UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MICROBIAL COMMUNITY ASSEMBLY, DISSOLVED ORGANIC MATTER INTERACTIONS AND BACTERIAL METABOLISM ALONG A BOREAL AQUATIC NETWORK

THESIS

PRESENTED

AS PARTIAL REQUIREMENT

OF THE DOCTORATE IN BIOLOGY

BY

MASUMI SUGA STADLER

JUNE 2025

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ASSEMBLAGE DES COMMUNAUTÉS MICROBIENNES, INTERACTIONS AVEC LA MATIÈRE ORGANIQUE DISSOUTE ET MÉTABOLISME BACTÉRIEN LE LONG D'UN RÉSEAU AQUATIQUE BORÉAL

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

MASUMI SUGA STADLER

JUIN 2025

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Paul, for enabling me to move to Montréal and pursue a Ph.D. under his guidance. I still remember our video call from my microscopy room in Austria as if it were yesterday. Since then, many years have passed, and I am deeply grateful for your thorough and honest feedback on my work and for always bringing the focus back to the bigger picture. You are a truly gifted scientist, and I have learned an incredible amount from working with you. Whatever comes next in my journey, I will cherish my experience with the CarBBAS group. I would also like to take this opportunity to thank the funding agencies that supported the del Giorgio lab and the CarBBAS research program (NSERC, Hydro-Québec), which allowed me to continue my studies without external financial support.

Having spent so many years in the CarBBAS lab, it was always a pleasure to learn from and collaborate with so many peers. A big thank you to Yves Prairie, who together with Paul created a beautiful atmosphere for young international scientists to thrive and exchange at UQAM. The goodbyes were always especially hard, as we were surrounded by such wonderful, kind-hearted people. Special thanks to Sophie Crevecoeur, Mario Muscarella, Sara Soria Pìriz, Michaela de Melo, Trista Vick-Majors and Clara Ruiz-González for all the microbe-related discussions. A big thank you to Felipe Rust, who did a phenomenal job guiding the field campaigns during the years I was privileged to participate and helping me in my very early stages of my Ph.D. I was truly fortunate to work alongside Pedro Maia Barbosa, Gabriel Duran, Joan Pere Casas-Ruiz, Maximilian Lau, Mathilde Couturier, Sofia Baliña, Bianca Rodrìguez-Cardona, Pascal Bodmer, Amir Shahabinia, Mariana Peifer Bezerra, Camille St-Arneault Sergerie, Candice Aulard, Facundo Smufer, Caroline Fink-Mercier, Sara Mercier-Blais, Yudhistir Reddy, Teresa Aguirrezabala, Cynthia Soued, Karelle Desrosiers, Marie Gérardin, Martin Demers, Gabriel Bastien-Beaudet, Tonya DelSontro, and Shoji Thottathil. Thank you for being wonderful lab and office mates - I cannot count all the cherished memories we've shared together! A special shout-out to Flora Mazoyer for aiding in the translation of this thesis' abstract into French. I am also incredibly grateful to my collaborators Ryan Hutchins, Trista Vick-Majors, and François Guillemette, who have been immensely helpful in clarifying the complexities of FT-ICR MS for me. Special thanks to my closest and longest Montréal

friends, Jihyeon Kim and Paula Reis, for the fun times, scientific exchanges, and emotional support over the years.

The La Romaine project would not have been possible without the exceptional work of Alice Parkes, Annick St-Pierre, and Sergue Paquet. Thank you for coordinating all the field and lab work over so many years. And a big thanks to all the summer students who have joined and helped in the field campaigns. Another heartfelt thank you to Alice - you are a true gem, and I cannot express how much I admire all that you do for us. I would like to thank the research associates of GRIL, Katherine Velghe and Maryline Robidoux, for helping me navigate the GRIL labs and with so many of the lab analyses.

I would also like to express my gratitude to a few past mentors who were crucial to my journey. Ms. Emi Hasegawa, thank you for being an exceptional teacher who guided me through difficult times. To three great scientists who profoundly influenced my academic journey and my fascination with microbes across ecosystem realms - Tom Battin, Andreas Richter, and Gerhard Herndl - I was truly privileged to learn about aquatic and terrestrial ecosystem sciences from you. I am also immensely grateful to the University of Vienna for providing me with a free and comprehensive education in biology. I would not be here without my two Master's supervisors: Martin Kainz, thank you for taking me in as a graduate student and teaching me about scientific philosophy and the importance of peer-review, values I continue to cherish; and, to my good friend and mentor, Elisabet Ejarque, thank you for emphasising the importance of work-life balance and for always having an open door for my stats questions. I couldn't have dreamt of a better duo to guide me in the early stages of my academic career.

I feel incredibly privileged to have grown up with a loving and supportive family. Mama and Papa, you showed me the beauty of the world, raised me bilingually, forgave my bad grades, and let me be a nerdy, indoorsy kid. Thank you for allowing me to find my passion, choose my own path, and supporting me financially so I could focus on my undergraduate and Master's studies. Knowing I had a family to fall back on made my Ph.D. less daunting than it might have been. Thank you for not pressuring me to continue in academia, for calling me out when I overworked, and for bringing

me back home almost every year to recharge and be with you. Thank you, Haji, for being the best brother. I've looked up to you every step of my life, and I don't even know what I would have done without you. Thank you for being my best friend growing up, having my back during tough times, and sharing the joy of hiking, travelling, and being silly together. Thank you for bringing Moeko and Ronja into our family - I can't wait to spend more time with all of you! Our weekly family calls have helped me stay connected and feel close to you all despite the distance and time difference. I also want to thank my grandparents - Oma, Opa, Ōmama, Ōpapa and Obāchan. A special thanks to Ōmama, who always waited for me to come home with a big, bright smile. I miss you all tremendously and wish we had more memories together, but I hope I have made you proud. Thank you all for loving me unconditionally and accepting me as I am. Dankeschön and ありがとう!

And last but not least, thank you to my new family members: my partner, Gabo, and my feline child, Evie. Gabo, you have brought so much warmth, joy, and brightness into my life. You fill me with hope, excitement, and positivity. You've been my biggest fan, and I cannot express how thankful I am for your belief in me - even more than I believe in myself. We've celebrated wins together, and you've helped me get out of bed and overcome tough days. Thank you for always being there for me and taking care of me. Meeting you in this lab is the best thing that has ever happened to me. Thank you, Evie, for letting me cuddle, play, and marvel at you. Thank you for helping me rediscover the beauty of nature through your curious eyes whenever we go for walks. You both have no idea how much the little mental breaks we have with each other have meant to me.

DEDICATION

I dedicate this work to Studio Ghibli.

Thank you for telling little me that

"It is good to be alive".

- Miyazaki Hayao

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ADs Abundance groups.

AIC Akaike information criterion.

AI_{mod} Modified aromaticity index.

ASV Amplicon sequence variant.

AWCD Average well colour development.

BA Bacterial abundance.

BCD Bacterial carbon demand.

BGE Bacterial growth efficiency.

BP Bacterial production.

BR Bacterial respiration.

CA Catchment area.

CI Confidence interval.

CSS Cumulative sum scaling.

DADA2 Divisive Amplicon Denoising Algorithm 2.

DNA Deoxyribonucleic acid.

DOC Dissolved organic carbon.

DOM Dissolved organic matter.

ESI Electrospray ionisation.

FT-ICR MS Fourier-Transform ion cyclotron resonance mass spectrometry.

FWWA Flow-weighted water age.

GAM Generalised additive model.

GIS Geographic information system.

GP Growth potential.

GR Growth rate.

GRIL Groupe de recherche interuniversitaire en limnologie.

GTDB Genome Taxonomy Database.

IOS Island of stability.

IPS Inflection point slope.

IQR Inter-quartile range.

MA Maximum absorbance.

$\mathbf{m_B}$ Bray-Curtis' distance in multi-dimensional space.

MF Molecular formula; *pl.* molecular formulae.

$\mathbf{m_S}$ Sørensen's distance in multi-dimensional space.

NMDS Non-metric multidimensional scaling.

NMS Nuclear magnetic resonance.

NOSC Nominal oxidation state of carbon.

NRT Net relatedness index.

NTI Nearest taxon index.

OTU Operational taxonomic unit.

PARAFAC Parallel-factor analyses.

PC Principal component/coordinate.

PCA Principal component analysis.

PCoA Principal coordinate analysis.

PERMANOVA Permutational analysis of variance.

PERMDISP Permutational test of multivariate dispersion.

PRT Potential reactivity threshold.

ρ Pearson's correlation coefficient.

RNA Ribonucleic acid.

rRNA Ribosomal ribonucleic acid.

RT Response time.

RQ Respiratory quotient.

SADs Species abundance distributions.

SD Standard deviation.

SPR Specific production rate.

SRR Specific respiration rate.

TR Total respiration.

Tukey's HSD Tukey's honestly significant difference test.

UT Utilisation time.

WRT Water residence time.

RÉSUMÉ

Les communautés microbiennes environnementales figurent parmi les plus complexes de la planète. En consommant la matière organique dissoute (MOD), les microbes hétérotrophes constituent la base des réseaux trophiques aquatiques et contribuent de manière significative aux émissions de gaz à effet de serre des écosystèmes aquatiques. Malgré leur importance pour les cycles biogéochimiques, notre compréhension de l'assemblage des communautés microbiennes, de leurs interactions avec la MOD et de leurs liens avec les processus métaboliques dans l'ensemble du réseau aquatique reste limitée. Cette th'ese propose une étude approfondie du réseau aquatique du bassin versant de la rivière La Romaine, situé dans la région boréale du Québec, Canada, et vise à relever plusieurs défis dans le domaine de l'écologie des communautés microbiennes localisées le long du continuum terrestre-aquatique. Premièrement, nous avons cherché à décrire l'assemblage des communautés microbiennes dans les réseaux aquatiques (chapitre 1). Deuxièmement, nous avons tenté d'identifier les unités les plus réactives au sein des assemblages de microbes et de MOD, et d'évaluer leurs interactions (chapitre 2). Enfin, nous avons mesuré différents aspects du métabolisme microbien et identifié comment leurs variations sont liées à celles de l'environnement, de l'hydrologie, de la composition des assemblages microbiens et de la MOD dans différents milieux aquatiques (chapitre 3).

Dans le chapitre 1, nous avons étudié l'assemblage des communautés microbiennes le long d'un continuum intégrant le milieu terrestre, les milieux d'eau douce et l'estuaire, en mesurant où et quand, le long de ce continuum, les communautés basées sur l'ADN et l'ARN divergeaient ou convergeaient. Ce système dynamique nous a permis de proposer une approche novatrice pour évaluer à quel moment les communautés étaient plutôt dominées par des effets de masse ou sélectionnées par l'environnement. Les résultats ont montré que les effets de masse étaient toujours présents, mais que la saisonalité modulait l'importance de la sélection environnementale dans les différentes sections du continuum. Les taxons provenaient de la totalité de la courbe d'abondance par rang, indiquant que les taxons abondants et les taxons rares jouent tous deux un rôle majeur dans la formation des communautés microbiennes actives. La portion active des communautés microbiennes provenait surtout des taxons d'origine terrestre, tandis qu'une grande fraction des microbes inactifs des eaux douces se retrouvait passivement transportée le long du cours d'eau, héritages des processus à l'oeuvre en amont.

Dans le chapitre 2, nous avons développé une approche novatrice pour modéliser les patrons spatiaux individuels des formules moléculaires de la MOD et des unités taxonomiques opérationnelles microbiennes le long d'un gradient d'âge de l'eau pondéré par le débit, incluant sols, ruisseaux, rivières, réservoirs et lacs. Nous avons identifié qu'environ 44% des formules de la MOD et seulement 7.5% des taxons microbiens semblaient réactifs aux changements environnementaux. La réactivité de la MOD était variable selon les saisons et les années, contrairement aux fractions microbiennes, qui restaient relativement stables, suggérant que le pool stable/persistant de la MOD est plus variable qu'on ne le pensait auparavant. Nous avons ensuite établi des corrélations entre les fractions réactives extraites de la MOD et les unités microbiennes afin d'évaluer les potentielles relations de cause à effet. Après avoir éliminé les corrélations suspectes ou douteuses, nous avons extrait les corrélations significatives entre les deux ensembles, révélant des variations saisonnières

dans les interactions MOD-microbes : les composés aromatiques et frais de la MOD semblaient consommés au printemps, tandis que les composés partiellement dégradés étaient préférentiellement utilisés en été. En outre, les groupes de composés ayant un état de dégradation similaire mais des tailles moléculaires variées semblaient sélectionner des phylums microbiens spécifiques dans les eaux plus vieilles, suggérant que les processus de sélection varient selon la saison et la disponibilité des ressources.

Enfin, le chapitre 3 explore les facteurs qui gouvernent les différents aspects du métabolisme microbien le long du continuum aquatique. Nous avons mesuré plusieurs indicateurs métaboliques, tels que la respiration bactérienne, la production bactérienne et les profils d'utilisation du carbone (BIOLOG EcoPlates). Des analyses multivariées des facteurs environnementaux et hydrologiques, de la structure des communautés microbiennes globale (ADN) et actives (ARN), ainsi que de la composition totale et chimiquement détaillée de la MOD, nous ont permis d'extraire leurs gradients saisonniers et spatiaux respectifs et de les relier aux variations du métabolisme bactérien. Cette approche a montré que les changements dans la production bactérienne sont associés à des modifications saisonnières de la composition des communautés microbiennes actives et globale. Les taux de respiration spécifiques étaient liés aux variations des facteurs environnementaux et de la composition de la MOD, en particulier aux variations de deux pools relativement frais, de taille moyenne et à faible teneur en azote. Ces résultats suggèrent que les changements de composition des communautés microbiennes et de la MOD le long des gradients hydrologiques influencent divers aspects du métabolisme bactérien.

En résumé, ces résultats révèlent l'influence déterminante de la connectivité hydrologique et de la saisonnalité sur l'assemblage des communautés microbiennes, les interactions MOD-microbes et le métabolisme microbien dans les systèmes aquatiques. Cette thèse éclaire des aspects fondamentaux de l'écologie et du métabolisme des communautés microbiennes discernables uniquement par la prise en compte de la connectivité intrinsèque des réseaux aquatiques et de leurs dynamiques saisonnières. Ainsi, ces résultats fournissent un cadre contextuel hydrologique et saisonnier aux nombreuses études à venir qui seront limitées à un écosystème isolé. De plus, cette thèse propose plusieurs cadres conceptuels et approches novateurs pour faire progresser notre compréhension de l'interconnexion naturelle des communautés microbiennes et des assemblages de la MOD le long des gradients hydrologiques et saisonniers.

Mots-clés: assemblage des communautés microbiennes, matière organique dissoute, métabolisme microbien, connectivité hydrologique, écologie des bassins versants, continuum terrestre-aquatique

ABSTRACT

Environmental microbial communities are among the most complex on Earth. By consuming dissolved organic matter (DOM), heterotrophic microbes form the base of aquatic food webs and significantly contribute to aquatic greenhouse gas emissions. Despite their importance in biogeochemical cycling, our understanding of how microbial communities assemble, interact with DOM, and link to metabolic processes across the entire aquatic network remains limited. This thesis provides a comprehensive aquatic network study of La Romaine River watershed located in boreal Québec, Canada, and we aimed to address multiple challenges in microbial community ecology along the terrestrial-aquatic continuum. Firstly, we sought to uncover microbial community assembly within aquatic networks (Chapter 1). Secondly, we attempted to discern the most reactive units of both microbial and DOM assemblages and evaluate their interactions (Chapter 2). Finally, we measured various dimensions of microbial metabolism and identified how their shifts relate to environmental, hydrological, microbial and DOM compositional shifts across multiple aquatic habitats (Chapter 3).

In chapter 1, we focused on understanding microbial community assembly along a terrestrial-freshwater-estuarine continuum by measuring when and where along the continuum, DNA and RNA-based communities diverged or converged. This dynamic framework enabled us to provide a novel approach to evaluate when communities are dominated by mass effects or environmental selection. The results showed that mass effects are always present, however, where in the aquatic continuum environmental selection becomes strongest was modulated by seasonality. Taxa were recruited across the entire rank abundance curve, indicating that not only abundant but also rare taxa play a major role in forming the active microbial community. Furthermore, the active portion of microbial communities was mostly traced back to terrestrially derived taxa, whereas a large fraction of inactive freshwater microbes was carried downstream, encompassing the legacy of upstream assembly processes.

In chapter 2, we developed a novel framework to model individual spatial patterns of both DOM molecular formulae and microbial operational taxonomic units along a flow-weighted water age gradient, from soils, streams, rivers to reservoirs and lakes. We identified that approximately 44% of DOM formulae and only 7.5% of microbial taxa appeared to be reactive to environmental changes. DOM reactivity varied significantly with seasons and years, unlike microbial fractions, which remained relatively stable, implying that the stable/persistent DOM pool is more variable than previously thought. We subsequently correlated the extracted reactive DOM and microbial moieties against each other to evaluate which relationships may potentially be causal. After removal of many suspect spurious correlations, we were able to extract a small set of meaningful correlations between the two assemblages, which revealed seasonal shifts in DOM-microbe interactions: fresh, aromatic DOM compounds appeared to be consumed in spring, while partially decomposed compounds appeared to be preferentially utilised in summer. Additionally, DOM clusters of similar degradation state but varying molecular size appeared to select for distinct microbial phyla in systems with older water ages, suggesting selective processes that vary by season and resource availability.

Finally, chapter 3 explores how various dimensions of microbial metabolism are governed along the aquatic continuum. To do so, we measured multiple metabolism metrics that included bacterial respiration, production, and carbon utilisation profiles (BIOLOG EcoPlates). Multivariate analyses of environmental and hydrological factors, bulk (DNA) and active (RNA) community structure, and bulk and chemically distinct DOM cluster compositions, allowed us to extract their individual seasonal and habitat gradients and relate them to changes in bacterial metabolism. This approach showed that shifts in bacterial production occur with seasonal changes in bulk and active microbial community composition. Whereas specific respiration rates linked to changes in environmental factors and DOM composition, particularly to two relatively fresh, mid-sized and low nitrogen DOM clusters. These results imply that microbial and DOM compositional changes along hydrological gradients influence distinct aspects of bacterial metabolism.

Taken together, these findings reveal the strong influence of hydrological connectivity and seasonality on microbial community assembly, DOM-microbe interactions, and microbial metabolism in aquatic systems. This thesis elucidates crucial aspects of microbial community ecology and metabolism that only emerge when the inherent connectedness of aquatic networks and their seasonal dynamics are considered. Hence, the collective results provide hydrological, and seasonal context to many future single-ecosystem studies. Additionally, this thesis provides several novel conceptual frameworks and approaches to advance our understanding of the interconnected nature of microbial communities and dissolved organic matter assemblages along hydrological and seasonal gradients.

Keywords: microbial community assembly, dissolved organic matter, microbial metabolism, hydrological connectivity, watershed ecology, terrestrial-aquatic continuum

INTRODUCTION

0.1 Context

0.1.1 Microorganisms and their role in freshwater carbon cycling

In the era of human-induced climate change (IPCC Core Writing Team, 2023), efforts to evaluate global carbon (C) budgets and unravel the mechanisms driving their dynamics have intensified (Cole *et al.*, 2007; Battin *et al.*, 2009; Tranvik *et al.*, 2009). Traditionally, streams and rivers were viewed as passive conduits, transporting primarily 'recalcitrant' carbon from terrestrial ecosystems to the oceans (e.g., Moran *et al.* (1991); Smith and Hollibaugh (1993)). However, this perspective has been re-examined, as inland waters are now recognised to transfer approximately 0.9 Pg C annually (Drake *et al.*, 2018), while simultaneously mineralising, transforming, and storing a significant fraction of this carbon during downstream movement (Regnier *et al.*, 2013; Drake *et al.*, 2018). Microbial communities play a crucial role in the transformation, consumption, and production of such organic carbon (Lennon and Pfaff, 2005; Guillemette and del Giorgio, 2012), however, the DOM pool also influences what microbial taxa may thrive (Judd *et al.*, 2006; Zhou *et al.*, 2024). This bi-directional relation of microbial community composition, their metabolism and dissolved organic matter (DOM) has been difficult to unravel over decades of studies (Tanentzap *et al.*, 2019).

Freshwater systems encompass a hydrological continuum that flows from upland sources through streams, rivers, and lakes before eventually reaching the ocean. Along this journey, extended water retention times influence the biochemical characteristics of DOM (Massicotte and Frenette, 2011; Casas-Ruiz *et al.*, 2020), as well as the structure, function, and abundance of microbial communities (Lindström *et al.*, 2006; Crump *et al.*, 2007; Besemer *et al.*, 2013). Variations in catchment characteristics, hydrology, environmental factors, and biological interactions all influence DOM (Kothawala *et al.*, 2015; Orlova *et al.*, 2024) and microbial community composition (Niño-García *et al.*, 2016a; Kraemer *et al.*, 2020) across the entire aquatic network. Variations in seasons additionally add complexity to understanding whole network dynamics (Crump *et al.*, 2003; Singh *et al.*, 2014; Paruch *et al.*, 2020). However, studies that examine microbial community assembly, DOM dynamics and microbial metabolism across large spatial and temporal gradients remain limited.

0.1.2 Microbial community assembly

0.1.2.1 Deterministic and stochastic processes

Since the concept of 'everything is everywhere, but the environment selects' was coined (Becking, 1934; de Wit and Bouvier, 2006), niche-theory based selection such as environmental filtering (i.e., deterministic processes) has been commonly assumed to be the main driver of microbial community assembly. However, recent evidence suggests a combination of deterministic and stochastic processes (Hanson *et al.*, 2012; Zhou and Ning, 2017) governing how communities come to be. Which of these processes is proportionally more important will also largely depend on the physical characteristics of the system studied (Zhou *et al.*, 2014), and there have been increasing efforts to attribute relative contributions of these contrasting processes to observed community patterns (Stegen *et al.*, 2013, 2015).

Microbial community assembly has been defined into four fundamental processes – selection, dispersal, diversification and drift – by adopting concepts from population genetics and community ecology (Vellend, 2010; Hanson *et al.*, 2012; Vellend, 2016). These four processes can be characterised by their degree of determinism and stochasticity, which harbour multiple facets of processes within themselves (Zhou and Ning, 2017). Although many approaches have been utilised to study deterministic and stochastic processes such as multivariate approaches and variance partitioning, these mostly fail to quantify their individual contribution (Zhou and Ning, 2017). An increasingly popular approach to quantify the partitioning of all four processes is to combine metrics derived from taxonomic and phylogenetic diversity (Stegen *et al.*, 2013, 2015; Zhou and Ning, 2017; Jia *et al.*, 2018). By testing observed phylogenetic diversity patterns against a null model, the βNTI (β nearest-taxon index) indicates whether a community shows tendencies of phylogenetic clustering (i.e., homogeneous selection such as environmental filtering) or overdispersion (i.e., heterogeneous selection such as competitive exclusion). The fraction that has been identified as random by the βNTI can be further partitioned by the β-diversity metric RC_{Bray} (modified Raup-Crick index) into homogenising dispersal and dispersal limitation. Other fractions that were not classified as any of the above processes are considered as 'weak' selection, dispersal, diversification and/or drift processes (Stegen *et al.*, 2013, 2015; Zhou and Ning, 2017).

Although this approach is the first to enable quantification of these ecological processes, it is underlain by major assumptions. Partitioning of selection processes based on phylogenetic diversity assumes niche conservatism, meaning that species relatedness correlates with their functional traits. Recently, many microbial traits have been shown to be phylogenetically conserved (Martiny *et al.*, 2015) but it was also noted that more complex traits encoded by many genes are more likely to be conserved and simpler traits that involve fewer genes are rather phylogenetically dispersed (Martiny *et al.*, 2013). Simultaneously, horizontal gene transfer (Rainey and Travisano, 1998; Ochman *et al.*, 2000; Papke and Gogarten, 2012) can mask phylogenetic signals (Doolittle, 1999) and phylogenetic distance estimations can be prone to a variety of methodological uncertainties such as sequencing errors and differing tree construction approaches (Martin, 2013). Additionally, the increasing literature on potential 'ecotypes' within a single species that show intraspecies niche-segregation can blur patterns based on phylogenetic conservatism (Hahn and Pöckl, 2005; Achtman and Wagner, 2008; Ackermann, 2015; Chase *et al.*, 2018). Finally, recent evidence suggests that subsets of microbial communities can underlie differing assembly processes, which is completely neglected in a whole-community approach (Niño-García *et al.*, 2016b; Monard *et al.*, 2016). However, instead of applying these metrics on the whole-community, there is a great potential in applying this framework to subsets of microbial communities (e.g., reactive vs. unreactive).

0.1.2.2 Spatio-temporal patterns in microbial communities

With increasing availability of molecular techniques to microbial ecology, several studies have been investigating spatial and more rarely temporal patterns. Similarly to many macro-biological observations, microbial communities in freshwaters exhibit a distance-decay distribution of community similarity and a taxa-area relationship (Green *et al.*, 2004; Martiny *et al.*, 2006; Soininen *et al.*, 2011). While spatial structuring of microbial communities in freshwaters seems to be less influenced by dispersal limitation (Crump *et al.*, 2007; der Gucht *et al.*, 2007; Jones and McMahon, 2009; Nelson *et al.*, 2009) including boreal systems (Niño-García *et al.*, 2016a), a fine-scale horizontal lake study has shown that spatial structuring is also not necessarily habitat specific (i.e., littoral vs pelagic). It was rather a true distance-decay relationship where compositional similarity halves across 4 km (Jones *et al.*, 2012). The authors concluded that the rates of biological and eco-

logical interactions drive assembly more rapidly than rates of water movement and turbulence in lakes, while other authors suggest that spatio-temporal microbial turnover is driven by both intrinsic (e.g., body size, dispersal rate) and extrinsic (e.g., ecosystem size, isolation) factors (Soininen, 2010). However, a true distance-decay could result solely from a common source of these microorganisms (e.g., terrestrial advection; Ruiz-González *et al.* (2015b)) that are gradually diluted and passively transported along the continuum rather than being environmentally selected (Niño-García *et al.*, 2017). Indeed, a gradual decline of soil-derived taxa has been found along a river continuum (Savio *et al.*, 2015). Furthermore, functional changes in communities resulted from a gradual microbial succession along the river continuum that is mainly driven by water residence time (Reed *et al.*, 2014).

Overall, the observed spatio-temporal patterns in microbial composition are to a large extent driven by dynamics within the reactive members that are especially sensitive to physical and chemical changes in the environment (Niño-García *et al.*, 2017; Nelson *et al.*, 2009), whereas stochastic or even chaotic patterns (Fernández *et al.*, 1999) in the passive fraction of microbes adds noise to compositional changes. Whether determinism or stochasticity drives microbial assembly and community composition will largely depend on the hydrological conditions given in a certain location at a certain time (Niño-García *et al.*, 2016a). To understand the high spatio-temporal variation of microbial communities, the vast differences in physical conditions that can affect residence time, light availability and consequently changing food-web dynamics have to be accounted for.

0.1.2.3 Death and dormancy

The inability of DNA based approaches to assess the activity of microorganisms is a matter of concern in the scientific community given the presence of dormant and dead microbes in environmental communities (Jones and Lennon, 2010; Lennon and Jones, 2011). Hence, several methods to gain insight into microbial activity have been proposed (Singer *et al.*, 2017). Although understanding of activity is often essential for certain scientific questions, most approaches involve intensive laboratory work such as incubation experiments with labelled substrate amendments (e.g., DNA-Stable Isotope probing) and/or marking microbes with fluorescent stains taken up by active members of the community for microscopy or flow cytometry analyses (e.g., 5-cyano-2,3-ditolyl

tetrazolium chloride (CTC), Fluorescence In-Situ Hybridisation).

At the same time, the only ambient whole-community approach that allows to extract also the species identity of microbes is to sequence the 16S rRNA transcription as complementary DNA (cDNA, hereafter RNA). RNA was found to correlate with microbial growth rates, thus RNA:DNA ratios have been utilised to measure the potential level of activity of microbial members within a natural sample. The major advantage of RNA-based approaches is the potential of carrying out parallel DNA and RNA extractions from the same environmental sample, and therefore provides an idea of the potential degree of activity on all the taxa within the community without need for further intervention. As any method, RNA:DNA ratios have been criticised for potential caveats that arise from evidence indicating that different microbes can harbour differing numbers of rRNA copies (Klappenbach *et al.*, 2000), and the fact that some dormant microorganisms may accumulate rRNA as a strategy to quickly respond to environmental changes (Sukenik *et al.*, 2012). Nevertheless, it was also demonstrated that RNA:DNA ratios perform worse in classifying the dormant rather than the active fraction of microbial communities (Steven *et al.*, 2017). Still, few approaches can investigate the potential of activity or - to be more conservative - protein synthesis on a whole community level of environmental samples (Blazewicz *et al.*, 2013). Recently, a soil study found similar patterns when comparing CTC staining, a marker for respiratory activity, and RNA:DNA ratios as activity metrics. They also reported a low correlation of ribosomal operon number to the measured RNA:DNA ratios (Bowsher *et al.*, 2019).

Accordingly, evidence is ambivalent and a further debate is expected to continue revolving the applicability of RNA as a measure of activity. Indeed, as a non quantitative measure, RNA:DNA ratios cannot be used to infer actual rates of metabolism and thus a direct linkage to rates of ecosystem processes is impossible (Blazewicz *et al.*, 2013). Nevertheless, consideration of RNA will certainly provide a more dynamic perspective on microbial communities compared to solely DNA approaches, which include for example relic DNA (Carini *et al.*, 2016; Lennon *et al.*, 2018). RNA:DNA ratios will be especially meaningful when applied within a spatio-temporal framework. For example, a marine study found that the rare microbial fraction had a significantly higher RNA:DNA ratio compared to the abundant fraction suggesting a higher activity potential that challenges our con-

ventional understanding that abundant members are mainly responsible for ecosystem dynamics and functioning (Campbell *et al.*, 2011; Wilhelm *et al.*, 2014; Morrissey *et al.*, 2016). RNA:DNA ratios encompass past, present and future activity and thus emerging spatio-temporal patterns accompanied with environmental fluctuations provide valuable insights into differing bacterial life strategies that are widely unexplored.

0.1.3 Assembly of dissolved organic matter

DOM is defined as the fraction of dissolved compounds that pass through a 0.7 μm or 0.2 μm filter (varies between studies). It not only contains C but other essential elements such as nitrogen (N), phosphorous (P) and sulfur (S). Natural DOM is considered one of the most complex chemical mixtures and thus represent a major challenge in analytical chemistry. Much of the insight into DOM has been achieved through bulk measurements of DOC and DON. Although bulk measurements give valuable results for numerical models, it does not decipher fine-scale changes that may interact with their heterotrophic consumers (Repeta, 2015). To understand DOM composition, freshwater studies have largely characterised the chromophoric fraction of DOM with absorbance and fluorescence approaches, however, in recent years high-resolution approaches such as fourier-transform ion cyclotron mass spectrometry (FT-ICR MS) have become more accessible to biogeochemists and ecologists.

With the advent of FT-ICR MS, ecological concepts, such as the notion of DOM assembly have recently been applied within the field of DOM research (Danczak *et al.*, 2020; Stadler *et al.*, 2023). The afore mentioned assembly metrics used in microbial ecology could potentially be applied to DOM molecules, which essentially share similar assembly processes as microorganisms. Certain DOM molecules underlie selection as of being selectively consumed or produced in specific habitats by biological and photochemical processes (Medeiros *et al.*, 2017; Seidel *et al.*, 2015). Differing molecules disperse along landscapes with the movement of water (dispersal; Hutchins *et al.* (2017)) and can form new structures by microbial and/or photochemical transformation (diversification; Osterholz *et al.* (2015); Cory and Kling (2018); Noriega-Ortega *et al.* (2019)). Finally, processes such as drift may occur through random chemical reactions (Chin *et al.*, 1998). This indeed very ecological approach to study DOM in a spatio-temporal manner has interpretational

challenges but it nevertheless provides an opportunity to employ and explore the detailed data obtained from fine-scale DOM approaches.

0.1.3.1 Persistence

From early DOC measurements, studies have been inspired by the relatively stable manner of DOC across ecosystems. Especially in the light of climate change, understanding not only the formation but the maintenance of this pool has become of major interest. FT-ICR MS approaches revealed a fraction in molecular formulae that were stable across the Atlantic and Southern Ocean, which was termed the 'island of stability' (IOS) due to its location in the commonly depicted Van-Krevelen space (Lechtenfeld *et al.*, 2014). The island describes compounds that are less saturated and more oxygenated as the most resistant to degradation, however, they do overlap substantially with the compounds classified as CRAM. Additionally, they are characterised with low N and no S and thus are thought to be of low nutritive value for microorganisms.

The river continuum concept (Vannote *et al.*, 1980) described how DOM diversity decreases from small streams to higher stream-order rivers. Degradation along the continuum leaves only refractory material behind that persists subsequently. Independently of high resolution molecular approaches, fluorescence and parallel-factor analyses (PARAFAC) have shown re-occuring PARAFAC components characterised as more humic-like components across environments (Ishii and Boyer, 2012). Since, Kellerman *et al.* (2015) has described a link between the molecular elemental composition and structural features to persistence within lakes. The authors describe a potential reactivity continuum from aromatic to aliphatic and from a high to low nominal oxidation state of carbon (NOSC). Although reactivity may be first driven by extrinsic factors such as availability of light, sorption to mineral particles and hydraulic retention time (Catalán *et al.*, 2016), the major control of degradation may switch to intrinsic properties of DOM. Recently, more evident results were found through a comprehensive study using FT-ICR MS analysis for 37 sites sampled along the aquatic continuum and the ocean (Kellerman *et al.*, 2018). The authors were able to show the existence of the IOS across the aquatic realm. However, they were also able to show that two autochthonous end-members (Antarctic lake, Pacific Ocean) were distinct from each other in terms of age, percentage of N and degradation state. Thus, they conclude that time is the major driving

force of DOM diversity and degradation state, which is also in agreement to what Catalán *et al.* (2016) have proposed on DOM decay rates.

While the IOS pre-dominantly describes stable compounds of low molecular weight, Repeta *et al.* (2002) showed via NMR that there is a persistent fraction across freshwater and marine samples within the high-molecular weight pool characterised by seven neutral sugars (i.e., glucose, galactose, mannose, xylose, arabinose, fucose, rhamnose), acylheteropolysaccharides (APS) and aminosugars (Aluwihare *et al.*, 1997; Repeta *et al.*, 2002). While the authors argue that both freshwater and marine microbiomes seem to be capable of producing this universal pool of HMW molecules, it is interesting how HMW molecules persist across aquatic ecosystems despite the size-reactivity continuum concept that postulates a correlation of size to its degradability (Benner and Amon, 2015). Recently, size fractionation approaches have shown that HMW has an aged signature of 135 - 2,700 years, while LMW hydrophilic and hydrophobic compounds were even older with 2,850 - 15,000 years and 2,470 - 6,680 years, respectively (Zigah *et al.*, 2017). They were also able to show that the major neutral sugars described by (Repeta *et al.*, 2002) had similar $\triangle^{14}$C values to the dissolved inorganic carbon (DIC) pool, thus even though showing ubiquitous behaviour, these compounds may be preferentially metabolised. The HMW mixture was described as 90% semi-labile and 10% 'refractory' humic carbon. However, evidence also suggests that a significant fraction of the humic substances is relatively young (Zigah *et al.*, 2017).

0.1.4 Interactions between bacterioplankton and dissolved organic matter

Hydrology mainly influences the release and production of DOM (Fasching *et al.*, 2016; Butturini *et al.*, 2016; Casas-Ruiz *et al.*, 2017), which leads to varying DOM composition, ultimately also affecting the dominant source and degradation pathways from streams to the ocean (Lapierre and Giorgio, 2014; Jones *et al.*, 2016; Massicotte *et al.*, 2017; Hutchins *et al.*, 2017; Cory and Kling, 2018; Kellerman *et al.*, 2018). Especially in streams and rivers, substantial amounts of terrestrial OM characterised by a dominance of humic and fulvic acids enters the aquatic environment. It is suggested that terrestrial chromophoric DOM is more prone to photolysis (Cory and Kling, 2018), although terrestrial DOM has also been reported to be available for biodegradation in the pelagic (Fasching *et al.*, 2014) and sediments (Freixa *et al.*, 2016) of fluvial systems. With increasing light

availability and residence time in lakes, in-system primary production generates fresh DOM that is commonly assumed to be more labile and readily available for microorganisms (Cole *et al.*, 1982; Fogg, 1983; Baines and Pace, 1991). Furthermore, partial photodegradation of coloured DOM may enhance the mineralisation of terrestrial DOM by microorganisms (Cory and Kling, 2018) and further evidence suggests a higher biodegradability of fresh terrestrial DOM than previously believed (Fasching *et al.*, 2014).

Although microbial-DOM interactions have been intensively studied, linkages between the quantity and quality of DOM and composition of bacterial communities have been difficult to establish. Some empirical evidence has suggested a connection between the quantity and quality of DOM with bacterial taxonomic community composition (e.g., Crump *et al.* (2003); Logue and Lindström (2008); Muscarella *et al.* (2019)), additionally an association with the functional composition was found in boreal systems (Ruiz-González *et al.*, 2015b). On the search for causal interactions, various incubation experiments found different bacterial groups to exhibit preferences in quantitative and/or qualitative DOM properties (Cottrell and Kirchman, 2000; Alonso and Pernthaler, 2006; Attermeyer *et al.*, 2014; Amaral *et al.*, 2016). Utilisation of low molecular weight compounds (LMWC) seemed to be a rather common functional trait among diverse microbes, whereas various high molecular weight compounds (HMWC) were degraded only by certain groups of microbes (Logue *et al.*, 2016). Similarly, it was also observed that more labile DOM could be removed by a single strain of bacteria, in contrary to degradation of less available compounds, which was only observed when various microbes were present (Pedler *et al.*, 2014). Interestingly, degradation of complex molecules does not seem to be phylogenetically conserved (Logue *et al.*, 2016), but was distributed across the microbial phylogenetic tree, hinting towards a potential of micro-diversity (Zimmerman *et al.*, 2013). Thus, DOM degradability is not solely a property of DOM itself (Nelson and Wear, 2014), but also a function of what microbes are present, how they are arranged and what their activity is (Osterholz *et al.*, 2016).

Advances in the resolution of technical approaches in both molecular and DOM research were argued to finally enable us to entangle this intertwined relationship (Herlemann *et al.*, 2014; Logue *et al.*, 2016). Yet, similar to the increasing resolution of molecular technologies, enhanced reso-

lution of DOM mixtures comes with computational and analytical challenges. DOM technologies providing information on a molecular level may not yield apparent correlation patterns because the DOM pool as a whole may comprise both reactive and non-reactive moieties that potentially form a universal background (Herlemann *et al.*, 2014; Kellerman *et al.*, 2015; Zark and Dittmar, 2018; Kellerman *et al.*, 2018). Using different levels of resolution between DOM and microbial communities can provide valuable insights (Amaral *et al.*, 2016), however, it has been shown that broad scale patterns between different DOM resolutions (e.g., fluorescence and absorbance analysis vs. FT-ICR MS) remain similar (Kellerman *et al.*, 2015). The true value in high-resolution data should not lie in whole mixture approaches that solely visualise differences among samples (e.g., Van Krevelen diagrams) but in finding dynamic partitions and subsets of molecules within the bulk DOM to link it with shifts in the microbial community composition. As a first step, aggregating high-resolution DOM data using diversity indices (Mentges *et al.*, 2017) and/or functional clustering (Ide *et al.*, 2017) rather than using individual molecular data should provide useful metrics to study DOM molecule assembly. Investigation of microbial and DOM co-variation should provide valuable insights into the complex relationship of microorganisms and their resource.

0.1.5 Emergent properties: Microbial community metabolism

Heterotrophic microbial community metabolism can be examined using various metrics that each provide a different dimension on how communities respond to differences in environmental factors (Hall and Cotner, 2007; Berggren *et al.*, 2007), resource conditions (Hall and Cotner, 2007; Berggren and del Giorgio, 2015), nutrient stoichiometry (Smith and Prairie, 2004; Berggren *et al.*, 2023) and biological constrains (Pradeep Ram *et al.*, 2016). In essence, the balance between bacterial respiration (BR) and bacterial production (BP) determines how much of the available organic carbon is channelled through the aquatic food-web or respired as CO_2 into the environment.

Linking microbial community composition to the processes and functions these communities mediate has been a constant challenge (Hall *et al.*, 2018). Most studies attempting to link microbial community composition to their function involve building mathematical models in experimental systems (Widder *et al.*, 2016). However, the applicability of results from short-term experiments to natural conditions is often limited (i.e., bottle effects), especially when single strains are

used instead of natural communities (Staley and Konopka, 1985). Furthermore, most microbial metabolism metrics as well as functional processes are often not a sum of individual microbial populations but rather an emergent property of complex communities (Hall *et al.*, 2018).

Metabolism such as bacterial production as been linked to quantity and quality of dissolved organic matter in aquatic systems (Judd *et al.*, 2006; Berggren *et al.*, 2009), however, little is known how metabolism changes across the entire aquatic network. Most metabolism studies focus on single ecosystem types and do not examine the entire range of hydrological scenarios present in our landscapes. Additionally, anthropogenic alterations such as damming that modify residence times along a river continuum may also affect community composition and consequently also microbial carbon processing (Ruiz-González *et al.*, 2013; Proia *et al.*, 2016; Maavara *et al.*, 2017). Hence, differences in community processes between two sites or times may still be related to rates of community restructuring between the investigated two points in space and time (Comte and del Giorgio, 2009, 2010). The lack of successful establishment between microbial community composition and community-level metabolism (Langenheder *et al.*, 2005; Fonte *et al.*, 2013), may be due to limited environmental gradients sampled in previous studies that did not allow the microbial community composition to change enough to find links to shifts in metabolism.

A recent analysis showed that the explanatory power of empirical models especially of carbon processing within soil systems was significantly improved when environmental and microbial (biomass and composition) data sets were combined (Graham *et al.*, 2016). Given that this evidence is based on a whole community approach, it is expected that the explanatory power will further improve if we can find and incorporate a clearer correlation between community composition, dissolved organic matter and community processing.

0.2 Thesis objectives

The overarching objective of this thesis is to advance our understanding of microbial and DOM assembly across a watershed scale. This thesis aims to provide both computational and conceptual insights into the intricate relationship between these two complex assemblages. By leveraging natural environmental gradients that emerge along aquatic continua - encompassing habitats such

as streams, rivers, lakes, and reservoirs - I seek to quantify and explore shifts in microbial assembly and links to DOM assembly. Moreover, this work endeavours to link various microbial processes to a comprehensive set of environmental, hydrological, and biogeochemical variables, enriching our understanding of how microbial community composition and availability of resources influence microbial metabolism.

The general aim is addressed through three main objectives, each explored in a dedicated chapter (Fig. 0.1):

1. Examine shifts in bacterial community structure along a terrestrial-aquatic continuum, assess how the balance between mass effects and species selection changes, and determine which taxa contribute to the reactive fraction within microbial assemblages.
2. Identify spatially reactive components within microbial and DOM assemblages and investigate their associations as well as analogies along a continuous hydrological gradient.
3. Investigate shifts in microbial metabolism along the aquatic continuum, focusing on identifying the environmental, hydrological, and biogeochemical drivers associated with these processes.

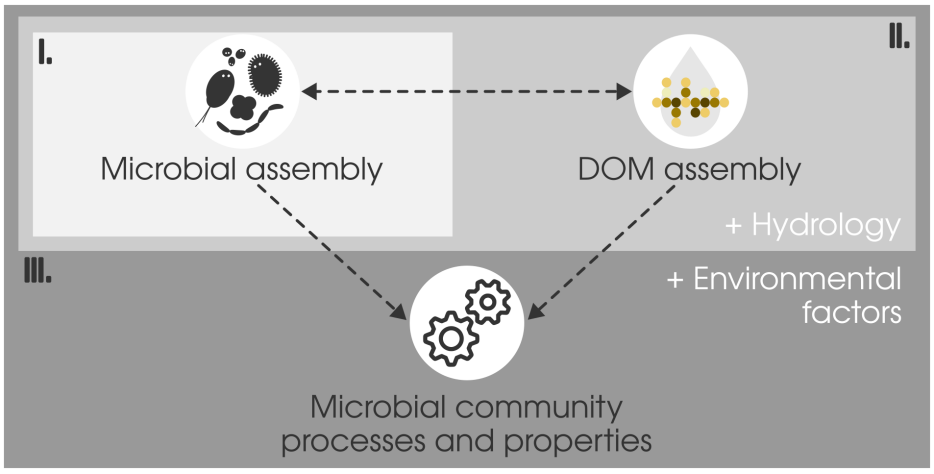

Figure 0.1: Conceptual diagram visualising the structure of this thesis.

0.3 General approach

To address the overall objectives of this thesis, a watershed-scale spatial and temporal survey was needed that consistently sampled both the microbial as well as DOM assemblage across multiple inter-connected aquatic habitats. The La Romaine project part of the CarBBAS industrial research chair had already kicked-off before the thesis' objectives were conceived. However, the unique sampling design of following the construction and damming of three reservoirs over three years starting in 2015, which was complemented with vast sampling of the entire watershed represented an optimal opportunity to address this thesis' objectives in a semi-pristine hydrological continuum. Although seasonal sampling was not consistent for all datasets (i.e., spring, summer, autumn), all seasons were sampled in the course of the project. Hence, the existing DNA, RNA and DOM samples between 2015-2017 where complemented with another full seasonal sampling design (spring, summer, autumn) and addition of laboratory incubations to measure microbial metabolism in 2018.

In chapter 1, microbial assembly along the entire continuum was examined using a dataset including the years 2015-2017 to cover the entire terrestrial-land-ocean continuum as denoted in the sampled habitats: soil, soilwaters, streams, rivers, reservoirs and the estuary. It provided a three year DNA dataset, and two year RNA dataset to address the question of how DNA and RNA samples diverge along the hydrological continuum. Bacterioplankton community composition was characterised using DNA and RNA via Illumina sequencing of the 16S rRNA gene, where RNA represents the potentially active microbial members. Detailed bioinformatic processing is outlined in the material and methods of Chapter 1, and the same approach was used for all subsequent chapters. The chapter introduces a novel approach to examine the balance between assembly processes, namely environmental selection and mass effects, by calculating the divergence of DNA and RNA samples within multivariate space. Inferences on assembly process shifts were based on the degree of divergence along the continuum.

In chapter 2, a smaller subset of samples was used covering years 2015-2016 spring and summer and focusing on all land and freshwater samples obtained. In addition to sequencing of DNA and RNA of microbial 16S rRNA, DOM was characterised using a high-resolution techinque (FT-ICR

MS). To model spatial patterns of individual DOM and microbial moieties in relation to hydrology, extensive hydrological estimation was conducted using empirically-derived discharge and velocity models. ArcGIS and custom programming in R was used to calculate the average time water spends to reach any point within the aquatic network by taking into account slower travel times through lakes (i.e., flow-weighted water age). Using high-resolution analyses for both assemblages we were able to model individual spatial patterns of microbial taxa and DOM molecular formulae and classify them by their general spatial behaviour. The spatial patterns were further categorised into reactive and unreactive components by using the model significance and assessing whether slopes obtained were above bootstrapped slopes. By correlating the individual reactive moieties of DOM and microbial communities, we were able to interpret their correlational patterns and attempt to discern mere correlational from causal relationships.

Finally, in chapter 3, measurements of microbial metabolism were added to the sampling campaigns in 2018. This chapter focuses on changes in microbial metabolism solely within aquatic habitats were streams and rivers ranging Strahler orders 1-7, lakes, reservoirs and reservoir hypolimnions were compared in how the changes in microbial processes relate to shifts in microbial community composition (DNA, RNA), DOM composition (FT-ICR MS), environmental variables and hydrological estimates. A range of environmental variables were measured *in situ* including temperature, dissolved oxygen, pH, conductivity; and others were measured in the laboratory: alkalinity, cholorphyll *a* (Chl *a*), DOC, dissolved inorganic carbon (DIC), total nitrogen (TN), ammonium concentration (NH_4), and total phosphorous (TP). In addition, we extracted and calculated the same hydrological variables as in chapter 2, namely catchment area, Strahler order, discharge, velocity, flow-weighted water age, and water residence time. As an *in-situ* metric of hydrology the degree of evaporation was estimated using d-excess, which can be calculated using isotope measurements of oxygen and deuterium. Multivariate analyses were conducted on each dataset (i.e., DNA, RNA, DOM, environment, hydrology) and the resulting axes were correlated to each microbial process measured to evaluate whether shifts in each microbial process are related to any of the biological, chemical and environmental matrices.

CHAPTER 1

TERRESTRIAL CONNECTIVITY, UPSTREAM AQUATIC HISTORY AND SEASONALITY SHAPE BACTERIAL COMMUNITY ASSEMBLY WITHIN A LARGE BOREAL AQUATIC NETWORK

Masumi Stadler[1] and Paul A. del Giorgio[1]

[1] Groupe de Recherche Interuniversitaire en Limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC, Canada.

Published in *The ISME Journal* (2022) DOI: 10.1038/s41396-021-01146-y

Keywords:

Microbial community assembly, land-freshwater-estuary continuum, upstream history, mass effects, species selection, DNA-RNA divergence, 16S rRNA gene sequencing, boreal watershed

N.B. References cited in this chapter are presented at the end of the thesis.

1.1 Abstract

During transit from soils to the ocean, microbial communities are modified and re-assembled, generating complex patterns of ecological succession. The potential effect of upstream assembly on downstream microbial community composition is seldom considered within aquatic networks. Here, we reconstructed the microbial succession along a land-freshwater-estuary continuum within La Romaine river watershed in Northeastern Canada. We captured hydrological seasonality and differentiated the total and reactive community by sequencing both 16S rRNA genes and transcripts. By examining how DNA- and RNA-based assemblages diverge and converge along the continuum, we inferred temporal shifts in the relative importance of assembly processes, with mass effects dominant in spring, and species selection becoming stronger in summer. The location of strongest selection within the network differed between seasons, suggesting that selection hotspots shift depending on hydrological conditions. The unreactive fraction (no/minor RNA contribution) was composed of taxa with diverse potential origins along the whole aquatic network, while the majority of the reactive pool (major RNA contribution) could be traced to soil/soilwater-derived taxa, which were distributed along the entire rank-abundance curve. Overall, our findings highlight the importance of considering upstream history, hydrological seasonality and the reactive microbial fraction to fully understand microbial community assembly on a network scale.

1.2 Introduction

Microbial communities across ecosystems are characterised by rank abundance distributions that vary in shape, yet we still know relatively little about how these structures come to be. Distribution shapes are thought to provide insight on community assembly (McGill *et al.*, 2007), with dominant and rare taxa assumed to be locally successful and transient, respectively (Magurran and Henderson, 2003; Nakadai *et al.*, 2020), but these interpretations have seldom been explicitly confirmed. Inherited from macroecology, microbial assembly processes have been defined into four fundamental categories - selection, dispersal, diversification, and drift - (Vellend, 2010), which vary in their degree of determinism and stochasticity (Zhou and Ning, 2017). Whereas diversification and drift usually manifest on longer, evolutionary, time scales, selection as well as dispersal are more relevant on ecological time scales. Regardless, there is always a historical aspect to commu-

nity assembly as local microbial communities reflect the balance between selection and dispersal processes that have occurred locally and in connected habitats in the past (Fukami, 2004). Hence, accounting for community history is vital to understand community assembly and the shape of the rank abundance distribution. Studies that investigated the relevance of history mainly followed microbes across time within an ecosystem (Shade and Gilbert, 2015; Comte *et al.*, 2017). Temporal history does shape local communities (i.e., legacy effects; Fukami (2015)), however, within an aquatic network, the uni-directional flow of water links temporal and spatial histories. Hydrology is a major driver of aquatic microbial community composition (Niño-García *et al.*, 2016a; Ruiz-González *et al.*, 2017b), as evidenced by soil microbes being flushed into and representing large proportions of aquatic communities (Ruiz-González *et al.*, 2015a; Crump *et al.*, 2012; Besemer *et al.*, 2013; Wisnoski *et al.*, 2020). As such, community structure at any given site within a hydrological network is the net result of upstream assembly processes (Nelson *et al.*, 2009), and network connectivity is further modulated by seasonal hydrological fluctuations (de Melo *et al.*, 2019; Caillon *et al.*, 2021). Therefore, spatial history is particularly relevant in highly interconnected freshwater networks (Vass and Langenheder, 2017), and there have been various studies that investigated the spatial context of aquatic microbial community assembly. Stegen *et al.* (2013) quantified major assembly processes based on spatial patterns of phylogenetic as well as taxonomic dispersion, which assumes that phylogenetically related organisms have similar niche requirements. Others have used spatial numerical distributions to infer the relative importance of selection versus passive transport across separate watersheds (Niño-García *et al.*, 2016a). While the importance of mixing and interacting communities between different ecosystems is now amply recognised (i.e., community coalescence; Mansour *et al.* (2018)), few studies consider interfaces between multiple ecosystems or ecosystem domains (e.g., terrestrial-aquatic; Nemergut *et al.* (2011); Shade *et al.* (2013)). A spatially connected, true aquatic continuum has mostly been evaluated on local scales within lakes (Logue and Lindström, 2010; Adams *et al.*, 2014; Langenheder *et al.*, 2017), along a single river mainstem (Winter *et al.*, 2007; Savio *et al.*, 2015; Hauptmann *et al.*, 2016; Doherty *et al.*, 2017; Gweon *et al.*, 2020) or on interconnected upstream networks (Besemer *et al.*, 2013; Nelson *et al.*, 2009; Widder *et al.*, 2014; Read *et al.*, 2015; Hassell *et al.*, 2018; Wisnoski and Lennon, 2021) and rarely have surrounding terrestrial ecosystems been considered as potential sources (Ruiz-González *et al.*, 2015a; Crump *et al.*, 2012; Wisnoski *et al.*, 2020). Moreover, active

and passive assembly processes are difficult to resolve as cell death and dormancy blur interpretations based on DNA patterns alone (Cole, 1999; Jones and Lennon, 2010). Indicative of recent protein synthesis, RNA sequencing has helped to disentangle active from unreactive microbial members (Bowsher *et al.*, 2019), however, only few freshwater studies have included both (Wisnoski *et al.*, 2020; Logue and Lindström, 2010; Székely *et al.*, 2013; Aanderud *et al.*, 2016; Denef *et al.*, 2016; Muscarella *et al.*, 2019; Peter *et al.*, 2018). All of these studies have separately yielded useful insight on microbial community assembly in freshwater systems, and they collectively point to the challenges ahead. The processes shaping community assembly are dynamic; selection and mass effects will vary in relative importance along complex aquatic networks as a function of the degree of connectivity to surrounding ecosystems, and upstream history. In order to capture the shifting balance of assembly processes and link those to the underlying rank abundance structure, we first need to examine a true hydrologic continuum that includes source communities and exchanges between various aquatic as well as terrestrial habitats as potential sources. Secondly, seasonality needs to be accounted for as the degree of connectivity depends largely on various hydrological scenarios in these networks. And lastly, DNA has to be accompanied by some indication of reactivity as selection and passive dispersal cannot be fully distinguished otherwise. In this study, we attempted a more holistic approach to aquatic microbial community assembly by addressing the three aforementioned critical dimensions.

1.2.1 Conceptual framework

Our overall aim was to follow shifts in bacterial community structure along a terrestrial-aquatic continuum and assess how the relative importance of mass effects versus species selection changes as communities traverse through varying environmental conditions and degrees of connectivity to the surrounding catchment. We carried out this study within La Romaine river watershed in the Northeastern region of boreal Québec, Canada, over several years and seasons. Starting from upstream sources such as soils, soilwaters, and headwater streams, we continued to follow the extant river orders (Strahler order 0–7) up to the estuarine plume. In addition, three reservoirs have been consecutively flooded mid-river over the sampling period. The sampling design covers various interfaces (terrestrial-aquatic, stream-river, river-reservoir, freshwater-estuary), and

other ecosystems within the watershed (e.g., headwater ponds, tributaries, lakes) that provide a further meta-community context. We first assess how the 16S rRNA gene (DNA-based) assemblage structure shifts along the terrestrial-freshwater-estuary continuum; we determine the general patterns of the spatial succession and its relation to different hydrologic seasonality. Furthermore, we differentiate the reactive from the total bacterial assemblage by additionally examining the 16S ribosomal RNA patterns relative to DNA. In this regard, RNA is not being used as an indication of absolute activity (Blazewicz *et al.*, 2013), rather we interpret the patterns of convergence and divergence between assemblage structures based on DNA and RNA (hereafter, DNA-RNA-based assemblage structures) along the continuum to infer shifts in the relative importance of species selection versus mass effects. To quantify divergence between DNA-RNA-based assemblage structures, we computed the distance between each DNA and RNA pair within multivariate space based on either incidence (presence-absence) or abundance dissimilarities. In a null scenario where DNA and RNA follow the exact same patterns, DNA-RNA-based assemblage structures remain equidistant, which would indicate no influx of unreactive bacteria (i.e., only detectable in DNA), and no changes in the reactivity of taxa within the community (no inactivation and activation of active and dormant taxa, respectively).

Local divergence in DNA-RNA-based assemblage structures in the continuum, on the other hand, may result from an influx of bacteria unreactive to local conditions (low RNA detectability), which would strongly influence the incidence-based distance, or from local shifts in the reactivity of specific taxa within the community (local activation/inactivation of taxa), influencing mostly the abundance-based distance. Shifts in how the incidence- and abundance-based metrics relate to each other across space and time, enabled us to gain insight into when and where selection or mass effects outweigh the other along the continuum. Finally, we explore where taxa potentially originated along the continuum and what fraction within the rank abundance curve the unreactive and reactive taxa commonly occupy.

1.3 Material and methods

1.3.1 Catchment characteristics and sampling

To follow the movement of microbial communities within a watershed, samples were taken along La Romaine river (Strahler order 7, Côte-Nord region, Québec, Canada) (Fig. S1.1a-b) from 2015-2017. La Romaine catchment belongs to the eastern black spruce-moss bioclimatic domain and has an area of 14,500 km^2. For detailed catchment characteristics refer to the supplementary methods (hereafter, SM) (SM 1.8.1, Fig. S1.1). In brief, the mainstem of the river (main trunk of riverine network) flows through a series of large, shallow lakes (hereafter, riverine lakes), emerging as Strahler order 6, and is subsequently dammed in a series of three hydroelectric reservoirs that were consecutively built in 2015 (RO2), 2016 (RO1), and 2017 (RO3). We refer to the river sections before and after the reservoir complex as upriver and downriver, respectively. The river has a total distance from the northern headwaters to the river mouth of 475 km.

In order to follow a terrestrial-aquatic continuum, various habitat types were sampled (Table S1.1). To capture a headwater network with soils, soilwaters, streams and ponds, we sampled the Petite Romaine sub-catchment (PR, *A*: 310.73 km^2, elevation: 580 masl, Strahler orders 0–4, Fig. 1.1c) due to the remoteness and inaccessibility of the northernmost headwaters. By sampling this headwater sub-catchment, we were able to follow a true continuum from an example headwater stream to the mainstem river into the estuarine plume. In addition, we sampled the reservoirs that are located along the mainstem. Other sites such as groundwaters, tributaries (Strahler orders 1–5), lakes and sediments in the catchment were sampled for a meta-community context. Tributaries refer to streams and rivers that were sampled at the confluence of the mainstem, in contrast to streams within the PR sub-catchment that represent a headwater network. Overall, 389 samples were collected for DNA (D) and 201 for RNA (R), covering spring (156-D, 66-R), summer (199- D, 101-R) and autumn (34-D,34-R) (Table S1.1). RNA samples were collected from 2016 onwards.

For detailed sampling procedures and sample preparation for each sample type refer to SM 1.8.2. In brief, all water samples were filtered onto a 0.22 μm polycarbonate filter with a peristaltic pump and homogenised soil and sediments were stored in aliquots. RNA samples were submerged in RNAlater and Lifeguard Soil Preservation Solution (QIAGEN, Hilden, Germany) for water and non-

water samples, respectively. RNA samples were stored at 4°C overnight to allow stabilisation and were subsequently frozen. All DNA and RNA samples were frozen at -20°C at the field station and further stored at -80°C at the university laboratory until extraction. DNeasy and RNeasy PowerWater and PowerSoil kits (QIAGEN, Hilden, Germany) were used following the manufacturer's instructions. RNA extracts were reversely transcribed to cDNA with a high capacity cDNA Reverse Transcription Kit (Applied Biosystems™, Foster City, CA, USA) and all samples were sent to Génome Québec Innovation Center (Montréal, QC, Canada) for paired-end sequencing of the 16S rRNA V4 region using the primers 515 F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806 R (5'GGACTACHVGGGTWTCTAAT-3') on a MiSeq platform (PE250, Illumina, San Diego, CA, USA; details in SM 1.8.2).

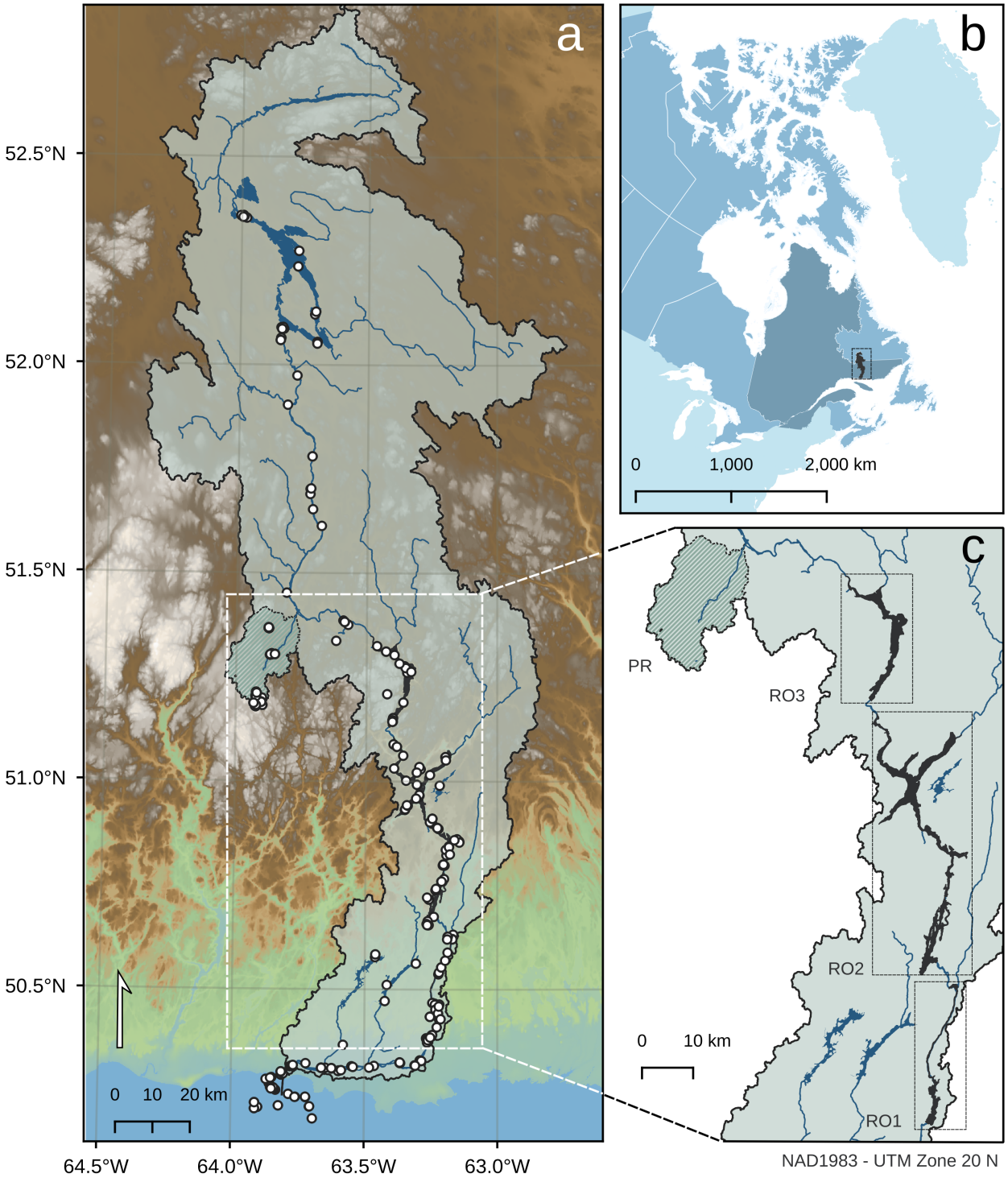

Figure 1.1: Location and overview of the La Romaine catchment. a) Scale and overview of the whole La Romaine catchment. Samples are represented as points. b) Location of the catchment within Canada and Québec. c) Focus on all built reservoirs RO1 (2015), RO2 (2014) and RO3 (2017) and the headwater stream sub-catchment Petite Romaine (PR).

