankton community. Indeed, the parasitic activity is one of the causes of the phytoplankton bloom termination (Frenken et al. 2016), indicating a strong parasitic activity leading to the observed high co-occurrence between the phytoplankton and parasitic fungi ASVs observed in the FULL model. Fungi infection promotes the leaking of cellular material that can act as alternative food source for Proteobacteria and Bacteroidota (Agha et al. 2016; Haraldsson et al. 2018). Klawonn *et al*., 2021 showed that the Bacteroidota and Proteobacteria biomass associated with diatoms is 2 to 4 times higher when the diatoms is infected by *Chytridiomycota*, showing a tripartite interaction between the phytoplankton, fungi and bacterioplankton (Klawonn et al. 2021). Finally, *Chytridiomycota* infections fragment the host in smaller pieces more easily digestible by predators like Ciliophora, and the fungal zoospore can also act as food sources for grazers (Agha et al. 2016; Gerphagnon et al. 2019; Frenken et al. 2020). This process is called the mycoloop and is well established for zooplankton. Recent evidence showed that microeukaryote grazers like Ciliophora also benefit from it (Kagami et al. 2014; Farthing et al. 2021).

For some associations observed in the model we lack information from which to infer the process responsible, like the association between the previously described TRA3-20 and Nitrospirota. Another unexpected observation was Actinobacteria showing the highest connection within itself, the vast majority positive, and this was present in both the NULL and FULL models, indicating both ecological niche overlap and direct interactions. In addition, it showed high interactivity with other members of the microbial community, like the strong connection between acI and acIV ASVs with a specific subgroup of Chytridiomycota. In freshwater ecosystems, Actinobacteria are mostly studied through correlative analysis (Bunse et al. 2016; Woodhouse et al. 2016; Woodhouse et al. 2018; Arandia-Gorostidi et al. 2022) as isolation of dominant freshwater Actinobacteria strains, affiliated to the acI lineage, was only made for the first time in 2019 (Hahn 2009; Kim et al. 2019c). However, Actinobacteria are well studied in other environments and showed a strong mediation of microorganisms' interactions *via* secondary metabolism (Protasov et al. 2020). This still remains to be observed for freshwater strains.

Conclusions

Bayesian inference allows to overcome natural limitations of biological surveys applying frequentist statistics. The HMSC approach applied here to high dimensional compositional data, albeit still in development, showed robust and valid results and is a promising approach for microbial ecology analysis. This study brought an interesting picture of the forces driving the microbial plankton temporal dynamic in the epilimnion of Lake Constance. The vertical mixing proved to mainly negatively impact a diversity of ASVs that composed the microbial plankton community. Motility and diverse feeding behavior seemed to be an important factor in the capability of microbial plankton ASVs to overcome the environment constraints imposed by vertical mixing. In addition, the impacted ASVs likely show taxa-dependent capabilities and traits that cannot be assessed by our amplicon-based dataset. The ASVs association matrices allowed us to observe an intricate network of interactions between microbial plankton ASVs. The central place of the phytoplankton could be confirmed, with bacterioplankton and other non-photosynthetic microorganisms gravitating around them. Likewise, we could observe the recently described important role of the parasitic Chytridiomycota on the microbial plankton community structure and its alternative pathway in the trophic network. Finally, interactions not described in the literature were also observed, for example for the case of the gammaproteobacterial TR3A3-20 highly connected with Nitrospirota. This could be linked to a role in the nitrogen cycle. Equally novel is the high interactivity of freshwater Actinobacteria taxa, for which experimental information are scarce. JSDM is a correlative method, so it does not model the processes responsible for the abiotic and biotic interactions but provides information on the statistical relation of such processes (Ovaskainen and Abrego 2020). Our observations should be used as hypothesis driven material to isolate specific microorganisms from Lake Constance and design experiments to test our observations.

Funding: This research was funded by DFG Research Training Group R3 - Responses to biotic and abiotic Changes, Resilience and Reversibility of Lake Ecosystems (GRK 2272).

Data Availability Statement: The Illumina sequence data presented in this study will be available online upon publication.

Acknowledgments: We would like to highlight the excellent support received from Alfred Sulger, Julia Schmidt, Angelika Seifried, Pia Mahler and Josef Halder at the Limnological Institute of University of Konstanz for managing the ship cruises. Further, we would like to thank all members of the RTG-R3, especially Tina Romer, and members of the AG Schleheck for their support. We would like to thank Otso Ovaskainen for his input in the HMSC approach.

Conflicts of Interest: The authors declare no conflict of interest.

General discussion

Raising our knowledge of the sequencing technology

The development of -omics techniques has allowed for a massive step forward in microbial ecology (Kozińska et al. 2019). The independence from the cultivation and the use of genomic information via Next-Generation Sequencing (NGS) led to the discovery of a huge hidden biodiversity (Pedrós-Alió 2012; Rossi et al. 2016). However, it also brought in light the discrepancy between dependent and independent cultivation analysis. The number of taxa identified was in average 8 times less for the cultivation methods, with low sharing of the detected microorganisms between the methods and differences of abundance based on the microorganisms taxonomy and the type of medium used (Stefani et al. 2015; Youseif et al. 2021). While these -omics technologies are well established in the scientific landscape, they continue to evolve to fit the always increasing requirement of current science and try to overcome their existing limitation and bias (Hugoni et al. 2017; Salmaso et al. 2018; Salmaso et al. 2020; Abrego et al. 2020; Tedersoo et al. 2021).

Indeed, such refined and more sensitive techniques are coming at a cost. The increase of sensitivity (number of true microorganisms detected), coming from the development of the sequencing techniques, is coupled with a decrease the specificity (number of false positive detected) (Parikh et al. 2008; Prodan et al. 2020). Furthermore, numerous studies also showed that results of microbial community analysis using NGS are highly dependent of the type of protocol used in the samples preparation (Padilla et al. 2015; Parada et al. 2016; Krakat et al. 2017). This decrease of specificity coupled with the presence of bias in the methodology raise the need of a better understanding of these technologies as well as localise and study these biases to normalize them or take them in account in the analysis. Chapter II already developed the connection between filtration procedure and NGS-based plankton community results. Yet more are present.

Bias introduced by biological variability

Indeed, another influence on the relative abundances observed for a microbial community by NGS is the number of targeted gene copies. 16S rRNA gene copy number variation is quite low (between 1 to 15) (Stoddard et al. 2015) while 18S rDNA

copy can be much higher in variation between species (Wang et al. 2017). Relative abundance variation could be attributed to the actual relative abundance variation as well as the difference of gene copy number. But gene copy number is not the only influence. Intragenomic polymorphism caused by single nucleotide difference also affects the relative abundances observed for a microbial community(Acinas et al. 2004; Wang et al. 2017; Lavrinienko et al. 2021). Intragenomic polymorphism is positively correlated with the gene copy number, making it more prevalent for 18S rDNA than 16S rDNA (Acinas et al. 2004; Gong et al. 2013). The 18S rDNA copy - cell size linear relationship was proposed as a possibility to normalize abundance estimation (Godhe *et al*., 2008). But the intragenomic variability coupled with the species variability of the relationship could introduce more variance in the dataset (Godhe *et al*., 2008; Wang *et al*., 2017), more tests on a larger sample size and taxonomic diversity are required.

Bioinformatic normalization methods are currently being developed but still need to be taken with caution, as results could be affected by the use of different reference databases or by the targeted gene region (Stoddard et al. 2015). Furthermore, the rDNA copy number variation is an adaptive evolution linked to differences in metabolisms, growth rate, habitat specialization, response to stress, etc. (Klappenbach et al. 2000; Shrestha et al. 2007; Nemergut et al. 2016; Salim et al. 2017). The capability for microorganisms to adapt their gene copy number means that in changing environment, a high variation in gene copy number could be present and might continuously impact microbial community analysis and, thus, require to adapt the tools.

Bias introduced by methodological limitations

All molecular biology steps required for the preparation present methodological flaws (Berry et al. 2011; Krakat et al. 2017). DNA yield can vary up to 9-fold depending on the DNA extraction method. This yield difference impacts the observed community with differentially abundant microbial species as estimation showed that 3 to 16% of the community variability can be explained by the extraction method (Sui *et al*., 2008; Vaidya *et al*., 2018). Similar observations were made for the PCR amplification step, where up to 10-fold difference of relative abundance can be observed when comparing between expected and observed mock community composition (Parada et al. 2016).

Correlation between the expected and observed community composition can vary from 0.5, indicating poor representation of the community by these primers, to 0.95, indicating good representation of the community depending of the primers pair used.

Assessing primer pair efficiency has shown that the widely used 515F-Y/806R primer pair is a poor choice when studying marine taxa as it led to under and overestimation of some major marine taxa like SAR11 (Parada et al., 2016), making this primer pair not suitable in my study, as marine sister clade of SAR11 and other impacted taxa are also major taxa in freshwater environments (Salcher et al. 2011b; Henson et al. 2018). The primer pair 515F-Y/926R, targeting the V4-V5 hypervariable regions, showed the best correlation between expected and observed marine mock community and was a good starter choice for our studies.

The development of an Illumina sequencing chemistry delivering longer paired end reads gave the opportunity to use a longer amplicon. The 357F/926R primers pair, targeting the V3-V5 hypervariable regions, showed high concordance between observed and expected mock communities reads (Kim et al. 2011; Group 2012). The estimated amplicon size of 569 bp is within, but at the limit, of the maximum range of Illumina Miseq 2x300 bp as 30 bp of overlapping between forward and reverse read is advised for pairing of the reads into a complete sequence for analysis. Amplicons derived from these primers were used in the analysis of Chapters II, III and IV. While the paired reads generated for Chapter II where of sufficient quality, the read pairing for Chapter III and IV was not and only the forward reads could be used. Although less phylogenetic information is held by the forward read only, the 300 bp length allowed for an overlapping of the V3 and part of the V4 hypervariable regions, making it still suitable for analyzing the bacterioplankton community.

The dependence to bioinformatic and its problematic

The exponential increase in the number of bioinformatic software allows for virtually any kind of analysis but also increases the probability of introducing variability. Although two software theoretically do the same process, their calculation method can differ, making comparison of results difficult. Bioinformatic treatment of amplicon data has two major steps: sequences clustering and taxonomic affiliation. The taxonomic affiliation requires a reference database and a taxonomic affiliation software. General classification database like RDP, SILVA, NCBI and Greengenes are the most widely

used in microbial community analysis and a comparison of the taxonomic names found that an astonishing 63% to 90% of the names are unique to one of these four references databases (Balvočiūtė and Huson 2017). These differences influence the affiliation of the amplicons to an extend high enough to be statistically significant, resulting in under or overestimation of the number of detected taxa (Sierra et al. 2020; Poorlin Ramakodi 2021).Introduction of ecosystem-specific reference databases like Freshtrain, which was used in the analyses for the Chapters II, III and IV, allows for a much finer taxonomic affiliation, at the cost of decreased compatibility with studies not using it (Rohwer et al. 2018).

The use of operational taxonomic units (OTU), in which sequences are placed under the same OTU based on a similarity threshold, has been used since its introduction in 1963 (Long 1965). Currently, the OTU concept is less used in favour of the amplicon sequences variant (ASV) concept based on sequencing error correction algorithms and exact sequence variants (Callahan et al. 2017). But even if these OTU and ASV each fall into a single definition, multiple algorithms have been developed to calculate them, each with their own specificities. These specificities affect community analysis by introducing deviation in the number of detected taxa and their relative abundance (Nearing et al. 2018; Prodan et al. 2020; Chiarello et al. 2022; Jeske and Gallert 2022).

I decided to work with the dada2 based ASV in the studies presented in Chapter I to IV. The choice of the dada2 algorithm came from its easy integration in both R and QIIME2 and its high sensitivity compared to other ASV based software like deblur and UNOISE3 (Prodan et al. 2020). From a mock community, dada2 was the only software able to detect all taxa and taxa variants while creating few spurious ASVs that can be removed afterwards (Prodan et al. 2020). Furthermore, when different ASV-based techniques are used on the same dataset, the proportion of shared ASVs can range from 16 to 61% depending of the filtration threshold (see next paragraph), hence there is the necessity to think and specify which software was used (Prodan et al. 2020). A misconception can be the interpretation of ASV as species when they are oligotype, as the taxonomic separation and identity is based on genetic information (Eren et al. 2013). Oligotype have been proven effective in analysing microbial community composition and ecology (Hugoni et al. 2017; Salmaso et al. 2018; Salmaso et al. 2020; Abrego et al. 2020; Odriozola et al. 2021). However, the hypervariable region(s)

targeted during the amplification greatly influence the phylogenetic affiliation of ASV (Krakat et al. 2017; Bukin et al. 2019; Park et al. 2021), sometime leading to misinterpretation of ecologically and phylogenetically cohesive populations (Berry et al. 2017b). While the development of accessible and efficient bioinformatics tools helped its democratization, we should always keep a critical view on their use and interpretation.

A good analysis starts by a good data preparation

The choice of data treatment as well as the statistics usage have a tremendous impact on the reliability of the results. For this discussion, I will focus on NGS dataset which consist of count reads per sample and ASV. Due to the high capability detection of NGS, the datasets are sparse (zero inflated) with a high number of ASVs (high dimension), the majority being in low abundance. With dada2 creating some spurious ASVs and the library preparation also integrating mistakes, it leads to the presence of very low abundant ASVs that are not biologically relevant (Prodan et al. 2020). These need to be filtered and removed before analysing the data. The filtration threshold depends on several parameters, including the type of analysis one wants to do, and the computational power available. Community analyses that do not look at single ASVs can apply lower filtration thresholds, keeping meaningful low abundant ASVs as well as potential error ASVs not impacting the community analysis due to low representation (Odriozola et al. 2021).

However, when the analysis focus on the ASVs as well, such as comparing ASVs abundance between sample sites in Chapter III (ULC *versus* LLC) or ASVs temporal co-occurrence in Chapter IV, a more stringent filtration needs to be done to keep ASVs holding enough reads, or information, for the statistics to be robust. A filtration of the ASVs below 0.01% of relative abundance in at least 5% of the samples was applied in Chapter III, hence removing very low abundant ASVs (Prodan *et al*., 2020 applied a filtration of 0.002%) while keeping low abundant ASVs that could represent a unique and rare community in ULC or LLC. Although I might have lost biologically relevant ASV, higher thresholds of filtration were shown to not have an impact on the overall structure of the community (Odriozola et al. 2021). In the case of Chapter IV, the filtration I applied was even more stringent, as ASVs below 0.5% of relative abundance present in less than 20% of the samples were removed. ASVs

associations analysis require a high amount of information to get statistically well supported correlation and joint species distribution models (JSDM) with Bayesian inference require a much higher computational power compared to frequentist statistic. Good and thoughtful preparation of NGS dataset is a pivotal step of microbial community analysis in order to have robust and reliable statistic results.

The nature of NGS datasets?

After the filtration, the reads count difference between samples needs to be normalized. For a long time, they were reduced in all samples to an equal number. Called rarefaction, this data treatment is highly criticised because it removes valid and usable data from the analysis (McMurdie and Holmes 2014). Several different, alternative reads count normalizations have been developed (Robinson and Oshlack 2010; Anders and Huber 2010) but they inherit limitations in the case of highly dimensional and sparse datasets (Gloor et al. 2016b). An important factor to take into account is also the compositional nature of NGS datasets resulting from the arbitrary total reads number imposed by the sequencing machine (Gloor et al. 2017). As a result, NGS datasets give the proportional representation of each ASV in the community rather than a quantitative microorganisms´ representation in the original sample (Vandeputte et al. 2017).

Most of the analyses performed for microbial communities like ordination, network study and differential abundance calculation, rely on correlation calculation. In the context of compositional analysis, correlation includes severe problems like inflation of negative associations or the horseshoe effect (Gloor et al. 2017; Morton et al. 2017). The starting point of compositional analysis is a ratio transformation like log ratio or centered log ratio (Aitchison 2003; Silverman et al. 2017) that can be used for multivariate hypothesis testing like PermANOVA (van den Boogaart and Tolosana-Delgado 2013). Methods accounting for the compositional nature of the data like the PhilR distance calculation or aldex2 for differential abundance calculation, are being push forward and have been used for the analysis presented in Chapter III (Fernandes et al. 2013; Gloor et al. 2016a; Silverman et al. 2017). Accounting for the compositional nature of the data tends to decrease the false positive detection rate in ASV abundance comparison (Thorsen et al. 2016; Hawinkel et al. 2019), which can lead to

less “impressive” albeit much more accurate results. But understanding well how to prepare one’s data and its nature is not the only consideration to take in account.

Limits of statistics and a way to overcome them?

Statistical analyses in microbial ecology overuse the p-value, which measures the strength of evidence against the null hypothesis (Halsey et al. 2015). However, p-value is not absolute as the statistical power, which measure of the capacity of an experiment to find an effect when there is one, represent the repeatability of the p-value results (Halsey et al. 2015). It depends on several parameters that include the expected effect size, population variation and sample size. In order to control parameters that affect the statistical power, statistical analysis should be considered at the experimental design stage. But let's be honest, this is almost never the case in biology, leading to poor statistical power and low reproducibility of the p-value (Halsey et al. 2015; Pollard et al. 2019).

Furthermore, the way Ronald Fisher formulated the p-value in 1925 was as a continuous variable indicating the likelihood of a null hypothesis (Fisher 1956; Fisher 1992; Biau et al. 2010). However, the current use of the p-value is derived from the null hypothesis significance testing (NHST) approach (Neyman and Pearson 1933). This approach introduced an arbitrary threshold (usually 0.05), representing the compromise between type I and type II error rates, and transformed the p-value into a binary variable with a significant or non-significant result. Although 0.05 was considered a sufficiently stringent threshold to avoid noise affecting the results, in reality it has a high risk of false positives, leading to misinterpretation and low reproducibility of p-value results. (Colquhoun 2017). Researchers now emphasise the use of other statistical indicators, either to complement p-value results or as an alternative to using the p-value (Nakagawa and Cuthill 2007).

Chapter III offered an interesting showcase, as the results of the PermANOVA and ANOSIM, testing the microbial plankton biodiversity community difference between ULC and LLC, showed p-values that could contradict with the other values of the tests. p-value indicated difference in biodiversity for the three microbial plankton communities (18S-NP, 18S-PP and 16S-PP) but the other statistics values indicated otherwise for the 18S-NP and 18S-PP community, leading to a more discussed outcome (Figure III.2 and Table III.4). When applying statistics, it is extremely

important to understand how the statistics test works and use all their values given for interpreting the results or else the conclusion derived from the interpretation can be wrong. Other values can be used as alternative to the p-value. Aldex2, used in Chapter III to study ASVs differentially abundant between ULC and LLC, can measure the effect size, the mean difference between two groups, and the confidence interval of the effect size rather than its "statistical significance" (p-value) (Nakagawa and Cuthill 2007). It is more informative for our analysis as there is no significance in what I measured but a degree of difference of abundance between ULC and LLC. As a result, the number of positive results, false positive, decrease highly (Figure 2) (Fernandes et al. 2013; Thorsen et al. 2016).

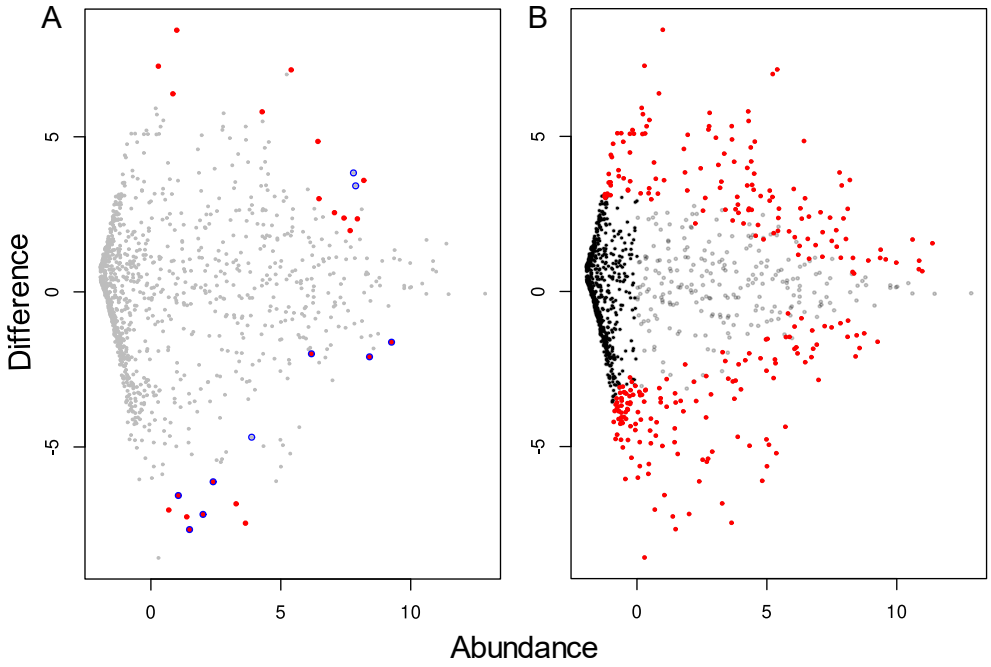

Figure 2: Relationship between relative abundance and differences between group of ASVs. The relationship was measured using aldex2 and the analysis was based on effect size and confidence interval (A) or q-value (B). Each dot represents a single ASVs. Red dot represents an ASVs with effect size above 1.5 (A) or p-value below 0.05 (B). Blue circle represents a confidence interval above 95% (A). These graphics shows the comparison of ASV relative abundance between ULC and LLC of the 16S-PP database during summer done in Chapter III.

Bayesian statistics: the future?

An alternative is to move from the frequentist statistics to the Bayesian inference. The JSDM used in Chapter IV, called hierarchical modelling of species communities (HMSC), to analyse the temporal microbial plankton dynamics and interactions is based on Bayesian inference. JSDM are correlative models, looking at

community-level patterns of how species react to their environment, relating such pattern to species traits and phylogenies and observe species co-occurrence (Ovaskainen and Soininen 2011; Ovaskainen et al. 2016; Ovaskainen et al. 2017a; Ovaskainen et al. 2019; Ovaskainen and Abrego 2020; Tikhonov et al. 2020b). Such models have been used in ecology successfully to model community ecology (Clark et al. 2014; Lei et al. 2015; Inoue et al. 2017; Wagner et al. 2020; Morgado et al. 2021) but stay limited in microbial ecology (Minard et al. 2019; Odriozola et al. 2021).

The first limitation is the high computational power required which is linked to the dataset size, and NGS dataset are much larger than datasets from field observations. The sparsity and high dimension of NGS data makes it more difficult for JSDM to achieve convergence of parameters, which is the indicator of results well supported by the probability (Ovaskainen and Abrego 2020; Tikhonov et al. 2020b). Big advancement has been made in both computer power, and the improvement of JSDM in modelling high dimensional datasets is an intense field of research, allowing such models to be used in microbial ecology (F. Dormann et al. 2007; Warton et al. 2015; Minard et al. 2019; Odriozola et al. 2021; Doser et al. 2022). One of the main limitations in microbial environmental studies is the sample size, which is usually not high enough to have good statistical power, as explained above.

Sample size is not as critical for building a Bayesian model (van de Schoot et al. 2014; van de Schoot et al. 2015). An experimental design created without statistical background, or limited by external factor like time, founding or human factor, will result in more robust results when using Bayesian inference rather than frequentists statistic. But I shall not go into a dichotomic thought of frequentists statistics and p-value is wrong and Bayesian inference right. Those two fields of analysis are each bringing their own strength and weakness and are valid tools to be used. Difficulties arise from the misunderstanding of these tools and unplanned experimental design. A better integration of the statistics teaching in biology and environmental science cursus as well as a collaboration with statistician should be push forward in the future. Despite the “dark” presentation of the statistics made here, the underlying message the underlying message is positive. The increase in the number of articles addressing these issues and proposing alternative solutions or approaches demonstrates a mindsets evolution and an awareness of the importance of statistics in biology.

Concept of resilience and stability in microbial ecology

Like observed in Chapter IV, the vertical mixing has a large impact on the microbial plankton community (Figure IV.2). But this impact is not constrained to the plankton community as the vertical mixing is an important component of the lake ecosystem. Weak vertical mixing means that the nutrient regeneration of the epilimnion will be weak. This low regeneration coupled with low water column physical turbulence can lead to shift in the phytoplankton community in favor of organisms with smaller cells size, hence potentially promoting the establishment of cyanobacteria, and potentially toxic cyanobacteria, replacing the diatoms as major phytoplankton bloom members (Winder and Hunter 2008; Fraisse et al. 2015).

Weaker vertical mixing associated with climate warming has a much larger effect than just on the diatom and cyanobacterial communities. From 2013 to 2017, the vertical mixing of Lake Constance was relatively weak, leading to the low nutrient regeneration in the epilimnion and by extent accumulation of nutrients and a decrease in oxygen concentration in *Überlinger See* (Limnologischer Zustand des Bodensees). If weaker vertical mixing occurs over a long period of time, one may observe the appearance of an oxygen-depleted hypolimnion accumulating nutrients and of a highly nutrient-depleted epilimnion. A sudden loading of the epilimnion by strong vertical mixing after a long period of thermal stratification, could trigger very high autotrophic activity and cyanobacterial blooms (Duarte et al. 2000; Jochimsen et al. 2013), severely affecting lake fauna and flora or increasing the intensity of seasonal hypoxic event like observed in *Zeller see* (Conroy et al. 2011). Such an event probably would not last, as the nutrient loading is not constant, but the resilience from such a traumatic event at the ecosystem level could be long and incomplete.

When the disturbance is 'short term', analyses have found relatively similar community compositions before and after, despite changing during the disturbance (Shade et al. 2011; Shade et al. 2012b; Ma et al. 2020), like what can be observed during the natural hypoxic event happening in *Zeller see* (Chapter III). However, resilience to long-term disturbances showed communities that either did not recover or could be significantly different from the one before the disturbance (Meisner et al. 2018; de Vries et al. 2018; Ibrahim et al. 2021). The extent to which high microbial community biodiversity is important for ecosystem functioning in response to a

disturbance is still debated, particularly in relation to the concept of functional redundancy (van der Plas 2019). Higher biodiversity increases the probability of finding species capable of resisting and recovering from a disturbance (Tilman and Downing 1994; Naeem and Li 1998; Lynch 2002; Aoki 2003), but contrasting results have been observed with observations of more stable environments hosting less diverse biodiversity (Givnish 1994; Andrén et al. 1995; Ulanowicz 2003).

Functional redundancy and metabolic flexibility of the community have been found to be independent of the taxonomic composition of the community (Avila-Jimenez et al. 2020). Therefore, the relationship between stability and diversity should not be limited to biodiversity, and functional diversity has been proposed to develop the diversity-stability relationship (Botton et al. 2006; Philippot et al. 2021). Functional diversity means that biotic interactions must be taken into account when studying the stability of a community (Griffiths and Philippot 2013; Palit et al. 2022). Amplicon sequencing based community analyses, as in this thesis, are the first necessary steps to observe community composition, infer biotic interactions and response to environmental parameters from a taxonomic point of view. However, it is too limited from a resilience and predictive point of view. Using the observation of amplicon sequencing as prior knowledge, metagenomic coupled with transcriptomic or even proteomic techniques would bring a more holistic approach to community stability analyses.

Conclusions

The black box of Lake Constance's microbial plankton community is finally observable. The microbial plankton community showed a typical composition based on observations in other pre-Alpine lakes. However, the comparison of the community composition based on the sampling site showed extraordinary differences that could be related to the environmental specificity of the sampling site, even though the two sites are connected and therefore share the same seed of microbial plankton. The specific community also seemed to influence the microbial plankton feeding on it, such as predators and parasites, indicating that the dynamics of the community is a combination of both abiotic factors and biotic interactions. This was confirmed when the model analysing the temporal community dynamics showed a high proportion of the community affected by vertical mixing, while the ASVs associator matrices revealed a huge network of connections between ASVs.

These observations and analyses have been made possible by the development of Next-Generation Sequencing and subsequent bioinformatics and biostatistics. NGS is a wonderful tool that has allowed biology to enter a whole new era. However, it seems that the advent of these new approaches has been so rapid that researchers are currently drowning in an unfathomable amount of information and a ridiculously large sea of possible analytical methods. Information is often analysed without really thinking about how, why and how to do it properly. Hopefully, with the democratisation of these approaches, there will be a greater awareness of these problems and a real envy of the learning and development of this research in the right way.

Future prospects

We have now established a DNA repository of the seasonal succession of the microbial plankton community of Lake Constance. This DNA bank covering Lower Lake Constance (2018-2019) and Upper Lake Constance (2018-present) is accessible for many kind of analyses. Climate change and anthropogenic activities will continue to alter our environment, either through acute disturbances such as increased occurrence of extreme weather events, or through chronic disturbances such as nutrient runoff. Such repositories can be used as data to feed and train models, thereby increasing their predictive power. The HMSC-based model used in Chapter IV to model the temporal dynamics of the microbial plankton community could be applied to the multi-year data set. By incorporating a difference between strong and weak vertical mixing and an emphasis on the cyanobacterial community for the model predictive power, the prediction of future community dynamics would increase. Re-analysis of the DNA bank by metagenomics would provide the community taxonomy and genomic capabilities of the species. Physiological and metabolic capabilities inferred from the metagenome could be incorporated into HMSC models, providing a correlation between environmental factors and metabolic capabilities, a correlation that could later be tested in laboratory after isolating microbial plankton species.

Observations from the models could lead to laboratory experiments to confront the observations. The important role of parasitic fungi in the microbial plankton community via the mycoloop is relatively new, Google scholar only giving 303 results, and represent an exciting field of work to pursue if fungi could be isolated from Lake Constance. The model found a high number of Actinobacteria associations and since their isolation from freshwater environments is now possible, a whole new avenue of experiments opens up as a huge amount of correlative information is available on this phylum and needs to be confronted with experimental data. Our studies focused on the microbial plankton community of the epilimnion of the lake. However, some species live at specific depths within this zone, depending on their pigment composition or environmental conditions, such as the hypoxic zone in the Lower Lake Constance. A vertical transect of the epilimnion when pigment peaks are observed or during the hypoxic event would give a more accurate resolution of the microbial plankton community.

Record of achievement

Chapter I

Eva Riehle did the toxin extraction and analysis, the DNA extraction, qPCR and RT-PCR with Sanger sequencing and the evaluation of the Moldaenke probe data.

I did the DNA library preparation for amplicon sequencing, the bioinformatic and biostatistics of the sequencing data and the phylogenetic analysis.

This study is a shared first co Authorship between Eva Riehle and I. Manuscript writing was done evenly by both.

Manuscript revision was done by Prof. Dr. David Schleheck and Prof. Dr. Daniel Dietrich.

Chapter II

Prof. Dr. David Schleheck designed and built the flowmeter device and wrote its program.

Alexander Fiedler worked as a Hiwi student. He participated to the calibration of the flowmeter device and corrected some part of the program in accordance.

I did the water sampling, DNA extraction, DNA library preparation, bioinformatic and biostatistics data treatment.

Manuscript was written by me.

Manuscript revision was done by Prof. Dr. David Schleheck.

Chapter III

I did the water sampling, DNA extraction and DNA library preparation, bioinformatic treatment and biostatistics analysis.

Manuscript was written by me.

Manuscript revision were done by David Schleheck.

Chapter IV

I did the water sampling, DNA extraction and DNA library preparation, bioinformatic treatment and biostatistics analysis.

Manuscript was written by me.

Manuscript revision were done by Dr. Laura Epp.

Supplementary Files

Chapter I: Are toxin-producing *Planktothrix* spp. an emerging species in Lake Constance?

Figure SI.1: Representative illustration of the different coloration of glass fiber filters loaded with biomass collected from chlorophyll-a maximum (left) and DRM (right). Two liters of water taken at 4 m or 20 m water depth, respectively, were filtered through GF6 filters. The samples were taken on June 9th, 2021.

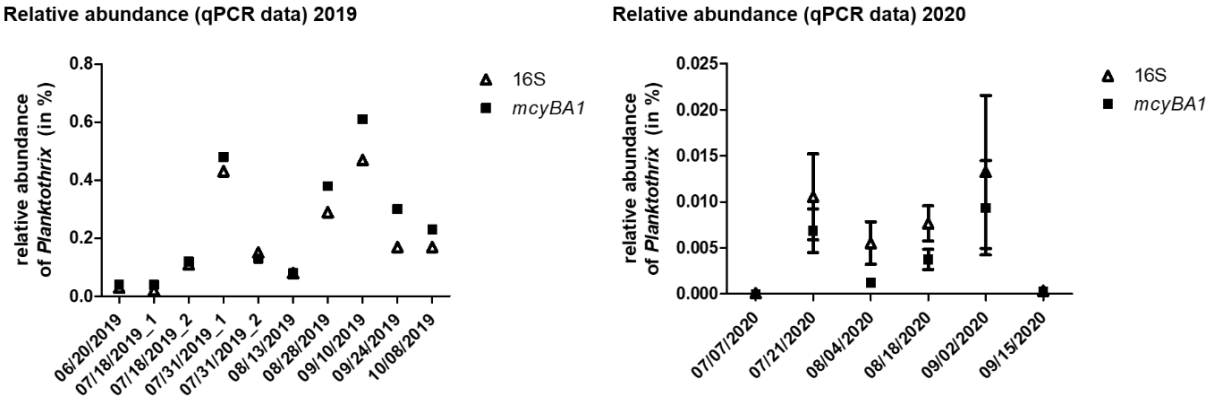

Figure SI.2: Relative abundance of *Planktothrix* spp. as determined by quantitative PCR. A *Planktothrix* 16S rDNA gene fragment and a *mcyBA1* domain were detected using the primer pairs from Ostermaier et al.. 2009 and analysed as described in Materials and Methods section. Squares represent relative abundance of *Planktothrix*-specific 16S rDNA gene fragments and triangles represent *Planktothrix*-specific *mcyBA1*. which amplified the first adenylation domain of *mcyB*. one part of the microcystin gene cluster (Ostermaier and Kurmayer 2009). 2019 data are represented as single determinations. while 2020 data are represented as mean of biological triplicates (n = 3) ± SEM.

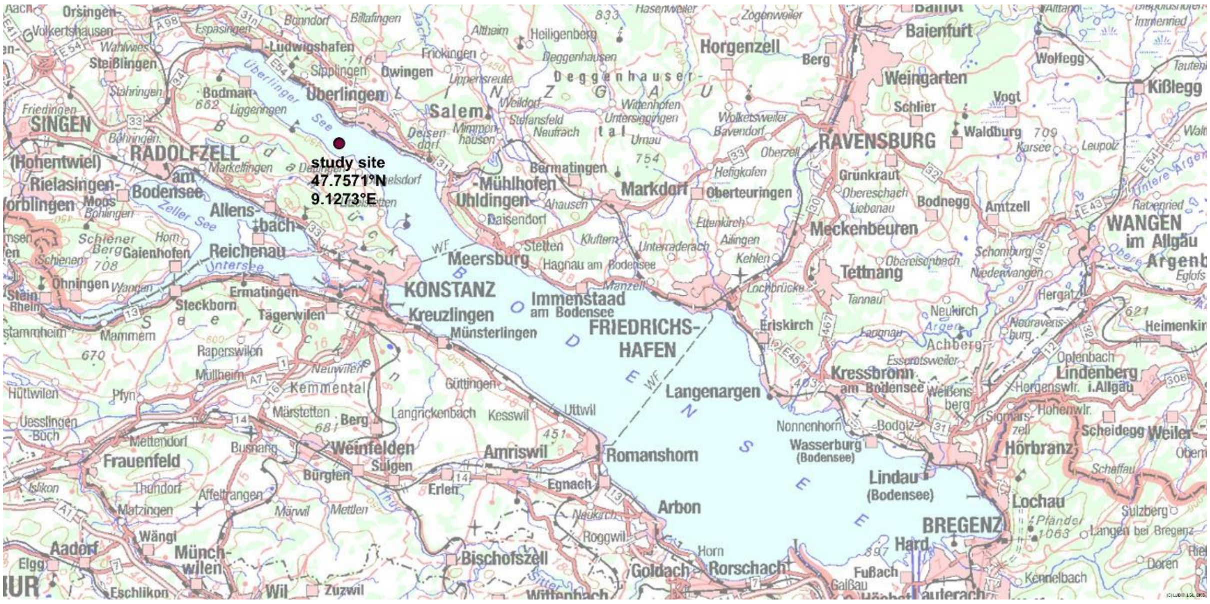

Figure SI.3: Localization of the study site in the *Überlingen* embayment of Lake Constance. The map was modified using the online tool UDO and is courtesy of LUBW Baden-Wuerttemberg [2]. The study site and the corresponding coordinates (47.757°N 9.1273°E) are indicated on the map.