1.3.2 Bioinformatic analysis

A detailed description of the bioinformatic treatment can be found in SM 1.8.3. In brief, primers were removed from 16S rRNA DNA and cDNA (hereafter, RNA) reads using *cutadapt* (v1.18; Martin (2013)). To identify amplicon sequence variants (ASVs), 16S rRNA amplicon reads were analysed through the DADA2 (Divisive Amplicon Denoising Algorithm 2) pipeline (v1.14.1; Callahan *et al.* (2017)). Taxonomy was assigned with the *DECIPHER* package (v2.14.0; Wright (2016)) implementing the IDTAXA algorithm (Murali *et al.*, 2018) and the GTDB database (Release 95; Parks *et al.* (2018)). To account for slight differences that may have emerged between DNA and RNA ASVs and potential differences among 16S rRNA copies within a single genome, ASVs were merged into OTUs by a 99% similarity threshold (Větrovský and Baldrian, 2013) with the *DECIPHER* package (Wright, 2016). OTUs only found in RNA ('phantom' taxa) were corrected for by replacing all observations with RNA>0 and DNA=0 with DNA=1 (Bowsher *et al.*, 2019). An observation (i.e., read count of an OTU within a sample) that only appeared in a single sample within each habitat type, season and nucleic acid type combination (i.e., singleton within a factorial combination) was considered unreliable if the singleton OTU had less than 10 reads within the sample. This approach not only removes singletons across the whole database but also singletons within each sampling campaign that had too few reads to be considered reliable. Furthermore, *metagenomeSeq* was used to transform and stabilise variation in library sizes with cumulative sum scaling (CSS; Paulson *et al.* (2013)) (hereafter: CSS reads). CSS results were compared with results achieved with various rarefaction thresholds and no substantial differences were observed (Fig. S1.2). A few minor differences are discussed in SM 1.8.4.

1.3.3 Data exploration and statistical analysis

A detailed version of this section is included in SM 1.8.5. To explore differences in microbial community composition across habitat types and seasons, a Principal Coordinates Analysis (PCoA) was conducted with Bray-Curtis dissimilarities (D_{BC}) (Bray and Curtis, 1957; Legendre and Legendre, 1998) based on all DNA samples with the function *pcoa* in the *ape* package (Paradis and Schliep, 2018)(n = 389, 16,322 OTUs). To evaluate statistical differences in habitat type and season a PERMANOVA was computed with 9,999 permutations with the *adonis* function. Multivariate homo-

geneity was tested for with *betadisper* and *permutest* (*vegan* package; Oksanen *et al.* (2019)). For all statistical analyses, an α level of 0.05 was chosen prior to analysis. To further evaluate whether sampled RNA-based assemblages were different from the DNA-based assemblages, we performed a second PCoA with both DNA and RNA samples (with D_{BC}, n = 590, 16,322 OTUs). Again, statistically different groups were investigated with a PERMANOVA (9,999 permutations), where habitat type, season and nucleic acid type (DNA versus RNA) formed the groups. To quantify how different DNA-RNA-based assemblages of the same sample are, the Bray-Curtis distance (m_{BC}) of each DNA-RNA sample pair within the PCoA ordination space was computed across *n*-dimensional space (Tabak, 2004); a similar approach to other studies that extracted the magnitude of change in multivariate space between two samples of interest (Muscarella *et al.*, 2019; Brown, 2003):

$$m(p,q) = \sqrt{(\mid p_1 - q_1 \mid)^2 + (\mid p_2 - q_2 \mid)^2 + \cdots + (\mid p_n - q_n \mid)^2}$$

where *p* and *q* represent DNA and RNA site scores, respectively, of each sample and *n* is the used maximum number of dimensions. We focused on the first axes that cumulatively explain 75% of the variation for each ordination ($n_{75\%}$), similar to Osterholz *et al.* (2016). This approach was implemented as it was evident from the PCoA that essential variation within non-aquatic samples was captured outside the first three axes and to exclude noise that may be captured when using all dimensions (Fig. 1.3).

To gain further insight into the processes shaping assemblage dissimilarities, we computed a PCoA with the Sørensen dissimilarity (D_S), which is the incidence-based equivalent of D_{BC} (Legendre and Legendre, 1998; Sørensen, 1948) (Fig. S1.4). By comparing incidence- and relative abundance-based dissimilarities, we can further distinguish in which samples DNA-RNA-based assemblages diverge primarily due to different present taxa or their abundances, respectively. We further applied the same framework of calculating the distance among DNA and RNA pairs across $n_{75\%}$ axes resulting in the Sørensen-based distance (m_S) (Fig. S1.4). In order to examine where shifts in the relative importance between incidence- and relative abundance-based distances were happening along the continuum, we calculated the difference between m_{BC} and m_S (Δ-distance). To explore

the interpretability of the Δ-distance approach, we simulated theoretical communities and computed Δ-distances for this mock dataset (details in SM 1.8.6). In brief, four species abundance distributions (SADs) with various levels of evenness were created and randomly sampled to generate DNA assemblages. We hypothesised that these various SADs represent a gradient from mass effects to selection, where less even communities are linked to stronger selection. Subsequently, DNA assemblages were duplicated for each site to create a base for the corresponding RNA assemblage. We hypothesised that the higher the number of OTUs in DNA without RNA, the stronger the mass effect. To create an additional range of mass effects, different numbers of OTUs were removed from the RNA assemblage. Results indicated that lower Δ-distance values correspond to stronger mass effects as indicated by higher replacement and higher evenness. Inversely, higher Δ-distance values indicate stronger selection with lower replacement and lower evenness (Fig. S1.5, Fig. S1.6). Results obtained during our rarefaction test (SM 1.8.4) showed that absolute numbers of Δ-distances varied across rarefaction thresholds, while relative patterns across seasons and habitat types remained consistent. Hence, absolute values in Δ-distances are likely to hold little meaning, and it is rather the relative change in Δ-distances and the resulting pattern across gradients or between habitats that is informative and comparable across studies.

1.3.4 Abundance groups

In order to explore where within a rank abundance curve community reshuffling is occurring, abundance groups (e.g., abundant, moderate, rare) were defined based on the shape of rank abundance curves per habitat type. Abundance thresholds are defined as the first and second moment of maximum acceleration along the rank abundance curve (Fig. S1.7). This approach classified all OTUs with >= 72 CSS reads as abundant, <72 and >= 10 CSS reads as moderate, and <10 CSS reads as rare (details in SM 1.8.7). This classification method was implemented as abundance thresholds commonly used are rather inconsistent across the literature, with little confidence in whether a particular abundance threshold is suitable for a given dataset. The implemented approach is not different from common fixed relative abundance thresholds (e.g., 1%), the only difference lies in the fact that the abundance threshold is derived empirically from the species abundance distributions of the studied dataset.

1.3.5 Classification of OTU origin and potential reactivity

OTUs were classified by the habitat in which they were first detected along the terrestrial-aquatic continuum (regardless of season) to have a proxy of origin for each OTU (hereafter, potential origin) (Ruiz-González *et al.*, 2015a; Crump *et al.*, 2012). The classification followed the order of soil, soilwater, stream, upriver, reservoirs, downriver, and finally, the estuary. To further explore patterns in the OTUs' DNA and RNA relationship, we correlated the contribution of individual OTUs to each local DNA and RNA pool (e.g., a local pool was defined for each season and habitat). The OTU contributions to each local pool were first averaged for each potential origin, abundance group and four RNA contribution categories (<25% confidence interval (CI), < median, > median and >75% CI) to enhance visibility. On first attempt, all OTUs falling in the <25% CI and < median categories showed striking invariability, while >75% CI and > median categories followed a linear relationship between an OTU's contribution to the DNA and RNA pools. This pattern was observed across most local pools, and hence, the contribution categories were reduced to two groups (>median and <median). This categorisation threshold is referred to as the 'potential reactivity threshold' (hereafter PRT, median: 0.067%), based on the apparent decoupling of DNA and RNA of OTUs below the median. We infer that the absence of a relationship between DNA and RNA likely reflects that these taxa below the PRT are present but may be generally unreactive to the environment, given that their numerical abundance is largely unrelated to their apparent potential activity. Hence, we categorised the taxa below the PRT together with taxa that did not have any RNA as 'unreactive'. It is important to note here that the PRT was applied for each habitat and season, hence, an OTU classified as reactive within for example soils, may become unreactive or stay reactive in subsequent habitats or seasons. All analyses have been conducted in R v3.4.2 (R Core Team, 2024) and RStudio v1.3.1073 (RStudio Team, 2024)(package details in SM 1.8.8).

1.4 Results

Sampled sites covered a large range of habitat types from soils, soilwater, over streams, the main river, lakes, reservoirs and the estuary. We recovered 51,901,843 quality filtered reads, with 119,109 identified ASVs. After 99% similarity OTU clustering, there were 35,995 unclassified OTUs that were removed in the downstream analyses. After sub-sampling only bacteria, 48,927,604 reads

and 16,322 OTUs were retained. The smallest and largest library size were found in a sample of riverine lakes and soilwater with 1,470 and 81,716 reads, respectively. On average, the lowest library sizes were found in sediment, soil, soilwater and estuary with less than 20,000 reads. In contrast, most freshwater samples had a library size larger than 20,000 reads (Fig. S1.8). There were 56 phyla, 130 classes, 316 orders, 571 families, and 1,027 genera represented in the dataset. Relative abundances of phyla varied across habitat types (Fig. S1.8) but on average, the metacommunity across all ecosystems was composed of Proteobacteria (38.4%), Verrucomicrobiota (9.6%), Patescibacteria (7.9%), Acidobacteriota (5.3%), Myxococcota (5.2%), Actinobacteriota (5.2%), Bac- teroidota (4.5%), Bdellovibrionota (3.8%), Planctomycetota (3.2%) and Cyanobacteria (2.7%).

1.4.1 Gradual change of assemblage structures along the terrestrial-aquatic continuum

We observed a clear directional pattern in community composition based on DNA, from the most terrestrially-influenced habitats such as soil, soilwater and sediment to the mainstem river and reservoir sites and the estuary, which was captured in the first PCoA axis (Fig. 1.2) and statistically supported by a PERMANOVA analysis (Table 1.1). Groundwaters, streams, tributaries, headwater ponds, and lakes were clearly aligned between the two endpoint clusters formed by terrestrial and riverine/reservoir samples (Fig. 1.2).

Table 1.1: PERMANOVA and PERMDISP results based on DNA alone and DNA and RNA community matrices.

		PERMANOVA				PERMDISP		
Dataset	Group	*df*	*F*-statistic	R^2	*p*-value	*df*	*F*-statistic	*p*-value
DNA	Habitat	12	17.5	0.35	< 0.0001	12	38.87	< 0.0001
	Season	2	12.05	0.04	< 0.0001	2	58.36	< 0.0001
	Combined					27	20.72	< 0.0001
DNA and RNA	Habitat	12	19.98	0.28	< 0.0001	12	28.88	< 0.0001
	Season	2	15.64	0.04	< 0.0001	2	62.03	< 0.0001
	Nucleic Acid Type	1	25.64	0.03	< 0.0001	1	1.20	0.27
	Combined					49	9.60	< 0.0001

The directional trajectory along the terrestrial-aquatic continuum had a striking seasonality that was especially marked for the downstream river and reservoir sites, with a clear separation between spring and summer/autumn samples along the second PCoA axis (Table 1.1). Soil, sediment, soilwater and groundwater sites, however, did not exhibit a clear seasonality (Fig. 1.2) and seasonality did not emerge as a strong driver even within a PCoA performed only with terrestrially-influenced samples (Fig. S1.9). Although PERMANOVA results strongly supported habitat type and seasonal clustering, the results could be affected by different dispersion of data within multivariate space, which interferes with a straight forward interpretation of the results as a PERMANOVA cannot distinguish among-group from within-group variation if data dispersion is variable (Anderson and Walsh, 2013). Differences in dispersion were found by habitat type alone, solely season and both habitat and season combined (Table 1.1). While dispersion between spring and summer was not statistically different (Tukey's HSD: $p > 0.05$), dispersion was always different when comparing autumn with other seasons (Tukey's HSD: $p < 0.0001$). The average distance of samples within the autumn cluster to its median was smaller compared to other seasons (0.46 vs 0.63/0.62) likely due to a smaller sample size in autumn ($n_{\text{autumn}} = 34$ versus $n_{\text{spring}} = 156$, $n_{\text{summer}} = 199$). Among habitat types, dispersion was larger in terrestrially-influenced sites such as soil (Distance to median: 0.61), soilwater (0.63), stream (0.63) and tributary (0.62) samples, compared to riverine (0.51), reservoir (0.52) and estuary (0.52) samples. The observed differences in the heterogeneity within habitat types likely reflects inherent characteristics of these ecosystems, with sites of stronger terrestrial influence exhibiting stronger spatial variance.

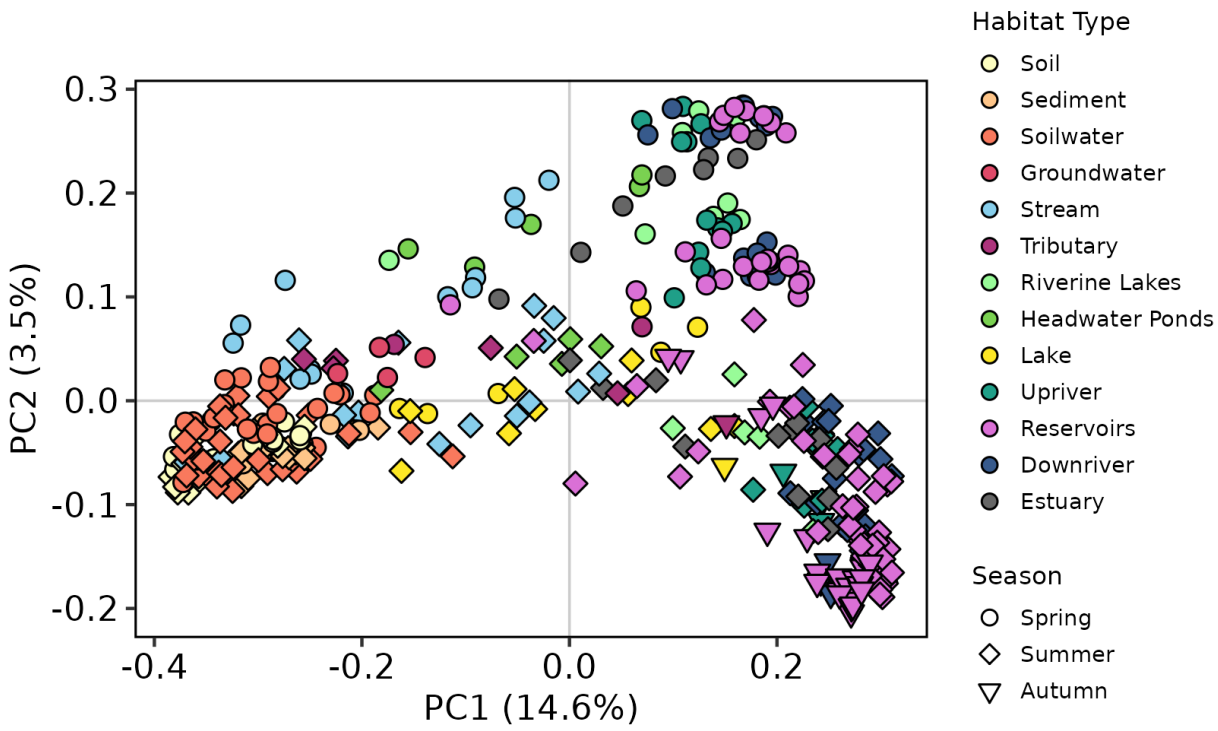

Figure 1.2: Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons. Overall PCoA analysis of DNA samples. The PCoA reveals microbial community shifts from terrestrial to freshwater samples. Habitat types ranged from soils, soilwaters, streams and headwater ponds (sampled in headwater stream network, Stahler orders 0-4), tributaries (sampled at the confluence to the mainstem, Strahler orders 1–5), riverine lakes (northernmost lakes through which the Romaine river flows), reservoirs, sections along the mainstem (Strahler order 6–7) upstream and downstream of the reservoir complex (upriver and downriver, respectively), and the estuary. Spring and summer/autumn show distinct paths in multivariate space. Percentage of variance explained are given in brackets for the first and second axes.

1.4.2 Patterns in RNA and DNA divergence

When DNA and RNA samples were combined in a second PCoA analysis, the three main axes of variation were habitat type (PC1), nucleic acid type (PC2) and seasons (PC3), and these three first

axes captured in summary 18.9% of the dissimilarity variance (Fig. 1.3). Strong DNA-RNA divergence emerged as soon as the continuum enters the upstream aquatic sites (i.e., streams) and amplified along the continuum (Fig. 1.3a). Seasonality was most pronounced in aquatic sites in both DNA and RNA (Fig. 1.3b) and was the second-strongest driver after the spatial continuum in a PCoA only with RNA samples (Fig. S1.10). There was no clear seasonality and differentiation between nucleic acid types in terrestrially-influenced samples (Figs. 1.3b, S1.10). However, it is noteworthy that visual inspection indicated that much of the terrestrially-influenced site dissimilarity was split upon additional axes (data not shown). Overall, PERMANOVA analysis indicated significant clustering by habitat type, season and nucleic acid type (Table 1.1). Similar to the DNA only PCoA, homogeneity of dispersion was mostly not fulfilled (Anderson and Walsh, 2013). According to PERMDISP, dispersion differed by all factorial combinations, habitat type and season, however, not for nucleic acid type alone (Table 1.1). Dispersion patterns among seasons and habitats were similar to the DNA only PERMDISP results, where dispersion was smaller in autumn compared to spring and summer and terrestrially influenced sites generally had a larger dispersion.

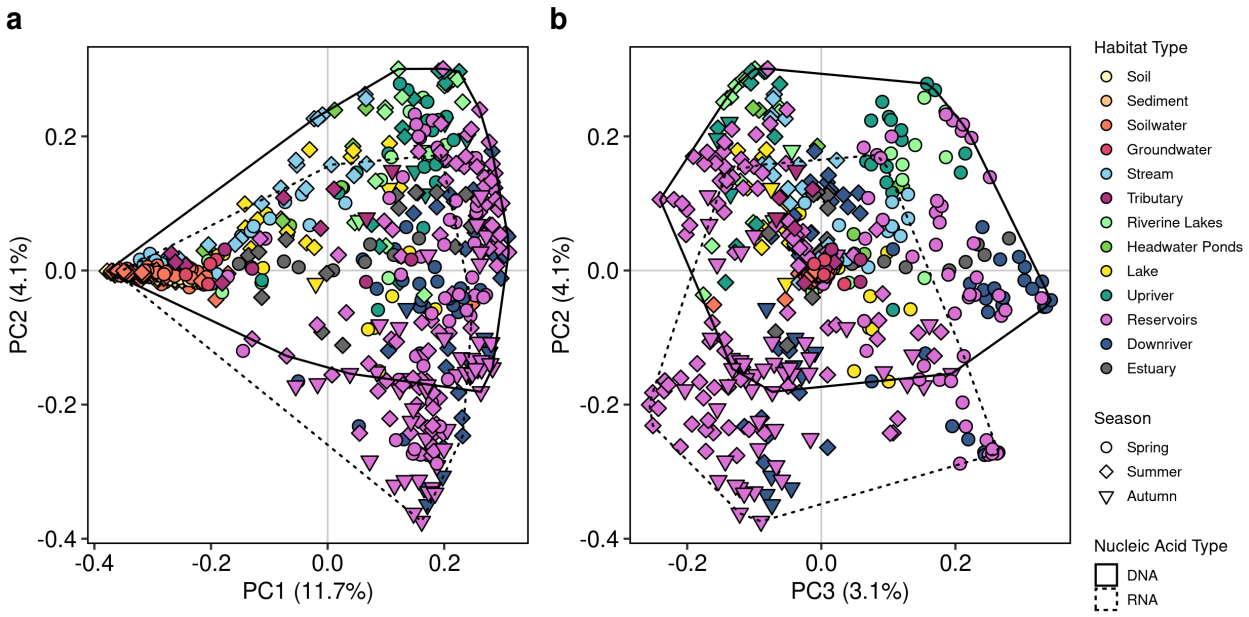

Figure 1.3: RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially-influenced habitats. PCoA analysis including RNA samples. a) Visualisation of first and second axes of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis using the second and third axes, differentiating nucleic acid type and seasons, respectively. The distribution of DNA and RNA samples are highlighted with polygons. Percentage of variance explained by the corresponding axes are given in brackets.

1.4.3 Inferring assembly dynamics from DNA-RNA pair-wise distance patterns along the continuum

To further explore the patterns in DNA-RNA-based assemblage structure differences along the continuum, we calculated the distance in PCoA ordination space between DNA and RNA of each sample based on abundance (m_{BC}) and incidence (m_S) along the number of axes capturing 75% of variance ($n_{75\%}$) as the depicted first three axes in Fig. 1.3 only captured a limited fraction of the total DNA-RNA divergence ($n_{75\%}$) axes: Sørensen = 204, Bray-Curtis = 192. Based on the incidence metric m_S, the largest average distances between DNA and RNA were found in spring (0.43 $\pm$ 0.29 (mean $\pm$ standard deviation)) especially in soilwaters (0.63 $\pm$ 0.33). In contrast, summer m_S distances were in general lower (0.18 $\pm$ 0.12) and more similar among habitat types (Fig. 1.4a). The abundance-based distance (m_{BC}) was in general very similar among seasons with spring be-

ing slightly higher (0.61 $\pm$ 0.20) than summer (0.53 $\pm$ 0.15) and autumn (0.50 $\pm$ 0.06). Lowest m_{BC} distances were found in soil (0.34) and sediment (0.30) samples. Overall, m_{BC} values were always larger than m_S, as m_{BC} captures both abundance and incidence differences and hence, add abundance-based differences to the distance observed with m_S.

To investigate relative changes between incidence and abundance-based distances, we computed the difference between m_{BC} and m_S (Δ-distance)(Fig. 1.4b). The Δ-distance can theoretically range between 1 and -1, both exemplifying extreme cases where m_{BC} = 1 and m_S = 0 and vice versa. In general, lower (or even negative) Δ-values indicate comparably higher incidence-based distances, where dissimilarity is largely driven by the occurrence of different taxa between DNA and RNA suggesting prevalence of mass effects (SM 1.8.6). On the other hand, higher Δ-values indicate relatively low incidence-based and high abundance-based distance, representing selection-driven dissimilarity with more taxa common between DNA and RNA but rather large numerical differences (SM 1.8.6). Hence, we interpret positive shifts in Δ-distances as transition from mass effects to selection dominated habitats and vice versa. Overall, there were different trajectories among seasons in Δ-distances, with lower values in spring remaining relatively stable around 0.17 $\pm$ 0.17, and higher average values in summer (0.35 $\pm$ 0.16), suggesting relatively higher overall mass effects in spring and selection in summer. In addition, there was a clear spatial pattern in Δ-distance in summer along the continuum, with an increase in Δ-distance from terrestrially-influenced sites (i.e., soils, soilwaters, streams) towards the river followed by a sharp decline downstream of the reservoir, suggesting increasing mass effects downstream of the reservoir. In autumn, selection gradually increases along the mainstem from the upstream river over reservoirs to the downstream river sites. Coinciding with these patterns, other sampled high residence time habitats within the watershed (i.e., riverine lakes, lakes) also had a higher average Δ-distance compared to tributaries (Fig. 1.4b), suggesting strong selection especially in summer. Whereas lakes remained relatively stable across seasons, riverine lakes had a distinct pattern comparable to tributaries with relatively low Δ-distances in spring indicative of high mass effects.

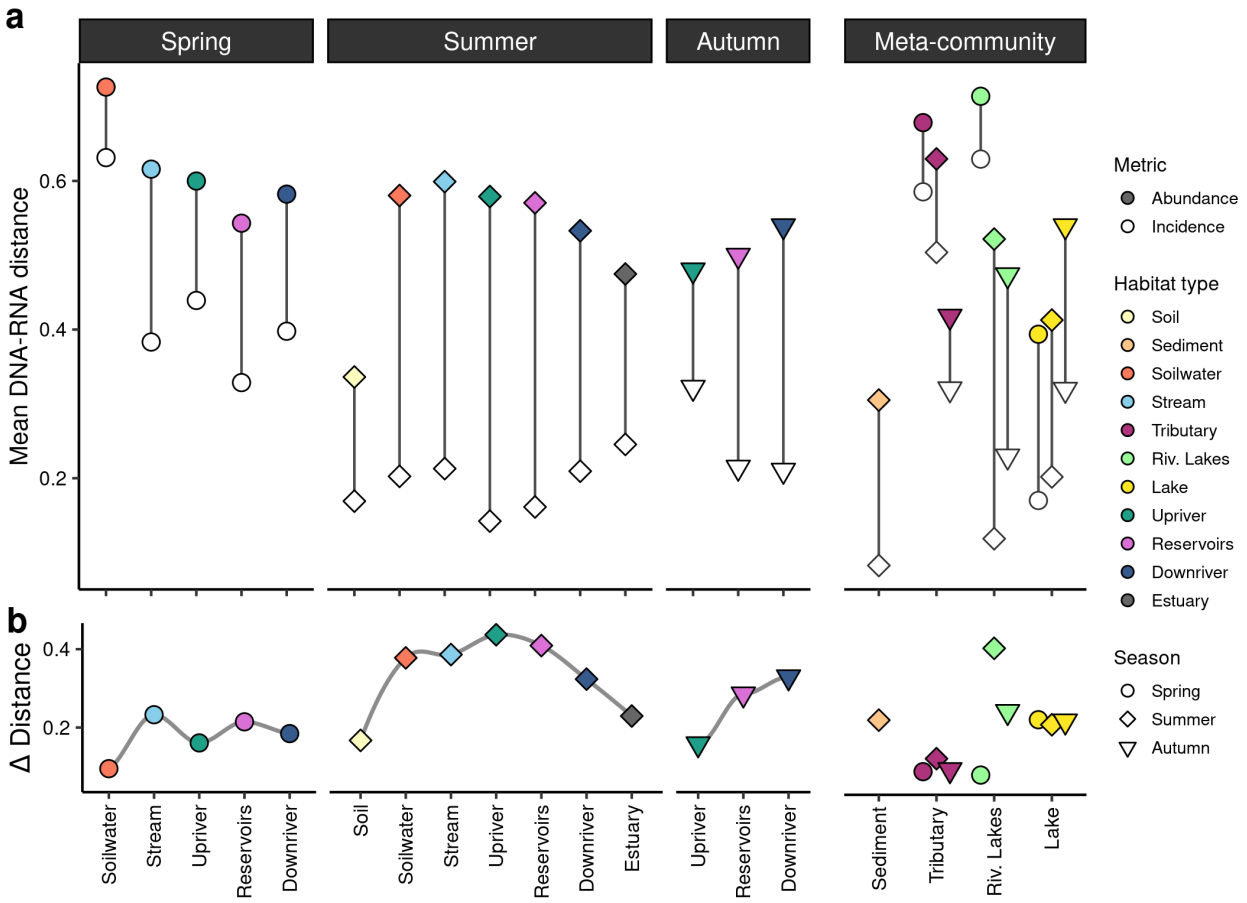

Figure 1.4: Patterns between abundance and incidence-based distance reveal seasonal shifts in assembly processes. a) Distances between DNA and RNA of the same sample within ordination space were averaged by habitat type and season. Abundance-based (coloured points, m_{BC}) and incidence-based distance (hollow points, m_S) indicate spatio-seasonal trends in strong mass effects, especially in spring with high incidence-based distances. b) Δ-distances between m_{BC} and m_S indicate shifts in the relative contribution of incidence- versus abundance-based distances along the continuum, with high Δ-distances indicating stronger selective forces. Habitats sampled outside the direct continuum are given as an additional meta-community context.

1.4.4 Taxa across the rank abundance curve contribute to the reactive pool

In the previous sections we have established that there were patterns of divergence and convergence of DNA-RNA-based assemblage structure that were spatially and temporally structured. To further explore what fractions within the community contribute to mass effects and selection, re-

spectively, we classified OTUs by the habitat in which they were first detected along the terrestrial-aquatic continuum, from soils to the estuary. In particular, by focusing on taxa that were only detected in DNA (termed unreactive) and thus strongly contribute to the observed mass effects, it became evident that a relatively large proportion potentially originated in the upstream habitats such as soils, soilwaters and streams. Yet new unreactive taxa were also gained along the entire continuum (Fig. 1.5a).

In order to understand the coherence between DNA and RNA patterns of reactive taxa that contribute to the observed patterns in selection, we examined the relative contribution of individual taxa to the DNA and RNA sequence pools for each habitat and season (Fig. 1.5b). For this, we also grouped OTUs according to where they were first detected in DNA along the continuum to identify subsets of taxa that activated and developed locally. To simplify the analysis, we further classified OTUs based on their mean DNA habitat abundance into locally abundant, moderate and rare taxa (see 'Methods'). Inspection of the various plots in Fig. 1.5b revealed a recurrent pattern in most habitats with a subset of taxa present across all DNA abundance groups, which nevertheless contributed negligibly to the local RNA pool (between 0.00005 and 0.0067% of sequences) and whose RNA contribution was decoupled from their contribution to the DNA pool. In contrast, there was another subset of taxa with a consistently higher contribution to %RNA sequences (above 0.0067% sequences, hereafter termed reactive) and whose contribution to RNA and DNA pools appeared to be linearly coupled. Although we classify these two groups as 'reactive' and 'unreactive' out of simplicity, we do not know why some taxa have a decoupled DNA and RNA relationship. We acknowledge that taxa within the 'unreactive' fraction may be taxa with disproportionately low rRNA albeit being still active. Reactive taxa were represented across DNA abundance groups, and in most habitats the overall relationship between %DNA and %RNA contribution averaged around a log-log slope of 1.18, suggesting a roughly proportional contribution (Fig. 1.5b), except for upriver spring samples where numerically rare taxa appeared to have disproportionately high RNA. On the other hand, taxa that were unreactive had a striking invariance, hovering around 0.0025% RNA contribution which may indicate a lower activity threshold. These taxa may be somewhat analogous to taxa that were entirely not detected in RNA and only in DNA. Taxa without a single detected RNA copy can have RNA in the environment, however, their RNA

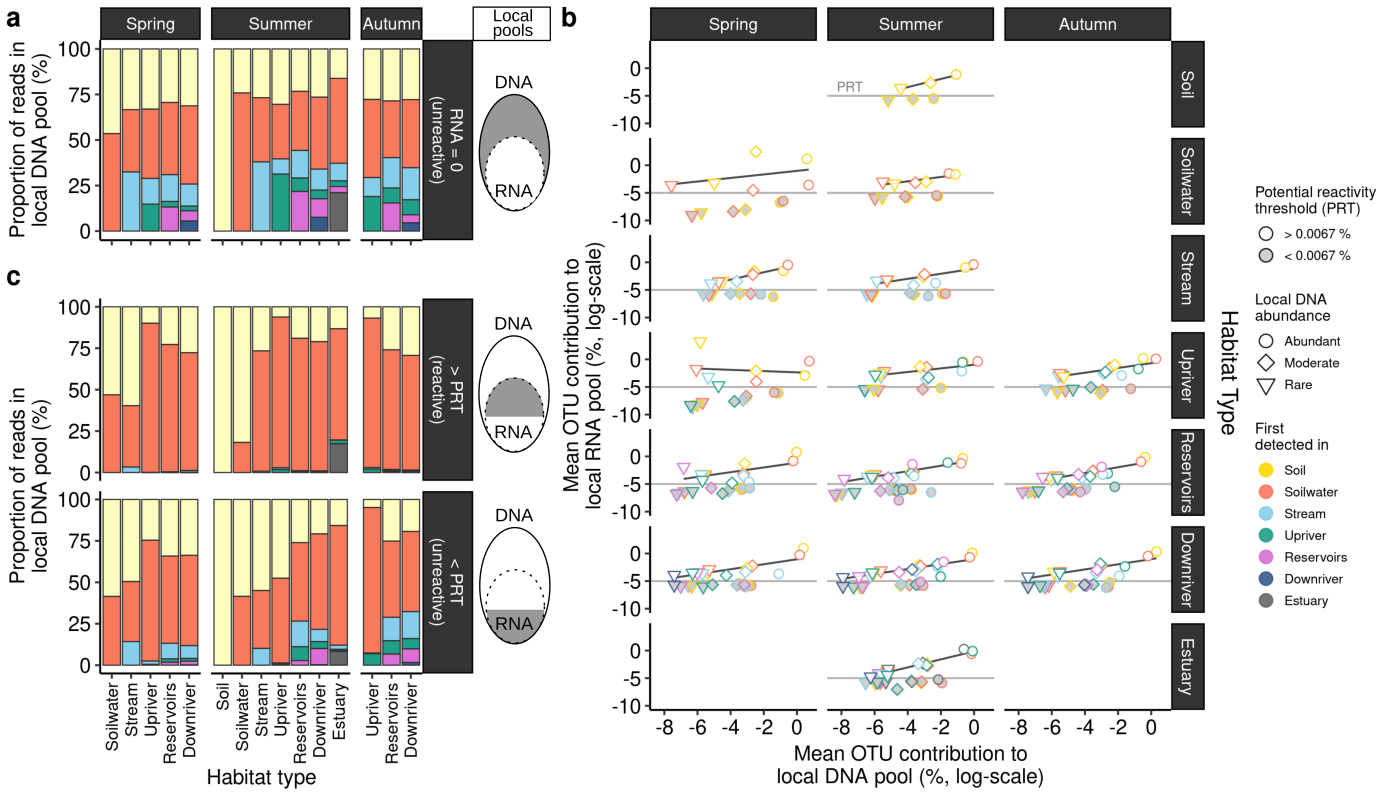

Figure 1.5: Taxa along the whole rank abundance curve contribute to the reactive pool. a) Contribution of taxa (operational taxonomic units (OTUs)) detected in DNA without any RNA observation (RNA = 0) to the local number of DNA reads for each habitat type and season. OTUs were binned by (1) their first detected habitat in DNA along the continuum, (2) local DNA abundance (e.g., abundant, moderate, rare, see methods). b) Each panel corresponds to an individual OTU's reads contribution to the local DNA and RNA pool, which was defined for each season and habitat type. Taxa were classified and averaged by (1) first detected habitat and (2) local DNA abundance and lastly, (3) by whether they are above or below the identified potential reactivity threshold (PRT) of 0.0067% RNA contribution to the total local RNA pool. Linear regression slopes were plotted for the averaged OTU contributions above the PRT. Due to the absence of a linear relationship of OTUs < PRT, OTUs > PRT were interpreted to be reactive, while OTUs < PRT were considered unreactive. c) Based on the observations of (b) and the consequent classification of reactive (upper panel) and reactive (lower panel) taxa, individual OTU's contribution to the total local DNA pool were visualised. Values are expressed as a fraction of the total number of DNA reads for each habitat type and season.

content may be too low to be captured with our sequencing depth. The reactive and unreactive taxa were both distributed across the entire range of %DNA contribution. The former indicates that there are locally reactive OTUs across the entire rank abundance curve, including rare taxa. Likewise, there were unreactive taxa across the entire rank abundance curve that may be numerically important in DNA. Furthermore, reactive taxa occupy on average 45.9% of the OTU pool in each habitat, whereas the unreactive taxa (<0.0067% contribution (22.7%) and RNA = 0 (31.4%)) together comprise 54.1% (Fig. S1.11).

Within the reactive fraction, OTUs that were first detected in soil had the largest overall contributions to the local RNA pool of reactive taxa across all habitats and seasons (overall mean of 0.97%) followed by soilwater-derived OTUs (0.22%), but OTUs that were first detected in a given local habitat were found within the reactive fraction of all DNA abundance groups (Fig. 1.5b). Soilwater and soil-derived taxa were less prevalent in the estuary, where riverine and estuarine-derived taxa become numerically more important in RNA. The overwhelming majority (>90%) of DNA sequences of reactive taxa along the entire continuum (except for estuarine sites) could be retraced to OTUs that were first detected in soils and soilwaters (Fig. 1.5c, top panel). In contrast, a relatively large proportion of DNA sequences of unreactive taxa were contributed by a more diverse pool of OTUs that originated in various habitats along the continuum (Fig. 1.5c, bottom panel), a pattern similar to that of OTUs undetected in RNA (Fig. 1.5a). The latter would suggest a high level of influx and persistence of OTUs along the continuum that were seemingly unreactive. Interestingly, there was not a single OTU that remained abundant or moderately abundant along the entire continuum, suggesting that the soil/soilwater-derived taxa, which consistently dominate the reactive fraction, nevertheless shift along the rank abundance curve along the continuum.

1.5 Discussion

In this study, we attempted to address three major challenges identified in understanding microbial community assembly within aquatic networks: First, to incorporate the upstream history of local communities, secondly, to capture a variety of hydrological scenarios and thirdly, to capture any indication of reactivity to changing environmental conditions. Here, we followed an interconnected, large scale continuum that extended from upstream soils into the estuary and sam-

pled across seasons to address shifts in assembly processes linked to hydrological fluctuations. In addition, we accompanied DNA with RNA sequencing to distinguish numerical responses that could be linked to passive transport of dormant or inactive bacteria from those associated to reactive taxa (Jones and Lennon, 2010; Wilhelm *et al.*, 2014). In this regard, we are not using RNA or RNA/DNA ratios as an index of absolute activity (Wilhelm *et al.*, 2014; Campbell *et al.*, 2011), but rather utilise spatial patterns in the degree of coupling between DNA-RNA-based assemblages to identify taxa that appear to react to local conditions, and distinguish where selection was most dominant along the network. We further classified OTUs by their habitat in which they were first observed to assess whether taxa from upstream habitats persist along the continuum (i.e., upstream history) and to further examine how unreactive and reactive taxa are distributed along the rank abundance curves.

1.5.1 Terrestrial influx and aquatic legacy shape reactive and unreactive fractions of bacterioplankton communities

The high connectivity and unidirectional flow within aquatic networks have often been neglected, and the effect of upstream selection history and dispersal among aquatic water bodies as well as from the surrounding terrestrial habitats has rarely been studied together at a whole watershed scale. Based on DNA observations, previous studies that linked terrestrial to aquatic ecosystems have converged to report large contributions of terrestrially-derived taxa within aquatic microbial assemblages (Ruiz-González *et al.*, 2015a), especially in systems with short residence times and higher connectivity to the surrounding terrestrial milieu (Crump *et al.*, 2012; Besemer *et al.*, 2013; Ruiz-González *et al.*, 2015a). Based solely on the analysis of DNA, we similarly observed a high prevalence of terrestrially-derived taxa along the entire aquatic continuum, but also a clear divergence in community structure between terrestrially-influenced (soil, soilwaters, groundwater) and larger aquatic water bodies (river, reservoirs) along this continuum. Streams, headwater ponds, tributaries and small lakes represented intermediate states between the two endpoint community structures. Notably, community structure of lakes was extremely heterogeneous and possibly reflected variations in the combination of network position (Carrara *et al.*, 2013) and residence time (Logue and Lindström, 2010). Clustering of reservoirs with riverine lakes and some larger lakes may indicate that residence time is likely a strong driver of community structure (Lindström

et al., 2006; Ruiz-González *et al.*, 2015a; Niño-García *et al.*, 2016a).

DNA alone, however, only provides a partial view of the underlying assembly processes as it has been unclear whether the observed strong terrestrial signature in aquatic systems (Ruiz-González *et al.*, 2015a) represents passive transport or active selection. After combining DNA-RNA-patterns, we observed that this strong terrestrial imprint is not limited to the unreactive fraction but was even more pronounced in the reactive taxa that we identified along the whole continuum. These terrestrially-derived taxa are likely a mixture of true terrestrial taxa and aquatic taxa that were once dispersed into and subsequently persisted in soils (i.e., seeds) (Ruiz-González *et al.*, 2015a, 2017b). The only aquatic habitat where the terrestrially-derived taxa did not overwhelmingly dominate the reactive pool was the estuary, where local estuarine taxa became more relevant. Although most reactive taxa could be retraced to soils and soilwaters in all habitats, there were nevertheless taxa that were locally recruited along the aquatic continuum that contributed to the reactive portion, and some taxa first appearing in streams, rivers and reservoirs became reactive downstream, including in the estuary. The comparably small but still relevant contribution of these potentially aquatic taxa to the reactive portion became more evident during our rarefaction test (details not shown, SM 1.8.4). We observed a larger proportion of stream and upriver taxa (10 - 25%) in the reactive fraction with lower rarefaction thresholds, indicating that many of the terrestrial taxa are indeed very rare and drop out of the analysis with low rarefaction thresholds. Remarkably, recruitment into the reactive fraction in each habitat occurred across the rank abundance curve, indicating that rare taxa may be highly responsive to the environment and therefore contribute to local ecosystem processes (Jones and Lennon, 2010; Campbell *et al.*, 2011; Hausmann *et al.*, 2019). Overall, the vast majority of taxa that showed local increases in abundance along the aquatic continuum mostly remained unreactive, indicating that mass effects are not only limited to terrestrially-derived taxa but apply similarly to other taxa first detected in aquatic habitats that are carried along the network as a historical imprint (Hauptmann *et al.*, 2016). These other taxa may not necessarily be strictly aquatic but could be taxa recruited from other source habitats that were not sampled within this study such as aeolian dispersed, wetlands, biofilms, and/or host-associated microorganisms. These taxa collectively form the historical imprint, which grows as the water traverses through various habitats along the flow path, and hence at any time within the

network a fraction of the bulk community embodies previous migration and selection that has happened in upstream habitats.

1.5.2 Spatial and seasonal shifts in dominant assembly processes

DNA- and RNA-based community structures were generally coherent, showing both seasonal as well as spatial gradients in almost a mirroring pattern within the PCoA, which has been observed previously (Logue and Lindström, 2010; Wilhelm *et al.*, 2014), however, their degree of similarity varied greatly in both time and space. As observed in our incidence and abundance-based dissimilarity comparison (Fig. 1.4), there is a very strong seasonal pattern and a clear spatial structure along the continuum in how DNA and RNA community structures relate to each other. Together with the pattern in reactive versus unreactive taxa, we observed a clear shift in the overall dominant assembly process at the whole network scale between seasons (Fig. 1.6). Dominance of mass effects in spring and increases in species selection in summer/autumn were likely driven by seasonality in hydrology (e.g., seasonal shifts in discharge (Figs. S1.12, S1.13) and reservoir residence time (Table S1.2; Luo *et al.* (2020)), temperature and other environmental factors (Paruch *et al.*, 2020). Where within the network species selection was most prominent, however, also differed between seasons, as evidenced in the network patterns of dissimilarity between DNA and RNA community structure and in the distribution of reactive and unreactive taxa (Fig. 1.6). In particular, we observed that reservoirs and subsequent downstream habitats were sites of more intense selection in spring, whereas upstream riverine sites became selection hotspots in summer, albeit selection continued to occur in downstream habitats (Fig. 1.6). Increasing selective pressure along a riverine residence time gradient has been hypothesised before (Read *et al.*, 2015), however, we have shown that the location of strongest selection shifts depending on the hydrological conditions. As such, high flow conditions push selection hotspots downstream, while low flow pulls selection upstream within the network. This seasonally moving window of selection hotspots is likely driven by a balance between the time a water parcel travels along the network and the time a taxon needs to react and grow, which is itself related to temperature and other environmental factors (Ruiz-González *et al.*, 2017b).

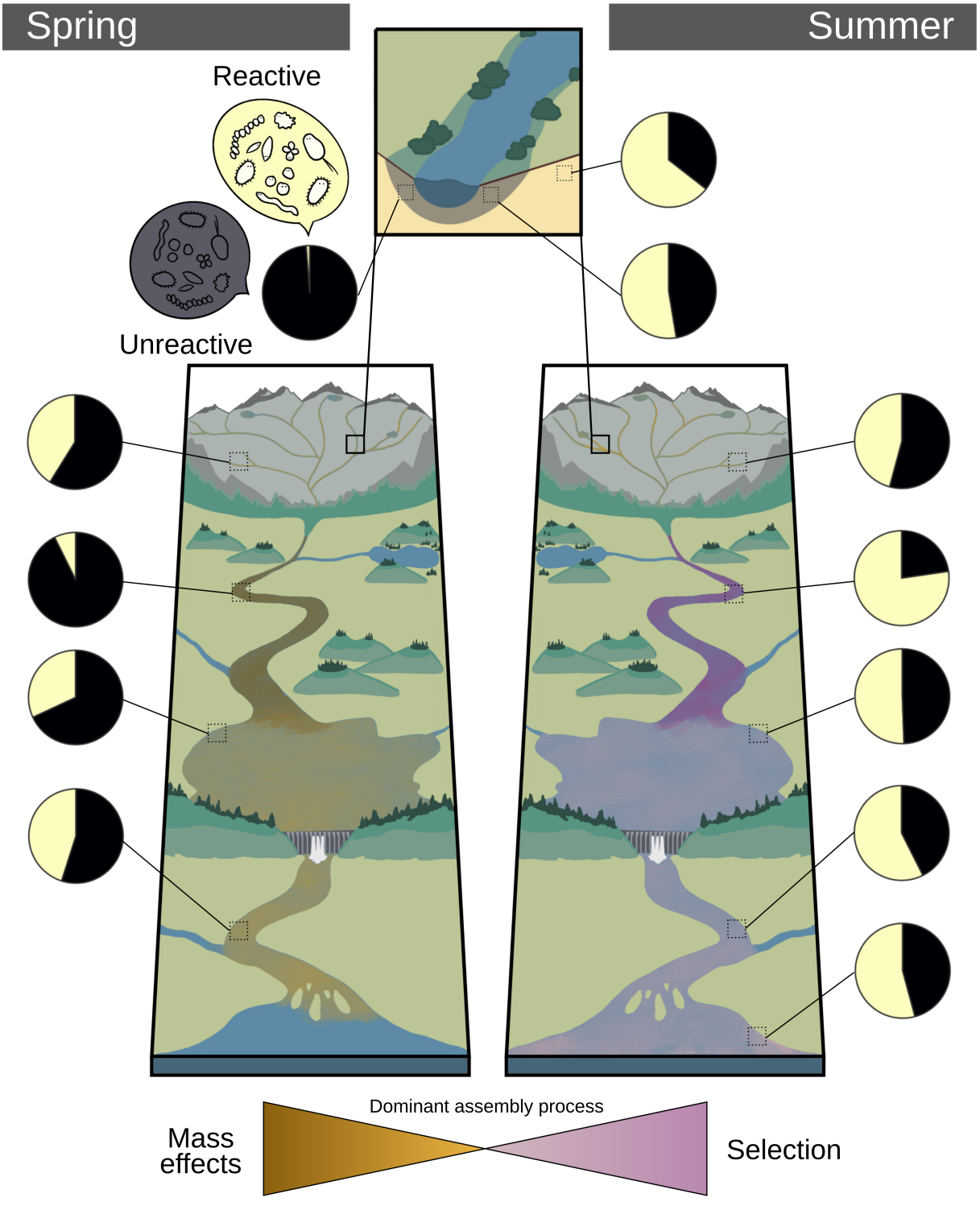

Figure 1.6: Conceptual figure of seasonally dominant assembly processes along the terrestrial-aquatic continuum. Pie charts visualise the proportion of bacterial OTUs (%) that were identified as reactive and unreactive, respectively (Fig. S1.11).

The only exception from this strong seasonal and shifting spatial pattern were small order streams, which are largely influenced by inputs from the surrounding catchment, as has been shown before (Crump *et al.*, 2012; Besemer *et al.*, 2013; Ruiz-González *et al.*, 2015a; Caillon *et al.*, 2021). Yet, streams were also found to be sites of strong selection independent of season, with consistently high proportions of reactive taxa relative to the connected soilwater communities and relatively large divergence between DNA-RNA-based community structure. Higher growth rates of stream bacterioplankton have been observed previously compared to higher residence time habitats (Stadler, 2021), and low order streams have been shown to be sites of intense processing of terrestrially-derived dissolved organic carbon (Hutchins *et al.*, 2017). The shallow environment with a plethora of fresh dissolved organic matter, likely provides unique niches that favor growth and selection of taxa that are washed in from the soil. Stream biofilms may help explain the relatively high selection despite low residence time as they are known to adapt their structures depending on the force of flow (Besemer *et al.*, 2007) and create microhabitats of enhanced microbial growth and biogeochemical processing (Battin *et al.*, 2003). It is outside the scope of this study to evaluate biofilms as an additional source for bacterioplankton communities (McDougald *et al.*, 2011), however, biofilms may similarly recruit terrestrial taxa (Besemer *et al.*, 2012). Our results suggest that the selection processes within these streams likely operates mostly by reshuffling terrestrially-derived taxa rather than selectively recruiting freshwater taxa, as we have observed an overwhelming proportion of terrestrially-derived taxa within the reactive fraction. This stream-filtered, terrestrially-derived community is what becomes the core of the aquatic historical imprint that will be transported throughout the network, representing the basis of dispersal and mass effect to the downstream habitats. This historical imprint is also a source of recruitment along the continuum, thus shaping the whole network scale community all the way into the estuary.

Together, our results suggest a framework wherein low order streams are sites of both intense mass effects and selection across seasons, while further hotspots of selection downstream are modulated by seasonality. As such, high flow conditions only allow for selection hotspots to occur once the network enters longer residence time habitats (i.e., larger reservoirs, lakes) and selection continues to prevail in subsequent downstream habitats such as the river. In contrast, low flow scenarios, which often correspond to higher temperatures and other environmental conditions

that favour growth (except under winter low flow conditions, not covered in this study), allow selection hotspots to occur earlier (i.e., upstream river) as travel time decreases. This framework suggests that mass effects are present at all times - regardless of high or low flow - within the network, resulting from both influx of bacteria from the surrounding terrestrial soils and a fluvial carry-over of aquatic bacteria that encompasses the legacy of upstream assembly processes. It is rather the degree and location of selection that can vary significantly in magnitude within the network depending on seasonality. The resulting recruitment of taxa occurs across the rank abundance curve, contradicting the common assumption that only the most abundant bacteria are reactive to the environment and contribute to ecosystem processes. Given the large role that the historical imprint plays in shaping microbial communities within the network, it is not surprising that the various attempts to link microbial processes to the bulk community composition often yield indecisive results (Hall *et al.*, 2018). In conclusion, our study highlights the importance of conceptually and empirically considering the potential downstream effect of upstream habitats and the seasonality of these influences when examining the assembly of aquatic microbial communities and the ecological underpinnings of the rank abundance structures observed in aquatic habitats.

1.6 Acknowledgements

We would like to begin by acknowledging that the land on which the field work was conducted is on the traditional and unceded territory of the Innu community (Nitassinan), and the laboratory work was conducted on the land of the Kanien'keha: ka people, as well as of the Métis people. We are especially grateful to Alice H. Parkes, Annick St-Pierre, and Serge Paquet, who maintained and oversaw the La Romaine project over the years. Collection and analysis of all variables would have not been possible without the support from members of the CarBBAS team including great contributions from undergraduate students. Therefore, we would like to thank Felipe Rust, Clara Ruiz-Gonzàlez, Trista Vick-Majors, Alexandre Ducharme, Roy Nahas, Ryan Hutchins, Marie Laure Gérardin, Erin Hotchkiss, Karelle Desrosiers, Martin Demers, Sara Mercier Blais, Julia Jakobsson, Francesca del Giorgio, Brenden Chabot, Sebastian Dugas, and Pascale Ouimet. We would also like to thank Katherine Velghe and Marilyne Robidoux for laboratory assistance, Sophie Crevecoeur

and Mario Muscarella for insightful comments on the initial draft of this manuscript and Yves Prairie for statistical advice. Finally, we would like to thank two anonymous reviewers and the editors for their constructive comments that greatly improved the manuscript. This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Québec.

1.7 Data availability

The raw 16S rRNA gene sequences, both DNA and cDNA are available at the public NCBI Sequence Read Archive (SRA) as part of the BioProject PRJNA693020. The code is available on Github (`https://github.com/CarBBAS/Paper_Stadler-delGiorgio_ISMEJ_2021`) and both code and processed microbial data were separately archived on Zenodo (Stadler, 2021; Stadler *et al.*, 2021).

1.8 Supplementary Information

1.8.1 Catchment characteristics

The river springs between the Atlantic and Saint Lawrence watersheds (52°52'20"N 63°36'55"W; elevation: 702 masl), and consequently flows through a series of lakes (hereafter riverine lakes) including the biggest lake in the catchment – Lake Brûlé (*A*: 127.11 km^2, elevation: 470 masl). The river mainly flows towards the South with a maximum distance from the northern headwaters to the river mouth expanding to approximately 475.1 km.

The catchment was glaciated 7,000 – 10,000 years ago and left mostly a till blanket and veneer as surficial material. It is mainly dominated by acid rocks (e.g., granodiorite, granite, quart diorite) with granitised sedimentary and volcanic rock, and has isolated patches of permafrost (0 - 10%)(Natural Resources Canada). The soil is composed of roughly 61.4% sand, 31.9% silt, 6.7% clay and stores approximately 140.4 t ha^{-1} of organic carbon (in top 5 cm; given are catchment averages) (Lehner and Grill, 2013; Hengl *et al.*, 2014).

The northern part of the catchment is characterised by a flat open black spruce (*Picea mariana*)-

lichen forest with shrubs and moss-lichen (Fig. S1.1a). As one follows the river downstream, the relief changes drastically to a steep mountainous stretch that forms sections of canyons (Fig. S1.1b). The river looses 330 m of elevation from the mountainous section until it makes a sharp turn to the west into the lower coastal plain. The coastal plain is characterised by peatland areas with swamps and shallow waters that are completely permafrost free (Fig. S1.1c). There are two larger tributaries in the coastal plain that flow through the lakes Puyjalon (*A*: 13.10 km^2) and Allard (*A*: 19.24 km^2). Nearly half of the catchment is covered by coniferous forests (*P. mariana*-moss), with mixed forests being rather minor (11%) and deciduous stands with white birch (*Betula papyrifera*) and trembling aspen (*Populus tremuloides*) are even more rare (2%) (Bureau d'audiences publiques sur l'environnement and Canadian Environmental Assessment Agency, 2009).

The Romaine river was dammed during the sampling period, forming a reservoir cascade complex with 4 reservoirs by 2020 after the sampling period. The reservoirs Romaine 2 (RO2, *A*: 81.15 km^2, mean depth: 61 m), Romaine 1 (RO1, *A*: 13.22 km^2, mean depth: 22 m) and Romaine 3 (RO3, *A*: 35.18 km^2, mean depth: 66 m) were flooded in the years 2014 (winter), 2015 (winter), and 2017 (spring), respectively.

A weather station located in the lower coastal plain (50° 16'55.000" N, 63° 36'41.000" W, Havre-Saint-Pierre Airport, Natural Resources Canada) recorded an annual precipitation of 810.77 $\pm$ 35.25 mm and 1.18 $\pm$ 0.73°C, -32.63 $\pm$ 1.36°C, and 25.8 $\pm$ 0.66°C for mean, minimum and maximum temperature over the sampled years.

Figure S1.1: Landscape within the Romaine catchment. a) Northern area with shrubs and moss-lichen, b) Mountainous section close to Reservoir 3, c) Lower coastal plain with peatland areas, d) Example of a sampled stream.

1.8.2 Sample processing and sequencing

Surface water samples were directly collected into a pre-rinsed carboy bottle at a depth of 0.5 m, close to the shore for stream samples and diverse locations within the river and reservoirs. Surface soil samples were collected by mixing three randomly selected cores (30 cm) that were taken in proximity of installed piezometers to sample soilwater. The upper 5 cm including surface vegetation were removed before the soil was transferred into a sterile plastic bag. Three piezometers were randomly installed in proximity (30 - 100 cm) to a sampled stream with an average depth of 50 $\pm$ 20 cm. However, if the piezometers were installed too close to the stream main channel, hyporheic water was sampled instead. Piezometers were emptied 3 times (1 - 2 h) with a peristaltic pump before sample water was collected. The water from the piezometers was pooled for

each site. Groundwater was directly collected from constructed wells with submersible pumps. Lake sediment samples were collected with sediment cores (1 - 2 m depth), and the upper 10 cm were collected and mixed for subsequent processing. All samples were stored in cooler boxes until return to the laboratory (maximum duration 7h), and were subsequently stored at 4°C upon arrival at the laboratory until further processing on the same day of sampling. A minimum of 25 mL and 250 mL of soil-/hyporheic-water and surface water, respectively, were filtered through 0.22 μm polycarbonate membrane filters (Merck Millipore, Darmstadt, Germany). Homogenised soil and sediment samples were transferred to aliquots of 0.25 g. After filtration, samples for RNA extraction were submerged in RNAlater and LifeGuard Soil Preservation solution (QIAGEN, Hilden, Germany) for water and humic samples (soil, soilwater, hyporheic water), respectively. To allow stabilisation in the buffer, samples were left at 4°C overnight and were subsequently stored frozen. All DNA and RNA samples were frozen at -20°C at the field station and further stored at -80°C at the university laboratory until extraction.

For extractions, PowerWater and PowerSoil DNA and RNA extraction kits (MoBio, Carlsbad, CA, USA) were used to extract water and soil/soil-/hyporheic-water/sediment samples, respectively. In 2017, the equivalent DNeasy and RNeasy PowerWater Kits (QIAGEN, Hilden, Germany) were used for DNA and RNA samples, respectively, due to discontinuation of the MoBio kit series.

Prior to cDNA reverse transcription, RNA extracts were checked for DNA contamination with a negative PCR test. Subsequently, cDNA was synthesised with a high capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). Successful DNA extraction and cDNA synthesis was evaluated via PCR amplification of the 515F-806R primers (IDT Technologies, Coralville, IA, USA) and DNA concentration was measured with a NanoDrop 2000c (Thermo Fisher Scientific Inc., Waltham, MA, USA).

1.8.3 Bioinformatic analysis

Primers were removed from 16S rRNA DNA and cDNA (hereafter RNA) data sets using the software *cutadapt* (v1.18, Martin (2013)), which allows for the removal of the primer sequence and its variants in their true and complement orientations. Additionally, all reads shorter than 125

nucleotides were removed as they cannot achieve a minimum overlap necessary for paired-end merging in downstream processing.

To identify amplicon sequence variants (ASVs), 16S rRNA amplicon reads were analysed through the DADA2 (Divisive Amplicon Denoising Algorithm 2) pipeline (v1.14.1, Callahan *et al*. (2017)) in R (v3.6.3, R Core Team (2024)). Read qualities were evaluated for each sequencing plate separately and read length was trimmed according to their quality scores. Samples were pooled by plate, season and sequencing depth for learning the error rates. DADA2 runs on a sample by sample basis, and thus removes observed singletons by sample to avoid inclusion of false-positive sequencing errors. To retain more rare taxa within a sampling campaign (year-season combinations) along the continuum, samples were 'pseudo'-pooled for the *dada()* step. This step enables the removal of singletons by pool but retains singletons within a sample. Paired-ends were merged after successful inference of amplicon variants. Chimeras were removed (*removeBimeraDenovo()* function) and, finally, taxonomy was assigned with the *DECIPHER* package (v2.14.0, Wright (2016)) implementing the increased accuracy IDTAXA algorithm (Murali *et al.*, 2018) and the provided trained classifier of the GTDB database (Release 95, Parks *et al*. (2018)). Only ASVs that were classified as Bacteria and not as Mitochondria or Chloroplast were evaluated in this study. Several ASVs were found to be highly abundant only in RNA. To account for slight differences that may have emerged between DNA and RNA ASVs and also to merge potential differences among 16S rRNA copies within a single genome, ASVs were merged into OTUs by a 99% similarity threshold (Větrovský and Baldrian, 2013) with the *DECIPHER* package (*AlignSeqs()*, *DistanceMatrix()*, *IdClusters()* functions, Wright (2016)). The sequence of the most abundant ASV within a OTU cluster was kept as a reference sequence if it was classified as at least at the domain level as 'Bacteria'. If the most abundant ASV within a cluster did not have any taxonomic classification, the ASV that had a taxonomic classification was chosen to represent the OTU cluster.

1.8.4 Effects of rarefaction

Cumulative sum scaling (CSS) results were compared with results achieved with various rarefaction thresholds. There were no substantial differences in the results between CSS and various rarefaction thresholds on α diversity estimates, which is believed to be most susceptible to library

size differences (Fig. S1.2).

We further re-run all analyses in the manuscript with the examined rarefaction thresholds. Both PCoAs showed negligible differences while the patterns in Sørensen and Bray-Curtis distances showed variation across rarefaction thresholds. Depending on the rarefaction threshold used, we observed different patterns in when incidence and when abundance-based distances were greater than the other, which was likely introduced by the random sampling procedure imposed by rarefaction affecting both incidences as well as abundances. As the threshold becomes higher (e.g., 10,000), all patterns in the absolute numbers of Sørensen and Bray-Curtis distances approach our originally observed patterns with CSS. We mainly use the Δ-distances to interpret shifts in mass effects and selection, and while the absolute numbers in Δ-distances changed, the patterns across habitat types and seasons remained fairly consistent across rarefaction thresholds with the highest rarefaction threshold resembling the original results the most. Hence, these results showed that the absolute numbers in Δ-distances do not hold meaning per se; it is the spatio-temporal relative change that gives our analysis meaning. Rarefaction does change the DNA-RNA distances themselves likely due to the loss of rare taxa and making abundance differences among OTUs smaller, however, the final interpretations remain the same.

Finally, we explored the effects of rarefaction on how reactive and unreactive OTUs were identified. Across all rarefaction thresholds, we could observe a clear differentiation between OTUs that were classified within each habitat as 'unreactive'. Firstly, due to their absence in any RNA and secondly, by the absence in a DNA to RNA contribution relationship. OTUs that were categorised by these two 'unreactive' categories were clearly different from those that were named 'reactive' due to the presence of a linear relationship in their DNA and RNA contribution. Within each rarefaction threshold, unreactive OTUs were characterised by diverse origins, and reactive taxa were mainly dominated by soil and soilwater taxa. We could also observe that with lower rarefaction thresholds (e.g., 1,470), more taxa from 'aquatic' habitats seemed to be reactive (i.e., taxa first detected in streams and upriver), however, as the rarefaction threshold increases, their proportion reduces (e.g., 10,000). These results point to the fact that rare taxa are important contributors to our observed patterns, and hence we strongly believe that rarefaction does rather bias

our observations by removing taxa essential to natural processes in microbial assembly (e.g., seed banks; Lennon and Jones (2011)). Rarefaction does not affect our observation that there were taxa that have a decoupled DNA-RNA relationship (below potential activity threshold), and those that do (above potential activity threshold). Consequently, CSS results were used for our manuscript.

1.8.5 Data exploration and statistical analyses

To explore differences in microbial community composition across habitat types and seasons, a Principal Coordinates Analysis (PCoA) was conducted with Bray-Curtis dissimilarities (D_{BC}) (Bray and Curtis, 1957; Legendre and Legendre, 1998) based on all DNA samples with the function *pcoa* in the *ape* package (Paradis and Schliep, 2018). The community matrix was Hellinger transformed to resolve a horse-shoe effect (Legendre and Gallagher, 2001). To correct any negative eigenvalues problematic for PERMANOVA analysis, the D_{BC} matrix was square-root transformed to Euclidean distance (Legendre and Legendre, 1998; Borcard *et al.*, 2011). To evaluate statistical differences in habitat type and season a PERMANOVA was computed with 9,999 permutations with the *adonis* function. A PERMANOVA cannot distinguish among-group from within-group variation if data dispersion is variable among groups (Anderson and Walsh, 2013), therefore, an analysis of multivariate homogeneity was computed with *betadisper*. Using *permutest*, we finally tested whether dispersion differs between groups.

Secondly, to evaluate whether sampled RNA-based assemblages were different from the DNA-based assemblages, we performed a second PCoA (D_{BC} with square-root transformation) with both DNA and RNA samples. Again, statistically different groups were investigated with a PERMANOVA (9,999 permutations), where habitat type, season and nucleic acid type (DNA versus RNA) formed the groups. The same framework explained above to check for dispersions was applied. To quantify how different DNA-/RNA-based assemblages of the same sample are, the Bray-Curtis distance (m_{BC}) of each DNA-RNA sample pair within the PCoA ordination space was computed across *n*-dimensional space (Tabak, 2004):

$$m(p,q) = \sqrt{(\mid p_1 - q_1 \mid)^2 + (\mid p_2 - q_2 \mid)^2 + \cdots + (\mid p_n - q_n \mid)^2}$$

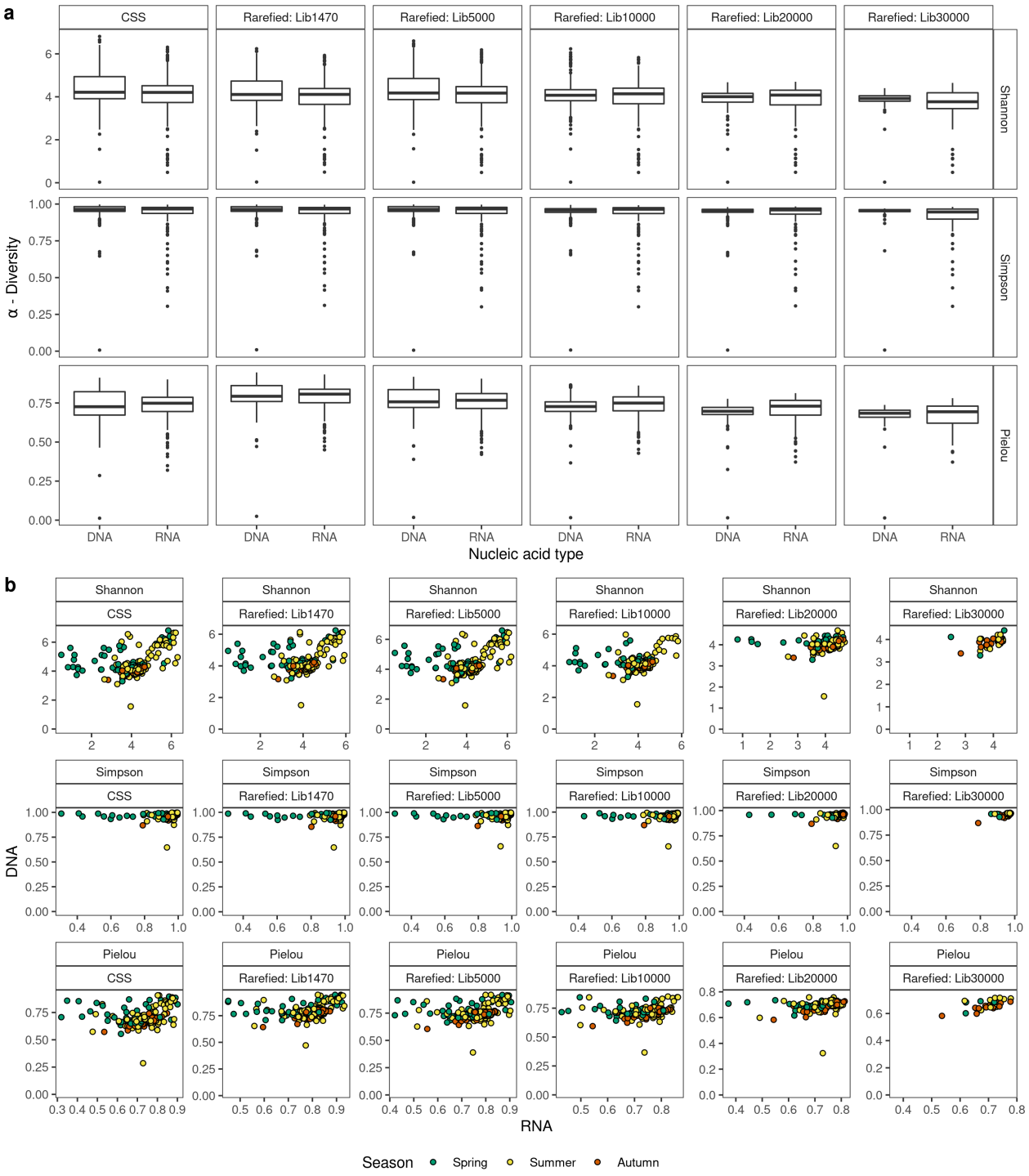

Figure S1.2: Comparing the effect of rarefaction on α diversity patterns. Three different α diversity indices were investigated: Shannon-Wiener index (H'), Simpson's index (λ) and Pielou's evenness (*J*). CSS = cumulative sum scaling. Rarefied datasets with the applied minimum library size threshold are indicated as 'Lib[threshold]'.

where *p* and *q* represent DNA and RNA site scores, respectively, of each sample and *n* is the used maximum number of dimensions. We focused on the first axes that cumulatively explain 75% of the variation for each ordination ($n_{75\%}$), similar to Osterholz *et al.* (2016). This approach was implemented as it was evident from the PCoA that essential variation within non-aquatic samples was captured outside the first three axes. The distance across all PCoA axes equals the initial pair-wise dissimilarity on which the PCoA is based on (Fig. S1.3). As such, the distance across $n_{75\%}$ axes extracts the proportion of the initial pair-wise dissimilarity that is captured by the axes cumulatively explaining 75% of the PCoA.

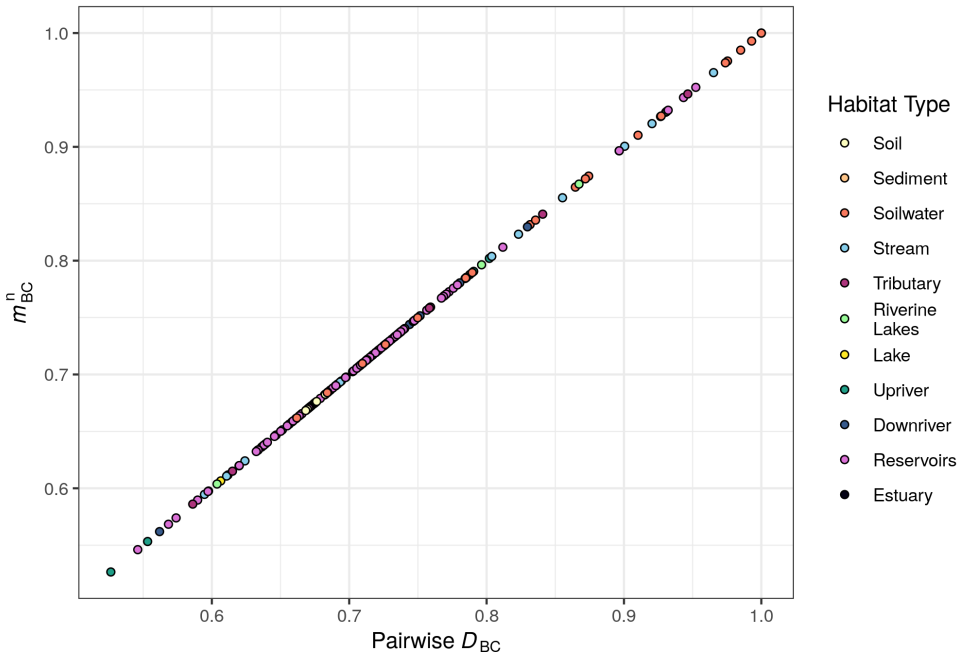

Figure S1.3: Distance along all PCoA axes equal pair-wise dissimilarity. As a proof of concept, the computed distance (e.g., m_{BC}) extracts a proportion of the individual pair-wise dissimilarities, summation of the distances across all PCoA dimensions equals the initial pair-wise Bray-Curtis dissimilarities, which is the input matrix into the PCoA. Thus, computing the pair-wise distance among a sub-selection of axes of the PCoA captures a proportion of the overall pair-wise dissimilarity that is explained by the variation and drivers of the selected axes.

To gain further insight into the processes shaping assemblage dissimilarities, we computed a PCoA with the Sørensen dissimilarity (D_S), which is the incidence based equivalent of D_{BC} (square-

root transformed to achieve Euclidean space) (Legendre and Legendre, 1998; Sørensen, 1948) (Fig. S1.4). By comparing incidence and abundance based dissimilarities, we can further distinguish in which samples DNA-/RNA-based assemblages diverge primarily due to different present taxa or their abundances, respectively. We further applied the same framework of calculating the distance among DNA and RNA pairs across $n_{75\%}$ axes resulting the Sørensen-based distance (m_S).

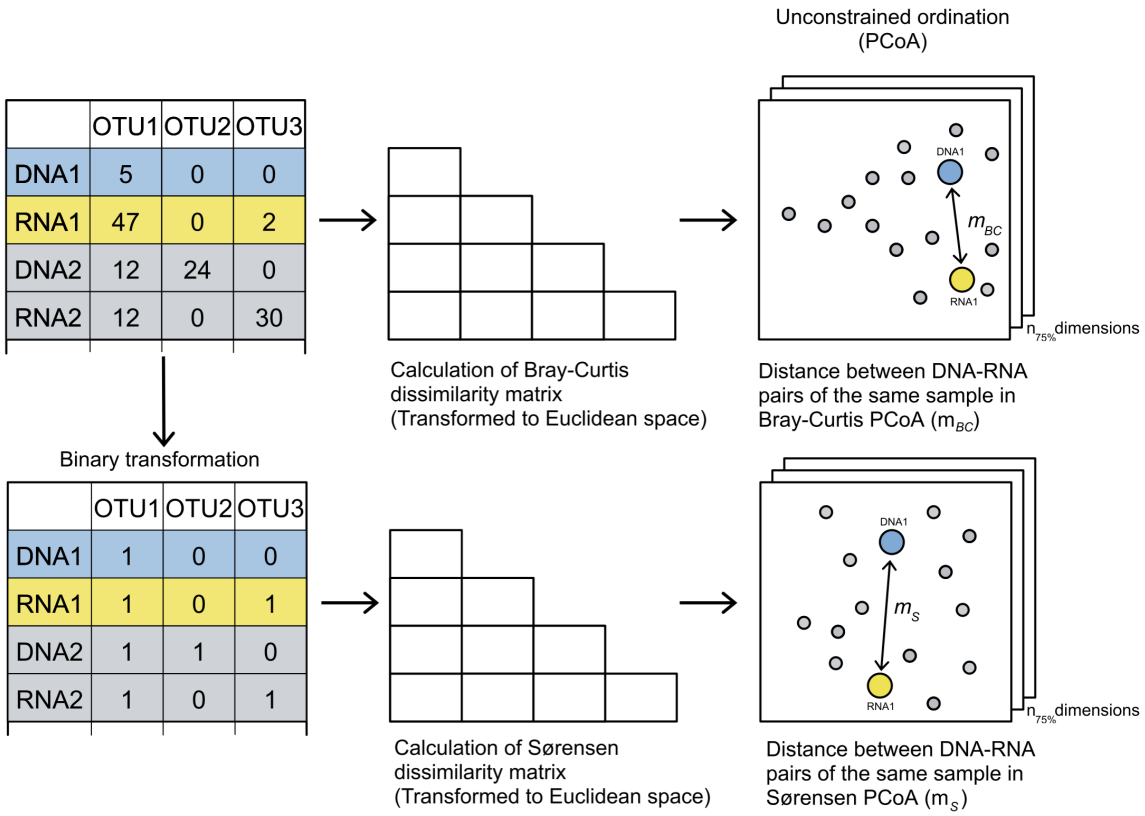

Figure S1.4: Schematic representation of n-distance calculation.

1.8.6 Simulation of theoretical communities to understand underlying patterns of Δ-distances

In order to support our approach to examine assembly processes using the Δ-distances approach between Sørensen and Bray-Curtis dissimilarity based PCoAs, we simulated different scenarios and calculated the Δ-distances on these theoretical communities.

We used four different species abundance distributions (SADs) of varying evenness (Pielou's *J*)

to mimic the DNA assemblages of different habitat types. We hypothesised that the degree of selection increases with lower evenness (higher Δ-distance), as only a few taxa are selected for and dominate the community. Each SAD was composed of 10,000 OTUs (Lennon *et al.*, 2018), which is close to our overall OTU pool of 16,322 OTUs found across the dataset. We used log-normal distributions to create SADs and modified the evenness by changing the scale parameter between 0.9 and 2.7. The higher the parameter is, the steeper and less even the community becomes. To create the DNA assemblages, we randomly sampled 25,000 times from each SAD, which was based on our average library size across the dataset (mean: 24,687.21).

As a second step, we duplicated the randomly sampled DNA assemblages to retrieve a base community for the corresponding RNA assemblages. To introduce additional mass effects, we implemented a gradient of OTU removal from the created DNA assemblages. The number of OTUs removed from the RNA assemblages were either 1/2, 1/3, 1/6 or 1/9 of all sampled OTUs in a site (hereafter, replacement). We hypothesised that the higher the replacement, the stronger the mass effect will be (lower Δ-distance). The number of reads that were removed as a result of the random OTU removal were re-sampled from the OTUs that were not present in the DNA assemblage of a particular site, leading to equal library sizes across sites. We further corrected for phantom taxa (RNA > 0, DNA = 0) with DNA = 1, following the analysis of our empirical dataset. Overall, we implemented 4 SAD and 4 replacement treatments that were run 9 times to avoid random sampling biases and compute standard deviations of the resulting metrics (overall n = 288).

Once the OTU matrix with all sites were set-up, we re-ran our analysis by calculating the PCoAs for both Sørensen and Bray-Curtis based dissimilarities, extracted the axes that cumulatively explain 75% of the variation and calculated the Δ-distances. Furthermore, two metrics were calculated for each site: 1) the number of OTUs that do not have any RNA, which were classified as 'unreactive' taxa in our study, and 2) the mean read difference between individual OTU's DNA and RNA, which indicates the discrepancy between the DNA and RNA. We hypothesised that higher DNA-RNA discrepancies indicate stronger selection.

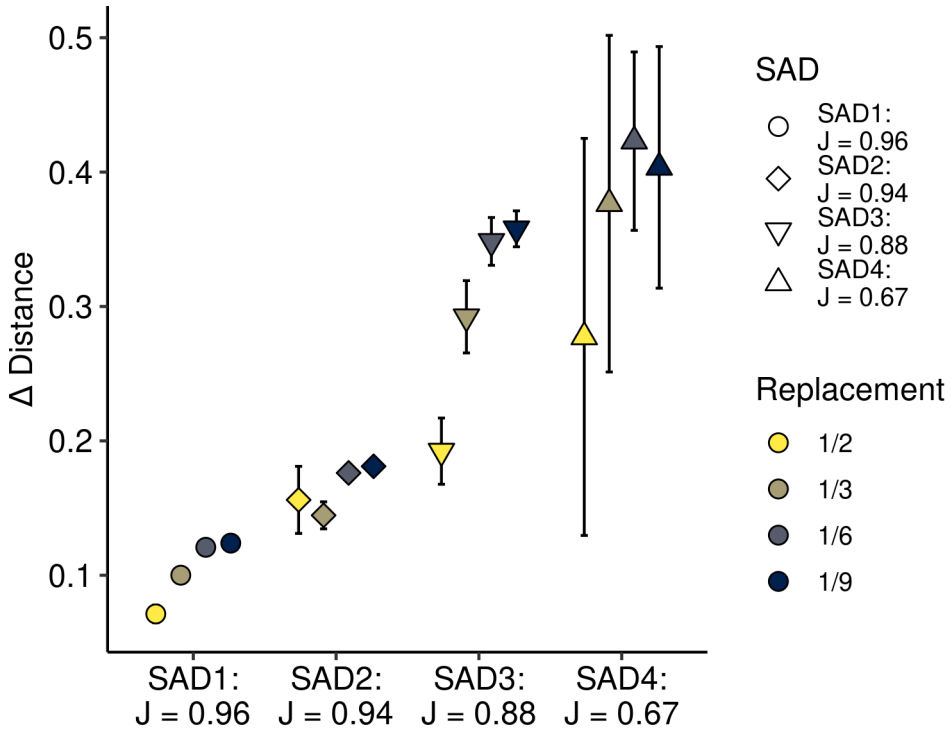

Figure S1.5: Simulated Δ-distances. Points represent the arithmetic mean of 9 independent simulations of SAD and replacement categorical combinations. Error bars indicate the standard deviation from the arithmetic mean.

Our simulation results indicate that there is indeed a clear trend in our evenness and replacement treatment on Δ-distances with lower Δ-distances observed in even communities (i.e., stronger mass effect) and higher Δ-distances in uneven communities (i.e., stronger selection) (Fig. S1.5). Additionally, higher replacement values lower the Δ-distances within each SAD treatment indicating stronger mass effects.

We also regressed the two above mentioned metrics with the Δ-distances (Fig. S1.6), and we found a negative relationship of the number of 'unreactive' OTUs with Δ-distances in line with our interpretation of mass effects and its dependency on a higher proportion of unreactive taxa. Furthermore, mean abundance differences were positively related to Δ-distances indicating stronger selection with higher abundance differences between DNA and RNA. Although our empirical relationships (Fig. S1.6c-d) are less stronger than the simulation based estimates (Fig. S1.6a-b), the overall trends are the same.

We would like to end this simulation section by highlighting that the absolute number of Δ-distances is likely to hold little meaning. Absolute Δ-distance values are not interpretable on their own, it is the comparison across habitats or gradients that create patterns and gives this approach room for interpretation.

1.8.7 Abundance classification

Traditionally, abundance groups (AGs) such as 'abundant' and 'rare' have been defined by various relative abundance thresholds ranging from 0.1 to 1% within the literature. While inconsistencies hinder comparisons among studies, we additionally are working with variance stabilised read numbers, thus traditional thresholds based on relative abundances are not applicable. In order to classify OTUs into AGs, we developed a new framework to classify OTUs into AGs based on the shape of rank abundance curves of each habitat. We initiate the framework by calculating the mean abundance of each OTU by habitat type. Subsequently, for each habitat type a smoothed rank abundance curve was generated with the function *smooth.spline* with the smoothing parameter set to 0.7 in R (*stats* package; R Core Team (2024); Fig. S1.7). Ranks that correspond to moments of acceleration along the curve were identified by taking the second derivative of the log(x + 1) transformed abundance curve (Fig. S1.7c). All OTUs ranked below the second maximum acceleration were defined as rare. OTUs falling above the first maximum acceleration were defined as abundant, while the section in between the two maxima represents moderately abundant OTUs (Fig. S1.7a). The CSS reads corresponding to the ranks identified for the AGs were extracted for each habitat type separately. Subsequently, the average CSS reads for each abundance threshold was calculated. This approach classified all OTUs with >= 72 CSS reads as abundant, < 72 and >=

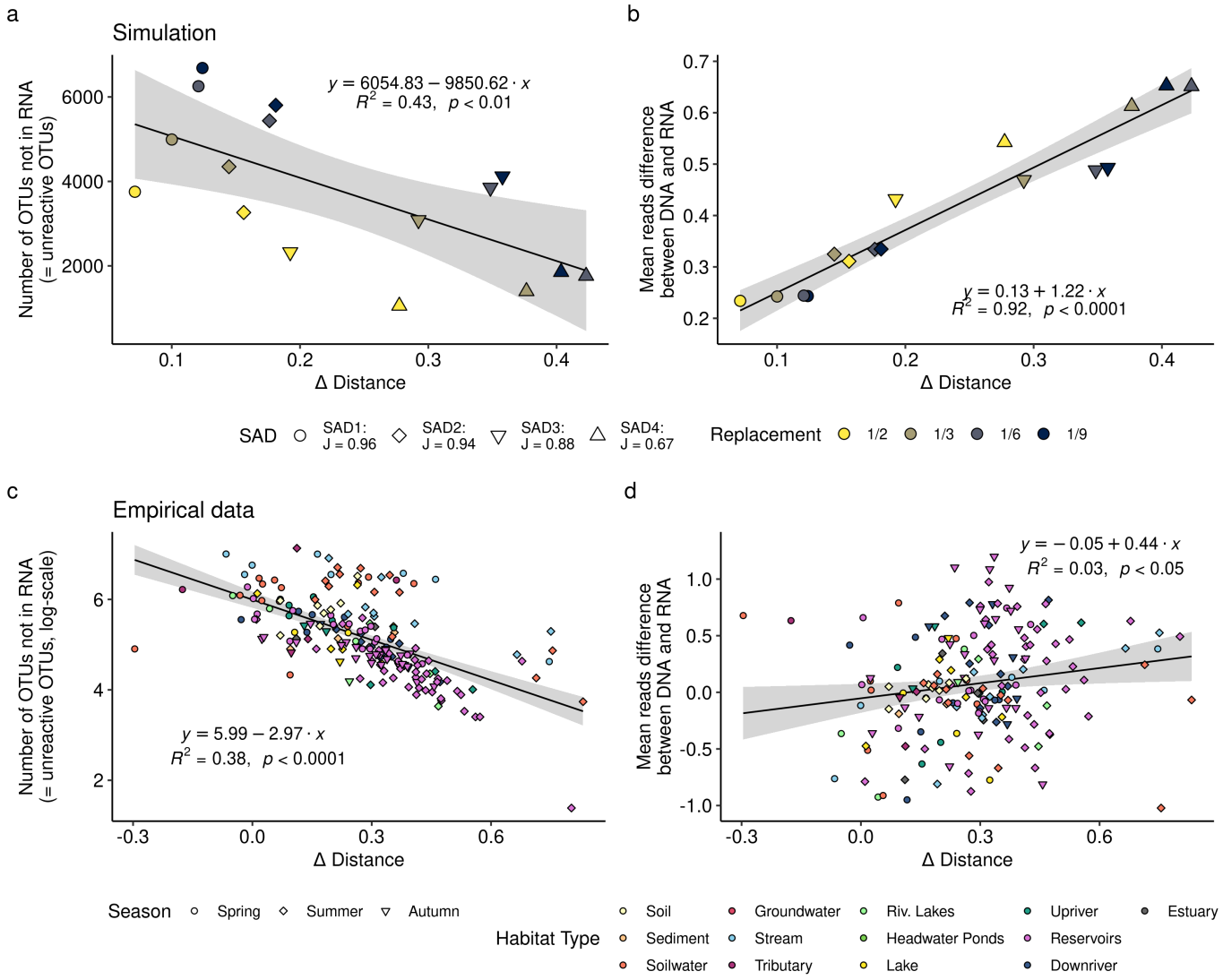

Figure S1.6: Linear regressions of Δ-distances and computed metrics. Points are arithmetic means for simulations (a-b) and individual samples for empirical data (c-d) (simulation: n = 16, empirical data: n = 193). Assumptions of normality and homogeneity were checked for each linear regression. Lines and grey areas represent regression slope and confidence intervals, respectively. Regression formulae and statistics are given within the individual plots.

10 CSS reads as moderate, and < 10 CSS reads as rare.

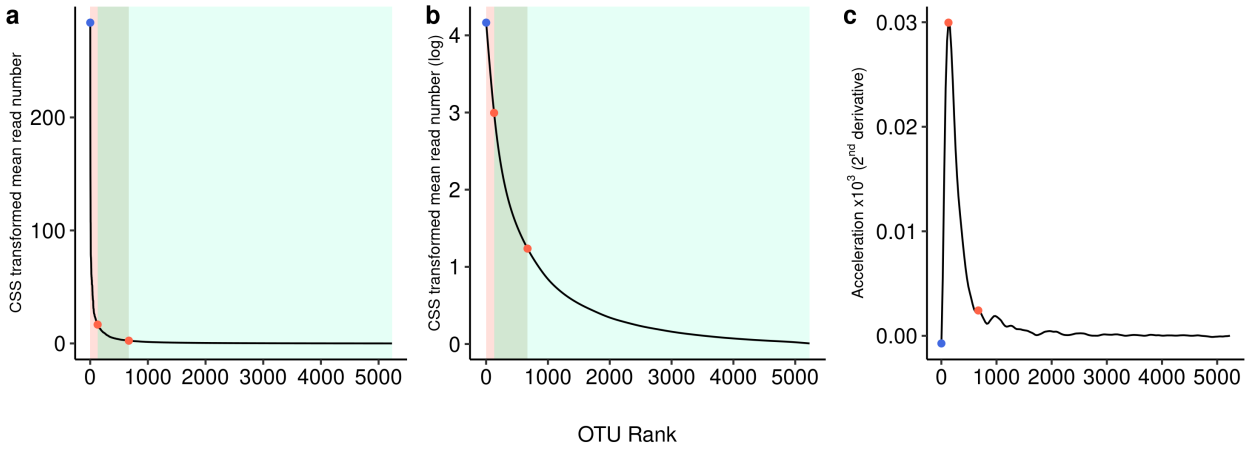

Figure S1.7: Classification of abundance groups. Schematic representation of the used approach to classify abundance groups based on derivative approximation of the log-transformed rank abundance curve by habitat type, where a) represents the original rank abundance curve and b) the log-transformed equivalent to derive points of maximum and minimum acceleration. The blue point represents the first minimum of the second derivative, red points are the first and second maxima of the second derivative. Pink, green and blue ranges visualise abundant, medium and rare classifications, respectively.

1.8.8 Software details

The packages *phyloseq*, *tidyverse*, *plyr* and *data.table* were used for data wrangling and transformation (McMurdie and Holmes, 2013; Wickham *et al.*, 2019; Wickham, 2011; Barrett *et al.*, 2024), and *doMC* and *parallel* enabled parallel processing (Revolution Analytics and Weston, 2019; R Core Team, 2024). *ggplot2*, *ggpubr*, *ggnewscale* and *cowplot* were used to visualise the results (Wickham, 2016; Kassambara, 2020a; Campitelli, 2020; Wilke, 2019). For statistical analyses, *vegan* and *rstatix* were used (Oksanen *et al.*, 2019; Kassambara, 2020b).

Maps were created with QGIS (v3.12) and a digital elevation model provided by Natural Resources Canada. Watersheds were delineated with ArcMap (v10.5.1, ESRI Inc., Redland, CA) and the Spatial Analyst Toolbox.

1.8.9 Additional supplementary figures

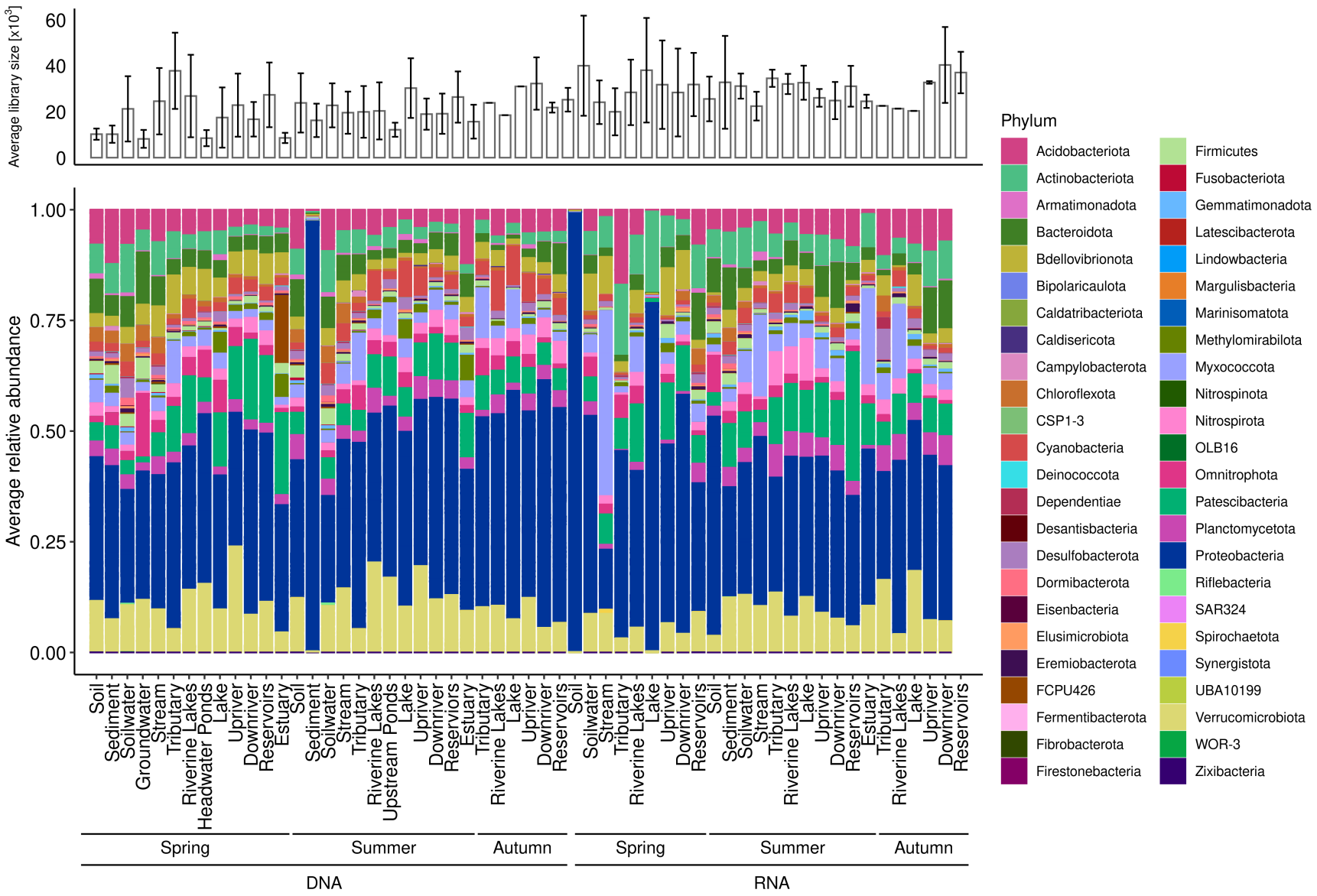

Figure S1.8: Taxonomic composition of habitat types. Given are averages of phyla found across habitat types. Upper panel shows average library sizes across habitat types. Error bars indicate the standard deviation from the mean.

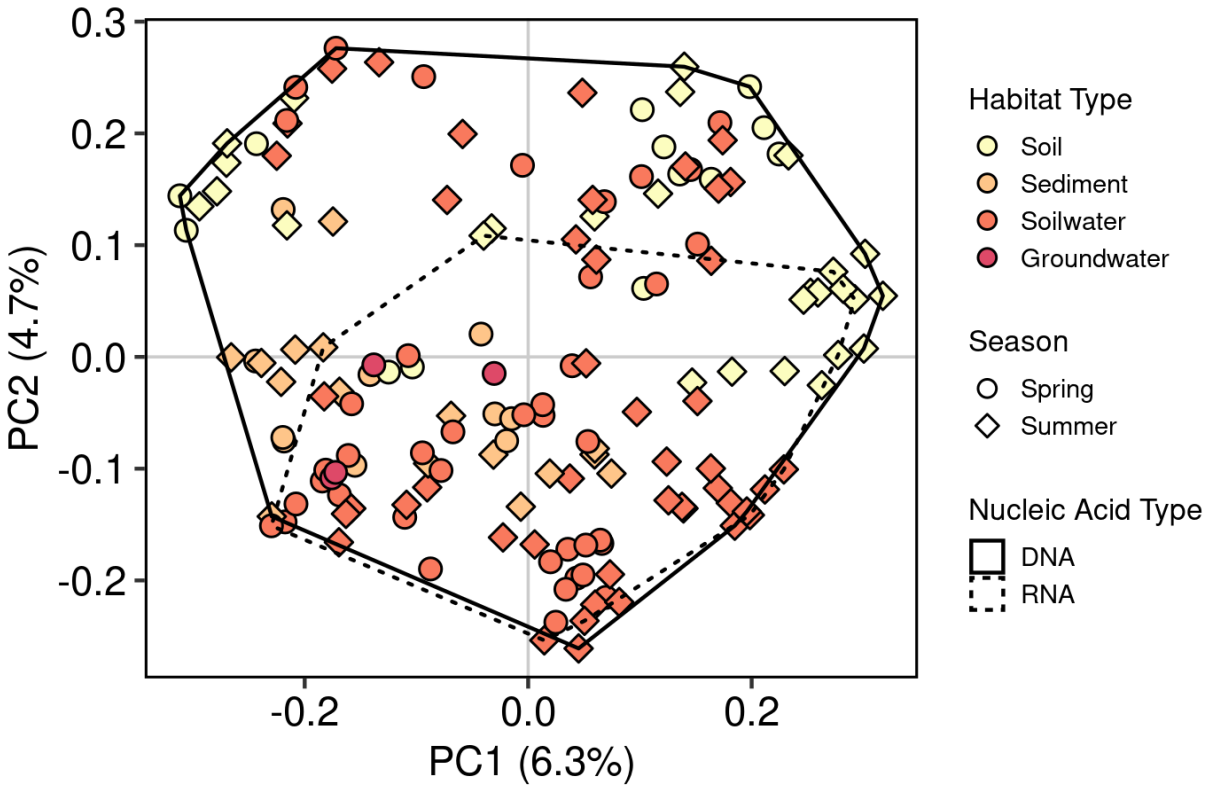

Figure S1.9: No clear seasonal clustering within terrestrial samples. PCoA analysis with square rooted Bray-Curtis dissimilarity on Hellinger transformed community matrix with only terrestrial samples (n = 156, 11,047 OTUs). Habitat types are distinguished by colour, seasons are indicated by shapes and nucleic acid type are visualised by different line colour.

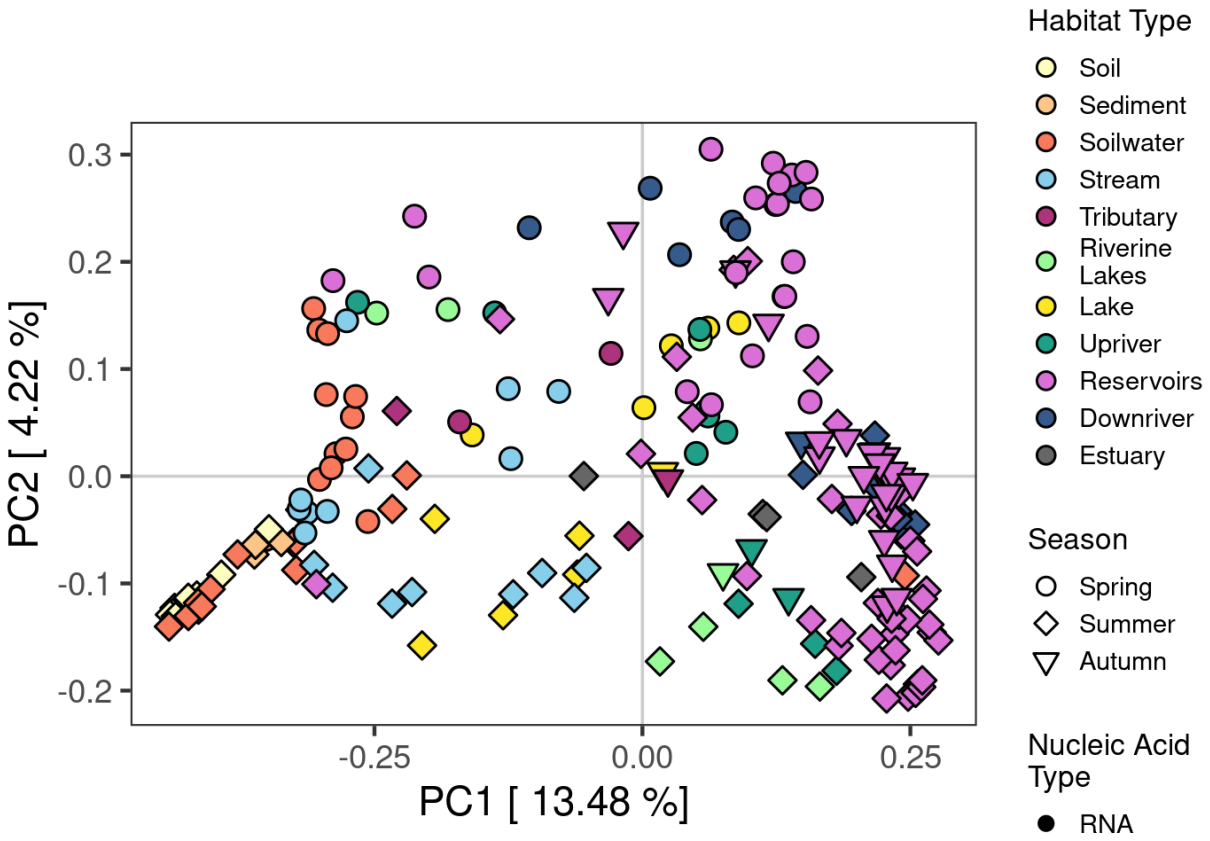

Figure S1.10: Habitat type and seasonal separation of RNA. PCoA analysis with square rooted Bray-Curtis dissimilarity on Hellinger transformed community matrix of all RNA samples (n = 201, 7 549 OTUs). Habitat types are distinguished by colour and seasons are indicated by shapes.

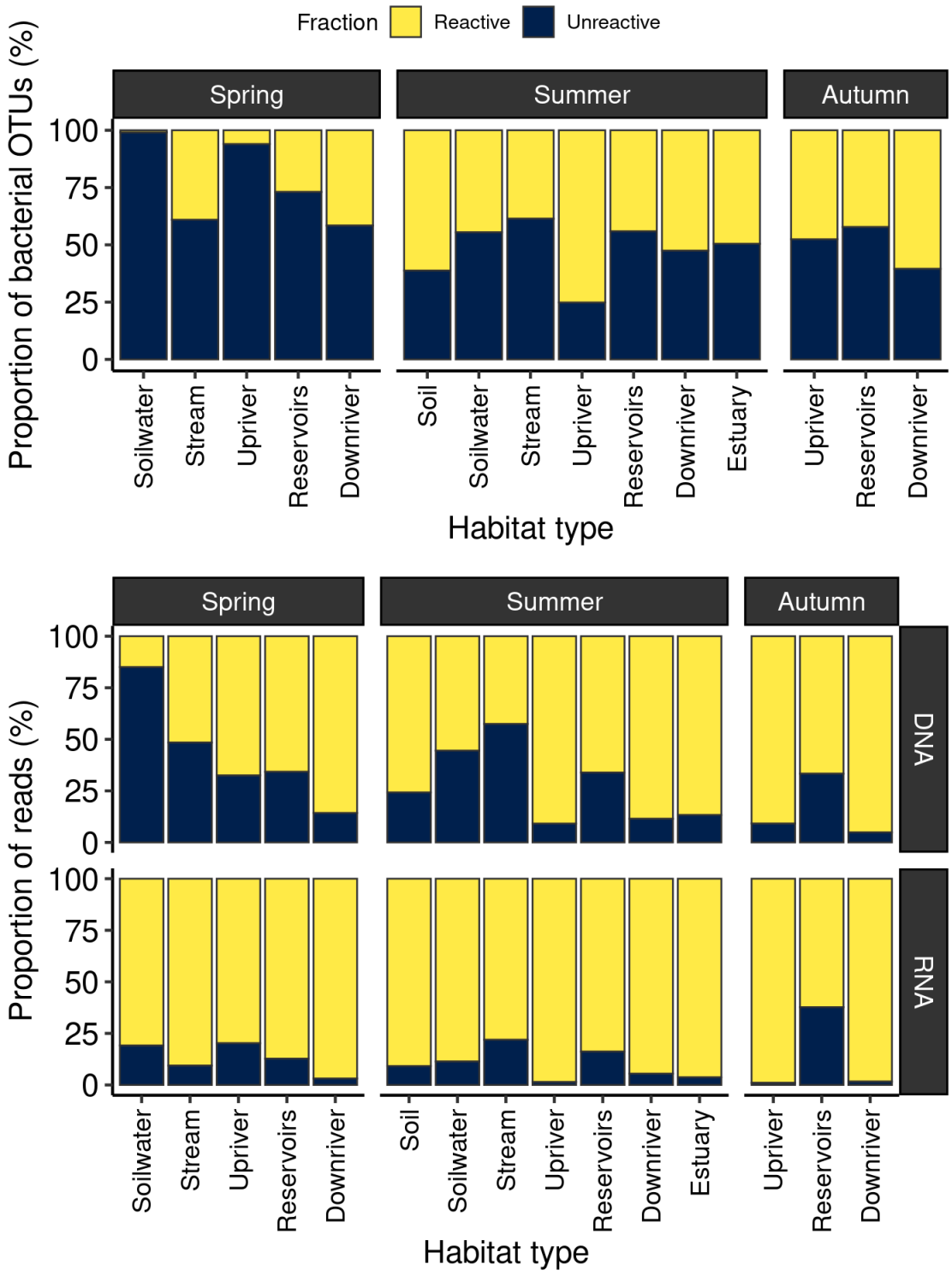

Figure S1.11: Unreactive and reactive fractions of OTUs. Proportion of bacterial OTUs (%) within each habitat type and season that can be attributed to the unreactive (RNA = 0 and < Median) and reactive (> Median) fraction.

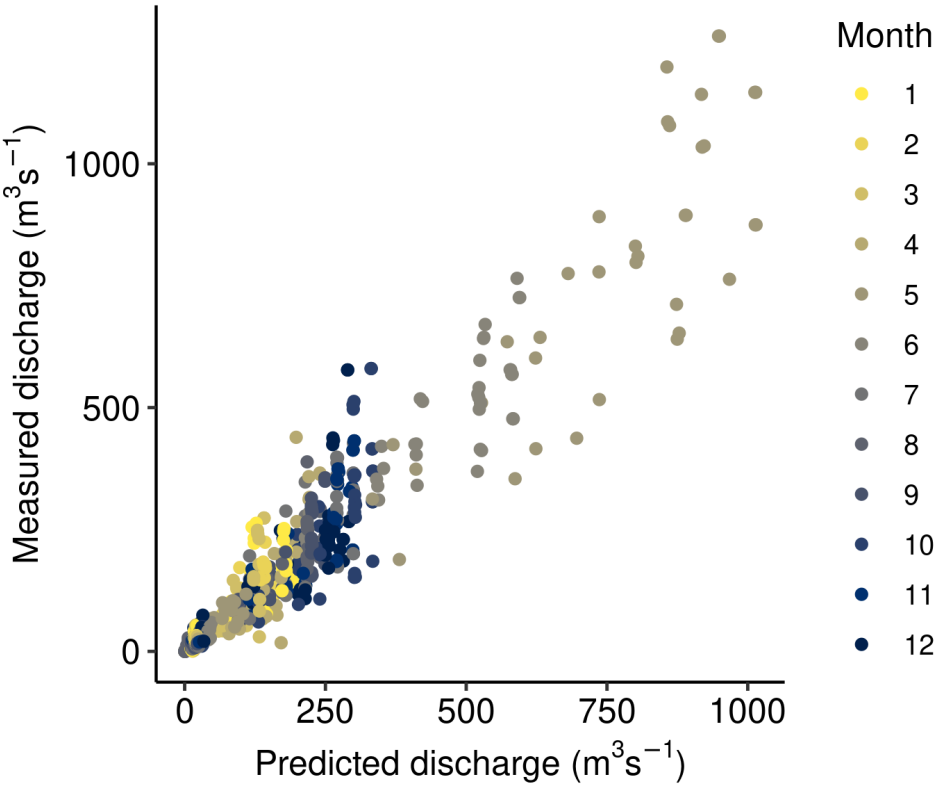

Figure S1.12: Modelled and measured discharge across the watershed. Discharge was continuously measured at point locations at various stream orders by Hydro-Québec. A model was created to model the discharge of all streams and rivers within the watershed using mean monthly air temperatures (weather station at Havre-Saint-Pierre airport), monthly difference in snow melt (retrieved from Brown and Brasnett (2010)) and flow accumulation (derived via GIS). This model is part of a separate manuscript in preparation, hence, the script and data are not available on Github.

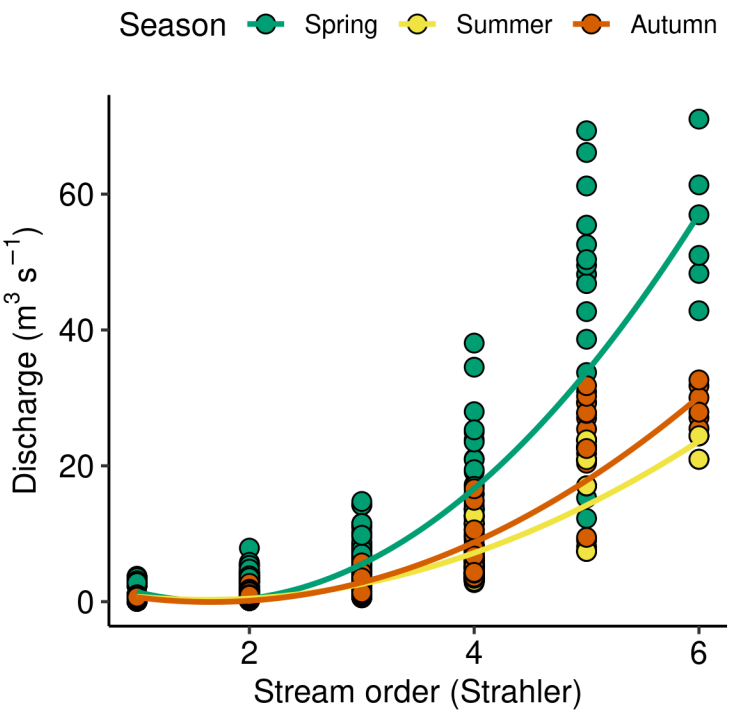

Figure S1.13: Seasonal differences in discharge across the watershed. Given are predicted discharges of all lotic systems within the watershed for the years 2015-2017. Spring, Summer and Autumn correspond to the months June, August and October, during which the field campaigns were carried out. Lines represent polynomial linear regressions. These results are part of a separate manuscript in preparation, hence, the script and data are not available on Github.

1.8.10 Supplementary tables

Table S1.1: Number of samples per habitat type.

Nucleic Acid Type	Soil	Sediment	Soilwater	Groundwater	Stream	Tributary	Riverine Lakes	Headwater Ponds	Lake	Upriver	Downriver	Reservoirs	Estuary	Sum
DNA	29	22	56	4	33	8	17	10	16	22	50	102	20	389
RNA	8	4	26	0	18	5	8	0	11	12	21	84	4	201

Table S1.2: Estimated seasonal variation in water residence time (d) across reservoirs. Water residence time is given in days. Water residence time was estimated from reservoir volume and continously measured discharge data at the reservoir outflow by Hydro-Québec. Monthly reservoir volume was estimated from a water level to volume relationship derived by Hydro-Québec. The underlying data are part of a separate manuscript in preparation, hence, the script and data are not available on Github.

Reservoir	Spring	Summer	Autumn
RO1	3.9	6.5	7.4
RO2	90.8	144.2	190.5
RO3	43.8	86.4	115.6

CHAPTER 2

UNRAVELING REACTIVE UNITS WITHIN MICROBIAL AND DISSOLVED ORGANIC MATTER ASSEMBLAGES ON A WATERSHED SCALE

Masumi Stadler[1], Trista J. Vick-Majors[2], François Guillemette[3] and Paul A. del Giorgio[1]

[1] Groupe de Recherche Interuniversitaire en Limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC, Canada.

[2] Department of Biological Sciences, Michigan Technological University, Houghton, Michigan, USA.

[3] Département des Sciences de l'Environnement, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

Keywords:

Dissolved organic matter, microbe-DOM interactions, terrestrial-aquatic continuum, spatial modelling, flow-weighted water age, correlation analysis

N.B. References cited in this chapter are presented at the end of the thesis.

2.1 Abstract

Aquatic microbial communities and dissolved organic matter (DOM) exhibit complex, bidirectional interactions within hydrological networks. Despite advances in high-resolution characterisation techniques, disentangling active versus seemingly passive components in these assemblages remains an on-going challenge. Here, we followed individual spatial patterns of microbial and DOM units along a large interconnected boreal watershed ($\sim$14,500 km^2) across two years and two seasons. We developed a novel framework using flow-weighted water age (FWWA) to model the movement of microbial and DOM components along a hydrological continuum encompassing soils, streams, rivers, reservoirs, and lakes. Our findings reveal distinct spatial units of microbial and DOM moieties that either increase, consistently decline, or exhibit local peaks in abundance with increasing water age. Overall, approximately 7.5% and 44% of microbial and DOM assemblages, respectively, were found to be statistically reactive within the watershed. Correlations between the reactive members in both DOM and microbial assemblages allowed us to identify a few potentially causal relationships, where fresh terrestrial compounds and mid-decomposed DOM were consumed in spring and summer, respectively. At the same time, decomposed DOM of different chemical signatures seemed to select for two distinct pools of microbial phyla in higher water age systems among the two seasons examined. This study provides new insights into the dynamic co-assembly of microbes and DOM, offering a framework for future research to better understand the underlying processes governing these interactions on a watershed scale.

2.2 Introduction

Heterotrophic microbes are both consumers, transformers and producers of dissolved organic matter (Guillemette and del Giorgio, 2012). This results in a complex, bi-directional relationship between aquatic microbial communities and dissolved organic matter, which has been the target of studies over several decades (Azam *et al.*, 1983; del Giorgio and Cole, 1998). Thousands of microbial taxa inhabit the world's aquatic ecosystems (Thompson *et al.*, 2017) and co-exist with an ever-changing DOM pool composed of thousands of molecules. It is thought that heterotrophic microorganisms have shared substrate preferences (Mou *et al.*, 2008), but individual taxa may also have unique preferences for certain DOM substrates (Lauro *et al.*, 2009). Therefore, the am-

bient DOM composition may directly influence which taxa within the microbial community can grow at a certain time and location (Muscarella *et al.*, 2019). At the same time, molecules are broken down and transformed by microorganisms, where by-products of degradation are often less degradable than their precursors (Jiao *et al.*, 2010) and persist in the DOM pool (Kellerman *et al.*, 2018; Zark and Dittmar, 2018). This constant processing and transforming of the DOM pool leads to molecular diversification and renders natural dissolved organic matter one of the most complex chemical mixtures on Earth. Hence, understanding the inherently complex relationship of microbes and DOM remains an ongoing challenge (Nelson and Wear, 2014).

High-resolution approaches held much promise to help resolve this intertwined relationship between microbes and DOM (Herlemann *et al.*, 2014; Logue *et al.*, 2016), and there is an increasing number of studies that attempt to link microbes and DOM using high-resolution approaches to characterise both assemblages. Yet, biogeochemists and microbial ecologists find themselves facing similar computational and conceptual challenges (Zinger *et al.*, 2012). An increased resolution does not necessarily provide more insights, particularly if we capture a large fraction of either passive or unreactive units. For example, it has been shown that a large fraction of microbial communities in soils and aquatic systems is not active, but rather dormant or dead (Lennon and Jones, 2011; Stadler and del Giorgio, 2022). Similarly, a large pool of molecules within ambient DOM has been observed to be remarkably persistent in time and space, and seemingly unreactive (Herlemann *et al.*, 2014; Kellerman *et al.*, 2015; Zark and Dittmar, 2018). This implies that within both assemblages, there is a substantial fraction that may not contribute to local microbe-DOM interactions. Most studies assessing microbe-DOM interactions have focused on linking bulk diversity patterns (Muscarella *et al.*, 2019; Tanentzap *et al.*, 2019; Kajan *et al.*, 2023), and only a handful of studies have considered assessing whether there are different associations between the bulk versus the active microbial fractions with the DOM pool (Osterholz *et al.*, 2016; Muscarella *et al.*, 2019). To date, however, none considered parsing out the reactive fractions in both microbial and DOM assemblages. It is therefore unsurprising that findings are often ambiguous and even contradictory, when the entirety of the two assemblages is used to explore interactions between microbes and DOM (Hall *et al.*, 2018).

In addition to the methodological and conceptual challenges of parsing out reactive moieties within both assemblages, one must consider the inherent connectivity of inland waters. Connectivity across realms (e.g. terrestrial versus aquatic) and within aquatic networks has been highlighted as a major driver of microbial (Mansour *et al.*, 2018; Ruiz-González *et al.*, 2015a) and DOM (Kothawala *et al.*, 2021) assembly. Ample evidence has been gathering that hydrology is a key driver of diversity and composition patterns of both DOM (Lynch *et al.*, 2019; Casas-Ruiz *et al.*, 2020; Ryan *et al.*, 2024) and microbial communities (Read *et al.*, 2015; Niño-García *et al.*, 2016b), and of DOM degradation potential (Catalán *et al.*, 2016; Attermeyer *et al.*, 2018; Peter *et al.*, 2020). There have been some studies following DOM (Hutchins *et al.*, 2017; Casas-Ruiz *et al.*, 2020; Peter *et al.*, 2020) and microbial communities (Savio *et al.*, 2015; Gweon *et al.*, 2020; Stadler and del Giorgio, 2022) along interconnected hydrologic networks at the watershed scale, but these have been carried out separately. High-resolution DOM and microbial data have been assessed together at the headwater portion of terrestrial-aquatic continua (Freeman *et al.*, 2024), and sections of large rivers (Kamjunke *et al.*, 2022) but never at a whole watershed scale. It is crucial to place these two complex assemblages within a framework that considers their co-movement along the entire hydrologic continuum, from soils, headwaters to downstream portions of the network in order characterise their spatial behaviours, identify passive versus reactive components, and unravel potential links that exist between them.

In this regard, it has been shown that a significant fraction of aquatic microbes, and DOM molecules are of allochthonous (terrestrial) origin (Boyer *et al.*, 1996; Crump *et al.*, 2012), and these allochthonous moieties are most prevalent in small order streams, where soil-to-water exchanges are most intense (Creed *et al.*, 2015; Ruiz-González *et al.*, 2015a). As these terrestrial sources decline or shift along the aquatic continuum, these allochthonous moieties may be degraded, diluted or simply persist in the case of DOM (McLaughlin and Kaplan, 2013; Creed *et al.*, 2015; Casas-Ruiz *et al.*, 2020), and inactivate, die or grow in the case of microbes (Ruiz-González *et al.*, 2015a; Stadler and del Giorgio, 2022). As the streams become larger rivers or even merge into lakes, water residence time consequently increases, favouring the selection and growth of aquatic organisms and the production of autochthonous organic material contributing to the emergence of novel molecules (Battin *et al.*, 2008; Sleighter and Hatcher, 2008; Casas-Ruiz *et al.*, 2020). All

these processes generate spatial patterns of microbes and molecules along the aquatic continuum, which can be characterised, quantified and linked. Some of these spatial patterns between microbial and DOM moieties may be causal. For example, some microbes consume a certain pool of DOM along the continuum, and hence, are being locally selected for as a function of the DOM pool that may be consumed (Findlay, 2003; Docherty *et al.*, 2006; Bambakidis *et al.*, 2024). Other relationships may not be causal, for example, microbes and DOM molecules may simply be correlated because they share the same source (e.g. terrestrial), or because they are merely travelling and being diluted together. The co-occurring spatial patterns of microbes and DOM molecules along complex aquatic landscapes have yet to be explored, despite the fundamental insights these patterns could provide into microbe and DOM co-assembly and interactions.

Here we present the spatial patterns of microbial and DOM molecular assembly along an interconnected aquatic network within a large boreal watershed ($\sim$14,500 km^2), encompassing soils, headwater streams, rivers, lakes and reservoirs, across multiple seasons and years. To address the challenge of integrating high-resolution DOM and microbial data, we developed a framework to model spatial patterns of microbes and molecules along a direct hydrologic continuum within the same watershed. To represent the hydrologic continuum in our modelling exercise, we computed flow-weighted water age (FWWA, adapted from flow-weighted travel time in Peter *et al.* (2020)) to approximate the average duration a molecule or microbe spent in the watershed to reach a certain point in the network. This modelling approach allowed us to distinguish between unreactive moieties and those showing significant shifts (positive or negative) in abundance along the continuum. We classified them into spatial units that increase (growth or production), consistently decline (death, degradation or dilution) with increasing water age, or exhibit local peaks or no discernible spatial pattern. We hypothesised that the microorganisms and molecules that fall within each of these spatial categories reflect similar ecological and biogeochemical properties, respectively. Additionally, we aimed to link microbial and DOM molecular spatial patterns to uncover potential relationships that might be overlooked when analysing bulk assemblages.

2.3 Material and methods

2.3.1 Study area and sampling

To follow the movement of bacteria and molecules along a terrestrial-aquatic continuum, a headwater stream network to reservoir continuum within La Romaine River watershed (Côte-Nord region, Québec, Canada) was sampled. Details of the catchment and network characteristics can be found in Stadler and del Giorgio (2022). In brief, La Romaine catchment is dominated by boreal coniferous forest, and the main river itself expands to approximately 475 km. The studied continuum starts in a sub-watershed of La Romaine River, where soil waters and streams were sampled (ranging Strahler orders 1-3). The headwater stream network merges into the main channel of the river, which subsequently flows into a reservoir (Romaine 2, storage reservoir, area (*A*) = 85.8 km^2, mean depth = 61 m, mean water residence time (WRT) = 141 d) which was flooded late 2014. A second reservoir downstream of Romaine 2 was flooded early 2016 (Romaine 1, run-of-the-river reservoir, *A* = 12.6 km^2, mean depth = 22 m, mean WRT = 6 d) (Rust *et al.*, 2022). Sampled lakes within the catchment had a surface area ranging from 0.1 km^2 to 19.2 km^2 and WRT from 113 d to 1601 d (Messager *et al.*, 2016). The main channel of La Romaine and multiple tributaries were sampled (Strahler orders 4-7). Field campaigns were conducted in 2015-2016 in two seasons each (June and August representing spring and summer and high and low flow, respectively) to capture variations in hydrology. Overall, 285 samples were collected for DNA, 125 for RNA and 172 for FT-ICR MS over the entire sampling period (Fig. 2.1a).

In a previous paper we provided details of the sampling procedures and general sample preparation for each sample type (soil / soil water versus stream / lake / reservoir) (Stadler and del Giorgio, 2022). In brief, DNA and RNA water samples were filtered onto a 0.22 μm polycarbonate filter and homogenised soils were stored in aliquots. RNA samples were submerged in RNAlater and Lifeguard Soil Preservation Solution (QIAGEN, Hilden, Germany) for water and non-water samples, respectively. RNA samples were stabilised overnight and were subsequently frozen. For Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) samples, water was filtered through a 0.45 μm polyethersulfone syringe filter (Sarstedt, Germany) and stored in 125 mL polycarbonate bottles. Bottles were frozen at -20°C immediately after filtration and were stored at the

same temperature until extraction.

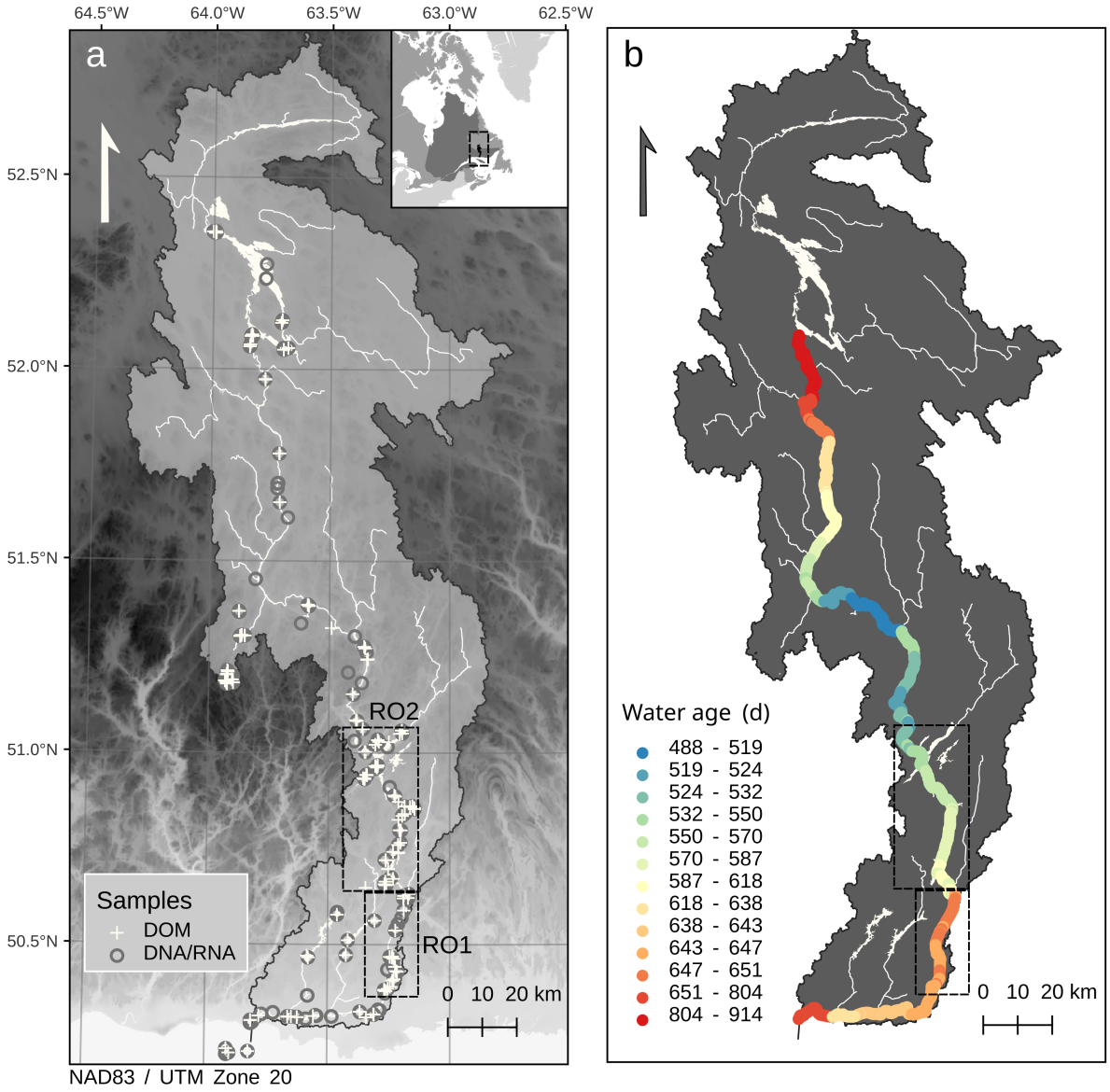

Figure 2.1: Estimated flow weighted water age along the main channel. a) Sampled locations for molecular and microbial samples within the watershed (Inserted figure: watershed location within Québec, Canada). b) Flow weighted water age estimated in August 2016 along the main river channel – La Romaine. Rectangles highlight reservoir positions within the watershed.

2.3.2 Sample processing

To determine bacterial abundance, samples were preserved on the same day of collection by adding para-formaldehyde (PFA) to a final concentration of 1% and glutaraldehyde (G) to 0.05%, then stored at -80°C until they were ready for analysis (del Giorgio *et al.*, 1996). Before analysis, the samples were thawed and stained with SYTO 13 (diluted in dimethyl sulfoxide (DMSO), 2.5 μM; Invitrogen, Waltham, MA, USA) at 0.025% of the sample volume. The stained samples were then analysed using an Accuri C6 flow cytometer (BD Bioscience, San José, CA, USA) at a flow rate of 14 L min^{-1}, utilising side scatter and green fluorescence (FL1-H) detection.

PowerWater and PowerSoil DNA and RNA extraction kits (MoBio, Carlsbad, CA, USA) were used to extract water and soil water samples, respectively. Subsequently, cDNA was synthesised with a high-capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). All samples were sent to Génome Québec Innovation Centre (Montréal, QC, Canada) for paired-end sequencing of the 16S rRNA V4 region using the primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'GGACTACHVGGGTWTCTAAT-3') on a MiSeq platform (PE250, Illumina, San Diego, CA, USA; details in Stadler and del Giorgio (2022).