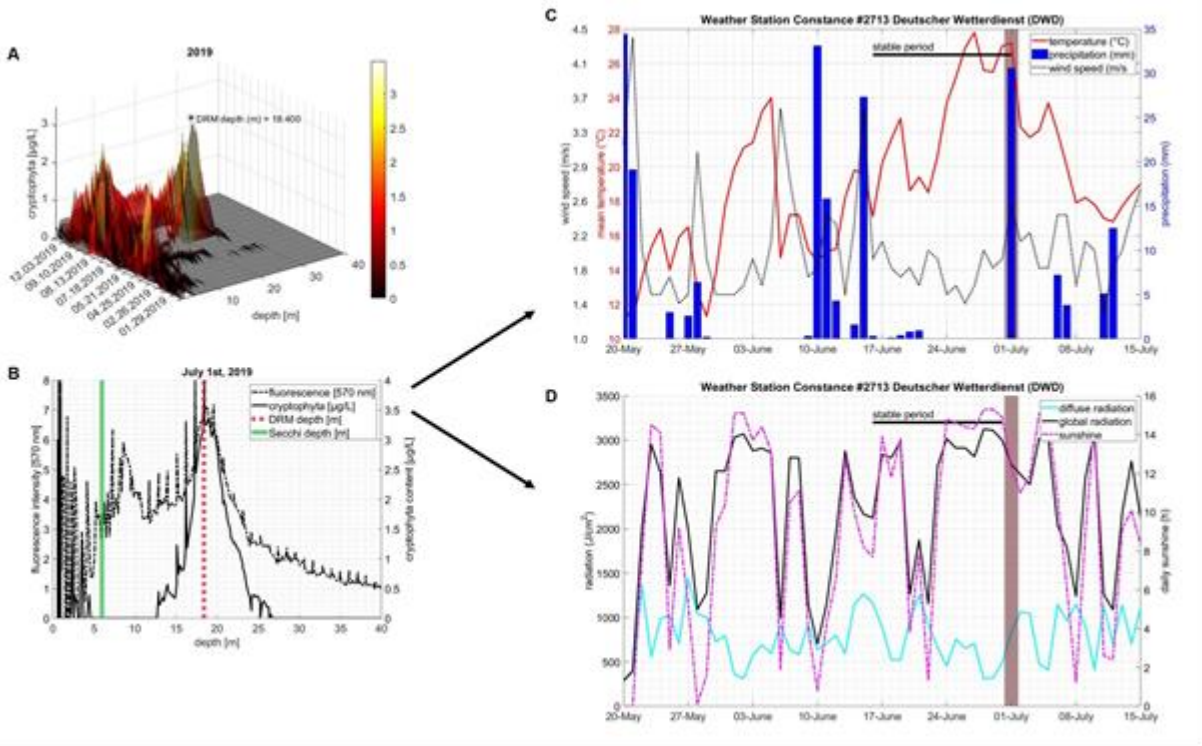

Figure SI.4: Detailed analysis of weather data in 2019. A: 3D plot of cryptophyta content (µg/L) from Jan-Dec of 2019, DRM is indicated by asterisk and marks the depth of maximal red pigment abundance (FluoroProbe 'cryptophyta' abundance, see main text); B: detailed analysis of 07/01/2019 FluoroProbe data, green bar indicates Secchi depth, red dotted bar indicates DRM depth; C: weather datasets six weeks prior and two weeks after the peak of red pigment abundance, including wind speed (m/s, black dotted line), mean temperature (°C, red line) and precipitation (mm, blue bars); D: radiation datasets six weeks prior and two weeks after July 1st, 2019 including diffuse (cyan line) and global radiation (black line, J/cm^2) and daily sunshine (h, pink dotted line). Chocolate colored bars in C and D indicate July 1st and the peak in cryptophyta concentration, stable weather period is indicated as black bar from June 17th to June 30th, 2019.

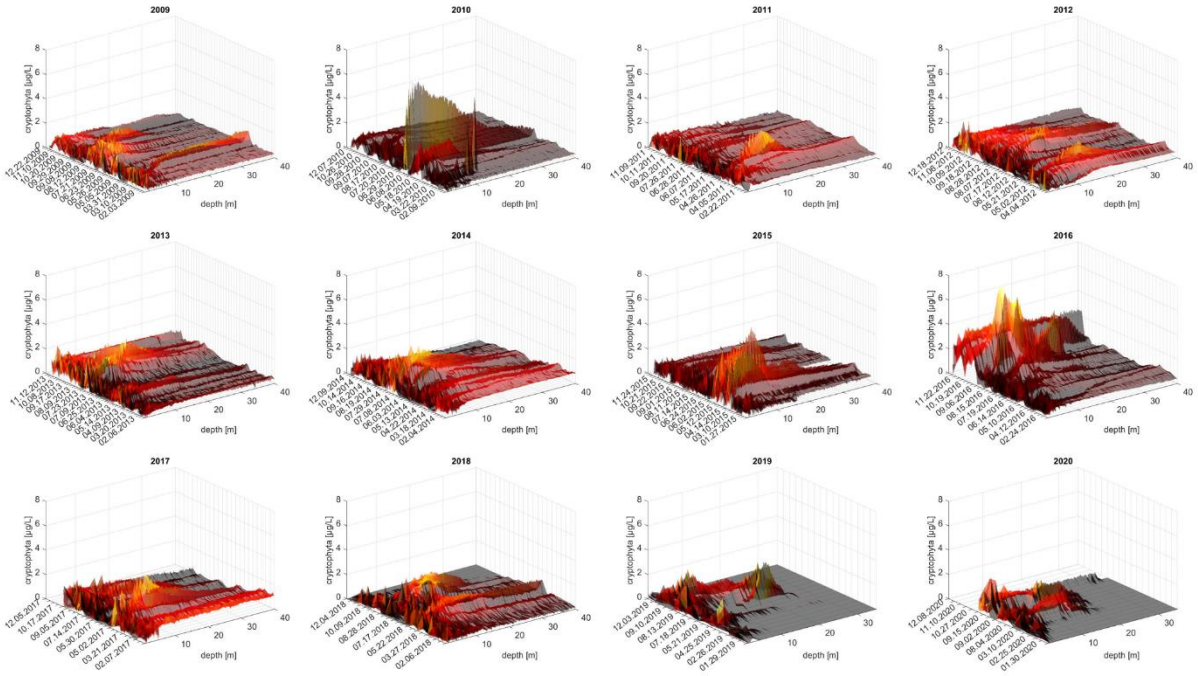

Figure SI.5: Depth-profiles recorded with the FluoroProbe from 2009-2020. Shown are the FluoroProbe profiles for 'cryptophyta' abundance recorded as proxy of red-pigment abundance in the water column from 0-40 m depth at the routine sampling site '*Wallhausen'* in the Lake *Überlingen* embayment of Upper Lake Constance. Coordinates of the routine sampling site, 47.7571°N 9.1273°E. Note that in this figure the dates follow American-style date format.

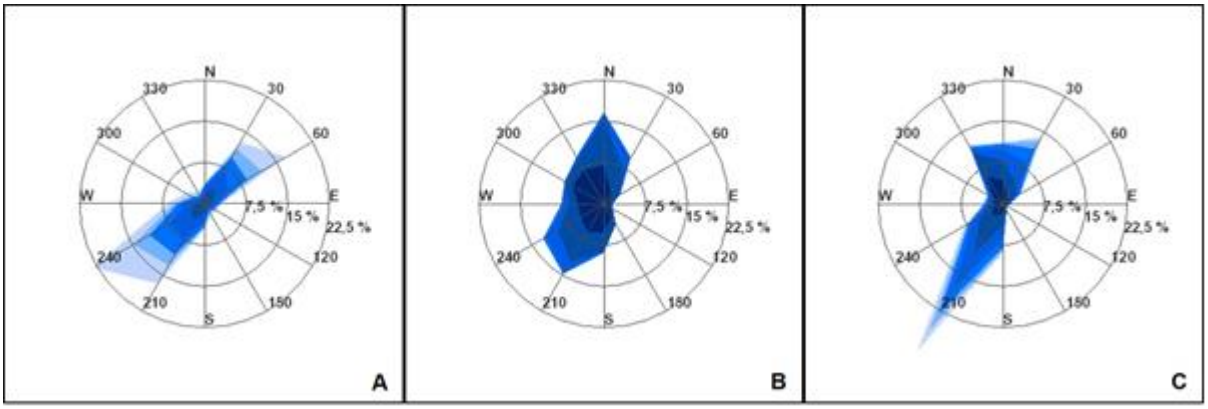

Figure SI.6: Mean wind speed and directions at three measurement locations near the study site. A: Illmensee wind station, B: Pfullendorf LUBW station, C: Singen LUBW station. Source: LUBW Baden-Wuerttemberg.

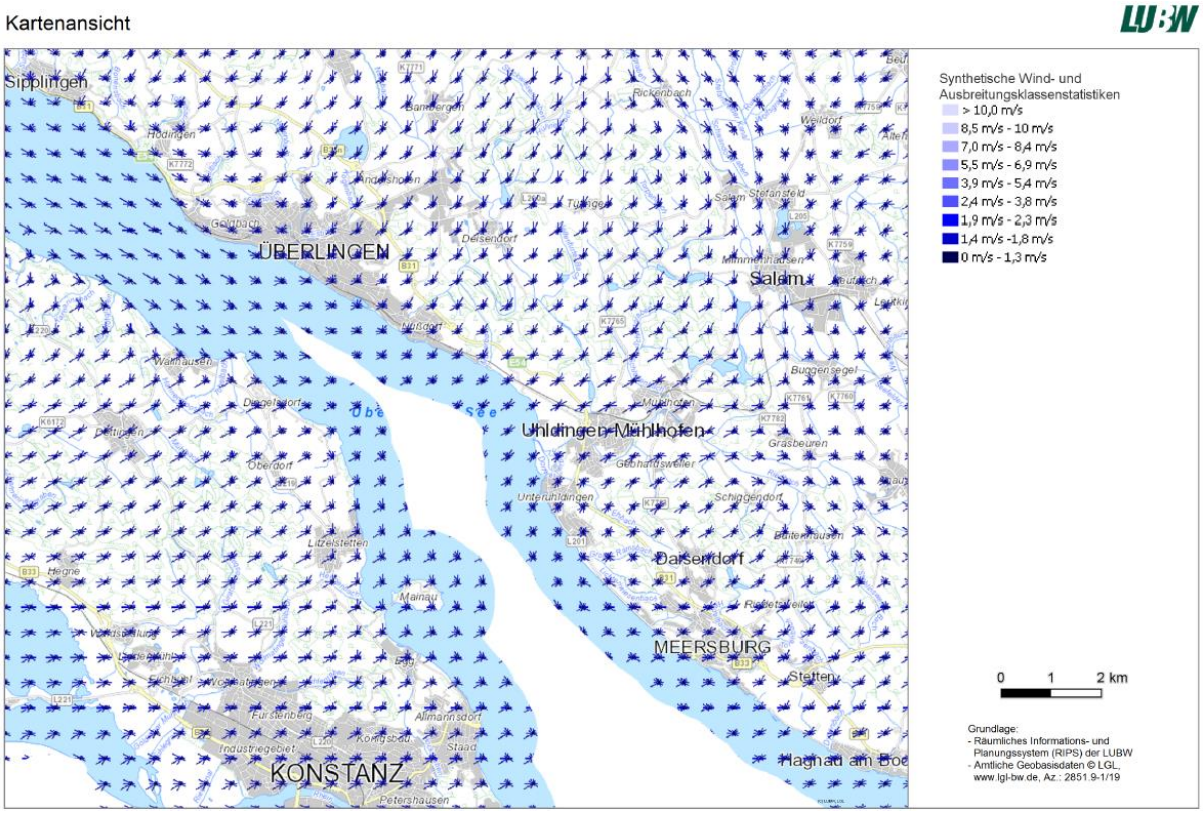

Figure SI.7: Overview map of mean wind speeds around the study site (47.757°N 9.1273°E). Data are courtesy of LUBW Baden-Wuerttemberg.

Table SI.1: Assignment of ASVs with SILVA and Greengenes databases

	SILVA_138						Greengenes				
ASVs	Class	Order	Family.Lineage	Genus.Clade	Species.tribe	Class	Order	Family.Lineage	Genus.Clade	Species.tribe
ASV1	Cyanobacteria	Cyanobacteriales	Microcystaceae	Microcystis_PCC-7914	Radiocystis_sp._JJ30-12	Oscillatoriophycideae	Chroococcales	Microcystaceae	Microcystis	NA
ASV8	Cyanobacteria	Cyanobacteriales	Phormidiaceae	Planktothrix_NIVA-CYA_15	NA	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	Planktothrix	NA
ASV9	Cyanobacteria	Cyanobacteriales	Phormidiaceae	Planktothrix_NIVA-CYA_15	Planktothrix_rubescens_CCAP_1459/14	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	Planktothrix	NA
ASV13	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV14	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV15	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV17	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV18	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV22	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV28	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV29	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV30	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV33	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA

	SILVA_138						Greengenes				
ASVs	Class	Order	Family.Lineage	Genus.Clade	Species.tribe	Class	Order	Family.Lineage	Genus.Clade	Species.tribe
ASV4	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV7	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV8	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_Synechococcus_sp.	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV10	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV12	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV17	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV21	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV22	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	uncultured_bacterium	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV25	Cyanobacteria	Synechococcales	Cyanobiaceae	Cyanobium_PCC-6307	NA	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	NA
ASV32	Cyanobacteria	Cyanobacteriales	Phormidiaceae	Planktothrix_NIVA-CYA_15	NA	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	Planktothrix	NA
ASV33	Cyanobacteria	Cyanobacteriales	Phormidiaceae	Planktothrix_NIVA-CYA_15	Planktothrix_rubescens_CCAP_1459/14	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	Planktothrix	NA
ASV34	Cyanobacteria	Cyanobacteriales	Microcystaceae	Microcystis_PCC-7914	NA	Oscillatoriophycideae	Chroococcales	Microcystaceae	Microcystis	NA

Table SI.2a: NCBI megablast results for *Cyanobium gracile / Synechococcus rubescens* affiliated sequences from 2019

	ASVs	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
2019	ASV13	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	584	584	99%	1E-166	98.78	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	529	529	99%	5E-150	95.73	1476	NR_102447.1
	ASV14	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	590	590	99%	2E-168	99.09	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	529	529	99%	5E-150	95.73	1476	NR_102447.1
	ASV15	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	610	610	100%	2E-174	100	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	521	521	100%	8E-148	95.15	1476	NR_102447.1
	ASV17	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	560	560	100%	2E-159	97.27	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	538	538	100%	8E-153	96.06	1438	NR_114406.1
	ASV18	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	560	560	100%	2E-159	97.27	1438	NR_114406.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	560	560	100%	2E-159	97.27	1456	NR_125481.1
	ASV22	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	588	588	100%	7E-168	98.79	1438	NR_114406.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	510	510	100%	2E-144	94.55	1456	NR_125481.1
	ASV28	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	577	577	100%	2E-164	98.18	1476	NR_102447.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	532	532	100%	4E-151	95.76	1456	NR_125481.1
	ASV29	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	566	566	100%	3E-161	97.58	1438	NR_114406.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	544	544	100%	2E-154	96.36	1456	NR_125481.1
	ASV30	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	577	577	100%	2E-164	98.18	1438	NR_114406.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	538	538	100%	8E-153	96.06	1456	NR_125481.1
	ASV33	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	549	549	100%	3E-156	96.67	1438	NR_114406.1
		Prochlorococcus marinus subsp. pastoris strain PCC 9511 16S ribosomal RNA, partial sequence	Prochlorococcus marinus subsp. pastoris	544	544	100%	2E-154	96.37	1465	NR_125480.1

Table S2b: NCBI megablast results for *Cyanobium gracile* / *Synechococcus rubescens* affiliated sequences from 2020

	ASVs	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
2020	ASV4	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	706	706	100%	0	100	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	617	617	100%	1E-176	95.81	1476	NR_102447.1
	ASV7	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	684	684	100%	0	98.95	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	623	623	100%	2E-178	96.07	1476	NR_102447.1
	ASV8	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	678	678	100%	0	98.69	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	623	623	100%	2E-178	96.07	1476	NR_102447.1
	ASV10	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	656	656	100%	0	97.64	1476	NR_102447.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	656	656	100%	0	97.64	1456	NR_125481.1
	ASV12	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	640	640	100%	0	96.86	1476	NR_102447.1
		Prochlorococcus marinus subsp. pastoris strain PCC 9511 16S ribosomal RNA, partial sequence	Prochlorococcus marinus subsp. pastoris	628	628	100%	5E-180	96.34	1465	NR_125480.1
	ASV17	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	684	684	100%	0	98.95	1438	NR_114406.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	606	606	100%	2E-173	95.29	1456	NR_125481.1
	ASV21	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	656	656	100%	0	97.64	1476	NR_102447.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	634	634	100%	0	96.6	1456	NR_125481.1
	ASV22	Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	667	667	100%	0	98.17	1476	NR_102447.1
		Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	623	623	100%	2E-178	96.07	1456	NR_125481.1
	ASV25	Synechococcus rubescens strain SAG 3.81 16S ribosomal RNA, partial sequence	Synechococcus rubescens	656	656	100%	0	97.64	1456	NR_125481.1
		Cyanobium gracile PCC 6307 16S ribosomal RNA, partial sequence	Cyanobium gracile PCC 6307	634	634	100%	0	96.6	1438	NR_114406.1

Table SI.3a: Conover-Iman test on *Synechococcus* ASVs of 2019: Conover-Iman False Discovery rate results after Benjamini Yekutieli p-value correction for multiple comparisons between the main *Synechococcus* ASVs of 2019 presented in the article in Fig. I.4A and I.4B.

		ASV13	ASV15	ASV14	ASV29	ASV28	ASV22	ASV30	ASV18	ASV33	ASV17
2019	ASV13		1.0000	1.0000	0.1417	0.0569	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	ASV15			1.0000	0.1312	0.0503	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	ASV14				0.8084	0.3926	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**
	ASV29					1.0000	**0.0009**	**0.0117**	**0.0005**	**0.0000**	**0.0000**
	ASV28						**0.0031**	**0.0357**	**0.0017**	**0.0000**	**0.0000**
	ASV22							1.0000	1.0000	1.0000	0.7186
	ASV30								1.0000	0.2297	0.1330
	ASV18									1.0000	0.9717
	ASV33										1.0000
	ASV17										

Table SI.3b: Conover-Iman test on *Synechococcus* ASVs of 2020: Conover-Iman False Discovery rate results after Benjamini Yekutieli p-value correction for multiple comparisons between the main *Synechococcus* ASVs of 2020 presented in the article in Fig. I.4C and I.4D.

		ASV4	ASV7	ASV21	ASV22	ASV8	ASV10	ASV12	ASV17	ASV25
2020	ASV4		0.8320	0.0952	**0.0039**	**0.0009**	**0.0001**	**0.0000**	**0.0000**	**0.0000**
	ASV7			1.0000	0.1426	**0.0447**	**0.0029**	**0.0001**	**0.0001**	**0.0000**
	ASV21				1.0000	0.4315	0.0537	**0.0026**	**0.0022**	**0.0006**
	ASV22					1.0000	0.6496	0.0715	0.0548	**0.0187**
	ASV8						1.0000	0.2312	0.1822	0.0677
	ASV10							1.0000	1.0000	0.5650
	ASV12								1.0000	1.0000
	ASV17									1.0000
	ASV25									

Table SI.4a: Relative abundance of *Synechococcus* ASVs in 2019: Relative abundance in percentage of the main *Synechococcus* taxa presented in Fig. I.4 in 2019. The color scheme is equivalent of what is represented in the heatmap in Fig. I.4A on the Log10 transformed data.

	01.07.2019	18.07.2019	31.07.2019	13.08.2019	28.08.2019	10.09.2019	24.09.2019	08.10.2019
ASV13	29.96	36.51	24.64	15.04	20.64	16.83	10.58	10.86
ASV15	11.47	16.34	22.17	22.30	18.76	21.78	16.68	15.29
ASV14	8.52	17.42	15.53	21.27	9.29	18.43	21.05	14.75
ASV29	15.64	6.23	10.33	4.87	27.54	9.52	6.69	15.86
ASV28	21.66	8.19	9.27	10.96	5.22	14.00	7.92	7.10
ASV22	0.00	0.22	0.79	3.25	3.55	1.74	14.54	16.15
ASV30	8.30	4.05	4.65	5.14	2.89	3.74	2.39	2.92
ASV18	0.69	2.03	2.07	4.21	1.07	4.05	7.11	4.71
ASV33	0.13	4.20	2.37	7.16	2.73	0.10	0.21	0.26
ASV17	1.85	2.61	1.83	1.20	1.90	2.24	1.03	1.35

Table S4b: Relative abundance of *Synechococcus* ASVs in 2020: Relative abundance in percentage of the main *Synechococcus* taxa presented in Fig. I.4 in 2020. The color scheme is equivalent of what is represented in the heatmap in Fig. I.4C on the Log10 transformed data.

	07.07.2020	21.07.2020	04.08.2020	18.08.2020	02.09.2020	15.09.2020
ASV4	35.36	40.57	56.21	67.61	49.61	23.16
ASV7	48.06	20.52	8.62	5.95	12.18	25.12
ASV21	0.97	10.00	8.88	16.01	13.84	19.42
ASV22	7.37	7.13	2.62	2.09	3.96	10.45
ASV8	3.30	5.01	3.47	1.84	3.82	6.76
ASV10	2.87	3.17	1.13	0.90	2.86	7.45
ASV12	0.53	5.64	7.62	0.56	0.94	0.70
ASV17	0.15	1.02	2.51	1.71	4.67	0.75
ASV25	0.32	1.35	1.03	0.98	1.17	1.67

Chapter II: Description of a “plankton filtration bias” and of an Arduino microcontroller-based flowmeter device that can help to resolve it

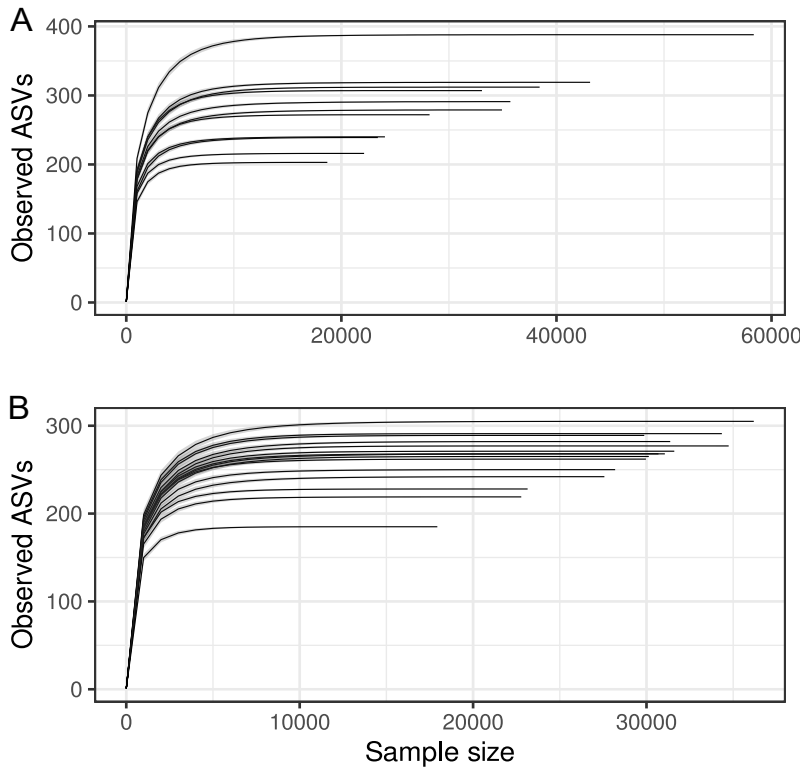

Figure SII.1: Rarefaction curve of the samples of experiment 1 (A) and 2 (B). X-axis: number of sub-sampled sequences. Y-axis: number of detected ASVs

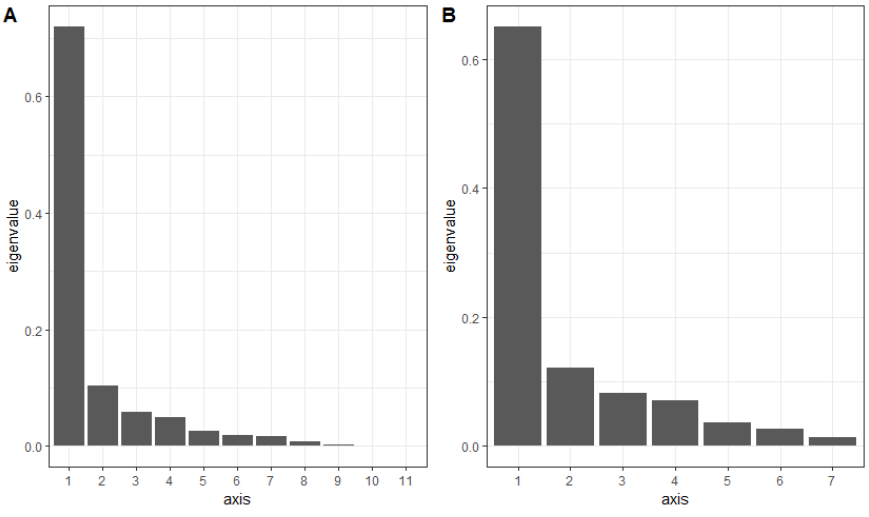

Figure SII.2: Eigenvalue of Principal Coordinate Analysis axis. Percentage of explanation of each axis of the PCoA. A: Axis of PCoA Figure.II.4. of the first experiment. B: Axis of PCoA Figure.II.9. of the second experiment.

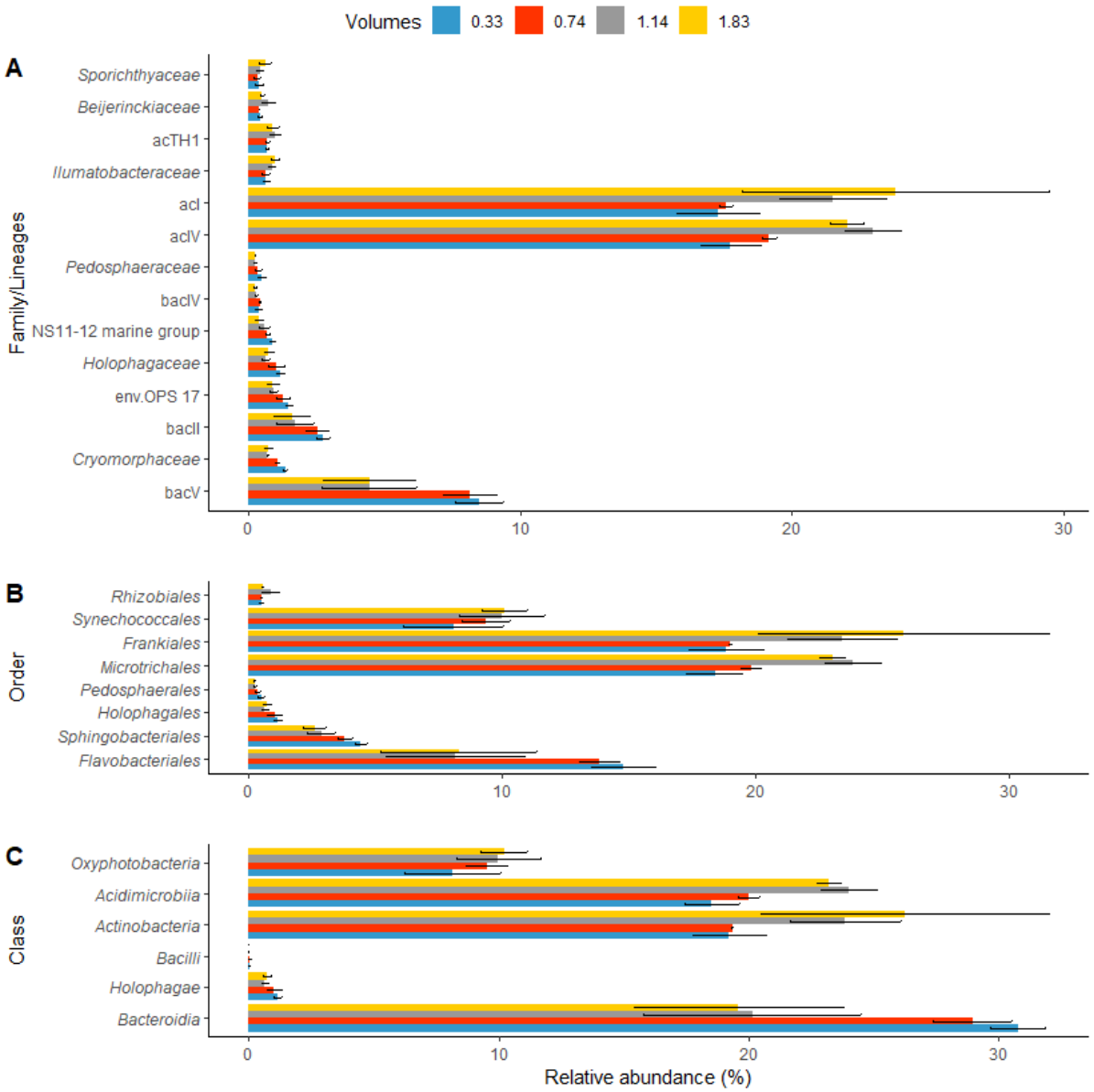

Figure SII.3: Taxa that have their relative abundance significatively impacted by the filtration bias. A: Family/Lineage taxonomic rank, B: Order taxonomic rank, C: Class taxonomic rank Colours indicate the different Volumes/Conditions. The relative abundance is expressed as the mean of the percentage of the proportion of reads affiliated to the taxa of the condition´s replicates. Error bars represent the standard deviation as a measure of the dispersion of the relative abundance between replicates. Note the variation of y axis scales.

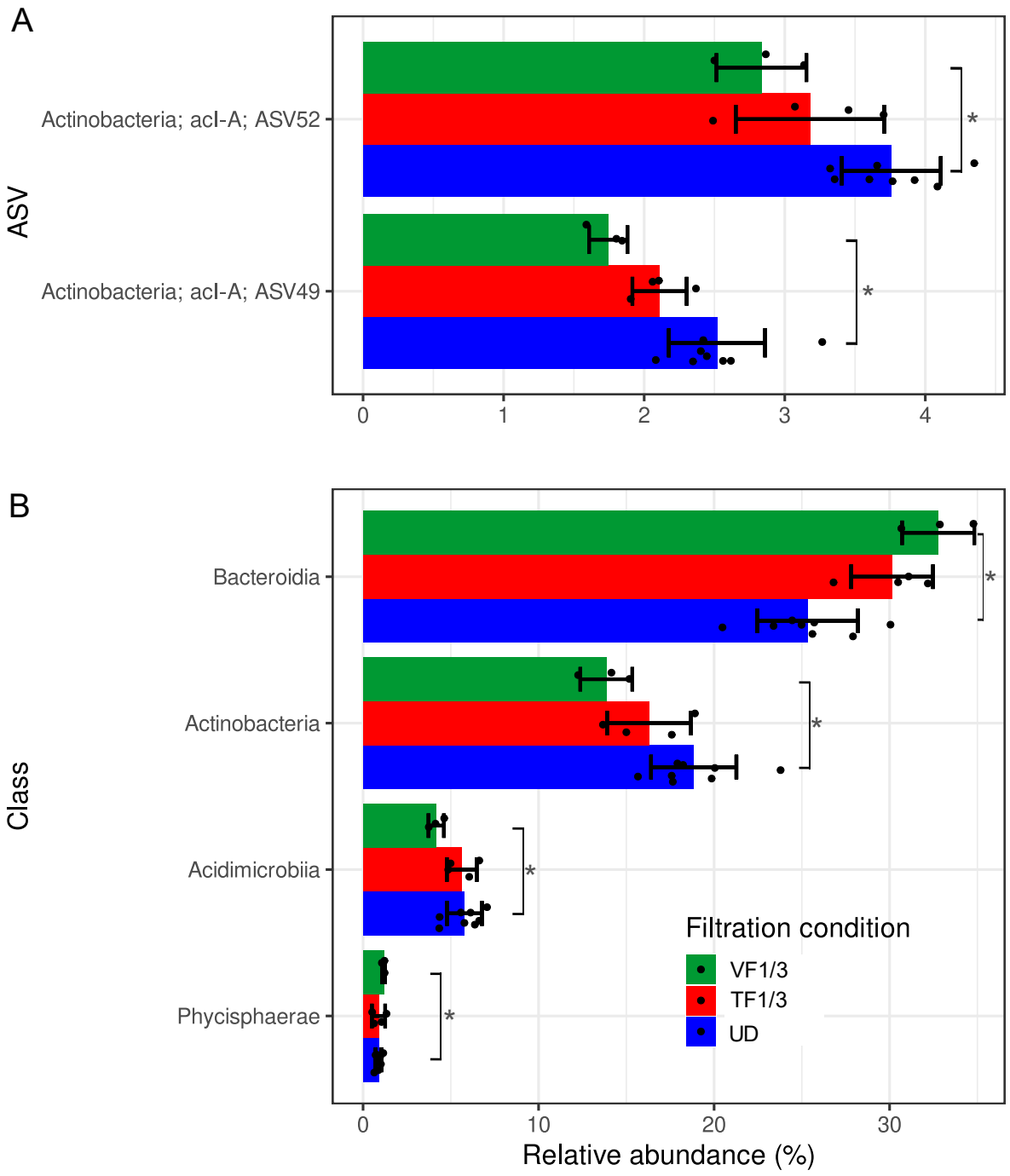

Figure SII.4: Comparison of taxa with their relative abundance significatively change for the fixed volume condition and not for the flowrate filtration device. A: Family affiliated phylotype taxonomic rank, B: Order taxonomic rank, C: Class taxonomic rank. Colours indicate the different conditions. The relative abundance is expressed as the mean of the percentage of the proportion of reads affiliated to the taxa of the condition´s replicates. Error bars represent the standard deviation as a measure of the dispersion of the relative abundance between replicates. Statistical significance is flagged with one star (*). Note the variation of y axis scales.