For FT-ICR MS analyses, samples were acidified to pH 2 with HCl, and subsequently solid-phase extracted with 50 mL Bond Elut PPL columns (Agilent Technologies, Santa Clara, CA, USA; Dittmar *et al.* (2008)). Columns were rinsed with 0.01 N HCl three times the cartridge volume (3 mL) and dried with N_2 before DOM was eluted using methanol (final concentration 50 mg C L^{-1}). Extracts were stored at -20°C in amber glass vials (acid-washed, pre-combusted at 550°C for 5h) until analysis. Samples were analysed in a randomised order with a custom-built 9.4 tesla FT-ICR MS to measure ions produced in negative-ion mode by electrospray ionisation (ESI) at the National High Magnetic Field Laboratory (Tallahassee, FL, USA; Blakney *et al.* (2011); Kaiser *et al.* (2011).

2.3.3 Data processing

2.3.3.1 Bioinformatic processing

A detailed version of the bioinformatic treatment can be found in the supplementary material of Stadler and del Giorgio (2022). In brief, primers were removed from 16S rRNA DNA and cDNA (hereafter, RNA) reads using cutadapt (v1.18, Martin (2013)). Operational taxonomic units (OTUs) were identified by first processing the reads through the DADA2 pipeline to retrieve amplicon sequence variants (100% similarity; v1.14.1, Callahan *et al.* (2017). Paired ends were merged, chimera removed, and taxonomy was assigned using the IDTAXA algorithm (Murali *et al.*, 2018) and the provided trained classifier of the GTDB database (Release 95, Parks *et al.* (2018)) in the DECIPHER package (Wright, 2016). ASVs that were identified as anything else than bacteria were removed from subsequent analyses. Due to slight differences that may have emerged between DNA and RNA, and potential differences between 16S rRNA copies (Větrovský and Baldrian, 2013), ASVs were further merged into OTUs by a 99% similarity threshold. All bioinformatic processing was conducted in R (v3.6.3, R Core Team (2024)) and RStudio (RStudio Team, 2024). A phylogenetic tree was constructed for the entire La Romaine database (2015-2018) and was trimmed down to the subset used in this manuscript. First, each OTU was matched to 10 relatives using a 95% minimum identity threshold using the SINA software (v1.7.2, Pruesse *et al.* (2012)). Relatives with a 100% goodness of fit were kept for subsequent tree construction. 9,303 unique relatives were retained and used to construct a tree with RAxML-NG (v1.0.1, Kozlov *et al.* (2019)) using the GTR+G model. Sample sequences were aligned to the relatives' sequences using PaPaRa (v2.5, Berger and Stamatakis (2011)), aligned to the relative tree using EPA-NG (v0.3.8, Barbera *et al.* (2019)) and finally placed on the tree using GAPPA (v0.7.1, Czech *et al.* (2020)). Estimated 16S rRNA copy numbers for each OTU were retrieved from the rrnDB database (Stoddard *et al.*, 2017). Estimates were assigned based on their best match in their taxonomic identification.

2.3.3.2 FT-ICR MS processing

Samples from 2015 and 2016 were measured in separate runs, and hence, were analysed independently. Signals with magnitudes greater than 6σ of the root mean square baseline noise were assigned molecular formulas using an in-house software (EnviroOrg©™, Corilo (2015)) created at

the National High Magnetic Field Laboratory. The software applied an internal 'walking' calibration for each mass spectrum, following the approach by Savory *et al.* (2011). Only formulae with elemental compositions within the ranges $C_{1\text{-}45}H_{1\text{-}92}N_{0\text{-}4}O_{1\text{-}25}S_{0\text{-}2}$ were considered for assignment, with a mass error capped at 200 ppb. The modified aromaticity index (AI_{mod}) was computed to evaluate the degree of unsaturation based on the molecular formula (Koch and Dittmar, 2006). The nominal oxidation state of carbon (NOSC) was computed to assess the degree of energy that can be yielded from oxidising a C unit (Riedel *et al.*, 2012).

We aggregated molecular formulae into groups based on their chemical similarity relative to 8 core structural and chemical metrics, using a hierarchical clustering approach similar to Danczak *et al.* (2020). These metrics included mass (*mz*), number of C, H, O, N, and S atoms, AI_{mod} and NOSC that were extracted for each molecular formula. The metrics were scaled and used to calculate a distance matrix using Euclidean distances (*dist()* function, *stats* package, R Core Team (2024)) and subsequently used in hierarchical clustering (*hclust()* function, *stats* package, R Core Team (2024)). Five optimal clusters were retained after evaluation with the *fviz_nbclust()* function (*factoextra* package, Kassambara and Mundt (2020)), and by examining the characteristic distributions and visualising the categorisation in Van Krevelen space. The cluster order was re-categorised by descending median NOSC values. The cluster groups derived from hierarchical clustering were visually compared with traditional molecular formulae classification following Hawkes *et al.* (2020) in Van Krevelen space (Fig. 2.2).

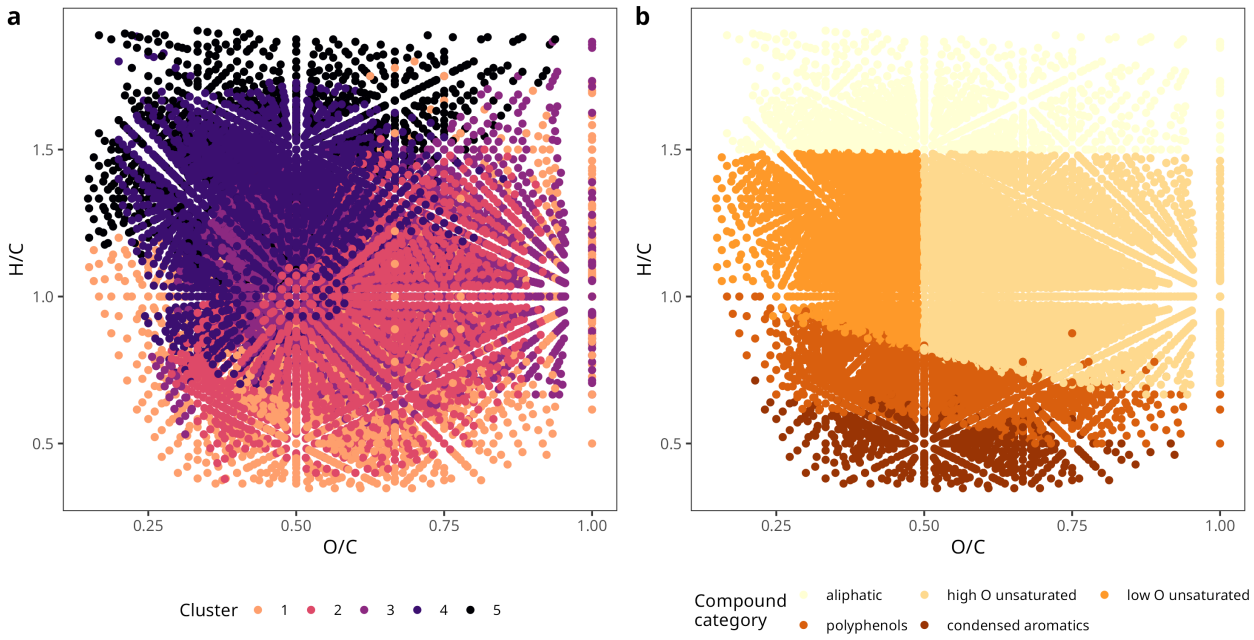

Figure 2.2: Comparison of molecular formulae categorisation methods in Van-Krevelen space. a) Five molecular formulae clusters emerged from hierarchical clustering on chemical attributes (i.e., mass, number of C, H, O, N, S atoms, AI_{mod} and NOSC). b) Molecular formulae were categorised into broad compound classes using the following criteria: aliphatics (H/C $\geq$ 1.5), low oxygen (O) unsaturated (H/C < 1.5, O/C < 0.5, AImod < 0.5) and high O unsaturated compounds (HC < 1.5, O/C $\geq$ 0.5, AI_{mod} < 0.5), polyphenols (0.5 < AI_{mod} < 0.67) and condensed aromatics (AI_{mod} $\geq$ 0.67) (Hawkes *et al.*, 2020). Broad classes were chosen to avoid over-interpretation of the limited structural resolution of the FT-ICR MS approach.

2.3.4 Flow-weighted water age

To visualise and study the spatial behaviours of microbes and molecules along a true hydrologic continuum, flow-weighted water age (FWWA) was estimated for the studied watershed. FWWA represents the average time water has travelled to arrive at any given point in the hydrologic network. A detailed description of how FWWA estimates were derived can be found in the supplementary methods (SM 2.8).

While flow-weighted travel time within stream networks have been estimated before (Peter *et al.*, 2020; Hosen *et al.*, 2021), lentic systems have been represented in a separate metric (i.e. upstream lake volume; Hosen *et al.* (2021)). To our knowledge there are no attempts to integrate the im-

pact of lentic systems on travel time. To derive a single metric that captures both lentic and lotic systems within a hydrologic network, we computed flow-weighted water age by first identifying which pixels were lentic or lotic systems by overlaying the HydroLAKES (Messager *et al.*, 2016) and reservoir (Hydro-Québec) polygons to the stream network (Fig. S2.1).

Water residence time in each pixel (WRT_{px}) of lotic systems was estimated by 1) estimating velocity by flow condition (high versus low; Spring versus Summer) using catchment area (km^2) and 2) dividing the length of the pixel by the estimated velocity in each pixel. The lotic velocity model was built on measured data from velocity measurements ranging Strahler orders 1-7. Although the R^2 was not particularly high (n = 103, R^2 = 0.46, Fig. S2.2b), the magnitude of the measured velocity as well as the relationship of velocity among Strahler orders found in the empirical data was captured with this model (Fig. S2.3). WRT_{px} within the reservoir were derived by 1) averaging daily WRT estimates for the entire reservoir for each month and 2) dividing the overall WRT by the number of pixels along the main channel of the reservoir. Similarly to reservoir estimates, the overall WRT in the lakes was divided by the number of pixels along the main channel flowing through the lake to estimate WRT_{px} (data retrieved from HydroLakes; Messager *et al.* (2016)). Both fluvial and reservoir WRT_{px} estimates reflect hydrologic changes in seasons, and only reservoir WRT_{px} additionally reflects changes by year. Additionally, discharge was modelled as a function of catchment area, using the same empirical dataset that was used to derive the velocity model (n = 103, R^2 = 0.88, estimated by flow condition, Fig. S2.2a). Details can be found in the supplementary material (SM 2.8).

A modified version of Peter *et al.* (2020) was implemented to compute flow-weighted water age (FWWA) along the stream network. FWWA within the same reach was calculated as follows:

$$FWWA_i = FWWA_{i-1} + WRT_i \times \frac{(1 - (Q_i - Q_{i-1}))}{Q_i}$$

where *i* indicates each pixel along a reach, *i-1* is the pixel that flows into *i*, WRT the estimated water residence time in the pixel in minutes, *Q* is the modeled discharge in $m_3\ s^{-1}$. First order streams

were first tracked downstream until a confluence was reached. Once a confluence was reached, FWWA was computed as follows:

$$FWWA_c = \frac{\sum_{i=1}^{n} FWWA_i \times Q_i}{\sum_{i}^{n} Q_i}$$

where *c* indicates each confluence, *n* the number of reaches merging into the confluence pixel, FWWA represents flow-weighted water age in minutes, *Q* is modeled discharge in m_3 s^{-1} and WRT is water residence time in the confluence pixel in minutes. Given the complexity of the stream network, an iterative loop was utilised to track down and calculate FWWA by Strahler order. A confluence was only processed once all merging reaches had a calculated FWWA.

Water age for soil and soil water samples was difficult to estimate and hence were assigned a value of -50 d. A negative value was selected to represent our conceptual view that aquatic continua start with surface water run-off and groundwater flowing from the surrounding terrestrial landscape into aquatic networks. Estimated water ages reflect differences by flow condition (high versus low flow), seasonal changes in reservoir WRT and the number of reservoirs present each year. Whole network FWWA at the river mouth in the St. Lawrence estuary was estimated at 752 and 755 days in spring and summer 2015, and as 756 and 819 days in spring and summer of 2016 (Fig. 2.1b, example of 2016 summer).

2.3.5 Modelling and classifying spatial patterns

A detailed version of the modelling exercise can be found in the supplementary material (SM 2.8). Prior to the modelling exercise, observations were filtered and z-scaled by campaign and binned and averaged at a 50-day interval along the water age gradient. To allow various dynamic patterns to be modelled, we established a decision tree that selects the best model for individual molecular formulae (MF) and operational taxonomic units (OTUs) along the FWWA gradient (illustrated in Fig. S2.4). The decision tree selects the best fitting model for each spatial behaviour from linear, and non-linear (polynomial versus GAM) options based on the smallest Akaike Information Criterion (AIC). Once the best model was selected, model statistics such as *p*-value and R^2 were

extracted. The slope of the initially fit linear model was extracted regardless of which model type was selected to aid in spatial pattern classification and null model comparison described below. For all non-linear models, the 2nd derivative was utilised to find peak locations along the FWWA gradient.

All linear models that had a positive slope were classified as 'increase', while all linear models with negative slopes were identified as 'decrease'. All non-linear models were first classified by their number of peaks. Models with a single peak were classified by where their peak was located (i.e., left, centre, right of FWWA gradient). If a peak was located within the centre area, the model was classified as 'unimodal'. All other single-peak models were classified as 'non-linear decrease' if their peak was located below the centre (right) and models with their peak located higher than the centre were classified as 'non-linear increase' (left). Any non-linear models with more than one peak were classified as multimodal increase and decrease when their linear model slope was positive and negative, respectively. Single-peak and multimodal spatial patterns were merged for all subsequent analyses. All models that had a slope of 0 or did not return a *p*-value due to insufficient sample size, were removed from downstream analyses.

2.3.6 Identifying reactive moieties

Most multiple comparison *p*-value correction methods focus on removing false positives (Type I error; Jafari and Ansari-Pour (2019)) but potentially increase the false negative rate (Type II error) and tend to exacerbate the removal of potentially valid observations (Lee, 2010). For this reason, we utilised a bootstrapping approach to evaluate which model results can be considered true positive and therefore reactive rather than relying on *p*-value correction methods. In theory, model outputs with a higher slope than slopes generated by a random reshuffling of observations indicate whether the spatial pattern is indeed distinct from random. Hence, for each OTU/MF, abundance values (reads and peak intensity, respectively) were shuffled over FWWA point observations 999 times per campaign (without replacement) and linear models were constructed for each randomisation. From the random models, linear slopes were extracted, and a 95% confidence interval was computed. To compare the randomisation model output to the empirical models, regardless of which model (linear versus non-linear) was selected as the best fit to the spatial pattern, the slope

from the first fitted linear model was always used. If an empirical model had a higher or smaller slope than the upper and lower limit of the randomisation confidence interval, respectively, it was deemed a true reactive pattern. To avoid any over-interpretation, we decided on a conservative method where only models that passed both the randomisation filter and were statistically significant (p-value < 0.05) were considered as 'reactive' spatial patterns.

2.3.7 Statistical analyses

To explore phylogenetic signals across the bulk and reactivity pools (reactive versus unreactive), UniFrac distances were computed on OTU matrices that were transformed into presence-absence data (*distance()* function, *phyloseq* package, McMurdie and Holmes (2013)). The sites were represented by combining the observations of OTUs by year, season and spatial pattern for the bulk pool (e.g., '2015 - Spring - decrease'), while the reactivity group (i.e., reactive versus unreactive) was added when additionally testing for phylogenetic differences among type of reactivity (e.g., '2015 - Spring - decrease - reactive'). A non-metric dimensional scaling (NMDS) was computed (k = 2) for the bulk dataset and when reactivity types were separated. An additional Principal Coordinates Analysis (PCoA) was computed on the reactivity type dataset alone (*ordinate()* function, *phyloseq* package, McMurdie and Holmes (2013)). To evaluate statistical differences in phylogenetic signal by spatial pattern, year, season and reactivity type, PERMANOVAs were computed with 9,999 permutations with the *adonis2()* function. PERMANOVA assumptions were tested with the *betadisper()* and *anova()* functions (*vegan* package, Oksanen *et al.* (2019)).

To further explore how the phylogenetic and chemical similarity signal was manifested within spatial patterns, the nearest taxon index (NTI) and net relatedness index (NRI) were computed using *ses.mntd()* and *ses.mpd()* functions, respectively, using the *picante* package (Kembel *et al.*, 2010). As a phylogenetic tree equivalent for DOM, the hierarchical clustering dendrogram was transformed into a tree (*as.phylo()* function; *ape* package, Paradis and Schliep (2018)). Both metrics are standardised against a null model. NTI is more sensitive to phylogenetic clustering or overdispersion occurring closer to the tips of a tree while NRI represents an overall pattern of phylogenetic structuring across the entire tree. Computations were conducted on OTU/MF matrices where year, season, spatial pattern and reactivity type combinations were treated as sites.

Statistical differences among phylogenetic and/or functional signals in spatial patterns were evaluated using Kruskal-Wallis Rank Sum tests (*kruskal.test()* function) and pair-wise differences were evaluated using Dunn's test (*dunn_test()* function, *rstatix* package; Kassambara (2020b)). An α level of 0.05 was chosen for all statistical analyses.

All analyses were conducted in R (v4.3.3, R Core Team (2024)) and RStudio (v4.2.764, RStudio Team (2024)). *tidyverse* was used for data cleaning, and unifying (Wickham *et al.*, 2019). *data.table* was used for data wrangling (Barrett *et al.*, 2024) and *plyr*, *foreach* and *doMC* were used to enable parallel processing (Wickham, 2011; Microsoft and Weston, 2022; Revolution Analytics and Weston, 2019). *ggplot2*, *ggpubr*, *plotly* and *iTOL* were used to visualise results (Wickham, 2016; Kassambara, 2020a; Sievert, 2020; Letunic and Bork, 2006). The map was created with QGIS (v3.28).

2.4 Results

Dissolved organic carbon (DOC) concentrations in La Romane river, Romaine 1 and 2 reservoirs and in the lakes within the watershed ranged from 5.5 to 7 mg C L^{-1}, and where higher in some of the tributaries, up to 10 mg C L^{-1} (Barbosa *et al.*, 2023). Bacterial abundance (ml^{-1}) was highest in soilwaters (4.7 x10^7 $\pm$ 8.6 x10^7; means $\pm$ standard deviation), followed by streams Strahler order 2 (1.6 x10^6 $\pm$ 1.1 x10^6). Rivers of orders 6 and 7, lakes and reservoirs had similar bacterial abundances between 1.4 x10^6 and 1.5 x10^6. Abundance in streams of orders 1 to 5 ranged between 1.0 and 1.2 x10^6. Bacterial density in aquatic samples were generally higher in summer (1.6 x10^6 $\pm$ 9.4 x10^5) over spring (1.3 x10^6 $\pm$ 5.5 x10^5).

The modelled flow-weighted water age (FWWA) represents the average time water spends from headwaters until it reaches any given point in the watershed, and it integrates the water age of the hydrologic elements (e.g., size and steepness of streams, depth, size, and number of lakes) that make up the sub-catchments. The FWWA ranged between 0.002 and 1,355 and 0.004 and 2,114 days for the sampled sites within the watershed in spring and summer, respectively. Stream FWWA increased along the Strahler order gradient with lowest FWWA in orders 1 (0.002 - 35.8 d), 2 (0.06 - 287.0 d), 3 (71.6 - 172.9 d), and higher values found in orders 4 (166.7 - 1,104.8 d), 5 (457.4 - 1,355.9 d), 6 (La Romaine upstream of reservoir cascade, 518.4 - 978.8 d) and 7 (La Romaine

downstream of reservoirs, 523.0 - 820.1 d). Systems with long water residence time, such as lakes and reservoirs, led to increased FWWA, with reservoir FWWA ranging between 522.5 and 726.3 days, and lake FWWA between 3.0 and 2,114.3 days. These patterns emerge due to the lower water velocities within lakes and reservoirs. The outlet of a lake and reservoir always reflects this longer water age through the system, yet as streams and rivers with low FWWA merge into these outlets further downstream, the FWWA starts to decrease once again. Hence, depending on the upstream history of each stream and river, streams of the same order can have very different FWWA.

Across the entire dataset 13,251 molecular formulae (MF) and 3,333 OTUs remained after cleaning and filtering procedures. We did not observe a clear seasonal pattern in the number of examined entities (richness) for each assemblage. For the microbial community, there were in general more OTUs in 2015 (2,135 $\pm$ 35) than 2016 (1,532 $\pm$ 25). Across sampling campaigns, the number of OTUs was mostly stable within years with no seasonal differences. In contrast, the DOM assemblage exhibited seasonal and inter-annual differences in the number of MF. While the MF count was stable within the year 2015 (7,375 $\pm$ 21), 2016 exhibited in general a higher number of MF (10,616 $\pm$ 3,469) and an exceptionally high number of MF in summer ($\sim$13,000). The number of unique OTUs and MF within each campaign followed in general the same trend as the overall number and ranged between 176 - 482 for OTUs, and 0 - 4,030 for MF (Table S2.1).

2.4.1 Characterisation of reactive microbial and DOM fractions

Overall, 12,386 spatial models were built for the microbial community with 9,259 and 3,126 DNA and RNA models, respectively. Out of these, 1,924 DNA and 466 RNA models were removed since they either did not report a *p*-value or their standard deviation was zero. We built 36,193 spatial models based for the ensemble of DOM molecular formulae, with only 211 models being removed due to the lack of a *p*-value.

Models that passed our slope randomisation filter and were significant ($p < 0.05$) were considered as 'reactive' spatial patterns. All other models that were not different from the randomisation slope or were not significant were categorised as 'unreactive' spatial patterns. Hence, the modelling exercise allowed us to apportion each assemblage, DOM as well as microbial, into reactive

and unreactive fractions. We first evaluated the proportion of the reactive pool within each assemblage. For the microbial community, there was no clear seasonal nor inter-annual pattern in the proportion of reactive moieties, which comprised between 13% to 17% of the total community. The proportion of reactive moieties within the RNA assemblage was similar at $\sim$13%. The DOM assemblage exhibited no clear seasonal trend of molecular reactive moieties, however, there was a clear inter-annual difference. Reactive MF consistently comprised $\sim$36% in spring, but there were substantial differences among the two years in summer, with reactive MF comprising 16% and 70% in 2015 and 2016, respectively. Across all years and seasons, the reactive fraction within microbial and DOM assemblages averaged 7.5 $\pm$ 1.8% and 44 $\pm$ 22.5%, respectively (Fig. 2.3a).

We then categorised and apportioned spatial patterns within the reactive pool. All models were categorised by their general spatial behaviour into increasing (linear or non-linear), unimodal, and decreasing (linear or non-linear) OTUs and MF (Fig. 2.3b). Within the reactive microbial fraction, there were proportionally more microbes that exhibited increasing (65.9 $\pm$ 3.1%) than decreasing patterns (28.5 $\pm$ 4.6%, Fig. 2.3c). There were no seasonal nor inter-annual differences, with proportions being remarkably stable. Likewise, there was no consistent seasonal DOM pattern in reactive pools, yet there were major inter-annual differences in these DOM spatial patterns: 2015 was dominated by decreasing MF (54.9 $\pm$ 11.0%), whereas 2016 was strongly dominated by increasing MF (75.1 $\pm$ 11.5%, Fig. 2.3c).

We examined whether the two studied years were different in their spatial range along the continuum to explore reasons underlying the observed differences in inter-annual patterns of MF spatial patterns. The spatial patterns in OTUs and MF that we identified were not constrained spatially and could occur in portions anywhere along the hydrological continuum. Hence, it is important to understand when and where along the network the spatial patterns preferentially occurred, and the network source of the moieties involved. We identified where the OTUs and MF first emerged (i.e., soil versus stream), and where they were last detected along the FWWA continuum (i.e., La Romaine river mouth versus lakes) by season and year. Larger lakes represent the endpoint of our hydrological water age continuum, since water age is reduced in fluvial systems beyond the lake outlet due to influx of smaller streams and rivers. Most spatial patterns of MF spanned the entire

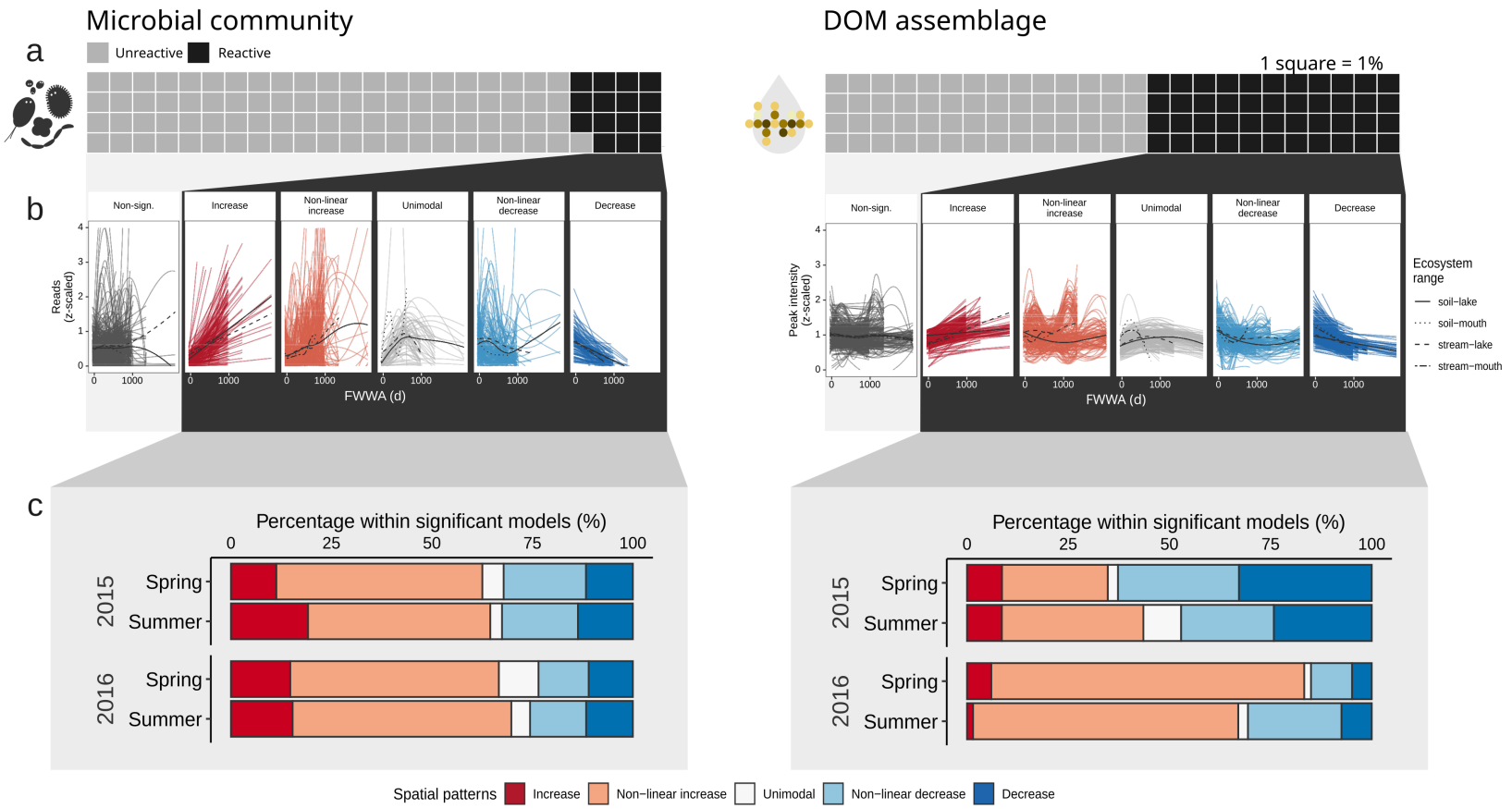

Figure 2.3: Fraction of unreactive and reactive moieties in microbial and DOM assemblages. a) Waffle plot illustrating the overall proportion of unreactive and reactive microbial (left) and molecular (right) moieties across the entire dataset. One square represents 1%. b) Examples of identified models within the unreactive fraction (grey) and reactive moieties (coloured) that were categorised as increase, non-linear increase, unimodal, non-linear decrease and decrease. Spatial patterns were modelled as a function of flow-weighted water age (FWWA) and scaled read numbers and peak intensity for microbes and molecular formulae, respectively. c) Seasonal and inter-annual proportion of identified spatial patterns within reactive moieties.

continuum from soils to lakes, except for Summer 2015 and Spring 2016 as no soil and soil water samples were retained after initial quality filtration (Fig. S2.5). Given that the spatial pattern distribution is very dissimilar between Summer 2015 and Spring 2016, with higher proportions of decreasing MF and increasing MF, respectively (Fig. 2.3c), it is unlikely that the absence of terrestrial samples in both campaigns is contributing to the observed trends. Overall, the majority (>90%) of MF could be retraced to soils in both spring and summer, and the majority of these MF were detected along the entire range of FWWA (95% in summer and 77% in spring), such that most of the spatial patterns ranged from soils to larger lakes (Fig. S2.5). For the microbial assemblage the spatial range was very consistent between years, with 75% - 97% of DNA and RNA spatial ranges originating in soils regardless of season and year, yet the endpoint of detection within the network varied greatly. A greater proportion of OTUs ($\sim$60%) were only found until the river mouth especially in summer, indicating that not all OTUs are present in lakes. These results collectively indicate that microbes are patchier in their spatial distribution, while molecular formulae are more continuous along the network. Regardless, both assemblages could be overwhelmingly traced back to soils.

2.4.2 Phylogenetic and functional underpinning of microbial spatial patterns

To evaluate whether there was a phylogenetic signal underlying the spatial patterns, a heatmap representing the reactive spatial patterns (i.e., increase, unimodal, and decrease) was plotted onto the microbial phylogenetic tree (Fig. 2.4). It was visually evident that many microbial OTUs had very few reactive representatives (empty spaces), however, those that were identified as reactive, generally seemed to have the same spatial pattern across campaigns and between DNA and RNA. To directly test whether there is phylogenetic signal in the distribution of spatial patterns, UniFrac distances were computed within two OTU matrices: bulk relatedness and the relatedness among reactivity groups (i.e., reactive versus non-reactive). Within the bulk pool, a strong phylogenetic signal among spatial patterns was found (Fig. S2.6a), which was statistically supported by a PERMANOVA analysis (F_6 = 2.03, R^2 = 0.37, p < 0.0001). However, there was no phylogenetic signal by season or year (p > 0.8). Decreasing and increasing linear OTUs were closely related, while non-linear spatial patterns started to exhibit phylogenetic divergence with unimodal OTUs being

most phylogenetically dispersed. When we considered the reactive and non-reactive fractions in a separate analysis, it was also found that there was a phylogenetic signal between spatial patterns ($F_6 = 1.75$, $R^2 = 0.16$, $p < 0.0001$) and that they also differed by reactivity groups ($F_1 = 5.53$, $R^2 = 0.09$, $p < 0.0001$). Similarly to the bulk pool, no seasonal or annual pattern was found ($p > 0.7$) (visually represented in a NMDS and PCoA in Fig. S2.6b-c, respectively). Within the phylogenetic multivariate space, linear patterns of the unreactive fraction were clustered, and the unreactive non-linear patterns were phylogenetically similar to the linear reactive spatial patterns. There was a general trend across the reactive fraction, where increasing spatial patterns were more phylogenetically similar to each other, while decreasing patterns were more dispersed in multivariate space (Fig. S2.6b). Phylogenetic clustering of unreactive non-linear patterns with reactive patterns may indicate that our modelling and filtering approach used to determine the reactive fractions may be conservative. Non-linear models may not be statistically significant due to their highly dynamic nature, and hence, a greater sample size and deeper sequencing depth may be needed to detect more reads for these OTUs to achieve 'reactiveness' as defined in our approach.

We computed indices of phylogenetic relatedness by spatial pattern to assess whether some spatial groups are more phylogenetically constrained than others. Tests were carried on the nearest taxon index (NTI) and net relatedness index (NRI) for both the bulk and reactive pool separately, to examine whether parsing out the reactive portion helps in identifying a phylogenetic signal (Kruskal-Wallis Rank Sum Test). Only NRI was significantly different among spatial patterns within the reactive pool (NRI: *d.f.* = 2, $\chi^2 = 7.42$, $p < 0.05$; NTI: *d.f.* = 2, $\chi^2 = 5.69$, $p = 0.058$). Dunn's test revealed that decreasers were statistically more over-dispersed than increasers and unimodals ($p < 0.05$, adjusted via Benjamini-Hochberg). In contrast, there were no phylogenetic differences among spatial patterns within the bulk pool (NTI: *d.f.* = 2, $\chi^2 = 2$, $p = 0.37$; NRI: *d.f.* = 2, $\chi^2 = 3.5$, $p = 0.17$) (Fig. 2.5a, data of NTI not shown). These results suggest that identification of the reactive pool on the basis of spatial patterns enhances our insight or understanding of the ecological and phylogenetic basis of microbial community assembly, which is not forthcoming from the analysis of the bulk community.

Furthermore, we focused on one specific functional trait - the estimated number of 16S rRNA copy

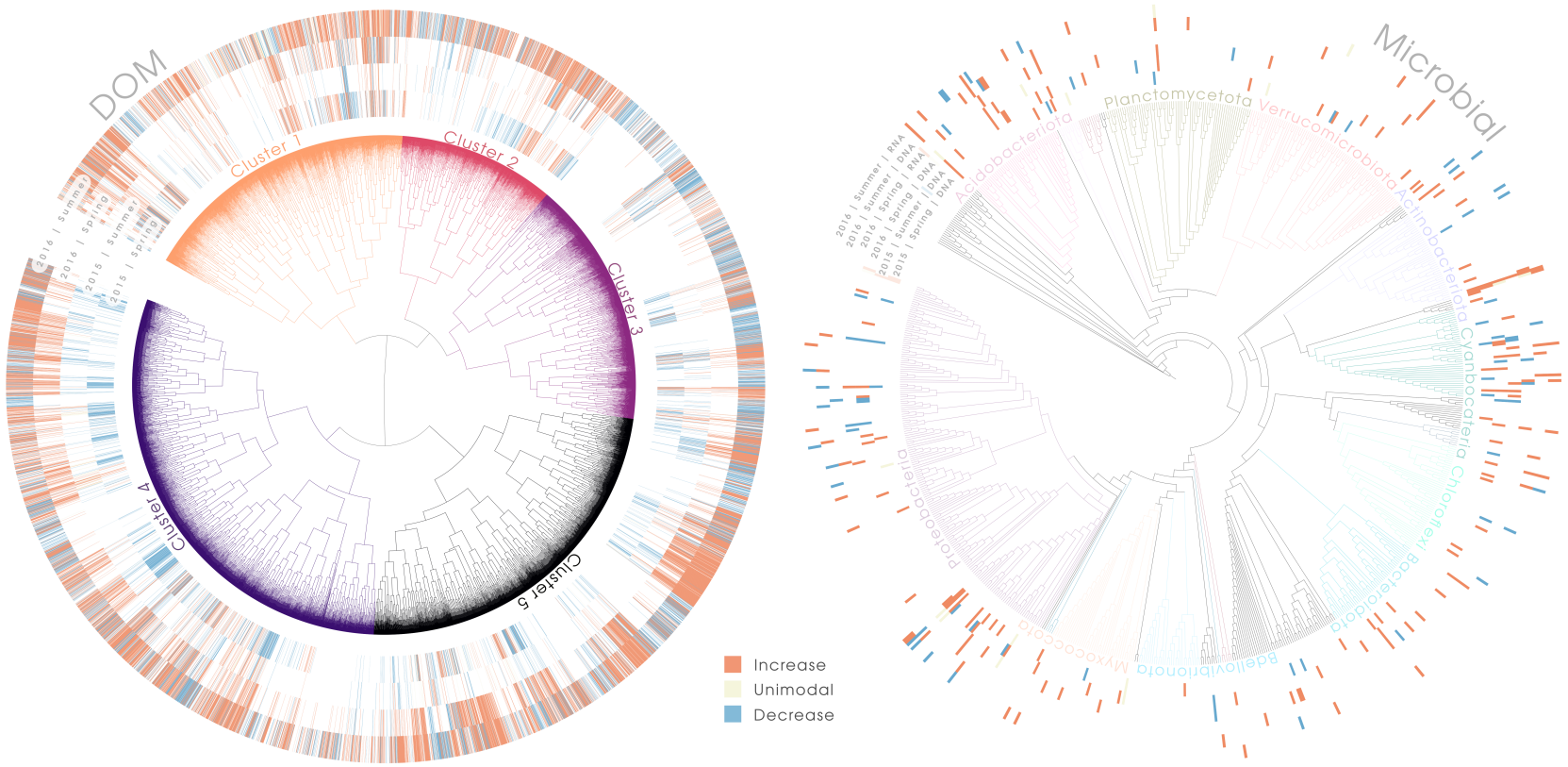

Figure 2.4: Heatmap of reactive spatial patterns on molecular and microbial tree. Left) Molecular tree derived from hierarchical clustering coloured by the identified 4 clusters. Heatmap rings represent 2015 Spring, 2015 Summer, 2016 Spring to 2016 Summer from inner to outer most ring. Right) Microbial phylogenetic tree coloured by phyla. Heatmap rings indicate the same temporal order as DOM tree, however, additional rings representing RNA for each campaign are added after DNA for 2016. Any spatial pattern that was identified as unreactive in a specific campaign is left blank.

number for each OTU – to examine whether the observed phylogenetic clustering coincided with functional differences among the same spatial groups and pools. However, no statistical differences were found among spatial patterns either within the reactive (*d.f.* = 2, χ^2 = 2.98, *p* = 0.22) or bulk pools (*d.f.* = 2, χ^2 = 1.17, *p* = 0.56) (Fig. 2.5b). These results indicate that the capacity of a given OTU to grow and/or accumulate or persist along the continuum does not depend on the number of 16S rRNA copy numbers.

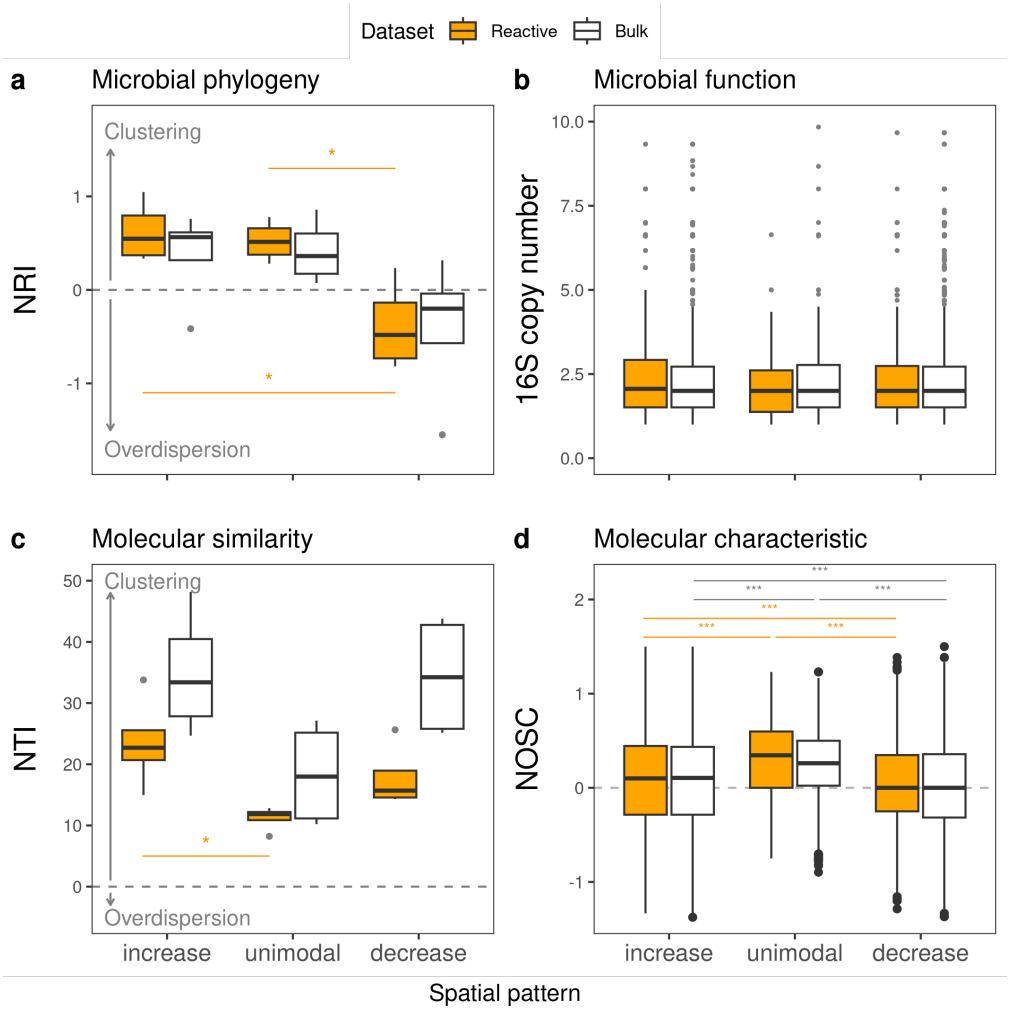

Figure 2.5: Differences in microbial phylogeny, molecular similarity and functional characteristics between spatial groups. a) Differences in the Nearest relative index (NRI) for microbial communities between spatial patterns of the reactive and bulk pool. b) Differences in 16S copy number among spatial patterns. c) Nearest Taxon Index (NTI) for molecular DOM assemblages based on a hierarchical clustering dendrogram. Positive NRI and NTI indicate phylogenetic clustering of OTUs and MF on the tree, while negative values represent overdispersion. NTI is more sensitive to patterns closer to the tips of the tree, while NRI represents structuring across the entire branches. d) Differences in nominal oxidation state of carbon (NOSC) across spatial patterns and pools. NOSC values above 0 indicate higher energy yield when a molecule is broken down, while low values indicate the need of additional energy to degrade a compound. The number of asterisks increase with lower *p*-values (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$) according to pair-wise comparisons with Dunn's tests. Statistical tests were only conducted to test differences between spatial patterns within the same pool (i.e. reactive and bulk).

2.4.3 Chemical and functional similarity underlying molecular spatial patterns

We used chemical similarity of DOM formulae to build a hierarchical clustering of MF, as an analogous analysis of microbial phylogenetic similarity. The unsupervised clustering technique revealed five functional groups that were characterised by different mass, aromaticity (AI_{mod}) and metabolic potential (NOSC). In essence, the clustering order (1-5) represents an ascending gradient for H/C and descending gradient for O/C, AI_{mod} and NOSC (Fig. S2.7, Table S2.2). All clusters were significantly different from each other in median H/C, O/C, C/N, AI_{mod}, NOSC and in the number of C, H, O atoms. Clusters 1 and 5, and cluster 2 and 4 were not different from each other in mass. Clusters 1 and 2 had no statistical difference in the number of N. The retrieved chemical dendrogram from the hierarchical clustering approach was plotted together with the reactive spatial pattern heatmap (Fig. 2.4) to visualise the distribution of MF spatial patterns by chemical similarity. In contrast to the microbial heatmap, many more MF exhibited a reactive spatial pattern, yet the direction of the pattern was often not consistent for a given MF and varied mostly by year.

To similarly test whether spatial patterns in MF were associated to chemical similarity, we applied the same phylogenetic indices (NTI and NRI) to the chemical dendrogram and tested whether there was a difference by spatial pattern within bulk and reactive pools (Kruskal-Wallis Rank Sum test). No difference was found in how the spatial patterns were spread across the dendrogram within the bulk pool (NTI: *d.f.* = 2, χ^2 = 4.9, p = 0.09; NRI: *d.f.* = 2, χ^2 = 3.5, p = 0.17). However, within the reactive pool a statistical difference was found for NTI (*d.f.* = 2, χ^2 = 8.0, $p < 0.05$) but not for NRI (*d.f.* = 2, χ^2 = 1.0, p = 0.59). For NTI, Dunn's test revealed that increasers were more clustered than unimodal patterns ($p < 0.05$, adjusted via Benjamini-Hochberg) (Fig. 2.5c). It is noteworthy that in general all spatial groups were clustered (NTI > 0) rather than overdispersed, with the least clustering observed in unimodal MF.

We further tested whether the spatial patterns were associated with functional differences in their nominal oxidation state of carbon (NOSC) and aromaticity index. There were statistical differences in both metrics among spatial patterns (KW-test, bulk: *d.f.* = 2, χ^2 = 355.58, $p < 0.0001$; reactive: *d.f.* = 2, χ^2 = 114.1, $p < 0.0001$) regardless of the examined pool ($p < 0.0001$, Fig. 2.5d). NOSC increased from decreasing (0.02 $\pm$ 0.48, mean $\pm$ SD), to increasing (0.08 $\pm$ 0.48) to unimodal

MF (0.23 $\pm$ 0.39). The spatial patterns were also significantly different in their aromaticity regardless of the reactivity pool (p < 0.0001), with AI_{mod} increasing from decreasing (0.28 $\pm$ 0.18), to increasing (0.32 $\pm$ 0.20) to unimodal (0.39 $\pm$ 0.20) MF. These results indicate that there are strong functional differences among spatial patterns even when the bulk pool is considered.

2.4.4 Linking spatial patterns in DOM molecular formulae and microbial OTUs

To evaluate individual spatial relationships of DOM MF and microbial OTUs, Spearman's correlation analyses were conducted along the FWWA continuum. Microbial OTU cumulative-sum scaled reads and molecular formulae peak intensities were z-scaled, observations were binned in a 50-day interval along the water age gradient and averaged before correlation analysis was carried out by campaign. Overall, 101,280,418 correlations were returned across the 4 campaigns. Correlations were filtered by whether a p-value was returned (no p-value was returned when there were not enough observations) and correlation coefficients of 0 or 1 were removed. A final dataset with 54,472,074 correlations was used for downstream analyses. p-values were not corrected for multiple comparisons since a correction only left correlations with extremely high correlation coefficients ($|\rho > 0.9|$), likely resulting in the exclusion of many valid cases. Of the 4,265,211 and 11,979,127 correlations that were retained in Spring and Summer 2015, only 4.9% and 5.6% respectively were significant (p < 0.05). Likewise, of the 13,084,715 and 25,143,021 correlations retained for Spring and Summer 2016, only 4.6% and 7.8% respectively were significant. In theory, all correlations between increasing OTUs and MF, and decreasing OTUs and MF should be positive. Likewise, all correlations between decreasing MF and increasing OTUs and vice versa should be negative. We further assessed whether the sign of the individual correlation coefficients aligned with the direction of the spatial patterns of the OTU and MF involved in the correlation, by identifying how many positive and negative significant correlations were observed for each spatial pattern combination of OTU and MF across the bulk and reactive dataset (Fig. S2.8).

Overall, we found coherence in the sign of the correlation and the combination of OTU and MF spatial pattern, with over 68% (bulk) and 80% (reactive) of coefficients having the expected sign (Fig. S2.8). The combinations that involved linear patterns of OTU and MF yielded consistently higher proportions of coherent correlations than the non-linear spatial patterns, particularly those

involving non-linearly increasing MF. This would suggest that a higher proportion of these non-linear spatial patterns may be less informative. Incoherent relationships may arise when MF and OTUs only coincide within limited portions along the continuum. For example, a non-linearly increasing MF may be negatively correlated to an increasing OTU when they only match within the range where the non-linear pattern is temporarily declining. Importantly, the reactive pools of both OTUs and MF yielded correlations that were consistently more coherent with the expected sign than the bulk pools, with proportions of expected signs among linear combinations exceeding 90% and with an improvement in the proportion of correct relationships in 19 out of 25 spatial pattern combinations (Fig. S2.8).

2.4.5 DOM properties underlying the spatial correlation between DOM formulae and microbial OTUs

In the previous section we showed that defining and extracting the most reactive microbial and molecular moieties results in the reduction of spurious patterns and in an improvement in our capacity to distinguish coherent relationships between DOM MF and microbial assemblages. Four correlation categories were identified: increasing OTU x increasing MF (correlation category 1), increasing OTU x decreasing MF (correlation category 2), decreasing OTU x increasing MF (correlation category 3), decreasing OTU x decreasing MF (correlation category 4). We further explored if the correlations among reactive moieties showed any biogeochemically coherent patterns in terms of distribution of DOM properties (Fig. 2.6). We calculated the proportion of significant correlations within each correlation category that involved MF in each of the 5 DOM clusters that we had previously defined. To simplify this analysis, we removed any correlations involving unimodal patterns, and those correlations that did not match our expected correlation signs, and we only focused on statistically significant correlations.

Unlike the proportion in spatial patterns, which did not show any clear seasonality (Fig. 2.2c), the relative contribution of DOM clusters to the total number of significant correlations by correlation category did (Fig. 2.6). The four correlation categories between MF and OTUs appear to be preferentially associated to specific DOM clusters, but these associations varied seasonally and inter-annually. Overall, the patterns of association were different between spring and summer,

and inter-annual differences were especially evident in spring. In spring 2015, an overwhelming dominance of Cluster 1 on all four correlation categories was observed ($\sim$66%), however, overall dominance shifted to Cluster 2 in 2016 ($\sim$42%). This pattern of dominance of a few clusters in spring was systematically different from the pattern observed in summer. There is remarkable consistency in the summer patterns across years, where contributions of Clusters 2, 3 and 4 are much higher in correlation categories involving declining MF. In contrast, correlation categories involving increasing MF showed larger contributions of Cluster 1, 2 and depending on the year, cluster 4. Despite the observed consistency in the patterns of contribution of specific clusters across seasons, it is also clear that the microbial-DOM links were very dynamic – and there were inter-annual shifts in the contribution of certain clusters. For example, cluster 2 and 4 are almost non-existent in spring 2015, however, they contribute largely to the correlation categories in spring 2016 (42 and 36%, respectively). These results indicate that there are seasonal consistencies as well as inter-annual differences in the pools involved in microbial-DOM interactions that may be linked to hydrology, climate and watershed shifts (i.e., number/age of reservoirs in watershed).

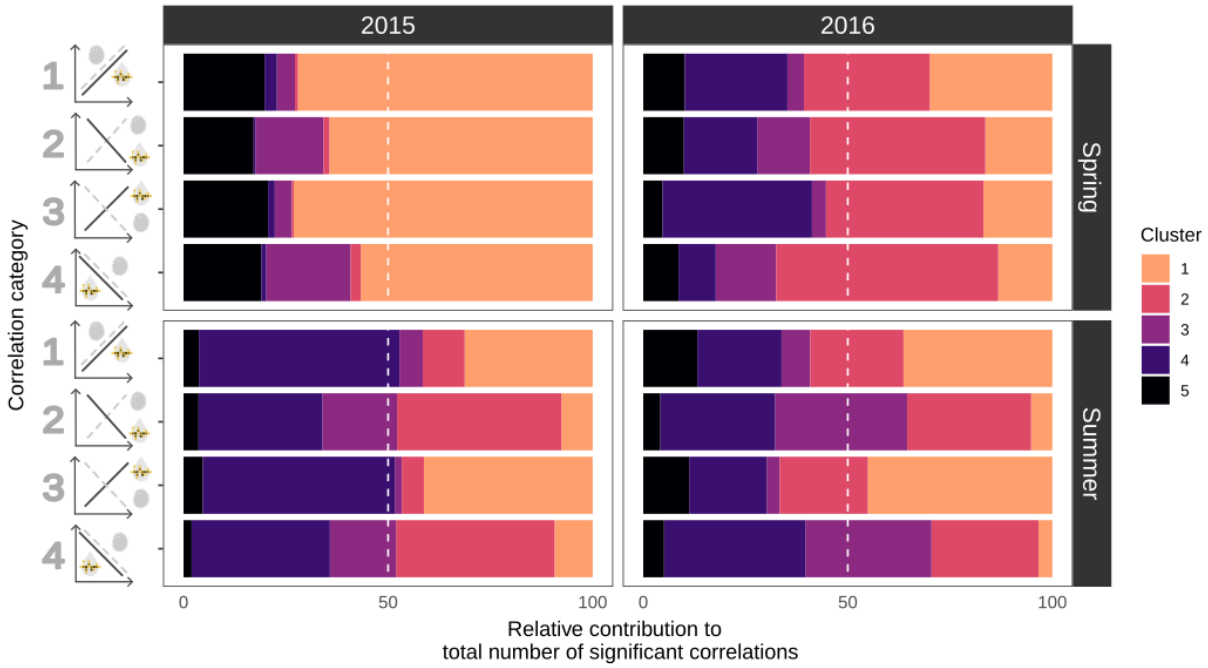

Figure 2.6: Proportion of significant correlations attributed to DOM clusters by correlation category. Proportions were computed only for significant correlations between reactive MF and OTUs. Unimodal spatial patterns and correlation relationships not consistent with our theoretical interpretation were excluded. Non-linear and linear patterns were merged. Colours represent DOM clusters identified via hierarchical clustering.

2.4.6 Beyond correlations: Using phyla-DOM associations to interpret the spatial correlations between DOM compounds and microbial OTUs

After having quantified the contribution of each DOM cluster to the four correlation categories across seasons, we further evaluated whether the observed contributions were shared across all phyla or whether there were disproportionately stronger relations between specific phyla and DOM clusters within certain correlation categories and seasons. To do this, we extracted the number of significant correlations for each phylum within each correlation category (and for the different sampling periods) and computed the proportion of these correlations attributable to each of the five DOM clusters (Fig. 2.7). As described in the previous section, we excluded OTUs and MF that had an unimodal spatial pattern, kept only significant correlations that matched the expected sign, and focused only on the reactive pool. We then computed a bootstrapped average and con-

fidence interval for the contribution of each DOM cluster to the ensemble of phyla across the four correlation categories. We used deviations from these confidence intervals to identify significantly higher contributions of DOM clusters to correlations with specific phyla within the four correlation categories (Fig. 2.7). Additionally, we identified whether these disproportional contributions occurred in spring or summer.

Figure 2.7 can be read horizontally as well as vertically, and both directions provide different perspectives on the relationship between phyla, DOM clusters and correlation categories. Horizontally, the figure provides similar information to what we extracted in figure 2.6, however, it additionally provides insights into which phyla are disproportionality related to each DOM cluster within each correlation category. Vertically, each column identifies significantly stronger relations between a DOM cluster and specific phyla, and how these relationships vary between correlation categories. Contrasting the contribution of each DOM cluster across each correlation category where above-average links were observed, enabled us to infer which DOM compounds are potentially merely correlated to microbes due to shared hydrology and origins, versus which groups were more strongly aligned to potential causal relationships such as microbial degradation as well as microbial production.

When the figure is examined vertically, it is clear that there are very different contribution patterns across DOM clusters to each of the four correlation categories. At one extreme, cluster 2 contributes abundantly to correlation category 4 (declining MF and OTUs) in spring but barely contributes to any other correlation category. This correlation category does not have a straightforward causal interpretation, unless it reflects the loss of microbially-mediated production of the DOM compounds involved. Yet, this cluster does not contribute to correlation category 1 (increasing DOM and OTUs), which would have supported the previous result's theory that the DOM is being produced by microbes. Together, these results suggest that this cluster is unlikely to be produced by microbes but is rather merely correlating to phyla that have a similar pattern of decay along the flow path, without any casual underpinnings.

Cluster 4 contributes mostly to correlation categories 1 and 3, both of which involve increasing

spatial patterns of MF, albeit with slightly different phyla. This pattern could be interpreted as this DOM pool positively selecting for specific phyla or being produced by the phyla that are increasing (correlation category 1). It also may indicate that the same DOM cluster is involved in correlations where the MF are being produced by phyla that are decaying along the continuum (correlation category 3). In both cases, there is the possibility of a causal underpinning.

Cluster 3 contributes abundantly to correlation category 2 (declining MF and increasing OTUs) with links to a wide range of phyla, and this would suggest that this DOM component might be consumed by a wide range of microbes and contribute to their growth along the continuum. This cluster secondarily contributes to correlation category 4 (declining MF and OTUs) with links to a subset of phyla that decline along the continuum, suggesting that this subset of relationships of cluster 3 with declining microbes might be mostly correlational as was the case for cluster 2. It is important to note that both cluster 3 and 4 contributed to correlations mostly in summer, suggesting that these DOM components are associated to DOM present at higher temperatures and low flow conditions.

Cluster 5 contributes to all correlation categories in spring but more noticeably to categories 1 and 4, which always involve co-directional spatial patterns in MF and OTUs (increasing MF and OTU; decreasing MF and OTU). Yet, across the two correlation categories the phyla involved are mostly different. If the phyla involved across the two correlation categories were the same, it may have indicated that these MF are produced by microbes as they increase with microbial growth and decline when the corresponding microbes disappear. Since the phyla involved are different, it may imply that the MF involved in correlation category 1 lead to the positive selection of the related phyla, whereas the links in category 4 are likely correlational co-decay relationships.

Cluster 1 contributed the most, and to all four correlation categories, with above-average links to many phyla regardless of the correlation category. Additionally, those links appear always in the same season (i.e., spring). This indicates that cluster 1 contributes to both declining and increasing DOM patterns in spring, yet the phyla involved in each correlation category are different, and these differences suggest fundamentally different mechanisms. The phyla that are related to cluster 1

are shared between correlation categories 1 and 2 and are different from those that are related with correlation categories 3 and 4, and the two phyla groups have opposing microbial spatial behaviours. When examined closely, the phyla involved in correlation categories 1 and 2 are not identical but clearly overlap and are related, since they appear closely along the phyla axis, which is ordered by their phylogenetic relatedness. With the highest number of above-average links in correlation category 2 (which may represent potential microbial consumption of DOM), these results may collectively indicate that cluster 1 represents a pool of compounds that is ubiquitous in spring, and where a fraction is potentially biodegradable and linked to microbial growth. In general, there was a strong seasonality in these above-average DOM/microbial relationships. Clusters 1, 2, and 5 contributed to correlations mostly in spring, whereas clusters 3 and 4 dominated correlations in summer. These results collectively indicate that different DOM clusters are associated with distinct correlation categories regardless of the phylogeny of the microbes involved. In addition, there is evidence that within correlation categories and seasons, specific phyla appear to be more related or reactive to specific DOM clusters than others. Overall, we infer that clusters 1 and 3 are potentially involved in microbially mediated degradation in spring and summer, respectively, whereas cluster 2 is only correlated to decaying microbes perhaps because they share the same origin (i.e., terrestrial). Increases in Cluster 4 and 5 are likely associated to the growth of certain microbial groups in summer and spring, respectively whereas cluster 4 may additionally be produced by a specific set of decaying microbes.

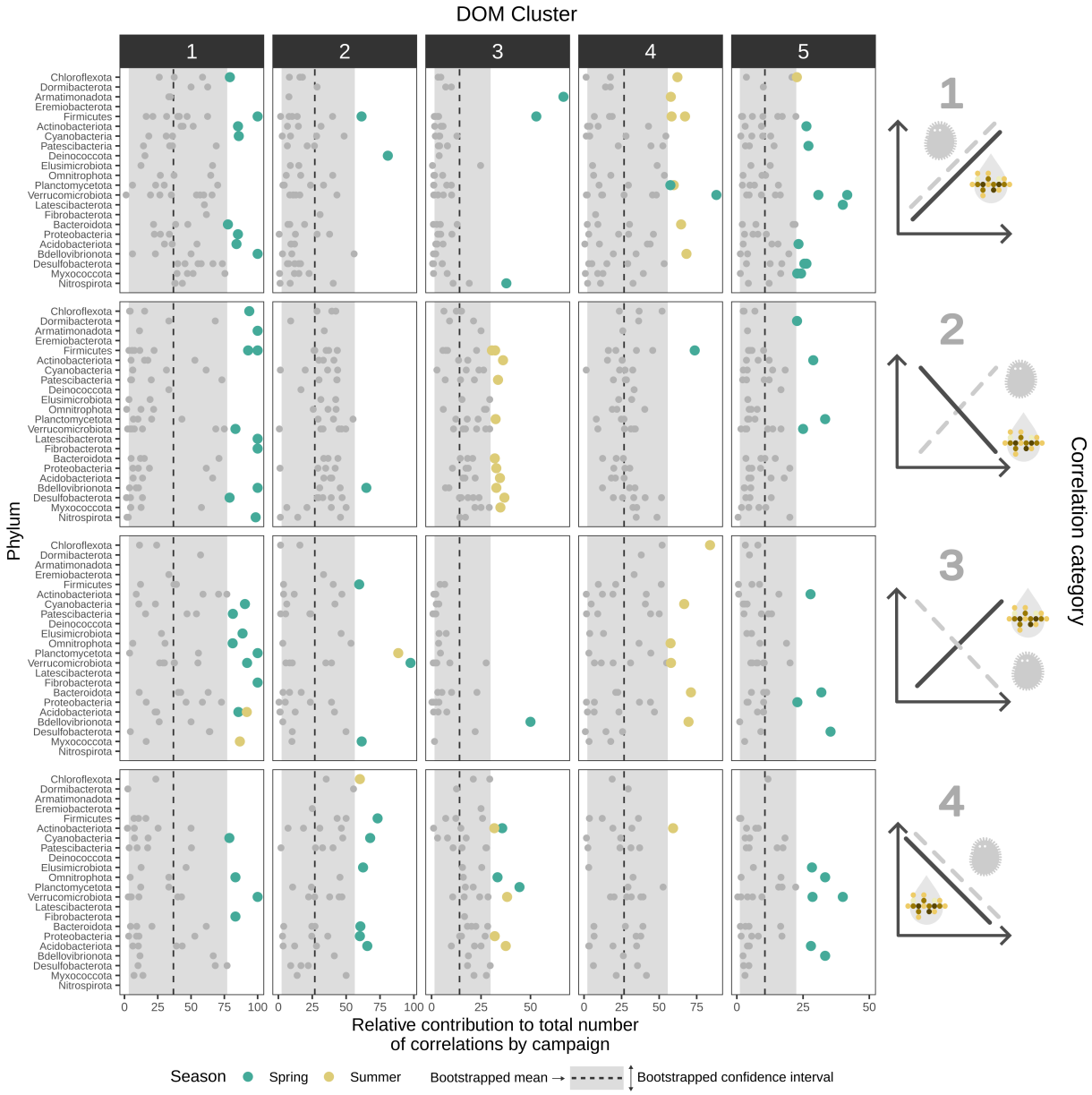

Figure 2.7: Proportion of significant correlations attributed to DOM clusters by phylum and correlation category. Each point represents correlations of phyla by season and year. Phyla subdivisions (e.g., Verrucomicrobiota A and B) are visualised in the same phyla row, however, were treated as separate phyla in the analysis. Phyla were sorted by phylogenetic relatedness according to the GTDB tree. Vertical line denotes bootstrapped mean by DOM cluster (R = 1,000) and grey area denotes lower and upper bootstrapped confidence interval (95%). Unimodal spatial patterns and correlation relationships not consistent with our theoretical interpretation were excluded. Non-linear and linear patterns were merged.

2.5 Discussion

In this study, we used spatial patterns of dissolved organic matter (DOM) molecular formulae (MF) and microbial operational taxonomic units (OTUs) along a flow-weighted water age (FWWA) gradient in a complex aquatic network as an index of reactivity. Reactivity here refers to indicators of change, distinct from traditional definitions tied to DOM degradability or microbial activity. By categorising MF and OTUs into increasing, unimodal, or decreasing trends, we identified the most reactive fractions of both assemblages, those presumably responding to environmental changes along the aquatic continuum and across seasons and years. This approach allowed us to extract ecologically relevant relationships by filtering out spurious correlations that were abundant when linking bulk assemblages and which obscured DOM-microbe interactions.