Table SII.1: Taxa with relative abundance significatively impacted by the filtration bias introduce by the different volume of water filtrated in the first experiment using EdgeR

Taxonomy	logFC	logCPM	F	p-value	FDR
Phylum					
Bacteroidetes	-0.47	17.79	18.44	0.00	0.00
Actinobacteria	0.46	18.61	12.90	0.00	0.01
Class					
Bacteroidia	-0.51	17.89	17.73	0.00	0.00
Actinobacteria	0.49	17.79	15.89	0.00	0.00
Holophagae	-0.58	13.09	15.65	0.00	0.00
Acidimicrobiia	0.40	17.73	10.91	0.00	0.01
Oxyphotobacteria	0.32	16.54	7.60	0.01	0.04
Bacilli	-6.11	7.52	7.00	0.01	0.05
Order					
Flavobacteriales	-0.69	16.73	32.75	0.00	0.00
Frankiales	0.53	17.75	19.42	0.00	0.00
Sphingobacteriales	-0.46	15.04	14.79	0.00	0.00
Microtrichales	0.45	17.71	13.96	0.00	0.00
Holophagales	-0.53	13.09	11.69	0.00	0.01
Pedosphaerales	-0.74	11.68	11.52	0.00	0.01
Synechococcales	0.38	16.52	9.97	0.00	0.02
Rhizobiales	0.50	12.73	9.00	0.00	0.03
Family/Lineage					
bacV	-0.81	15.90	30.61	0.00	0.00
Ilumatobacteraceae	0.72	13.17	22.48	0.00	0.00
Cryomorphaceae	-0.64	13.22	21.04	0.00	0.00
bacVI	-0.53	13.65	17.14	0.00	0.01
acI	0.53	17.65	15.40	0.00	0.01
acIV	0.43	17.67	13.78	0.00	0.02
Pedosphaeraceae	-0.74	11.70	13.69	0.00	0.02
env.OPS 17	-0.51	13.47	13.52	0.00	0.02
bacII	-0.56	14.35	13.33	0.00	0.02
Holophagaceae	-0.54	13.11	12.13	0.00	0.02
NS11-12 marine group	-0.61	12.64	11.76	0.00	0.02
acTH1	0.56	13.04	11.70	0.00	0.02
Beijerinckiaceae	0.64	12.45	10.84	0.00	0.03
Sporichthyaceae	0.74	12.48	10.81	0.00	0.03

Table SII.1: Taxa with relative abundance significatively impacted by the filtration bias introduce by the different volume of water filtrated in the first experiment using EdgeR

Taxonomy	logFC	logCPM	F	p-value	FDR
ASV					
detes*; Bacteroidia; Flavobacteriales*; bacV; unclassified; unclassified; ASV847	-1.00	14.60	137.97	0.00	0.00
Bacteroidetes*; Bacteroidia; Flavobacteriales*; bacV; unclassified; unclassified; ASV857	-0.83	14.82	122.06	0.00	0.00
Actinobacteria*; Actinobacteria; Frankiales*; acI; acI-B; acI-B1; ASV244	0.40	15.96	118.95	0.00	0.00
Actinobacteria*; Acidimicrobiia; Microtrichales*; acIV; acIV-A; lluma-A1. ASV121	0.29	16.28	94.53	0.00	0.00
Bacteroidetes*; Bacteroidia; Chitinophagales*; bacI; bacI-A; bacI-A1. ASV703	-0.47	14.71	34.30	0.00	0.00
Actinobacteria*; Actinobacteria; Frankiales*; acI; acI-A; acI-A7; ASV215	0.42	14.55	22.76	0.00	0.00
Actinobacteria*; Acidimicrobiia; Microtrichales*; acIV; acIV-A; lluma-A2; ASV154	0.22	15.46	19.47	0.00	0.00
Actinobacteria*; Actinobacteria; Frankiales*; acI; acI-A; acI-A6. ASV236	0.48	14.12	17.38	0.00	0.00
Actinobacteria*; Actinobacteria; Frankiale*s; acI; acI-B; acI-B1; ASV250	0.49	14.09	17.25	0.00	0.00
Cyanobacteria*; Oxyphotobacteria; Synechococcales; Cyanobiaceae*; Cyanobium PCC-6307; ASV15	-8.08	9.05	16.83	0.00	0.00
Acidobacteria*; Holophagae; Holophagales; Holophagaceae*; marine group; uncultured bacterium; ASV4	-7.82	8.82	14.10	0.00	0.02
Bacteroidetes*; Bacteroidia; Flavobacteriales; Cryomorphaceae*; uncultured; uncultured bacterium; ASV788	-0.79	13.10	13.63	0.00	0.02
Proteobacteria*; Gammaproteobacteria; Betaproteobacteriales*; betI; betI-A; unclassified; ASV353	-7.68	8.71	12.88	0.00	0.02
Proteobacteria*; Gammaproteobacteria; Betaproteobacteriales*; betI; betI-B; Rhodo; ASV319	-3.48	10.27	12.72	0.00	0.02
Actinobacteria*; Actinobacteria; Frankiales*; acI; acI-A; acI-A7; ASV217	2.99	10.59	11.74	0.00	0.04

Table SII.2: Taxa having their relative abundance significatively changed by the difference of microbial plankton density in the fixed volume conditions of the second experiment using EdgeR

Taxonomy	logFC	logCPM	F	PValue	FDR
Phylum					
Actinobacteria	-0.44	17.80	15.44	0.00	0.01
Bacteroidetes	0.38	18.08	10.46	0.00	0.02
Planctomycetes	0.36	13.76	8.69	0.01	0.03
Class					
Phycisphaerae	0.48	13.24	11.20	0.00	0.03
Bacteroidia	0.43	18.07	10.82	0.00	0.03
Acidimicrobiia	-0.42	15.70	9.77	0.00	0.03
Actinobacteria	-0.38	17.41	8.97	0.00	0.03
Family/Lineage					
acI	-0.59	17.40	16.87	0.00	0.02
acIV	-0.60	15.80	16.71	0.00	0.02
bacV	0.47	13.58	13.75	0.00	0.03
Flavobacteriaceae	0.44	14.80	12.54	0.00	0.04
ASV					
Actinobacteria; *Actinobacteria; Frankiale*; acI; acI-A; ASV215	-0.45	15.10	23.08	0.00	0.048
Actinobacteria; *Actinobacteria; Frankiale*; acI; acI-A; ASV236	-0.57	14.50	22.80	0.00	0.048

Chapter III: Nano- and pico-plankton succession in Upper and Lower Lake Constance followed by 18S and 16S rDNA amplicon sequencing: Same seasonal dynamics but different communities

Table SIII.1: Date of sampling campaign by sampling sites.

Sampling date		Seasons
ULC	LLC	
13.03.2018	15.03.18	Winter
27.03.2018	29.03.18	Spring
10.04.2018	14.04.18	Spring
27.04.2018	26.04.18	Spring
08.05.2018	09.05.18	Spring
22.05.2018	24.05.18	Spring
05.06.2018	07.06.18	Spring
19.06.2018	21.06.18	Summer
03.07.2018	05.07.18	Summer
17.07.2018	19.07.18	Summer
31.07.2018	02.08.18	Summer
14.08.2018	16.08.18	Summer
28.08.2018	30.08.18	Summer
11.09.2018	13.09.18	Summer
21.09.2018	01.10.18	Autumn
09.10.2018	11.10.18	Autumn
23.10.2018	25.10.18	Autumn
06.11.2018	08.11.2018	Autumn
22.11.2018	20.11.2018	Autumn
04.12.2018	06.12.2018	Autumn
18.12.2018	20.12.2018	Autumn
11.01.2019	10.01.2019	Winter
29.01.2019	31.01.2019	Winter
12.02.2019	14.02.20199	Winter
26.02.2019	27.02.2019	Winter
12.03.2019	19.03.2019	Winter

Table SIII.2: **Average total relative abundance of the main phyla composing the three plankton community calculated in the full dataset.**

Size class	Microbial plankton	Phyla	Relative abundance (%)		Relative abundance (%)
			ULC	LLC	
180 - 5 µm	18S-NP	Ciliophora	41.23	41.02	41.13
		Myzozoa	19.63	10.85	15.24
		Bacillariophyta	16.95	12.29	14.62
		Ochrophyta	7.04	13.14	10.09
		Cryptophyta	2.52	6.73	4.63
		Cercozoa	3.14	5.34	4.24
		Chlorophyta	4.42	3.44	3.93
5 - 0.2 µm	18S-PP	Cryptophyta	17.32	23.9	20.61
		Ciliophora	16.73	14.74	15.74
		Myzozoa	15.96	11.31	13.63
		Ochrophyta	16.97	9.47	13.22
		Haptophyta	7.38	11.01	9.2
		Cercozoa	7.59	7.46	7.53
		Chlorophyta	3.78	6.55	5.17
	16S-PP	Actinobacteria	43.62	47.1	45.36
		Bacteroidota	21.96	26.54	24.25
		Proteobacteria	14.52	15.68	15.1
		Cyanobacteria	11.13	5.53	8.33
		Planctomycetes	2.55	1.49	2.02
		Chloroflexi	1.74	1.12	1.43
		Verrucomicrobia	1.14	1.4	1.27
		Acidobacteria	1.94	0.31	1.13

Table SIII.3: **Number of unique ASVs and their minimal and maximal relative abundance at a specific date grouped by phyla and sampling site.** ULC: upper lake Constance. LLC: lower lake Constance. Relative abundance is calculated in percentage of the total plankton community.

Size class	Microbial plankton	Phyla	Sampling site	ASV number	Relative abundance (%)	
					Minimal	Maximal
180 - 5 µm	18S-NP	Myzozoa	ULC	98	0.24	17.51
			LLC	66	0.02	7.08
		Ochrophyta	ULC	50	0.02	4.89
			LLC	77	0.11	4.21
		Ciliophora	ULC	59	0.1	7.43
			LLC	71	0.14	7.5
		Cercozoa	ULC	31	0	2.51
			LLC	44	0.04	8.28
		Blastocladiomycota	ULC	18	0	3.36
			LLC	10	0	0.79
		Chlorophyta	ULC	12	0	0.42
			LLC	85	0.13	9.72
5 - 0.2 µm	18S-PP	Myzozoa	ULC	123	0.58	23.36
			LLC	93	0.1	12.05
		Ochrophyta	ULC	107	0.41	11.43
			LLC	143	0.26	6.92
		Ciliophora	ULC	42	0.05	3.42
			LLC	69	0.5	3.73
		Blastocladiomycota	ULC	16	0	2.68
			LLC	6	0	0.64
		Chlorophyta	ULC	18	0.01	0.94
			LLC	76	0.21	6.29
		Choanozoa	ULC	21	0	9.26
			LLC	63	0.01	2.39
		Bigyra	ULC	11	0	0.7
			LLC	93	0.16	9.44
	16S-PP	Acidobacteriota	ULC	35	0	3.08
			LLC	3	0	0.14
		Actinobacteriota	ULC	22	0	0.6
			LLC	35	0.05	1
		Bacteroidota	ULC	41	0.09	3.88
			LLC	123	0.12	5.52
		Cyanobacteria	ULC	4	0	0.05
			LLC	7	0	2.94
		Proteobacteria	ULC	126	0.1	4.48
			LLC	206	0.23	6.69
		Verrucomicrobiota	ULC	20	0.01	0.72
			LLC	29	0.07	1.16

Table SIII.4: Number of taxa differentially abundant at the taxonomic rank tested. Results are separated by plankton type, season and by sampling site.

Size class	Microbial plankton	Seasons	Sampling site	Number of differentially abundant taxa				
				ASVs	Family	Order	Class	Phyla
180 - 5 µm	18S-NP	Winter	ULC	11	1	1	1	1
			LLC	0	0	0	0	0
		Spring	ULC	12	3	3	1	1
			LLC	2	0	0	0	0
		Summer	ULC	7	1	2	0	1
			LLC	9	0	1	0	0
		Autumn	ULC	7	2	3	1	2
			LLC	9	2	2	2	1
5 - 0.2 µm	18S-PP	Winter	ULC	7	2	2	0	0
			LLC	6	1	0	1	0
		Spring	ULC	0	0	0	0	0
			LLC	4	0	0	0	0
		Summer	ULC	4	1	1	1	1
			LLC	11	0	0	0	0
		Autumn	ULC	0	0	0	0	0
			LLC	13	0	0	0	0
	16S-PP	Winter	ULC	65	14	9	5	2
			LLC	45	16	14	8	4
		Spring	ULC	0	1	1	0	0
			LLC	6	2	1	0	0
		Summer	ULC	0	1	1	0	0
			LLC	7	0	0	0	0
		Autumn	ULC	12	1	0	0	0
			LLC	17	2	1	0	0

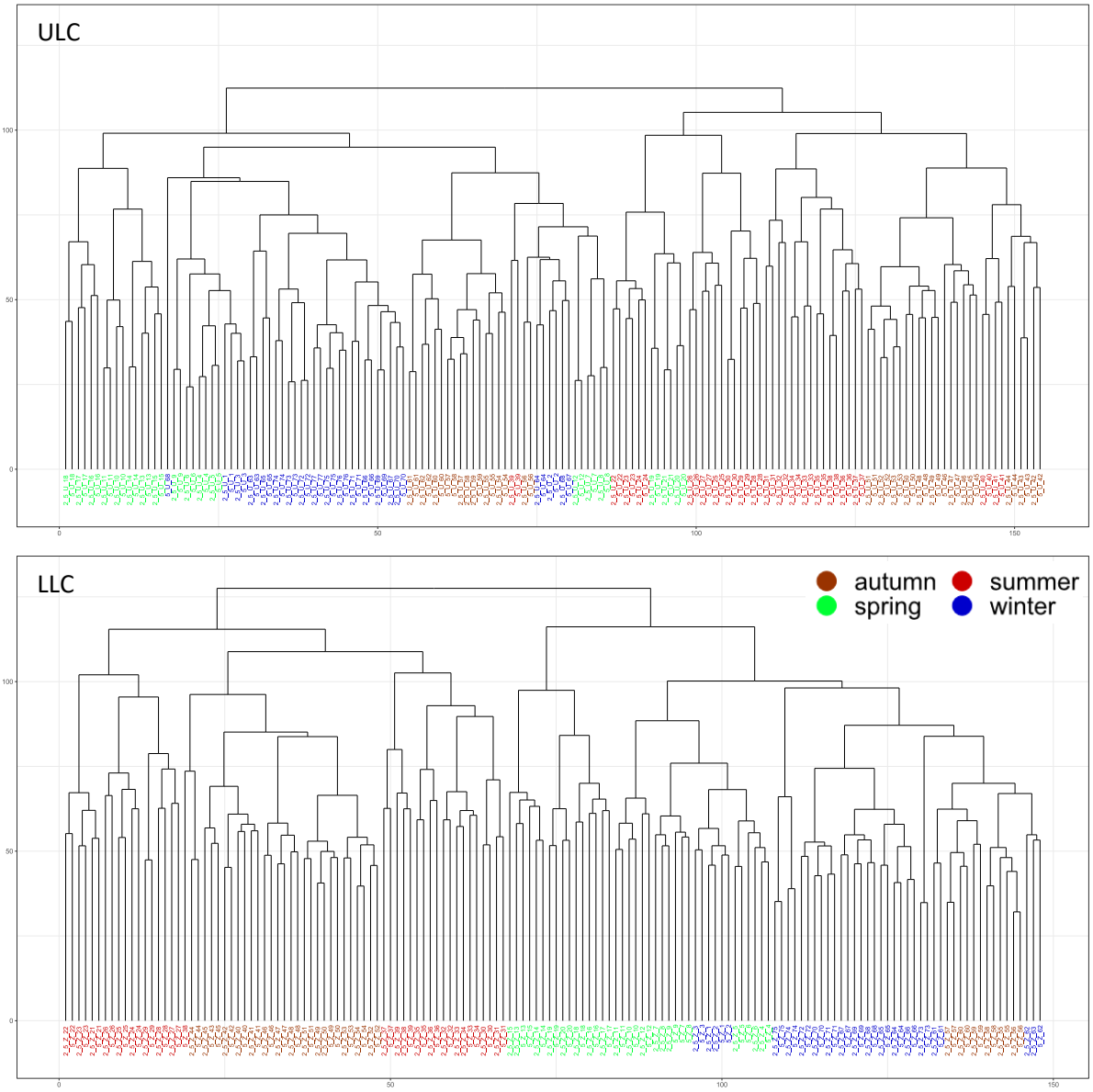

Figure SIII.1: Dendrogram presenting the similarity between samples constituting the 18S-NP dataset sequenced by two independent sequencing runs. The top panel correspond to ULC and bottom panel LLC. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

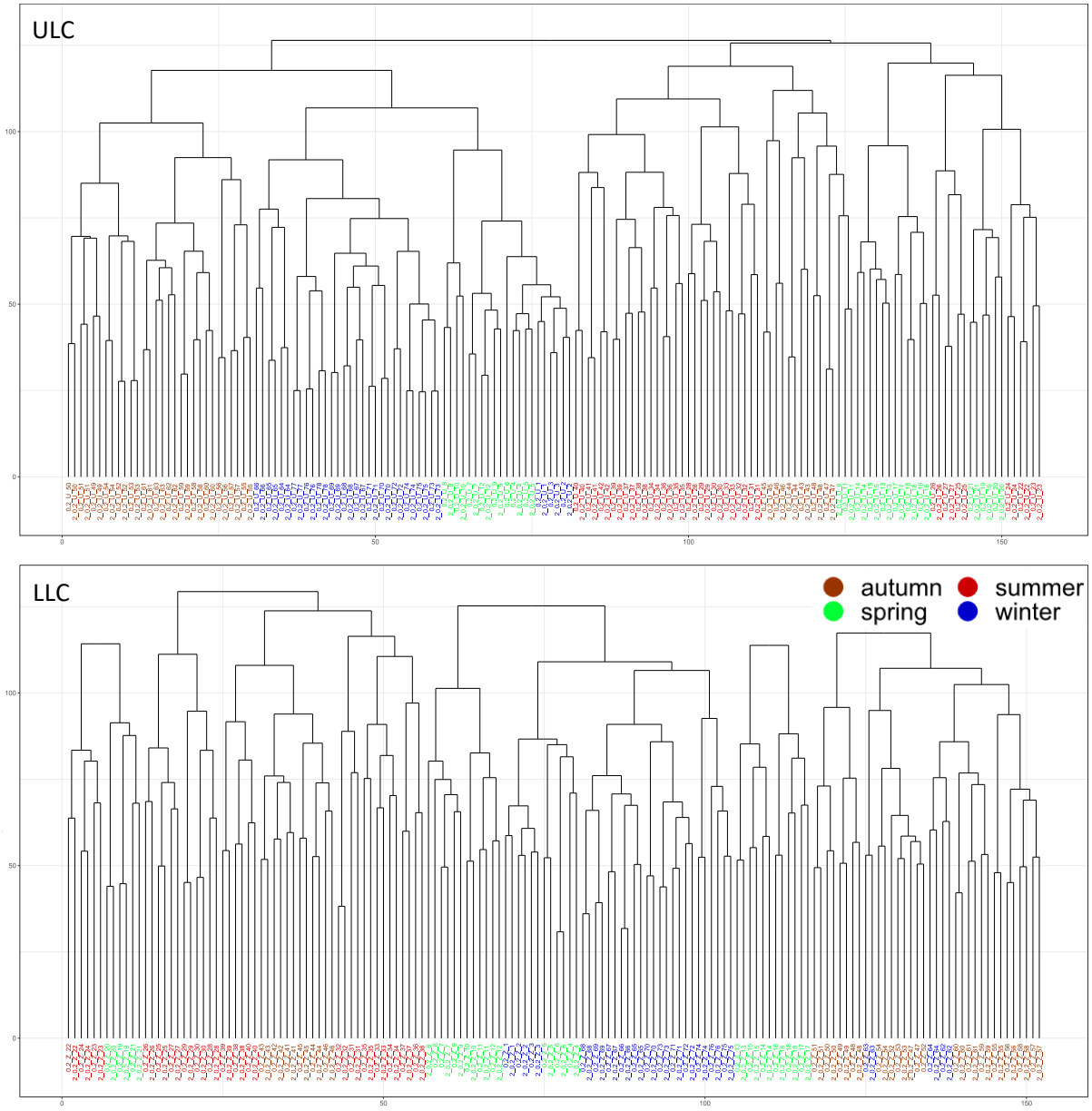

Figure SIII.2: Dendrogram presenting the similarity between samples constituting the 18S-PP dataset sequenced by two independent sequencing runs. The top panel correspond to ULC and bottom panel LLC. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

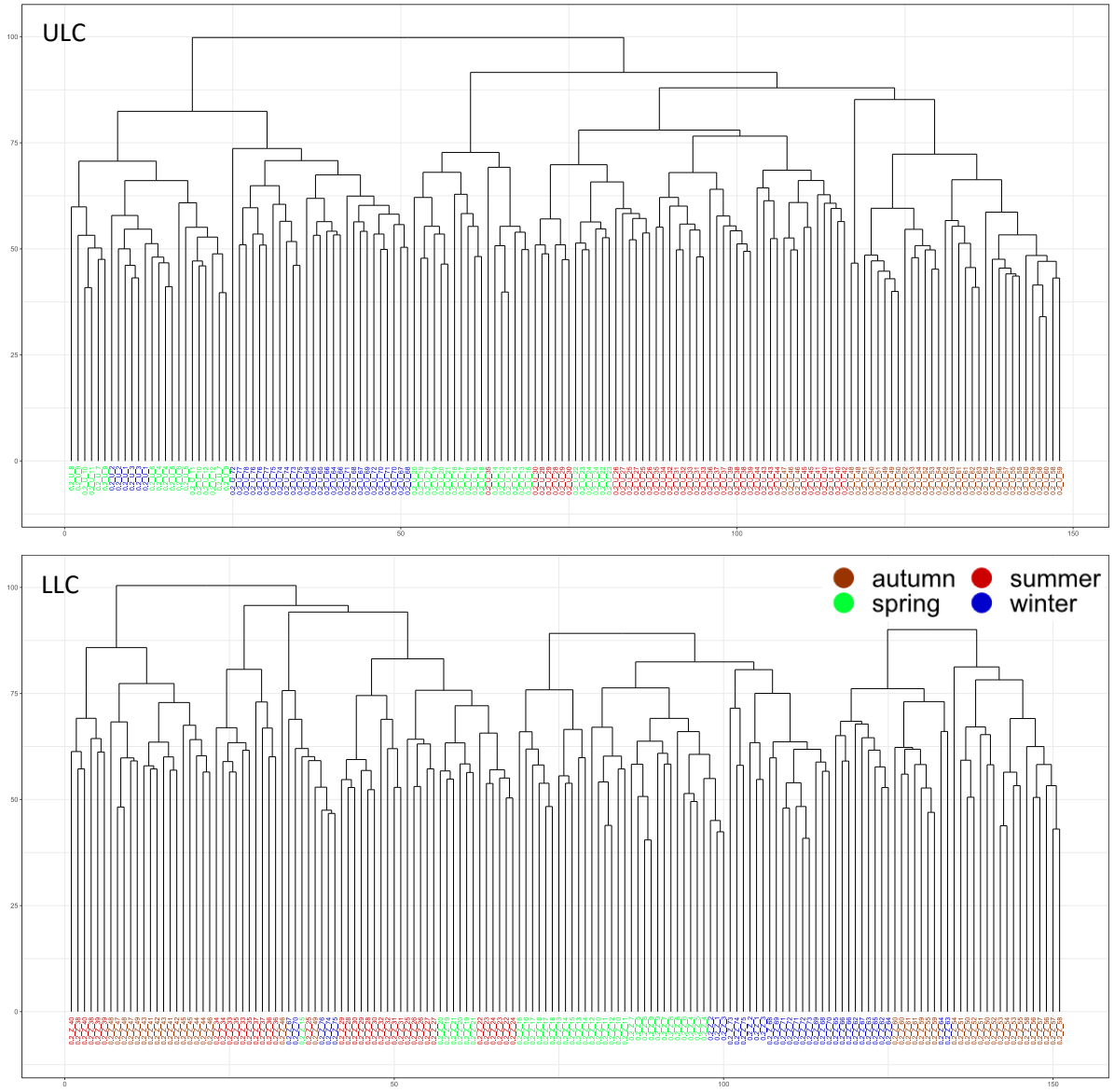

Figure SIII.3: Dendrogram presenting the similarity between samples constituting the 16S-PP dataset sequenced by two independent sequencing runs. The top panel correspond to ULC and bottom panel LLC. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

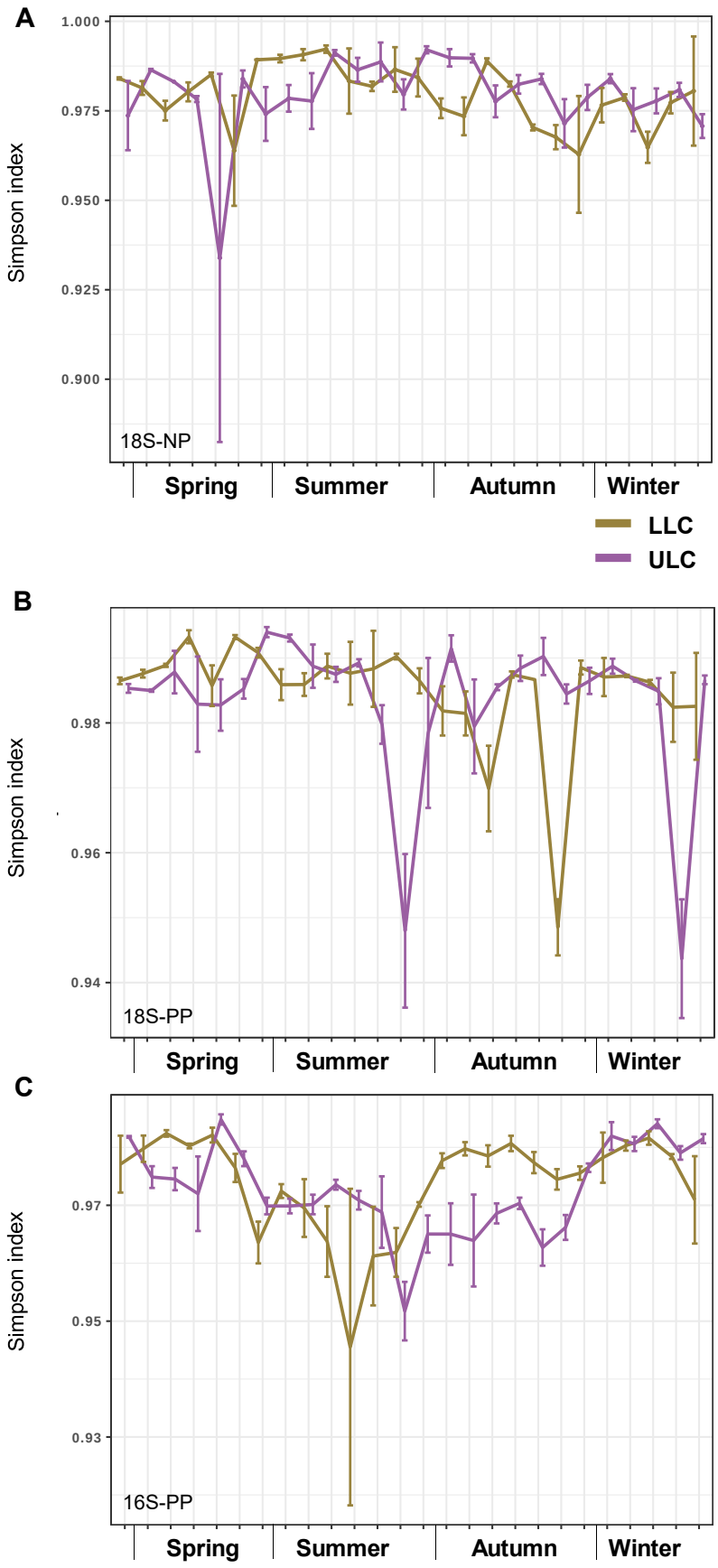

Figure SIII.4: Simpson diversity index diversity calculated per sampling site and across the sampling dates. The purple lines correspond to the Upper Lake Constance (ULC) and the gold lines correspond to the Lower Lake Constance (LLC) sampling site. The upper panel (A) shows the indices for the eukaryotic nanoplankton (18S-NP dataset), the central panel (B) for the eukaryotic picoplankton (18S-PP dataset) and the lower panel (C) for the prokaryotic picoplankton (16S-PP dataset). The x-axis represents the different seasons (for exact sampling dates, please see Table SIII.1). Note that the y-axes have different scale for each of the three graphs.

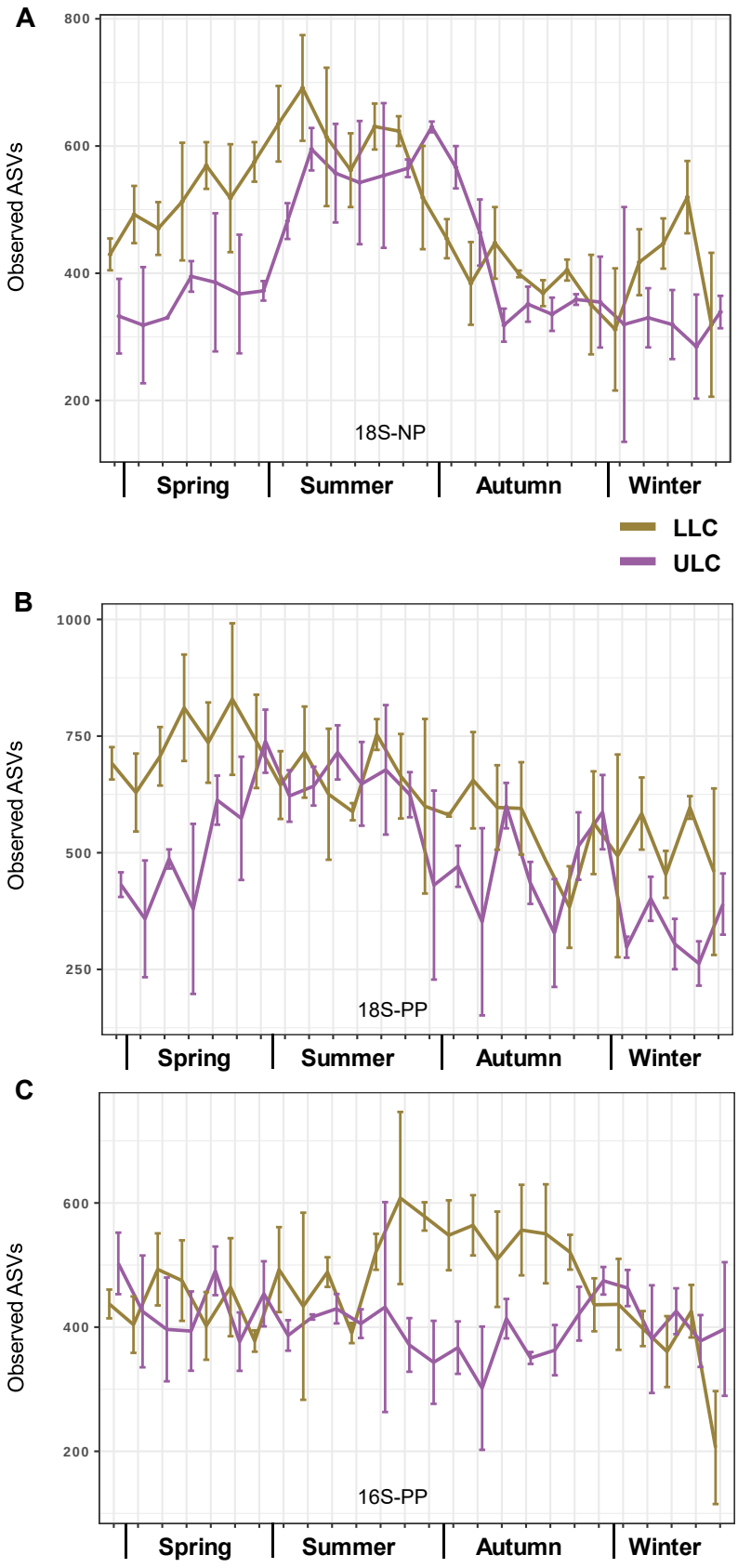

Figure SIII.5: Observed ASVs diversity calculated per sampling site and across the sampling dates. The purple lines correspond to the Upper Lake Constance (ULC) and the gold lines correspond to the Lower Lake Constance (LLC) sampling site. The upper panel (A) shows the indices for the eukaryotic nanoplankton (18S-NP dataset), the central panel (B) for the eukaryotic picoplankton (18S-PP dataset) and the lower panel (C) for the prokaryotic picoplankton (16S-PP dataset). The x-axis represents the different seasons (for exact sampling dates, please see Table SIII.1). Note that the y-axes have different scale for each of the three graphs.

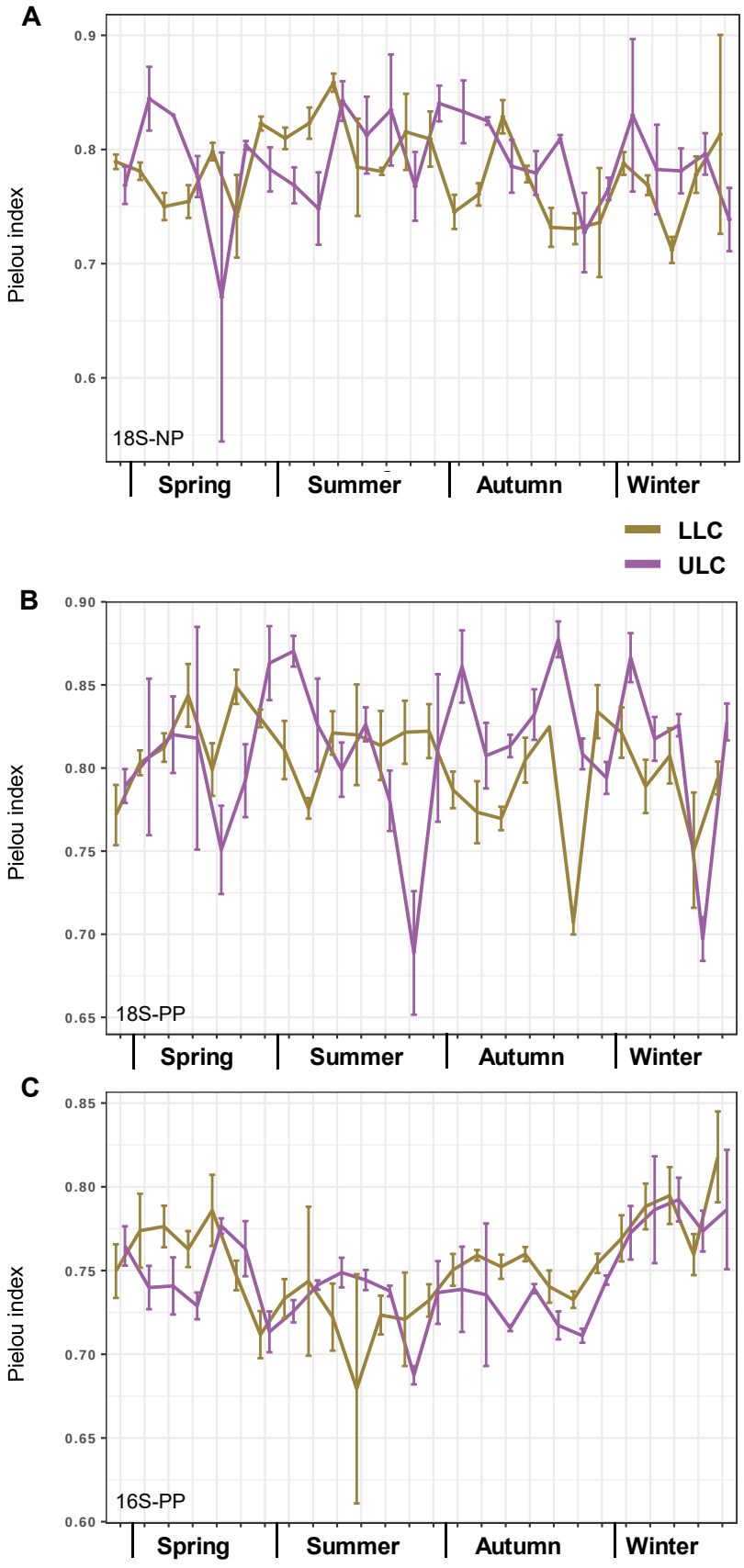

Figure SIII.6: Pielou evenness index diversity calculated per sampling site and across the sampling dates. The purple lines correspond to the Upper Lake Constance (ULC) and the gold lines correspond to the Lower Lake Constance (LLC) sampling site. The upper panel (A) shows the indices for the eukaryotic nanoplankton (18S-NP dataset), the central panel (B) for the eukaryotic picoplankton (18S-PP dataset) and the lower panel (C) for the prokaryotic picoplankton (16S-PP dataset). The x-axis represents the different seasons (for exact sampling dates, please see Table SIII.1). Note that the y-axes have different scale for each of the three graphs.

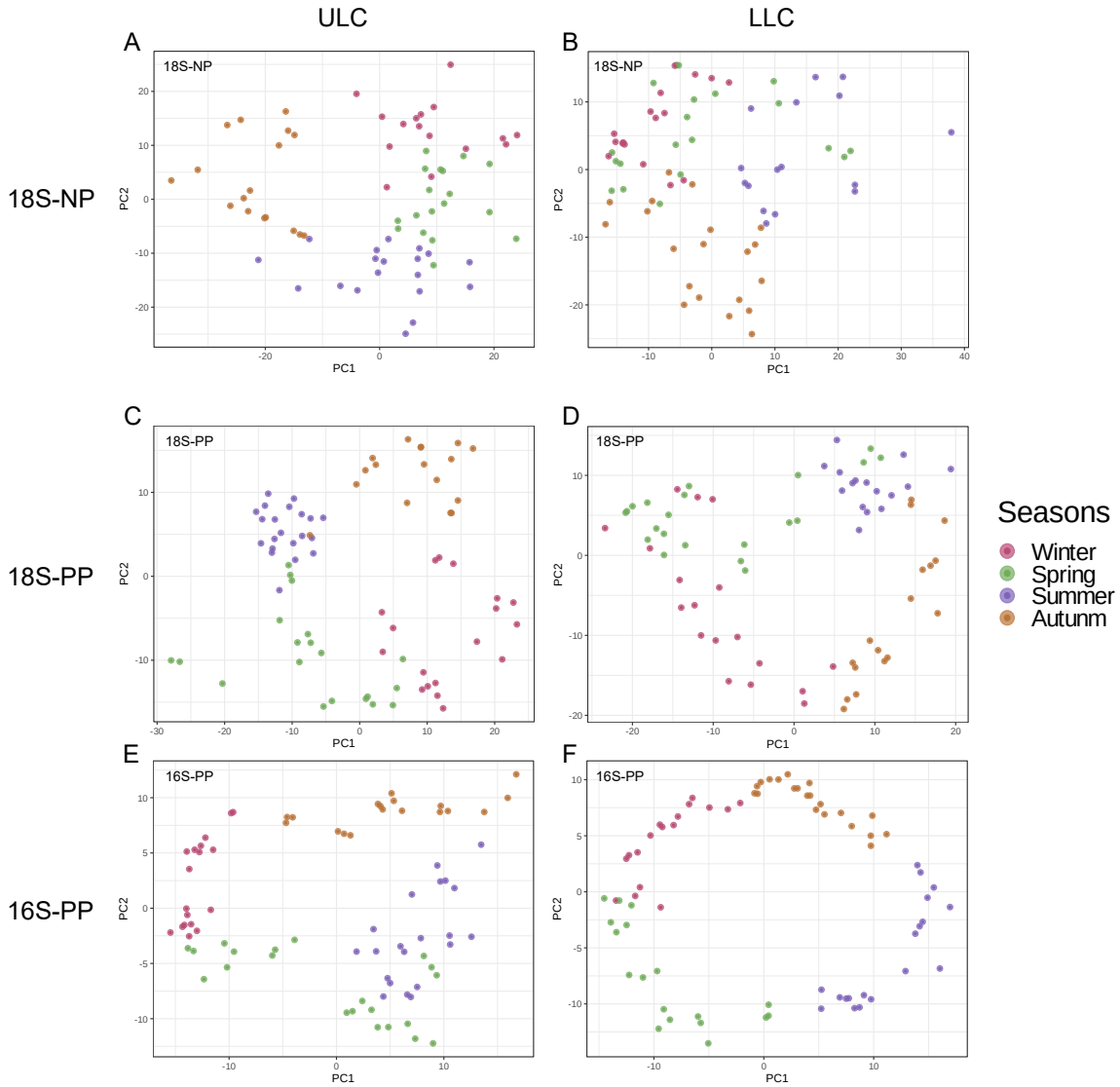

Figure SIII.7: Temporal dynamic of the microplankton community observed in ULC and LLC independently. Dissimilarity matrix calculated based on the phylogenetic isometric log-ratio (PhilR) transformed Euclidean distance matrix and visualize by Principal Component Analysis (PCA). Colors indicate the seasons. The top panels (AB) show the data for the eukaryotic nanoplankton (18S-NP). The centrals panel (CD) show the data for the eukaryotic picoplankton (18S-PP). The lower panels (EF) show the data for the prokaryotic picoplankton (16S-PP). The right panels (A - C - E) show the ULC temporal biodiversity and the left panels (B - D - F) show the LLC temporal biodiversity.

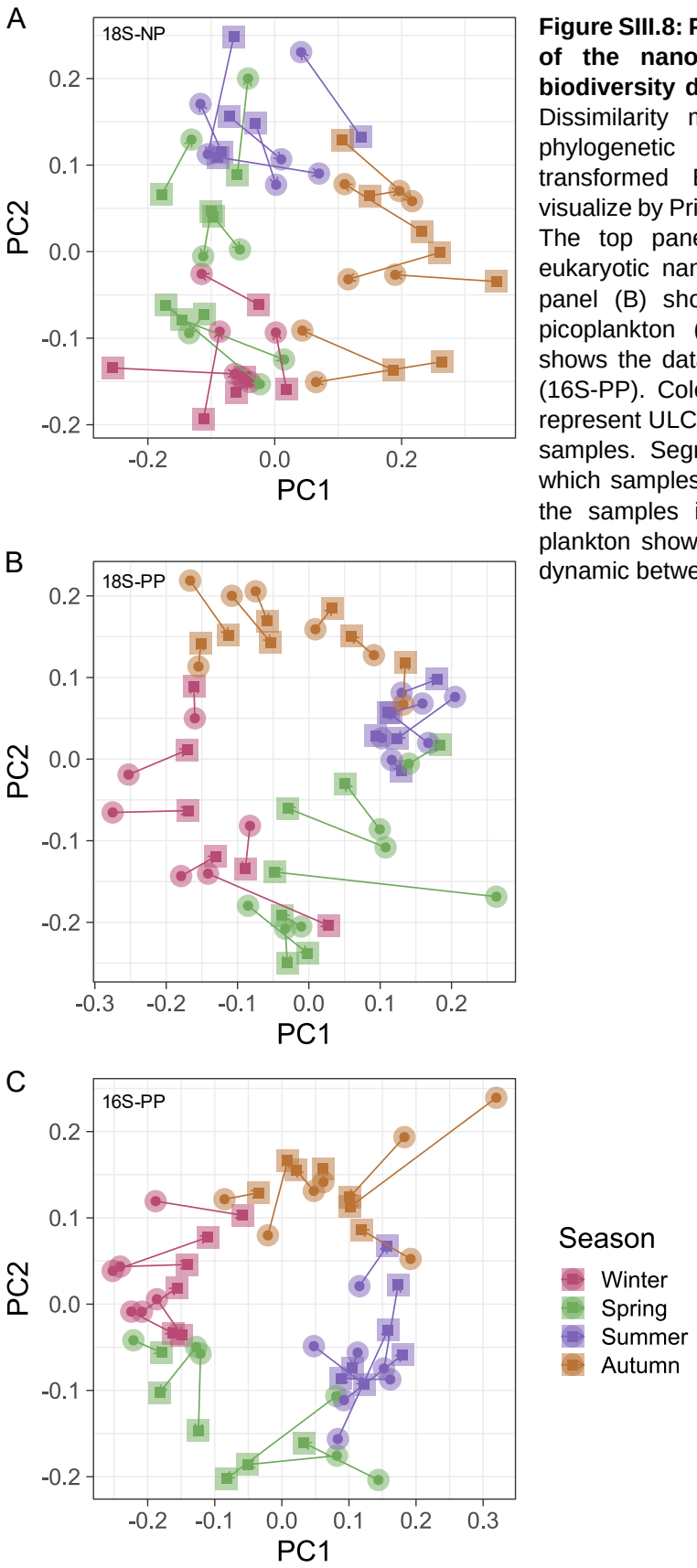

Figure SIII.8: Procrustes analysis visualization of the nano- and picoplankton temporal biodiversity dynamic between ULC and LLC. Dissimilarity matrix calculated based on the phylogenetic isometric log-ratio (PhilR) transformed Euclidean distance matrix and visualize by Principal Component Analysis (PCA). The top panel (A) shows the data for the eukaryotic nanoplankton (18S-NP). The central panel (B) shows the data for the eukaryotic picoplankton (18S-PP). The lower panel (C) shows the data for the prokaryotic picoplankton (16S-PP). Colors indicate the seasons. Square represent ULC samples and circle represent LLC samples. Segment connecting points indicates which samples from ULC dataset correspond to the samples in the LLC dataset. Each three plankton showed high concordance of temporal dynamic between ULC and LLC

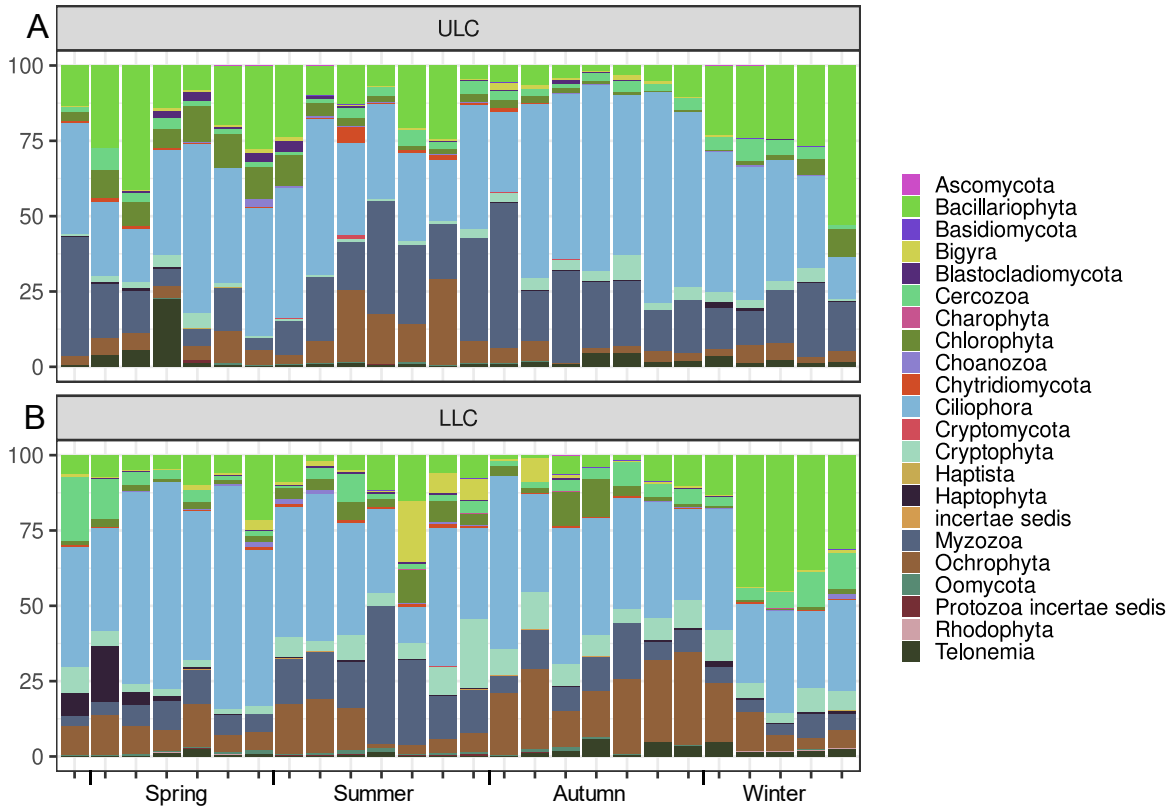

Figure SIII.9: Temporal relative abundance dynamic of the total 18S-NP community at the phylum taxonomic rank. Y-axis represent the relative abundance in percentage and x-axis represent the temporality with the different seasons. The panel A correspond to the community composition of ULC and panel B correspond to the community composition of LLC. Colors correspond to the phylum and the color pattern between 18S-NP and 18S-PP dataset (Figure SIII.9 and Figure SIII.10) is the same.

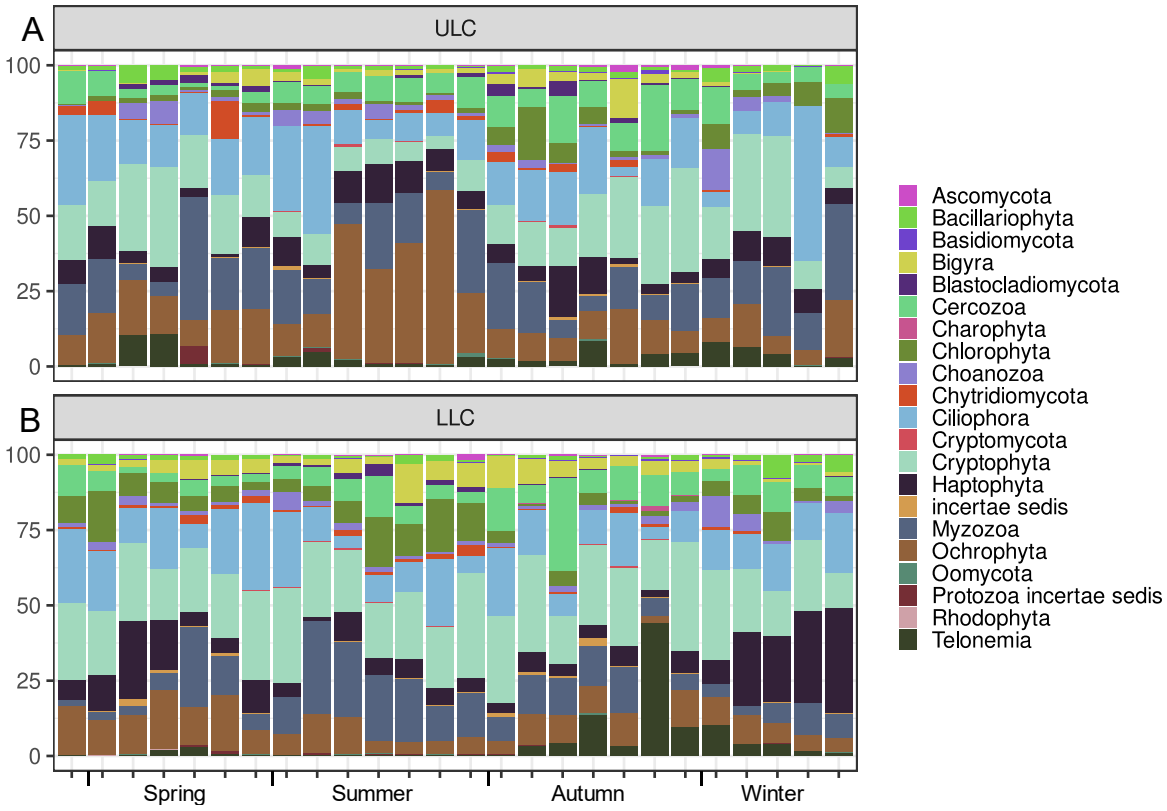

Figure SIII.10: Temporal relative abundance dynamic of the total 18S-NP community at the phylum taxonomic rank. Y-axis represent the relative abundance in percentage and x-axis represent the temporality with the different seasons. The panel A correspond to the community composition of ULC and panel B correspond to the community composition of LLC. Colors correspond to the phylum and the color pattern between 18S-NP and 18S-PP dataset (Figure SIII.9 and Figure SIII.10) is the same.

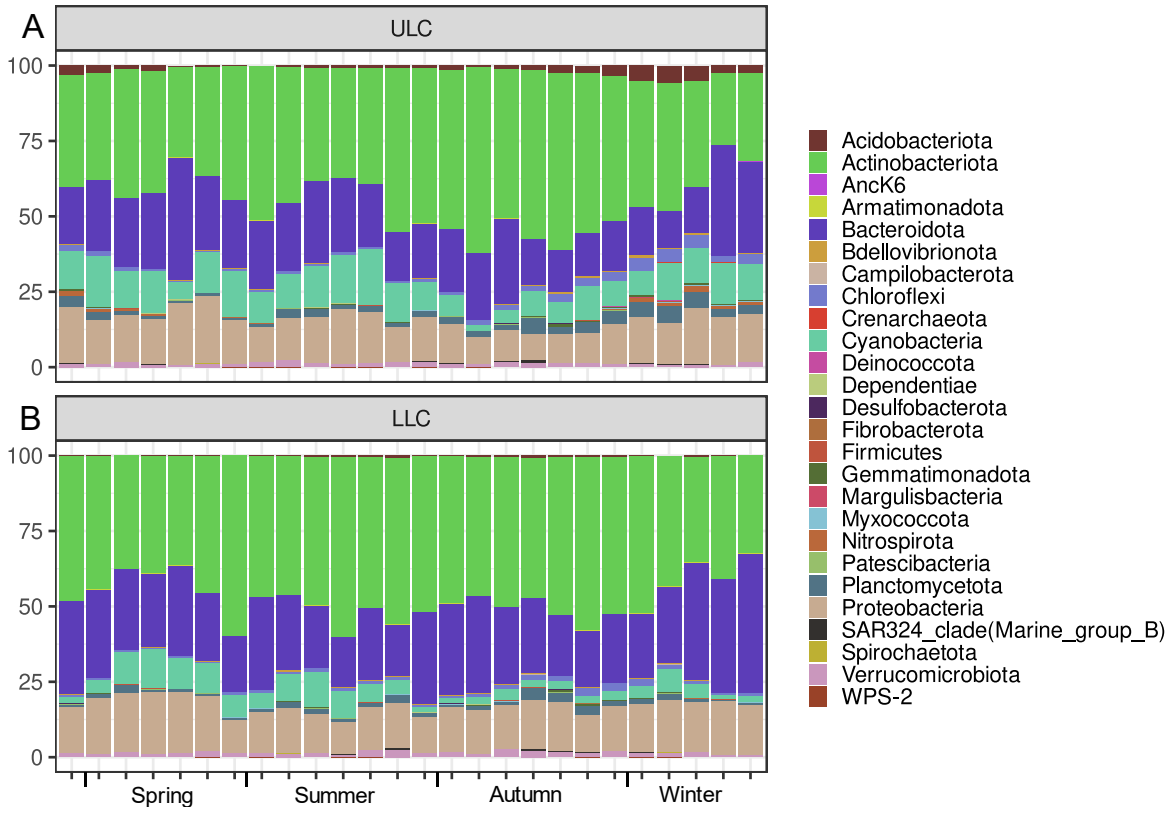

Figure SIII.11: Temporal relative abundance dynamic of the total 16S-PP community at the phylum taxonomic rank. Y-axis represent the relative abundance in percentage and x-axis represent the temporality with the different seasons. The panel A correspond to the community composition of ULC and panel B correspond to the community composition of LLC. Colors correspond to the phylum.

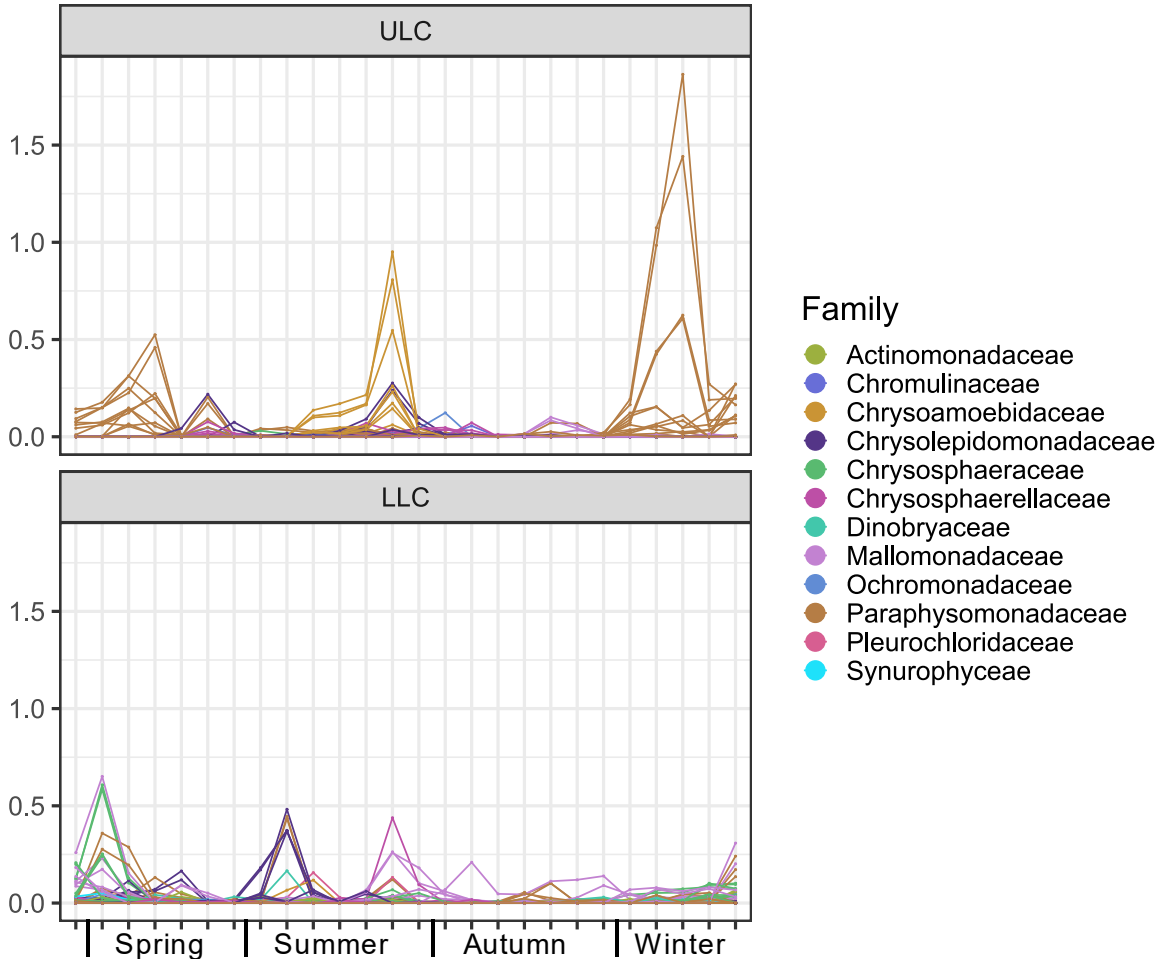

Figure SIII.12: Relative abundance temporal variation of the 18S-NP unique ASVs affiliated to Ochrophyta. Each line represented a single ASV. ASVs are colored by family taxonomic affiliation. The panel A correspond to ULC and panel B correspond to LLC

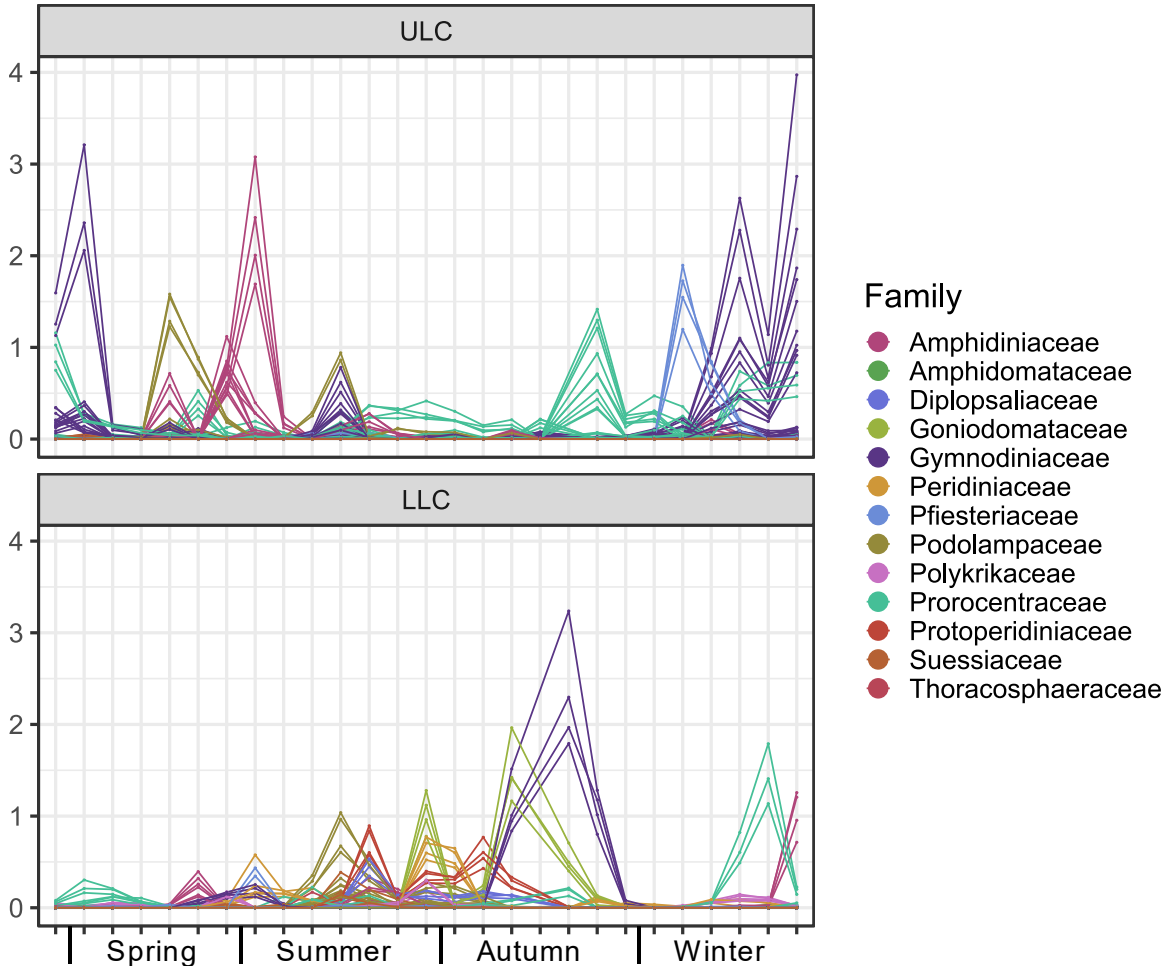

Figure SIII.13: Relative abundance temporal variation of the 18S-PP unique ASVs affiliated to Myzozoa. Each line represented a single ASV. ASVs are colored by family taxonomic affiliation. The panel A correspond to ULC and panel B correspond to LLC

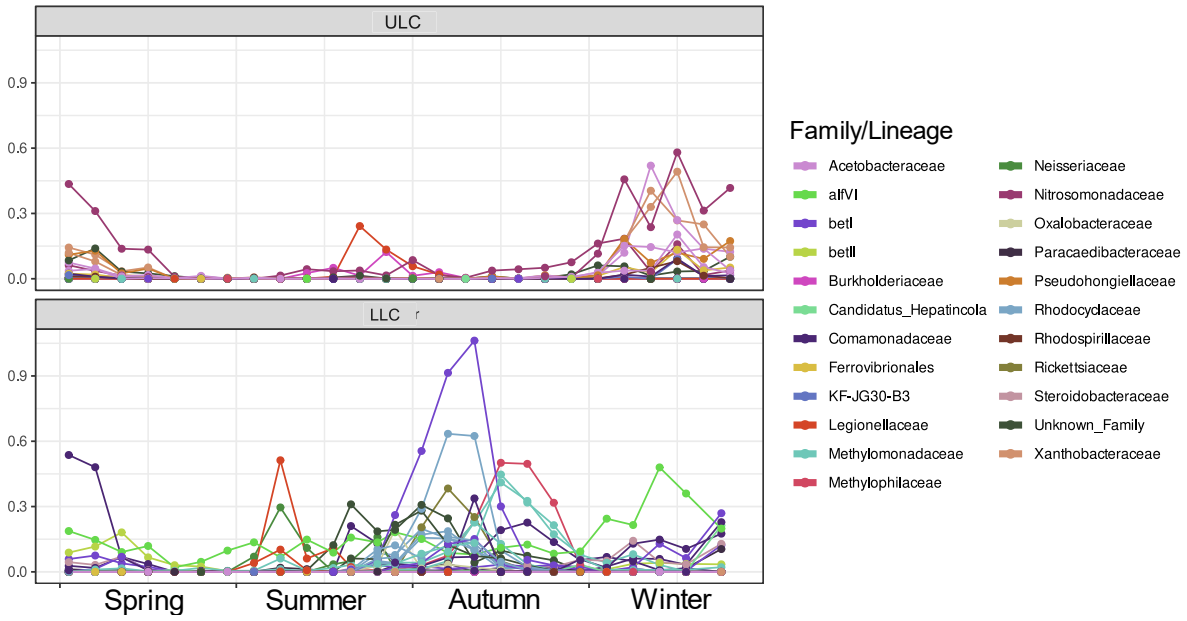

Figure SIII.14: Relative abundance temporal variation of the 16S-PP unique ASVs affiliated to Proteobacteria. Each line represented a single ASV. ASVs are colored by family taxonomic affiliation. The panel A correspond to ULC and panel B correspond to LLC The unique ULC and LLC proteobacteria community were composed of a high number of ASVs (Table SIII.3) affiliated to a total of 71 families. For a better visualization, a filtration of the ASVs not presenting a relative abundance above 0.2 at one time point was applied

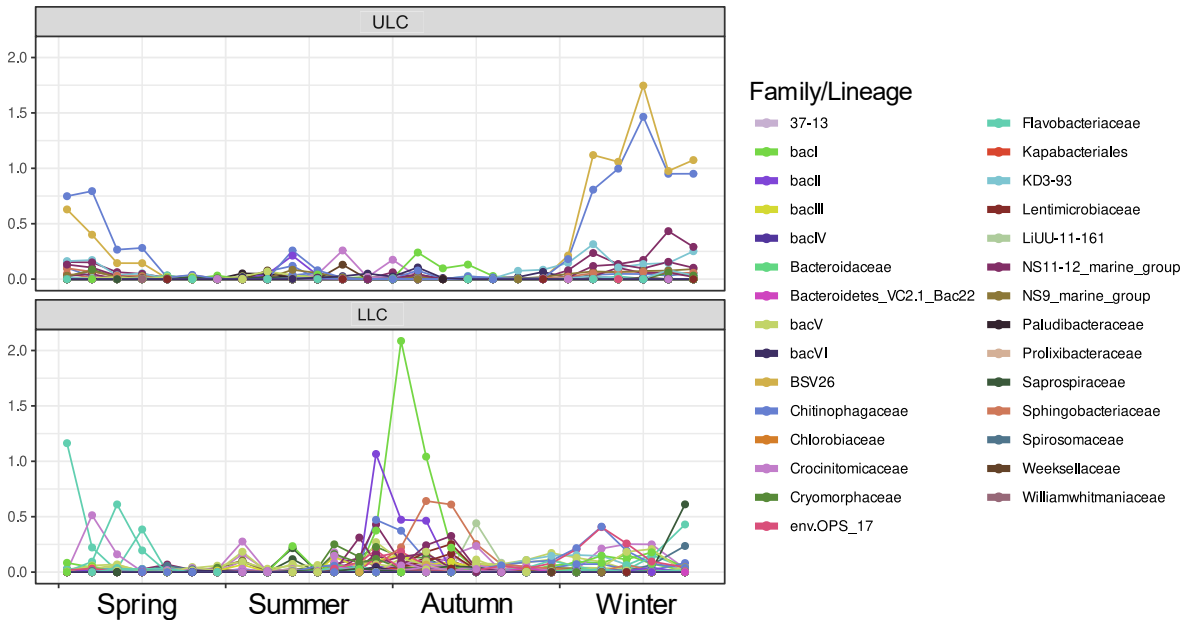

Figure SIII.15: Relative abundance temporal variation of the 16S-PP unique ASVs affiliated to Bacteroidota. Each line represented a single ASV. ASVs are colored by family taxonomic affiliation. The panel A correspond to ULC and panel B correspond to LLC

Chapter IV: Lake Constance microbial plankton dynamic is heavily impacted by the vertical mixing and ASVs associations: a Joint species distribution modelling approach

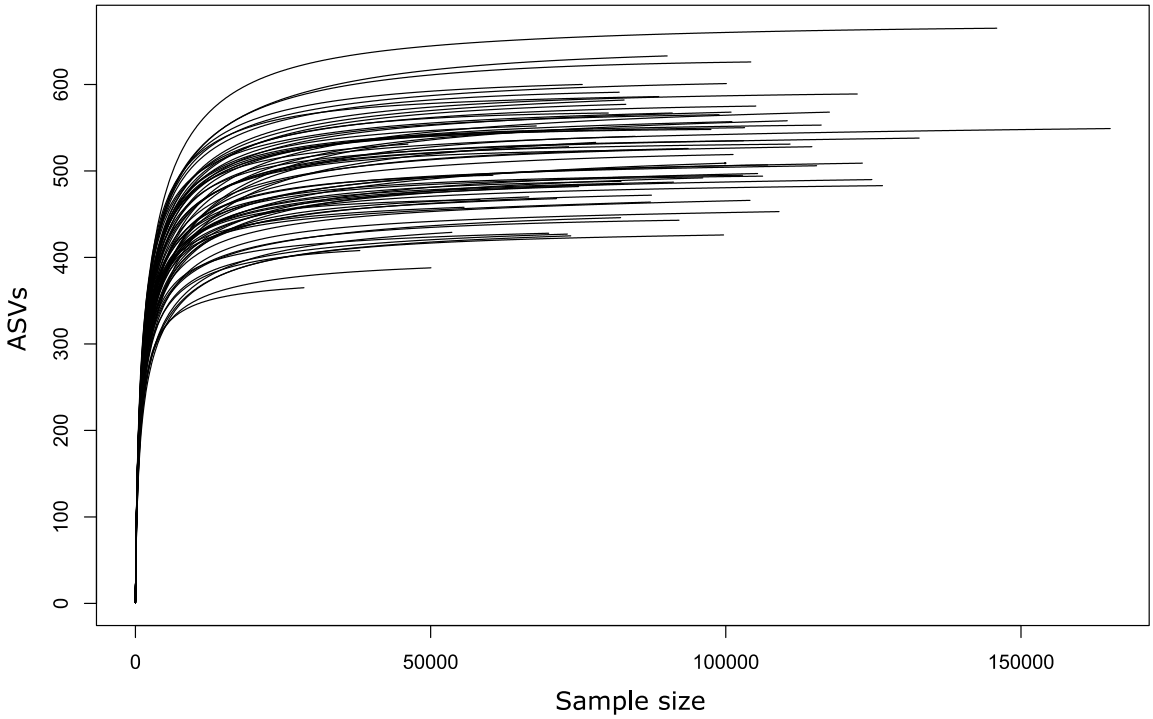

Figure SIV.1: Rarefaction curve of the samples. X-axis indicate the number of sub-samples reads randomly chosen and y-axis the number of unique ASVs detected. The minimum number of reads in a sample is 28540, which is enough to capture the majority of the sequence ASVs in our samples as all samples are close or reach a plateau phase.