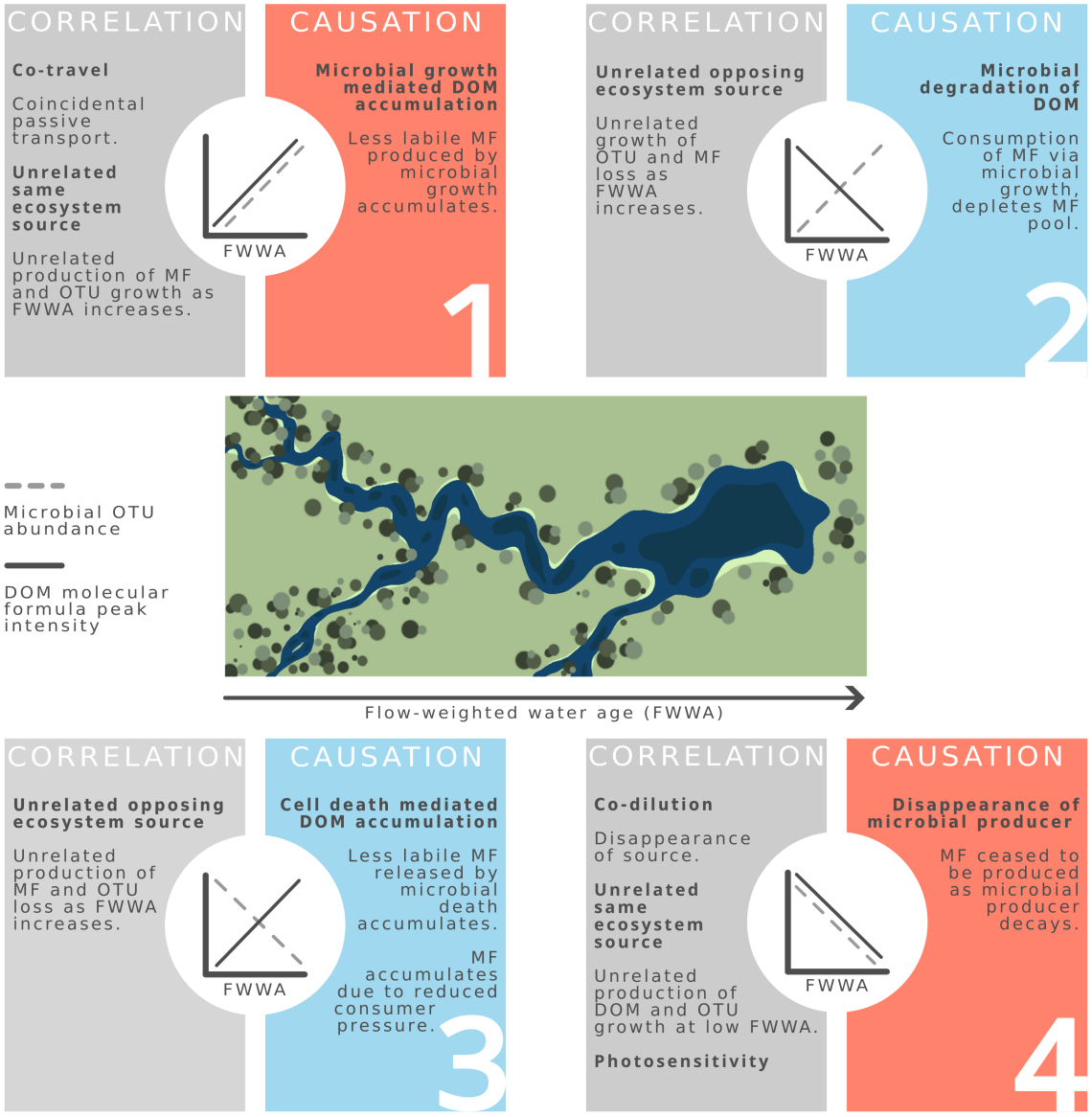

Figure 2.8: Conceptual diagram contrasting correlational and causal relationships between microbes and DOM. Dashed, and solid lines represent microbial and molecular spatial patterns, respectively. Listed potential relationships are hypothetical interpretations and not based on data.

For decades, it has been known that a portion of microbial assemblages are unreactive, either dead or dormant (Lennon and Jones, 2011). Hence, it has been hypothesised that indecisive findings between microbial and DOM assemblages may potentially arise due to the presence of a vast

pool of taxa that are not contributing to local processes (Hall *et al.*, 2018). Previous estimates suggest that around 30% of planktonic OTUs in freshwaters may be inactive (Lennon and Jones, 2011), however, our findings of 93% unreactive OTUs based on their spatial patterns of abundance are notably higher. Nevertheless, these estimates approximate previously modelled ranges for highly selective environments (Cole, 1999). These differences may arise from both conceptual and methodological factors. Our approach was conservative - defining reactivity through statistically significant spatial patterns - likely overestimated unreactive OTUs, whereas earlier estimates relied on fluorescence staining techniques to differentiate between live and inactive or dead cells (Gasol *et al.*, 2008). Our conservative threshold may be difficult to meet, particularly when spatial patterns are subtle, or sample sizes are insufficient, hence the lack of a significant spatial relationship does not necessarily imply inactivity or death. For microbial sequencing, detection limits add another layer of sampling effects (Smith and Peay, 2014). Especially for moderately abundant and rare microbes, more sampling points or deeper sequencing may be needed to achieve statistical significance in our spatial modelling framework (Pedrós-Alió, 2012). This methodological limitation was also reflected in our RNA analysis. We found that the reactive RNA fraction was comparable to that of DNA, although we had expected a higher proportion since RNA is often considered to represent recently active microbes (Stadler and del Giorgio, 2022). On the one hand, these results may indicate that our approach extracts the most evident spatial patterns, which naturally are reflected in the same OTUs being extracted as reactive moieties in both DNA and RNA. On the other hand, the short half-life of RNA (Steiner *et al.*, 2019) may reduce the sampling effort and hence prevent the successful modelling of spatial patterns, leading to an underestimation of the reactive RNA moieties. To avoid any over-interpretation, we decided to focus on the DNA spatial patterns in all analyses since sampling effects are likely smaller compared to RNA.

Similar to the unreactive pool of microbes, there is a fraction within DOM assemblages that is seemingly persistent across environmental changes (Kellerman *et al.*, 2015), whereas other molecules are available for photo- and biodegradation (Cory and Kling, 2018). Persistence within DOM assemblages is yet another facet of microbial-DOM interactions that may prevent the successful establishment of meaningful links between the two. Only a few studies have examined the proportion of reactive DOM molecular formulae. While not directly comparable, one study estimated

that approximately 11% of DOM MF were ubiquitous across multiple North American rivers (Stadler *et al.*, 2023), while approximately 21% of compounds were shared between soilwaters and headwater streams (Freeman *et al.*, 2024). Similarly, another study found that 47% of MF in lakes co-occurred with multiple marine environments (Zark and Dittmar, 2018). Our approach identified an average of 56% of molecular formulae as unreactive to environmental changes. Unlike co-occurrence-based approaches, our method captures MF that exhibit dynamic changes in peak intensities along a hydrologic continuum, rather than simply identifying common MF - a key distinction from previous studies. Unlike the static core pool described in earlier studies (Kellerman *et al.*, 2015; Freeman *et al.*, 2024), our results suggest that the unreactive fraction is dynamic, influenced by inter-annual and seasonal variations. For example, summer proportions were vastly different among years by almost 50%. Since there were no clear hydrological differences between the two years, these results may be the consequence of the flooding of a newly constructed reservoir in 2016. The reservoir had a larger proportion of flooded peatland compared to the previously flooded reservoir (2014; Rust *et al.* (2022)), which may have introduced fresh terrestrial compounds, altering DOM composition (Larson *et al.*, 2014). Such findings underscore the role of large-scale watershed disturbances, like damming or wildfires, in shaping DOM dynamics (Lambert *et al.*, 2016; Xenopoulos *et al.*, 2021).

The interplay between intrinsic factors (e.g., microbial adaptability (Ruiz-González *et al.*, 2017b; Savio *et al.*, 2015), DOM degradability (Catalán *et al.*, 2021; Cory and Kling, 2018)) and extrinsic drivers (e.g., hydrology, temperature (Niño-García *et al.*, 2016a; Read *et al.*, 2015; Massicotte *et al.*, 2017; Lambert *et al.*, 2016)) determines the fate of microbes and DOM. Our data revealed that microbes with increasing trends along the continuum tended to dominate in aquatic systems, suggesting a consistent capacity for growth, despite seasonal shifts in taxa. Previous findings in the same watershed highlighted that environmental selection processes become stronger in high water residence time systems (Stadler and del Giorgio, 2022). Together, our findings indicate that even when there are vastly different hydrologic conditions, a large fraction of microbes can grow at least at some point within the continuum as evidenced from the many microbes following non-linearly increasing spatial patterns. Phylogenetic analysis showed that increasing taxa were more closely related among each other than decreasing taxa, hinting at shared traits among aquatic-

adapted microbes. Interestingly, these phylogenetic signals were only found in deeper levels of the tree rather than the tips. These findings may suggest that there are general preferences among phyla for growing in aquatic systems (Schmidt *et al.*, 2016), however, there may also be ecotypes within the same phyla, where highly related taxa can occupy very different niches to avoid competition (Pernthaler, 2017). In contrast, microbes that declined were widely spread across the phylogenetic tree, indicating that failure to establish in any freshwater ecosystem is likely not a preserved trait. It is unclear from our study what functions manifest in the ability to grow along the aquatic landscape since we did not measure functional genes, however, it clearly did not depend on the potential growth rate (16S gene copy number). Other functional traits may be more important for aquatic growth such as biofilm- and algae-associations (Dow *et al.*, 2020; Graham *et al.*, 2015) and/or nutrient-uptake efficiencies of limiting nutrients in freshwaters (Currie and Kalf, 1984).

Unlike the microbial community, the DOM assemblage showed alternation in the dominance of spatial patterns, however, this variation was inter-annual and not seasonal. As discussed earlier, the discrepancy between the two years is likely a result of the flooding of a new reservoir in early 2016. The damming of rivers is known to change DOM composition through the flooding and release of terrestrial material (Wang *et al.*, 2020) and increased algal biomass (Oliver *et al.*, 2016). There were no algal blooms observed in the sampled years; hence, it is likely that the plethora of observed non-linear spikes of compounds is attributable to larger terrestrial contributions of fresh, highly oxidised compounds with low nitrogen (N) content (cluster 2 derived from FT-ICR MS hierarchical clustering, Fig. 2.2). In contrast, the previous year exhibited a larger proportion of MF declining along the continuum attributable to large, olefinic, and degraded compounds with low N content (cluster 4). Unlike microbes, DOM assemblages exhibited greater functional coherence within spatial categories, emphasising the role of intrinsic molecular properties in shaping their fate within the network (Kellerman *et al.*, 2015). Compounds were most functionally alike among increasing MF, with relatively high aromaticity and NOSC values representing thermodynamically favourable compounds. Compounds that decreased were also relatively similar to each other, and tended to have lower aromaticity and NOSC values. Compounds that were the most thermodynamically favourable and aromatic were found within the unimodal pool, which likely

reflects that these compounds are the most available in terms of photo- and biodegradability. Differences among spatial patterns in these chemical metrics were also evident when analysing the bulk pool. These findings contrast results of a recent study showing that compounds with higher NOSC values were preferentially removed along a stream network in summer (Peter *et al.*, 2020). Additionally, it has also been shown that aromaticity and photo-reactivity decrease with increasing water residence time (Kellerman *et al.*, 2014; Grasset *et al.*, 2024), while aliphatics were positively correlated to water residence time. Patterns similarly opposing previous literature were found in a boreal soil-stream-river continuum (Hutchins *et al.*, 2017), and it was argued that the results are not necessarily contradictory as the studies cover different portions of the continuum. A similar scenario can be envisioned here, and it seems that the inclusion of terrestrial ecosystems in the continuum is a key element that explains these seemingly contradictory patterns to studies that have only sampled aquatic ecosystems (Kellerman *et al.*, 2014; Peter *et al.*, 2020; Grasset *et al.*, 2024). Additionally, we evaluated changes along a water age continuum, which integrates both water residence time as well as an aspect of water history in the watershed. It is difficult to interpret mere DOM spatial patterns when there are no microbial data associated to the patterns. Hence, our dataset represents a unique opportunity to inch a step closer to answer why certain DOM molecules exhibit the spatial patterns we observe along a terrestrial-aquatic continuum.

When microbes and dissolved organic matter interactions are evaluated in empirical settings, it is extremely difficult to distinguish correlational patterns from causal interactions. DOM that has been characterised via absorbance, fluorescence or high-resolution approaches have been described to govern microbial community composition or diversity patterns many times before (Crump *et al.*, 2003; Jones and McMahon, 2009; Ruiz-González *et al.*, 2015b; Muscarella *et al.*, 2019; Ávila *et al.*, 2019), however, their complex bi-directional relationship has not often been addressed (Hu *et al.*, 2022). When put together in a conceptual framework, it becomes apparent that vastly different correlational and causal scenarios may result in the same relationships between microbes and DOM molecular formula along hydrological gradients (Fig. 2.8). Microbial growth-mediated DOM production (i.e., DOM transformation, degradation by-products) may be less labile than their precursor (Koch *et al.*, 2014) and would consequently accumulate as water ages (Fig. 2.8, correlation category 1). Yet, the same patterns may emerge simply due to co-travel and constant influx

of terrestrially sourced persistent microbes (Crump *et al.*, 2012) and less labile DOM (Massicotte *et al.*, 2017), or unrelated production and growth of MF and OTUs, respectively, in high FWWA systems (i.e., MF produced in lakes (Stedmon and Markager, 2005; Liu *et al.*, 2020), OTUs adapted to lakes (Comte *et al.*, 2017)). DOM degradation by microbes (Fig. 2.8, correlation category 2) is the most assumed causal relationship between heterotrophic microorganisms and DOM (Wiegner and Seitzinger, 2001; Fasching *et al.*, 2014), however, along a hydrologic continuum, the same negative relationship may emerge due to unrelated loss of MF (e.g., photodegradation of allochthonous DOM (Grasset *et al.*, 2024), or dilution of terrestrially derived DOM sourced at specific points in watershed), and OTU influx or growth in high FWWA systems (i.e., high WRT adapted microbes; Niño-García *et al.* (2016a)). The reverse scenario of opposing sources results in another negative relationship, when autochthonous MF and allochthonous OTUs are correlated (Fig. 2.8, correlation category 3). Yet, in a causal relationship, we might be dealing with cell death mediated release of less-labile DOM that may accumulate with water age (Kawasaki and Benner, 2006). Finally, the least plausible causal scenario is when specific microbes produce certain MF and they may consequently produce synchronously declining patterns as the microbial producer disappears (Fig. 2.8, correlation category 4; Kawasaki and Benner (2006)). This relationship is the most likely to have more purely correlational underpinnings such as joint dilution away from a common (likely terrestrial) source (i.e., allochthonous; Ruiz-González *et al.* (2015a)), or unrelated DOM production in and microbial adaptation to low FWWA systems (Zeglin, 2015; Boodoo *et al.*, 2020). It may also potentially indicate shared photosensitive characteristics of specific DOM and microbes as UV stress increases in high WRT systems (Herndl *et al.*, 1999; Grasset *et al.*, 2024). Among the potential causal relationships, some patterns are driven by bacteria (i.e., bacterial consumption of DOM, Fig. 2.8, correlation category 2, Kamjunke *et al.* (2020); Freeman *et al.* (2024)), while others are driven more by the DOM compounds (i.e., DOM selecting for certain bacterial composition, Fig. 2.8, correlation category 1 and 4, Findlay (2003); Docherty *et al.* (2006); Bambakidis *et al.* (2024)).

In our analyses, we first identified ecologically and statistically meaningful relations between reactive microbes and DOM MF along the sampled FWWA gradient out of the vast numbers of potential correlations that exist between all MF and OTUs across the watershed (Fig. S2.8). We subsequently

attempted to parse out which of these meaningful relationships were potentially causal patterns versus merely correlational (Fig. 2.7). Many spurious relationships were filtered out by our conservative modelling approach and correlation filtration step that removed relationships that did not have the same relationship sign as would be expected (e.g., increasing OTU x increasing MF = positive relationship). Although these conservative filtration steps were put in place, it is impossible to completely remove significant relationships that emerge due to strong spatial or seasonal patterns observed along a gradient in both MF and OTUs. For example, when a MF has a strong positive trend along the FWWA gradient and is correlated against ubiquitous microbial OTUs that have a slight downward or upward trend, significant correlations will be found by virtue of strong spatial gradients. At the same time, many causal relationships may not be captured due to the transient nature of highly reactive compounds (Cory and Kaplan, 2012; Pollard, 2013). What we are able to measure in the environment are likely products of partial-degradation (i.e., by-products or leftovers, Tranvik (1992); Fasching *et al.* (2014); Cory and Kling (2018)). Hence, we are more likely to identify mere correlational patterns over true causal relationships. The correlational patterns that we often examine using nodes and links in network analyses (Zhou *et al.*, 2021; Hu *et al.*, 2022), cannot distinguish between such nuanced DOM-microbial relationships.

By examining in detail the above-average contributions between correlations of various DOM clusters and microbial phyla, we were able to gain some insight on which relationships that we identified could be potentially causal versus correlational. The clearest observation we made from our analysis is that seasonality plays a major role in determining the nature of the links that exist between microbes and DOM compounds. In spring, we found that some microbial phyla were potentially degrading fresh, aromatic DOM with high N content and small molecular size (cluster 1), especially in high water age systems generating a causal pattern typical of microbial degradation (Fig. 2.8, panel 2). Although terrestrial DOM has long been thought to be mostly recalcitrant (Moran *et al.*, 1991; Smith and Hollibaugh, 1993), our results add to evidence that suggests widespread microbial consumption of terrestrial DOM (McLaughlin and Kaplan, 2013; Wiegner *et al.*, 2015; Fitch *et al.*, 2018), especially of small molecular size (Berggren *et al.*, 2010). These terrestrial inputs have been shown to be elevated during the spring freshet (Spencer *et al.*, 2008; Berggren *et al.*, 2010). It is noteworthy, however, that DOM cluster 1 contributed to all correlation

categories. These results may indicate that only some compounds within cluster 1 are available for microbial degradation, whereas others evade degradation for reasons unclear in our study and may be continuously loaded, therefore generating correlational patterns with microbes that both increase or decline along the continuum (Fig. 2.8, panels 1 and 3). DOM cluster 2 was also only associated to microbes in spring, and its appearance in only one correlation category may imply that this large molecular weight, low N, fresh and phenolic/highly oxygenated pool likely degrades as non-aquatic microbes decay along the water age gradient. These compounds may be diluted or photo-degraded along the water age continuum (Cory and Kling, 2018; Grasset *et al.*, 2024), generating likely a correlational pattern (Fig. 2.8, panel 4). Lastly, DOM cluster 5, characterised by high N, small molecular size, aliphatic-like and decomposed nature was found to select for certain microbial phyla (Fig. 2.8, panel 1), implying that these compounds may not be available to a wide array of microbial taxa (Gómez-Consarnau *et al.*, 2012; Logue *et al.*, 2016). Given that they increase synchronously along the water age gradient, it may also represent the endpoint of a DOM degradation state (Hutchins *et al.*, 2017; Kellerman *et al.*, 2018) that corresponds to a spatial succession of microbial communities (Read *et al.*, 2015; Hassell *et al.*, 2018; Wisnoski and Lennon, 2021) associated to the DOM processing state in spring, where succession senescence occurs in high WRT systems with the accumulation of more decomposed DOM (Hosen *et al.*, 2021). Regardless, these potential relations would represent a causal relationship (Fig. 2.8, panel 1). In contrast to spring, summer DOM correlations were clearly dominated by clusters 3 and 4. Our results suggest degradation of cluster 3 by multiple phyla, indicating that these medium sized, high N, high NOSC, olefinic and mid-decomposed compounds may stimulate microbial growth, and this underlies what we interpret as a potentially causal relationship (Fig. 2.8, panel 2). The lack of correlations with fresher, less decomposed matter in summer may indicate that such compounds are much more rapidly taken up and transformed (Pollard, 2013), hence, we are only able to observe a DOM-microbial degradation relationship with mildly decomposed material that may have already undergone bio- and photo-degradation (Lapierre and Giorgio, 2014; Cory and Kling, 2018). Lastly, DOM cluster 4 (large molecular mass, olefinic-like, mildly decomposed, low NOSC) seems to select for certain microbial taxa in summer (Fig. 2.8, panel 1), similar to how cluster 5 acted in spring. However, since the phyla involved in correlation category 1 (potentially DOM selection of taxa) for cluster 4 and 5 are completely different, these DOM pools appear to be operating on a different set

of phyla (Gómez-Consarnau *et al.*, 2012; Logue *et al.*, 2016). Both DOM clusters are decomposed, stable, thermodynamically unfavourable compounds, however, they differ in their molecular size and tendency to contain N. These results indicate that DOM of varying size may select for distinct portions of the microbial community as has been described before in marine systems (Amon and Benner, 1996; Varela *et al.*, 2020). Additionally, cluster 4 may also be produced via microbial cell death due to its positive correlation with the decline of some phyla (Fig. 2.8, panel 4). This finding adds to evidence of compounds produced with microbial cell death, such as amino acids (Kawasaki and Benner, 2006) or humic-like compounds (Stadler *et al.*, 2020). These compounds that may be related to cell death are rather stable (low O/C) and thermodynamically unfavourable (low NOSC), and hence may be rather recalcitrant as previously proposed (Nagata and Kirchman, 1999; Ogawa *et al.*, 2001).

Together, our results shed light on the intertwined nature of dissolved organic matter and microbial community composition and address the conceptual difficulties of deciphering their interactions along a hydrological continuum. By extracting the most reactive units within both assemblages and examining their interactions along a water age gradient, we were able to discern a few potential causal relationships between portions of the DOM pool with various microbial phyla. It seems that the high influx of fresh terrestrial organic matter in spring, allowed us to identify microbial-DOM degradation dynamics involving fresh, highly available compounds, while the remaining decomposed, stable compounds selected for a specific community composition in high water age and WRT systems. In summer, however, no direct link of fresh DOM was found with microbes, but rather with mid-sized, partially decomposed compounds. These results may imply that higher temperatures, lower flow rates and longer periods since a last pulse of fresh terrestrial DOM (Raymond *et al.*, 2016) may have exhausted the fresh terrestrial DOM pool, whereas fresh autochthonous primary production is too quickly processed to be detectable in our sampling timescales (Pollard, 2013). Hence, degradation correlations found in summer likely represent transformation products of the most degradable DOM compounds, that are still accessible to microbial metabolism. Our comprehensive, watershed-scale study exemplifies how individual relationships between high-resolution microbial and DOM assemblages can be extracted and interpreted, when put together into a holistic conceptual framework that accounts for ecological,

and hydrological dynamics in both DOM and microbial assemblages.

2.6 Acknowledgements

We would like to begin by acknowledging that the land on which the field work was conducted is on the traditional and unceded territory of the Innu community (Nitassinan), and the laboratory work in Montréal was conducted on the land of the Kanien'keha:ka people, as well as of the Métis people. We are especially grateful to Alice H. Parkes, Annick St-Pierre, and Serge Paquet, who maintained and oversaw the La Romaine project over the years. For field and laboratory work we would like to thank Felipe Rust, Clara Ruiz-Gonzàlez, Alexandre Ducharme, Roy Nahas, Ryan Hutchins, Marie Laure Gérardin, Erin Hotchkiss, Karelle Desrosiers, Martin Demers, Sara Mercier Blais, Julia Jakobsson and Francesca del Giorgio. We thank Marie Laure Gérardin, Gabriel Duran and Hydro-Québec for providing valuable discharge data to build our hydrological models. We would also like to thank Katherine Velghe and Maryline Robidoux for laboratory assistance, and Yves Prairie for aiding in the water age calculations. This research was in part enabled by computing resources provided by Calcul Québec (`https://www.calculquebec.ca/`) and the Digital Research Alliance of Canada (`https://alliancecan.ca/`). This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Québec.

2.7 Data availability

The raw 16S rRNA gene sequences, both DNA and cDNA are available at the public NCBI Sequence Read Archive (SRA) as part of the BioProject PRJNA693020. The code will be made available on Github after successful publication and both code and processed microbial data will be separately archived on Zenodo.

2.8 Supplementary Information

2.8.1 Hydrological estimates

To visualise and study the spatial patterns of microbes and molecules along a true hydrological continuum, flow-weighted water age (FWWA) was estimated for the studied watershed. FWWA represents the average time water takes to arrive at any given point in the hydrological network. A digital elevation model (18 x 18 m) was obtained from GeoGratis Canada (Natural Resources Canada, 2017) to delineate the watershed and calculate metrics such as flow accumulation, flow length, and pixel area using the Spatial Analyst Toolbox in ArcMap (v10.7.1, ESRI Inc., Redland, CA, USA). A flow accumulation threshold of 3,000 pixels was used to identify the stream network. Flow accumulation was converted to catchment area by multiplying it by the pixel area

2.8.1.1 Reconstruction of a true hydrological network

To our knowledge there are no attempts to estimate FWWA within a watershed by considering the various ecosystems (i.e., lentic versus lotic systems) within a hydrological network, and they commonly assume that there are no lentic systems. To reconstruct a 'true' hydrological network, we first identified which pixels were lentic or lotic systems by overlaying the HydroLAKES (Messager *et al.*, 2016) and reservoir (Hydro-Québec) polygons to the stream network and applied a buffer of 0.004 degrees. To identify the downstream pixel for each pixel in the watershed, we sorted the stream network by flow length (*Flow length* tool in ArcMap) within unique reach IDs assigned to stream vertices (*Stream to Feature* tool in ArcMap). To identify downstream pixels at confluences, unique node connection IDs that connect two stream reaches were utilised (extracted from *Stream to Feature* tool in ArcMap). Once the ecosystems along the stream network and the flow path were identified, tributaries into lakes and reservoirs were identified by finding pixels of ecosystem shifts (e.g., fluvial to lake) along the stream network. The lake/reservoir main inlet was identified as the tributary with the longest flow length. Similarly, the lake/reservoir outlet was identified as the pixel within the lake/reservoir with the longest flow length. The main channel for each lake and reservoir was subsequently identified by tracking the inlet reach downstream until the outlet. Other reaches that flow within lentic systems were not considered for subsequent water age calculations, and hence the identified tributaries' confluence was moved

onto the main channel of the lentic system (Fig. S2.2). Hydrological network sorting, and identifying the flow paths, tributaries, lake inlets and outlets were programmed in R (v.4.3.1, R Core Team (2024)).

2.8.1.2 Fluvial discharge, velocity and residence time

Discharge and velocity were measured in a few streams ranging Strahler orders 1-4 within two headwater sub-watersheds (Petite Romaine and Bernard) in 2015/2016 and 2021/2022, respectively, using a 2-D Acoustic Doppler Velocimeter (FlowTracker, Sontek, San Diego, CA, USA). Additionally, a few hydrological stations located at Strahler orders 5-7 were periodically measured for discharge and velocity using a vessel mounted Acoustic Doppler current profiler (data provided by Hydro-Québec, Montréal, QC, Canada). All May/June measurements were categorised as high flow, while all other months were identified as low flow conditions. Combining these datasets spanning over the years 2010-2022 and Strahler orders 1-7 (n = 103), we identified a model to estimate discharge (Q) using catchment area (km^2) by flow condition (Fig. S2.3a):

$$log\ Q_{high} = 0.26 + 14.03 \times log\ CA - 1.96 \times log\ CA^2$$

$$log\ Q_{low} = 0.26 + 14.03 - 0.28 \times log\ CA - 1.96 \times log\ CA^2$$

where *CA* is catchment area in km^2 and *Q* is discharge in $m^3\ s^{-1}$ (R^2 = 0.88). Discharge and catchment area were log-transformed to fulfil model assumptions. Velocity was also estimated using catchment area and flow condition as follows (Fig. S2.3b):

$$log\ v_{high} = -0.49 + 3.22 \times log\ CA - 2.17 \times log\ CA^2$$

$$log\ v_{low} = -0.49 + 3.22 - 0.29 \times log\ CA - 2.17 \times log\ CA^2$$

where *CA* is catchment area in km^2 and *v* is velocity in $m\ s^{-1}$. Although the R^2 was not particularly high (R^2 = 0.46), the magnitude of the measured velocity as well as the relationship of velocity

among Strahler orders found in the empirical data was captured with this model (Fig. S2.4). These two models were the most accurate and parsimonious solutions in comparison to directly calculating velocity from estimated channel cross-sectional area and discharge. WRT in each fluvial pixel (WRT_{px}) was estimated by dividing the pixel length by the estimated local velocity.

2.8.1.3 Reservoir water residence time

To estimate WRT_{px} within the reservoir, the overall water residence time within the reservoir was calculated for each month. Daily water level measurements for the reservoir were transformed into volume estimates using a conversion key provided by Hydro-Québec. The conversion key was implemented using a general additive model (GAM) approach as the relationship between water level and volume was not linear (data not provided). Discharge stations located at the turbine and overflow (water not used for electricity production) of the reservoir were used to calculate the daily outflow from the reservoir. Daily estimates were averaged by month and monthly water residence time was computed as:

$$WRT_{Reservoir} = \frac{V}{Q_{out}}$$

where V represents reservoir volume in m^3 and Q_{out} the discharge at the outflow in $m^3\ s^{-1}$. Subsequently, the monthly reservoir WRT was divided by the number of pixels along the main channel of the reservoir to estimate the WRT_{px} along the main channel.

2.8.1.4 Lake water residence time

Due to the lack of data for remote lakes sampled in the watershed, WRT estimates were taken from the HydroLAKES dataset (Messager *et al.*, 2016). Similarly to reservoir estimates, the overall WRT in the lakes was divided by the number of pixels along the main channel flowing through the lake to estimate WRT_{px}.

2.8.2 Modelling and classifying spatial patterns

Before modelling spatial patterns, the dataset was filtered by each sampling campaign. Extreme outlier observations of an operational taxonomic unit (OTU) or molecular formula (MF) were removed (values above or below 3 times the interquartile range), MF and OTUs with less than 8 and 7 observations, respectively, within a campaign were removed (based on histogram observations). Observations were z-scaled for each OTU and MF per campaign.

Prior to the modelling exercise, observations were binned and averaged at a 50-day interval along the water age gradient. A MF and OTU was only considered for modelling if more than 3 bins recorded an actual observation. To allow various dynamic patterns to be modelled, we established a decision tree that selects the best model for individual MF and OTUs along the FWWA gradient (illustrated in Fig. S2.5). The decision tree starts with first fitting a linear model and then comparing it to polynomial (2nd and 3rd order) and GAM models using maximum likelihood estimation (*gam()* function; *mgcv* package; Wood (2011)). Each bin was weighed by its corresponding number of observations. The best model was selected based on the smallest Akaike Information Criterion (AIC). If the model with the smallest AIC was a non-linear model, non-linearity was justified by testing it against the linear model using a χ^2-test (*anova()*; *mgcv* package; Wood (2011)). Once the best model was selected, model statistics such as *p*-value and R^2 were extracted. The slope of the initially fit linear model was extracted regardless of which model type was selected to aid in spatial pattern classification described below. For all non-linear models, the 2^{nd} derivative was utilised to find peak locations along the FWWA gradient.

All linear models that had a positive slope were classified as 'increase', while all linear models with negative slopes were identified as 'decrease'. All non-linear models were first classified by their number of peaks. Models with a single peak were classified by where their peak was located. For each model, the centre along their respective water age range was identified. Subsequently, a buffer area was defined around the centre by adding and subtracting 1/6 of the FWWA range to the centre. If a peak was located within the buffer area, the model was classified as 'unimodal'. All other one-peak models were classified as 'non-linear decrease' if their peak was located below the FWWA range centre and models with their peak located higher than the FWWA range centre was

classified as 'non-linear increase'. Any non-linear models with more than one peak were classified as multimodal increase and decrease when their linear model slope was positive and negative, respectively. Non-linear and multimodal spatial patterns were merged depending on the analysis. All models that had a slope of 0 or did not return a *p*-value were removed from downstream analyses.

2.8.3 Supplementary figures

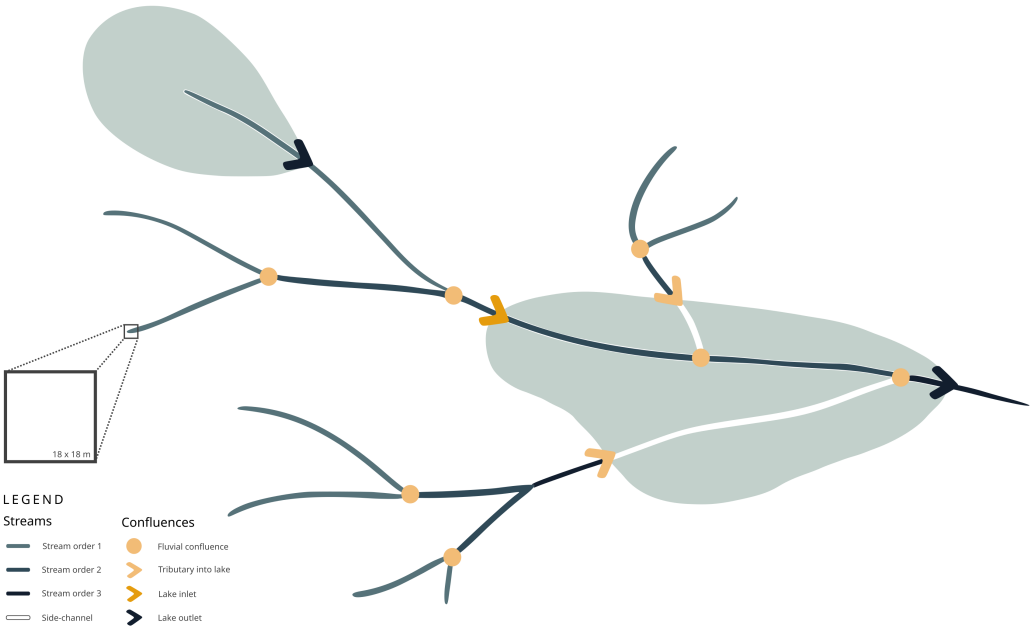

Figure S2.1: Schematic representation of flow-weighted water age calculation. Stream colours represent Strahler order. Within lakes only the coloured channel was considered when calculating flow-weighted water age (= main channel). The white channels' pixels were skipped during the cumulative calculation. Hence, the FWWA until the confluence to the lake (yellow arrow) was summed to the main channel where the side channel merges into the main channel (yellow point).

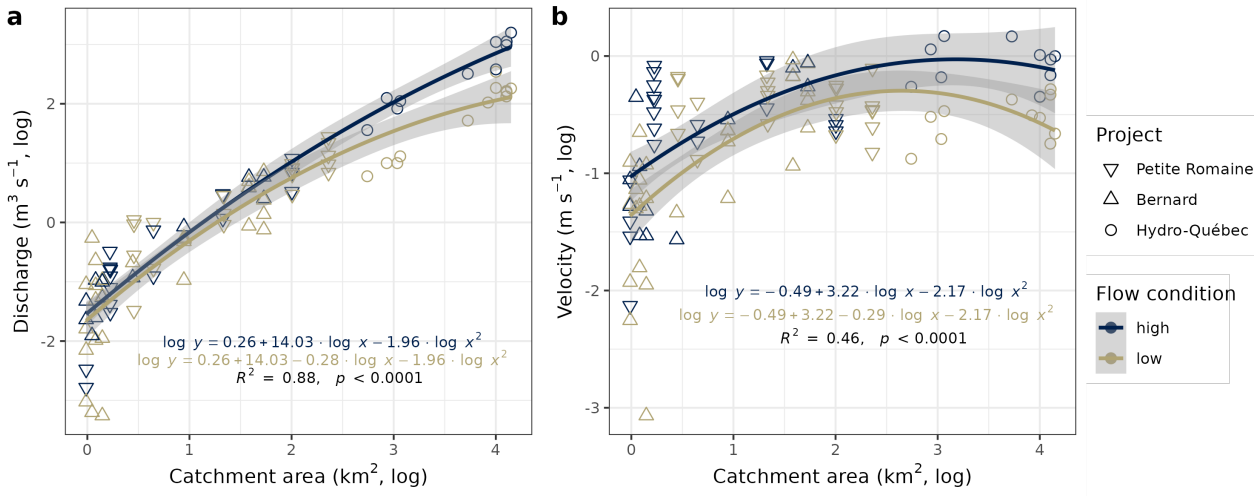

Figure S2.2: Hydrological models used to estimate water age in the watershed. a) Log-transformed discharge as a function of log catchment area in km^2. Model equations by flow condition are given in blue for high and brown for low flow. b) Log-transformed velocity as a function of log catchment area. Model equations are likewise given by flow condition. Various point shapes indicate the source of empirical measurements used to construct models. Petite Romaine and Bernard are sub-watersheds of La Romaine watershed, representing small headwater watersheds. Hydro-Québec data capture larger rivers within the watershed.

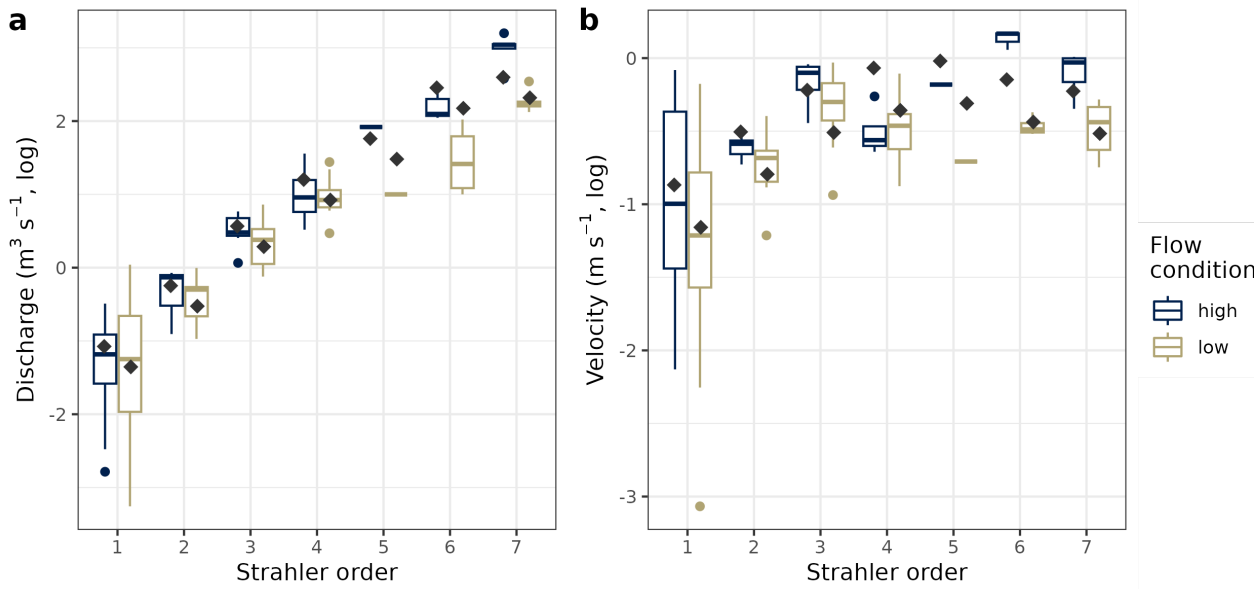

Figure S2.3: Measured versus estimated discharge and velocity by Strahler order. Diamonds indicate the median estimated value for each Strahler order and flow condition using the models presented in figure S2.2. Boxplots represent the measured discharge (a) and velocity (b) within the watershed. Blue and brown colours indicate high and low flow, respectively. The boxplot middle line represents the median, lower and upper hinges correspond to the 25th and 75th percentiles. Upper and lower whiskers expand to the largest and smallest value, respectively, no further than 1.5 times the inter-quartile range (IQR) from the hinge. Outliers are depicted as points.

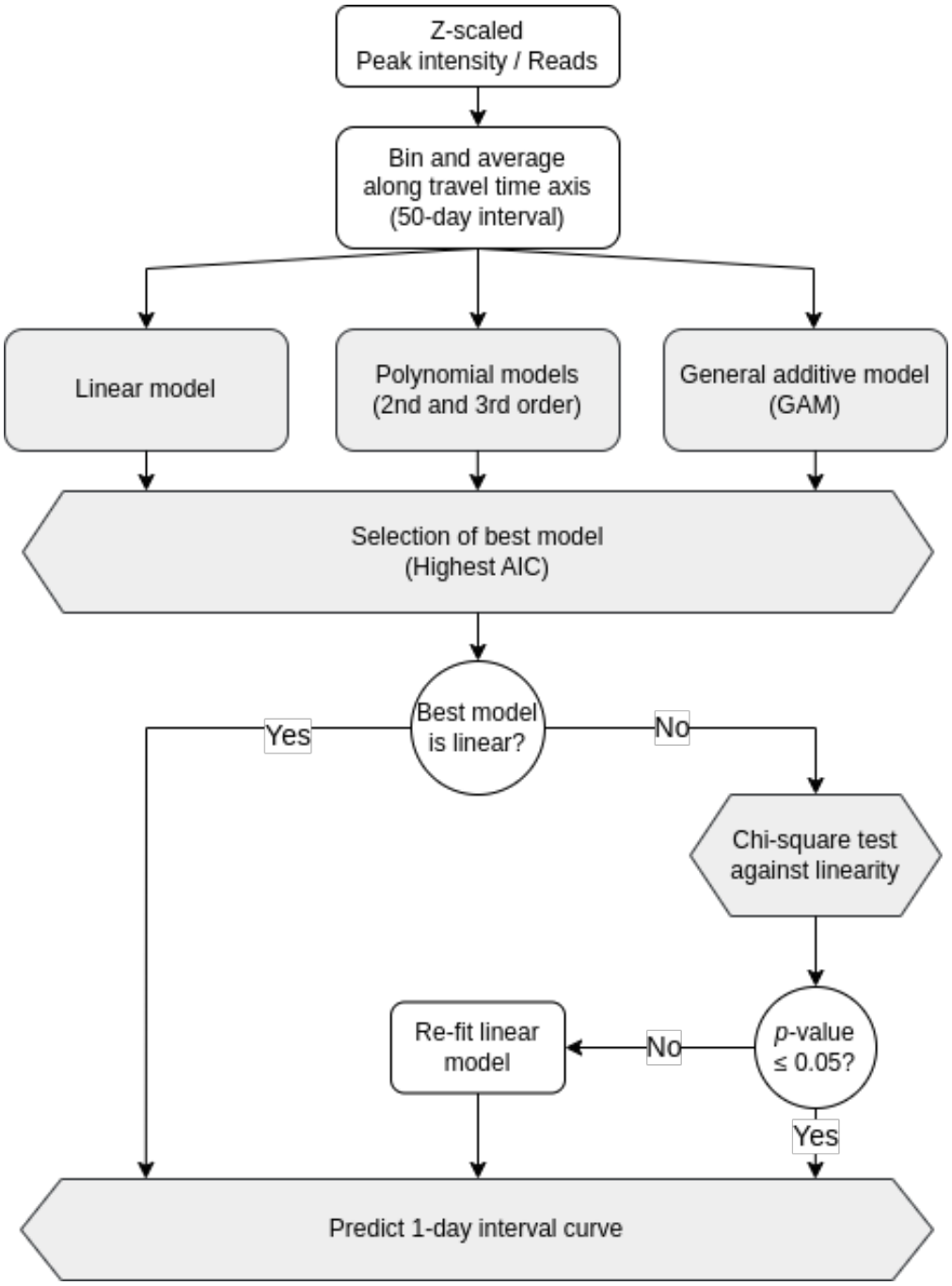

Figure S2.4: Flow chart illustrating decision tree utilised in modelling framework. Steps in the depicted decision tree were followed to find the best fitting model to characterise the spatial pattern for each operational taxonomic unit (OTU) and molecular formulae (MF) for each season and year. Grey boxes indicate steps involving decisions based on statistical information.

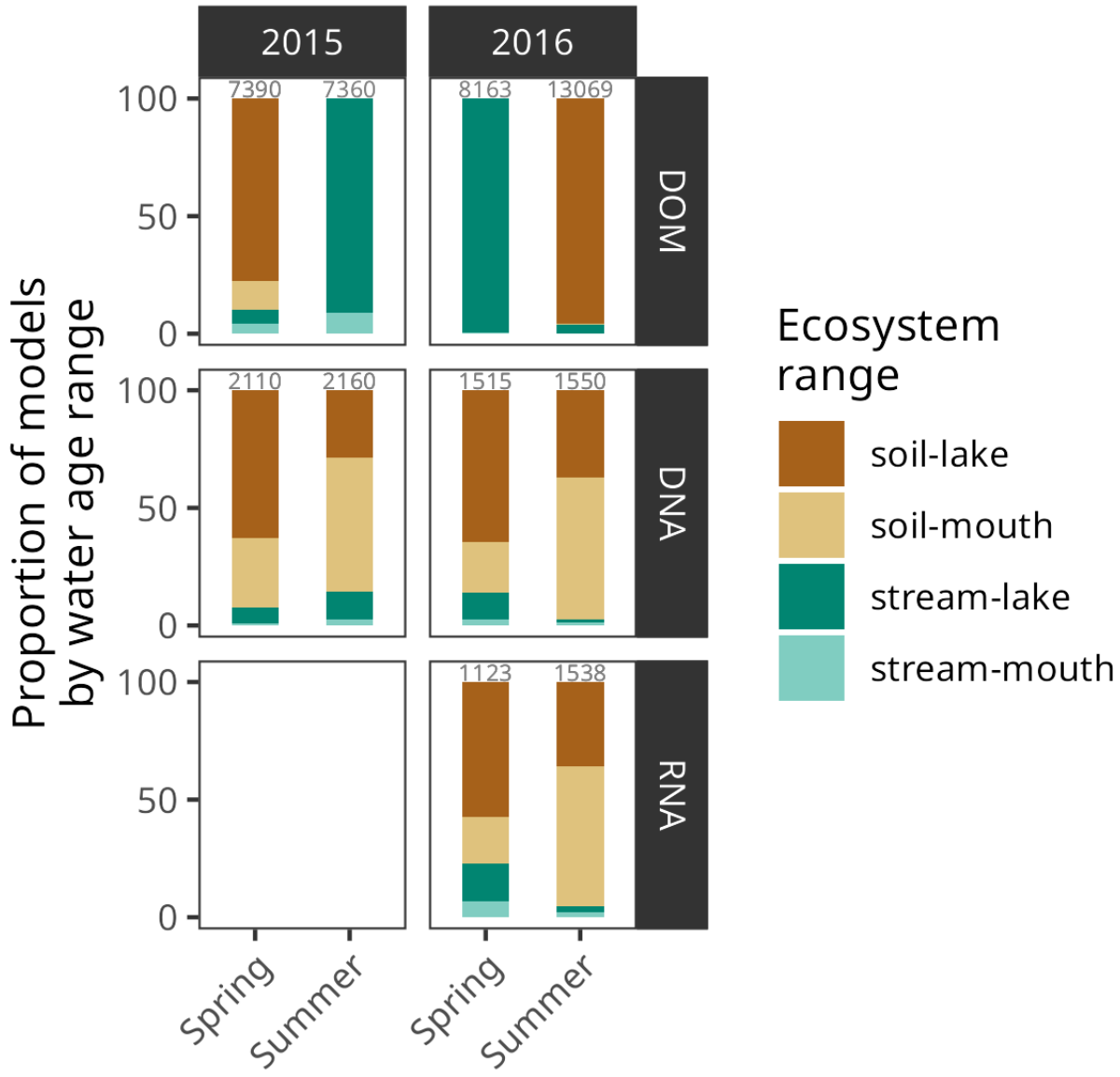

Figure S2.5: Distribution by ecosystem range for all spatial patterns. Soil includes soilwater and groundwater sites. Lake indicates any MF/OTU observed beyond the flow-weighted water age at the river mouth.

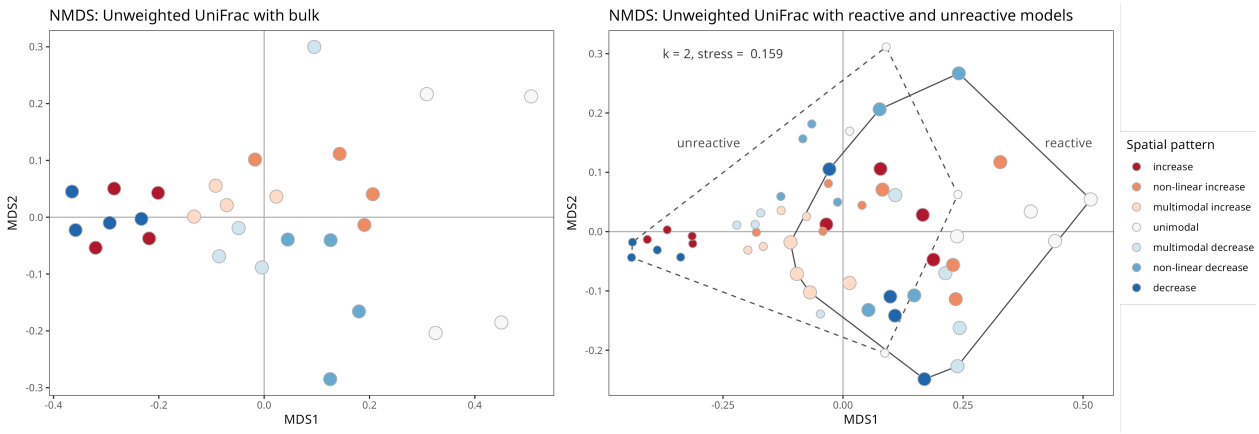

Figure S2.6: Unweighted UniFrac phylogenetic distance among spatial patterns in bulk and reactivity pools of microbial dataset. Non-metric multidimensional scaling (NMDS) of unweighted UniFrac distance on presence-absence transformed community matrix of microbial OTUs. Spatial patterns are distinguished by colour, reactivity pools (unreactive versus reactive) are depicted as different sizes in points as well as surrounded by polygons.

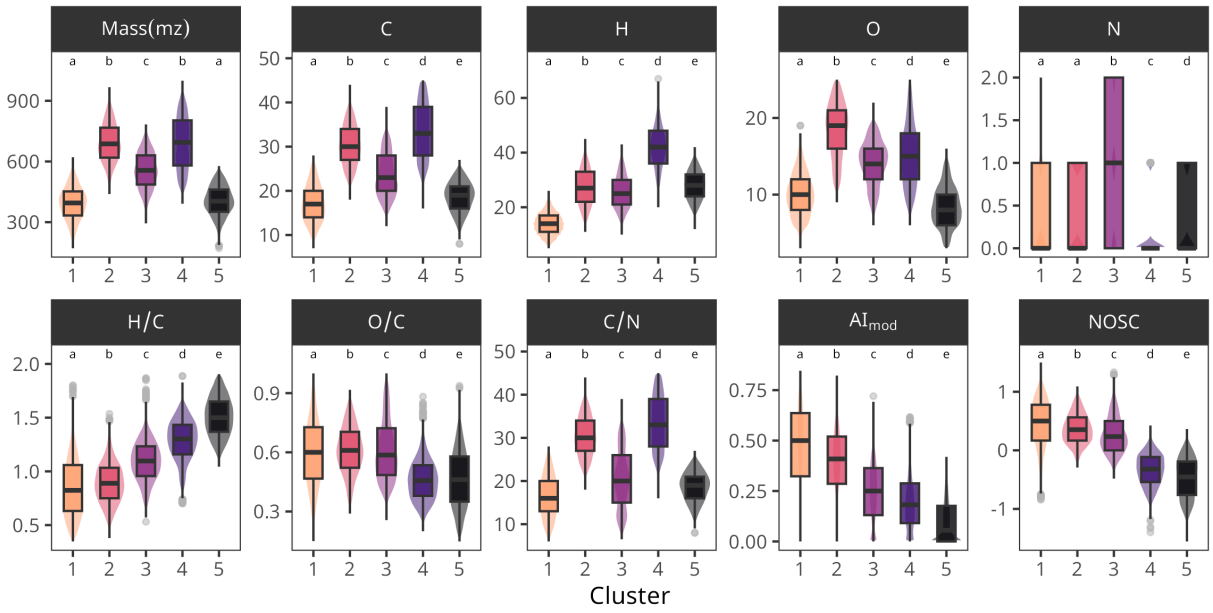

Figure S2.7: Chemical metrics used in hierarchical clustering analysis and their distribution among identified clusters. Given are chemical metrics such as the number of elements within a molecular formula (C = carbon, H = hydrogen, O = oxygen, N = nitrogen), mass (in *mz*), elemental ratios (H/C, O/C, C/N) and indicators of aromaticity (AI_{mod}) as well as nominal oxidation state of carbon (NOSC). Middle lines of boxplots represent the median, while the upper and lower hinges represent the 25^{th} and 75^{th} percentiles. Upper and lower whiskers expand to the largest and smallest value, respectively, no further than 1.5 times the inter-quartile range (IQR) from the hinge. Outliers are depicted as points. Clusters are identified as colours. The distribution of data is additionally depicted in the cluster colours around the boxplots.

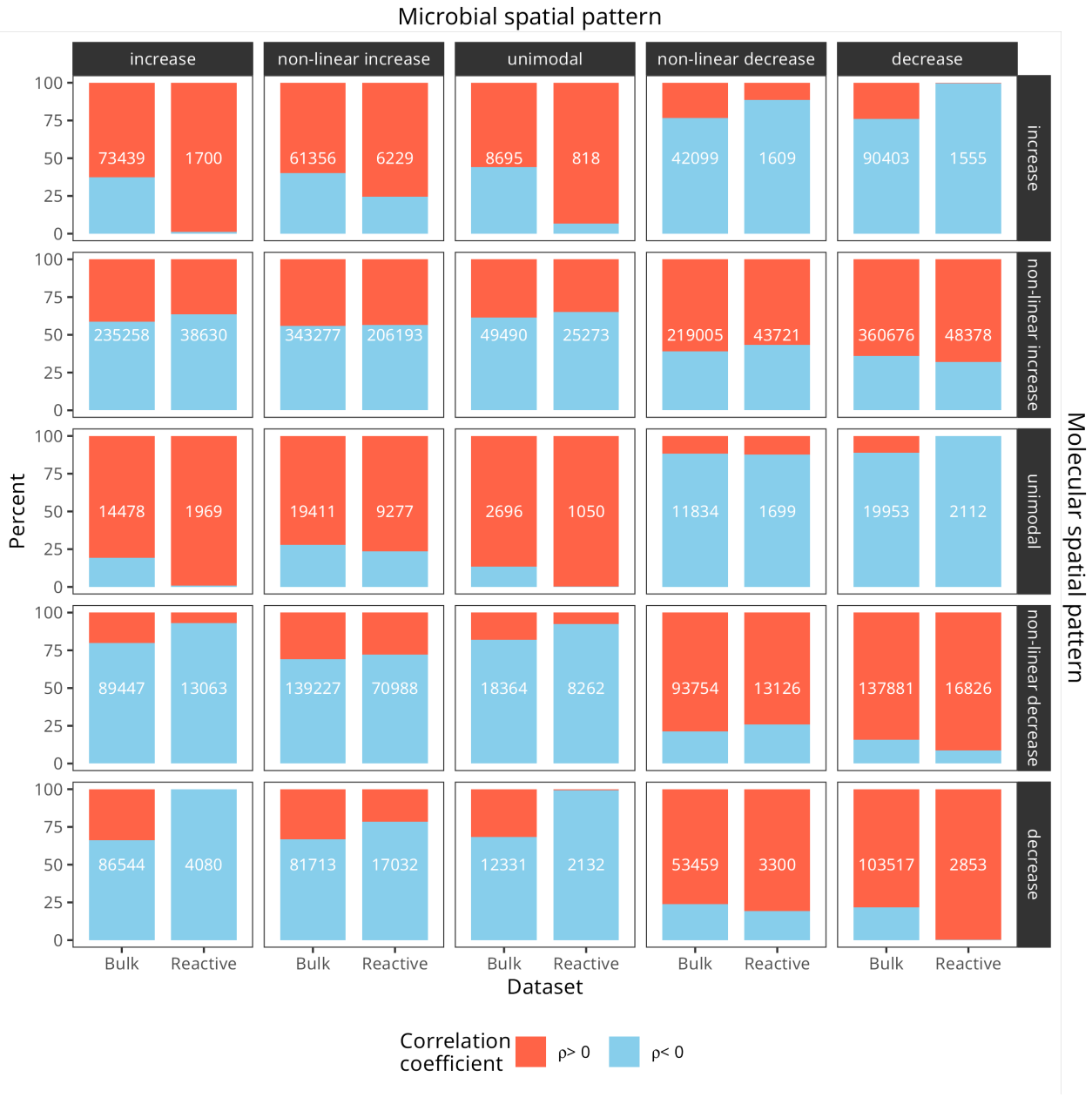

Figure S2.8: Proportion of significant positive and negative relationships between microbial and molecular spatial patterns. Percentages are given for the total number of correlations by pool (bulk versus reactive) and spatial pattern combinations.

2.8.4 Supplementary tables

Table S2.1: Number of overall (n) and unique microbial OTUs and DOM molecular formulae per campaign.

		Microbial		DOM	
Year	Season	n	unique	n	unique
2015	Spring	2,110	447	7,390	66
2015	Summer	2,160	482	7,360	92
2016	Spring	1,515	176	8,163	0
2016	Summer	1,550	238	13,069	4,030

Table S2.2: Averages and standard deviations of chemical indices for identified molecular formulae clusters.

Cluster	Mass (m/z)	H/C	O/C	C/N	AI_{mod}	NOSC
1	392.6 ± 86.5	0.9 ± 0.3	0.6 ± 0.2	16.5 ± 5.1	0.5 ± 0.2	0.5 ± 0.4
2	691.9 ± 102.7	0.9 ± 0.2	0.6 ± 0.1	30.5 ± 4.9	0.4 ± 0.2	0.4 ± 0.3
3	556.2 ± 95.3	1.1 ± 0.2	0.6 ± 0.2	20.5 ± 7.1	0.2 ± 0.2	0.3 ± 0.3
4	693.0 ± 136.8	1.3 ± 0.2	0.5 ± 0.1	33.4 ± 6.6	0.2 ± 0.1	-0.5 ± 0.4
5	402.7 ± 79.7	1.5 ± 0.2	0.5 ± 0.2	18.6 ± 3.6	0.1 ± 0.1	-0.5 ± 0.4

CHAPTER 3

EXPLORING LINKS BETWEEN BACTERIAL METABOLISM, DISSOLVED ORGANIC MATTER AND MICROBIAL COMMUNITY COMPOSITION ALONG A BOREAL AQUATIC CONTINUUM

Masumi Stadler[1], Ryan Hutchins[2] and Paul A. del Giorgio[1]

[1] Groupe de Recherche Interuniversitaire en Limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC, Canada.

[2] Department of Chemistry and Biology, Toronto Metropolitan University, Toronto, ON, Canada.

Keywords:

Microbial metabolism, microbial community composition, dissolved organic matter, aquatic network, seasonal dynamics, bacterial respiration, bacterial growth efficiency

N.B. References cited in this chapter are presented at the end of the thesis.

3.1 Abstract

Microbial metabolism is a key driver of global carbon cycles, yet its regulation across aquatic networks remains poorly understood. Here, we explore links between microbial metabolism, community composition, and dissolved organic matter (DOM) across streams, rivers, lakes, and reservoirs in a boreal watershed over three seasons. Results reveal that various dimensions of microbial metabolism are differently modulated by environmental and hydrological changes, as well as microbial community composition, with weakest associations found to DOM composition. The whole aquatic network was characterised by generally higher bacterial respiration (BR), extremely low bacterial production (BP) and consequently low bacterial growth efficiency (BGE) in spring. In contrast, summer and autumn exhibited elevated BP and BGE, particularly in fluvial and lake ecosystems, while BR remained stable in these systems across all seasons. Mid-molecular sized, highly oxygenated DOM compounds were associated with higher respiration and BGE, whereas bulk and active microbial community composition linked to BP. Substrate utilisation patterns (Biolog EcoPlates) varied across habitats and seasons, with lake and reservoir communities generally favouring slower, sustained consumption, whereas streams and rivers showed rapid utilisation of many substrates. These patterns correlated with seasonal shifts in microbial community structure and terrestrial DOM composition. Collectively, our findings emphasise the interconnected roles of microbial communities, DOM composition, and environmental drivers in shaping metabolic processes across the entire aquatic network, providing insights for biogeochemical modelling.

3.2 Introduction

Each year, approximately 5.4 petagrams of carbon (C) enter inland waters, mostly transported from land to aquatic ecosystems (Drake *et al.*, 2018). Through transformation, degradation, utilisation and trophic transfer, the microbial communities that inhabit these aquatic ecosystems collectively influence the processing and fate of this carbon and hence play a key role in the C budget on a global scale (Tranvik, 1992; Battin *et al.*, 2009). Often referred to as engines of ecosystems (Falkowski *et al.*, 2008), heterotrophic bacteria have yet to be effectively represented in earth system models despite major advances in molecular approaches (American Society for Microbiology, 2023; Lennon *et al.*, 2024). Microbial gene abundances (Rocca *et al.*, 2015) and community

composition (Bier *et al.*, 2015) often fail to predict biogeochemical processes better than environmental factors alone (Hall *et al.*, 2018). Whereas efforts to document microbial diversity (Locey and Lennon, 2016; Thompson *et al.*, 2017) or reconstruct genomes (Garner *et al.*, 2023; Borton *et al.*, 2025) across temporal and spatial scales have expanded greatly, there has been much less focus on quantifying heterotrophic microbial metabolism and its underlying drivers in recent years. Studies that have attempted to link microbial community composition metrics to process rates have had the tendency to focus on microbial processes that are tied to specific functional taxa (i.e., phylogenetic conservatism; Schimel *et al.* (2005)). For instance, methanogenesis and methanotrophy are primarily performed by narrow phylogenetic groups, enabling direct linkage of process rates to the presence and abundance of the specific microbial taxa involved (e.g., Reis *et al.* (2020); Bertolet *et al.* (2019). However, predicting broader metabolic processes that are shared among most microbial taxa using genomic data has proven a major challenge (Langenheder *et al.*, 2005; Severin *et al.*, 2014).

Bacterial respiration (BR) and biomass production (BP) are among the most fundamental microbial processes in all aquatic ecosystems. Both are key components of the C cycle, the former contributing to CO_2 emissions that are relevant on a global scale (Berggren *et al.*, 2012), and the latter building the basis of aquatic food-webs (Cole *et al.*, 1988). In turn these two processes represent different fates of the total organic matter consumed by aquatic microbial communities (often referred to as bacterial C demand, or BCD), and which represents a direct link between bacterial metabolism and organic carbon cycling in aquatic ecosystems. This is a complex and bi-directional link: microbes influence the organic matter pools in aquatic ecosystems through their selective consumption and subsequent transformation (Guillemette *et al.*, 2016; Logue *et al.*, 2016), and production of new substrates (Berggren *et al.*, 2019; Bello *et al.*, 2021). Yet, the size and nature of the ambient DOM pools influence microbial C consumption (Eiler *et al.*, 2003; Logue *et al.*, 2016; Casas-Ruiz *et al.*, 2016; Catalán *et al.*, 2021), community composition (Docherty *et al.*, 2006; Muscarella *et al.*, 2019) and various other aspects of microbial metabolism (Lennon and Pfaff, 2005; Fasching *et al.*, 2020; Rodibaugh *et al.*, 2020). Since DOM source and composition may also influence bacterial structure (Judd *et al.*, 2006; Kritzberg *et al.*, 2006), it is possible that the link found between microbial metabolism and DOM may be mediated rather by shifts in community com-

position. The links between the three (i.e., bacterial metabolism, community composition and DOM), however, may emerge only at levels that resolve their complexity in more detail. For example, most studies that link DOM composition to bacterial metabolism and community structures use either bulk (i.e., total DOC; Findlay *et al.* (1998); Eiler *et al.* (2003); Maranger *et al.* (2005); Jansson *et al.* (2008)) or coarse measurements of DOM quality (i.e., allochthonous versus autochthonous; Kritzberg *et al.* (2005); Attermeyer *et al.* (2013); Rodibaugh *et al.* (2020); Berggren *et al.* (2023); Mao *et al.* (2025)). Many studies that link microbial structure to metabolism have recently incorporated high-resolution approaches (Roiha *et al.*, 2016; Mao *et al.*, 2025) compared to traditionally used coarse genomic methods (Eiler *et al.*, 2003; Kritzberg *et al.*, 2006). However, very few have distinguished between the bulk and active community composition (Severin *et al.*, 2014). To our knowledge, there are no studies that have attempted to link the three components using high-resolution approaches to answer if a more detailed understanding of microbial community and DOM composition can explain variations in microbial metabolism.

Another often overlooked aspect of microbial metabolism are the various dimensions associated to it. Beyond BP, BR and BCD, the performance of aquatic microbial communities can be characterised by community-level physiological and bioenergetic properties such as substrate utilisation capabilities, bacterial growth efficiency (BGE) and bacterial growth rates, which provide additional information on how communities utilise and channel organic C resources (Comte and del Giorgio, 2011). Some of these metrics are measured from ambient samples, and hence, represent realised rates and properties in the environment. Other properties assess the community's potential to grow and produce biomass when ecological constraints such as predation and within-community competition for resources are reduced (i.e., increases in biomass and changes in protein synthesis as measured in filtered or diluted incubations). Many studies to date only look at one metabolism metric (Cole *et al.*, 1988; Judd *et al.*, 2006; Lindström *et al.*, 2010), or one C pathway (i.e., production metrics; Berger *et al.* (1995); Gasol *et al.* (2002)) at a time. Studies that have assessed BGE and BCD, on the other hand, have generally measured both BP and BR (Kritzberg *et al.*, 2006; Hall and Cotner, 2007; Lennon and Cottingham, 2008; Berggren *et al.*, 2009), yet very few studies evaluate multiple metabolism metrics beyond these classic variables (Comte and del Giorgio, 2009; del Giorgio *et al.*, 2011). The ensemble of metabolic properties provides additional infor-

mation on how communities utilise and allocate organic C resources, and these multiple aspects of microbial metabolism and physiology are not necessarily regulated by the same factors (del Giorgio *et al.*, 2011). Each metabolic component may link to microbial community composition, DOM and environmental factors differently, along environmental gradients, and the coupling or de-coupling of these metabolic dimensions to each other and to the various drivers provide insight on community-level constraints, and survival, resource acquisition, and bioenergetic strategies (Gasol *et al.*, 2008).

The various dimensions of microbial metabolism often do not vary drastically within the same ecosystem (Tranvik and Höfle, 1987; Kroer, 1993). Hence, it is often necessary to evaluate metabolism across large spatial (Benner *et al.*, 1995; Biddanda *et al.*, 2001; del Giorgio *et al.*, 2006) and/or temporal gradients (Roland and Cole, 2000; Baña *et al.*, 2020) or compare multiple types of ecosystems (del Giorgio and Cole, 1998), to establish how changes in metabolism may link to microbial community or DOM composition. Previous attempts that did not find strong links between microbial metabolism and DOM (Berggren *et al.*, 2023) or microbial composition (Comte and del Giorgio, 2011; Fonte *et al.*, 2013; Becker *et al.*, 2017) may simply have lacked sufficient environmental breadth and therefore large enough changes in metabolism to detect any such links. Interconnected streams, rivers and lakes represent very different hydrological, environmental and ecological scenarios for microbial communities, and often present large DOM gradients (Raymond *et al.*, 2016; Hosen *et al.*, 2021), and therefore, likely result in large shifts in bacterial metabolism as well. Hence, aquatic networks within watersheds at a regional scale represent an effective unit to examine shifts in bacterial metabolism due to their large environmental and hydrological gradients, and their coverage of multiple ecosystem types (i.e., lotic versus lentic). To date, comprehensive studies evaluating changes in the various dimensions of microbial metabolism across a diverse set of interconnected aquatic ecosystems and seasons are scarce (Comte and del Giorgio, 2009). To gain a holistic understanding of heterotrophic microbial metabolism in aquatic ecosystems and its potential links to DOM and microbial community composition, it is crucial to examine how the various aspects of microbial metabolism change across multiple hydrological, environmental, and biological conditions.

Here we present the results of an integrated, watershed-scale study of how the various aspects of aquatic heterotrophic microbial metabolism (i.e., bacterioplankton production, respiration, growth efficiency, growth rates, among others) relate to changes in environment, hydrology, DOM and microbial community composition along an entire aquatic network. The sampling design covered streams and rivers of Strahler orders 1-7 and lentic systems including lakes and reservoirs, and sampling was carried out over three seasons. To characterise microbial community composition, we sequenced DNA and RNA taxonomic markers to evaluate shifts in both bulk and active communities. A high-resolution approach was additionally utilised to examine DOM composition using Fourier-transform ion cyclotron mass spectrometry (FT-ICR MS). DOM and microbial composition analyses were complemented with measurements of microbial consumption patterns of various carbon substrates (Biolog EcoPlates). Our comprehensive study culminated in seven data matrices, namely environmental factors, hydrology, DNA, RNA, DOM molecular composition, carbon utilisation profile (CUP) and microbial metabolism metrics. Multivariate analyses allowed us to relate shifts across habitats and seasons within each matrix to the various dimensions of microbial metabolism and elucidated that the various dimensions of microbial metabolism link to microbial community, and DOM composition differently.

3.3 Material and methods

3.3.1 Sampling and environmental characterisation

Samples were taken within La Romaine river watershed (Côte-Nord, QC, Canada) in 2018 3.1. For detailed catchment characteristics refer to Stadler and del Giorgio (2022). In brief, the catchment has an area of $\sim$14,500 km^2, the main river extends to 475 km and is part of the boreal black spruce-moss bioclimatic domain. The river emerges as a Strahler order 6 from the largest lake in the northern part of the catchment and flows through a series of recently commissioned hydroelectric reservoirs, RO3, RO2 and RO1 (ordered from most upstream to downstream). In 2018, RO2, RO1 and RO3 were three, two and one year old, respectively. The river turns into a Strahler order 7 during its passage through RO3. Several smaller streams ranging Strahler orders 1-5 were sampled before their confluence into the main river. Three seasonal campaigns were conducted yielding in 48, 55 and 55 samples taken for DOM, DNA and RNA in spring, summer and autumn,

respectively (total n = 158; 3.1). A subset of 102 samples was used to determine respiration, 150 for bacterial production, 131 for carbon substrate utilisation, 84 for DOM biodegradability, and 77 for DOM photodegradability assays. Detailed processing and incubation protocols are outlined within the microbial methods section.

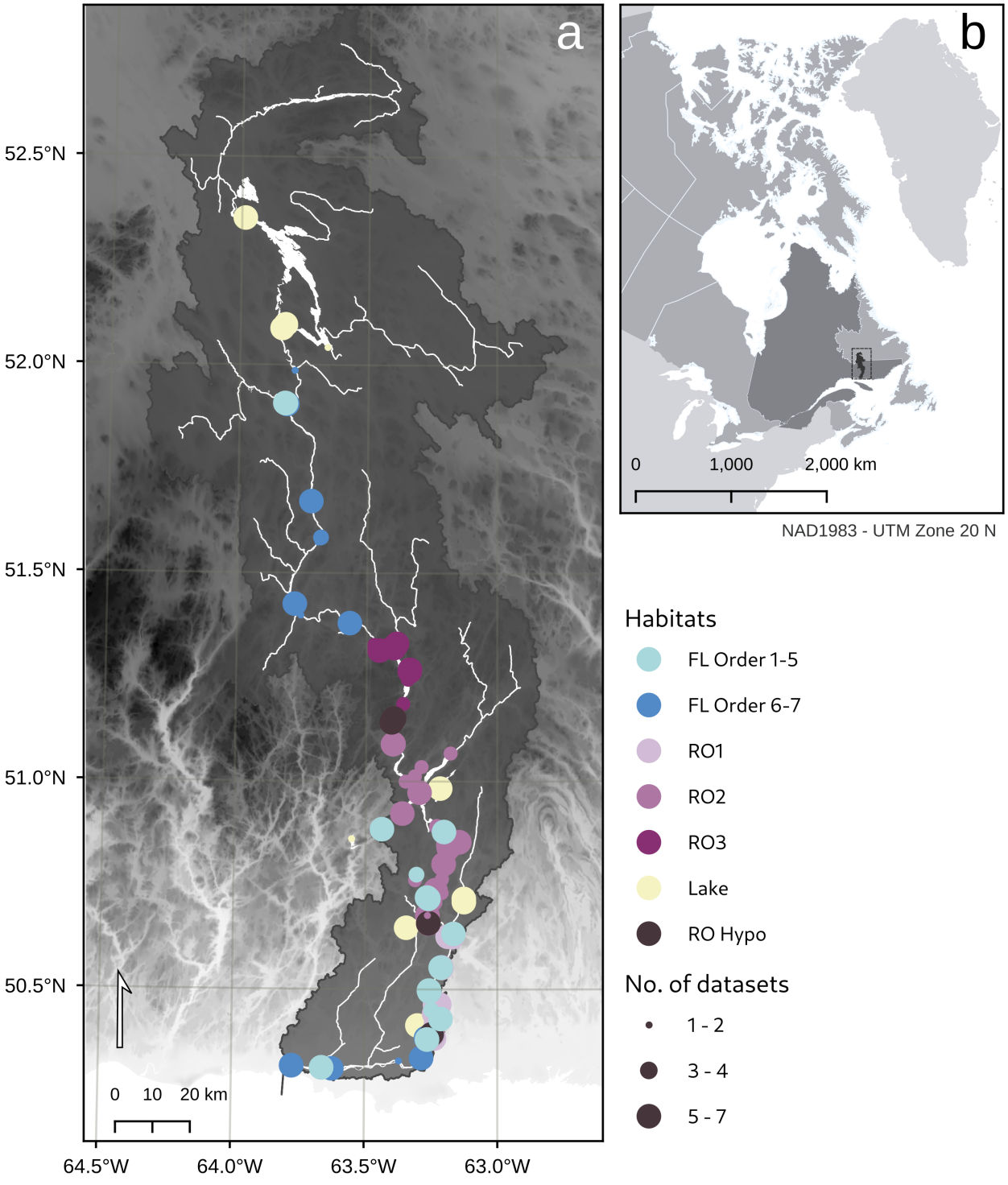

Figure 3.1: Map of La Romaine watershed with samples taken in 2018. Colours indicate habitat types and point size represents how many datasets were collected for each site. The datasets were: Environment, DNA, RNA, DOM, carbon utilisation profile and microbial metabolism processes.

Surface water temperature (°C), conductivity (μS cm^{-1}), pH, and dissolved oxygen (O_2) saturation (%) were measured at each sampling site using a multiparameter probe (600XLV2-M, Yellow Springs Instruments, OH, USA). Air temperature (°C), wind speed (m s^{-1}), relative humidity (%) and atmospheric pressure (Mb), measured at 1 metre above the water surface, were recorded with a handheld weather meter (Kestrel Meter 4000, PA, USA). Additionally, surface water samples were collected at each site in 5-litre polypropylene bottles for all downstream water analyses. These samples were processed on the same day and stored under refrigeration until analysis. Alkalinity (mEq L^{-1}) was determined via titration with HCl within two days of sampling. All other chemical analyses were conducted at the analytical laboratory of the Groupe de recherche interuniversitaire en limnologie (GRIL) at Université du Québec à Montréal, Montréal, Québec, Canada. For chlorophyll *a* (Chl*a*; μg L^{-1}) analysis, 1L of water was filtered in duplicates using GF/F filters (Whatman, Kent, UK), stored at –20°C and extracted with 90% ethanol, and quantified by spectrophotometry following acidification (measured at 665 and 750 nm, Ultrospec 2100 pro, Thermo Fisher Scientific Inc., Waltham, MA, USA; Lorenzen (1967)). Total phosphorus (TP; μg L^{-1}) and total nitrogen (TN; mg L^{-1}) concentrations were analysed in duplicates and determined through spectrophotometric methods after digestion with potassium persulfate (Ultrospec 2100 pro, Biochrom Ltd., Cambridge, UK; Wetzel and Likens (2000)) and alkaline persulfate (Flow Solution 3100, OI Analytical, College Station, TX, USA); Patton and Kryskalla (2003)), respectively. Ammonium (NH_4; μg L^{-1}) samples were analysed using a chloramine reaction with salicylate to form indophenol blue dye (Flow Solution 3100, OI Analytical, College Station, TX, USA; EPA Method 350.1). Dissolved organic carbon (DOC; mg L^{-1}) and dissolved inorganic carbon (DIC; mg L^{-1}) samples were filtered through 0.45 μm polyethersulfone cartridges (Sarstedt AG & Co., Germany), stored at 4°C in the dark, and analysed via high-temperature persulfate oxidation using a total carbon analyzer (TOC1010, OI Analytical, USA). Additionally, the water was analyzed for the isotopes ^{18}O and ^{2}H using a laser spectroscopic technique designed for liquid water analysis at the Light stable isotope geochemistry laboratory of the GEOTOP at the Université du Québec à Montréal (LGR DT-100 Liquid Water Stable Isotope Analyser, Los Gatos Research Inc., Mountain View, CA, USA).

3.3.2 Hydrological characterisation

To separate environmental drivers from hydrology, several hydrologic metrics were extracted or estimated for the watershed. A digital elevation model (18 x 18 m resolution) was retrieved from GeoGratis Canada (Natural Resources Canada, 2017) to define the watershed boundaries and compute metrics such as flow accumulation, flow length, Strahler order and pixel area using the Spatial Analyst Toolbox in ArcMap (v10.7.1, ESRI Inc., Redland, CA, USA). The stream network was delineated using a flow accumulation threshold of 3,000 pixels. To derive catchment area (CA), flow accumulation values were multiplied by the pixel area.

Discharge and velocity measurements were collected from streams of Strahler orders 1-4 within two headwater sub-watersheds (Petite Romaine and Bernard) during the years 2015/2016 and 2021/2022, using a 2-D Acoustic Doppler Velocimeter (FlowTracker, Sontek, San Diego, CA, USA). Additional hydrological measurements at stations located in rivers of Strahler orders 5-7 were performed periodically using a vessel-mounted Acoustic Doppler current profiler, with data provided by Hydro-Québec (Montréal, QC, Canada). High flow conditions were assigned to measurements taken in May and June, while measurements from other months were categorised as low flow conditions. By integrating data from 2010 to 2022 across Strahler orders 1-7 (n = 103), a model was developed to estimate discharge (Q) based on catchment area (km^2) under different flow conditions ($log\ Q_{high} = 0.26 + 14.03 \times log\ CA - 1.96 \times log\ CA^2$; $log\ Q_{low} = 0.26 + 14.03 - 0.28 \times log\ CA - 1.96 \times log\ CA^2$; R^2 = 0.88). Velocity (v) was also modelled using the same dataset using catchment area and flow conditions ($log\ v_{high} = -0.49 + 3.22 \times log\ CA - 2.17 \times log\ CA^2$; $log\ v_{low} = -0.49 + 3.22 - 0.29 \times log\ CA - 2.17 \times log\ CA^2$; R^2 = 0.46). For details refer to supplementary material in chapter 2 (SM 2.8).

Flow-weighted water age (FWWA) was calculated by adapting Peter *et al.* (2020) to incorporate both lentic and lotic systems into the concept of flow-weighted travel time. The procedure involved distinguishing lentic and lotic systems along the river network, estimating water residence time (WRT) for each ecosystem separately, cumulatively summing the residence time along the flow path and weighing the FWWA at each stream confluence by the discharge of the merging streams (details in supplementary material of chapter 2; SM 2.8). WRT for lakes was retrieved from the Hy-

droLakes database (Messager *et al.*, 2016) and was calculated for reservoirs using daily water level data (Hydro-Québec, Montréal, Canada). The overall ecosystem WRT was divided by the number of pixels along the main channel to retrieve WRT within each pixel (WRT_{px}). To estimate FWWA in lotic systems, WRT_{px} was calculated by estimating velocity by catchment area (see models in chapter 2), and dividing the length of the pixel by the modelled velocity in each pixel. The velocity, WRT and FWWA reflect seasonal changes, as well as local ecosystems (i.e., lentic versus lotic). To evaluate the average age of water in each ecosystem and sampling point, an isotopic approach using oxygen and hydrogen was implemented. The delta (δ) values for isotopic composition were calculated using the formula:

$$\delta^{18}O \; or \; \delta^{2}H = (\frac{R_{sample} - R_{standard}}{R_{standard}}) \times 10,000$$

Here, R_{sample} and $R_{standard}$ represent the ratio of heavy to light isotopes (^{18}O:^{16}O or ^{2}H:^{1}H) in the samples and standards (Vienna Standard Mean Ocean Water – VSMOW), respectively. The isotopic composition of the water was then used to calculate the deuterium excess (d-excess; Dansgaard (1964)):

$$d - excess = \delta^{2}H - 8 \times \delta^{18}O$$

D-excess serves as an index for water evaporation, with higher d-excess indicating less evaporated water, which is characteristic of more recent precipitation (Turner *et al.*, 2014). During evaporation, water becomes enriched in deuterium (resulting in low d-excess) compared to precipitation, which is less enriched in deuterium (resulting in high d-excess).

3.3.3 DOM characterisation using FT-ICR MS

For high-resolution characterisation of the DOM pool, samples were filtered using 0.45 μm polyethersulfone filters (Sarstedt, Germany) and subsequently stored in 125 mL polypropylene bottles at -20°C until they were ready for solid-phase extraction. The volume of the sample processed

through the PPL cartridges was adjusted based on the DOC concentration to ensure that 40 μg of carbon could be extracted. The extraction process involved rinsing the cartridges with Milli-Q water at pH 2 (acidified with concentrated HCl) and conditioning them with methanol overnight. A series of methanol and acidified Milli-Q water steps followed, concluding with the elution of DOM using methanol into pre-combusted amber glass vials (Dittmar *et al.*, 2008). The vials were then stored at -20°C in the dark until the DOM solid-phase extracts could be analysed at the Water Quality Centre at Trent University (Peterborough, Canada), using a 7 tesla Bruker SolariX XR Fourier transform-ion cyclotron resonance mass spectrometer (FT-ICR MS, Billerica, MA, USA).

External calibration was performed daily before analysis, employing electrospray ionisation (ESI) and sodium trifluoroacetate (NaTFA, 0.1 mg mL^{-1} in methanol) in negative ion modes. Data acquisition was controlled with Bruker *ftms Control* software (v2.1.0) across a mass range of m/z 200 to 1,000, and the data were exported using the Bruker *Compass DataAnalysis* software (v5.0). Each sample was injected at a flow rate of 120 μl h^{-1} to acquire 300 spectra scans. Molecular formulae were assigned to quality-inspected masses using the protocols outlined by the ICBM-OCEAN tool, a platform for processing DOM mass spectra (Merder *et al.*, 2020). Only molecular formulae listed in the theoretical molecular formulae list ranging the elemental composition of $C_{4\text{-}100}$, $H_{4\text{-}200}$, $O_{1\text{-}70}$, $N_{0\text{-}4}$, $S_{0\text{–}2}$, $P_{0\text{-}1}$ were considered. All samples were analysed using the same equipment, following the same protocol, within the same month. To ensure the quality of our samples, we excluded samples with sum intensities below 50 and summed peak intensities of isotopic isomers. Molecular formulae with O/C $\leq$ 1.2 and H/C $\leq$ 2.2 (Hawkes *et al.*, 2020), and double bond equivalents minus oxygen (DBE-O) $\leq$ 10 (Herzsprung *et al.*, 2014) were kept for downstream analyses. The contribution of each formula to the total composition of a sample was determined by rescaling all formula peak intensities so that the sum of peak intensities equalled 1.

To assess the degree of unsaturation in the samples, the modified aromaticity index (AI_{mod}) was calculated based on the molecular formulae (Koch and Dittmar, 2006). The metric aids in identifying aromatic structures (AI_{mod} > 0.5) or condensed aromatic structures ($AI_{mod} \geq 0.67$). Additionally, the nominal oxidation state of carbon (NOSC) was determined, where higher values signify compounds that are more oxidised and thermodynamically favourable (Riedel *et al.*, 2012).

Molecular formulae were grouped into clusters according to their chemical similarities across nine core structural and chemical metrics using unsupervised machine learning (i.e., hierarchical clustering). These metrics included mass (*mz*), the number of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulfur (S) and phosphorous (P) atoms, as well as AI_{mod} and NOSC, which were extracted for each formula. The data were scaled, a distance matrix was computed using Euclidean distances (*dist()* function, *stats* package (R Core Team, 2024)) and was then used for hierarchical clustering (*hclust()* function, *stats* package (R Core Team, 2024)). After evaluating the clusters with the *fviz_nbclust()* function (*factoextra* package, Kassambara and Mundt (2020)) and examining the characteristic distributions (Fig. S3.1), four optimal clusters were selected. The clusters were reorganised based on descending median NOSC values.

In addition, we conducted standardised biological and photochemical degradation experiments to determine a concentration of degradable DOC that was independent of varying incubation conditions, ensuring comparability across ecosystems and over time (Lapierre *et al.*, 2013). The experiments started on the same day as the sample collection. For biological degradation, water was filtered through pre-combusted 2.7 μm pore-size GF/D filters (Whatman) to preserve the natural bacterial community and used to fill a 500 mL acid washed glass bottle in duplicates. The water samples were then incubated in the dark for 16 days at a constant temperature of 20°C, in the dark, and samples were taken every 2 days and fixed with concentrated sulfuric acid (5N). DOC lability was estimated as the average slope of the change in DOC concentration as a function of the replicate incubations (mg C L^{-1} d^{-1}). A portion of the filtered water was used for photochemical degradation experiments, conducted in a solar simulator (Qsun XE1-BC, Qlab, FL, USA) under a standardised light dose in 24 mm diameter quartz glass tubes arranged horizontally. It is assumed that the intense UV exposure inhibited most bacterial growth and DOM utilisation within the tubes. The exposure time (24 hours) and temperature (24°C) were consistent across all experiments (Lapierre *et al.*, 2013). Photodegradability is presented as the rate of DOC removed in one day under these conditions (mg C L^{-1} d^{-1}).

3.3.4 Microbial community characterisation

Samples to enumerate bacterial abundance were fixed on the same day of sampling to a final concentration of 1% Paraformaldehyde (PFA) and 0.05% Glutaraldehyde (G) and stored at -80°C until analysis (del Giorgio *et al.*, 1996). Prior to analysis, samples were thawed and stained with SYTO 13 (Invitrogen, Waltham, MA, USA) diluted in dimethyl sulfoxide (DMSO; Stock concentration 2.5 μM). Samples were stained by adding 0.025% of the sample volume as SYTO 13 stock solution, and subsequently vortexed and incubated in the dark at room temperature for 10 min (del Giorgio *et al.*, 1996). Stained samples were analysed on an Accuri C6 flow cytometer (BD Bioscience, San Jose, CA, USA) with a sample flow rate of 14 μL min^{-1} using side scatter and green fluorescence (FL1-H). If a sample exceeded the event counts of 800 s^{-1}, the sample was diluted. Any measurement above 700 on the FL1-H channel was considered as background noise.

For DNA and RNA community characterisation, water samples were filtered using a 0.22 μm polycarbonate filter and a peristaltic pump. RNA filters were submerged in RNAlater (Qiagen, Hilden, Germany) and allowed to stabilise overnight at room temperature. All DNA and RNA samples were initially frozen at -20°C at the field station, then transferred to -80°C storage at the university laboratory until extraction. DNA and RNA were extracted using DNeasy and RNeasy PowerWater kits (Qiagen, Hilden, Germany) according to the manufacturer's protocols. RNA extracts were converted to cDNA using a high-capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). Subsequently, all samples were sent to the Génome Québec Innovation Center (Montréal, QC, Canada) for paired-end sequencing of the 16S rRNA V4 region. Sequencing was performed on a MiSeq platform (PE250, Illumina, San Diego, CA, USA) using primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3').

Detailed bioinformatic processing steps are described in Stadler and del Giorgio (2022). Briefly, primers were removed from the 16S rRNA DNA and cDNA (referred to as RNA) sequences using cutadapt (v1.18, Martin (2013)). Amplicon sequence variants (ASVs) were identified by processing the 16S rRNA amplicon sequences through the DADA2 pipeline (v1.14.1, Callahan *et al.* (2016)). Taxonomic classification was carried out using the DECIPHER package (v2.14.0, Wright (2016)), which applies the IDTAXA algorithm (Murali *et al.*, 2018) and utilises the GTDB database (Release 95, Parks

et al. (2018)). To address any minor discrepancies between DNA and RNA ASVs, as well as potential differences in 16S rRNA gene copies within a single genome (Větrovský and Baldrian, 2013), ASVs were clustered into operational taxonomic units (OTUs) based on a 99% similarity threshold. Any OTUs present only in RNA (referred to as 'phantom' taxa) were adjusted by replacing all instances where RNA > 0 and DNA = 0 with DNA = 1 (Bowsher *et al.*, 2019).

Observations, defined as the read count of an OTU within a sample, that were present in only a single sample across each habitat type, season, and nucleic acid type combination (i.e., singletons within a factorial combination) were deemed unreliable if they contained fewer than 10 reads. This method not only filters out singletons across the entire dataset but also removes those within each sampling campaign that had insufficient read counts to be considered reliable. Additionally, metagenomeSeq was employed to normalise and stabilise variations in library sizes using cumulative sum scaling (CSS) (Paulson *et al.*, 2013), resulting in what is referred to here as CSS reads.

3.3.5 Bacterial community metabolism

Respiration was measured for a subset of samples for which DNA and RNA had been extracted. Water from each site was incubated in 500 mL Erlenmeyer flasks sealed with silicone stoppers, parafilm and electronic tape to prevent gas exchange. For each site, five flasks were prepared, two to measure total respiration (unfiltered), two for bacterial respiration (filtered) and one to sample water for bacterial abundance and production. Water to measure bacterial respiration was filtered by 1 μm Pall A/E glass-fibre filters (VWR, Radnor, PA, USA, pre-combusted at 450°C for 4h) using a peristaltic pump. The respiration flasks were immersed in temperature controlled dark water baths. The temperature was kept constant with circulating water at mean ambient epilimnetic temperatures. O_2 concentration was measured with an optode system in a dark chamber, where oxygen-sensitive optical sensors are placed within the Erlenmeyer flasks and read by a fiber optic meter (Fibox 3, PreSens, Regensburg, Germany). Per reading point, a minimum of 10 consecutive measurements were recorded to ensure a stable value, while avoiding any temperature shifts during measurements. Temperature was measured within a separate Erlenmeyer flask incubated in the same bath and treated the same way as the samples prior to O_2 readings and which served as a control to correct the readings for variations in temperature due to manipulation (Marchand

et al., 2009). O_2 concentration in the flasks was measured twice a day, once in the morning, and a second time after 6h. An incubation was terminated as soon as a significant slope in the O_2 versus time relationship was observed (min = 4, max = 12 time-points). A respiratory quotient (RQ) of 1.2 was applied to convert O_2 consumption to CO_2 production (Berggren *et al.*, 2012) as follows:

$$R = (\beta_{O_2} \times -1) \times (1/M_{O_2}) \times RQ \times M_{CO_2}$$

where β_{O_2} is the slope of measured oxygen consumption in μg O_2 L^{-1} d^{-1}, M is the molar mass of O_2 and CO_2, respectively. Water was sampled during incubations at least once per day to measure bacterial abundance and production, as described below.

Bacterial biomass production (BP) was measured as rates of [^{3}H]-leucine incorporation following the protocol described in Kirchman (1993). Ambient samples as well as samples during the incubation were used to measure BP. Triplicate samples were exposed to [^{3}H]-leucine concentrations between 30 and 70 nmol L^{-1} during 1h at measured surface water or incubation temperature. Killed controls were pre-treated with 5% w/v of TCA (trichloroacetic acid). Leucine incorporation into protein was determined by precipitation with TCA followed by centrifugation, and scintillation counting (Beckman LS6500, Beckman Coulter, Brea, CA, USA). Leucine uptake (average, control-corrected) was integrated over time by dividing the incubation time and was converted to carbon production rates by using a leucine to C conversion factor of 3.1 (Kirchman, 1993). Whenever BP was measured, samples to measure bacterial abundances (BA) were taken as well.

We determined several other metabolic metrics: Bacterial growth efficiency (BGE) was computed as $BGE = BP/(BP+BR)$ (del Giorgio and Cole, 1998), and bacterial carbon demand (BCD) computed as BCD = BP + BR, where BP is the average bacterial production measured in the successive time points of the bacterial respiration incubations. Two metrics representing physiological traits of the community were additionally computed: 1) Specific production rate (SPR; μ) and 2) Specific respiration rate (SRR), where ambient (bulk) BP and BR determined in incubations were divided by bacterial density in ambient waters (BA), respectively. These metrics provide a perspective of

two key dimensions of cellular activity (biosynthesis and catabolism) at the community level.

Two alternative growth rates were also quantified from the incubations: 1) Growth potential (GP), and 2) Growth rate (GR). GP was estimated as the slope of the relationship between the leucine uptake (BP) at various time points as a function of time during the bacterial respiration incubations, and it represents the potential for growth of the community (del Giorgio *et al.*, 2011). GR was similarly estimated by regressing the change in BA (ln-transformed) during the incubation period and represents net population growth within the incubations (del Giorgio *et al.*, 2011). These two metrics reflect the response of the community to reduction in both bacterial abundance (i.e., dilution and reduction of competition) and in the abundance of bacterial predators (i.e., reduction in predation) resulting from the filtration of the samples prior to the respiration incubations (del Giorgio *et al.*, 2011).

3.3.6 Carbon substrate utilisation

Carbon substrate utilisation patterns were measured with Biolog EcoPlates (Hayward, CA, USA; Garland and Mills (1991)). These plates contain 31 different carbon sources in triplicates. 125 μL of unfiltered water were directly inoculated into the wells. Substrate use was measured as colour development from a tetrazolium dye that reacts to the CO_2 produced by bacterial respiration. The development of the dye was measured as absorbance at 595 nm with a microplate reader (Tecan GENios, Männedorf, Germany). One plate per sample was incubated in the dark at room temperature and the blank-corrected overall plate colour development (average well colour development, AWCD) was monitored two to three times a day after inoculation. The mean colour development of each compound was calculated as the blank-corrected mean absorbance of each substrate measured at the time when the AWCD was closest to 0.5 (Garland *et al.*, 2001). An AWCD of 0.5 was usually reached within 2-5 days, after which daily measurements were reduced to once per day until the AWCD stabilised and reached the maximum average rate of substrate utilisation.

In addition to the traditional way of extracting one-single value per plate by choosing an AWCD closest to 0.5 (Berggren and del Giorgio, 2015; Garland *et al.*, 2001; Ruiz-González *et al.*, 2017a), several additional parameters were extracted by fitting a logistic regression to the change in AWCD

over the time of incubation. The package *drc* (Ritz *et al.*, 2015) was used to fit a logistic regression with the equation:

$$c + \frac{d - c}{1 + e^{b \times (x - f)}}$$

where *b* is the slope at the inflection point, *c* the minimum (lower asymptote), *d* the maximum (higher asymptote), and *f* the 'dose' at the time of half response. Estimates for coefficients were obtained with the self-starting function *L.4()* (*drc* package, Ritz *et al.* (2015)). A four-parameter model was chosen to keep the model fit simple and assuming that the curve is symmetric around the inflection point. The second derivative was taken to find the point where C consumption started and ended, respectively.

$$c + \frac{d - c}{e^{bx - fb} + 1}$$

Four metrics were selected to describe the community's response to each substrate: 1) Slope at inflection point (IPS) indicating the speed of substrate consumption. 2) Response time (RT) representing how long the community took to start consuming the substrate. 3) Utilisation time (UT) indicating the length of substrate consumption period, and 4) maximum absorbance (MA) which describes how much of the substrate was transformed into CO_2. These four metrics were retained after a correlation matrix revealed a lack of correlation among them, ensuring that each metric represents a different aspect of the community in terms of substrate affinity and consumption capacity. To simplify the result interpretations, the 31 substrates were grouped into Alcohols (alc), Amines (ami), Esters (est), hydrophilic amino acids (aa+), hydrophobic amino acids (aa-), lipids (lip), organic acids with a respiratory quotient (RQ) > 1 (oa^), organic acids with a RQ $\leq$ 1 (oa_), polymers (poly), proteins (prot), and sugars (sug) following Berggren *et al.* (2012).

3.3.7 Statistical analyses

To simplify the sampling design, habitat categories were merged based on a principal coordinates analysis (PCoA) on the DNA of the microbial community composition. Initially, all stream, river and lake samples were categorised by their respective Strahler orders. Each reservoir had both surface and hypolimnion categories. A PERMANOVA analysis was conducted on the PCoA and statistical differences among habitats were evaluated. All streams/rivers of Strahler orders 1-5 as well as Strahler orders 6-7 were merged into two categories, since the orders within each category were not statistically significant from each other (Table S1). Only one lake category was retained, as there was no statistical difference by lake order. Reservoir hypolimnion categories were also merged, since they were not statistically different from each other.

Once habitat categories were established, we examined the similarities between habitats and season in 1) environmental variables (e.g., pH, DO, DOC, TN, TP, and others) and 2) hydrological metrics (e.g. catchment area (*CA*), velocity (*v*), discharge (*Q*) and others) using a principal component analysis (PCA; *prcomp()* function; *stats* package (R Core Team, 2024)). Only relevant and non-highly correlated variables were retained for the final analyses. Variables were z-scaled before conducting the PCAs. Furthermore, habitat and seasonal differences in microbial community composition (DNA and RNA), DOM composition as well as the composition of the four individual DOM clusters identified via hierarchical clustering were evaluated using principal coordinates analyses (PCoA) with Bray-Curtis distances (*cmdscale()* function; *stats* package (R Core Team, 2024)). The DNA and RNA abundances were Hellinger-transformed prior to analyses (*decostand()* function in *vegan* package (Oksanen *et al.*, 2019)). Finally, differences in carbon utilisation profile (CUP) and microbial metabolism metrics among seasons and habitats were evaluated using two separate PCAs. Both matrices were separately z-scaled before PCA analyses.

These analyses resulted in a total of 11 multivariate output matrices. For each multivariate analysis, habitat-season combination centroids in multivariate space were calculated by computing the averages of site scores along the first two axes for all samples in each habitat-season category. Each multivariate output was tested for habitat and seasonal differences using PERMANOVA analyses (*adonis2()* function in the *vegan* package (Oksanen *et al.*, 2019)) and tested for multivariate

dispersions (*betadisper()* and *anova()* function in *vegan* package (Oksanen *et al.*, 2019)). Both tests were run with 9,999 permutations and parallelised with the *parallel* package (R Core Team, 2024).

To further explore the relation of CUP to the DOM and microbial composition, the first two CUP axes were correlated to the first two axes of bulk DOM, individual DOM clusters, DNA and RNA PCoAs (Spearman's correlations; *cor.test()* function in *stats* package (R Core Team, 2024)). Additionally, the first two axes of PCoAs on bulk DOM, DOM clusters, DNA, RNA and CUP were correlated with habitat-season averaged lability and photodegradability measurements to assess whether DOM and/or microbial compositional changes are associated to shifts in DOM photo- and biodegradability.

Finally, in order to test the relation among various microbial metabolism metrics to all the various dimensions (i.e., environment, hydrology, resource, microbial community) that may affect these metrics, we correlated the first two axes of the microbial metabolism PCA to the first two axes of all the 10 other multivariate analyses. Additionally, we correlated the individual microbial metabolism metrics (i.e., BR, BP, BGE and others) to the same first two axes extracted from the 10 multivariate analyses to evaluate how the individual metabolism metrics relate to the various environmental and biological components.

For all the above-mentioned analyses, only significant correlations were evaluated and visualised. An α level of 0.05 was selected prior to all statistical analyses. All analyses were conducted in R (v4.4.2, R Core Team (2024)) and RStudio (v9.1.394, RStudio Team (2024)). For data cleaning and wrangling, *data.table* and *tidyverse* were used (Barrett *et al.*, 2024; Wickham *et al.*, 2019). *plyr* was used to parallelise the workflow (Wickham, 2011). *ggplot2* and *patchwork* were used to visualise the results (Wickham, 2016; Pedersen, 2024). The map was created with QGIS (v3.28).

3.4 Results

3.4.1 Environmental and hydrological gradients within the watershed

Principal component analysis (PCA) of environmental variables resulted in clear seasonal and cross-habitat patterns within La Romaine watershed (Fig. 3.2a). The first principal component (PC1),

accounting for 17.5% of the variation, clearly separated spring from summer samples, with autumn positioned in between (supported by PERMANOVA; Table S3.2). Spring stood out with its fast winds and high dissolved oxygen saturation, whereas summer was characterised by warmer air temperatures, higher nitrogen concentrations (TN, NH_4), and a surge in dissolved organic carbon (DOC) and chlorophyll *a*. In contrast, PC2, capturing 15.6% of the variation, revealed stronger habitat differences (supported by PERMANOVA; Table S3.2). Ecosystems with longer water residence times, such as lakes and reservoirs, were linked to warmer surface water temperatures , whereas streams and rivers were characterised by higher conductivity and total phosphorus (TP) levels. Although spring samples seemed to cluster closely together regardless of habitat, dispersions in multivariate space were not significantly different among seasons (Table S3.2). In contrast, there were significant differences in dispersions in multivariate space among habitats. In particular, streams and rivers were more dispersed in multivariate space in comparison to the surface reservoir sites, which were more clustered.

In contrast to the environmental variables, hydrological variables were more closely tied to habitat than to season, with an almost 40% difference in the variance explained between the two categories (Fig. 3.2b; PERMANOVA results Table S3.2). PERMANOVA results, however, did reveal both habitat and seasons to be significantly affecting hydrological differences. PC1 (40%) distinctly separated streams and lakes from reservoirs and larger rivers. Reservoirs and high-order rivers (Strahler order 6-7) were associated with large catchment areas (CA) and high discharge (*Q*), whereas the second axis (17.4%) distinguished lakes with their higher water residence times (WRT) from low-order, fast-flowing streams (Strahler orders 1-5). It is important to note here that seasonal differences are only reflected between spring and summer/autumn samples for the metrics discharge, velocity and water residence time because the models used to estimate these variables were built to only separately predict low and high flow (low flow including both summer and autumn). Variables that potentially were able to reflect differences across all seasons were only d-excess, and water residence time estimates of reservoirs. WRT of lakes were estimated on an annual scale and hence did not reflect any seasonality. These methodological constraints could explain the tight clustering of summer and autumn samples in the hydrological dataset, highlighting its limitations in capturing finer seasonal nuances.

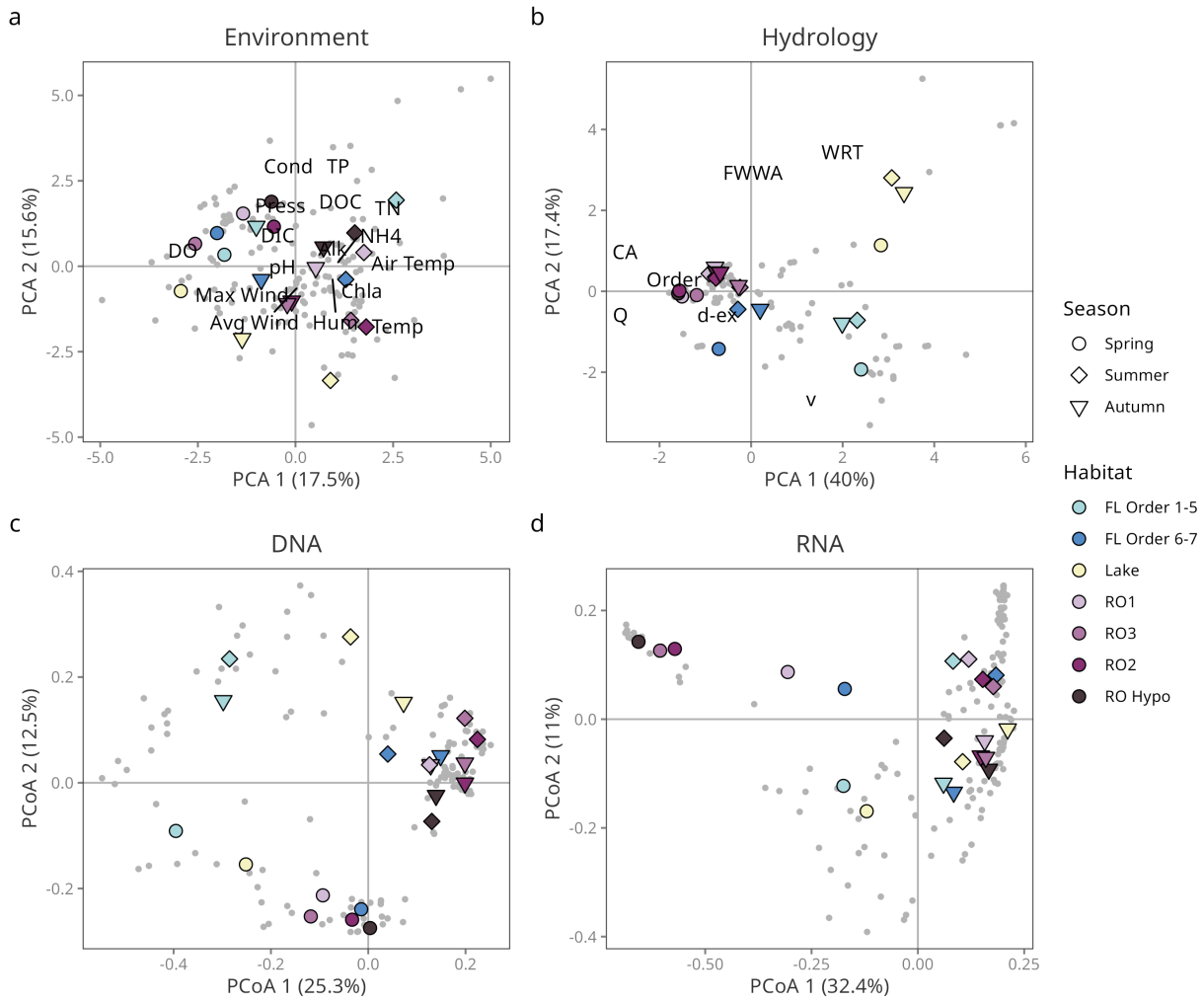

Figure 3.2: Habitat and seasonal differences in environmental variables, hydrology and microbial community composition. PCA analyses revealing environmental (a) and hydrological (b) differences among habitats across seasons. Bray-Curtis dissimilarity based PCoAs reveal clear differences between DNA (c) and RNA (d) community composition across habitats and seasons. Percentage of variance explained by the corresponding axes are given in the axis titles.

3.4.2 DNA and RNA seasonal successions along the aquatic continuum

The analysis of DNA composition, using Bray-Curtis dissimilarity and principal coordinates analysis (PCoA), revealed a distinct gradient of microbial succession across various aquatic habitats and seasons (Fig. 3.2c). The first axis, which accounted for 25.3% of the variation, shows a clear a progression from smaller streams to lakes, and ultimately to larger rivers and reservoirs. Mir-

roring the seasonal shifts in the environmental dataset, the second axis distinguished between spring and the combined seasons of summer and autumn, suggesting that the microbial taxa driving this succession change across seasons. This seasonal and habitat succession was statistically supported by a PERMANOVA analysis (Table S3.2).Notably, succession pathways varied by season. During spring, reservoirs more closely resembled smaller streams and lakes, with the progression culminating in higher Strahler order rivers and hypolimnetic reservoir samples. In contrast, during summer and autumn, succession concluded with reservoir samples, whereas larger rivers showed greater similarity to lakes and streams.

Similarly, RNA composition displayed minor habitat-related but much stronger seasonal variations, in relation to what was observed for DNA (PERMANOVA; Table S3.2),. Seasonal differences appeared across both axes, with the first axis (explaining 32.4% of the variation) separating spring from summer and autumn. The second axis (11%) further distinguished between summer and autumn samples, with the exception of lake and hypolimnion autumn samples, which clustered with other summer samples. The most pronounced distinctions occurred between spring and summer/autumn reservoir samples, except for RO1 spring samples, which aligned more closely with higher-order river samples (Strahler orders 6-7).

Together, these results suggest that microbial communities in both the bulk (DNA) and active (RNA) fractions undergo seasonal successions along the aquatic continuum, with RNA reflecting stronger seasonality and DNA showing stronger habitat differentiation. Although these patterns were supported by PERMANOVAs, multivariate dispersions differed significantly in DNA habitats and RNA seasons, complicating the distinction between among-group and within-group variation. Among DNA habitats, smaller streams and lakes showed the largest dispersions, indicating high microbial community heterogeneity within these habitat types. For RNA seasons, spring samples exhibited broader dispersions, suggesting that the active microbial community is more heterogeneous across habitats in spring, whereas summer and autumn samples generally had more similar active taxa across habitats.

3.4.3 Patterns in DOM composition among seasons and along the continuum

The dissolved organic matter (DOM) pool was categorised using hierarchical clustering, an unsupervised machine learning approach. Metrics representing the chemical composition (e.g., atomic number of elements) and character of each molecular formula (e.g., aromaticity and nominal oxidation state of carbon (NOSC)) were used to classify the molecular formulae into clusters. This process identified four statistically distinct clusters within the dataset, which represent an ascending order in mass, H/C, and C/N, and a descending order in aromaticity and NOSC (Fig. S3.1). The O/C ratio followed a descending order across clusters 2, 1, 3, and 4, with cluster 4 exhibiting the lowest O/C ratios. Each cluster also displayed unique distribution patterns in Van Krevelen space (Fig. 3.3a).

The overall DOM composition did not show a clear seasonal trend; rather, it was separated primarily by habitat type (Fig. 3.3b; PERMANOVA, Table S3.2). The first DOM axis (21.9%) separated RO3 and lower-order stream samples from all other habitats. While no strong seasonality was apparent across the entire multivariate space, a subtle pattern emerged in which summer and autumn RO3 samples were separated from spring RO3 and all lower-order stream samples along axis 2 (12.9%). This suggests that the spring DOM composition in RO3 was similar to that in small streams. The habitat differentiation observed in the bulk DOM composition seems to be mostly driven by clusters 2 and 3 (Fig. 3.3d-e), both of which have high oxygen content and moderate mass. Clusters 1 and 4, in contrast, had patterns that were distinct from bulk DOM and clusters 2 and 3 (Fig. 3.3b-f).

Clusters 2 and 3 showed minor seasonal variation within the majority of habitats. In lakes and higher-order rivers, spring and summer samples were more similar to each other in the composition of cluster 3 (Fig. 3.3e). Conversely, in reservoir samples, spring and autumn samples appeared more similar in cluster 3 composition. Within cluster 2, a more pronounced seasonal separation was evident in multivariate space: spring samples clustered together, whereas summer and autumn samples also formed a separate grouping, with the exception of lakes (Fig. 3.3d). Despite these visual trends, seasonal differences were not statistically significant in the bulk DOM and in clusters 1, 2, and 3, likely due to only a subset of habitats exhibiting seasonality while the statistical

test evaluates the entire dataset.

Clusters 1 and 4 showed the lowest variance explained among all clusters, with no clear visual differentiation by habitat or season. In cluster 1, the only samples diverging from the central trend were autumn lower-order stream samples and spring samples from lakes, RO2, and hypolimnion samples along axis 1 (14.8%). However, these differences were not statistically significant (PERMANOVA; Table S3.2). Interestingly, cluster 4, despite low explained variance, was the only DOM pool with statistically significant seasonal and habitat clustering (PERMANOVA; Table S3.2). Cluster 4 comprises high-mass, aliphatic-like molecules that are highly decomposed and less energetically favourable. These results suggest that while some DOM pools are influenced by habitat (e.g., clusters 2, 3, and 4) and season (e.g., cluster 4), one pool appears to be remarkably stable across ecosystems and seasons (i.e., cluster 1).

To evaluate if the bulk DOM molecular composition, the composition of individual DOM clusters, and microbial community composition (DNA, RNA) related to bio- and photodegradation rates, we extracted the scores of the first two axes from the PCoAs conducted on each of the matrices, and correlated these to the corresponding habitat-averaged lability and photodegradability rates (Fig. S3.2). Habitat lability was significantly correlated to shifts in cluster 1, 2 and 3 (Cluster 3 PCoA2 is not shown; ρ = 0.48, *p* value < 0.05), DOM, and DNA composition. Habitat photodegradability was significantly correlated to compositional changes in cluster 1, 2, and 4, and also to DNA. It is noteworthy, however, that these correlations were mostly driven by shifts in lower-order stream DOM and DNA composition, coupled with very low lability and high photodegradability rates in those sites. The only relationships that remained significant after removing lower-order stream samples were cluster 2 (ρ = 0.5), DOM (ρ = 0.53) and DNA (ρ = 0.65) with lability. However, only DNA remained significant when RO3 was additionally removed (ρ = 0.71). Overall, the results indicate that cluster 2, DOM and DNA shifts are most likely to be linked with DOM lability, whereas photodegradability is likely related to cluster 2 and 4 to some degree and undergo similar shifts as DNA along the continuum.

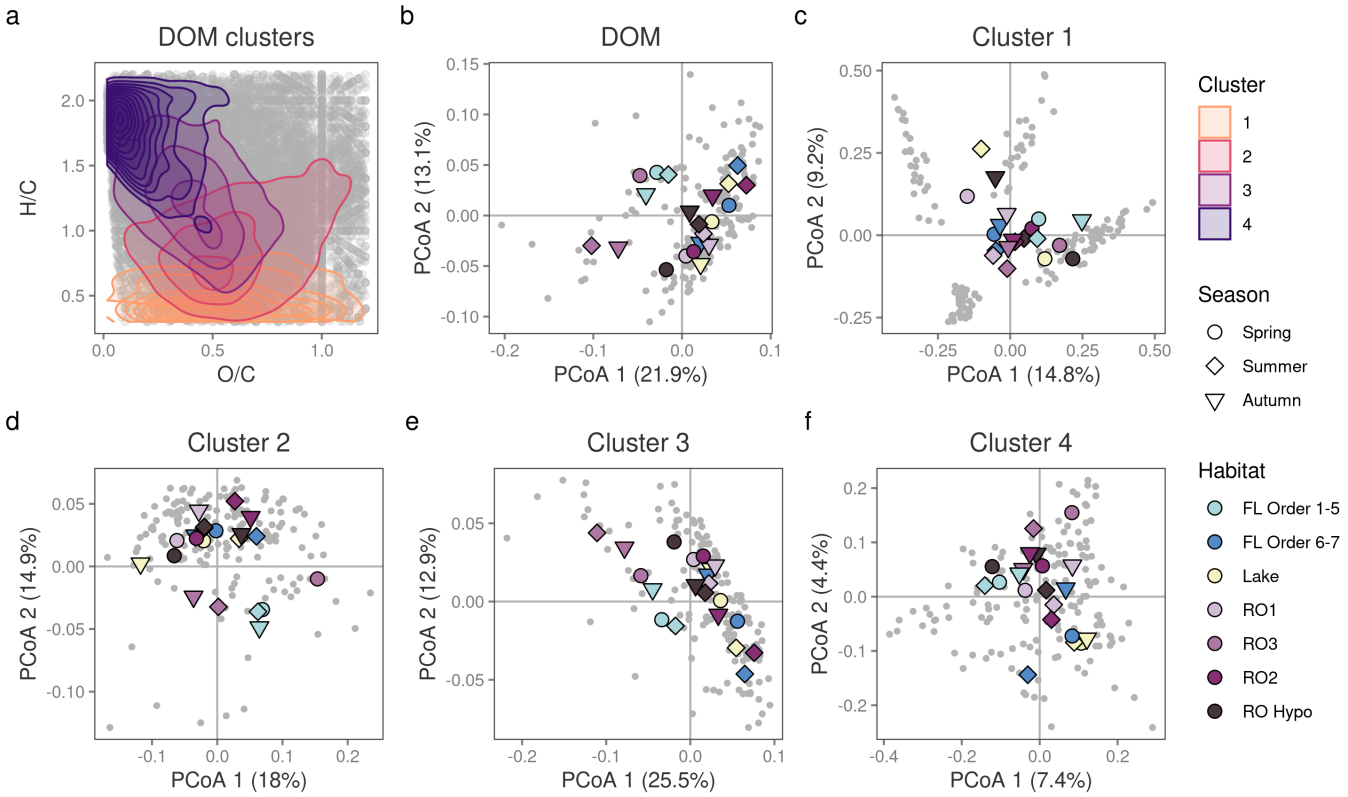

Figure 3.3: Comparison of bulk DOM composition and composition within DOM clusters identified via hierarchical clustering. a) Density plot highlighting the areas in Van-Krevelen space associated with each DOM cluster identified via hierarchical clustering. Bray-Curtis dissimilarity based PCoAs visualise differences in bulk DOM composition (b) and DOM clusters 1-4 (c-f) across sampled habitats and seasons. Percentage of variance explained by each PCoA axis are given in axis titles.

3.4.4 Seasonal and habitat specific shifts in microbial carbon utilisation profile

Carbon Utilisation Profile (CUP) enables us to evaluate the capacity of local microbial communities to degrade a fixed panel of substrates that are ecologically relevant. This yields a substrate utilisation profile for each community that reflects both a community capacity and the potential influence of ambient DOM on selecting these community functionalities. To understand how this carbon utilisation profile differs between habitats and seasons, we conducted a multivariate analysis on consumption metrics derived from the Biolog EcoPlates incubations. Specifically, we examined four primary metrics, all based on the colorimetric determination of the reduction of tetrazolium salts as an index of respiration (i.e., substrate utilisation) within each well: 1) the slope at the inflection point (IPS) of the consumption curve, which represents the rate of consumption; 2) response time (RT), which marks the onset of substrate consumption; and 3) utilisation time (UT), which indicates the duration of consumption for each substrate, and 4) maximum absorbance (MA), which was used as an indicator of the overall amount of substrate consumed. Our results indicate that CUP metrics capture both seasonal and habitat-specific variations in microbial substrate utilisation profiles (Fig. 3.4a), explaining 40.6% of the overall variation along the first two axes (PERMANOVA; Table S3.2). In spring, reservoirs, and higher-order rivers cluster together on both axes, while the summer and autumn samples of the same habitats cluster on the opposing diagonal side, indicating strong seasonal differences in their carbon utilisation profiles. Additionally, lakes and hypolimnetic reservoir samples show strong seasonal trends, where the seasons spread in opposing directions across the first two axes. In contrast, smaller-order streams cluster together along the first axis, indicating lower seasonal differences.

These habitat and seasonal differences in CUP are underlain by differences in which and how these substrates are consumed. Figure 3.4b shows how the individual consumption metrics drive the observed habitat and seasonal distribution of C utilisation profiles. Axis 1, which explained 30.8% of the variation, suggests that communities with high response times (RT), tended to sustain high utilisation times (UT) but overall lower consumption rates (IPS). In contrast, Axis 2, explaining 9.8% of the variation, captured mostly differences in the amount and type of substrate consumed (MA). As an example, alcohols were preferentially consumed in summer and autumn whereas amino acids, polymers and proteins were preferentially consumed in spring. Likewise, this axis distinguished

between habitat selectivity, with amino acids, proteins and polymers preferentially consumed in the hypolimnion of reservoirs, and little or no preference in small order streams.

These results suggest trade-offs in microbial consumption dynamics. Interestingly, these observed trade-offs between consumption rate, maximum substrate consumption, response time, and utilisation time were significantly correlated with specific components of DOM and to microbial community composition (Fig. 3.5). PCoA2 of bulk DOM (from Fig. 3.5b) and of DOM cluster 3 (from Fig. 3.5e) as well as PCoA1 of cluster 2 (from Fig. 3.5d), were associated with consumption trade-offs linked to habitat differences captured in PC1 of CUP (Fig. 3.5a-c). This suggests that DOM Cluster 2 and 3 (high O, low N, and medium mass) may be influencing the carbon utilisation strategies of the microbial communities. Furthermore, CUP PC2 was significantly correlated with PCoA2 of the community DNA (Fig. 3.2c), which captured seasonal distinctions between spring and summer/autumn communities (Fig. 3.5d). These findings imply that observed substrate preferences may be linked to microbial community composition, with spring communities being capable of consuming the largest amounts of any substrate.

3.4.5 Microbial metabolism and drivers along the continuum

Microbial metabolism provides a glimpse into multiple facets of microbial functioning across various environments. These processes are regulated by different combinations of microbial community assembly, microbial composition, environmental conditions, resource and nutrient availability and ecological processes such as competition and predation relationships. Hence, most microbial processes cannot be predicted by who is present and are therefore often referred to as 'emergent' community properties (Hall *et al.*, 2018). Variables that can be measured to characterise microbial communities can be broadly categorised in four major groups: metrics that capture microbial C pools (e.g., bacterial biomass); metrics that capture how much carbon is incorporated within the community (e.g., growth rate, bacterial production), and others that quantify how much carbon is released (e.g., respiration rate). The fourth category includes metrics that integrate both processes (i.e., bacterial carbon demand, bacterial growth efficiency) (Table 3.1).

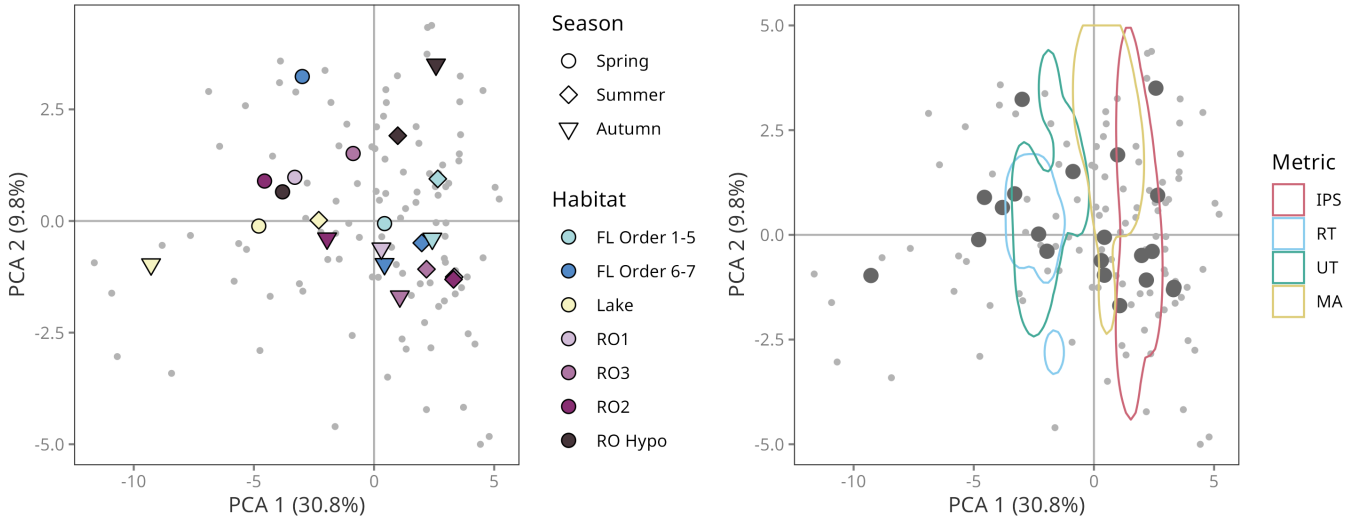

Figure 3.4: Seasonal and habitat differences in carbon substrate utilisation. a) PCA visualises habitat and seasonal differences in carbon utilisation profile metrics extracted from BIOLOG EcoPlates incubations. b) Density distribution of CUP metrics in multivariate space. Metrics used included: Slope at inflection point (IPS), response time (RT), utilisation time (UT) and maximum absorption (MA). Metrics were averaged by substrate groups listed in the methods section (3.3). Distribution of substrates in PCA space is depicted in Fig. S3.3

.

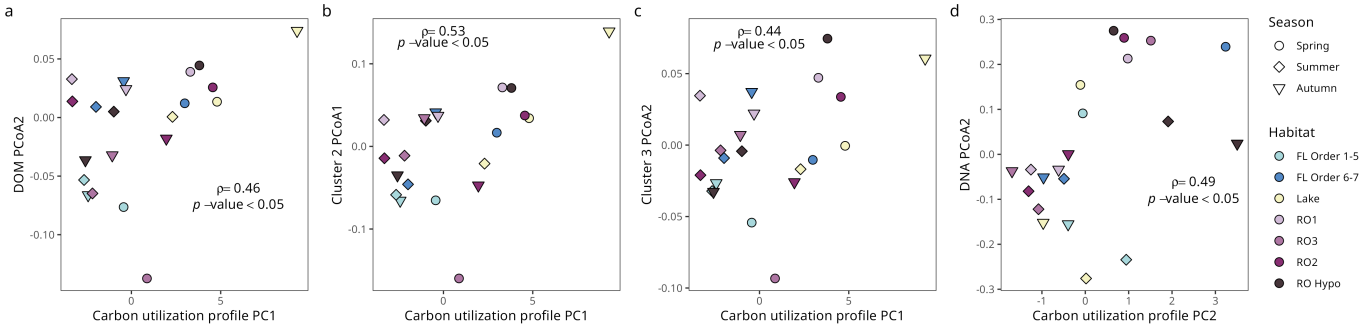

Figure 3.5: Spearman correlation of carbon utilisation potential (CUP) axes with DOM and microbial community composition. CUP axes were correlated with bulk DOM, cluster DOM, DNA and RNA PCoA axes and only significant correlations were plotted. DOM PCoA2 (a), Cluster 2 PCoA1 (b) and DNA PCoA2 were the only axes correlated with patterns in carbon utilisation potential. Correlation coefficients (ρ) and *p*-values are annotated within the plots.

Table 3.1: Microbial metabolism metrics and their ranges in boreal aquatic networks.

Metric	Abbr.	Unit	Description	Range
Bacterial abundance	BA	Cells L^{-1}	Ambient bacterial population	$1.2\times10^{9} \pm 3.9\times10^{8}$
Bacterial production	BP	μg C L^{-1} d^{-1}	Estimation of protein synthesis	6.2 ± 5.5
Specific production rate	SPR	μg C $cell^{-1}$ d^{-1}	Growth normalised by ambient population size	$5.6\times10^{-9} \pm 4.6\times10^{-9}$
Bacterial respiration	BR	μg C L^{-1} d^{-1}	Respiration attributable to bacteria	48.7 ± 29.7
Specific respiration rate	SRR	μg C $cell^{-1}$ d^{-1}	Respiration normalised by ambient population size	$4.5\times10^{-8} \pm 3.2\times10^{-8}$
Bacterial carbon demand	BCD	μg C L^{-1} d^{-1}	Total uptake and processing of carbon	54.9 ± 29.7
Bacterial growth efficiency	BGE	%	Balance in community respiration and production	23 ± 18
Growth rate	GR	μ d^{-1}	Abundance growth during incubations	0.10 ± 0.22
Growth potential	GP	μg C L^{-1} d^{-1} d^{-1}	Change in protein synthesis during incubations	5.3 ± 9.2
Total respiration	TR	μg C L^{-1} d^{-1}	Planktonic respiration	63.1 ± 32.6

A multivariate analysis on the entire dataset of microbial metabolic processes across the aquatic continuum reveals both habitat as well as seasonal patterns on how these communities' function (Fig. 3.6). There seems to be a slightly stronger effect of seasons (PERMANOVA: R^2 = 0.18, Table S3.2), with clustering of spring samples along the first PC axis capturing 40.5% of variation. Overall, spring samples showed higher growth rates, growth potential and higher specific respiration rates in comparison to summer and autumn samples. Whereas autumn samples had still relatively high growth potential, they were additionally characterised by higher BGE. Additionally, lake and low-order stream samples had high specific production rates, and bacterial production

in autumn. In contrast, summer was characterised by high respiration rates and bacterial carbon demand (BCD), and lower BGE.

Habitat effects were statistically significant as evaluated via a PERMANOVA (Table S3.2), however, patterns were more difficult to evaluate visually. Overall, reservoir samples seem to be slightly separated from stream, river and lake samples along the second principal component. Most reservoir samples were characterised by higher growth rates and growth potentials (PC2 > 0), while the majority of stream, river and lake samples were on the opposing side in multivariate space and characterised by higher bacterial abundance, and higher production rates.

It seems that the two principal component axes capture two distinct microbial trade-offs: the first axis describes a trade-off between respiration/BCD and BGE, indicating that when microbial (and total plankton) respiration rates and microbial carbon demand are high, growth efficiency tends to be low. Low BGE coincides not only with high total bacterial respiration, but also with high specific respiration rates, perhaps linking to scenarios of high maintenance costs or to the use of specific DOM pools.

Axis 2 (18.8%) contrasts the realised bacterial production and growth in ambient waters (BP and SPR) with the growth response of these same communities during incubations of filtered sample (GR and GP), which represents a form of community reset. Axis 2 suggests that habitats where ambient bacterial production and growth were high, measured the lowest potential growth responses (GR and GP) during *in-vitro* filtered incubation. In contrast, habitats with strong environmental or biological constraints to ambient bacterial growth (low ambient production and growth rate), had the strongest responses to the dilution and removal of predators generated during the incubation set-up. Overall, it appears that variations in BGE play a relatively minor role in determining BP and growth rates in these habitats. BGE is very orthogonal to BP, and BP does not appear to be a function of BCD either.