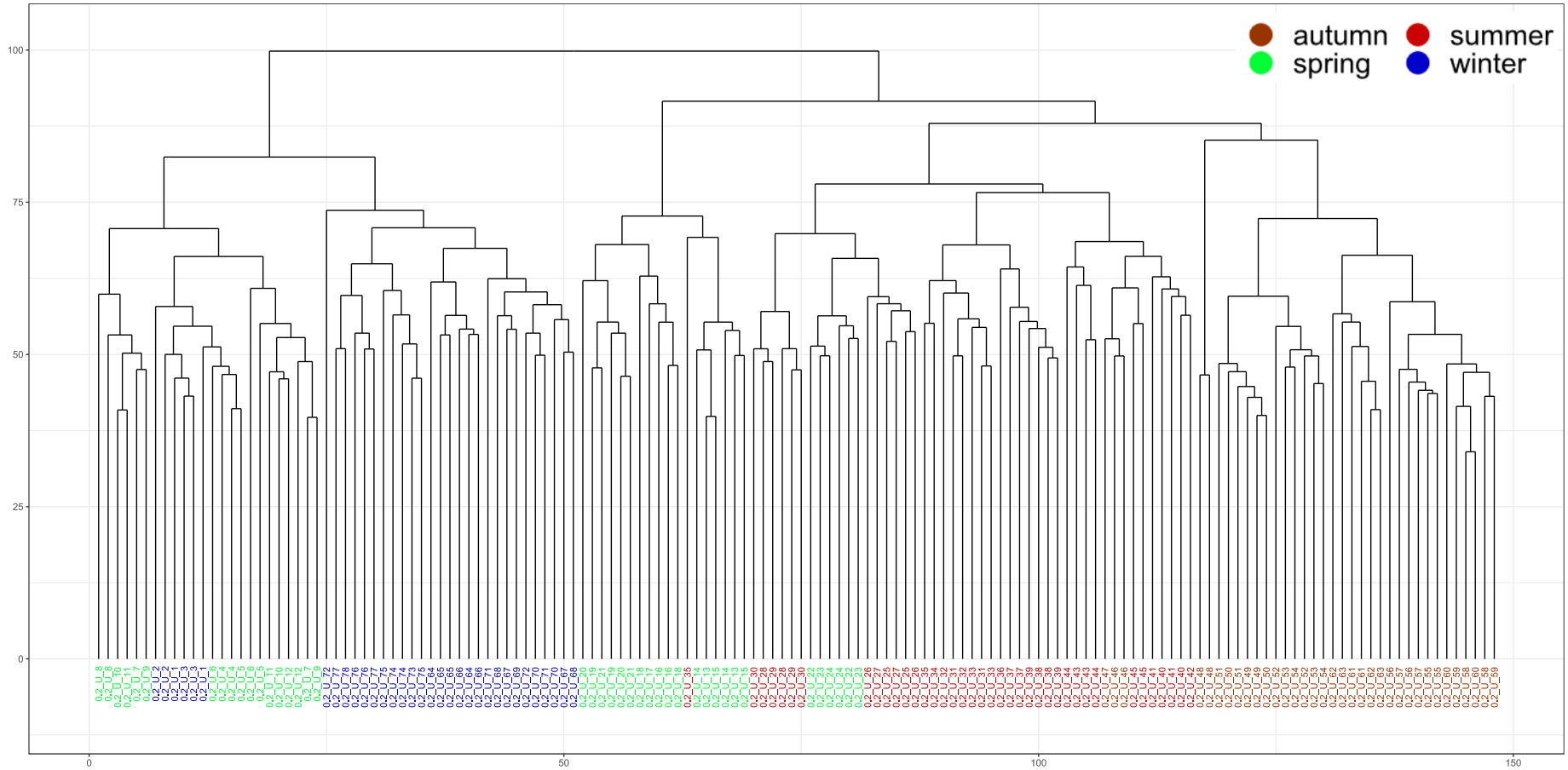

Figure SIV.2: Dendrogram presenting the similarity between 16S rDNA picoplankton samples sequenced by two independent sequencing runs. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

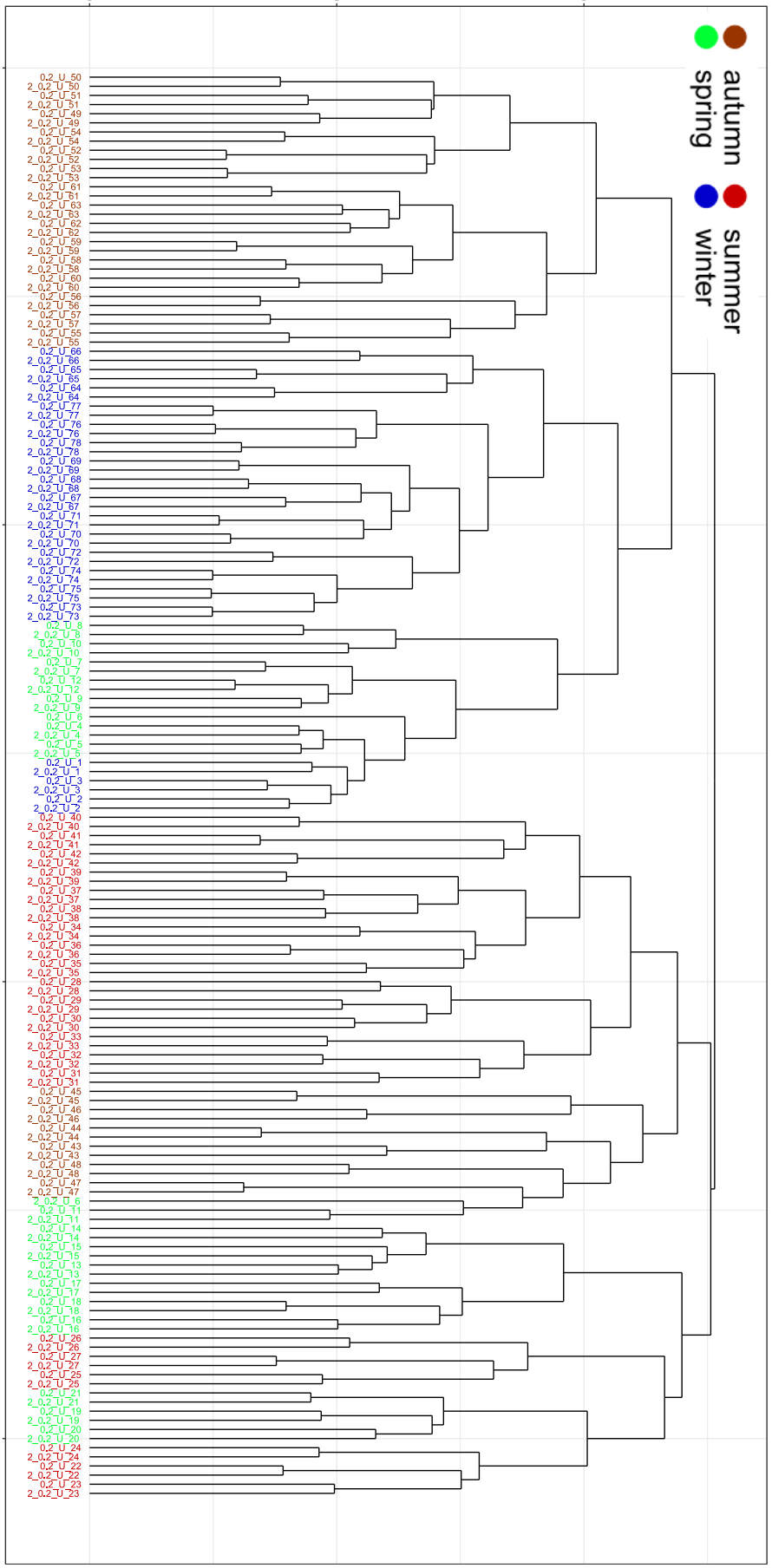

Figure SIV.3: Dendrogram presenting the similarity between 18S rDNA picoplankton samples sequenced by two independent sequencing runs. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

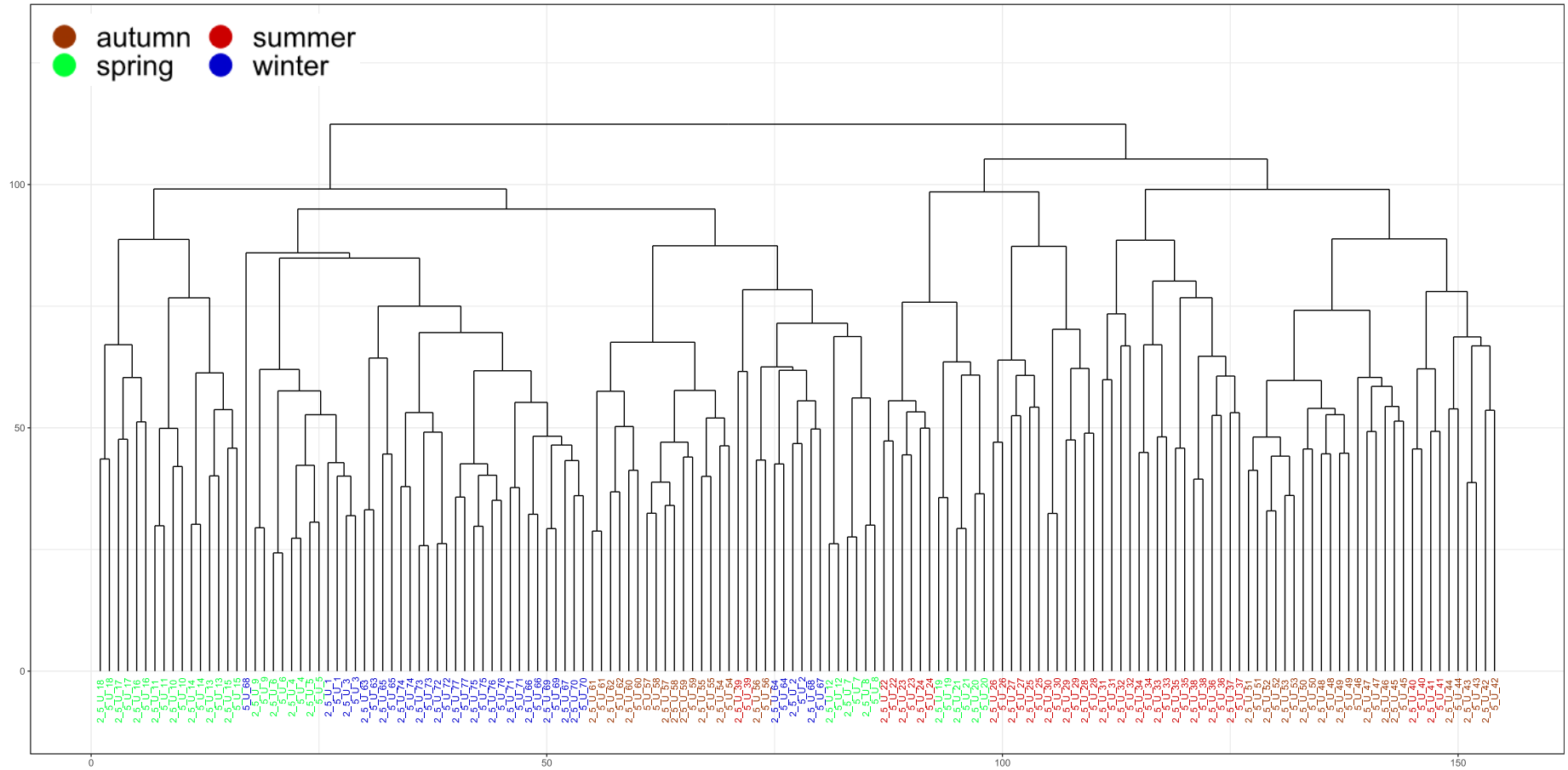

Figure SIV.4: Dendrogram presenting the similarity between 18S rDNA nanoplankton samples sequenced by two independent sequencing runs. Most of the samples are clustering closely to their equivalent from the other sequencing run and almost no single samples can be observed, indicating homogeneity of sequencing between the two runs. Each sample is named twice in the dendrogram, one for each sequencing run. The colors correspond to the seasons, representing the seasonality and visualizing if samples are not homogenous.

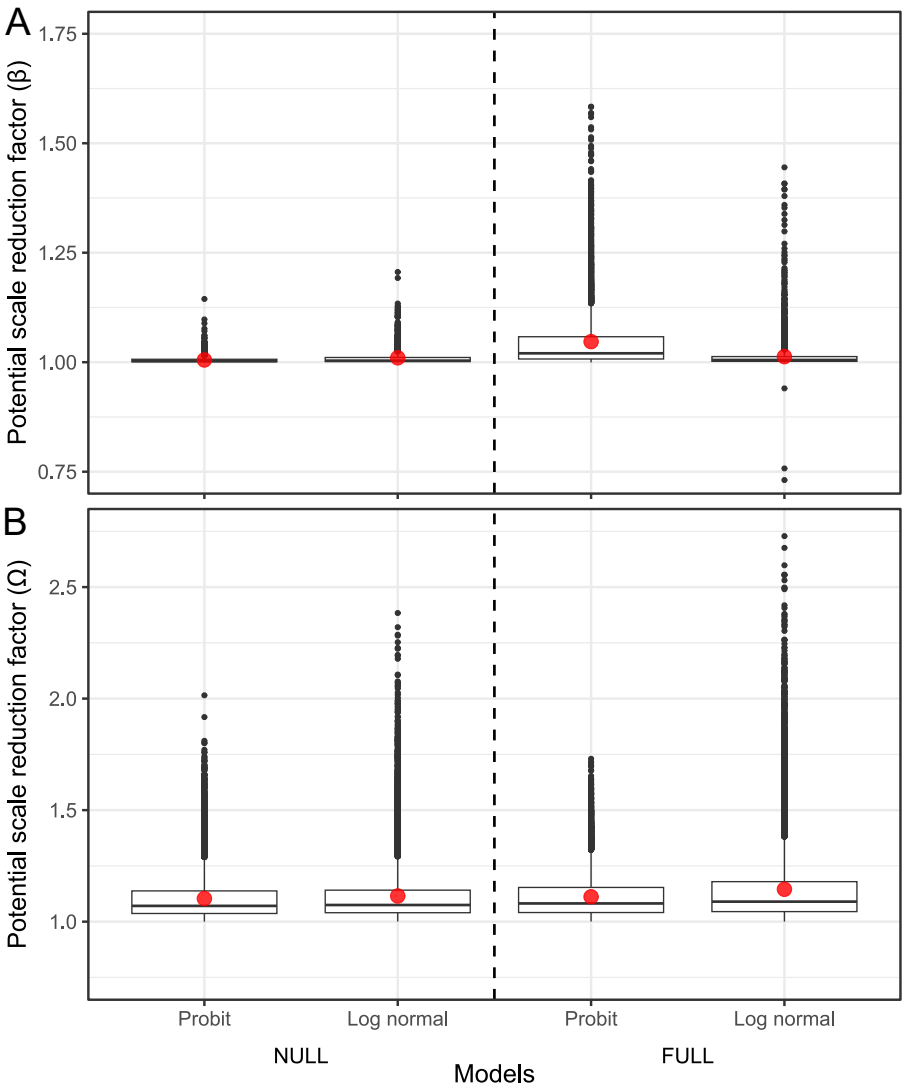

Figure SIV.5: Boxplot of the potential scale reduction factor of the Beta (A) and Omega (B) parameters. Red dot indicates the average

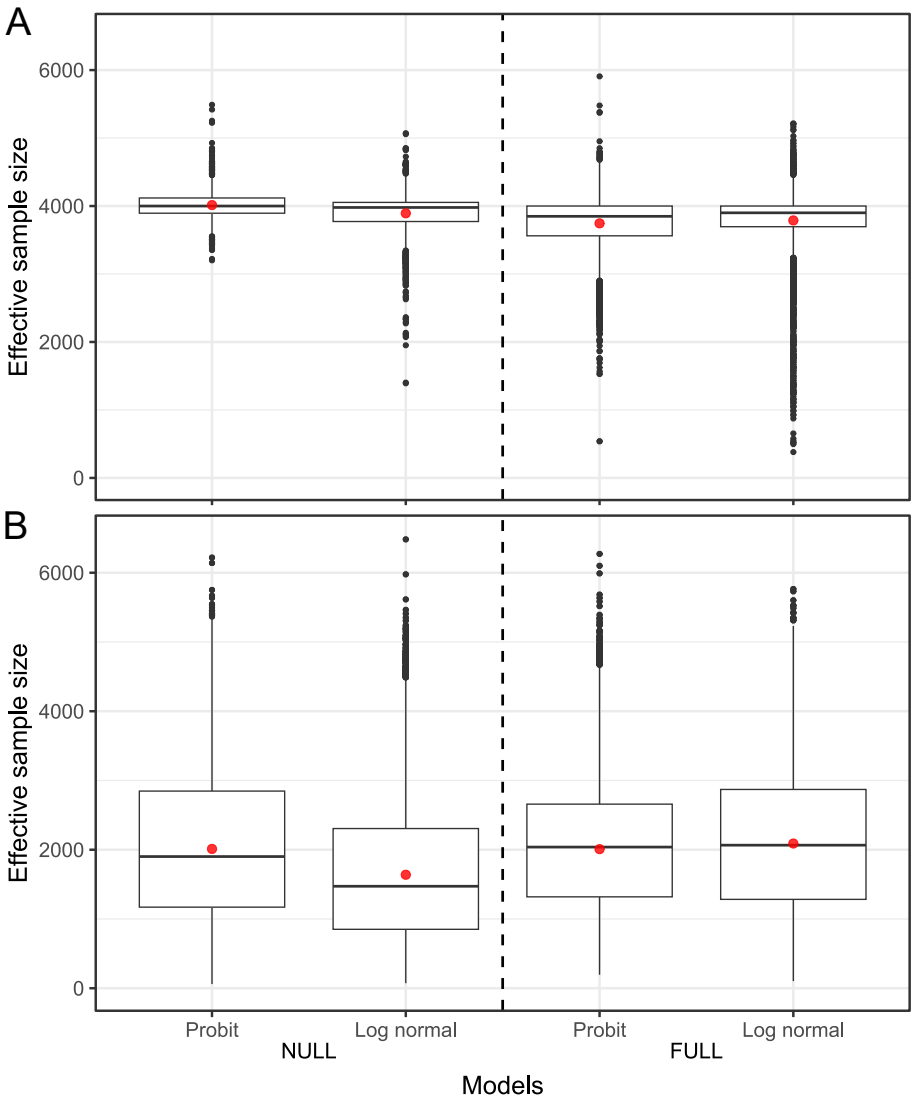

Figure SIV.6: Boxplot of the effective sample size of the Beta (A) and Omega (B) parameters. Red dot indicates the average

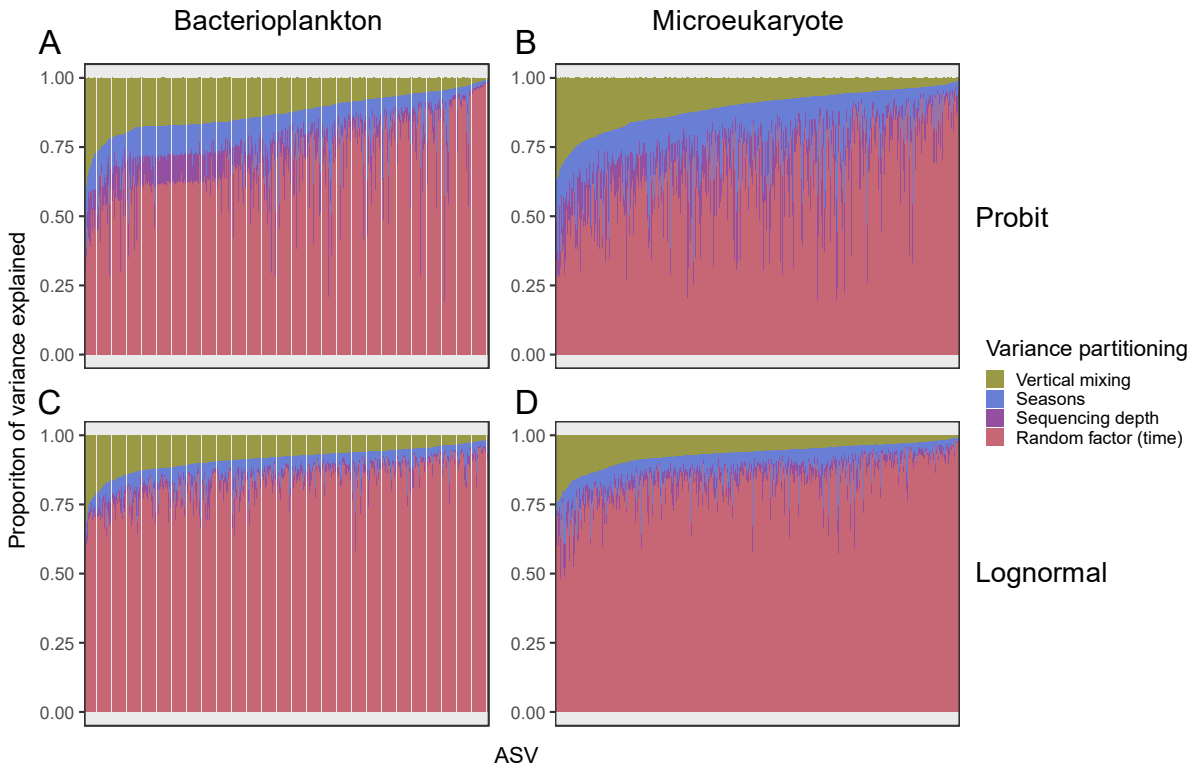

Figure SIV.7: variance partitioning of bacterioplankton (A-C) and microeukaryote (B-D) community composition of the probit (A-B) and the Lognormal models (C-D) ordered by descending seasonality proportion. The barplot represent the proportion of variance explained by each variable (y axis) for each ASVs (x axis). The fixed factors comprise the sequencing depth and the seasonality (Seasons and cold-water mixing) and the species-to-species association is represented by the random. ASVs order is not the same between the probit and lognormal model for the bacterioplankton and microeukaryote. ASVs are ordered based on the proportion of variance explained by the time random factor.

id	replicat	time	date
U_1	1	1	13.03.2018
U_3	1	1	13.03.2019
U_4	2	15	27.03.2018
U_5	2	15	27.03.2019
U_6	2	15	27.03.2020
U_7	3	29	10.04.2018
U_8	3	29	10.04.2019
U_9	3	29	10.04.2020
U_10	4	46	27.04.2018
U_13	5	57	08.05.2018
U_14	5	57	08.05.2019
U_15	5	57	08.05.2020
U_16	6	71	22.05.2018
U_17	6	71	22.05.2019
U_18	6	71	22.05.2020
U_19	7	85	05.06.2018
U_20	7	85	05.06.2019
U_21	7	85	05.06.2020
U_22	8	99	19.06.2018
U_23	8	99	19.06.2019
U_24	8	99	19.06.2020
U_25	9	113	03.07.2018
U_26	9	113	03.07.2019
U_27	9	113	03.07.2020
U_28	10	127	17.07.2018
U_29	10	127	17.07.2019
U_30	10	127	17.07.2020
U_31	11	141	31.07.2018
U_32	11	141	31.07.2019
U_34	12	155	14.08.2018
U_35	12	155	14.08.2019
U_36	12	155	14.08.2020
U_37	13	169	28.08.2018
U_38	13	169	28.08.2019
U_40	14	183	11.09.2018
U_41	14	183	11.09.2019
U_42	14	183	11.09.2020
U_43	15	193	21.09.2018
U_44	15	193	21.09.2019
U_45	15	193	21.09.2020
U_46	16	211	09.10.2018
U_48	16	211	09.10.2018
U_49	17	225	23.10.2018
U_50	17	225	23.10.2018
U_51	17	225	23.10.2018
U_52	18	239	06.11.2018
U_53	18	239	06.11.2018
U_54	18	239	06.11.2018
U_55	19	255	22.11.2018
U_57	19	255	22.11.2018
U_58	20	267	04.12.2018
U_59	20	267	04.12.2018
U_60	20	267	04.12.2018
U_61	21	281	18.12.2018
U_62	21	281	18.12.2018
U_63	21	281	18.12.2018
U_65	22	305	11.01.2019
U_66	22	305	11.01.2019
U_67	23	323	29.01.2019
U_69	23	323	29.01.2019
U_70	24	337	12.02.2019
U_71	24	337	12.02.2019
U_72	24	337	12.02.2019
U_73	25	351	26.02.2019
U_74	25	351	26.02.2019
U_75	25	351	26.02.2019
U_76	26	365	12.03.2019
U_77	26	365	12.03.2019
U_78	26	365	12.03.2019

Table SIV.1: Study design oof the HMSC model pipeline with the sample id (id), replicates per sample date (replicat) time used as a latent variable for the ASVs associations (time) and sampling date (date). The time variable used as latent variable for the ASVs associations represent the day of sampling with the first sampling date as the original one.

Table SIV.2: total number of reads after the major bioinformatic and biostatistics steps for the two independent sequencing run in the three first steps and then merged together. Raw paired end reads correspond to the number of reads after merging of the forward and reverse sequences. Denoised is the number of reads after filtration and denoising performed by dada2. Non chimeric is the final number of reads after the chimera detection performed by dada2. Filtration of the low abundant ASVs indicate the number of reads left after keeping only ASVs with a relative abundance (in percentage) above 0.5% in at least 20% of the samples and correspond to the total number of reads that was used to perform the analysis.

	Raw paired end merged	Denoised	Non chimeric	Sequencing run merge	Filtration low abundant ASVs
Bacterioplankton	2673067	1410919	1398600	4487577	3581517
	4828422	3218032	3065075		
Microeukaryote	4588482	4490502	4317695	8055908	4865859
	3946453	3870588	3738213		

Table SIV.3: Percentage of associations associated to each phylum by models with their total relative abundance and richness (number of ASVs associated to the phylum)

Phylum	NULL		FULL		relative abundance	richness
	Probit	Lognormal	Probit	Lognormal		
Acidobacteriota	1.79	1.25	2.31	1.58	0.99	14
Actinobacteriota	2.65	11.56	2.20	11.99	25.97	62
Armatimonadota	0.13	**N/A**	0.07	**N/A**	0.01	1
Bacteroidota	11.48	10.29	11.74	9.79	12.56	112
Bdellovibrionota	0.21	0.10	0.25	0.10	0.04	1
Cercozoa	4.01	4.31	4.14	4.54	3.46	48
Chloroflexi	0.69	1.32	0.83	1.59	0.96	7
Chlorophyta	3.61	3.36	3.49	3.77	1.35	34
Choanoflagellida	1.38	1.49	1.07	0.82	0.68	17
Chrompodellids	0.46	0.33	0.74	0.20	0.17	6
Ciliophora	16.48	12.53	16.57	12.61	8.13	156
Cryptophyta	3.35	2.95	3.19	2.43	5.72	29
Cyanobacteria	0.53	1.80	0.28	1.48	6.51	13
Dependentiae	**N/A**	0.08	**N/A**	0.09	0.03	1
Desulfobacterota	0.11	0.00	0.14	**N/A**	0.02	1
Dinoflagellata	10.02	7.78	9.77	7.88	4.26	98
Fungi	3.00	2.44	3.09	2.44	0.75	32
Gemmatimonadota	0.33	0.32	0.41	0.52	0.13	2
Haptophyta	1.82	1.44	1.70	1.62	3.77	19
Katablepharidophyta	1.74	1.59	1.94	1.68	2.37	19
Margulisbacteria	0.16	0.01	0.06	**N/A**	0.02	1
Mesomycetozoa	0.40	0.57	0.53	0.29	0.14	4
Nitrospirota	0.42	0.55	0.54	0.62	0.24	2
Ochrophyta	14.16	11.75	13.46	12.04	7.17	146
Opalozoa	1.39	1.49	1.73	1.67	0.46	16
Perkinsea	2.65	2.83	2.81	2.76	2.55	28
Planctomycetota	1.13	2.12	1.15	2.48	1.42	14
Proteobacteria	11.99	12.69	11.61	11.95	7.67	128
Pseudofungi	0.60	0.65	0.62	1.15	0.47	6
SAR324_clade	0.40	0.18	0.21	0.17	0.08	3
Telonemia	1.39	0.88	1.75	0.86	1.43	12
Verrucomicrobiota	1.52	1.35	1.59	0.89	0.49	15

Table SIV.4: proportion of positive and negative associations in the FULL models

	FULL models			
	Probit		Lognormal	
	Negative	Positive	Negative	Positive
Actinobacteriota	0.44	0.56	0.37	0.63
Bacteroidota	0.41	0.59	0.43	0.57
Cercozoa	0.47	0.53	0.46	0.54
Chlorophyta	0.34	0.66	0.52	0.48
Choanoflagellida	0.39	0.61	0.58	0.42
Ciliophora	0.40	0.60	0.44	0.56
Cryptophyta	0.37	0.63	0.38	0.62
Cyanobacteria	0.39	0.61	0.43	0.57
Dinoflagellata	0.39	0.61	0.43	0.57
Fungi	0.41	0.59	0.39	0.61
Haptophyta	0.42	0.58	0.42	0.58
Katablepharidophyta	0.44	0.56	0.38	0.62
Ochrophyta	0.37	0.63	0.43	0.57
perkinsea	0.43	0.57	0.42	0.58
Proteobacteria	0.42	0.58	0.44	0.56
Verrucomicrobiota	0.45	0.55	0.45	0.55

Table SIV.5: number of associations between phytoplankton phyla and bacterioplankton taxa

	Ochrophyta	Dinoflagellata	Chlorophyta	Haptophyta	Cryptophyta	Cyanobacteria
bacII	16	16	15	9	16	5
Flavobacteriaceae	9	9	9	9	9	4
Rhodobacterales	4	4	4	3	4	1
Cytophagales	4	4	3	3	3	2

Phyla	Actinobacteriota		Bacteroidota		Cercozoa		Chlorophyta		Choanoflagellida	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	32	20	151	160	54	52	51	50	14	13
Bacteroidota	151	160	974	554	301	245	352	136	94	55
Cercozoa	54	52	301	245	78	80	77	67	28	28
Chlorophyta	51	50	352	136	77	67	104	36	39	14
Choanoflagellida	14	13	94	55	28	28	39	14	8	2
Ciliophora	190	207	1375	723	420	360	478	192	129	68
Cryptophyta	39	41	263	161	86	64	86	40	25	13
Cyanobacteria	3	2	28	13	5	8	6	1	2	2
Dinoflagellata	122	101	763	477	241	203	256	126	73	44
Fungi	36	43	252	149	87	64	94	33	17	14
Haptophyta	25	19	126	85	42	43	40	21	15	5
Katablepharidophyta	28	17	133	106	51	45	32	26	11	9
Ochrophyta	175	136	1134	635	323	287	391	150	106	52
Perkinsea	23	37	235	141	65	59	75	40	23	11
Proteobacteria	207	79	743	755	260	256	244	193	71	64
Verrucomicrobiota	32	12	97	122	34	48	23	27	11	8

Table SIV.6: total number of positive and negative associations between the major phyla of the FULL probit occurrence model

Phyla	Ciliophora		Cryptophyta		Cyanobacteria		Dinoflagellata		Fungi		Haptophyta	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	190	207	39	41	3	2	122	101	36	43	25	19
Bacteroidota	1375	723	263	161	28	13	763	477	252	149	126	85
Cercozoa	420	360	86	64	5	8	241	203	87	64	42	43
Chlorophyta	478	192	86	40	6	1	256	126	94	33	40	21
Choanoflagellida	129	68	25	13	2	2	73	44	17	14	15	5
Ciliophora	1970	1050	373	211	43	15	1152	670	367	197	172	129
Cryptophyta	373	211	60	36	7	3	235	125	77	46	34	22
Cyanobacteria	43	15	7	3	NA	NA	25	8	3	2	3	2
Dinoflagellata	1152	670	235	125	25	8	628	418	197	124	111	78
Fungi	367	197	77	46	3	2	197	124	54	30	25	25
Haptophyta	172	129	34	22	3	2	111	78	25	25	12	18
Katablepharidophyta	206	159	39	28	3	4	119	98	40	33	23	19
Ochrophyta	1655	856	319	155	28	12	973	494	278	170	168	94
Perkinsea	337	185	66	39	8	3	196	119	57	37	32	24
Proteobacteria	1026	1050	249	157	13	23	670	543	195	207	132	84
Verrucomicrobiota	129	173	26	20	2	3	92	84	22	34	18	13

Table SIV.6: total number of positive and negative associations between the major phyla of the FULL probit occurrence model

Phyla	Haptophyta		Katablepharidophyta		Ochrophyta		Perkinsea		Proteobacteria		Verrucomicrobiota	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	25	19	28	17	175	136	23	37	207	79	32	12
Bacteroidota	126	85	133	106	1134	635	235	141	743	755	97	122
Cercozoa	42	43	51	45	323	287	65	59	260	256	34	48
Chlorophyta	40	21	32	26	391	150	75	40	244	193	23	27
Choanoflagellida	15	5	11	9	106	52	23	11	71	64	11	8
Ciliophora	172	129	206	159	1655	856	337	185	1026	1050	129	173
Cryptophyta	34	22	39	28	319	155	66	39	249	157	26	20
Cyanobacteria	3	2	3	4	28	12	8	3	13	23	2	3
Dinoflagellata	111	78	119	98	973	494	196	119	670	543	92	84
Fungi	25	25	40	33	278	170	57	37	195	207	22	34
Haptophyta	12	18	23	19	168	94	32	24	132	84	18	13
Katablepharidophyta	23	19	10	22	172	118	30	27	138	103	19	17
Ochrophyta	168	94	172	118	1348	678	266	175	967	712	134	99
Perkinsea	32	24	30	27	266	175	48	28	148	195	17	26
Proteobacteria	132	84	138	103	967	712	148	195	1148	450	149	50
Verrucomicrobiota	18	13	19	17	134	99	17	26	149	50	14	4

Table SIV.6: total number of positive and negative associations between the major phyla of the FULL probit occurrence model

Phyla	Actinobacteriota		Bacteroidota		Cercozoa		Chlorophyta		Choanoflagellida	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	658	70	286	107	69	92	28	97	4	15
Bacteroidota	286	107	260	98	65	90	39	82	5	20
Cercozoa	69	92	65	90	44	40	45	31	12	4
Chlorophyta	28	97	39	82	45	31	24	12	7	7
Choanoflagellida	4	15	5	20	12	4	7	7	NA	NA
Ciliophora	234	241	212	221	118	82	105	78	19	20
Cryptophyta	30	26	43	48	26	12	26	12	6	5
Cyanobacteria	35	11	20	23	9	7	9	14	2	2
Dinoflagellata	141	165	128	120	94	50	65	38	12	12
Fungi	29	34	30	47	39	10	30	10	4	4
Haptophyta	41	25	23	31	16	9	14	11	3	1
Katablepharidophyt	33	36	27	30	23	10	17	6	3	2
Ochrophyta	207	267	168	188	130	73	94	73	18	13
Perkinsea	51	45	43	50	28	22	24	13	4	4
Proteobacteria	358	137	291	121	58	100	42	99	12	26
Verrucomicrobiota	41	6	16	13	3	6	2	9	NA	2

Table SIV.7: total number of positive and negative associations between the major phyla of the FULL lognormal abundance model

Phyla	Ciliophora		Cryptophyta		Cyanobacteria		Dinoflagellata		Fungi	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	234	241	30	26	35	11	141	165	29	34
Bacteroidota	212	221	43	48	20	23	128	120	30	47
Cercozoa	118	82	26	12	9	7	94	50	39	10
Chlorophyta	105	78	26	12	9	14	65	38	30	10
Choanoflagellida	19	20	6	5	2	2	12	12	4	4
Ciliophora	382	180	85	36	33	49	219	136	75	26
Cryptophyta	85	36	14	10	11	5	46	26	16	5
Cyanobacteria	33	49	11	5	18	4	17	20	5	5
Dinoflagellata	219	136	46	26	17	20	148	74	53	23
Fungi	75	26	16	5	5	5	53	23	12	6
Haptophyta	47	27	11	8	4	3	16	20	9	8
Katablepharidophyt	59	15	13	3	3	5	26	14	12	3
Ochrophyta	349	193	68	32	35	35	234	116	101	35
Perkinsea	79	38	20	2	7	9	46	29	22	7
Proteobacteria	172	305	55	45	33	11	130	171	28	55
Verrucomicrobiota	22	16	1	8	1	5	11	10	1	3

Table SIV.7: total number of positive and negative associations between the major phyla of the FULL lognormal abundance model

Phyla	Haptophyta		Katablepharidophyta		Ochrophyta		Perkinsea		Proteobacteria		Verrucomicrobiota	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actinobacteriota	41	25	33	36	207	267	51	45	358	137	41	6
Bacteroidota	23	31	27	30	168	188	43	50	291	121	16	13
Cercozoa	16	9	23	10	130	73	28	22	58	100	3	6
Chlorophyta	14	11	17	6	94	73	24	13	42	99	2	9
Choanoflagellida	3	1	3	2	18	13	4	4	12	26	NA	2
Ciliophora	47	27	59	15	349	193	79	38	172	305	22	16
Cryptophyta	11	8	13	3	68	32	20	2	55	45	1	8
Cyanobacteria	4	3	3	5	35	35	7	9	33	11	1	5
Dinoflagellata	16	20	26	14	234	116	46	29	130	171	11	10
Fungi	9	8	12	3	101	35	22	7	28	55	1	3
Haptophyta	2	2	8	2	46	21	14	5	30	32	2	4
Katablepharidophyt	8	2	4	2	55	13	20	7	29	37	NA	3
Ochrophyta	46	21	55	13	326	194	86	40	182	266	11	12
Perkinsea	14	5	20	7	86	40	12	6	40	66	3	6
Proteobacteria	30	32	29	37	182	266	40	66	488	148	18	14
Verrucomicrobiota	2	4	NA	3	11	12	3	6	18	14	6	NA

Table SIV.7: total number of positive and negative associations between the major phyla of the FULL lognormal abundance model

References

Abrego N, Crosier B, Somervuo P, Ivanova N, Abrahamyan A, Abdi A, Hämäläinen K, Junninen K, Maunula M, Purhonen J, Ovaskainen O (2020) Fungal communities decline with urbanization—more in air than in soil. ISME J 14:2806–2815. https://doi.org/10.1038/s41396-020-0732-1

Achenbach LA, Michaelidou U, Bruce RA, Fryman J, Coates JD 2001 Dechloromonas agitata gen. nov., sp. nov. and Dechlorosoma suillum gen. nov., sp. nov., two novel environmentally dominant (per)chlorate-reducing bacteria and their phylogenetic position. International Journal of Systematic and Evolutionary Microbiology 51:527–533. https://doi.org/10.1099/00207713-51-2-527

Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF (2004) Divergence and Redundancy of 16S rRNA Sequences in Genomes with Multiple rrn Operons. Journal of Bacteriology 186:2629–2635. https://doi.org/10.1128/JB.186.9.2629-2635.2004

Adrian R, O'Reilly CM, Zagarese H, Baines SB, Hessen DO, Keller W, Livingstone DM, Sommaruga R, Straile D, Van Donk E, Weyhenmeyer GA, Winder M (2009) Lakes as sentinels of climate change. Limnology and Oceanography 54:2283–2297. https://doi.org/10.4319/lo.2009.54.6_part_2.2283