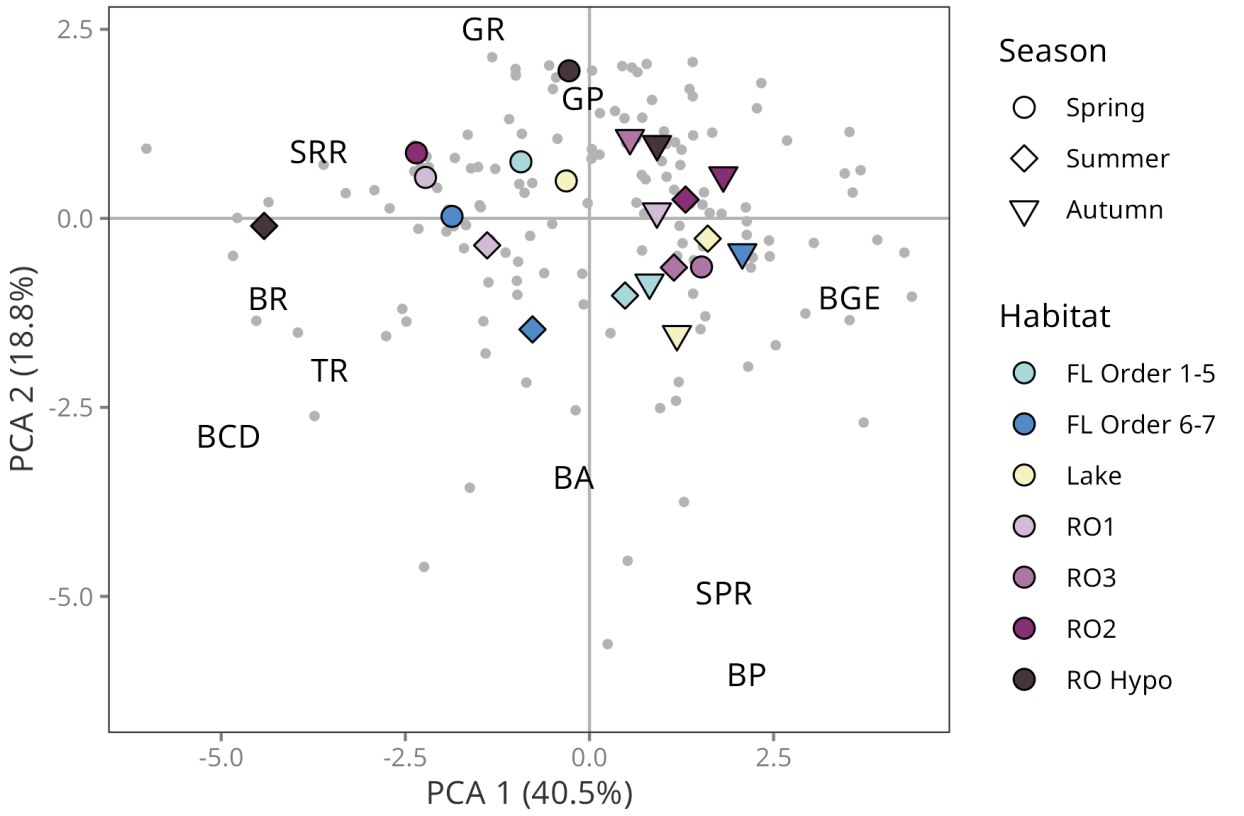

Figure 3.6: Habitat and seasonal differences in microbial metabolism. PCA analysis on all microbial metabolism metrics measured, where coloured points represent centroids of each habitat per season. Used metrics are bacterial abundance (BA), bacterial production (BP), specific production rate (SPR), bacterial respiration (BR), specific respiration rate (SRR), bacterial carbon demand (BCD), bacterial growth efficiency (BGE), growth rate (GR), growth potential (GP), and total respiration (TR).

These two dimensions of microbial metabolism captured by the two PCA axes represent emergent metrics of the microbial community. To evaluate whether these two dimensions were correlated to other habitat-season shifts in environment, hydrology, DOM and microbial community composition Pearson's correlations were conducted between the two microbial metabolism axes and all other multivariate analyses previously conducted. Axis 1, representing the BR/BCD-BGE trade-off was positively correlated to DNA PCoA2 and negatively correlated to the environmental PC2 (Fig. S3.4a-b), whereas axis 2, reflecting ambient and incubation differences in growth was positively correlated to DNA PCoA2 and RNA PCoA1 (Fig. S3.4c-d). These results indicate that the

emergent trade-offs captured in the microbial metabolism PCA, were significantly related to shifts in microbial community composition, and to the environment to a certain degree, but no clear correlation was found with DOM, hydrology or CUP.

To evaluate potential drivers influencing the individual microbial metabolic processes, we correlated the individual metabolic metrics to the previously described shifts (represented as scores in multivariate analyses) in environment, hydrology, DNA, RNA, DOM and CUP along the hydrological continuum and between seasons (Fig. 3.7). Bacterial abundance was only correlated with shifts in hydrology, indicating that bacterial density in a specific habitat may be linked to gradients in WRT (HYD2) as observed in lower-order streams and it gradually declines towards lakes. BP and SPR were high in smaller-order streams in spring and autumn, and low in summer. These two growth metrics were most strongly linked to seasonal shifts in DNA (DNA2) and RNA (RNA1), as well as network position as captured in (HYD1). SPR was further linked with ENV2, which captured habitat differences involving multiple environmental variables. These correlations collectively indicate that ambient bacterial abundance, production and growth (BP, SPR) reflect a combination of hydrology, network position and environmental conditions, and are linked to seasonal shifts in community composition and activity, as reflected in the patterns in community DNA and RNA, respectively. Bacterial respiration was not strongly related to any of the potential drivers. BR only related to ENV1, which reflects habitat differences in environmental conditions (mostly linked to conductivity and TP), whereas SRR was additionally linked to habitat shifts in DOM composition driven by changes from rivers/lakes/RO1-2 to streams/RO3. SRR additionally related slightly with shifts in DOM clusters 2 and 3, both of which showed a habitat/seasonal progression with the endpoints being summer RO3 and spring lake sites. These results suggest that microbial respiration is unrelated to microbial community composition but likely linked to substrates and nutrients that may be available for microbial consumption. In this regard, Bacterial carbon demand (BCD) followed the same patterns as bacterial respiration, and was related to a gradient of cluster 2 from streams/lakes to rivers/reservoirs, and it was further linked to habitat differences in conductivity and TP. These patterns suggest that the overall carbon demand of microbial metabolism is associated to resource availability, and DOM cluster 2 may play a major role in this regard. BGE, on the other hand, was strongly related to both DNA and RNA, and to a lesser extent to shifts in DOM

cluster 1 and CUP. These results suggest that BGE may be at least partly driven by seasonal shifts in microbial community composition, and reflecting shifts in the nature of the organic substrates consumed. The latter is further supported by the positive link that was found between BGE and CUP.

The metrics of growth *in-vitro* showed very different links to those that were found with *in-situ* measurements. Growth rate was linked to seasonal differences in the environment (ENV1), habitat shifts in RNA (RNA2) and DOM cluster 4 (Cl4_1), and most strongly associated to hydrology, which was strongly driven by network position and discharge (HYD1). Growth potential was only correlated to seasonal differences in environmental conditions (ENV1), habitat shifts in cluster 4 (Cl4_1) and to sites where substrate consumption slopes were high (CUP1). These metrics of microbial metabolism associated to *in-vitro* conditions appear to be strongly related to ambient environmental conditions and resource availability, which may reflect constraints in terms of growth limitation and carrying capacity. Finally, total plankton respiration, which we have added here to provide a more general context, was related to environmental and WRT conditions.

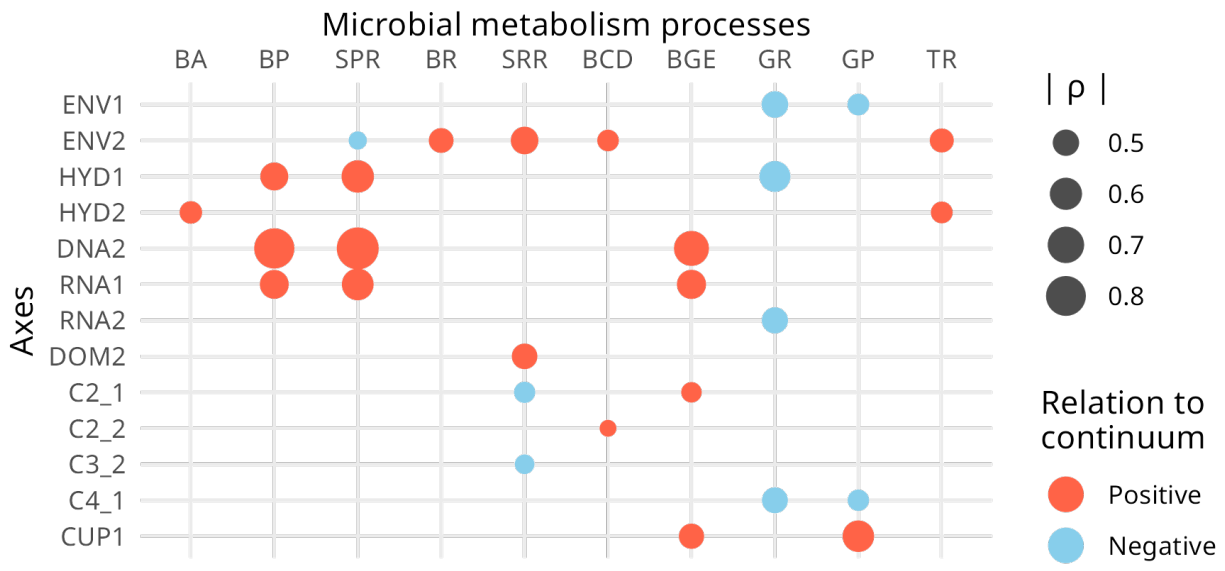

Figure 3.7: Correlation heatmap for each microbial metabolism metric to environmental, hydrological, microbial, dissolved organic matter, and carbon utilisation potential PCA or PCoA axes. Points are only visualised for significant differences in Spearman correlations ($p < 0.05$). If necessary, the direction of axes was inverted to represent an aquatic continuum from small order streams, large order rivers, lakes, reservoirs and hypolimnion. A unified direction of habitats across all axes enabled a clearer interpretation of correlation results. Axes that never showed a significant correlation to any metabolic variables are not shown.

3.5 Discussion

This study provides a comprehensive view into multiple aspects of microbial metabolism and their linkage to microbial and DOM composition across a wide range of aquatic ecosystems. By examining a wide range of ecosystems along the aquatic continuum, including small headwater streams, larger rivers, lakes and reservoirs, we observed that changes in microbial metabolism were most strongly associated to seasonal shifts in environment, hydrology and microbial community composition. We show how metabolism trade-offs measured in ambient waters and ecological constraints captured in *in-vitro* incubations emerge when the different microbial metabolism metrics are evaluated together. Although changes in the bulk DOM composition barely related to any metabolism metrics, shifts in specific portions of the DOM pool were associated to changes in specific respiration rates, bacterial growth efficiency, and growth potentials.

Along our extensive habitat and seasonal gradients, we observed a clear differentiation in microbial community composition by habitat and season in both bulk (DNA-based) and active (RNA-based) composition, respectively. Habitat differences were largest within bulk communities likely a result of varying assembly dynamics along the hydrological continuum (Stadler and del Giorgio, 2022), with smaller streams and lakes showing the largest variation in microbial community composition within the same ecosystem. Streams are likely influenced by strong terrestrial influx of microbes (Crump *et al.*, 2012; Besemer *et al.*, 2013) and wide hydrological shifts strongly shape biofilm- and planktonic-based bacteria depending on the season (Olapade and Leff, 2005; Besemer *et al.*, 2012; Battin *et al.*, 2016). Variations in microbial communities among lakes are likely a reflection of lake trophic state (Kolmonen *et al.*, 2011; Kraemer *et al.*, 2020), DOM composition (Crump *et al.*, 2003; Kritzberg *et al.*, 2006) and network connectivity (Yannarell and Triplett, 2004; Lindström *et al.*, 2006; Crump *et al.*, 2007) and position (Nelson *et al.*, 2009). Seasonality, however, influenced the active portion of microbial communities more than the bulk community. The active fraction of these communities across habitats was most different in spring, likely indicating stronger selection processes acting on the active fractions (Wisnoski *et al.*, 2020) in each habitat, whereas the active fractions were more alike in summer and autumn across habitats. These results highlight the dynamic changes of both bulk and active microbial communities across habitats (Crevecoeur *et al.*, 2022) and seasons (Stadler and del Giorgio, 2022) within the same aquatic network.

Contrary to the clear habitat and seasonal shifts in microbial community composition, bulk DOM composition did not show clear seasonal differences but rather differed by habitats. Lower order streams and a newly flooded reservoir (RO3) were most alike in their bulk DOM composition, likely indicating an influx of terrestrial material shared between these ecosystems (Planas *et al.*, 2005). This habitat differentiation was also reflected in two of the DOM clusters that we identified, namely clusters 2 and 3, which were characterised as olefinic, mid-molecular sized, lower N content, and higher oxygenation. These molecular formulae likely represent relatively fresh compounds that are mildly energetically favourable for microbial consumption. It is interesting to note that we observed a major shift in composition within clusters 2 and 3 from low order streams and RO3, which are heavily influenced by terrestrial inputs to rivers and lakes. Neither of these

pools was characterised as highly aromatic but rather partially overlap with compound groups commonly categorised as polyphenolic and unsaturated compounds with high (cluster 3) and low oxygen (cluster 2) (Hawkes *et al.*, 2020). On the one hand, these two clusters might represent mostly terrestrially derived compounds that are bio- and photodegraded or transformed (Kujawinski *et al.*, 2004; Hutchins *et al.*, 2017) along the aquatic continuum but maintaining their overall chemical similarity. Another possibility is that these pools are derived from multiple sources, including terrestrial and various aquatic that differ in chemical characteristics but whose leftover compounds tend to converge towards a common, mildly decomposed state, and thus their chemical fingerprint may become similar after partial degradation (Zark and Dittmar, 2018; Freeman *et al.*, 2024), which resulted in these compounds being clustered in the hierarchical clustering approach. Compositional shifts in clusters 2 and 3 correlated with DOM lability rates, with lability being generally higher in streams than in downstream habitats, suggesting that these pools are indeed terrestrially derived and tend to be both decomposed and transformed along the aquatic continuum, rather than generated within aquatic habitats. Shifts in cluster 2 also coincided with changes in photodegradability rates, which were also highest in streams, implying that cluster 2 additionally may be more susceptible to transformations via UV (Kujawinski *et al.*, 2004) and may contribute to photo-decay induced CO_2 emissions (Hutchins *et al.*, 2017). These results imply that shifts in these two pools coincide with an exhaustion of labile and photoreactive fractions of the DOM, suggesting the selective removal and transformation of a highly reactive pool of terrestrial compounds. Cluster 1 did not have clearly discernible spatial or temporal patterns either visually nor statistically, and despite this cluster being highly energetically favourable, and aromatic, this pool seems to remain remarkably uninfluenced by seasons and habitats. As such, this pool may represent DOM compounds that are either persistent in the aquatic environment (Kellerman *et al.*, 2018; Zark and Dittmar, 2018) or ubiquitously present due to an high influx (Wilkinson *et al.*, 2013). Lastly, cluster 4 exhibited both seasonal and habitat differentiation, indicating that the highly decomposed, large molecular mass, and energetically unfavourable pool seems to mainly separate between streams/reservoirs and rivers/lakes. Given that higher order river summer centroid was the most separated point from all other centroids, it may indicate that higher chlorophyll *a* concentration may coincide with shifts in this DOM compound pool. Higher primary production may imply higher degradation of such material on shorter timescales (Tranvik, 1998), and compounds

of Cluster 4 may be the less degradable leftovers (Zark and Dittmar, 2018).

Quantifying carbon utilisation profiles is an accessible way to obtain functional information of microbial communities (Miki *et al.*, 2018). However, the functional capacity of microbial communities may be determined by both, what community is present (Ruiz-González *et al.*, 2017b; D'Andrilli *et al.*, 2019) and what substrates the community has been exposed to (Berggren and del Giorgio, 2015). By capturing various community responses to the diverse substrates offered (i.e., consumption slope, response time, maximum absorbance), we were able to parse out how microbial community and DOM composition may affect their carbon utilisation capacities. Overall, it seemed that lakes in every season, and reservoirs, and higher order rivers in spring were characterised by microbial communities that responded slowly and consumed substrates for prolonged periods of time, whereas summer/autumn communities in reservoir and higher order rivers as well as in streams in any season responded and consumed substrates quickly. These differences in response and consumption time may be influenced by the ambient DOM composition as the axes capturing the utilisation differences mainly correlated to habitat shifts in DOM composition, in particular shifts in clusters 2 and 3. Differences in how long communities consumed varying DOM sources have been shown before, with algal DOM resulting in rapid consumption, whereas terrestrial DOM is consumed for longer (Pérez and Sommaruga, 2006). How much of each substrate was consumed in the CUP approach, and general substrate preferences mainly differed between spring and hypolimnion samples and all other surface samples, which was correlated with seasonal differences in microbial community composition. As such, the spring freshet may introduce a diversity of microbes (Crump *et al.*, 2012; Comte *et al.*, 2018; Caillon *et al.*, 2021) and DOM (Nelson *et al.*, 2009; Voss *et al.*, 2015; Kellerman *et al.*, 2014) and increase bacterial abundances (Battin *et al.*, 2004; Caillon and Schelker, 2020) in the aquatic network, which enables the utilisation of specific compounds and overall higher consumption of all substrates. In contrast, summer and autumn microbial communities may be less plastic than previously described (Comte *et al.*, 2013), and preferentially consume carbon substrates most similar to their resource in the environment (Battin *et al.*, 1999; Findlay, 2003). Furthermore, their generally lower bacterial abundances may have resulted in less substrate consumed compared to spring. Together, our functional profile results indicate that a detailed examination of substrate consumption dynamics can provide emergent properties

of community-level microbial C processing, beyond the traditional metric that captures an overall capacity of substrate utilisation (i.e., AWCD; Garland *et al.* (2001); Comte and del Giorgio (2010); Ruiz-González *et al.* (2017a)). These emergent properties were captured by the two PC axes, reflecting differing influences of past DOM selection processes and underlying community structures on the communities' capability to consume C substrates of various availability and complexity.

Microbial metabolism metrics provide insight into the intricate relationships between nutrient limitation (Smith and Prairie, 2004) and carbon allocation (Jansson *et al.*, 2006; Berggren *et al.*, 2007; Hall and Cotner, 2007) of microbial communities in the environment. When we examined the various microbial metabolism metrics together, we observed a clear pattern by season and habitat. In general, spring microbial communities had higher respiration rates and carbon demands, which resulted in lower BGE. Compared to many other studies where BGE was mainly modulated by BP (Kritzberg *et al.*, 2005; Berggren *et al.*, 2007; Amado *et al.*, 2013), their overall relation was relatively weak in our dataset and a relationship was only observed in spring or when individual habitat types were considered (Table S3.3 and S3.4). Across the entire dataset, BGE was rather strongly negatively correlated with BR, where increased BR led to lower BGE. This negative relationship was present in every season; however, it was strongest in spring (Table S3.4). These results indicate that relatively higher input of terrestrial material increases bacterial respiration rates, carbon demand and specific respiration rates, in spite of lower temperatures. In contrast, higher temperatures, TN, and Chl*a* concentrations may increase bacterial growth efficiencies and bacterial production to a certain degree. Previous studies observed a tendency for microbial communities to allocate the consumed carbon towards respiration when there is phosphorous limitation (Smith and Prairie, 2004; Jansson *et al.*, 2006; Godwin and Cotner, 2015), or when primarily autochthonous dissolved organic matter was consumed (Guillemette *et al.*, 2016). Neither of these factors are likely to contribute to the higher respiration found in our study, as TP concentrations and terrestrial DOM contributions were highest in spring. Interestingly, BR was relatively similar across streams and there was no relationship with BGE (Berggren *et al.*, 2007), hence, the strong association between BR and BGE was only found due to the various ecosystems sampled. BGE was similarly regulated by variations in BR in two network-scale studies (Becker *et al.*, 2017; Rodibaugh *et al.*, 2020), where higher temperatures and increased DOM allochthony

were found to be the main drivers, respectively. These studies were conducted in very different climatic regions (sub-tropical and arid), and the drivers underlying the same BR-BGE patterns may not necessarily be the same. Although our potential temperature effects on BR do not align with previous findings of higher BR at elevated temperatures (Jansson *et al.*, 2008; Hall and Cotner, 2007; Becker *et al.*, 2017), positive relations between allochthonous DOM inputs and BR (Lennon and Pfaff, 2005; Rodibaugh *et al.*, 2020) coincide with our findings in boreal regions. Increased BP rates were found to correlate with Chl*a* concentration (Rodibaugh *et al.*, 2020), which are in line with our results. Increased phytoplankton production links to bacterioplankton through the production of fresh autochthonous DOM (Crump *et al.*, 2003), however, such positive correlation and consequent bottom-up effect on microbial metabolism may depend on the nutrient availability such as nitrogen (Le *et al.*, 1994).

Although shifts in BP, BR and BGE have often been linked to changes in DOC quality (Findlay *et al.*, 1998; Crump *et al.*, 2003; Judd *et al.*, 2006; Jansson *et al.*, 2008; Lennon and Cottingham, 2008; Berggren and del Giorgio, 2015), we did not find any strong link between the ensemble of metabolism metrics with DOM. Our results from correlating the microbial metabolism PCA axes with other multivariate axes indicate that the relationships between microbial metabolism metrics are not directly influenced by DOM composition, but rather by the environment (i.e., temperature, TP, TN, Chl*a*) and microbial community composition. However, from the correlation analyses of individual metabolism metrics, we observed that changes in the DOM components likely of terrestrial origin did significantly but weakly link to specific respiration rates. In contrast, BP and specific production rates correlated with shifts in hydrology and both bulk and active microbial community composition. Together, these results imply variation in the regulation of different metabolic dimensions, where higher BR is modulated by shifts in terrestrial DOM, higher nutrient availability and DOC concentrations (i.e., seasonal variations), and BP linked to shifts in ecosystem properties such as network position, as well as habitat and seasonal shifts in bulk and active microbial community composition, respectively. This may imply that during the spring freshet, the entire aquatic network receives higher nutrient loadings, DOC concentrations and fresh terrestrial dissolved organic matter (Voss *et al.*, 2015; Burd *et al.*, 2018) that is rapidly consumed and respired by the microbial community, leading to reduced growth efficiencies (i.e., bacterial splurging). Once

the nutrient and fresh DOM supply becomes depleted, however, bacteria become more conservative in their carbon allocation (Jansson *et al.*, 2006), retaining more of the carbon consumed in bacterial biomass than respiration. This shift towards higher BP and BGE may potentially result from the community composition shifting towards microbial taxa (Eiler *et al.*, 2003) that have lower maintenance costs and are able to allocate more C towards biomass (Ramin and Allison, 2019; Muscarella *et al.*, 2020). The highest specific growth rates were found in streams and lakes in summer and autumn, and most samples of both habitats were located relatively early in the aquatic network. These results may indicate that headwater habitats have large influxes of a diverse pool of microbes (Crump *et al.*, 2012; Besemer *et al.*, 2013) that allow for the selection of taxa that are able to maximise the utilisation and consumption of available DOM resources, rather than the actual allocation of this C to biomass, since there was no relationship between growth rate and BGE in these habitats.

Finally, measurements of changes in growth rates and protein synthesis (i.e., GR and GP), additionally provided us with insights into how communities respond when diluted, and predators are reduced (Pomeroy *et al.*, 1994). Interestingly, *in-vitro* growth was most strongly upregulated in spring and autumn samples of reservoirs, in particular within the deeper hypolimnetic sites. These findings contrast previous results that found increased BGE with growth potentials (del Giorgio *et al.*, 2011). Increases in *in-vitro* growth related to shifts in the active microbial community composition and DOM composition of a relatively decomposed, energetically unfavourable pool, for reasons that are unclear. These results may also indicate that ambient reservoir microbial community production and growth is constrained by top-down effects such as predation (Pradeep Ram *et al.*, 2024).

In summary, our results highlight the variation of microbial metabolism that emerges when the entire aquatic network is considered. Although BGE has often been found to be modulated by variations in BP within single ecosystems, when examined across multiple habitats and seasons, it became clear that variations in BR influence BGE at the network scale. BR was mainly linked to enhanced nutrient and carbon loadings in spring, whereas specific respiration rate was additionally linked to shifts in DOM composition, in particular with DOM of terrestrial origin. Although BGE

was also weakly linked to shifts in terrestrial DOM, it was more strongly linked with shifts in bulk and active microbial community structure, which was most evidently reflected in the link of microbial community composition to production rates (BP, SPR). Similarly, carbon utilisation profiles were linked to both shifts in community composition as well as ambient terrestrial DOM composition, indicating how both resource and consumer structure can determine substrate affinities and the speed, duration and amount of substrate consumption patterns. Our study elucidates how varying bacterial metabolism metrics are differently modulated by DOM and microbial community composition. Given the recent scarcity of literature evaluating microbial metabolism, we are in need of more studies examining variations and drivers of the multiple dimensions of microbial metabolism across inland waters.

3.6 Acknowledgements

We would like to begin by acknowledging that the land on which the field work was conducted is on the traditional and unceded territory of the Innu community (Nitassinan), and the laboratory work in Montréal was conducted on the land of the Kanien'keha:ka people, as well as of the Métis people. We are especially grateful to Alice H. Parkes, Annick St-Pierre, and Serge Paquet, who maintained and oversaw the La Romaine project over the years. Collection and analysis of all variables would have not been possible without the support from members of the CarBBAS team including great contributions from undergraduate students. Therefore, we would like to thank Felipe Rust, Alexandre Ducharme, Roy Nahas, Rachelle Sauvé, Jean Christophe Sicotte-Brisson, Gabriel Bastien-Beaudet, Étienne Gautheir Dufour, Nathalie Larochelle, Joan P. Casas-Ruiz, Mathilde Couturier, Pascal Bodmer and Sophie Crevecoeur. We would also like to thank Katherine Velghe and Maryline Robidoux for laboratory assistance. This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Québec.

3.7 Data availability

The raw 16S rRNA gene sequences, both DNA and cDNA will be made available at the public NCBI Sequence Read Archive (SRA) as part of the BioProject PRJNA693020. The code will be made

available on Github after successful publication and both code and processed microbial data will be separately archived on Zenodo.

3.8 Supplementary Information

3.8.1 Supplementary figures

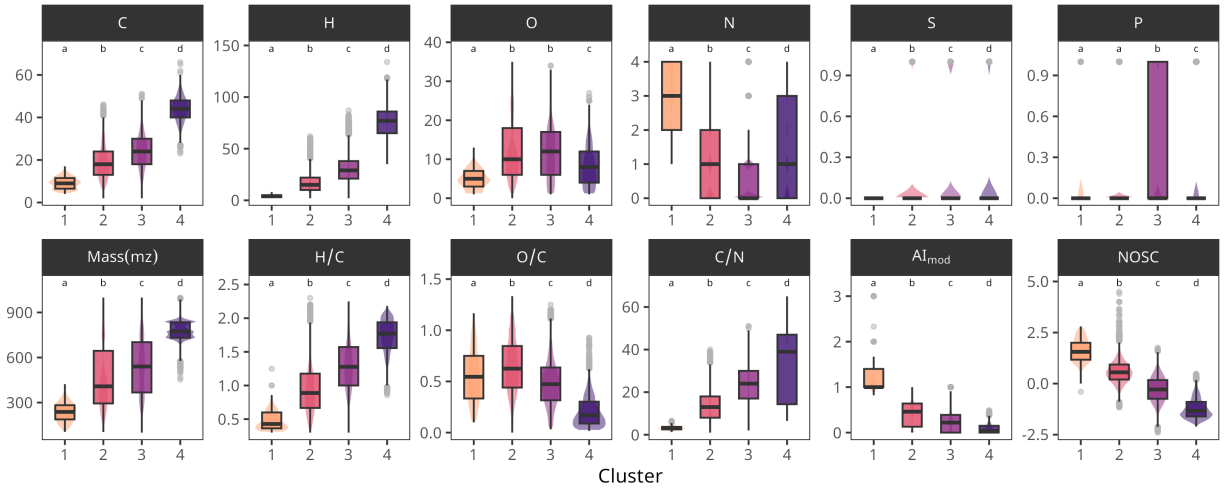

Figure S3.1: Chemical metrics used in hierarchical clustering analysis and their distribution among identified clusters. Given are chemical metrics such as the number of elements within a molecular formula (C = carbon, H = hydrogen, O = oxygen, N = nitrogen, S = sulfur, P = phosphorous), mass (in *mz*), elemental ratios (H/C, O/C, C/N) and indicators of aromaticity (AI_{mod}) as well as nominal oxidation state of carbon (NOSC). Middle lines of boxplots represent the median, while the upper and lower hinges represent the 25^{th} and 75^{th} percentiles. Upper and lower whiskers expand to the largest and smallest value, respectively, no further than 1.5 times the inter-quartile range (IQR) from the hinge. Outliers are depicted as points. Clusters are identified as colours. The distribution of data is additionally depicted in the cluster colours around the boxplots.

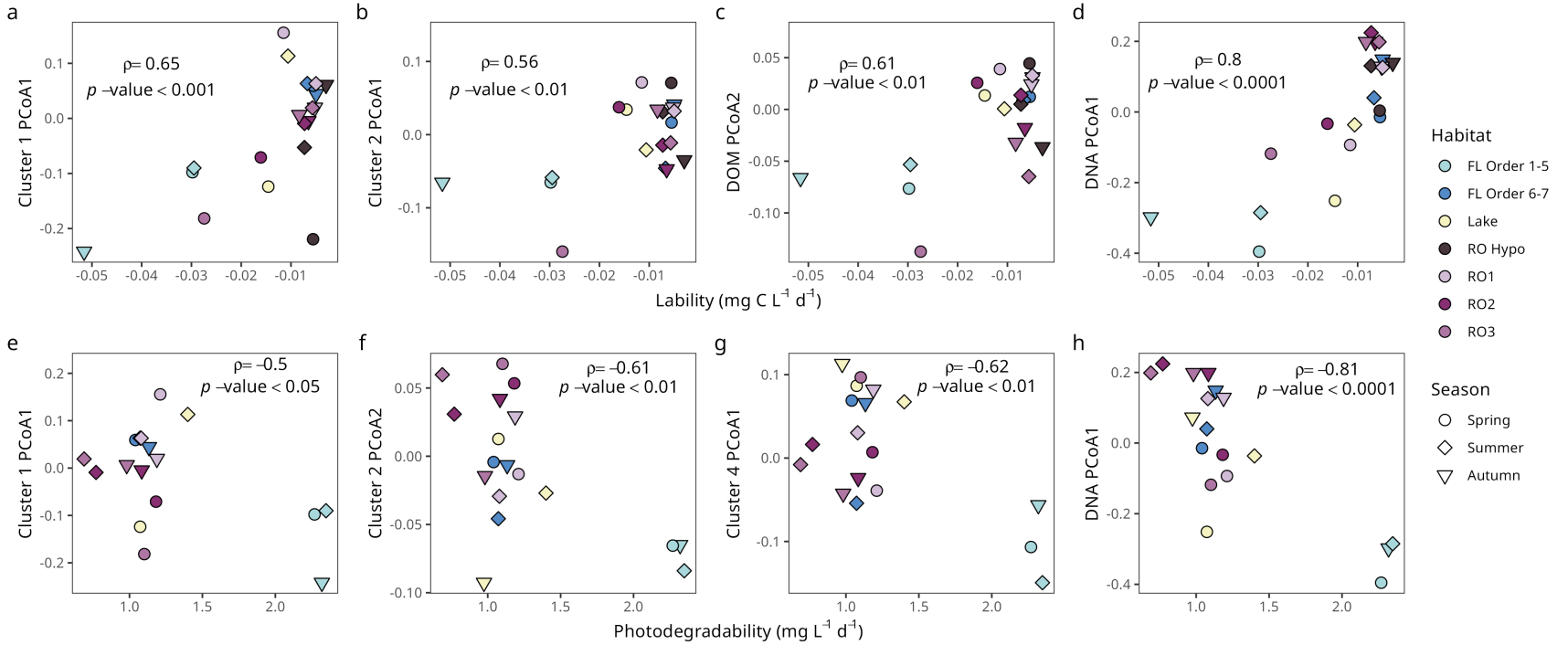

Figure S3.2: Spearman correlations of lability and photodegradability of DOM with axes extracted from PCoAs of DOM and microbial composition (DNA and RNA). Lability and photodegradability were correlated with bulk DOM, cluster DOM, DNA and RNA PCoA axes and only significant correlations are plotted. Correlation coefficient (ρ) and *p*-values are annotated within each correlation plot.

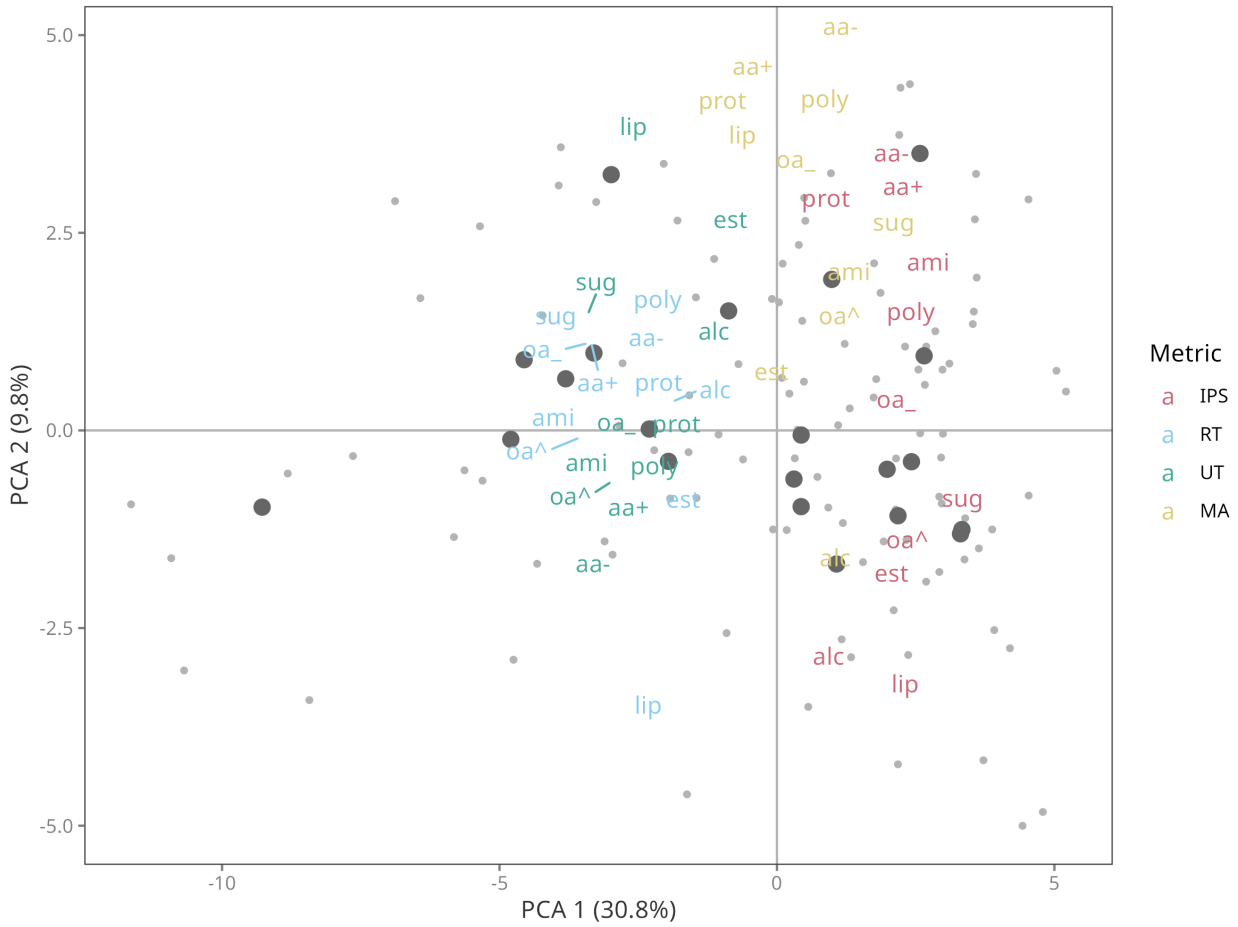

Figure S3.3: Substrate and CUP metric distribution in CUP PCA. Species distribution of PCA conducted on CUP metrics (Fig. 3.4). CUP metrics were averaged by substrate groups of alcohols (alc), amines (ami), esters (est), hydrophilic amino acids (aa+), hydrophobic amino acids (aa-), lipids (lip), organic acids with a respiratory quotient (RQ) > 1 (oa^), organic acids with a RQ $\leq$ 1 (oa_), polymers (poly), proteins (prot), and sugars (sug).

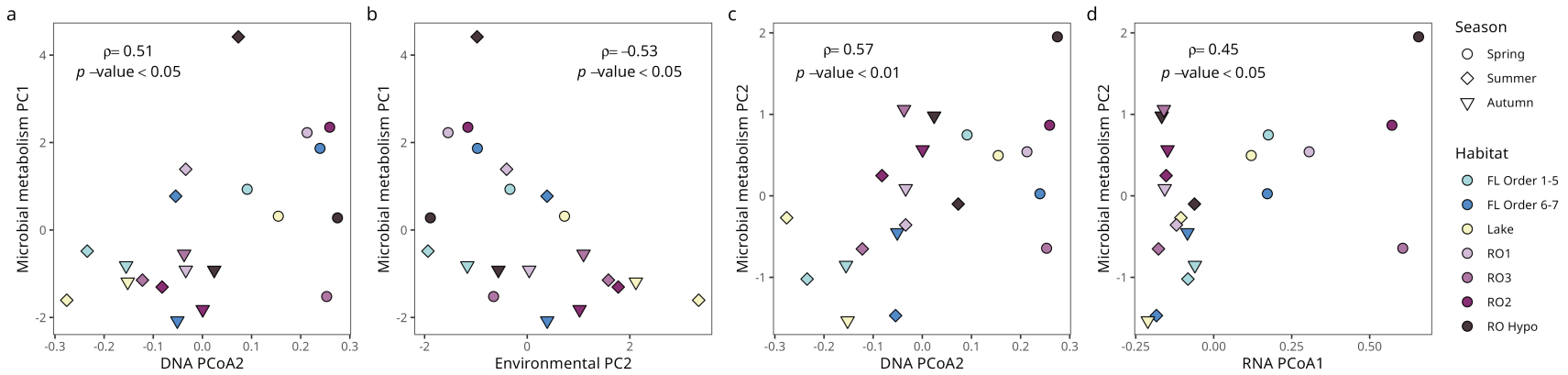

Figure S3.4: Spearman correlations of axes extracted from microbial metabolism PCA with all other axes extracted from multivariate analyses on different matrices. Axes of microbial metabolism PCA were correlated with environmental, hydrologic, and CUP PCA axes, as well as, bulk DOM, individual cluster DOM (1-4), DNA and RNA PCoA axes. Only significant correlations are plotted. Correlation coefficient (ρ) and *p*-values are annotated within each correlation plots.

3.8.2 Supplementary tables

Table S3.1: PERMDISP analyses on DNA PCoAs to simplify habitat categories. Three PCoAs were computed for habitat categories identified as either fluvial (FL), lakes (L) and reservoirs (RO). After PERMDISP model was calculated, Tukey's HSD test was conducted to evaluate habitat by habitat differences. The number of asterisks increase with lower *p*-values (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$) according to pair-wise comparisons with Tukey's HSD test.

Habitat	Group 1	Group 2	Pair-wise difference	Lower	Upper	Adj. *p*	Significance
Lake	L Order 2	L Order 1	-0.022	-0.268	0.225	0.998	
	L Order 4	L Order 2	0.068	-0.208	0.344	0.914	
	L Order 4	L Order 1	0.047	-0.229	0.322	0.977	
	L Order 5	L Order 4	-0.073	-0.375	0.230	0.922	
	L Order 5	L Order 1	-0.026	-0.302	0.250	0.997	
	L Order 5	L Order 2	-0.004	-0.280	0.271	1.000	
	L Order 6	L Order 4	-0.211	-0.472	0.051	0.130	
	L Order 6	L Order 1	-0.164	-0.395	0.067	0.202	
	L Order 6	L Order 2	-0.143	-0.373	0.088	0.306	
	L Order 6	L Order 5	-0.138	-0.400	0.123	0.440	
Reservoir	RO1 Deep	RO1	-0.052	-0.182	0.078	0.857	
	RO2	RO3 Deep	-0.096	-0.239	0.047	0.385	
	RO2	RO1	-0.036	-0.092	0.021	0.446	
	RO2	RO3	0.034	-0.041	0.110	0.776	
	RO2	RO1 Deep	0.016	-0.113	0.145	0.999	
	RO2 Deep	RO3	0.145	0.032	0.258	0.004	**
	RO2 Deep	RO2	0.110	0.011	0.210	0.021	*
	RO2 Deep	RO1 Deep	0.127	-0.027	0.281	0.171	

Continued on next page

Habitat	Group 1	Group 2	Pair-wise difference	Lower	Upper	Adj. *p*	Significance
	RO2 Deep	RO1	0.075	-0.026	0.175	0.273	
	RO2 Deep	RO3 Deep	0.015	-0.151	0.180	1.000	
	RO3	RO1	-0.070	-0.147	0.007	0.095	
	RO3	RO1 Deep	-0.018	-0.158	0.121	0.999	
	RO3 Deep	RO3	0.130	-0.022	0.283	0.140	
	RO3 Deep	RO1 Deep	0.112	-0.073	0.297	0.501	
	RO3 Deep	RO1	0.060	-0.084	0.204	0.832	
Stream	FL Order 2	FL Order 1	-0.009	-0.204	0.187	1.000	
	FL Order 3	FL Order 1	-0.004	-0.157	0.149	1.000	
	FL Order 3	FL Order 2	0.005	-0.205	0.215	1.000	
	FL Order 4	FL Order 1	-0.017	-0.193	0.158	1.000	
	FL Order 4	FL Order 3	-0.013	-0.205	0.178	1.000	
	FL Order 4	FL Order 2	-0.009	-0.235	0.218	1.000	
	FL Order 5	FL Order 1	-0.055	-0.201	0.092	0.914	
	FL Order 5	FL Order 3	-0.051	-0.216	0.114	0.965	
	FL Order 5	FL Order 2	-0.046	-0.251	0.159	0.993	
	FL Order 5	FL Order 4	-0.037	-0.223	0.149	0.996	
	FL Order 6	FL Order 1	-0.204	-0.329	-0.080	0.000	***
	FL Order 6	FL Order 3	-0.201	-0.347	-0.054	0.002	**
	FL Order 6	FL Order 4	-0.187	-0.357	-0.018	0.021	*
	FL Order 6	FL Order 5	-0.150	-0.289	-0.011	0.027	*
	FL Order 6	FL Order 2	-0.196	-0.386	-0.006	0.039	*
	FL Order 7	FL Order 1	-0.179	-0.291	-0.067	0.000	***

Continued on next page

Habitat	Group 1	Group 2	Pair-wise difference	Lower	Upper	Adj. *p*	Significance
	FL Order 7	FL Order 3	-0.175	-0.311	-0.040	0.004	**
	FL Order 7	FL Order 4	-0.162	-0.322	-0.002	0.046	*
	FL Order 7	FL Order 5	-0.125	-0.252	0.003	0.060	
	FL Order 7	FL Order 2	-0.170	-0.352	0.011	0.080	
	FL Order 7	FL Order 6	0.025	-0.077	0.128	0.988	

Table S3.2: PERMANOVA and PERMDISP results for each multivariate analysis conducted. Significant groups as revealed by PERMANOVA are highlighted in bold.

		PERMANOVA				PERMDISP		
Dataset	Group	*df*	*F*-statistic	R^2	*p* value	*df*	*F*-statistic	*p* value
Environment	**Habitat**	6	5.24	0.15	< 0.0001	6	5.07	< 0.0001
	Season	2	20.46	0.19	< 0.0001	2	0.09	= 0.92
	Combined	8	9.05	0.34	< 0.0001	20	3.10	< 0.0001
Hydrology	**Habitat**	6	23.10	0.46	< 0.0001	6	36.39	< 0.0001
	Season	2	10.97	0.07	< 0.0001	2	2.60	= 0.08
	Combined	8	20.06	0.53	< 0.0001	20	14.81	< 0.0001
DNA	**Habitat**	6	8.84	0.23	< 0.0001	6	18.42	< 0.0001
	Season	2	19.18	0.17	< 0.0001	2	0.88	= 0.42
	Combined	8	11.43	0.40	< 0.0001	20	8.78	< 0.0001
RNA	**Habitat**	6	1.68	0.06	< 0.01	6	0.63	= 0.7
	Season	2	21.00	0.23	< 0.0001	2	30.95	< 0.0001
	Combined	8	6.51	0.28	< 0.0001	20	1.51	= 0.09

Continued on next page

		PERMANOVA				PERMDISP		
Dataset	Group	*df*	*F*-statistic	R^2	*p* value	*df*	*F*-statistic	*p* value
DOM	**Habitat**	6	6.64	0.20	< 0.0001	6	11.32	< 0.0001
	Season	2	1.14	0.01	= 0.248	2	0.57	= 0.56
	Combined	8	5.27	0.21	< 0.0001	18	3.52	< 0.0001
Cluster 1	Habitat	6	1.18	0.04	= 0.09	6	2.13	= 0.05
	Season	2	0.84	0.01	= 0.76	2	0.73	= 0.48
	Combined	8	1.09	0.05	= 0.20	17	1.18	= 0.29
Cluster 2	**Habitat**	6	5.03	0.16	< 0.0001	6	7.37	< 0.0001
	Season	2	1.07	0.01	= 0.332	2	0.34	= 0.71
	Combined	8	4.04	0.17	< 0.0001	18	2.58	< 0.001
Cluster 3	**Habitat**	6	7.40	0.22	< 0.0001	6	12.57	< 0.0001
	Season	2	1.16	0.01	= 0.2443	2	0.61	= 0.54
	Combined	8	5.84	0.23	< 0.0001	18	3.86	< 0.0001
Cluster 4	**Habitat**	6	2.19	0.08	< 0.0001	6	5.83	< 0.0001
	Season	2	1.41	0.02	< 0.01	2	1.73	= 0.18
	Combined	8	2.00	0.09	< 0.0001	18	2.82	< 0.001
CUP	**Habitat**	6	3.14	0.14	< 0.0001	6	2.87	< 0.05
	Season	2	8.38	0.12	< 0.0001	2	1.40	= 0.25
	Combined	8	4.45	0.26	< 0.0001	20	1.73	< 0.05
Microbial metabolism	**Habitat**	6	3.76	0.11	< 0.0001	6	2.39	< 0.05
	Season	2	18.43	0.18	< 0.0001	2	7.77	< 0.001
	Combined	8	7.43	0.29	< 0.0001	20	2.87	< 0.001

Table S3.3: Spearman correlation results between BP, BR and BGE by habitat. Significant correlation coefficients are highlighted in bold.

Habitat	BP vs. BGE		BR vs. BGE		BP vs. BR	
	p-value	ρ	*p*-value	ρ	*p*-value	ρ
FL Order 1-5	0.11	0.31	0.10	-0.32	0.68	-0.08
FL Order 6-7	< 0.05	**0.48**	< 0.001	**-0.69**	0.97	0.01
Lake	0.13	0.47	< 0.01	**-0.73**	0.80	0.08
RO1	< 0.0001	**0.63**	< 0.0001	**-0.70**	< 0.05	**-0.39**
RO2	< 0.001	**0.60**	< 0.0001	**-0.94**	< 0.001	**-0.57**
RO3	< 0.05	**0.71**	< 0.05	**-0.63**	< 0.05	**-0.63**
RO Hypo	0.46	0.31	0.18	-0.53	0.51	0.28

Table S3.4: Spearman correlation results between BP, BR and BGE by season. Significant correlation coefficients are highlighted in bold.

Season	BP vs. BGE		BR vs. BGE		BP vs. BR	
	p-value	ρ	*p*-value	ρ	*p*-value	ρ
Spring	< 0.01	**0.39**	< 0.0001	**-0.70**	< 0.05	**-0.32**
Summer	0.89	0.02	< 0.0001	**-0.67**	0.83	0.03
Autumn	0.81	0.03	< 0.0001	**-0.70**	0.14	-0.20

CONCLUSION

The core of this thesis lies in understanding spatial and temporal patterns of microbial community assembly and functioning across an entire aquatic continuum. To do so, we investigated how microbial communities come together within an aquatic network, how they relate to spatio-temporal patterns in dissolved organic matter (DOM), and how their shifts potentially associate to changes in microbial metabolism. We have explored the relationships of microbial communities with DOM and microbial metabolism at an unprecedented spatio-temporal scale and complexity by examining a diverse array of ecosystems (e.g., streams, rivers, lakes, reservoirs), observing patterns across years and seasons, and distinguishing more reactive fractions within both microbial and DOM assemblages. The thesis not only provides an invaluable comprehensive dataset to the literature, which includes high-resolution measurements of DNA, RNA, DOM and microbial metabolism, but it also aimed to further our understanding of core principles in microbial ecology.

In chapter 1 we demonstrated that microbial community assembly within boreal aquatic networks was strongly influenced by terrestrial connectivity, upstream aquatic history, and seasonal variations. We observed that terrestrially-derived microorganisms dominated both, DNA- and RNA-based communities in every season. In contrast, microbes first detected in aquatic ecosystems were carried downstream within the DNA pool; representing assembly processes that had occurred in the past. These differences in microbial origin were a result of the variation in the predominant assembly process along the aquatic network. Mass effects were present at all times due to the unidirectional flow of water favouring the dispersal of terrestrial microbes and aquatic microbes in upstream ecosystems. However, environmental selection was also strong in lower-order streams. Furthermore, environmental selection became stronger again when the continuum arrived in higher-order rivers and reservoirs in spring and summer, respectively. The location of the strongest environmental selection across the aquatic continuum was determined by seasonal variation in hydrology and environmental factors. These assembly processes and recruitment of taxa into the active community occurred across the entire rank abundance curve, implying that the rare biosphere plays a vital role in microbial community functioning. These findings highlight the importance of considering the upstream aquatic network and seasonality when examining microbial

community dynamics within inland waters.

Chapter 2 advanced our understanding of microbial and DOM assemblages by identifying microbial taxa and DOM molecular formulae that are reactive to environmental changes. This approach enabled us to extract the most dynamic fractions within both assemblages across the hydrological continuum, which revealed that the majority of microbial communities and approximately half of the DOM pool do not exhibit statistically significant spatial patterns with flow-weighted water age. Although most microbes tended to increase, the DOM assemblage was dominated by either increasing or decreasing units depending on the examined year; variation that can be attributed to the creation and flooding of a new reservoir. Despite the lack of an apparent seasonality, a clear seasonality emerged when the reactive units of both microorganisms and DOM were correlated to each other. Influx of fresh terrestrially-derived DOM in spring led to many correlations indicative of microbial consumption along the water age gradient, whereas mid-sized, partially decomposed molecular formulae were seemingly consumed in summer. These findings may imply that the supply of fresh terrestrial DOM is too high for rapid consumption during the spring freshet, which enabled us to observe degradation dynamics of microbes with fresh DOM. In contrast, concentrations of fresh autochthonous DOM may not be high enough and its quick degradation and transformation led us to observe causal interactions only with partially degraded DOM molecular formulae. At the same time, different sets of microbial phyla in high water residence time systems in spring and summer were likely selected by decomposed DOM of low and high molecular size, respectively. The findings of this chapter highlighted the inherent complexity of microbe-DOM interactions, making it extremely difficult to discern any meaningful patterns. Only after multiple quality filters and statistical thresholds were put in place, a core set of relationships could be extracted and interpreted when put in a comprehensive ecological framework. The developed conceptual framework provides perspective to the scientific literature by highlighting the importance of considering ecological and hydrological dynamics when using correlational analyses.

In Chapter 3, we delved into microbial metabolism, illustrating how metabolic metrics vary and link to microbial community and DOM composition across habitats and seasons within the aquatic

network. Across the aquatic network microbial respiration rates and carbon demands peaked in spring. In particular, specific respiration rates were linked to shifts in terrestrial DOM pools. Bacterial production generally increased in summer and autumn, and was highest in streams and lakes that were located early within the aquatic network. Bacterial growth efficiency was rather modulated by variations in bacterial respiration rather than bacterial production, and linked to shifts in microbial community composition. Assays for evaluating the functional capacity of the microbial community to consume a variety of carbon substrates revealed that spring and lake communities have general substrate preferences and consume substrates slowly. Conversely, streams (all seasons) and reservoirs (summer and autumn) consumed a diverse set of substrates very quickly. Studies that evaluate the variation and drivers of microbial metabolism are lacking in recent years. This chapter emphasises the importance of examining microbial metabolism across spatio-temporal gradients to understand how metabolism responds to hydrological, environmental, DOM and microbial community shifts. In order for the scientific community to find coherent links to drivers and ultimately incorporate heterotrophic metabolism to earth system models in the future, more studies are encouraged to explore links of metabolism across large gradients.

The collective results of the three chapters paint a picture of microbial community ecology across the aquatic network with strong seasonal influences. During the spring freshet, both microbial OTUs and fresh DOM molecular formulae are carried from the terrestrial surroundings into the aquatic network (Crump *et al.*, 2012; Ruiz-González *et al.*, 2015c). Despite terrestrial microbes undergoing strong selective pressure, a large fraction of terrestrial microbes continue to thrive along the aquatic continuum. Influx of fresh terrestrial DOM and high bacterial abundances result in enhanced bacterial respiration, and lower bacterial growth efficiencies across the entire aquatic network. However, as the fresh DOM pool gets consumed and transformed into more degraded compounds, a pool of less available DOM accumulates and starts to act as a selective force on who remains active within the microbial community in higher residence time systems. In summer, terrestrial DOM supply becomes scarce, except for lower-order streams where terrestrial-aquatic exchanges remain strong. Hence, streams do not show large variations in community assembly nor respiration across seasons. However, their bacterial production and growth efficiency increases in summer likely due to the increase of available nutrients (Voss *et al.*, 2015; Burd *et al.*, 2018). With

warmer temperatures and nutrient supplies, DOM shifts to more autochthonous material, but the fresh material is too quickly consumed and transformed, allowing us to only measure DOM degradation links to microbes with mildly decomposed matter. Less available DOM accumulates with water age regardless of the season, affecting microbial community composition, which likely leads to the upstream legacy of microbes that cease to be active along the continuum. Shifts in the decomposed DOM pools may have resulted in the stronger selection effects observed in higher water residence time systems, allowing for terrestrial microbes to continue to thrive, while many aquatic microbes may lack in the functional capacity to degrade more decomposed matter. By examining microbial community assembly, DOM interactions and microbial metabolism within the same aquatic network across seasons, this thesis represents a novel contribution to the field of microbial ecology and biogeochemistry. The collective insights highlight the importance of evaluating these interdependent components of the aquatic carbon cycle together, in order to gain novel insights on how microbial ecology influences biogeochemical patterns across the entire aquatic network.

In summary, this thesis provides spatio-temporal insights into the intertwined nature of microbial communities, dissolved organic matter composition and bacterial metabolism. The results provide examples and raise attention on how connectivity and seasonality are crucial components into understanding aquatic biogeochemical processes within inland waters.

BIBLIOGRAPHY

Aanderud, Z. T., Vert, J. C., Lennon, J. T., Magnusson, T. W., Breakwell, D. P. and Harker, A. R. (2016). Bacterial dormancy is more prevalent in freshwater than hypersaline lakes. *Frontiers in Microbiology*, *7*, 853. http://dx.doi.org/10.3389/fmicb.2016.00853

Achtman, M. and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, *6*, 431–440. http://dx.doi.org/10.1038/nrmicro1872

Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, *13*, 497–508. http://dx.doi.org/10.1038/nrmicro3491

Adams, H. E., Crump, B. C. and Kling, G. W. (2014). Metacommunity dynamics of bacteria in an arctic lake: The impact of species sorting and mass effects on bacterial production and biogeography. *Frontiers in Microbiology*, *5*, 82. http://dx.doi.org/10.3389/fmicb.2014.00082

Alonso, C. and Pernthaler, J. (2006). Concentration-dependent patterns of leucine incorporation by coastal picoplankton. *Applied and Environmental Microbiology*, *72*, 2141–2147. http://dx.doi.org/10.1128/AEM.72.3.2141-2147.2006

Aluwihare, L. I., Repeta, D. J. and Chen, R. F. (1997). A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature*, *387*, 166–169. http://dx.doi.org/10.1038/387166a0

Amado, A. M., Meirelles-Pereira, F., Vidal, L. O., Sarmento, H., Suhett, A. L., Farjalla, V. F., Cotner, J. B. and Roland, F. (2013). Tropical freshwater ecosystems have lower bacterial growth efficiency than temperate ones. *Frontiers in Microbiology*, *4*. http://dx.doi.org/10.3389/fmicb.2013.00167

Amaral, V., Graeber, D., Calliari, D. and Alonso, C. (2016). Strong linkages between DOM optical

properties and main clades of aquatic bacteria. *Limnology and Oceanography*, *61*, 906–918. http://dx.doi.org/10.1002/lno.10258

American Society for Microbiology (2023). Microbes in models: Integrating microbes into earth system models for understanding climate change. in *Report on an American Academy of Microbiology Virtual Colloquium held on Dec. 6 and 8, 2022*. American Society for Microbiology. http://dx.doi.org/10.1128/aamcol.jun.2023. URL https://asm.org/reports/microbes-in-models-integrating-microbes-into-earth

Amon, R. M. W. and Benner, R. (1996). Bacterial utilization of different size classes of dissolved organic matter. *Limnology and Oceanography*, *41*, 41–51. http://dx.doi.org/10.4319/lo.1996.41.1.0041

Anderson, M. J. and Walsh, D. C. (2013). Anderson and walsh (2013) permanova, anosim, and the mantel test in the face of heterogeneous dispersions- what null hypothesis are you .pdf. *Ecological Monograph*, *83*, 557–574. http://dx.doi.org/10.1890

Attermeyer, K., Catalán, N., Einarsdottir, K., Freixa, A., Groeneveld, M., Hawkes, J. A., Bergquist, J. and Tranvik, L. J. (2018). Organic carbon processing during transport through boreal inland waters: Particles as important sites. *Journal of Geophysical Research: Biogeosciences*, *123*, 2412–2428. http://dx.doi.org/10.1029/2018jg004500

Attermeyer, K., Hornick, T., Kayler, Z. E., Bahr, A., Zwirnmann, E., Grossart, H. P. and Premke, K. (2014). Enhanced bacterial decomposition with increasing addition of autochthonous to allochthonous carbon without any effect on bacterial community composition. *Biogeosciences*, *11*, 1479–1489. http://dx.doi.org/10.5194/bg-11-1479-2014

Attermeyer, K., Premke, K., Hornick, T., Hilt, S. and Grossart, H.-P. (2013). Ecosystem-level studies of terrestrial carbon reveal contrasting bacterial metabolism in different aquatic habitats. *Ecology*, *94*, 2754–2766. http://dx.doi.org/10.1890/13-0420.1

Ávila, M. P., Brandão, L. P., Brighenti, L. S., Tonetta, D., Reis, M. P., Stæhr, P. A., Asmala, E., Amado, A. M., Barbosa, F. A., Bezerra-Neto, J. F. and Nascimento, A. M. (2019). Linking shifts in bacterial community with changes in dissolved organic matter pool in a tropical lake. *Science*

of the Total Environment, *672*, 990–1003. http://dx.doi.org/10.1016/j.scitotenv.2019.04.033

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L. and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, *10*, 257–263. http://dx.doi.org/10.1021/acs.joc.6b00938

Baines, S. B. and Pace, M. L. (1991). The production of dissolved organic matter by phytoplankton and its importance to bacteria : Patterns across marine and freshwater systems. *Limnology and Oceanography*, *36*, 1078–1090.

Bambakidis, T., Crump, B. C., Yoon, B., Kyzivat, E. D., Aho, K. S., Leal, C. F., Fair, J. H., Stubbins, A., Wagner, S., Raymond, P. A. and Hosen, J. D. (2024). Temperature, water travel time, and dissolved organic matter structure river microbial communities in a large temperate watershed. *Limnology and Oceanography*. http://dx.doi.org/10.1002/lno.12591

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A. (2019). EPA-ng: Massively parallel evolutionary placement of genetic sequences. *Systematic Biology*, *68*, 365–369. http://dx.doi.org/10.1093/sysbio/syy054

Barbosa, P. M., Bodmer, P., Stadler, M., Rust, F., Tremblay, A. and del Giorgio, P. A. (2023). Ecosystem metabolism is the dominant source of carbon dioxide in three young boreal cascade-reservoirs (la romaine complex, québec). *Journal of Geophysical Research: Biogeosciences*, *128*, 1–21. http://dx.doi.org/10.1029/2022JG007253

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T. and Schwendinger, B. (2024). data.table: Extension of 'data.frame'. URL https://CRAN.R-project.org/package=data.table

Battin, T. J., a. Kaplan, L., Findlay, S., Hopkinson, C. S., Marti, E., Packman, A. I., Newbold, J. D. and Sabater, F. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nature Geoscience*, *1*, 95–100. http://dx.doi.org/10.1038/ngeo602

Battin, T. J., a Kaplan, L., Newbold, J. D. and Hansen, C. M. E. (2003). Contributions of microbial

biofilms to ecosystem processes in stream mesocosms. *Nature*, *426*, 439–442. http://dx.doi.org/10.1038/nature02152

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M. and Packmann, A. I. (2016). The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology*, *14*, 251–263. http://dx.doi.org/10.1038/nrmicro.2016.15

Battin, T. J., Butturini, A. and Sabater, F. (1999). Immobilization and metabolism of dissolved organic carbon by natural sediment biofilms in a mediterranean and temperate stream. *Aquatic Microbial Ecology*, *19*, 297–305.

Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A. and Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, *2*, 598–600. http://dx.doi.org/10.1038/ngeo618

Battin, T. J., Wille, A., Psenner, R. and Richter, A. (2004). Large-scale environmental controls on microbial biofilms in high-alpine streams.

Baña, Z., Abad, N., Uranga, A., Azúa, I., Artolozaga, I., Unanue, M., Iriberri, J., Arrieta, J. M. and Ayo, B. (2020). Recurrent seasonal changes in bacterial growth efficiency, metabolism and community composition in coastal waters. *Environmental Microbiology*, *22*, 369–380. http://dx.doi.org/10.1111/1462-2920.14853

Becker, J. C., Rodibaugh, K. J., Hahn, D. and Nowlin, W. H. (2017). Bacterial community composition and carbon metabolism in a subtropical riverscape. *Hydrobiologia*, *792*, 209–226. http://dx.doi.org/10.1007/s10750-016-3058-2

Becking, L. G. M. B. (1934). *Geobiologie of Inleiding tot de Milieukunde*. W.P. Van Stockum and Zoon.