Agha R, Saebelfeld M, Manthey C, Rohrlack T, Wolinska J (2016) Chytrid parasitism facilitates trophic transfer between bloom-forming cyanobacteria and zooplankton (Daphnia). Sci Rep 6:35039. https://doi.org/10.1038/srep35039

Aitchison J (2003) The Statistical Analysis of Compositional Data. Blackburn Press

Alavi MR (2004) Predator/Prey Interaction between Pfiesteria piscicida and Rhodomonas Mediated by a Marine Alpha Proteobacterium. Microb Ecol 47:48–58. https://doi.org/10.1007/s00248-003-1018-7

Allen LZ, Allen EE, Badger JH, McCrow JP, Paulsen IT, Elbourne LD, Thiagarajan M, Rusch DB, Nealson KH, Williamson SJ, Venter JC, Allen AE (2012) Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. The ISME Journal 6:1403–1414. https://doi.org/10.1038/ismej.2011.201

Allgaier M, Grossart H-P (2006) Seasonal dynamics and phylogenetic diversity of free-living and particle-associated bacterial communities in four lakes in northeastern Germany. Aquatic Microbial Ecology 45:115–128. https://doi.org/10.3354/ame045115

Altaner S, Puddick J, Fessard V, Feurstein D, Zemskov I, Wittmann V, Dietrich DR (2019) Simultaneous Detection of 14 Microcystin Congeners from Tissue Samples Using UPLC- ESI-MS/MS and Two Different Deuterated Synthetic Microcystins as Internal Standards. Toxins (Basel) 11:388. https://doi.org/10.3390/toxins11070388

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biology 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106

References

Anderson MJ (2017) Permutational Multivariate Analysis of Variance (PERMANOVA). In: Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd, pp 1–15

Anderson MJ, Walsh DCI (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs 83:557–574. https://doi.org/10.1890/12-2010.1

Anderson MJ, Walsh DCI, Robert Clarke K, Gorley RN, Guerra-Castro E (2017) Some solutions to the multivariate Behrens–Fisher problem for dissimilarity-based analyses. Australian & New Zealand Journal of Statistics 59:57–79. https://doi.org/10.1111/anzs.12176

Anderson OR (1988) Ciliates (Phylum: Ciliophora). In: Anderson OR (ed) Comparative Protozoology: Ecology, Physiology, Life History. Springer, Berlin, Heidelberg, pp 94–107

Andrén O, Bengtsson J, Clarholm M (1995) Biodiversity and species redundancy among litter decomposers. In: Collins HP, Robertson GP, Klug MJ (eds) The Significance and Regulation of Soil Biodiversity: Proceedings of the International Symposium on Soil Biodiversity, held at Michigan State University, East Lansing, May 3–6, 1993. Springer Netherlands, Dordrecht, pp 141–151

Annenkova NV, Lavrov DV, Belikov SI (2011) Dinoflagellates Associated with Freshwater Sponges from the Ancient Lake Baikal. Protist 162:222–236. https://doi.org/10.1016/j.protis.2010.07.002

Aoki I (2003) Diversity–productivity–stability relationship in freshwater ecosystems: Whole-systemic view of all trophic levels. Ecol Res 18:397–404. https://doi.org/10.1046/j.1440-1703.2003.00564.x

Arandia-Gorostidi N, Krabberød AK, Logares R, Deutschmann IM, Scharek R, Morán XAG, González F, Alonso-Sáez L (2022) Novel Interactions Between Phytoplankton and Bacteria Shape Microbial Seasonal Dynamics in Coastal Ocean Waters. Frontiers in Marine Science 9

Aránguiz-Acuña A, Luque JA, Pizarro H, Cerda M, Heine-Fuster I, Valdés J, Fernández-Galego E, Wennrich V (2020) Aquatic community structure as sentinel of recent environmental changes unraveled from lake sedimentary records from the Atacama Desert, Chile. PLOS ONE 15:e0229453. https://doi.org/10.1371/journal.pone.0229453

Ávila MP, Staehr PA, Barbosa FAR, Chartone-Souza E, Nascimento AMA (2017) Seasonality of freshwater bacterioplankton diversity in two tropical shallow lakes from the Brazilian Atlantic Forest. FEMS Microbiol Ecol 93. https://doi.org/10.1093/femsec/fiw218

Avila-Jimenez M-L, Burns G, He Z, Zhou J, Hodson A, Avila-Jimenez J-L, Pearce D (2020) Functional Associations and Resilience in Microbial Communities. Microorganisms 8:951. https://doi.org/10.3390/microorganisms8060951

Azam F, Fenchel T, Field J, Gray JS, Meyer L, Thingstad TF (1983) The Ecological Role of Water-Column Microbes in the Sea. Marine Ecology Progress Series 10:257–263. https://doi.org/10.3354/meps010257

Baer J, Spiessl C, Auerswald K, Geist J, Brinker A (2022) Signs of the times: Isotopic signature changes in several fish species following invasion of Lake Constance by quagga mussels. Journal of Great Lakes Research 48:746–755. https://doi.org/10.1016/j.jglr.2022.03.010

References

Baltar F, Palovaara J, Unrein F, Catala P, Horňák K, Šimek K, Vaqué D, Massana R, Gasol JM, Pinhassi J (2016) Marine bacterial community structure resilience to changes in protist predation under phytoplankton bloom conditions. ISME J 10:568–581. https://doi.org/10.1038/ismej.2015.135

Balvočiūtė M, Huson DH (2017) SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? BMC Genomics 18:114. https://doi.org/10.1186/s12864-017-3501-4

Banerji A, Bagley M, Elk M, Pilgrim E, Marinson J, Santo Domingo J (2018) Spatial and temporal dynamics of a freshwater eukaryotic plankton community revealed via 18S rRNA gene metabarcoding. Hydrobiologia 818:71–86. https://doi.org/10.1007/s10750-018-3593-0

Barton AD, Ward BA, Williams RG, Follows MJ (2014) The impact of fine-scale turbulence on phytoplankton community structure. Limnology and Oceanography: Fluids and Environments 4:34–49. https://doi.org/10.1215/21573689-2651533

Bax N, Williamson A, Aguero M, Gonzalez E, Geeves W (2003) Marine invasive alien species: a threat to global biodiversity. Marine Policy 27:313–323. https://doi.org/10.1016/S0308-597X(03)00041-1

Bec A, Martin-Creuzburg D, Elert EV (2010) Fatty acid composition of the heterotrophic nanoflagellate Paraphysomonas sp.: influence of diet and de novo biosynthesis. Aquatic Biology 9:107–112. https://doi.org/10.3354/ab00244

Beck DAC, Kalyuzhnaya MG, Malfatti S, Tringe SG, Rio TG del, Ivanova N, Lidstrom ME, Chistoserdova L (2013) A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the Methylococcaceae and the Methylophilaceae. PeerJ 1:e23. https://doi.org/10.7717/peerj.23

Behnke A, Bunge J, Barger K, Breiner H-W, Alla V, Stoeck T (2006) Microeukaryote community patterns along an O2/H2S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). Appl Environ Microbiol 72:3626–3636. https://doi.org/10.1128/AEM.72.5.3626-3636.2006

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernát G, Boross N, Somogyi B, Vörös L, G.-Tóth L, Boros G (2020) Oligotrophication of Lake Balaton over a 20-year period and its implications for the relationship between phytoplankton and zooplankton biomass. Hydrobiologia 847:3999–4013. https://doi.org/10.1007/s10750-020-04384-x

Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification. Appl Environ Microbiol 77:7846–7849. https://doi.org/10.1128/AEM.05220-11

Berry MA, Davis TW, Cory RM, Duhaime MB, Johengen TH, Kling GW, Marino JA, Den Uyl PA, Gossiaux D, Dick GJ, Denef VJ (2017a) Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. Environ Microbiol 19:1149–1162. https://doi.org/10.1111/1462-2920.13640

Berry MA, White JD, Davis TW, Jain S, Johengen TH, Dick GJ, Sarnelle O, Denef VJ (2017b) Are Oligotypes Meaningful Ecological and Phylogenetic Units? A Case Study of Microcystis in Freshwater Lakes. Frontiers in Microbiology 8

References

Beversdorf LJ, Chaston SD, Miller TR, McMahon KD (2015) Microcystin mcyA and mcyE Gene Abundances Are Not Appropriate Indicators of Microcystin Concentrations in Lakes. PLOS ONE 10:e0125353. https://doi.org/10.1371/journal.pone.0125353

Biau DJ, Jolles BM, Porcher R (2010) P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers. Clin Orthop Relat Res 468:885–892. https://doi.org/10.1007/s11999-009-1164-4

Biderre-Petit C, Jézéquel D, Dugat-Bony E, Lopes F, Kuever J, Borrel G, Viollier E, Fonty G, Peyret P (2011) Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake. FEMS microbiology ecology 77:533–45. https://doi.org/10.1111/j.1574-6941.2011.01134.x

Bielewicz S, Bell E, Kong W, Friedberg I, Priscu JC, Morgan-Kiss RM (2011) Protist diversity in a permanently ice-covered Antarctic Lake during the polar night transition. ISME J 5:1559–1564. https://doi.org/10.1038/ismej.2011.23

Blottière L, Jaffar-Bandjee M, Jacquet S, Millot A, Hulot FD (2017) Effects of mixing on the pelagic food web in shallow lakes. Freshwater Biology 62:161–177. https://doi.org/10.1111/fwb.12859

Boers SA, Hays JP, Jansen R (2015) Micelle PCR reduces chimera formation in 16S rRNA profiling of complex microbial DNA mixtures. Sci Rep 5:14181. https://doi.org/10.1038/srep14181

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Silva RD, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, Hooft JJJ van der, Vargas F, Vázquez-Baeza Y, Vogtmann E, Hippel M von, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9

Botton S, van Heusden M, Parsons JR, Smidt H, van Straalen N (2006) Resilience of Microbial Systems Towards Disturbances. Critical Reviews in Microbiology 32:101–112. https://doi.org/10.1080/10408410600709933

Bouaïcha N, Miles CO, Beach DG, Labidi Z, Djabri A, Benayache NY, Nguyen-Quang T (2019) Structural Diversity, Characterization and Toxicology of Microcystins. Toxins (Basel) 11:714. https://doi.org/10.3390/toxins11120714

References

Bradford TM, Morgan MJ, Lorenz Z, Hartley DM, Hardy CM, Oliver RL (2013) Microeukaryote community composition assessed by pyrosequencing is associated with light availability and phytoplankton primary production along a lowland river. Freshwater Biology 58:2401–2413. https://doi.org/10.1111/fwb.12219

Bradley PH, Pollard KS (2017) Proteobacteria explain significant functional variability in the human gut microbiome. Microbiome 5:36. https://doi.org/10.1186/s40168-017-0244-z

Bray JR, Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs 27:325–349. https://doi.org/10.2307/1942268

Brown CT (2014) A History of Bioinformatics (in the Year 2039). In: Git Piper. https://gitpiper.com/. Accessed 20 Apr 2023

Brown MB (1975) 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. Biometrics 31:987–992. https://doi.org/10.2307/2529826

Brümmer F, Müller A (2015) Süßwasserschwämme (Porifera: Spongillidae) in Baden-Württemberg. Jahreshefte der Gesellschaft für Naturkunde in Württemberg 171:259–282. https://doi.org/10.26251/jhgfn.171.2015.259-282

Bubak I, Śliwińska-Wilczewska S, Głowacka P, Szczerba A, Możdżeń K (2020) The Importance of Allelopathic Picocyanobacterium Synechococcus sp. on the Abundance, Biomass Formation, and Structure of Phytoplankton Assemblages in Three Freshwater Lakes. Toxins (Basel) 12:259. https://doi.org/10.3390/toxins12040259

Buchan A, LeCleir GR, Gulvik CA, González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol 12:686–698. https://doi.org/10.1038/nrmicro3326

Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI (2019) The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data 6:190007. https://doi.org/10.1038/sdata.2019.7

Bunse C, Bertos-Fortis M, Sassenhagen I, Sildever S, Sjöqvist C, Godhe A, Gross S, Kremp A, Lips I, Lundholm N, Rengefors K, Sefbom J, Pinhassi J, Legrand C (2016) Spatio-Temporal Interdependence of Bacteria and Phytoplankton during a Baltic Sea Spring Bloom. Frontiers in Microbiology 7

Buratti FM, Manganelli M, Vichi S, Stefanelli M, Scardala S, Testai E, Funari E (2017) Cyanotoxins: producing organisms, occurrence, toxicity, mechanism of action and human health toxicological risk evaluation. Arch Toxicol 91:1049–1130. https://doi.org/10.1007/s00204-016-1913-6

Buskey EJ, Stoecker DK (1989) Behavioral responses of the marine tintinnid Favella sp. to phytoplankton: influence of chemical, mechanical and photic stimuli. Journal of Experimental Marine Biology and Ecology 132:1–16. https://doi.org/10.1016/0022-0981(89)90173-1

Cabello-Yeves PJ, Zemskaya TI, Zakharenko AS, Sakirko MV, Ivanov VG, Ghai R, Rodriguez-Valera F (2020) Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. Limnology and Oceanography 65:1471–1488. https://doi.org/10.1002/lno.11401

References

Cabrerizo MJ, Medina-Sánchez JM, González-Olalla JM, Sánchez-Gómez D, Carrillo P (2022) Microbial plankton responses to multiple environmental drivers in marine ecosystems with different phosphorus limitation degrees. Science of The Total Environment 816:151491. https://doi.org/10.1016/j.scitotenv.2021.151491

Cai X, Yao L, Hu Y, Dahlgren RA (2022) Effects of aquatic nitrogen pollution on particle-attached ammonia-oxidizing bacteria in urban freshwater mesocosms. World J Microbiol Biotechnol 38:64. https://doi.org/10.1007/s11274-022-03251-2

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11:2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

Callieri C, Hernández-Avilés S, Salcher MM, Fontaneto D, Bertoni R (2016) Distribution patterns and environmental correlates of Thaumarchaeota abundance in six deep subalpine lakes. Aquat Sci 78:215–225. https://doi.org/10.1007/s00027-015-0418-3

Carr A, Diener C, Baliga NS, Gibbons SM (2019) Use and abuse of correlation analyses in microbial ecology. ISME J 13:2647–2655. https://doi.org/10.1038/s41396-019-0459-z

Catherine A, Quiblier C, Yéprémian C, Got P, Groleau A, Vinçon-Leite B, Bernard C, Troussellier M (2008) Collapse of a Planktothrix agardhii perennial bloom and microcystin dynamics in response to reduced phosphate concentrations in a temperate lake. FEMS Microbiol Ecol 65:61–73. https://doi.org/10.1111/j.1574-6941.2008.00494.x

Chafee M, Fernàndez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, Amann RI (2018) Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J 12:237–252. https://doi.org/10.1038/ismej.2017.165

Chalar G (2009) The use of phytoplankton patterns of diversity for algal bloom management. Limnologica 39:200–208. https://doi.org/10.1016/j.limno.2008.04.001

Chiarello M, McCauley M, Villéger S, Jackson CR (2022) Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. PLoS One 17:e0264443. https://doi.org/10.1371/journal.pone.0264443

Chiu M-C, Ao S, He F, Resh VH, Cai Q (2020) Elevation shapes biodiversity patterns through metacommunity-structuring processes. Science of The Total Environment 743:140548. https://doi.org/10.1016/j.scitotenv.2020.140548

Chorus I, Welker M (eds) (2021) Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management, 2nd edn. CRC Press, London

Christiansen G, Fastner J, Erhard M, Börner T, Dittmann E (2003) Microcystin Biosynthesis in Planktothrix: Genes, Evolution, and Manipulation. Journal of bacteriology 185:564–72. https://doi.org/10.1128/JB.185.2.564-572.2003

References

Christiansen G, Kurmayer R, Liu Q, Börner T (2006) Transposons Inactivate Biosynthesis of the Nonribosomal Peptide Microcystin in Naturally Occurring Planktothrix spp. Applied and Environmental Microbiology 72:117–123. https://doi.org/10.1128/AEM.72.1.117-123.2006

Chun S-J, Cui Y, Lee JJ, Choi I-C, Oh H-M, Ahn C-Y (2020) Network analysis reveals succession of Microcystis genotypes accompanying distinctive microbial modules with recurrent patterns. Water Research 170:115326. https://doi.org/10.1016/j.watres.2019.115326

Clark JS, Gelfand AE, Woodall CW, Zhu K (2014) More than the sum of the parts: forest climate response from joint species distribution models. Ecological Applications 24:990–999. https://doi.org/10.1890/13-1015.1

Clément L, Emeric D, J GB, Laurent M, David L, Eivind H, Kristian V (2018) A data-supported history of bioinformatics tools

Cleveland CJ, Ayres RU (2004) Encyclopedia of energy. Elsevier Academic Press, Amsterdam

Cole JJ, Findlay S, Pace ML (1988) Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Marine Ecology Progress Series 43:1–10

Colquhoun D (2017) The reproducibility of research and the misinterpretation of p-values. Royal Society Open Science 4:171085. https://doi.org/10.1098/rsos.171085

Conceição LP, de Jesus Affe HM, da Silva DML, de Castro Nunes JM (2021) Spatio-temporal variation of the phytoplankton community in a tropical estuarine gradient, under the influence of river damming. Regional Studies in Marine Science 43:101642. https://doi.org/10.1016/j.rsma.2021.101642

Conroy JD, Boegman L, Zhang H, Edwards WJ, Culver DA (2011) "Dead Zone" dynamics in Lake Erie: the importance of weather and sampling intensity for calculated hypolimnetic oxygen depletion rates. Aquat Sci 73:289–304. https://doi.org/10.1007/s00027-010-0176-1

Costa MS, Costa M, Ramos V, Leão PN, Barreiro A, Vasconcelos V, Martins R (2015) Picocyanobacteria from a clade of marine Cyanobium revealed bioactive potential against microalgae, bacteria, and marine invertebrates. J Toxicol Environ Health A 78:432–442. https://doi.org/10.1080/15287394.2014.991466

Coutinho FH, Meirelles PM, Moreira APB, Paranhos RP, Dutilh BE, Thompson FL (2015) Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. PeerJ 3:e1008. https://doi.org/10.7717/peerj.1008

Cullen J, Macintyre J (1998) Behavior, physiology and the niche of depth-regulating phytoplankton

Culverhouse PF, Williams R, Reguera B, Herry V, González-Gil S (2003) Do experts make mistakes? A comparison of human and machine indentification of dinoflagellates. Marine Ecology Progress Series 247:17–25. https://doi.org/10.3354/meps247017

Cummings DE, Caccavo Jr. F, Spring S, Rosenzweig RF (1999) Ferribacterium limneticum, gen. nov., sp. nov., an Fe(III)-reducing microorganism isolated from mining-impacted freshwater lake sediments. Arch Microbiol 171:183–188. https://doi.org/10.1007/s002030050697

References

Cuypers Y, Vinçon-Leite B, Groleau A, Tassin B, Humbert J-F (2011) Impact of internal waves on the spatial distribution of Planktothrix rubescens (cyanobacteria) in an alpine lake. ISME J 5:580–589. https://doi.org/10.1038/ismej.2010.154

D'Angeli IM, Ghezzi D, Leuko S, Firrincieli A, Parise M, Fiorucci A, Vigna B, Addesso R, Baldantoni D, Carbone C, Miller AZ, Jurado V, Saiz-Jimenez C, Waele JD, Cappelletti M (2019) Geomicrobiology of a seawater-influenced active sulfuric acid cave. PLOS ONE 14:e0220706. https://doi.org/10.1371/journal.pone.0220706

Dayhoff MO, Ledley RS (1962) Comprotein: a computer program to aid primary protein structure determination. In: Proceedings of the December 4-6, 1962, fall joint computer conference. Association for Computing Machinery, New York, NY, USA, pp 262–274

de Vries FT, Griffiths RI, Bailey M, Craig H, Girlanda M, Gweon HS, Hallin S, Kaisermann A, Keith AM, Kretzschmar M, Lemanceau P, Lumini E, Mason KE, Oliver A, Ostle N, Prosser JI, Thion C, Thomson B, Bardgett RD (2018) Soil bacterial networks are less stable under drought than fungal networks. Nat Commun 9:3033. https://doi.org/10.1038/s41467-018-05516-7

Debroas D, Domaizon I, Humbert J-F, Jardillier L, Lepère C, Oudart A, Taïb N (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. FEMS Microbiology Ecology 93:1–14. https://doi.org/10.1093/femsec/fix023

del Campo J, Bass D, Keeling PJ (2020) The eukaryome: Diversity and role of microeukaryotic organisms associated with animal hosts. Functional Ecology 34:2045–2054. https://doi.org/10.1111/1365-2435.13490

Delhomme N, Sundström G, Zamani N, Lantz H, Lin Y-C, Hvidsten TR, Höppner MP, Jern P, Peer YV de, Lundeberg J, Grabherr MG, Street NR (2015) Serendipitous Meta-Transcriptomics: The Fungal Community of Norway Spruce (Picea abies). PLOS ONE 10:e0139080. https://doi.org/10.1371/journal.pone.0139080

Diehl S, Berger S, Ptacnik R, Wild A (2002) Phytoplankton, Light, and Nutrients in a Gradient of Mixing Depths: Field Experiments. Ecology 83:399–411. https://doi.org/10.1890/0012-9658(2002)083[0399:PLANIA]2.0.CO;2

Dietrich D, Hoeger S (2005) Guidance values for microcystins in water and cyanobacterial supplement products (blue-green algal supplements): a reasonable or misguided approach? Toxicol Appl Pharmacol 203:273–289. https://doi.org/10.1016/j.taap.2004.09.005

Díez B, Pedrós-Alió C, Massana R (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl Environ Microbiol 67:2932–2941. https://doi.org/10.1128/AEM.67.7.2932-2941.2001

Dodds WK (2006) Nutrients and the "dead zone": the link between nutrient ratios and dissolved oxygen in the northern Gulf of Mexico. Frontiers in Ecology and the Environment 4:211–217. https://doi.org/10.1890/1540-9295(2006)004[0211:NATDZT]2.0.CO;2

Doenz CJ, Seehausen O (2020) Rediscovery of a presumed extinct species, Salvelinus profundus, after re-oligotrophication. Ecology 101:e03065. https://doi.org/10.1002/ecy.3065

Doser JW, Finley AO, Banerjee S (2022) Joint species distribution models with imperfect detection for high-dimensional spatial data

References

Duarte C, Agustí S, Gasol J, Vaqué D, Vázquez-Domínguez E (2000) Effect of nutrient supply on the biomass structure of planktonic communities: An experimental test on a Mediterranean coastal community. Marine Ecology-progress Series - MAR ECOL-PROGR SER 206:87–95. https://doi.org/10.3354/meps206087

Ducklow HW, Purdie DA, Williams PJLeB, Davies JM (1986) Bacterioplankton: A Sink for Carbon in a Coastal Marine Plankton Community. Science 232:865–867. https://doi.org/10.1126/science.232.4752.865

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-Time DNA Sequencing from Single Polymerase Molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986

Eiler A, Bertilsson S (2004) Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environ Microbiol 6:1228–1243. https://doi.org/10.1111/j.1462-2920.2004.00657.x

Eiler A, Heinrich F, Bertilsson S (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J 6:330–342. https://doi.org/10.1038/ismej.2011.113

Eisenhauer N, Scheu S, Jousset A (2012) Bacterial Diversity Stabilizes Community Productivity. PLOS ONE 7:e34517. https://doi.org/10.1371/journal.pone.0034517

Elo M, Ketola T, Komonen A (2021) Species co-occurrence networks of ground beetles in managed grasslands. COMMUNITY ECOLOGY 22:29–40. https://doi.org/10.1007/s42974-020-00034-3

Eloe EA, Shulse CN, Fadrosh DW, Williamson SJ, Allen EE, Bartlett DH (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. Environmental Microbiology Reports 3:449–458. https://doi.org/10.1111/j.1758-2229.2010.00223.x

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML (2013) Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 4:1111–1119. https://doi.org/10.1111/2041-210X.12114

Eriksson JE, Toivola D, Meriluoto JA, Karaki H, Han YG, Hartshorne D (1990) Hepatocyte deformation induced by cyanobacterial toxins reflects inhibition of protein phosphatases. Biochem Biophys Res Commun 173:1347–1353. https://doi.org/10.1016/s0006-291x(05)80936-2

Ernst A, Becker S, Wollenzien UIA, Postius C (2003) Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis. Microbiology (Reading) 149:217–228. https://doi.org/10.1099/mic.0.25475-0

Ernst B, Hoeger SJ, O'Brien E, Dietrich DR (2009) Abundance and toxicity of Planktothrix rubescens in the pre-alpine Lake Ammersee, Germany. Harmful Algae 8:329–342. https://doi.org/10.1016/j.hal.2008.07.006

References

F. Dormann C, M. McPherson J, B. Araújo M, Bivand R, Bolliger J, Carl G, G. Davies R, Hirzel A, Jetz W, Daniel Kissling W, Kühn I, Ohlemüller R, R. Peres-Neto P, Reineking B, Schröder B, M. Schurr F, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Farthing HN, Jiang J, Henwood AJ, Fenton A, Garner TWJ, Daversa DR, Fisher MC, Montagnes DJS (2021) Microbial Grazers May Aid in Controlling Infections Caused by the Aquatic Zoosporic Fungus Batrachochytrium dendrobatidis. Frontiers in Microbiology 11

Faust K, Raes J (2016) CoNet app: inference of biological association networks using Cytoscape. F1000Res 5:1519. https://doi.org/10.12688/f1000research.9050.2

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7:129–143. https://doi.org/10.1038/nrmicro1949

Fenchel T (2008) The microbial loop – 25 years later. Journal of Experimental Marine Biology and Ecology 366:99–103. https://doi.org/10.1016/j.jembe.2008.07.013

Fenchel T, Finlay B (2008) Oxygen and the spatial structure of microbial communities. Biol Rev Camb Philos Soc 83:553–569. https://doi.org/10.1111/j.1469-185X.2008.00054.x

Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB (2013) ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. PLOS ONE 8:e67019. https://doi.org/10.1371/journal.pone.0067019

Fernández Castro B, Sepúlveda Steiner O, Knapp D, Posch T, Bouffard D, Wüest A (2021) Inhibited vertical mixing and seasonal persistence of a thin cyanobacterial layer in a stratified lake. Aquat Sci 83:38. https://doi.org/10.1007/s00027-021-00785-9

Feurstein D, Holst K, Fischer A, Dietrich DR (2009) Oatp-associated uptake and toxicity of microcystins in primary murine whole brain cells. Toxicol Appl Pharmacol 234:247–255. https://doi.org/10.1016/j.taap.2008.10.011

Feurstein D, Kleinteich J, Heussner AH, Stemmer K, Dietrich DR (2010) Investigation of Microcystin Congener–Dependent Uptake into Primary Murine Neurons. Environ Health Perspect 118:1370–1375. https://doi.org/10.1289/ehp.0901289

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science 281:237–240. https://doi.org/10.1126/science.281.5374.237

Figueiras FG, Teixeira IG, Froján M, Zúñiga D, Arbones B, Castro CG (2020) Seasonal Variability in the Microbial Plankton Community in a Semienclosed Bay Affected by Upwelling: The Role of a Nutrient Trap. Frontiers in Marine Science 7

Filker S, Sommaruga R, Vila I, Stoeck T (2016) Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. Mol Ecol 25:2286–2301. https://doi.org/10.1111/mec.13633

Finger D, Wüest A, Bossard P (2013) Effects of oligotrophication on primary production in peri-alpine lakes. Water Resources Research 49:4700–4710. https://doi.org/10.1002/wrcr.20355

References

Fischer A, Hoeger SJ, Stemmer K, Feurstein DJ, Knobeloch D, Nussler A (2010) The role of organic anion transporting polypeptides (OATPs/SLCOs) in the toxicity of different microcystin congeners in vitro: A comparison of primary human hepatocytes and OATP-transfected HEK293 cells. Toxicology and Applied Pharmacology 245. https://doi.org/10.1016/j.taap.2010.02.006

Fisher RA (1992) Statistical Methods for Research Workers. In: Kotz S, Johnson NL (eds) Breakthroughs in Statistics: Methodology and Distribution. Springer, New York, NY, pp 66–70

Fisher RA (1956) Statistical methods and scientific inference. Hafner Publishing Co., Oxford, England

Flieder M, Buongiorno J, Herbold CW, Hausmann B, Rattei T, Lloyd KG, Loy A, Wasmund K (2021) Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. ISME J 15:3159–3180. https://doi.org/10.1038/s41396-021-00992-0

Flynn KJ, Stoecker DK, Mitra A, Raven JA, Glibert PM, Hansen PJ, Granéli E, Burkholder JM (2013) Misuse of the phytoplankton–zooplankton dichotomy: the need to assign organisms as mixotrophs within plankton functional types. Journal of Plankton Research 35:3–11. https://doi.org/10.1093/plankt/fbs062

Fournier C, Riehle E, Dietrich DR, Schleheck D (2021) Is Toxin-Producing Planktothrix sp. an Emerging Species in Lake Constance? Toxins (Basel) 13:666. https://doi.org/10.3390/toxins13090666

Fraisse S, Bormans M, Lagadeuc Y (2015) Turbulence effects on phytoplankton morphofunctional traits selection. Limnology and Oceanography 60:872–884. https://doi.org/10.1002/lno.10066

Frenken T, Alacid E, Berger SA, Bourne EC, Gerphagnon M, Grossart H-P, Gsell AS, Ibelings BW, Kagami M, Küpper FC, Letcher PM, Loyau A, Miki T, Nejstgaard JC, Rasconi S, Reñé A, Rohrlack T, Rojas-Jimenez K, Schmeller DS, Scholz B, Seto K, Sime-Ngando T, Sukenik A, Van de Waal DB, Van den Wyngaert S, Van Donk E, Wolinska J, Wurzbacher C, Agha R (2017) Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. Environ Microbiol 19:3802–3822. https://doi.org/10.1111/1462-2920.13827

Frenken T, Velthuis M, de Senerpont Domis LN, Stephan S, Aben R, Kosten S, van Donk E, Van de Waal DB (2016) Warming accelerates termination of a phytoplankton spring bloom by fungal parasites. Glob Chang Biol 22:299–309. https://doi.org/10.1111/gcb.13095

Frenken T, Wolinska J, Tao Y, Rohrlack T, Agha R (2020) Infection of filamentous phytoplankton by fungal parasites enhances herbivory in pelagic food webs. Limnology and Oceanography 65:2618–2626. https://doi.org/10.1002/lno.11474

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, Delong EF (2008) Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci USA 105:3805–3810. https://doi.org/10.1073/pnas.0708897105

Fuchsman CA, Staley JT, Oakley BB, Kirkpatrick JB, Murray JW (2012) Free-living and aggregate-associated Planctomycetes in the Black Sea. FEMS Microbiol Ecol 80:402–416. https://doi.org/10.1111/j.1574-6941.2012.01306.x

Gaedeke A, Sommer U (1986) The influence of the frequency of periodic disturbances on the maintenance of phytoplankton diversity. Oecologia 71:25–28. https://doi.org/10.1007/BF00377315

References

Garcia SL, Salka I, Grossart H-P, Warnecke F (2013) Depth-discrete profiles of bacterial communities reveal pronounced spatio-temporal dynamics related to lake stratification. Environ Microbiol Rep 5:549–555. https://doi.org/10.1111/1758-2229.12044

Garneau M-È, Posch T, Pernthaler J (2015) Seasonal patterns of microcystin-producing and non-producing Planktothrix rubescens genotypes in a deep pre-alpine lake. Harmful Algae 50:21–31. https://doi.org/10.1016/j.hal.2015.10.001

Gaspar JM (2018) NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics 19:536. https://doi.org/10.1186/s12859-018-2579-2

Gauns M, Mochemadkar S, Pratihary A, Shirodkar G, Narvekar PV, Naqvi SWA (2020) Phytoplankton associated with seasonal oxygen depletion in waters of the western continental shelf of India. Journal of Marine Systems 204:103308. https://doi.org/10.1016/j.jmarsys.2020.103308

Gauthier J, Vincent AT, Charette SJ, Derome N (2019) A brief history of bioinformatics. Brief Bioinform 20:1981–1996. https://doi.org/10.1093/bib/bby063

Geelen JFM, Leuven RSEW (1986) Impact of acidification on phytoplankton and zooplankton communities. Experientia 42:486–494. https://doi.org/10.1007/BF01946686

Geng H, Belas R (2010) Molecular mechanisms underlying roseobacter–phytoplankton symbioses. Current Opinion in Biotechnology 21:332–338. https://doi.org/10.1016/j.copbio.2010.03.013

Gerphagnon M, Agha R, Martin-Creuzburg D, Bec A, Perriere F, Rad-Menéndez C, Gachon CMM, Wolinska J (2019) Comparison of sterol and fatty acid profiles of chytrids and their hosts reveals trophic upgrading of nutritionally inadequate phytoplankton by fungal parasites. Environmental Microbiology 21:949–958. https://doi.org/10.1111/1462-2920.14489

Givnish TJ (1994) Does diversity beget stability? Nature 371:113–114. https://doi.org/10.1038/371113b0

Gloor GB, Macklaim JM, Fernandes AD (2016a) Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. Journal of Computational and Graphical Statistics 25:971–979. https://doi.org/10.1080/10618600.2015.1131161

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome Datasets Are Compositional: And This Is Not Optional. Frontiers in Microbiology 8

Gloor GB, Macklaim JM, Vu M, Fernandes AD (2016b) Compositional uncertainty should not be ignored in high-throughput sequencing data analysis. Austrian Journal of Statistics 45:73–87. https://doi.org/10.17713/ajs.v45i4.122

Gong J, Dong J, Liu X, Massana R (2013) Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. Protist 164:369–379. https://doi.org/10.1016/j.protis.2012.11.006

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224. https://doi.org/10.1093/molbev/msp259

Grassl N, Kulak NA, Pichler G, Geyer PE, Jung J, Schubert S, Sinitcyn P, Cox J, Mann M (2016) Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. Genome Med 8. https://doi.org/10.1186/s13073-016-0293-0

References

Gribble KE, Anderson DM (2006) MOLECULAR PHYLOGENY OF THE HETEROTROPHIC DINOFLAGELLATES, PROTOPERIDINIUM, DIPLOPSALIS AND PREPERIDINIUM (DINOPHYCEAE), INFERRED FROM LARGE SUBUNIT rDNA1. Journal of Phycology 42:1081–1095. https://doi.org/10.1111/j.1529-8817.2006.00267.x

Griffiths BS, Philippot L (2013) Insights into the resistance and resilience of the soil microbial community. FEMS Microbiology Reviews 37:112–129. https://doi.org/10.1111/j.1574-6976.2012.00343.x

Griffiths BS, Ritz K, Bardgett RD, Cook R, Christensen S, Ekelund F, Sørensen SJ, Bååth E, Bloem J, De Ruiter PC, Dolfing J, Nicolardot B (2000) Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity–ecosystem function relationship. Oikos 90:279–294. https://doi.org/10.1034/j.1600-0706.2000.900208.x

Gronchi E, Straile D, Diehl S, Joehnk K, Peeters F (2022) Impact of climate warming on phenological asynchrony of plankton dynamics across Europe. Preprints

Grossart H-P, Simon M (1993) Limnetic macroscopic organic aggregates (lake snow): Occurrence, characteristics, and microbial dynamics in Lake Constance. Limnology and Oceanography 38:532–546. https://doi.org/10.4319/lo.1993.38.3.0532

Group JCHMPDGW (2012) Evaluation of 16S rDNA-Based Community Profiling for Human Microbiome Research. PLOS ONE 7:e39315. https://doi.org/10.1371/journal.pone.0039315

Grujcic V, Nuy JK, Salcher MM, Shabarova T, Kasalicky V, Boenigk J, Jensen M, Simek K (2018) Cryptophyta as major bacterivores in freshwater summer plankton. ISME J 12:1668–1681. https://doi.org/10.1038/s41396-018-0057-5

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41:D597–D604. https://doi.org/10.1093/nar/gks1160

Hahn MW (2009) Description of seven candidate species affiliated with the phylum Actinobacteria, representing planktonic freshwater bacteria. Int J Syst Evol Microbiol 59:112–117. https://doi.org/10.1099/ijs.0.001743-0

Hahn MW, Kasalický V, Jezbera J, Brandt U, Jezberová J, Šimek K (2010) Limnohabitans curvus gen. nov., sp. nov., a planktonic bacterium isolated from a freshwater lake. Int J Syst Evol Microbiol 60:1358–1365. https://doi.org/10.1099/ijs.0.013292-0

Hahn MW, Lünsdorf H, Wu Q, Schauer M, Höfle MG, Boenigk J, Stadler P (2003) Isolation of novel ultramicrobacteria classified as actinobacteria from five freshwater habitats in Europe and Asia. Appl Environ Microbiol 69:1442–1451. https://doi.org/10.1128/aem.69.3.1442-1451.2003

Hällfors MH, Antão LH, Itter M, Lehikoinen A, Lindholm T, Roslin T, Saastamoinen M (2020) Shifts in timing and duration of breeding for 73 boreal bird species over four decades. Proceedings of the National Academy of Sciences 117:18557–18565. https://doi.org/10.1073/pnas.1913579117

References

Halsey LG, Curran-Everett D, Vowler SL, Drummond GB (2015) The fickle P value generates irreproducible results. Nat Methods 12:179–185. https://doi.org/10.1038/nmeth.3288

Hammer AC, Pitchford JW (2006) Mixotrophy, allelopathy and the population dynamics of phagotrophic algae (cryptophytes) in the Darss Zingst Bodden estuary, southern Baltic. Marine Ecology Progress Series 328:105–115. https://doi.org/10.3354/meps328105

Haraldsson M, Gerphagnon M, Bazin P, Colombet J, Tecchio S, Sime-Ngando T, Niquil N (2018) Microbial parasites make cyanobacteria blooms less of a trophic dead end than commonly assumed. ISME J 12:1008–1020. https://doi.org/10.1038/s41396-018-0045-9

Haukka K, Kolmonen E, Hyder R, Hietala J, Vakkilainen K, Kairesalo T, Haario H, Sivonen K (2006) Effect of Nutrient Loading on Bacterioplankton Community Composition in Lake Mesocosms. Microb Ecol 51:137–146. https://doi.org/10.1007/s00248-005-0049-7

Hawinkel S, Mattiello F, Bijnens L, Thas O (2019) A broken promise: microbiome differential abundance methods do not control the false discovery rate. Brief Bioinform 20:210–221. https://doi.org/10.1093/bib/bbx104

He Y, Sen B, Zhou S, Xie N, Zhang Y, Zhang J, Wang G (2017) Distinct Seasonal Patterns of Bacterioplankton Abundance and Dominance of Phyla α-Proteobacteria and Cyanobacteria in Qinhuangdao Coastal Waters Off the Bohai Sea. Frontiers in Microbiology 8

Heil CA, Glibert PM, Fan C (2005) Prorocentrum minimum (Pavillard) Schiller: A review of a harmful algal bloom species of growing worldwide importance. Harmful Algae 4:449–470. https://doi.org/10.1016/j.hal.2004.08.003

Henson MW, Lanclos VC, Faircloth BC, Thrash JC (2018) Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. ISME J 12:1846–1860. https://doi.org/10.1038/s41396-018-0092-2

Herber J, Klotz F, Frommeyer B, Weis S, Straile D, Kolar A, Sikorski J, Egert M, Dannenmann M, Pester M (2020) A single Thaumarchaeon drives nitrification in deep oligotrophic Lake Constance. Environmental Microbiology 22:212–228. https://doi.org/10.1111/1462-2920.14840

Hewson I, Fuhrman JA (2004) Richness and Diversity of Bacterioplankton Species along an Estuarine Gradient in Moreton Bay, Australia. Applied and Environmental Microbiology 70:3425–3433. https://doi.org/10.1128/AEM.70.6.3425-3433.2004

Hingsamer P, Peeters F, Hofmann H (2014) The Consequences of Internal Waves for Phytoplankton Focusing on the Distribution and Production of Planktothrix rubescens. PLOS ONE 9:e104359. https://doi.org/10.1371/journal.pone.0104359

Hinode K, Yamaguchi H, Nishihara GN, Matsuoka K (2019) Changes in phytoplankton assemblage caused by anoxic conditions revealed with mesocosms in Omura Bay, western Japan. Nova Hedwigia 271–290. https://doi.org/10.1127/nova_hedwigia/2019/0557

Hoeger SJ, Dietrich DR, Hitzfeld BC (2002) Effect of ozonation on the removal of cyanobacterial toxins during drinking water treatment. Environ Health Perspect 110:1127–1132. https://doi.org/10.1289/ehp.021101127

References

Hoeger SJ, Shaw G, Hitzfeld BC, Dietrich DR (2004) Occurrence and elimination of cyanobacterial toxins in two Australian drinking water treatment plants. Toxicon 43:639–649. https://doi.org/10.1016/j.toxicon.2004.02.019

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. J Bacteriol 180:4765–4774

Hugoni M, Etien S, Bourges A, Lepère C, Domaizon I, Mallet C, Bronner G, Debroas D, Mary I (2013a) Dynamics of ammonia-oxidizing Archaea and Bacteria in contrasted freshwater ecosystems. Res Microbiol 164:360–370. https://doi.org/10.1016/j.resmic.2013.01.004

Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, Salter I, Agogué H, Mary I, Galand PE (2013b) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci U S A 110:6004–6009. https://doi.org/10.1073/pnas.1216863110

Hugoni M, Vellet A, Debroas D (2017) Unique and highly variable bacterial communities inhabiting the surface microlayer of an oligotrophic lake. Aquatic Microbial Ecology 79:115–125. https://doi.org/10.3354/ame01825

Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM (2018) Cyanobacterial blooms. Nat Rev Microbiol 16:471–483. https://doi.org/10.1038/s41579-018-0040-1

Huisman J, Sharples J, Stroom JM, Visser PM, Kardinaal WEA, Verspagen JMH, Sommeijer B (2004) Changes in Turbulent Mixing Shift Competition for Light Between Phytoplankton Species. Ecology 85:2960–2970. https://doi.org/10.1890/03-0763

Hunt DE, Lin Y, Church MJ, Karl DM, Tringe SG, Izzo LK, Johnson ZI (2013) Relationship between Abundance and Specific Activity of Bacterioplankton in Open Ocean Surface Waters. Appl Environ Microbiol 79:177–184. https://doi.org/10.1128/AEM.02155-12

Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PSG, Chen I-MA, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney LJ, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Kinder Haake S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Dwayne Lunsford R, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavromatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM,

Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Pop M, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, Ross MC, Russ C, Sanka RK, Sankar P, Fah Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock GM, Wilson RK, White O, The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214. https://doi.org/10.1038/nature11234

Ibrahim A, Capo E, Wessels M, Martin I, Meyer A, Schleheck D, Epp L (2021) Anthropogenic impact on the historical phytoplankton community of Lake Constance reconstructed by multimarker analysis of sediment-core environmental DNA. Molecular Ecology 30:1–17. https://doi.org/10.1111/mec.15696

Inoue K, Stoeckl K, Geist J (2017) Joint species models reveal the effects of environment on community assemblage of freshwater mussels and fishes in European rivers. Diversity and Distributions 23:284–296. https://doi.org/10.1111/ddi.12520

Jacquet S, Briand J-F, Leboulanger C, Avois-Jacquet C, Oberhaus L, Tassin B, Vinçon-Leite B, Paolini G, Druart J-C, Anneville O, Humbert J-F (2005) The proliferation of the toxic cyanobacterium Planktothrix rubescens following restoration of the largest natural French lake (Lac du Bourget). Harmful Algae 4:651–672. https://doi.org/10.1016/j.hal.2003.12.006

Jang SH, Jeong HJ, Jang SH, Jeong HJ (2020) Spatio-temporal distributions of the newly described mixotrophic dinoflagellate <italic>Yihiella yeosuensis</italic> (Suessiaceae) in Korean coastal waters and its grazing impact on prey populations. Algae 35:45–59. https://doi.org/10.4490/algae.2020.35.2.24

Jensen JP, Jeppesen E, Olrik K, Kristensen P (1994) Impact of Nutrients and Physical Factors on the Shift from Cyanobacterial to Chlorophyte Dominance in Shallow Danish Lakes. Can J Fish Aquat Sci 51:1692–1699. https://doi.org/10.1139/f94-170

Jeong HJ, Yoo YD, Kim JS, Kim TH, Kim JH, Kang NS, Yih W (2004) Mixotrophy in the Phototrophic Harmful Alga Cochlodinium polykrikoides (Dinophycean): Prey Species, the Effects of Prey Concentration, and Grazing Impact. Journal of Eukaryotic Microbiology 51:563–569. https://doi.org/10.1111/j.1550-7408.2004.tb00292.x

Jeong I-H, Kim K-H, Park J-S (2013) Analysis of bacterial diversity in sponges collected off Chujado, an Island in Korea, using barcoded 454 pyrosequencing: Analysis of a distinctive sponge group containing Chloroflexi. J Microbiol 51:570–577. https://doi.org/10.1007/s12275-013-3426-9

Jeske JT, Gallert C (2022) Microbiome Analysis via OTU and ASV-Based Pipelines-A Comparative Interpretation of Ecological Data in WWTP Systems. Bioengineering (Basel) 9:146. https://doi.org/10.3390/bioengineering9040146

Jetten MSM (2008) The microbial nitrogen cycle. Environmental Microbiology 10:2903–2909. https://doi.org/10.1111/j.1462-2920.2008.01786.x

References

Jia Y, Zhao S, Guo W, Peng L, Zhao F, Wang L, Fan G, Zhu Y, Xu D, Liu G, Wang R, Fang X, Zhang H, Kristiansen K, Zhang W, Chen J (2022) Sequencing introduced false positive rare taxa lead to biased microbial community diversity, assembly, and interaction interpretation in amplicon studies. Environmental Microbiome 17:43. https://doi.org/10.1186/s40793-022-00436-y

Jiang W, Tian X, Li L, Dong S, Zhao K, Li H, Cai Y (2019) Temporal bacterial community succession during the start-up process of biofilters in a cold-freshwater recirculating aquaculture system. Bioresource Technology 287:121441. https://doi.org/10.1016/j.biortech.2019.121441

Jochimsen MC, Kümmerlin R, Straile D (2013) Compensatory dynamics and the stability of phytoplankton biomass during four decades of eutrophication and oligotrophication. Ecology Letters 16:81–89. https://doi.org/10.1111/ele.12018

Jones RI (1998) Phytoplankton, Primary Production and Nutrient Cycling. In: Hessen DO, Tranvik LJ (eds) Aquatic Humic Substances: Ecology and Biogeochemistry. Springer, Berlin, Heidelberg, pp 145–175

Jungblut A, Neilan B (2006) Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. Archives of microbiology 185:107–14. https://doi.org/10.1007/s00203-005-0073-5

Kagami M, Miki T, Takimoto G (2014) Mycoloop: chytrids in aquatic food webs. Frontiers in Microbiology 5

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Kan J, Wang K, Chen F (2006) Temporal variation and detection limit of an estuarine bacterioplankton community analysed by denaturing gradient gel electrophoresis (DGGE). Aquatic Microbial Ecology 42:7–18. https://doi.org/10.3354/ame042007

Karczewski K, Riss HW, Meyer EI (2017) Comparison of DNA-fingerprinting (T-RFLP) and high-throughput sequencing (HTS) to assess the diversity and composition of microbial communities in groundwater ecosystems. Limnologica 67:45–53. https://doi.org/10.1016/j.limno.2017.10.001

Kim M, Morrison M, Yu Z (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. Journal of Microbiological Methods 84:81–87. https://doi.org/10.1016/j.mimet.2010.10.020

Kim S, Islam MdR, Kang I, Cho J-C (2021) Cultivation of Dominant Freshwater Bacterioplankton Lineages Using a High-Throughput Dilution-to-Extinction Culturing Approach Over a 1-Year Period. Frontiers in Microbiology 12

Kim S, Kang I, Seo J-H, Cho J-C (2019a) Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical "helper" catalase. ISME J 13:2252–2263. https://doi.org/10.1038/s41396-019-0432-x

Kim S, Kang I, Seo J-H, Cho J-C (2019b) Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. The ISME Journal 13. https://doi.org/10.1038/s41396-019-0432-x

Kim S, Kang I, Seo J-H, Cho J-C (2019c) Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J 13:2252–2263. https://doi.org/10.1038/s41396-019-0432-x

Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. Applied and Environmental Microbiology 66:1328–1333. https://doi.org/10.1128/AEM.66.4.1328-1333.2000

Klawonn I, Van den Wyngaert S, Parada AE, Arandia-Gorostidi N, Whitehouse MJ, Grossart H-P, Dekas AE (2021) Characterizing the "fungal shunt": Parasitic fungi on diatoms affect carbon flow and bacterial communities in aquatic microbial food webs. Proceedings of the National Academy of Sciences 118:e2102225118. https://doi.org/10.1073/pnas.2102225118

Klotz F, Kitzinger K, Ngugi DK, Büsing P, Littmann S, Kuypers MMM, Schink B, Pester M (2022) Quantification of archaea-driven freshwater nitrification from single cell to ecosystem levels. ISME J 16:1647–1656. https://doi.org/10.1038/s41396-022-01216-9

Knefelkamp B, Carstens K, Wiltshire KH (2007) Comparison of different filter types on chlorophyll-a retention and nutrient measurements. Journal of Experimental Marine Biology and Ecology 345:61–70. https://doi.org/10.1016/j.jembe.2007.01.008

Komárek Jiří, Anagnostidis K (2005) Cyanoprokaryota. 2. Teil/2nd Part, Oscillatoriales. Elsevier GmbH, München

Kosolapov D, Rogozin D, Gladchenko I, Kopylov A, Zakharova E (2003) Microbial sulfate reduction in a brackish meromictic steppe lake. Aquatic Ecology 37:215–226. https://doi.org/10.1023/A:1025871300917

Kovács AW, Tóth VR, Pálffy K (2018) The effects of interspecific interactions between bloom forming cyanobacteria and Scenedesmus quadricauda (chlorophyta) on their photophysiology. Acta Biol Hung 69:210–223. https://doi.org/10.1556/018.69.2018.2.9

Kozińska A, Seweryn P, Sitkiewicz I (2019) A crash course in sequencing for a microbiologist. J Appl Genet 60:103–111. https://doi.org/10.1007/s13353-019-00482-2

Krakat N, Anjum R, Demirel B, Schröder P (2017) Methodological flaws introduce strong bias into molecular analysis of microbial populations. J Appl Microbiol 122:364–377. https://doi.org/10.1111/jam.13365

Küchler-Krischun J, Kleiner J (1990) Heterogeneously nucleated calcite precipitation in Lake Constance. A short time resolution study. Aquatic Science 52:176–197. https://doi.org/10.1007/BF00902379

Kunzmann AJ, Ehret H, Yohannes E, Straile D, Rothhaupt K-O (2019) Calanoid copepod grazing affects plankton size structure and composition in a deep, large lake. Journal of Plankton Research 41:955–966. https://doi.org/10.1093/plankt/fbz067

Kurmayer R, Christiansen G, Fastner J, Börner T (2004) Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium Planktothrix spp. Environ Microbiol 6:831–841. https://doi.org/10.1111/j.1462-2920.2004.00626.x

Kurmayer R, Deng L, Entfellner E (2016) Role of toxic and bioactive secondary metabolites in colonization and bloom formation by filamentous cyanobacteria Planktothrix. Harmful Algae 54:69–86. https://doi.org/10.1016/j.hal.2016.01.004

Lachnit T, Meske D, Wahl M, Harder T, Schmitz R (2011) Epibacterial community patterns on marine macroalgae are host-specific but temporally variable. Environmental Microbiology 13:655–665. https://doi.org/10.1111/j.1462-2920.2010.02371.x

Lauterborn R (1925) Zur Kenntnis des Planktons des Bodensees und dei benachbarten Kleinseen. Mitteilungen des Badischen Landesvereins fuer Naturkunde 1:421–30

Lavrinienko A, Jernfors T, Koskimäki JJ, Pirttilä AM, Watts PC (2021) Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends in Microbiology 29:19–27. https://doi.org/10.1016/j.tim.2020.05.019

Leadbeater BSC (2015) The Choanoflagellates. Cambridge University Press

Lee S, Kang Y-C, Fuhrman JA (1995) Imperfect retention of natural bacterioplankton cells by glass fiber filters. Marine Ecology Progress Series 119:285–290

Lei F, Qu Y, Song G, Alström P, Fjeldså J (2015) The potential drivers in forming avian biodiversity hotspots in the East Himalaya Mountains of Southwest China. Integrative Zoology 10:171–181. https://doi.org/10.1111/1749-4877.12121

Leshem T, Letcher PM, Powell MJ, Sukenik A (2016) Characterization of a new chytrid species parasitic on the dinoflagellate, Peridinium gatunense. Mycologia 108:731–743. https://doi.org/10.3852/15-197

Lewis KM, van Dijken GL, Arrigo KR (2020) Changes in phytoplankton concentration now drive increased Arctic Ocean primary production. Science 369:198–202. https://doi.org/10.1126/science.aay8380

Li H, Xing P, Wu QL (2012) The high resilience of the bacterioplankton community in the face of a catastrophic disturbance by a heavy Microcystis bloom. FEMS Microbiology Ecology 82:192–201. https://doi.org/10.1111/j.1574-6941.2012.01417.x

Li J, Zhang J, Liu L, Fan Y, Li L, Yang Y, Lu Z, Zhang X (2015) Annual periodicity in planktonic bacterial and archaeal community composition of eutrophic Lake Taihu. Sci Rep 5:15488. https://doi.org/10.1038/srep15488

Li W-Q, Wu Z-J, Zong Y-Y, Wang GG, Chen F-S, Liu Y-Q, Li J-J, Fang X-M (2022) Tree species mixing enhances rhizosphere soil organic carbon mineralization of conifers in subtropical plantations. Forest Ecology and Management 516:120238. https://doi.org/10.1016/j.foreco.2022.120238

Li Y, Chen L (2014) Big Biological Data: Challenges and Opportunities. Genomics Proteomics Bioinformatics 12:187–189. https://doi.org/10.1016/j.gpb.2014.10.001

Limnologischer Zustand des Bodensees I-IG für den IGKB - Internationale Gewässerschutzkommission - Limnologischer Zustand des Sees (Grüne Berichte). https://www.igkb.org/oeffentlichkeitsarbeit/limnologischer-zustand-des-sees-gruene-berichte/. Accessed 16 Dec 2021

Lin Y-C, Chin C-P, Yang JW, Chiang K-P, Hsieh C, Gong G-C, Shih C-Y, Chen S-Y (2022) How Communities of Marine Stramenopiles Varied with Environmental and Biological Variables in the Subtropical Northwestern Pacific Ocean. Microb Ecol 83:916–928. https://doi.org/10.1007/s00248-021-01788-7

References

Lindeman RL (1991) The trophic-dynamic aspect of ecology. Bltn Mathcal Biology 53:167–191. https://doi.org/10.1007/BF02464428

Liu K, Ng HY-T, Gao Z, Liu H (2022) Selective Feeding of a Mixotrophic Dinoflagellate (Lepidodinium sp.) in Response to Experimental Warming and Inorganic Nutrient Imbalance. Frontiers in Microbiology 13

Liu K, Ng HY-T, Zhang S, Liu H (2021) Effects of temperature on a mixotrophic dinoflagellate (Lepidodinium sp.) under different nutritional strategies. Marine Ecology Progress Series 678:37–49. https://doi.org/10.3354/meps13865

Llorens-Marès T, Catalan J, Casamayor EO (2020) Taxonomy and functional interactions in upper and bottom waters of an oligotrophic high-mountain deep lake (Redon, Pyrenees) unveiled by microbial metagenomics. Sci Total Environ 707:135929. https://doi.org/10.1016/j.scitotenv.2019.135929

Llorens-Marès T, Triadó-Margarit X, Borrego CM, Dupont CL, Casamayor EO (2016) High Bacterial Diversity and Phylogenetic Novelty in Dark Euxinic Freshwaters Analysed by 16S Tag Community Profiling. Microb Ecol 71:566–574. https://doi.org/10.1007/s00248-015-0696-2

Logue JB, Langenheder S, Andersson AF, Bertilsson S, Drakare S, Lanzén A, Lindström ES (2012) Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species–area relationships. ISME J 6:1127–1136. https://doi.org/10.1038/ismej.2011.184

Long CA (1965) Sokal, Robert R., and Peter H. A. Sneath. Principles of Numerical Taxonomy. W. H. Freeman and Co., San Francisco and London. Pp. xvi + 359, illus. 1963. Price $8.50. Journal of Mammalogy 46:111–112. https://doi.org/10.2307/1377831

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. ISME J 5:169–172. https://doi.org/10.1038/ismej.2010.133

Lynch JM (2002) Resilience of the rhizosphere to anthropogenic disturbance. Biodegradation 13:21–27. https://doi.org/10.1023/A:1016333714505

Ma T, Jiang Y, Elbehery AHA, Blank S, Kurmayer R, Deng L (2020) Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake. Hydrobiologia 847:535–548. https://doi.org/10.1007/s10750-019-04118-8

MacArthur R (1955) Fluctuations of Animal Populations and a Measure of Community Stability. Ecology 36:533–536. https://doi.org/10.2307/1929601

MacKintosh C, Beattie KA, Klumpp S, Cohen P, Codd GA (1990) Cyanobacterial microcystin-LR is a potent and specific inhibitor of protein phosphatases 1 and 2A from both mammals and higher plants. FEBS Lett 264:187–192. https://doi.org/10.1016/0014-5793(90)80245-e

MacKintosh RW, Dalby KN, Campbell DG, Cohen PT, Cohen P, MacKintosh C (1995) The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1. FEBS Lett 371:236–240. https://doi.org/10.1016/0014-5793(95)00888-g

Madsen EL (2011) Microorganisms and their roles in fundamental biogeochemical cycles. Current Opinion in Biotechnology 22:456–464. https://doi.org/10.1016/j.copbio.2011.01.008

References

Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, Bihouée A, Jean G, Díaz FP, Fernández-Gómez B, Cabrera P, Gaete A, Latorre C, Gutiérrez RA, Maass A, Cambiazo V, Navarrete SA, Eveillard D, González M (2018) Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Sci Rep 8:5875. https://doi.org/10.1038/s41598-018-23931-0

Mangot J-F, Domaizon I, Taib N, Marouni N, Duffaud E, Bronner G, Debroas D (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. Environ Microbiol 15:1745–1758. https://doi.org/10.1111/1462-2920.12065

Mantzouki E, Campbell J, van Loon E, Visser P, Konstantinou I, Antoniou M, Giuliani G, Machado-Vieira D, Gurjão de Oliveira A, Maronić DŠ, Stević F, Pfeiffer TŽ, Vucelić IB, Žutinić P, Udovič MG, Plenković-Moraj A, Tsiarta N, Bláha L, Geriš R, Fránková M, Christoffersen KS, Warming TP, Feldmann T, Laas A, Panksep K, Tuvikene L, Kangro K, Häggqvist K, Salmi P, Arvola L, Fastner J, Straile D, Rothhaupt K-O, Fonvielle J, Grossart H-P, Avagianos C, Kaloudis T, Triantis T, Zervou S-K, Hiskia A, Gkelis S, Panou M, McCarthy V, Perello VC, Obertegger U, Boscaini A, Flaim G, Salmaso N, Cerasino L, Koreivienė J, Karosienė J, Kasperovičienė J, Savadova K, Vitonytė I, Haande S, Skjelbred B, Grabowska M, Karpowicz M, Chmura D, Nawrocka L, Kobos J, Mazur-Marzec H, Alcaraz-Párraga P, Wilk-Woźniak E, Krztoń W, Walusiak E, Gagala I, Mankiewicz-Boczek J, Toporowska M, Pawlik-Skowronska B, Niedźwiecki M, Pęczuła W, Napiórkowska-Krzebietke A, Dunalska J, Sieńska J, Szymański D, Kruk M, Budzyńska A, Goldyn R, Kozak A, Rosińska J, Szeląg-Wasielewska E, Domek P, Jakubowska-Krepska N, Kwasizur K, Messyasz B, Pełechata A, Pełechaty M, Kokocinski M, Madrecka B, Kostrzewska-Szlakowska I, Frąk M, Bańkowska-Sobczak A, Wasilewicz M, Ochocka A, Pasztaleniec A, Jasser I, Antão-Geraldes AM, Leira M, Hernández A, Vasconcelos V, Morais J, Vale M, Raposeiro PM, Gonçalves V, Aleksovski B, Krstić S, Nemova H, Drastichova I, Chomova L, Remec-Rekar S, Elersek T, Delgado-Martín J, García D, Cereijo JL, Gomà J, Trapote MC, Vegas-Vilarrúbia T, Obrador B, García-Murcia A, Real M, Romans E, Noguero-Ribes J, Duque DP, Fernández-Morán E, Úbeda B, Gálvez JÁ, Marcé R, Catalán N, Pérez-Martínez C, Ramos-Rodríguez E, Cillero-Castro C, Moreno-Ostos E, Blanco JM, Rodríguez V, Montes-Pérez JJ, Palomino RL, Rodríguez-Pérez E, Carballeira R, Camacho A, Picazo A, Rochera C, Santamans AC, Ferriol C, Romo S, Soria JM, Hansson L-A, Urrutia-Cordero P, Özen A, Bravo AG, Buck M, Colom-Montero W, Mustonen K, Pierson D, Yang Y, M H Verspagen J, de Senerpont Domis LN, Seelen L, Teurlincx S, Verstijnen Y, Lürling M, Maliaka V, Faassen EJ, Latour D, Carey CC, W Paerl H, Torokne A, Karan T, Demir N, Beklioğlu M, Filiz N, E Levi E, Iskin U, Bezirci G, Tavşanoğlu ÜN, Çelik K, Özhan K, Karakaya N, Koçer MAT, Yilmaz M, Maraşlıoğlu F, Fakioglu Ö, Soylu EN, Yağcı MA, Çınar Ş, Çapkın K, Yağcı A, Cesur M, Bilgin F, Bulut C, Uysal R, Köker L, Akçaalan R, Albay M, Alp MT, Özkan K, Sevindik TO, Tunca H, Önem B, Richardson J, Edwards C, Bergkemper V, O'Leary S, Beirne E, Cromie H, Ibelings BW (2018) A European Multi Lake Survey dataset of environmental variables, phytoplankton pigments and cyanotoxins. Sci Data 5:180226. https://doi.org/10.1038/sdata.2018.226

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. https://doi.org/10.1038/nature03959

Marshall W, Laybourn-Parry J (2002) The balance between photosynthesis and grazing in Antarctic mixotrophic cryptophytes during summer. Freshwater Biology 47:2060–2070. https://doi.org/10.1046/j.1365-2427.2002.00950.x

Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J, Not F (2009) Grazing rates and functional diversity of uncultured heterotrophic flagellates. ISME J 3:588–596. https://doi.org/10.1038/ismej.2008.130

MATLAB (2020) MATLAB MATLAB. version R2020a 2020. https://fr.mathworks.com/products/new_products/release2020a.html. Accessed 29 Apr 2023

Matsuo Y, Imagawa H, Nishizawa M, Shizuri Y (2005) Isolation of an Algal Morphogenesis Inducer from a Marine Bacterium. Science 307:1598–1598. https://doi.org/10.1126/science.1105486

McKindles KM, Jorge AN, McKay RM, Davis TW, Bullerjahn GS (2021) Isolation and Characterization of Rhizophydiales (Chytridiomycota), Obligate Parasites of Planktothrix agardhii in a Laurentian Great Lakes Embayment. Applied and Environmental Microbiology 87:e02308-20. https://doi.org/10.1128/AEM.02308-20

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8:e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLOS Computational Biology 10:e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meisner A, Jacquiod S, Snoek BL, ten Hooven FC, van der Putten WH (2018) Drought Legacy Effects on the Composition of Soil Fungal and Prokaryote Communities. Frontiers in Microbiology 9

Meriluoto J, Spoof L, Codd GA Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis | Wiley. In: Wiley.com. https://www.wiley.com/en-ie/Handbook+of+Cyanobacterial+Monitoring+and+Cyanotoxin+Analysis-p-9781119068686. Accessed 29 Apr 2023

Millette NC, Pierson JJ, Aceves A, Stoecker DK (2017) Mixotrophy in Heterocapsa rotundata: A mechanism for dominating the winter phytoplankton. Limnology and Oceanography 62:836–845. https://doi.org/10.1002/lno.10470

Minard G, Tikhonov G, Ovaskainen O, Saastamoinen M (2019) Variation in Melitaea cinxia gut microbiota is phylogenetically highly structured but only mildly driven by host plant microbiota, sex or parasitism. bioRxiv 510446. https://doi.org/10.1101/510446

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Moffat JF, Tompkins LS (1992) A quantitative model of intracellular growth of Legionella pneumophila in Acanthamoeba castellanii. Infection and Immunity 60:296–301. https://doi.org/10.1128/iai.60.1.296-301.1992

Morgado R, Pedroso R, Porto M, Herrera JM, Rego F, Moreira F, Beja P (2021) Preserving wintering frugivorous birds in agro-ecosystems under land use change: Lessons from intensive and

super-intensive olive orchards. Journal of Applied Ecology 58:2975–2986. https://doi.org/10.1111/1365-2664.14029

Morrissey EM, Mau RL, Schwartz E, Koch BJ, Hayer M, Hungate BA (2018) Taxonomic patterns in the nitrogen assimilation of soil prokaryotes. Environmental Microbiology 20:1112–1119. https://doi.org/10.1111/1462-2920.14051

Morton JT, Toran L, Edlund A, Metcalf JL, Lauber C, Knight R (2017) Uncovering the Horseshoe Effect in Microbial Analyses. mSystems 2:e00166-16. https://doi.org/10.1128/mSystems.00166-16

Munir S, Burhan Z, Naz T, Siddiqui PJA, Morton SL (2013) Morphotaxonomy and seasonal distribution of planktonic and benthic Prorocentrales in Karachi waters, Pakistan Northern Arabian Sea. Chin J Ocean Limnol 31:267–281. https://doi.org/10.1007/s00343-013-2150-y

Naeem S, Li S (1998) A more reliable design for biodiversity study? Nature 394:30–30. https://doi.org/10.1038/27815

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc 82:591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Naselli-Flores L (2014) Morphological analysis of phytoplankton as a tool to assess ecological state of aquatic ecosystems: the case of Lake Arancio, Sicily, Italy. Inland Waters 4:15–26. https://doi.org/10.5268/IW-4.1.686

Nearing JT, Douglas GM, Comeau AM, Langille MGI (2018) Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6:e5364. https://doi.org/10.7717/peerj.5364

Nekola JC, White PS (1999) The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

Nemergut DR, Knelman JE, Ferrenberg S, Bilinski T, Melbourne B, Jiang L, Violle C, Darcy JL, Prest T, Schmidt SK, Townsend AR (2016) Decreases in average bacterial community rRNA operon copy number during succession. ISME J 10:1147–1156. https://doi.org/10.1038/ismej.2015.191

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14–49. https://doi.org/10.1128/MMBR.00028-10

Newton RJ, Jones SE, Helmus MR, McMahon KD (2007) Phylogenetic Ecology of the Freshwater Actinobacteria acI Lineage. Applied and Environmental Microbiology 73:7169–7176. https://doi.org/10.1128/AEM.00794-07

Neyman J, Pearson ES (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses. Phil Trans Roy Soc Lond A 231:289–337. https://doi.org/10.1098/rsta.1933.0009

Novarino G (2003) A Companion to the Identification of Cryptomonad Flagellates (Cryptophyceae = Cryptomonadea). Hydrobiologia 502:225–270. https://doi.org/10.1023/B:HYDR.0000004284.12535.25

Nübel U, Garcia-Pichel F, Muyzer G (1997) PCR primers to amplify 16S rRNA genes from cyanobacteria. Appl Environ Microbiol 63:3327–3332

References

Odriozola I, Abrego N, Tláskal V, Zrůstová P, Morais D, Větrovský T, Ovaskainen O, Baldrian P (2021) Fungal Communities Are Important Determinants of Bacterial Community Composition in Deadwood. mSystems 6:e01017-20. https://doi.org/10.1128/mSystems.01017-20

Ogorelec Ž, Rudstam LG, Straile D (2022) Can young-of-the-year invasive fish keep up with young-of-the-year native fish? A comparison of feeding rates between invasive sticklebacks and whitefish. Ecology and Evolution 12:e8486. https://doi.org/10.1002/ece3.8486

Okazaki Y, Hodoki Y, Nakano S (2013) Seasonal dominance of CL500-11 bacterioplankton (phylum Chloroflexi) in the oxygenated hypolimnion of Lake Biwa, Japan. FEMS Microbiology Ecology 83:82–92. https://doi.org/10.1111/j.1574-6941.2012.01451.x

Okazaki Y, Nakano S-I (2016) Vertical partitioning of freshwater bacterioplankton community in a deep mesotrophic lake with a fully oxygenated hypolimnion (Lake Biwa, Japan). Environmental Microbiology Reports 8:780–788. https://doi.org/10.1111/1758-2229.12439

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara B, Simpson G, Solymos P, Stevens H, Wagner H (2015) Vegan: Community Ecology Package. R Package Version 22-1 2:1–2

Olrik K (1998) Ecology of mixotrophic flagellates with special reference to Chrysophyceae in Danish lakes. In: Alvarez-Cobelas M, Reynolds CS, Sánchez-Castillo P, Kristiansen J (eds) Phytoplankton and Trophic Gradients. Springer Netherlands, Dordrecht, pp 329–338

Ostermaier V, Kurmayer R (2009) Distribution and Abundance of Nontoxic Mutants of Cyanobacteria in Lakes of the Alps. Microb Ecol 58:323–333. https://doi.org/10.1007/s00248-009-9484-1

Ovaskainen O, Abrego N (2020) Joint Species Distribution Modelling: With Applications in R. Cambridge University Press, Cambridge

Ovaskainen O, Abrego N, Halme P, Dunson D (2016) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. Methods in Ecology and Evolution 7:549–555. https://doi.org/10.1111/2041-210X.12501

Ovaskainen O, Rybicki J, Abrego N (2019) What can observational data reveal about metacommunity processes? Ecography 42. https://doi.org/10.1111/ecog.04444

Ovaskainen O, Soininen J (2011) Making more out of sparse data: hierarchical modeling of species communities. Ecology 92:289–295. https://doi.org/10.1890/10-1251.1

Ovaskainen O, Tikhonov G, Dunson D, Grøtan V, Engen S, Saether B-E, Abrego N (2017a) How are species interactions structured in species-rich communities? A new method for analysing time-series data. Proceedings of the Royal Society B: Biological Sciences 284:20170768. https://doi.org/10.1098/rspb.2017.0768

Ovaskainen O, Tikhonov G, Norberg A, Guillaume Blanchet F, Duan L, Dunson D, Roslin T, Abrego N (2017b) How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters 20:561–576. https://doi.org/10.1111/ele.12757

Padilla CC, Ganesh S, Gantt S, Huhman A, Parris DJ, Sarode N, Stewart FJ (2015) Standard filtration practices may significantly distort planktonic microbial diversity estimates. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.00547

References

Paerl HW, Huisman J (2008) Climate. Blooms like it hot. Science 320:57–58. https://doi.org/10.1126/science.1155398

Pajdak-Stós A, Sobczyk M, Fiałkowska E, Kocerba-Soroka W, Fyda J (2017) The effect of three different predatory ciliate species on activated sludge microfauna. Eur J Protistol 58:87–93. https://doi.org/10.1016/j.ejop.2017.01.001

Palit K, Rath S, Chatterjee S, Das S (2022) Microbial diversity and ecological interactions of microorganisms in the mangrove ecosystem: Threats, vulnerability, and adaptations. Environ Sci Pollut Res 29:32467–32512. https://doi.org/10.1007/s11356-022-19048-7

Pålsson C, Granéli W (2003) Diurnal and seasonal variations in grazing by bacterivorous mixotrophs in an oligotrophic clearwater lake. Archiv fur Hydrobiologie 157:289–307. https://doi.org/10.1127/0003-9136/2003/0157-0289

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18:1403–1414. https://doi.org/10.1111/1462-2920.13023

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412

Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R (2008) Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 56:45–50

Park BS, Lee M, Shin K, Baek SH (2020) Response of the bacterioplankton composition to inorganic nutrient loading and phytoplankton in southern Korean coastal waters: A mesocosm study. Marine Ecology 41:e12591. https://doi.org/10.1111/maec.12591

Park C, Kim SB, Choi SH, Kim S (2021) Comparison of 16S rRNA Gene Based Microbial Profiling Using Five Next-Generation Sequencers and Various Primers. Frontiers in Microbiology 12

Patidar SK, Chokshi K, George B, Bhattacharya S, Mishra S (2014) Dominance of cyanobacterial and cryptophytic assemblage correlated to CDOM at heavy metal contamination sites of Gujarat, India. Environ Monit Assess 187:4118. https://doi.org/10.1007/s10661-014-4118-6

Pedrós-Alió C (2012) The Rare Bacterial Biosphere. Annual Review of Marine Science 4:449–466. https://doi.org/10.1146/annurev-marine-120710-100948

Peeters F, Straile D, Lorke A, Livingstone DM (2007) Earlier onset of the spring phytoplankton bloom in lakes of the temperate zone in a warmer climate. Global Change Biology 13:1898–1909. https://doi.org/10.1111/j.1365-2486.2007.01412.x

Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, Knetsch CW, Figueiredo C (2019) Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Frontiers in Microbiology 10

Peres-Neto PR, Jackson DA (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–178. https://doi.org/10.1007/s004420100720

Petri M (2006) Water Quality of Lake Constance. Handbook of Environmental Chemistry, Volume 5, Part L 127–138. https://doi.org/10.1007/698_5_018

References

Pettitt ME, Orme BAA, Blake JR, Leadbeater BSC (2002) The hydrodynamics of filter feeding in choanoflagellates. European Journal of Protistology 38:313–332. https://doi.org/10.1078/0932-4739-00854

Philippot L, Griffiths BS, Langenheder S (2021) Microbial Community Resilience across Ecosystems and Multiple Disturbances. Microbiol Mol Biol Rev 85:e00026-20. https://doi.org/10.1128/MMBR.00026-20

Pielou EC (1966) The measurement of diversity in different types of biological collections. Journal of Theoretical Biology 13:131–144. https://doi.org/10.1016/0022-5193(66)90013-0

Pierrou U (1976) The Global Phosphorus Cycle. Ecological Bulletins 75–88

Pierson DC (2012) Light and Primary Production in Lakes. In: Bengtsson L, Herschy RW, Fairbridge RW (eds) Encyclopedia of Lakes and Reservoirs. Springer Netherlands, Dordrecht, pp 485–492

Pilgrim EM, Smucker NJ, Wu H, Martinson J, Nietch CT, Molina M, Darling JA, Johnson BR (2022) Developing Indicators of Nutrient Pollution in Streams Using 16S rRNA Gene Metabarcoding of Periphyton-Associated Bacteria. Water 14:2361. https://doi.org/10.3390/w14152361

Pinhassi J, Berman T (2003a) Differential growth response of colony-forming alpha- and gamma-proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the eastern Mediterranean Sea, and the Gulf of Eilat. Appl Environ Microbiol 69:199–211. https://doi.org/10.1128/AEM.69.1.199-211.2003

Pinhassi J, Berman T (2003b) Differential Growth Response of Colony-Forming α- and γ-Proteobacteria in Dilution Culture and Nutrient Addition Experiments from Lake Kinneret (Israel), the Eastern Mediterranean Sea, and the Gulf of Eilat. Applied and Environmental Microbiology 69:199–211. https://doi.org/10.1128/AEM.69.1.199-211.2003

Poggiato G, Münkemüller T, Bystrova D, Arbel J, Clark J, Thuiller W (2021) On the Interpretations of Joint Modeling in Community Ecology. Trends in Ecology & Evolution. https://doi.org/10.1016/j.tree.2021.01.002

Pollard DA, Pollard TD, Pollard KS (2019) Empowering statistical methods for cellular and molecular biologists. MBoC 30:1359–1368. https://doi.org/10.1091/mbc.E15-02-0076

Pollet T, Humbert J-F, Tadonléké RD (2014) Planctomycetes in Lakes: Poor or Strong Competitors for Phosphorus? Appl Environ Microbiol 80:819–828. https://doi.org/10.1128/AEM.02824-13

Pollet T, Tadonléké RD, Humbert JF (2011) Spatiotemporal Changes in the Structure and Composition of a Less-Abundant Bacterial Phylum (Planctomycetes) in Two Perialpine Lakes. Applied and Environmental Microbiology 77:4811–4821. https://doi.org/10.1128/AEM.02697-10

Poorlin Ramakodi M (2021) Influence of 16S rRNA Reference Databases in Amplicon-Based Environmental Microbiome Research

Popa R, Popa R, Mashall MJ, Nguyen H, Tebo BM, Brauer S (2009) Limitations and benefits of ARISA intra-genomic diversity fingerprinting. Journal of Microbiological Methods 78:111–118. https://doi.org/10.1016/j.mimet.2009.06.005

References

Porter TM, Martin W, James TY, Longcore JE, Gleason FH, Adler PH, Letcher PM, Vilgalys R (2011) Molecular phylogeny of the Blastocladiomycota (Fungi) based on nuclear ribosomal DNA. Fungal Biology 115:381–392. https://doi.org/10.1016/j.funbio.2011.02.004

Portillo MC, Villahermosa D, Corzo A, Gonzalez JM (2011) Microbial Community Fingerprinting by Differential Display-Denaturing Gradient Gel Electrophoresis. Appl Environ Microbiol 77:351–354. https://doi.org/10.1128/AEM.01316-10

Posch T, Eugster B, Pomati F, Pernthaler J, Pitsch G, Eckert EM (2015) Network of Interactions Between Ciliates and Phytoplankton During Spring. Frontiers in Microbiology 6

Posch T, Köster O, Salcher MM, Pernthaler J (2012) Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. Nature Climate Change 2:809–813. https://doi.org/10.1038/nclimate1581

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE 5:e9490. https://doi.org/10.1371/journal.pone.0009490

Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E (2020) Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One 15:e0227434. https://doi.org/10.1371/journal.pone.0227434

Protasov ES, Axenov-Gribanov DV, Shatilina ZM, Timofeyev MA, Lane AL (2020) Freshwater Actinobacteria from sediments of the deep and ancient Lake Baikal (Russia) and their genetic potential as producers of secondary metabolites. Aquatic Microbial Ecology 84:1–14. https://doi.org/10.3354/ame01923

Pulvermüller AG, Kleiner J, Mauser W (1995) Calcite patchiness in Lake Constance as viewed by LANDSAT-TM. Aquatic Science 57:338–349. https://doi.org/10.1007/BF00878397

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41:D590–D596. https://doi.org/10.1093/nar/gks1219

R core team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing

Rajaratnam B, Sparks D (2015) MCMC-Based Inference in the Era of Big Data: A Fundamental Analysis of the Convergence Complexity of High-Dimensional Chains

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rocke E, Jing H, Liu H (2013) Phylogenetic composition and distribution of picoeukaryotes in the hypoxic northwestern coast of the Gulf of Mexico. MicrobiologyOpen 2:130–143. https://doi.org/10.1002/mbo3.57

References

Rodriguez F, Varela M, Zapata M (2002) Phytoplankton assemblages in the Gerlache and Bransfield Straits (Antarctic Peninsula) determined by light microscopy and CHEMTAX analysis of HPLC pigment data. Deep Sea Research Part II: Topical Studies in Oceanography 49:723–747. https://doi.org/10.1016/S0967-0645(01)00121-7

Rohwer RR, Hamilton JJ, Newton RJ, McMahon KD (2018) TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution. mSphere 3. https://doi.org/10.1128/mSphere.00327-18

Rosenberg E (2014) The Family Chitinophagaceae. In: The Prokaryotes: Other Major Lineages of Bacteria and The Archaea. pp 493–495

Rossi A, Boscaro V, Carducci D, Serra V, Modeo L, Verni F, Fokin SI, Petroni G (2016) Ciliate communities and hidden biodiversity in freshwater biotopes of the Pistoia province (Tuscany, Italy). European Journal of Protistology 53:11–19. https://doi.org/10.1016/j.ejop.2015.12.005

Rubbens P, Schmidt ML, Props R, Biddanda BA, Boon N, Waegeman W, Denef VJ (2019) Randomized Lasso Links Microbial Taxa with Aquatic Functional Groups Inferred from Flow Cytometry. mSystems 4:e00093-19. https://doi.org/10.1128/mSystems.00093-19

Ruber J, Geist J, Hartmann M, Millard A, Raeder U, Zubkov M, Zwirglmaier K (2018) Spatio-temporal distribution pattern of the picocyanobacterium Synechococcus in lakes of different trophic states: a comparison of flow cytometry and sequencing approaches. Hydrobiologia 811:1–16. https://doi.org/10.1007/s10750-017-3368-z

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers Y-H, Falcón LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77. https://doi.org/10.1371/journal.pbio.0050077

Sabel M, Eckmann R, Jeppesen E, Rösch R, Straile D (2020) Long-term changes in littoral fish community structure and resilience of total catch to re-oligotrophication in a large, peri-alpine European lake. Freshwater Biology 65:1325–1336. https://doi.org/10.1111/fwb.13501

Saito H, Ota T, Suzuki K, Nishioka J, Tsuda A (2006) Role of heterotrophic dinoflagellate Gyrodinium sp. in the fate of an iron induced diatom bloom. Geophysical Research Letters 33. https://doi.org/10.1029/2005GL025366

Salazar VW, Tschoeke DA, Swings J, Cosenza CA, Mattoso M, Thompson CC, Thompson FL (2020) A new genomic taxonomy system for the Synechococcus collective. Environ Microbiol 22:4557–4570. https://doi.org/10.1111/1462-2920.15173

Salcher M, Pernthaler J, Frater N, Posch T (2011a) Vertical and longitudinal distribution patterns of different bacterioplankton populations in a canyon-shaped, deep prealpine lake. Limnology and oceanography 56:2027–2039. https://doi.org/10.4319/lo.2011.56.6.2027

References

Salcher MM, Pernthaler J, Posch T (2011b) Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria "that rule the waves" (LD12). ISME J 5:1242–1252. https://doi.org/10.1038/ismej.2011.8

Salcher MM, Schaefle D, Kaspar M, Neuenschwander SM, Ghai R (2019) Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in Methylophilaceae. ISME J 13:2764–2777. https://doi.org/10.1038/s41396-019-0471-3

Salim D, Bradford WD, Freeland A, Cady G, Wang J, Pruitt SC, Gerton JL (2017) DNA replication stress restricts ribosomal DNA copy number. PLOS Genetics 13:e1007006. https://doi.org/10.1371/journal.pgen.1007006

Salmaso N (2010) Long-term phytoplankton community changes in a deep subalpine lake: responses to nutrient availability and climatic fluctuations. Freshwater Biology 55:825–846. https://doi.org/10.1111/j.1365-2427.2009.02325.x

Salmaso N, Albanese D, Capelli C, Boscaini A, Pindo M, Donati C (2018) Diversity and Cyclical Seasonal Transitions in the Bacterial Community in a Large and Deep Perialpine Lake. Microb Ecol 76:125–143. https://doi.org/10.1007/s00248-017-1120-x

Salmaso N, Boscaini A, Pindo M (2020) Unraveling the Diversity of Eukaryotic Microplankton in a Large and Deep Perialpine Lake Using a High Throughput Sequencing Approach. Frontiers in Microbiology 11

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Santhakumari S, Kannappan A, Pandian SK, Thajuddin N, Rajendran RB, Ravi AV (2016) Inhibitory effect of marine cyanobacterial extract on biofilmformation and virulence factor production of bacterial pathogens causing vibriosis inaquaculture. J Appl Phycol 28:313–324. https://doi.org/10.1007/s10811-015-0554-0

Santoferrara LF, McManus GB, Greenfield DI, Smith SA (2022) Microbial communities (bacteria, archaea and eukaryotes) in a temperate estuary during seasonal hypoxia. Aquatic Microbial Ecology 88:61–79. https://doi.org/10.3354/ame01982

Schmidt U, Conrad R (1993) Hydrogen, carbon monoxide, and methane dynamics in Lake Constance. Limnology and Oceanography 38:1214–1226. https://doi.org/10.4319/lo.1993.38.6.1214

Schuurman T, de Boer RF, Kooistra-Smid AMD, van Zwet AA (2004) Prospective Study of Use of PCR Amplification and Sequencing of 16S Ribosomal DNA from Cerebrospinal Fluid for Diagnosis of Bacterial Meningitis in a Clinical Setting. J Clin Microbiol 42:734–740. https://doi.org/10.1128/JCM.42.2.734-740.2004

Schwalbach MS, Brown M, Fuhrman JA (2005) Impact of light on marine bacterioplankton community structure. Aquatic Microbial Ecology 39:235–245. https://doi.org/10.3354/ame039235

Scoble JM, Cavalier-Smith T (2014) Scale evolution in Paraphysomonadida (Chrysophyceae): Sequence phylogeny and revised taxonomy of Paraphysomonas, new genus Clathromonas, and 25 new species. European Journal of Protistology 50:551–592. https://doi.org/10.1016/j.ejop.2014.08.001

Seyedsayamdost MR, Case RJ, Kolter R, Clardy J (2011) The Jekyll-and-Hyde chemistry of Phaeobacter gallaeciensis. Nature Chem 3:331–335. https://doi.org/10.1038/nchem.1002

References

Shade A (2023) Microbiome rescue: directing resilience of environmental microbial communities. Current Opinion in Microbiology 72:102263. https://doi.org/10.1016/j.mib.2022.102263

Shade A, Jones SE, McMahon KD (2008) The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. Environ Microbiol 10:1057–1067. https://doi.org/10.1111/j.1462-2920.2007.01527.x

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, Matulich KL, Schmidt TM, Handelsman J (2012a) Fundamentals of microbial community resistance and resilience. Front Microbiol 3:417. https://doi.org/10.3389/fmicb.2012.00417

Shade A, Read JS, Welkie DG, Kratz TK, Wu CH, McMahon KD (2011) Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. Environmental Microbiology 13:2752–2767. https://doi.org/10.1111/j.1462-2920.2011.02546.x

Shade A, Read JS, Youngblut ND, Fierer N, Knight R, Kratz TK, Lottig NR, Roden EE, Stanley EH, Stombaugh J, Whitaker RJ, Wu CH, McMahon KD (2012b) Lake microbial communities are resilient after a whole-ecosystem disturbance. ISME J 6:2153–2167. https://doi.org/10.1038/ismej.2012.56

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

Shiratori T, Thakur R, Ishida K (2017) Pseudophyllomitus vesiculosus (Larsen and Patterson 1990) Lee, 2002, a Poorly Studied Phagotrophic Biflagellate is the First Characterized Member of Stramenopile Environmental Clade MAST-6. Protist 168:439–451. https://doi.org/10.1016/j.protis.2017.06.004

Shrestha PM, Noll M, Liesack W (2007) Phylogenetic identity, growth-response time and rRNA operon copy number of soil bacteria indicate different stages of community succession. Environmental Microbiology 9:2464–2474. https://doi.org/10.1111/j.1462-2920.2007.01364.x

Sierra MA, Li Q, Pushalkar S, Paul B, Sandoval TA, Kamer AR, Corby P, Guo Y, Ruff RR, Alekseyenko AV, Li X, Saxena D (2020) The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community. Genes 11:878. https://doi.org/10.3390/genes11080878

Silverman JD, Bloom RJ, Jiang S, Durand HK, Dallow E, Mukherjee S, David LA (2021) Measuring and mitigating PCR bias in microbiota datasets. PLOS Computational Biology 17:e1009113. https://doi.org/10.1371/journal.pcbi.1009113

Silverman JD, Washburne AD, Mukherjee S, David LA (2017) A phylogenetic transform enhances analysis of compositional microbiota data. eLife 6:e21887. https://doi.org/10.7554/eLife.21887

Šimek K, Horňák K, Jezbera J, Nedoma J, Vrba J, Straškrábová V, Macek M, Dolan JR, Hahn MW (2006) Maximum growth rates and possible life strategies of different bacterioplankton groups in relation to phosphorus availability in a freshwater reservoir. Environmental Microbiology 8:1613–1624. https://doi.org/10.1111/j.1462-2920.2006.01053.x

References

Šimek K, Kasalický V, Zapomělová E, Horňák K (2011) Alga-Derived Substrates Select for Distinct Betaproteobacterial Lineages and Contribute to Niche Separation in Limnohabitans Strains. Applied and Environmental Microbiology 77:7307–7315. https://doi.org/10.1128/AEM.05107-11

Simon A (2010) Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 10 May 2020

Simpson EH (1949) Measurement of Diversity. Nature 163:688–688. https://doi.org/10.1038/163688a0

Six C, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, Scanlan DJ, Partensky F (2007) Diversity and evolution of phycobilisomes in marine Synechococcusspp.: a comparative genomics study. Genome Biology 8:R259. https://doi.org/10.1186/gb-2007-8-12-r259

SK (2016) Konstanz: Burgunderblut-Alge breitet sich im Bodensee aus. In: SÜDKURIER Online. https://www.suedkurier.de/region/kreis-konstanz/kreis-konstanz/Burgunderblut-Alge-breitet-sich-im-Bodensee-aus;art372432,8981028. Accessed 29 Apr 2023

Śliwińska-Wilczewska S, Barreiro Felpeto A, Możdżeń K, Vasconcelos V, Latała A (2019) Physiological Effects on Coexisting Microalgae of the Allelochemicals Produced by the Bloom-Forming Cyanobacteria Synechococcus sp. and Nodularia Spumigena. Toxins (Basel) 11:712. https://doi.org/10.3390/toxins11120712

Smayda TJ (2010) Adaptations and selection of harmful and other dinoflagellate species in upwelling systems. 2. Motility and migratory behaviour. Progress in Oceanography 85:71–91. https://doi.org/10.1016/j.pocean.2010.02.005

Sommer U (1986) The periodicity of phytoplankton in Lake Constance (Bodensee) in comparison to other deep lakes of central Europe. Hydrobiologia 138:1–7. https://doi.org/10.1007/BF00027228

Sommer U (1985) Seasonal Succession of Phytoplankton in Lake Constance. BioScience 35:351–357. https://doi.org/10.2307/1309903

Sommer U, Adrian R, De Senerpont Domis L, Elser JJ, Gaedke U, Ibelings B, Jeppesen E, Lürling M, Molinero JC, Mooij WM, van Donk E, Winder M (2012) Beyond the Plankton Ecology Group (PEG) Model: Mechanisms Driving Plankton Succession. Annual Review of Ecology, Evolution, and Systematics 43:429–448. https://doi.org/10.1146/annurev-ecolsys-110411-160251

Sommer U, Gaedke U, Schweizer A (1993) The first decade of oligotrophication of Lake Constance : II. The response of phytoplankton taxonomic composition. Oecologia 93:276–284. https://doi.org/10.1007/BF00317682

Sommer U, Gliwicz Z, Lampert W, Duncan A (1986) The PEG-model of seasonal succession of planktonic events in fresh waters. Archiv Fur Hydrobiologie 106

Spellerberg IF, Fedor PJ (2003) A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Global Ecology and Biogeography 12:177–179. https://doi.org/10.1046/j.1466-822X.2003.00015.x

Sperfeld M, Diekert G, Studenik S (2019) Anaerobic aromatic compound degradation in Sulfuritalea hydrogenivorans sk43H. FEMS Microbiology Ecology 95:fiy199. https://doi.org/10.1093/femsec/fiy199

References

Spiller H, Shanmugam KT (1987) Physiological conditions for nitrogen fixation in a unicellular marine cyanobacterium, Synechococcus sp. strain SF1. J Bacteriol 169:5379–5384

Stabel H-H (1986) Calcite precipitation in Lake Constance: Chemical equilibrium, sedimentation, and nucleation by algae1. Limnology and Oceanography 31:1081–1094. https://doi.org/10.4319/lo.1986.31.5.1081

Staden R (1979) A strategy of DNA sequencing employing computer programs. Nucleic Acids Research 6:2601–2610. https://doi.org/10.1093/nar/6.7.2601

Steen AD, Crits-Christoph A, Carini P, DeAngelis KM, Fierer N, Lloyd KG, Cameron Thrash J (2019) High proportions of bacteria and archaea across most biomes remain uncultured. ISME J 13:3126–3130. https://doi.org/10.1038/s41396-019-0484-y

Stefani FOP, Bell TH, Marchand C, Providencia IE de la, Yassimi AE, St-Arnaud M, Hijri M (2015) Culture-Dependent and -Independent Methods Capture Different Microbial Community Fractions in Hydrocarbon-Contaminated Soils. PLOS ONE 10:e0128272. https://doi.org/10.1371/journal.pone.0128272

Stickney HL, Hood RR, Stoecker DK (2000) The impact of mixotrophy on planktonic marine ecosystems. Ecological Modelling 125:203–230. https://doi.org/10.1016/S0304-3800(99)00181-7

Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res 43:D593–D598. https://doi.org/10.1093/nar/gku1201

Stoecker DK, Hansen PJ, Caron DA, Mitra A (2017) Mixotrophy in the Marine Plankton. Ann Rev Mar Sci 9:311–335. https://doi.org/10.1146/annurev-marine-010816-060617

Straile D (2021) Die limnologischen Institute am Bodensee: 100 Jahre Bodenseeforschung im Spiegel der Zeit-und Wissenschaftsgeschichte. Schriften des Vereins für Geschichte des Bodensees und seiner Umgebung 139:249–288

Straile D, Geller W (1998) Crustacean zooplankton in Lake Constance from 1920 to 1995: Response to eutrophication and re-oligotrophication. First publ in: Advances in Limnology 53 (1998), pp 255-274 53

Straile D, Kerimoglu O, Peeters F, Jochimsen MC, Kümmerlin R, Rinke K, Rothhaupt K-O (2010) Effects of a half a millennium winter on a deep lake – a shape of things to come? Global Change Biology 16:2844–2856. https://doi.org/10.1111/j.1365-2486.2009.02158.x

Strayer DL (2010) Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. Freshwater Biology 55:152–174. https://doi.org/10.1111/j.1365-2427.2009.02380.x

Su Y, Zhang C, Liu J, Weng Y, Li H, Zhang D (2016) Assessing the impacts of phosphorus inactive clay on phosphorus release control and phytoplankton community structure in eutrophic lakes. Environmental Pollution 219:620–630. https://doi.org/10.1016/j.envpol.2016.06.029

Taguchi S, Laws EA (1988) On the microparticles which pass through glass fiber filter type GF/F in coastal and open waters. J Plankton Res 10:999–1008. https://doi.org/10.1093/plankt/10.5.999

References

Tan X, Gu H, Ruan Y, Zhong J, Parajuli K, Hu J (2019a) Effects of nitrogen on interspecific competition between two cell-size cyanobacteria: Microcystis aeruginosa and Synechococcus sp. Harmful Algae 89:101661. https://doi.org/10.1016/j.hal.2019.101661

Tan X, Gu H, Zhang X, Parajuli K, Duan Z (2019b) Effects of Phosphorus on Interspecific Competition between two cell-size Cyanobacteria: Synechococcus sp. and Microcystis aeruginosa. Bull Environ Contam Toxicol 102:231–238. https://doi.org/10.1007/s00128-018-2527-x

Tang X, Chao J, Gong Y, Wang Y, Wilhelm SW, Gao G (2017) Spatiotemporal dynamics of bacterial community composition in large shallow eutrophic Lake Taihu: High overlap between free-living and particle-attached assemblages. Limnology and Oceanography 62:1366–1382. https://doi.org/10.1002/lno.10502

Tedersoo L, Albertsen M, Anslan S, Callahan B (2021) Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology. Applied and Environmental Microbiology 87:e00626-21. https://doi.org/10.1128/AEM.00626-21

Thakur R, Shiratori T, Ishida K (2019) Taxon-rich Multigene Phylogenetic Analyses Resolve the Phylogenetic Relationship Among Deep-branching Stramenopiles. Protist

Thébault E, Fontaine C (2010) Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. Science 329:853–856. https://doi.org/10.1126/science.1188321

Thornton DCO (2014) Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. European Journal of Phycology 49:20–46. https://doi.org/10.1080/09670262.2013.875596

Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA, Stokholm J, Al-Soud WA, Sørensen S, Bisgaard H, Waage J (2016) Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome 4:62. https://doi.org/10.1186/s40168-016-0208-8

Tikhonov G, Opedal ØH, Abrego N, Lehikoinen A, de Jonge MMJ, Oksanen J, Ovaskainen O (2020a) Joint species distribution modelling with the r-package Hmsc. Methods in Ecology and Evolution 11:442–447. https://doi.org/10.1111/2041-210X.13345

Tikhonov G, Opedal ØH, Abrego N, Lehikoinen A, de Jonge MMJ, Oksanen J, Ovaskainen O (2020b) Joint species distribution modelling with the r-package Hmsc. Methods in Ecology and Evolution 11:442–447. https://doi.org/10.1111/2041-210X.13345

Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA (2000) Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system. Chem Biol 7:753–764. https://doi.org/10.1016/s1074-5521(00)00021-1

Tilman D, Downing JA (1994) Biodiversity and stability in grasslands. Nature 367:363–365. https://doi.org/10.1038/367363a0

Torres-Beltrán M, Mueller A, Scofield M, Pachiadaki MG, Taylor C, Tyshchenko K, Michiels C, Lam P, Ulloa O, Jürgens K, Hyun J-H, Edgcomb VP, Crowe SA, Hallam SJ (2019) Sampling and Processing Methods Impact Microbial Community Structure and Potential Activity in a Seasonally Anoxic Fjord: Saanich Inlet, British Columbia. Front Mar Sci 6. https://doi.org/10.3389/fmars.2019.00132

References

Ulanowicz RE (2003) Some steps toward a central theory of ecosystem dynamics. Computational Biology and Chemistry 27:523–530. https://doi.org/10.1016/S1476-9271(03)00050-1

Urbach E, Vergin KL, Young L, Morse A, Larson GL, Giovannoni SJ (2001) Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. Limnology and Oceanography 46:557–572. https://doi.org/10.4319/lo.2001.46.3.0557

Urry L, Cain M, Wasserman S, Minorsky P, Reece J (2017) Campbell Biology (Campbell Biology Series). Pearson, New York

van de Schoot R, Broere JJ, Perryck KH, Zondervan-Zwijnenburg M, van Loey NE (2015) Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. Eur J Psychotraumatol 6:10.3402/ejpt.v6.25216. https://doi.org/10.3402/ejpt.v6.25216

van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MAG (2014) A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. Child Development 85:842–860. https://doi.org/10.1111/cdev.12169

van den Boogaart KG, Tolosana-Delgado R (2013) Analyzing Compositional Data with R. Springer, Berlin, Heidelberg

Van den Wyngaert S, Salcher MM, Pernthaler J, Zeder M, Posch T (2011) Quantitative dominance of seasonally persistent filamentous cyanobacteria (Planktothrix rubescens) in the microbial assemblages of a temperate lake. Limnology and Oceanography 56:97–109. https://doi.org/10.4319/lo.2011.56.1.0097

van der Plas F (2019) Biodiversity and ecosystem functioning in naturally assembled communities. Biological Reviews 94:1220–1245. https://doi.org/10.1111/brv.12499

van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C (2018) The Third Revolution in Sequencing Technology. Trends in Genetics 34:666–681. https://doi.org/10.1016/j.tig.2018.05.008

van Dijk EL, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: Tone down the bias. Experimental Cell Research 322:12–20. https://doi.org/10.1016/j.yexcr.2014.01.008

Vandeputte D, Kathagen G, D'hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J (2017) Quantitative microbiome profiling links gut community variation to microbial load. Nature 551:507–511. https://doi.org/10.1038/nature24460

Vaz-Moreira I, Egas C, Nunes OC, Manaia CM (2011) Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample. Antonie van Leeuwenhoek 100:245–257. https://doi.org/10.1007/s10482-011-9583-0

Verity PG, Robertson CY, Tronzo CR, Andrews MG, Nelson JR, Sieracki ME (1992) Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton. Limnology and Oceanography 37:1434–1446. https://doi.org/10.4319/lo.1992.37.7.1434

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-Generation Sequencing: From Basic Research to Diagnostics. Clinical Chemistry 55:641–658. https://doi.org/10.1373/clinchem.2008.112789

References

Vollenweider RA, Munawar M, Stadelmann P (1974) A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes. J Fish Res Bd Can 31:739–762. https://doi.org/10.1139/f74-100

Wagner T, Hansen GJA, Schliep EM, Bethke BJ, Honsey AE, Jacobson PC, Kline BC, White SL (2020) Improved understanding and prediction of freshwater fish communities through the use of joint species distribution models. Can J Fish Aquat Sci 77:1540–1551. https://doi.org/10.1139/cjfas-2019-0348

Walsby AE, Ng G, Dunn C, Davis PA (2004) Comparison of the Depth Where Planktothrix rubescens Stratifies and the Depth Where the Daily Insolation Supports Its Neutral Buoyancy. The New Phytologist 162:133–145

Walsby AE, Schanz F (2002) Light-dependent growth rate determines changes in the population of Planktothrix rubescens over the annual cycle in Lake Zürich, Switzerland. New Phytol 154:671–687. https://doi.org/10.1046/j.1469-8137.2002.00401.x

Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R (2016) Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems 1. https://doi.org/10.1128/mSystems.00009-15

Wang C, Zhang T, Wang Y, Katz LA, Gao F, Song W (2017) Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: impact of copy number variation and experimental error. Proc Biol Sci 284:20170425. https://doi.org/10.1098/rspb.2017.0425

Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FKC (2015) So Many Variables: Joint Modeling in Community Ecology. Trends in Ecology & Evolution 30:766–779. https://doi.org/10.1016/j.tree.2015.09.007

Weisbrod B, Riehle E, Helmer M, Martin-Creuzburg D, Dietrich DR (2020) Can toxin warfare against fungal parasitism influence short-term Dolichospermum bloom dynamics? - A field observation. Harmful Algae 99:101915. https://doi.org/10.1016/j.hal.2020.101915

Wickham H (2016) ggplot2. Springer International Publishing, Cham

William S, Helene F, Copeland A (2012) JGI-Bacterial-DNA-isolation-CTAB-Protocol. In: DOE Joint Genome Institute. https://jgi.doe.gov/user-programs/pmo-overview/protocols-sample-preparation-information/jgi-bacterial-dna-isolation-ctab-protocol-2012/. Accessed 29 Apr 2023

Williams S (2011) Free as in Freedom [Paperback]: Richard Stallman's Crusade for Free Software. O'Reilly Media, Inc.

Winder M, Hunter DA (2008) Temporal organization of phytoplankton communities linked to physical forcing. Oecologia 156:179–192. https://doi.org/10.1007/s00442-008-0964-7

Winter C, Köstner N, Kruspe C-P, Urban D, Muck S, Reinthaler T, Herndl GJ (2018) Mixing alters the lytic activity of viruses in the dark ocean. Ecology 99:700–713. https://doi.org/10.1002/ecy.2135

Woodhouse J, Ziegler J, Grossart H-P, Neilan B (2018) Cyanobacterial Community Composition and Bacteria–Bacteria Interactions Promote the Stable Occurrence of Particle-Associated Bacteria. Frontiers in Microbiology 9:777. https://doi.org/10.3389/fmicb.2018.00777

References

Woodhouse JN, Kinsela AS, Collins RN, Bowling LC, Honeyman GL, Holliday JK, Neilan BA (2016) Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. ISME J 10:1337–1351. https://doi.org/10.1038/ismej.2015.218

Woolway RI, Sharma S, Smol JP (2022) Lakes in Hot Water: The Impacts of a Changing Climate on Aquatic Ecosystems. BioScience 72:1050–1061. https://doi.org/10.1093/biosci/biac052

Wu J-Y, Jiang X-T, Jiang Y-X, Lu S-Y, Zou F, Zhou H-W (2010) Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiology 10:255. https://doi.org/10.1186/1471-2180-10-255

Wu P-F, Li D-X, Kong L-F, Li Y-Y, Zhang H, Xie Z-X, Lin L, Wang D-Z (2020) The diversity and biogeography of microeukaryotes in the euphotic zone of the northwestern Pacific Ocean. Science of The Total Environment 698:134289. https://doi.org/10.1016/j.scitotenv.2019.134289

Wu S, Zhao Y, Chen Y, Dong X, Wang M, Wang G (2019) Sulfur cycling in freshwater sediments: A cryptic driving force of iron deposition and phosphorus mobilization. Science of The Total Environment 657:1294–1303. https://doi.org/10.1016/j.scitotenv.2018.12.161

Wu X, Rensing C, Han D, Xiao K-Q, Dai Y, Tang Z, Liesack W, Peng J, Cui Z, Zhang F (2022) Genome-Resolved Metagenomics Reveals Distinct Phosphorus Acquisition Strategies between Soil Microbiomes. mSystems 7:e01107-21. https://doi.org/10.1128/msystems.01107-21

Xiao X, Sogge H, Lagesen K, Tooming-Klunderud A, Jakobsen KS, Rohrlack T (2014) Use of High Throughput Sequencing and Light Microscopy Show Contrasting Results in a Study of Phytoplankton Occurrence in a Freshwater Environment. PLOS ONE 9:e106510. https://doi.org/10.1371/journal.pone.0106510

Xing P, Tao Y, Luo J, Wang L, Li B, Li H, Wu QL (2020) Stratification of microbiomes during the holomictic period of Lake Fuxian, an alpine monomictic lake. Limnology and Oceanography 65:S134–S148. https://doi.org/10.1002/lno.11346

Yang J, Lv J, Liu Q, Nan F, Li B, Xie S, Feng J (2021) Seasonal and spatial patterns of eukaryotic phytoplankton communities in an urban river based on marker gene. Sci Rep 11:23147. https://doi.org/10.1038/s41598-021-02183-5

Yang X, Su J, Zheng X, Zhou Y, Tian Y, Ning X, Zheng T (2009) [16S rDNA clone library analysis of microbial diversity associated with the PSP-producing dinoflagellate Alexandrium tamarense]. Huan Jing Ke Xue 30:271–279

Yéprémian C, Gugger MF, Briand E, Catherine A, Berger C, Quiblier C, Bernard C (2007) Microcystin ecotypes in a perennial Planktothrix agardhii bloom. Water Research 41:4446–4456. https://doi.org/10.1016/j.watres.2007.06.028

Yoo YD, Seong KA, Jeong HJ, Yih W, Rho J-R, Nam SW, Kim HS (2017) Mixotrophy in the marine red-tide cryptophyte Teleaulax amphioxeia and ingestion and grazing impact of cryptophytes on natural populations of bacteria in Korean coastal waters. Harmful Algae 68:105–117. https://doi.org/10.1016/j.hal.2017.07.012

Youseif SH, Abd El-Megeed FH, Humm EA, Maymon M, Mohamed AH, Saleh SA, Hirsch AM (2021) Comparative Analysis of the Cultured and Total Bacterial Community in the Wheat Rhizosphere Microbiome Using Culture-Dependent and Culture-Independent Approaches. Microbiology Spectrum 9:e00678-21. https://doi.org/10.1128/Spectrum.00678-21

References

Yu G (2020) Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics 69:e96. https://doi.org/10.1002/cpbi.96

Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, Belk KE, Morley PS, McAllister TA (2018) Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep 8:5890. https://doi.org/10.1038/s41598-018-24280-8

Zavarzin GA, Stackebrandt E, Murray RG (1991) A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. Can J Microbiol 37:1–6. https://doi.org/10.1139/m91-001

Zeder M, Peter S, Shabarova T, Pernthaler J (2009) A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. Environ Microbiol 11:2676–2686. https://doi.org/10.1111/j.1462-2920.2009.01994.x

Zehr JP, Ward BB (2002) Nitrogen Cycling in the Ocean: New Perspectives on Processes and Paradigms. Applied and Environmental Microbiology 68:1015–1024. https://doi.org/10.1128/AEM.68.3.1015-1024.2002

Zhang C, Chen Y, Xu B, Xue Y, Ren Y (2020) Improving prediction of rare species' distribution from community data. Sci Rep 10:12230. https://doi.org/10.1038/s41598-020-69157-x

Zhang H, Jia J, Chen S, Huang T, Wang Y, Zhao Z, Feng J, Hao H, Li S, Ma X (2018) Dynamics of Bacterial and Fungal Communities during the Outbreak and Decline of an Algal Bloom in a Drinking Water Reservoir. International Journal of Environmental Research and Public Health 15:361. https://doi.org/10.3390/ijerph15020361

Zhou J, Richlen ML, Sehein TR, Kulis DM, Anderson DM, Cai Z (2018) Microbial Community Structure and Associations During a Marine Dinoflagellate Bloom. Frontiers in Microbiology 9

Zwisler W, Selje-Aßmann N, Simon M (2003) Seasonal patterns of the bacterioplankton community composition in a large mesotrophic lake. Aquatic Microbial Ecology - AQUAT MICROB ECOL 31:211–225. https://doi.org/10.3354/ame031211

R Core Team (2020). — European Environment Agency. https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006. Accessed 19 Dec 2022