Bello, M. D., Lee, H., Goyal, A. and Gore, J. (2021). Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nature Ecology and Evolution*, *5*, 1424–1434. http://dx.doi.org/10.1038/s41559-021-01535-8

Benner, R. and Amon, R. M. (2015). The size-reactivity continuum of major bioelements in the ocean. *Annual Review of Marine Science*, *7*, 185–205. http://dx.doi.org/10.1146/annurev-marine-010213-135126

Benner, R., Opsahl, S., Chin-Leo, G., Richey, J. E. and Forsberg, B. R. (1995). Bacterial carbon metabolism in the Amazon river system. *Limnology and Oceanography*, *40*, 1262–1270. http://dx.doi.org/10.4319/lo.1995.40.7.1262

Berger, B., Hoch, B., Kavka, G. and Herndl, G. J. (1995). Bacterial metabolism in the river Danube: parameters influencing bacterial production. *Freshwater Biology*, *34*, 601–616. http://dx.doi.org/10.1111/j.1365-2427.1995.tb00916.x

Berger, S. A. and Stamatakis, A. (2011). Aligning short reads to reference alignments and trees. *Bioinformatics*, *27*, 2068–2075. http://dx.doi.org/10.1093/bioinformatics/btr320

Berggren, M. and del Giorgio, P. A. (2015). Distinct patterns of microbial metabolism associated to riverine dissolved organic carbon of different source and quality. *Journal of Geophysical Research: Biogeosciences*, *120*, 989–999. http://dx.doi.org/10.1002/2015JG002963

Berggren, M., Gudasz, C., Guillemette, F., Hensgens, G., Ye, L. and Karlsson, J. (2019). Systematic microbial production of optically active dissolved organic matter in subarctic lake water. *Limnology and Oceanography*, 1–11. http://dx.doi.org/10.1002/lno.11362

Berggren, M., Lapierre, J. F. and Giorgio, P. A. D. (2012). Magnitude and regulation of bacterioplankton respiratory quotient across freshwater environmental gradients. *ISME Journal*, *6*, 984–993. http://dx.doi.org/10.1038/ismej.2011.157

Berggren, M., Laudon, H., Haei, M., Ström, L. and Jansson, M. (2010). Efficient aquatic bacterial metabolism of dissolved low-molecular-weight compounds from terrestrial sources. *ISME Journal*, *4*, 408–416. http://dx.doi.org/10.1038/ismej.2009.120

Berggren, M., Laudon, H. and Jansson, M. (2007). Landscape regulation of bacterial growth efficiency in boreal freshwaters. *Global*, *21*, 1–9. http://dx.doi.org/10.1029/2006GB002844

Berggren, M., Laudon, H. and Jansson, M. (2009). Aging of allochthonous organic carbon regulates bacterial production in unproductive boreal lakes. *Limnology and Oceanography*, *54*, 1333–1342. http://dx.doi.org/10.4319/lo.2009.54.4.1333

Berggren, M., Ye, L., Sponseller, R. A., Bergström, A. K., Karlsson, J., Verheijen, H. and Hensgens, G. (2023). Nutrient limitation masks the dissolved organic matter composition effects on bacterial metabolism in unproductive freshwaters. *Limnology and Oceanography*, *68*, 2059–2069. http://dx.doi.org/10.1002/lno.12406

Bertolet, B. L., West, W. E., Armitage, D. W. and Jones, S. E. (2019). Organic matter supply and bacterial community composition predict methanogenesis rates in temperate lake sediments. http://dx.doi.org/10.1002/lol2.10114

Besemer, K., Peter, H., Logue, J. B., Langenheder, S., Lindström, E. S., Tranvik, L. J. and Battin, T. J. (2012). Unraveling assembly of stream biofilm communities. *The ISME Journal*, *6*, 1459–1468. http://dx.doi.org/10.1038/ismej.2011.205

Besemer, K., Singer, G., Limberger, R., Chlup, A.-K., Hochedlinger, G., Hödl, I., Baranyi, C. and Battin, T. J. (2007). Biophysical controls on community succession in stream biofilms. *Applied and Environmental Microbiology*, *73*, 4966–4974. http://dx.doi.org/10.1128/AEM.00588-07

Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W. and Battin, T. J. (2013). Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proceedings of the Royal Society B*, *280*, 20131760. http://dx.doi.org/http://dx.doi.org/10.1098/rspb.2013.1760

Biddanda, B., Ogdahl, M. and Cotner, J. (2001). Dominance of bacterial metabolism in oligotrophic relative to eutrophic waters. *Limnology and Oceanography*, *46*, 730–739. http://dx.doi.org/10.4319/lo.2001.46.3.0730

Bier, R. L., Bernhardt, E. S., Boot, C. M., Graham, E. B., Hall, E. K., Lennon, J. T., Nemergut, D. R., Osborne, B. B., Ruiz-González, C., Schimel, J. P., Waldrop, M. P. and Wallenstein, M. D. (2015). Linking microbial community structure and microbial processes: An empirical and

conceptual overview. *FEMS Microbiology Ecology*, *91*. http://dx.doi.org/10.1093/femsec/fiv113

Blakney, G. T., Hendrickson, C. L. and Marshall, A. G. (2011). Predator data station: A fast data acquisition system for advanced ft-icr ms experiments. *International Journal of Mass Spectrometry*, *306*, 246–252. http://dx.doi.org/10.1016/j.ijms.2011.03.009

Blazewicz, S. J., Barnard, R. L., Daly, R. A. and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities : limitations and uses. *The ISME Journal*, *7*, 2061–2068. http://dx.doi.org/10.1038/ismej.2013.102

Boodoo, K. S., Fasching, C. and Battin, T. (2020). Sources, transformation and fate of dissolved organic matter in the gravel bar of a prealpine stream. *Journal of Geophysical Research: Biogeosciences*, *125*, e2019JG005604. http://dx.doi.org/10.1029/2019jg005604

Borcard, D., Gillet, F. and Legendre, P. (2011). *Numerical Ecology with R*. Springer. http://dx.doi.org/10.1007/978-1-4419-7976-6

Borton, M. A., McGivern, B. B., Willi, K. R., Woodcroft, B. J., Mosier, A. C., Singleton, D. M., Bambakidis, T., Pelly, A., Daly, R. A., Liu, F., Freiburger, A., Edirisinghe, J. N., Faria, J. P., Danczak, R., Leleiwi, I., Goldman, A. E., Wilkins, M. J., Hall, E. K., Pennacchio, C., Roux, S., Eloe-Fadrosh, E. A., Good, S. P., Sullivan, M. B., Wood-Charlson, E. M., Miller, C. S., Ross, M. R. V., Henry, C. S., Crump, B. C., Stegen, J. C. and Wrighton, K. C. (2025). A functional microbiome catalogue crowdsourced from north american rivers. *Nature*, *637*, 103–112. http://dx.doi.org/10.1038/s41586-024-08240-z

Bowsher, A. W., Kearns, P. J. and Shade, A. (2019). 16S rRNA/rRNA gene ratios and cell activity staining reveal consistent patterns of microbial activity in plant-associated soil. *mSystems*, *4*, e00003–19. http://dx.doi.org/10.1128/msystems.00003-19

Boyer, E. W., Hornberger, G. M., Bencala, K. E. and McKnight, D. (1996). Overview of a simple model describing variation of dissolved organic carbon in an upland catchment. *Ecological Modelling*, *86*, 183–188.

Bray, J. R. and Curtis, J. T. (1957). An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs*, *27*, 325–349. http://dx.doi.org/10.2307/1942268

Brown, B. L. (2003). Spatial heterogeneity reduces temporal variability in stream insect communities. *Ecology Letters*, *6*, 316–325. http://dx.doi.org/10.1046/j.1461-0248.2003.00431.x

Brown, R. D. and Brasnett, B. (2010). Canadian meteorological centre (cmc) daily snow depth analysis data. http://dx.doi.org/10.5067/W9FOYWH0EQZ3

Burd, K., Tank, S. E., Dion, N., Quinton, W. L., Spence, C., Tanentzap, A. J. and Olefeldt, D. (2018). Seasonal shifts in export of DOC and nutrients from burned and unburned peatland-rich catchments, Northwest Territories, Canada. *Hydrology and Earth System Sciences*, *22*, 4455–4472. http://dx.doi.org/10.5194/hess-22-4455-2018

Bureau d'audiences publiques sur l'environnement and Canadian Environmental Assessment Agency (2009). Romaine hydroelectric complex project: Investigation and public hearing report. URL https://iaac-aeic.gc.ca/050/documents/34211/34211E.pdf

Butturini, A., Guarch, A., Romaní, A. M., Freixa, A., Amalfitano, S., Fazi, S. and Ejarque, E. (2016). Hydrological conditions control in situ DOM retention and release along a mediterranean river. *Water Research*, *99*, 33–45. http://dx.doi.org/10.1016/j.watres.2016.04.036

Caillon, F., Besemer, K., Peduzzi, P. and Schelker, J. (2021). Soil microbial inoculation during flood events shapes headwater stream microbial communities and diversity. *Microbial Ecology*. http://dx.doi.org/10.1007/s00248-021-01700-3

Caillon, F. and Schelker, J. (2020). Dynamic transfer of soil bacteria and dissolved organic carbon into small streams during hydrological events. *Aquatic Sciences*, *82*, 1–11. http://dx.doi.org/10.1007/s00027-020-0714-4

Callahan, B. J., McMurdie, P. J. and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*, 2639–2643. http://dx.doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. and Holmes, S. P. (2016). DADA2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, *13*. http://dx.doi.org/10.1038/nmeth.3869

Campbell, B. J., Yu, L., Heidelberg, J. F. and Kirchman, D. L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences*, *108*, 12776–12781. http://dx.doi.org/10.1073/pnas.1101405108

Campitelli, E. (2020). ggnewscale: Multiple fill and colour scales in 'ggplot2'. URL https://cran.r-project.org/package=ggnewscale

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S. and Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology*, *2*, 1–6. http://dx.doi.org/10.1038/nmicrobiol.2016.242

Carrara, F., Rinaldo, A., Giometto, A. and Altermatt, F. (2013). Complex interaction of dendritic connectivity and hierarchical patch size on biodiversity in river-like landscapes. *The American Naturalist*, *183*, 13–25. http://dx.doi.org/10.1086/674009

Casas-Ruiz, J. P., Catalán, N., Gómez-Gener, L., von Schiller, D., Obrador, B., Kothawala, D. N., López, P., Sabater, S. and Marcé, R. (2017). A tale of pipes and reactors: Controls on the in-stream dynamics of dissolved organic matter in rivers. *Limnology and Oceanography*, *62*, S85–S94. http://dx.doi.org/10.1002/lno.10471

Casas-Ruiz, J. P., Spencer, R. G., Guillemette, F., von Schiller, D., Obrador, B., Podgorski, D. C., Kellerman, A. M., Hartmann, J., Gómez-Gener, L., Sabater, S. and Marcé, R. (2020). Delineating the continuum of dissolved organic matter in temperate river networks. *Global Biogeochemical Cycles*, *34*, 1–15. http://dx.doi.org/10.1029/2019GB006495

Casas-Ruiz, J. P., Tittel, J., von Schiller, D., Catalán, N., Obrador, B., Gómez-Gener, L., Zwirnmann, E., Sabater, S. and Marcé, R. (2016). Drought-induced discontinuities in the source and degradation of dissolved organic matter in a mediterranean river. *Biogeochemistry*, *127*, 125–139. http://dx.doi.org/10.1007/s10533-015-0173-5

Catalán, N., Marcé, R., Kothawala, D. N. and Tranvik, L. J. (2016). Organic carbon decomposition rates controlled by water retention time across inland waters. *Nature Geoscience*, *9*, 1–7. http://dx.doi.org/10.1038/ngeo2720

Catalán, N., Pastor, A., Borrego, C. M., Casas-Ruiz, J. P., Hawkes, J. A., Gutiérrez, C., von Schiller, D. and Marcé, R. (2021). The relevance of environment vs. composition on dissolved organic matter degradation in freshwaters. *Limnology and Oceanography*, *66*, 306–320. http://dx.doi.org/10.1002/lno.11606

Chase, A. B., Gomez-Lunar, Z., Lopez, A. E., Li, J., Allison, S. D., Martiny, A. C. and Martiny, J. B. (2018). Emergence of soil bacterial ecotypes along a climate gradient. *Environmental Microbiology*, *20*, 4112–4126. http://dx.doi.org/10.1111/1462-2920.14405

Chin, W.-C., Orellana, M. V. and Verdugo, P. (1998). Spontaneous assembly of marine dissolved organic matter into polymer gels. *Nature*, *391*.

Cole, J. J. (1999). Aquatic microbiology for ecosystem scientists: New and recycled paradigms in ecological microbiology. *Ecosystems*, *2*, 215–225. http://dx.doi.org/10.1007/s100219900069

Cole, J. J., Findlay, S. and Pace, M. L. (1988). Bacterial production in fresh and saltwater ecosystems: a cross-system overview.

Cole, J. J., Likens, G. E. and Strayer, D. L. (1982). Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. *Limnology and Oceanography*, *27*, 1080–1090.

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Downing, J. A., Middelburg, J. J. and Melack, J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems*, *10*, 171–184. http://dx.doi.org/10.1007/s10021-006-9013-8

Comte, J., Berga, M., Severin, I., Logue, J. B. and Lindström, E. S. (2017). Contribution of different bacterial dispersal sources to lakes: Population and community effects in different

seasons. *Environmental Microbiology*, *19*, 2391–2404. http://dx.doi.org/10.1111/1462-2920.13749

Comte, J., Culley, A. I., Lovejoy, C. and Vincent, W. F. (2018). Microbial connectivity and sorting in a high arctic watershed. *ISME Journal*, *12*, 2988–3000. http://dx.doi.org/10.1038/s41396-018-0236-4

Comte, J. and del Giorgio, P. A. (2009). Links between resources, c metabolism and the major components of bacterioplankton community structure across a range of freshwater ecosystems. *Environmental Microbiology*, *11*, 1704–1716. http://dx.doi.org/10.1111/j.1462-2920.2009.01897.x

Comte, J. and del Giorgio, P. A. (2010). Linking the patterns of change in composition and function in bacterioplankton successions along environmental gradients. *Ecology*, *91*, 1466–1476.

Comte, J. and del Giorgio, P. A. (2011). Composition influences the pathway but not the outcome of the metabolic response of bacterioplankton to resource shifts. *PLoS ONE*, *6*. http://dx.doi.org/10.1371/journal.pone.0025266

Comte, J., Fauteux, L., Giorgio, P. A. and Amado, A. M. (2013). Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities. *Frontiers in Microbiology*, *4*, 1–11. http://dx.doi.org/10.3389/fmicb.2013.00112

Corilo, Y. (2015). EnviroOrg.

Cory, R. M. and Kaplan, L. A. (2012). Biological lability of streamwater fluorescent dissolved organic matter. *Limnology and Oceanography*, *57*, 1347–1360. biolability FDOM, labile, semilabile, recalcitrant, http://dx.doi.org/10.4319/lo.2012.57.5.1347

Cory, R. M. and Kling, G. W. (2018). Interactions between sunlight and microorganisms influence dissolved organic matter degradation along the aquatic continuum. *Limnology and Oceanography Letters*, *3*, 102–116. http://dx.doi.org/10.1002/lol2.10060

Cottrell, M. T. and Kirchman, D. L. (2000). Natural assemblages of marine proteobacteria and members of the cytophaga-flavobacter cluster consuming low- and high-molecular-weight

dissolved organic matter. *Applied and Environmental Microbiology*, *66*, 1692–1697. http://dx.doi.org/10.1128/AEM.66.4.1692-1697.2000

Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B., Bergamaschi, B. A., Aiken, G. R., Burns, D. A., Findlay, S. E., Shanley, J. B., Striegl, R. G., Aulenbach, B. T., Clow, D. W., Laudon, H., McGlynn, B. L., McGuire, K. J., Smith, R. A. and Stackpoole, S. M. (2015). The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*, 1272–1285. http://dx.doi.org/10.1139/cjfas-2014-0400

Crevecoeur, S., Prairie, Y. T. and del Giorgio, P. A. (2022). Tracking the upstream history of aquatic microbes in a boreal lake yields new insights on microbial community assembly. *PNAS Nexus*, *1*, 1–12. http://dx.doi.org/10.1093/pnasnexus/pgac171

Crump, B. C., Adams, H. E., Hobbie, J. E. and Kling, G. W. (2007). Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology*, *88*, 1365–1378. http://dx.doi.org/10.1890/06-0387

Crump, B. C., Amaral-Zettler, L. A. and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME Journal*, *6*, 1629–1639. http://dx.doi.org/10.1038/ismej.2012.9

Crump, B. C., Kling, G. W., Bahr, M. and Hobbie, J. E. (2003). Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. *Applied and environmental microbiology*, *69*, 2253–2268. http://dx.doi.org/10.1128/AEM.69.4.2253

Currie, D. J. and Kalf, J. (1984). The relative importance of bacterioplankton and phytoplankton in phosphorus uptake in freshwater. *Limnology and Oceanography*, *29*, 311–321. http://dx.doi.org/10.4319/LO.1984.29.2.0311

Czech, L., Barbera, P. and Stamatakis, A. (2020). Genesis and gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics*, *36*, 3263–3265. http://dx.doi.org/10.1093/BIOINFORMATICS/BTAA070

Danczak, R. E., Chu, R. K., Fansler, S. J., Goldman, A. E., Graham, E. B., Tfaily, M. M., Toyoda, J. and Stegen, J. C. (2020). Using metacommunity ecology to understand environmental metabolomes. *Nature Communications*, *11*. http://dx.doi.org/10.1038/s41467-020-19989-y

Dansgaard, W. (1964). Stable isotopes in precipitation. *Tellus*, *16*, 436–468. http://dx.doi.org/10.1111/j.2153-3490.1964.tb00181.x

de Melo, M. L., Bertilsson, S., Amaral, J. H. F., Barbosa, P. M., Forsberg, B. R. and Sarmento, H. (2019). Flood pulse regulation of bacterioplankton community composition in an Amazonian floodplain lake. *Freshwater Biology*, *64*, 108–120. http://dx.doi.org/10.1111/fwb.13198

de Wit, R. and Bouvier, T. (2006). ‘ everything is everywhere, but, the environment selects ’; what did baas becking and beijerinck really say? *Environmental Microbiology*, *8*, 755–758. http://dx.doi.org/10.1111/j.1462-2920.2006.01017.x

del Giorgio, P. A., Bird, D. F., Prairie, Y. T. and Planas, D. (1996). Flow cytometric determination of bacterial abundance in lake plankton with the green nucleic acid stain syto 13. *Limnology and Oceanography*, *41*, 783–789. http://dx.doi.org/10.4319/lo.1996.41.4.0783

del Giorgio, P. A. and Cole, J. J. (1998). Bacterial growth efficiency in natural aquatic systems. *Annual Review of Ecology and Systematics*, *29*, 503–541. http://dx.doi.org/10.1146/annurev.ecolsys.29.1.503

del Giorgio, P. A., Condon, R., Bouvier, T., Longnecker, K., Bouvier, C., Sherr, E. and Gasol, J. M. (2011). Coherent patterns in bacterial growth, growth efficiency, and leucine metabolism along a northeastern pacific inshore-offshore transect. *Limnology and Oceanography*, *56*, 1–16. http://dx.doi.org/10.4319/lo.2011.56.1.0001

del Giorgio, P. A., Pace, M. L. and Fischer, D. (2006). Relationship of bacterial growth efficiency to spatial variation in bacterial activity in the Hudson river. *Aquatic Microbial Ecology*, *45*, 55–67.

Denef, V. J., Fujimoto, M., Berry, M. A. and Schmidt, M. L. (2016). Seasonal succession leads to habitat-dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. *Frontiers in Microbiology*, *7*, 1–13. http://dx.doi.org/10.3389/fmicb.2016.00606

der Gucht, K. V., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., Jeppesen, E., Conde-Porcuna, J.-M., Schwenk, K., Zwart, G., Degans, H., Vyverman, W. and Meester, L. D. (2007). The power of species sorting: Local factors drive bacterial community composition over a wide range of spatial scales. *Proceedings of the National Academy of Sciences*, *104*, 20404–20409. http://dx.doi.org/10.1073/pnas.0707200104

Dittmar, T., Koch, B., Hertkorn, N. and Kattner, G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (spe-DOM) from seawater. *Limnology and Oceanogaphy: Methods*, *6*, 230–235. http://dx.doi.org/10.1515/fabl.2002.019

Docherty, K. M., Young, K. C., Maurice, P. A. and Bridgham, S. D. (2006). Dissolved organic matter concentration and quality influences upon structure and function of freshwater microbial communities. *Microbial Ecology*, *52*, 378–388. http://dx.doi.org/10.1007/s00248-006-9089-x

Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A. V., Medeiros, P. M., Payet, J. P., Richey, J. E., Satinsky, B. M., Sawakuchi, H. O., Ward, N. D., Crump, B. C. and Seymour, J. R. (2017). Bacterial biogeography across the Amazon river-ocean continuum. *Frontiers in Microbiology*. http://dx.doi.org/10.3389/fmicb.2017.00882

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science*, *284*, 2124–2128. http://dx.doi.org/10.1126/science.284.5423.2124

Dow, L., Morrissey, K. L., Willems, A. and Kroth, P. G. (2020). Complete genome sequence of dyadobacter sp. 32, isolated from a culture of the freshwater diatom cymbella microcephala. *Marine Genomics*, *52*. http://dx.doi.org/10.1016/j.margen.2019.100720

Drake, T. W., Raymond, P. A. and Spencer, R. G. (2018). Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnology And Oceanography Letters*, *3*, 132–142. http://dx.doi.org/10.1002/lol2.10055

D'Andrilli, J., Junker, J. R., Smith, H. J., Scholl, E. A. and Foreman, C. M. (2019). DOM composition alters ecosystem function during microbial processing of isolated sources. *Biogeochemistry*, *142*, 281–298. http://dx.doi.org/10.1007/s10533-018-00534-5

Eiler, A., Langenheder, S., Bertilsson, S. and Tranvik, L. J. (2003). Heterotrophic bacterial growth efficiency and community structure at different natural organic carbon concentrations. *Applied and Environmental Microbiology*, *69*, 3701–3709. http://dx.doi.org/10.1128/AEM.69.7.3701-3709.2003

Falkowski, P. G., Fenchel, T. and Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science*, *320*, 1034–1039. http://dx.doi.org/10.1126/science.1153213

Fasching, C., Akotoye, C., Bižić, M., Fonvielle, J., Ionescu, D., Mathavarajah, S., Zoccarato, L., Walsh, D. A., Grossart, H. P. and Xenopoulos, M. A. (2020). Linking stream microbial community functional genes to dissolved organic matter and inorganic nutrients. *Limnology and Oceanography*, *65*, S71–S87. http://dx.doi.org/10.1002/lno.11356

Fasching, C., Behounek, B., Singer, G. A. and Battin, T. J. (2014). Microbial degradation of terrigenous dissolved organic matter and potential consequences for carbon cycling in brownwater streams. *Scientific Reports*, *4*, 1–7. http://dx.doi.org/10.1038/srep04981

Fasching, C., Ulseth, A. J., Schelker, J., Steniczka, G. and Battin, T. J. (2016). Hydrology controls dissolved organic matter export and composition in an alpine stream and its hyporheic zone. *Limnology and Oceanography*, *61*, 558–571. http://dx.doi.org/10.1002/lno.10232

Fernández, A., Huang, S., Seston, S., Xing, J., Hickey, R., Criddle, C. and Tiedje, J. (1999). How stable is stable? function versus community composition. *Applied and Environmental Microbiology*, *65*, 3697–3704. http://dx.doi.org/10.1016/j.jhazmat.2012.09.022

Findlay, S. (2003). *Bacterial Response to Variation in Dissolved Organic Matter*, volume 376, (p. 363–379).

Findlay, S., Sinsabaugh, R. L., Fischer, D. T. and Franchini, P. (1998). Sources of dissolved organic carbon supporting planktonic bacterial production in the tidal freshwater Hudson river. *Ecosystems*, *1*, 227–239.

Fitch, A., Orland, C., Willer, D., Emilson, E. J. and Tanentzap, A. J. (2018). Feasting on terrestrial organic matter: Dining in a dark lake changes microbial decomposition. *Global Change Biology*, *24*, 5110–5122. http://dx.doi.org/10.1111/gcb.14391

Fogg, G. E. (1983). The ecological significance of extracellular products of phytoplankton photosynthesis. *Botanica Marina*, *26*, 3–14.

Fonte, E. S., Amado, A. M., Meirelles-Pereira, F., Esteves, F. A., Rosado, A. S. and Farjalla, V. F. (2013). The combination of different carbon sources enhances bacterial growth efficiency in aquatic ecosystems. *Microbial Ecology*, *66*, 871–878. http://dx.doi.org/10.1007/s00248-013-0277-1

Freeman, E. C., Emilson, E. J., Dittmar, T., Braga, L. P., Emilson, C. E., Goldhammer, T., Martineau, C., Singer, G. and Tanentzap, A. J. (2024). Universal microbial reworking of dissolved organic matter along environmental gradients. *Nature Communications*, *15*. http://dx.doi.org/10.1038/s41467-023-44431-4

Freixa, A., Ejarque, E., Crognale, S., Amalfitano, S., Fazi, S., Butturini, A. and Romaní, A. M. (2016). Sediment microbial communities rely on different dissolved organic matter sources along a mediterranean river continuum. *Limnology and Oceanography*, *61*, 1389–1405. http://dx.doi.org/10.1002/lno.10308

Fukami, T. (2004). Assembly history interacts with ecosystem size to influence species diversity. *Ecology*, *85*, 3234–3242. http://dx.doi.org/10.1890/04-0340

Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 1–23. http://dx.doi.org/10.1146/annurev-ecolsys-110411-160340

Garland, J. L. and Mills, A. L. (1991). Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *57*, 2351–2359.

Garland, J. L., Mills, A. L. and Young, J. S. (2001). Relative effectiveness of kinetic analysis vs single point readings for classifying environmental samples based on community-level physiological profiles (clpp). *Soil Biology and Biochemistry*, *33*, 1059–1066. http://dx.doi.org/10.1016/S0038-0717(01)00011-6

Garner, R. E., Kraemer, S. A., Onana, V. E., Fradette, M., Varin, M. P., Huot, Y. and Walsh, D. A. (2023). A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nature Microbiology*, *8*, 1920–1934. http://dx.doi.org/10.1038/s41564-023-01435-6

Gasol, J. M., Comerma, M., García, J. C., Armengol, J., Casamayor, E. O., Kojecká, P. and Šimek, K. (2002). A transplant experiment to identify the factors controlling bacterial abundance, activity, production, and community composition in a eutrophic canyon-shaped reservoir. *Limnology and Oceanography*, *47*, 62–77. http://dx.doi.org/10.4319/lo.2002.47.1.0062

Gasol, J. M., Pinhassi, J., Alonso-Sáez, L., Ducklow, H., Herndl, G. J., Koblížek, M., Labrenz, M., Luo, Y., Morán, X. A. G., Reinthaler, T. and Simon, M. (2008). Towards a better understanding of microbial carbon flux in the sea. in *Aquatic Microbial Ecology*, volume 53, 21–38. Inter-Research. http://dx.doi.org/10.3354/ame01230

Godwin, C. M. and Cotner, J. B. (2015). Stoichiometric flexibility in diverse aquatic heterotrophic bacteria is coupled to differences in cellular phosphorus quotas. *Frontiers in Microbiology*, *6*. http://dx.doi.org/10.3389/fmicb.2015.00159

Graham, E. B., Knelman, J. E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J. M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J. C., Glanville, H. C., Jones, D. L., Angel, R., Salminen, J., Newton, R. J., Bürgmann, H., Ingram, L. J., Hamer, U., Siljanen, H. M. P., Peltoniemi, K., Potthast, K., Bañeras, L., Hartmann, M., Banerjee, S., Yu, R.-Q., Nogaro, G., Richter, A., Koranda, M., Castle, S. C., Goberna, M., Song, B., Chatterjee, A.,

Nunes, O. C., Lopes, A. R., Cao, Y., Kaisermann, A., Hallin, S., Strickland, M. S., Garcia-Pausas, J., Barba, J., Kang, H., Isobe, K., Papaspyrou, S., Pastorelli, R., Lagomarsino, A., Lindström, E. S., Basiliko, N. and Nemergut, D. R. (2016). Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? *Frontiers in Microbiology*, *7*. http://dx.doi.org/10.3389/fmicb.2016.00214

Graham, L. E., Knack, J. J., Graham, M. E., Graham, J. M. and Zulkifly, S. (2015). A metagenome for lacustrine cladophora (cladophorales) reveals remarkable diversity of eukaryotic epibionts and genes relevant to materials cycling. *Journal of Phycology*, *51*, 408–418. http://dx.doi.org/10.1111/jpy.12296

Grasset, C., Einarsdottir, K., Catalán, N., Tranvik, L. J., Groeneveld, M., Hawkes, J. A. and Attermeyer, K. (2024). Decreasing photoreactivity and concurrent change in dissolved organic matter composition with increasing inland water residence time. *Global Biogeochemical Cycles*, *38*. http://dx.doi.org/10.1029/2023GB007989

Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M. and Beattie, A. J. (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, *432*, 747–750.

Guillemette, F. and del Giorgio, P. A. (2012). Simultaneous consumption and production of fluorescent dissolved organic matter by lake bacterioplankton. *Environmental Microbiology*, *14*, 1432–1443. http://dx.doi.org/10.1111/j.1462-2920.2012.02728.x

Guillemette, F., McCallister, S. L. and Giorgio, P. A. (2016). Selective consumption and metabolic allocation of terrestrial and algal carbon determine allochthony in lake bacteria. *The ISME Journal*, *10*, 1373–1382. http://dx.doi.org/10.1038/ismej.2015.215

Gweon, H. S., Bowes, M. J., Moorhouse, H. L., Oliver, A. E., Bailey, M. J., Acreman, M. C. and Read, D. S. (2020). Contrasting community assembly processes structure lotic bacteria metacommunities along the river continuum. *Environmental Microbiology*. http://dx.doi.org/10.1111/1462-2920.15337

Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M. and Pinhassi, J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environmental Microbiology*, *14*, 2361–2378. http://dx.doi.org/10.1111/j.1462-2920.2012.02804.x

Hahn, M. W. and Pöckl, M. (2005). Ecotypes of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Applied and Environmental Microbiology*, *71*, 766–773. http://dx.doi.org/10.1128/AEM.71.2.766-773.2005

Hall, E. K., Bernhardt, E. S., Bier, R. L., Bradford, M. A., Boot, C. M., Cotner, J. B., del Giorgio, P. A., Evans, S. E., Graham, E. B., Jones, S. E., Lennon, J. T., Locey, K. J., Nemergut, D., Osborne, B. B., Rocca, J. D., Schimel, J. P., Waldrop, M. P. and Wallenstein, M. D. (2018). Understanding how microbiomes influence the systems they inhabit. *Nature Microbiology*, *3*, 977–982. http://dx.doi.org/10.1038/s41564-018-0201-z

Hall, E. K. and Cotner, J. B. (2007). Interactive effect of temperature and resources on carbon cycling by freshwater bacterioplankton communities. *Aquatic Microbial Ecology*, *49*, 35–45. http://dx.doi.org/10.3354/ame01124

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. and Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, *10*, 497–506. http://dx.doi.org/10.1038/nrmicro2795

Hassell, N., Tinker, K. A., Moore, T. and Ottesen, E. A. (2018). Temporal and spatial dynamics in microbial community composition within a temperate stream network. *Environmental Microbiology*, *20*, 3560–3572. http://dx.doi.org/10.1111/1462-2920.14311

Hauptmann, A. L., Markussen, T. N., Stibal, M., Olsen, N. S., Elberling, B., Bælum, J., Sicheritz-Pontén, T. and Jacobsen, C. S. (2016). Upstream freshwater and terrestrial sources are differentially reflected in the bacterial community structure along a small arctic river and its estuary. *Frontiers in Microbiology*, *7*, 1–16. http://dx.doi.org/10.3389/fmicb.2016.01474

Hausmann, B., Pelikan, C., Rattei, T., Loy, A. and Pester, M. (2019). Long-term transcriptional activity at zero growth of a cosmopolitan rare biosphere member. *mBio*, *10*, e02189–18.

Hawkes, J. A., D'Andrilli, J., Agar, J. N., Barrow, M. P., Berg, S. M., Catalán, N., Chen, H., Chu, R. K., Cole, R. B., Dittmar, T., Gavard, R., Gleixner, G., Hatcher, P. G., He, C., Hess, N. J., Hutchins, R. H., Ijaz, A., Jones, H. E., Kew, W., Khaksari, M., Lozano, D. C. P., Lv, J., Mazzoleni, L. R., Noriega-Ortega, B. E., Osterholz, H., Radoman, N., Remucal, C. K., Schmitt, N. D., Schum, S. K., Shi, Q., Simon, C., Singer, G., Sleighter, R. L., Stubbins, A., Thomas, M. J., Tolic, N., Zhang, S., Zito, P. and Podgorski, D. C. (2020). An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer? *Limnology and Oceanography: Methods*, *18*, 235–258. `http://dx.doi.org/10.1002/lom3.10364`

Hengl, T., de Jesus, J., MacMillan, R., Batjes, N., Heuvelink, G., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J., Walsh, M. and Gonzalez, M. (2014). Soilgrids1km—global soil information based on automated mapping. *PLoS ONE*, *9*, e105992. `http://dx.doi.org/10.1371/journal.pone.0105992`. URL `http://isric.org/explore/soilgrids`

Herlemann, D. P. R., Manecki, M., Meeske, C., Pollehne, F., Labrenz, M., Schulz-Bull, D., Dittmar, T. and Jürgens, K. (2014). Uncoupling of bacterial and terrigenous dissolved organic matter dynamics in decomposition experiments. *PLoS ONE*, *9*, e93945. `http://dx.doi.org/10.1371/journal.pone.0093945`

Herndl, G. J., Arrieta, J. M., Kaiser, E., Obernosterer, I., Pausz, C. and Reitner, B. (1999). Role of ultraviolet radiation in aquatic systems: interaction between mixing processes, photochemistry and microbial activity. in C. R. Bell, M. Brylinsky, and P. Johnson-Green (Eds.). *DOC Degradation in Freshwater and Marine Systems: Microbial versus Photochemical Processes*. Atlantic Canada Society for Microbial Ecology.

Herzsprung, P., Hertkorn, N., Tumpling, W. V., Harir, M., Friese, K. and Schmitt-Kopplin, P. (2014). Understanding molecular formula assignment of fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *An-*

alytical and Bioanalytical Chemistry, *406*, 7977–7987. http://dx.doi.org/10.1007/s00216-014-8249-y

Hosen, J. D., Allen, G. H., Amatuli, G., Breitmeyer, S., Cohen, M. J., Crump, B. C., Lu, Y. H., Payet, J. P., Poulin, B. A., Stubbins, A., Yoon, B. and Raymond, P. A. (2021). River network travel time is correlated with dissolved organic matter composition in rivers of the contiguous united states. *Hydrological Processes*, *35*. http://dx.doi.org/10.1002/hyp.14124

Hu, A., Choi, M., Tanentzap, A. J., Liu, J., Jang, K. S., Lennon, J. T., Liu, Y., Soininen, J., Lu, X., Zhang, Y., Shen, J. and Wang, J. (2022). Ecological networks of dissolved organic matter and microorganisms under global change. *Nature Communications*, *13*. http://dx.doi.org/10.1038/s41467-022-31251-1

Hutchins, R. H., Aukes, P., Schiff, S. L., Dittmar, T., Prairie, Y. T. and del Giorgio, P. A. (2017). The optical, chemical, and molecular dissolved organic matter succession along a boreal soil-stream-river continuum. *Journal of Geophysical Research: Biogeosciences*, *122*, 2892–2908. http://dx.doi.org/10.1002/2017JG004094

Ide, J., Ohashi, M., Takahashi, K., Sugiyama, Y., Piirainen, S., Kortelainen, P., Fujitake, N., Yamase, K., Ohte, N., Moritani, M., Hara, M. and Leena, F. (2017). Spatial variations in the molecular diversity of dissolved organic matter in water moving through a boreal forest in eastern finland. *Scientific Reports*, *7*, 1–12. http://dx.doi.org/10.1038/srep42102

IPCC Core Writing Team (2023). Ipcc, 2023: Climate change 2023: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change. http://dx.doi.org/10.59327/IPCC/AR6-9789291691647

Ishii, S. K. and Boyer, T. H. (2012). Behavior of reoccurring parafac components in fluorescent dissolved organic matter in natural and engineered systems: A critical review. *Environmental Science and Technology*, *46*, 2006–2017. http://dx.doi.org/10.1021/es2043504

Jafari, M. and Ansari-Pour, N. (2019). Why, when and how to adjust your p values? *Cell Journal*, *20*, 604–607. http://dx.doi.org/10.22074/cellj.2019.5992

Jansson, M., Bergström, A. K., Lymer, D., Vrede, K. and Karlsson, J. (2006). Bacterioplankton growth and nutrient use efficiencies under variable organic carbon and inorganic phosphorus ratios. *Microbial Ecology*, *52*, 358–364. http://dx.doi.org/10.1007/s00248-006-9013-4

Jansson, M., Hickler, T., Jonsson, A. and Karlsson, J. (2008). Links between terrestrial primary production and bacterial production and respiration in lakes in a climate gradient in subarctic sweden. *Ecosystems*, *11*, 367–376. http://dx.doi.org/10.1007/s10021-008-9127-2

Jia, X., Dini-Andreote, F. and Salles, J. F. (2018). Community assembly processes of the microbial rare biosphere. *Trends in Microbiology*, *26*, 738–747. http://dx.doi.org/10.1016/j.tim.2018.02.011

Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman, D. L., Weinbauer, M. G., Luo, T., Chen, F. and Azam, F. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology*, *8*, 593–599. http://dx.doi.org/10.1038/nrmicro2386

Jones, S. E., Cadkin, T. A., Newton, R. J. and McMahon, K. D. (2012). Spatial and temporal scales of aquatic bacterial beta diversity. *Frontiers in Microbiology*, *3*, 1–10. http://dx.doi.org/10.3389/fmicb.2012.00318

Jones, S. E. and Lennon, J. T. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences*, *107*, 5881–5886. http://dx.doi.org/10.1073/pnas.0912765107

Jones, S. E. and McMahon, K. D. (2009). Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environmental Microbiology*, *11*, 905–913. http://dx.doi.org/10.1111/j.1462-2920.2008.01814.x

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

Judd, K. E., Crump, B. C. and Kling, G. W. (2006). Variation in dissolved organic matter controls bacterial production and community composition. *Ecology*, *87*, 2068–2079.

Kaiser, N. K., Quinn, J. P., Blakney, G. T., Hendrickson, C. L. and Marshall, A. G. (2011). A novel 9.4 tesla fticr mass spectrometer with improved sensitivity, mass resolution, and mass range. *Journal of the American Society for Mass Spectrometry*, *22*, 1343–1351. http://dx.doi.org/10.1007/s13361-011-0141-9

Kajan, K., Osterholz, H., Stegen, J., Udovič, M. G. and Orlić, S. (2023). Mechanisms shaping dissolved organic matter and microbial community in lake ecosystems. *Water Research*, *245*. http://dx.doi.org/10.1016/j.watres.2023.120653

Kamjunke, N., Beckers, L. M., Herzsprung, P., von Tümpling, W., Lechtenfeld, O., Tittel, J., Risse-Buhl, U., Rode, M., Wachholz, A., Kallies, R., Schulze, T., Krauss, M., Brack, W., Comero, S., Gawlik, B. M., Skejo, H., Tavazzi, S., Mariani, G., Borchardt, D. and Weitere, M. (2022). Lagrangian profiles of riverine autotrophy, organic matter transformation, and micropollutants at extreme drought. *Science of the Total Environment*, *828*, 154243. http://dx.doi.org/10.1016/j.scitotenv.2022.154243

Kamjunke, N., Lechtenfeld, O. J. and Herzsprung, P. (2020). Quality of dissolved organic matter driven by autotrophic and heterotrophic microbial processes in a large river. *Water (Switzerland)*, *12*. http://dx.doi.org/10.3390/W12061577

Kassambara, A. (2020a). ggpubr: 'ggplot2' based publication ready plots.

Kassambara, A. (2020b). rstatix: Pipe-friendly framework for basic statistical tests. URL https://cran.r-project.org/package=rstatix

Kassambara, A. and Mundt, F. (2020). factoextra: Extract and visualize the results of multivariate data analyses.

Kawasaki, N. and Benner, R. (2006). Bacterial release of dissolved organic matter during cell growth and decline: Molecular origin and composition. *Limnology and Oceanography*, *51*, 2170–2180. http://dx.doi.org/10.4319/lo.2006.51.5.2170

Kellerman, A. M., Dittmar, T., Kothawala, D. N. and Tranvik, L. J. (2014). Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nature Communications*, *5*, 1–8. http://dx.doi.org/10.1038/ncomms4804

Kellerman, A. M., Guillemette, F., Podgorski, D. C., Aiken, G. R., Butler, K. D. and Spencer, R. G. (2018). Unifying concepts linking dissolved organic matter composition to persistence in aquatic ecosystems. *Environmental Science and Technology*, *52*, 2538–2548. http://dx.doi.org/10.1021/acs.est.7b05513

Kellerman, A. M., Kothawala, D. N., Dittmar, T. and Tranvik, L. J. (2015). Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nature Geoscience*, *8*, 454–457. http://dx.doi.org/10.1038/NGEO2440

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. and Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*, 1463–1464.

Kirchman, D. L. (1993). *Leucine incorporation as a measure of biomass production by heterotrophic bacteria*, (p. 509–512). CRC Press.

Klappenbach, J. A., Dunbar, J. M. and Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Envir. Microbiol.*, *66*, 1328–1333. http://dx.doi.org/10.1128/AEM.66.4.1328-1333.2000.Updated

Koch, B. P. and Dittmar, T. (2006). From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry*, *20*, 926–932. http://dx.doi.org/10.1002/rcm.2386

Koch, B. P., Kattner, G., Witt, M. and Passow, U. (2014). Molecular insights into the microbial formation of marine dissolved organic matter: Recalcitrant or labile? *Biogeosciences*, *11*, 4173–4190. quot;For example, DOM produced by marine microbial communities in culture experiments is very similar to DOM in the ocean regarding molecular formulae [..]quot;, http://dx.doi.org/10.5194/bg-11-4173-2014

Kolmonen, E., Haukka, K., Rantala-Ylinen, A., Rajaniemi-Wacklin, P., Lepistö, L. and Sivonen, K. (2011). Bacterioplankton community composition in 67 finnish lakes differs according to trophic status. *Aquatic Microbial Ecology*, *62*, 241–250. http://dx.doi.org/10.3354/ame01461

Kothawala, D. N., Ji, X., Laudon, H., Ågren, A. M., Futter, M. N., Köhler, S. J. and Tranvik, L. J. (2015). The relative influence of land cover, hydrology, and in-stream processing on the composition of dissolved organic matter in boreal streams. *Journal of Geophysical Research: Biogeosciences*, *120*, 1491–1505. http://dx.doi.org/10.1002/2015JG002946

Kothawala, D. N., Kellerman, A. M., Catalán, N. and Tranvik, L. J. (2021). Organic matter degradation across ecosystem boundaries: The need for a unified conceptualization. *Trends in Ecology and Evolution*, *36*, 113–122. http://dx.doi.org/10.1016/j.tree.2020.10.006

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*, 4453–4455. http://dx.doi.org/10.1093/BIOINFORMATICS/BTZ305

Kraemer, S. A., da Costa, N. B., Shapiro, B. J., Fradette, M., Huot, Y. and Walsh, D. A. (2020). A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. *ISME Journal*, *14*, 3011–3023. http://dx.doi.org/10.1038/s41396-020-0733-0

Kritzberg, E. S., Cole, J. J., Pace, M. M. and Granéli, W. (2005). Does autochthonous primary production drive variability in bacterial metabolism and growth efficiency in lakes dominated by terrestrial c inputs? *Aquatic Microbial Ecology*, *38*, 103–111. http://dx.doi.org/10.3354/ame038103

Kritzberg, E. S., Langenheder, S. and Lindström, E. S. (2006). Influence of dissolved organic matter source on lake bacterioplankton structure and function - implications for seasonal dynamics of community composition. *FEMS Microbiology Ecology*, *56*, 406–417. http://dx.doi.org/10.1111/j.1574-6941.2006.00084.x

Kroer, N. (1993). Bacterial growth efficiency on natural dissolved organic matter. *Limnology and Oceanography*, *38*, 1282–1290. http://dx.doi.org/10.4319/lo.1993.38.6.1282

Kujawinski, E. B., Vecchio, R. D., Blough, N. V., Klein, G. C. and Marshall, A. G. (2004). Probing molecular-level transformations of dissolved organic matter: Insights on photochemical degradation and protozoan modification of DOM from electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry*, *92*, 23–37. http://dx.doi.org/10.1016/j.marchem.2004.06.038

Lambert, T., Teodoru, C. R., Nyoni, F. C., Bouillon, S., Darchambeau, F., Massicotte, P. and Borges, A. V. (2016). Along-stream transport and transformation of dissolved organic matter in a large tropical river. *Biogeosciences*, *13*, 2727–2741. http://dx.doi.org/10.5194/bg-13-2727-2016

Langenheder, S., Lindström, E. S. and Tranvik, L. J. (2005). Weak coupling between community composition and functioning of aquatic bacteria. *Limnology and Oceanography*, *50*, 957–967.

Langenheder, S., Wang, J., Karjalainen, S. M., Laamanen, T. M., Tolonen, K. T., Vilmi, A. and Heino, J. (2017). Bacterial metacommunity organization in a highly connected aquatic system. *FEMS Microbiology Ecology*, *93*, 1–9. http://dx.doi.org/10.1093/femsec/fiw225

Lapierre, J. F. and Giorgio, P. A. D. (2014). Partial coupling and differential regulation of biologically and photochemically labile dissolved organic carbon across boreal aquatic networks. *Biogeosciences*, *11*, 5969–5985. http://dx.doi.org/10.5194/bg-11-5969-2014

Lapierre, J.-F., Guillemette, F., Berggren, M. and del Giorgio, P. A. (2013). Increases in terrestrially derived carbon stimulate organic carbon processing and CO_2 emissions in boreal aquatic ecosystems. *Nature Communications*, *4*. http://dx.doi.org/10.1038/ncomms3972

Larson, J. H., Frost, P. C., Xenopoulos, M. A., Williams, C. J., Morales-Williams, A. M., Vallazza, J. M., Nelson, J. C. and Richardson, W. B. (2014). Relationships between land cover and dissolved organic matter change along the river to lake transition. *Ecosystems*, *17*, 1413–1425. http://dx.doi.org/10.1007/s10021-014-9804-2

Lauro, F. M., Mcdougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., Demaere, M. Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A. C.,

Detter, C., Han, C. S., Brown, M. V., Robb, F. T., Kjelleberg, S. and Cavicchioli, R. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences*, *106*, 15527–15533.

Le, J., Wehr, J. D. and Campbell, L. (1994). Uncoupling of bacterioplankton and phytoplankton production in fresh waters is affected by inorganic nutrient limitation. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *60*, 2086–2093.

Lechtenfeld, O. J., Kattner, G., Flerus, R., McCallister, S. L., Schmitt-Kopplin, P. and Koch, B. P. (2014). Molecular transformation and degradation of refractory dissolved organic matter in the atlantic and southern ocean. *Geochimica et Cosmochimica Acta*, *126*, 321–337. `http://dx.doi.org/10.1016/j.gca.2013.11.009`

Lee, J. (2010). *Statistical bioinformatics* (1st ed.). John Wiley & Sons Inc.

Legendre, P. and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*, 271–280. `http://dx.doi.org/10.1007/s004420100716`

Legendre, P. and Legendre, L. (1998). *Numerical Ecology* (2nd ed.). Elsevier.

Lehner, B. and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*, 2171–2186. URL `www.hydrosheds.org`

Lennon, J. T., Abramoff, R. Z., Allison, S. D., Burckhardt, R. M., DeAngelis, K. M., Dunne, J. P., Frey, S. D., Friedlingstein, P., Hawkes, C. V., Hungate, B. A., Khurana, S., Kivlin, S. N., Levine, N. M., Manzoni, S., Martiny, A. C., Martiny, J. B., Nguyen, N. K., Rawat, M., Talmy, D., Todd-Brown, K., Vogt, M., Wieder, W. R. and Zakem, E. J. (2024). Priorities, opportunities, and challenges for integrating microorganisms into earth system models for climate change prediction. `http://dx.doi.org/10.1128/mbio.00455-24`

Lennon, J. T. and Cottingham, K. L. (2008). Microbial productivity in variable resource environments. *Ecology*, *89*, 1001–1014. `http://dx.doi.org/10.1890/07-1380.1`

Lennon, J. T. and Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, *9*, 119–130. http://dx.doi.org/10.1038/nrmicro2504

Lennon, J. T., Muscarella, M. E., Placella, S. A. and Lehmkuhl, B. K. (2018). How, when, and where relic DNA affects microbial diversity. *mBio*, *9*, e00637–18. http://dx.doi.org/10.1128/mBio.00637-18

Lennon, J. T. and Pfaff, L. E. (2005). Source and supply of terrestrial organic matter affects aquatic microbial metabolism. *Aquatic Microbial Ecology*, *39*, 107–119.

Letunic, I. and Bork, P. (2006). Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*, 127–128.

Lindström, E. S., Feng, X. M., Grane, W. and Kritzberg, E. S. (2010). The interplay between bacterial community composition and the environment determining function of inland water bacteria. *55*, 2052–2060. http://dx.doi.org/10.4319/lo.2010.55.5.2052

Lindström, E. S., Forslund, M., Algesten, G. and Bergström, A.-K. (2006). External control of bacterial community structure in lakes. *Limnology and Oceanography*, *51*, 339–342.

Liu, S., He, Z., Tang, Z., Liu, L., Hou, J., Li, T., Zhang, Y., Shi, Q., Giesy, J. P. and Wu, F. (2020). Linking the molecular composition of autochthonous dissolved organic matter to source identification for freshwater lake ecosystems by combination of optical spectroscopy and ft-icr-ms analysis. *Science of the Total Environment*, *703*. http://dx.doi.org/10.1016/j.scitotenv.2019.134764

Locey, K. J. and Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, *113*, 5972–5975. http://dx.doi.org/10.1073/pnas.1521291113

Logue, J. B. and Lindström, E. S. (2008). Biogeography of bacterioplankton in inland waters. *Freshwater Reviews*, *1*, 99–114. http://dx.doi.org/10.1608/FRJ-1.1.9

Logue, J. B. and Lindström, E. S. (2010). Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *The ISME Journal*, *4*, 729–738. http://dx.doi.org/10.1038/ismej.2009.156

Logue, J. B., Stedmon, C. A., Kellerman, A. M., Nielsen, N. J., Andersson, A. F., Laudon, H., Lindström, E. S. and Kritzberg, E. S. (2016). Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. *The ISME Journal*, *10*, 1–13. From Duplicate 2 (Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter - Logue, Jürg Brendan; Stedmon, Colin a; Kellerman, Anne M; Nielsen, Nikoline J; Andersson, Anders F; Laudon, Hjalmar; Lindström, Eva S; Kritzberg, Emma S)

DOC + abundance
log 10 (cells) ml-1

DOC 17.5 - 7.5, http://dx.doi.org/10.1038/ismej.2015.131

Lorenzen, C. (1967). Determination of chlorophyll and pheo-pigments: spectrophotometric equations. *Limnology and Oceanography*, *12*, 343–346.

Luo, X., Xiang, X., Yang, Y., Huang, G., Fu, K., Che, R. and Chen, L. (2020). Seasonal effects of river flow on microbial community coalescence and diversity in a riverine network. *FEMS Microbiology Ecology*, *96*, 1–13. http://dx.doi.org/10.1093/femsec/fiaa132

Lynch, L. M., Sutfin, N. A., Fegel, T. S., Boot, C. M., Covino, T. P. and Wallenstein, M. D. (2019). River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. *Nature Communications*, *10*. http://dx.doi.org/10.1038/s41467-019-08406-8

Maavara, T., Lauerwald, R., Regnier, P. and Cappellen, P. V. (2017). Global perturbation of organic carbon cycling by river damming. *Nature Communications*, *8*. http://dx.doi.org/10.1038/ncomms15347

Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*, 714–716. http://dx.doi.org/10.1038/nature01547

Mansour, I., Heppell, C. M., Ryo, M. and Rillig, M. C. (2018). Application of the microbial community coalescence concept to riverine networks. *Biological Reviews*, *93*, 1832–1845. http://dx.doi.org/10.1111/brv.12422

Mao, Z., Han, Y., Xun, F., An, S., Li, B., Wang, Y., Chen, H., Wu, Q. L. and Xing, P. (2025). Warming effects on pelagic carbon metabolism is related to substrate composition and bacterioplankton community history. *Water Research*, *270*. http://dx.doi.org/10.1016/j.watres.2024.122846

Maranger, R. J., Pace, M. L., Giorgio, P. A. D., Caraco, N. F. and Cole, J. J. (2005). Longitudinal spatial patterns of bacterial production and respiration in a large river-estuary: Implications for ecosystem carbon consumption. *Ecosystems*, *8*, 318–330. http://dx.doi.org/10.1007/s10021-003-0071-x

Marchand, D., Prairie, Y. T. and del Giorgio, P. A. (2009). Linking forest fires to lake metabolism and carbon dioxide emissions in the boreal region of northern québec. *Global Change Biology*, *15*, 2861–2873. http://dx.doi.org/10.1111/j.1365-2486.2009.01979.x

Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, pp. 10–12. http://dx.doi.org/10.14806/ej.17.1.200

Martiny, A. C., Treseder, K. and Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, *7*, 830–838. http://dx.doi.org/10.1038/ismej.2012.160

Martiny, J. B., Jones, S. E., Lennon, J. T. and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science*, *350*. http://dx.doi.org/10.1126/science.aac9323

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., louise Reysenbach, A., Smith, V. H. and Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, *4*, 102–112. http://dx.doi.org/10.1038/nrmicro1341

Massicotte, P., Asmala, E., Stedmon, C. and Markager, S. (2017). Global distribution of dissolved organic matter along the aquatic continuum : Across rivers , lakes and oceans. *Science of the Total Environment*, *609*, 180–191. http://dx.doi.org/10.1016/j.scitotenv.2017.07.076

Massicotte, P. and Frenette, J.-J. (2011). Spatial connectivity in a large river system : resolving the sources and fate of dissolved organic matter. *Ecological Applications*, *21*, 2600–2617.

McDougald, D., Rice, S. A., Barraud, N., Steinberg, P. D. and Kjelleberg, S. (2011). Should we stay or should we go : mechanisms and ecological consequences for biofilm dispersal. *Nature Publishing Group*, *10*, 39–50. http://dx.doi.org/10.1038/nrmicro2695

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I. and White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*, 995–1015. http://dx.doi.org/10.1111/j.1461-0248.2007.01094.x

McLaughlin, C. and Kaplan, L. A. (2013). Biological lability of dissolved organic carbon in stream water and contributing terrestrial sources. *Freshwater Science*, *32*, 1219–1230. http://dx.doi.org/10.1899/12-202.1

McMurdie, P. J. and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*, e61217.

Medeiros, P. M., Seidel, M., Gifford, S. M., Ballantyne, F., Dittmar, T., Whitman, W. B. and Moran, M. A. (2017). Microbially-mediated transformations of estuarine dissolved organic matter. *Frontiers in Marine Science*, *4*. http://dx.doi.org/10.3389/fmars.2017.00069

Mentges, A., Feenders, C., Seibt, M. and Blasius, B. (2017). Functional molecular diversity of marine dissolved organic matter is reduced during degradation. *Frontiers in Marine Science*, *4*, 1–10. http://dx.doi.org/10.3389/fmars.2017.00194

Merder, J., Freund, J. A., Feudel, U., Hansen, C. T., Hawkes, J. A., Jacob, B., Klaproth, K., Niggemann, J., Noriega-Ortega, B. E., Osterholz, H., Rossel, P. E., Seidel, M., Singer, G., Stubbins, A., Waska, H. and Dittmar, T. (2020). Icbm-ocean: Processing ultrahigh-resolution mass spectrometry data of complex molecular mixtures. *Analytical Chemistry*, *92*, 6832–6838. http://dx.doi.org/10.1021/acs.analchem.9b05659

Messager, M. L., Lehner, B., Grill, G., Nedeva, I. and Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*, 1–11. http://dx.doi.org/10.1038/ncomms13603

Microsoft and Weston, S. (2022). foreach: Provides foreach looping construct. URL https://CRAN.R-project.org/package=foreach

Miki, T., Yokokawa, T., Ke, P. J., Hsieh, I. F., hao Hsieh, C., Kume, T., Yoneya, K. and Matsui, K. (2018). Statistical recipe for quantifying microbial functional diversity from ecoplate metabolic profiling. *Ecological Research*, *33*, 249–260. http://dx.doi.org/10.1007/s11284-017-1554-0

Monard, C., Gantner, S., Bertilsson, S., Hallin, S. and Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports*, *6*, 37719. http://dx.doi.org/10.1038/srep37719

Moran, M. A., Pomeroy, L. R., Sheppard, E. S., Atkinson, L. P. and Hodson, R. E. (1991). Distribution of terrestrially derived dissolved organic matter on the southeastern u.s. continental shelf. *Limnology and Oceanography*, *36*, 1134–1149. http://dx.doi.org/10.4319/lo.1991.36.6.1134

Morrissey, E. M., Mau, R. L., Schwartz, E., Caporaso, J. G., DIjkstra, P., Gestel, N. V., Koch, B. J., Liu, C. M., Hayer, M., McHugh, T. A., Marks, J. C., Price, L. B. and Hungate, B. A. (2016). Phylogenetic organization of bacterial activity. *ISME Journal*, *10*, 2336–2340. http://dx.doi.org/10.1038/ismej.2016.28

Mou, X., Sun, S., Edwards, R. A., Hodson, R. E. and Moran, M. A. (2008). Bacterial carbon

processing by generalist species in the coastal ocean. *Nature*, *451*, 708–711. http://dx.doi.org/10.1038/nature06513

Murali, A., Bhargava, A. and Wright, E. S. (2018). Idtaxa: A novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, *6*, 1–14. http://dx.doi.org/10.1186/s40168-018-0521-5

Muscarella, M. E., Boot, C. M., Broeckling, C. D. and Lennon, J. T. (2019). Resource heterogeneity structures aquatic bacterial communities. *ISME Journal*, *13*, 2183–2195. http://dx.doi.org/10.1038/s41396-019-0427-7

Muscarella, M. E., Howey, X. M. and Lennon, J. T. (2020). Trait-based approach to bacterial growth efficiency. *Environmental Microbiology*, *22*, 3494–3504. http://dx.doi.org/10.1111/1462-2920.15120

Nagata, T. and Kirchman, D. L. (1999). Mortality of microbes in aquatic environments bacterial mortality: A pathway for the formation of refractory DOM? in C. R. Bell, M. Brylinsky, and P. Johnson-Green (Eds.). *Microbial Biosystems: New Frontiers. Proceedings of the 8th International Symposium on Microbial Ecology*. Atlantic Canada Society for Microbial Ecology.

Nakadai, R., Okazaki, Y. and Matsuoka, S. (2020). Describing macroecological patterns in microbes: Approaches for comparative analyses of operational taxonomic unit read number distribution with a case study of global oceanic bacteria. *Environmental DNA*, *2*, 535–543. http://dx.doi.org/10.1002/edn3.78

Natural Resources Canada (2017). Canadian digital elevation model. URL https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333

Nelson, C. E., Sadro, S. and Melack, J. M. (2009). Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. *Limnology and Oceanography*, *54*, 1292–1305. http://dx.doi.org/10.4319/lo.2009.54.4.1292

Nelson, C. E. and Wear, E. K. (2014). Microbial diversity and the lability of dissolved organic carbon.

Proceedings of the National Academy of Sciences, *111*, 7166–7167. http://dx.doi.org/10.1073/pnas.1405751111

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., Fierer, N., Townsend, A. R., Cleveland, C. C., Stanish, L. and Knight, R. (2011). Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology*, *13*, 135–144. http://dx.doi.org/10.1111/j.1462-2920.2010.02315.x

Niño-García, J. P., Ruiz-Gonzalez, C., del Giorgio, P. A., Ruiz-González, C. and del Giorgio, P. A. (2017). Exploring the ecological coherence between the spatial and temporal patterns of bacterioplankton in boreal lakes. *Frontiers in Microbiology*, *8*, 1–12. http://dx.doi.org/10.3389/fmicb.2017.00636

Niño-García, J. P., Ruiz-González, C. and del Giorgio, P. A. (2016a). Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal*, *10*, 1755–1766. http://dx.doi.org/10.1038/ismej.2015.226

Niño-García, J. P., Ruiz-González, C. and del Giorgio, P. A. (2016b). Landscape-scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. *Ecology Letters*, *19*, 1506–1515. http://dx.doi.org/10.1111/ele.12704

Noriega-Ortega, B. E., Wienhausen, G., Mentges, A., Dittmar, T., Simon, M. and Niggemann, J. (2019). Does the chemodiversity of bacterial exometabolomes sustain the chemodiversity of marine dissolved organic matter? *Frontiers in Microbiology*, *10*, 1–13. http://dx.doi.org/10.3389/fmicb.2019.00215

Ochman, H., Lawrence, J. G. and Grolsman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*, 299–304. http://dx.doi.org/10.1038/35012500

Ogawa, H., Amagai, Y., Koike, I., Kaiser, K. and Ronald, B. (2001). Production of refractory dissolved organic matter by bacteria. *Science*, *292*, 917–920. http://dx.doi.org/10.1126/science.1057627

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. (2019). vegan: Community Ecology Package. URL https://cran.r-project.org/package=vegan

Olapade, O. A. and Leff, L. G. (2005). Seasonal response of stream biofilm communities to dissolved organic matter and nutrient enrichments. *Applied and Environmental Microbiology*, *71*, 2278–2287. http://dx.doi.org/10.1128/AEM.71.5.2278âĂŞ2287.2005

Oliver, A. A., Spencer, R. G., Deas, M. L. and Dahlgren, R. A. (2016). Impact of seasonality and anthropogenic impoundments on dissolved organic matter dynamics in the Klamath river (oregon/california, USA). *Journal of Geophysical Research: Biogeosciences*, *121*, 1946–1958. http://dx.doi.org/10.1002/2016JG003497

Orlova, J., Amiri, F., Bourgeois, A. K., Buttle, J. M., Cherlet, E., Cuss, C. W., Devito, K. J., Emelko, M. B., Floyd, W. C., Foster, D. E., Hutchins, R. H. S., Jamieson, R., Johnson, M. S., McSorley, H. J., Silins, U., Tank, S. E., Thompson, L. M., Webster, K. L., Williams, C. H. S. and Olefeldt, D. (2024). Composition of stream dissolved organic matter across canadian forested ecozones varies in three dimensions linked to landscape and climate. *Water Resources Research*, *60*. http://dx.doi.org/10.1029/2023WR035196

Osterholz, H., Niggemann, J., Giebel, H. A., Simon, M. and Dittmar, T. (2015). Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nature Communications*, *6*. http://dx.doi.org/10.1038/ncomms8422

Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J. and Dittmar, T. (2016). Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *The ISME Journal*, *10*, 1717–1730. http://dx.doi.org/10.1038/ismej.2015.231

Papke, R. T. and Gogarten, J. P. (2012). How bacterial lineages emerge. *Science*, *336*, 45–46. http://dx.doi.org/10.1126/science.1219241

Paradis, E. and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A. and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*, 996. http://dx.doi.org/10.1038/nbt.4229

Paruch, L., Paruch, A. M., Eiken, H. G., Skogen, M. and Sørheim, R. (2020). Seasonal dynamics of lotic bacterial communities assessed by 16S rRNA gene amplicon deep sequencing. *Scientific Reports*, *10*, 16399. http://dx.doi.org/10.1038/s41598-020-73293-9

Patton, C. and Kryskalla, J. (2003). Methods of analysis by the u.s. geological survey national water quality laboratory – evaluation of alkaline persulfate digestion as an alternative to kjedahl digestion for determination of total and dissolved nitrogen and phosphorus.

Paulson, J. N., Stine, O. C., Bravo, H. C. and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*, 1200–1202. http://dx.doi.org/10.1038/nmeth.2658

Pedersen, T. L. (2024). patchwork: The composer of plots.

Pedler, B. E., Aluwihare, L. I. and Azam, F. (2014). Single bacterial strain capable of significant contribution to carbon cycling in the surface ocean. *Proceedings of the National Academy of Sciences*, *111*, 7202–7207. http://dx.doi.org/10.1073/pnas.1401887111

Pedrós-Alió, C. (2012). The rare bacterial biosphere. *Annual Review of Marine Science*, *4*, 449–466. http://dx.doi.org/10.1146/annurev-marine-120710-100948

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

Peter, H., Jeppesen, E., Meester, L. D. and Sommaruga, R. (2018). Changes in bacterioplankton community structure during early lake ontogeny resulting from the retreat of the greenland ice sheet. *The ISME Journal*, *12*, 544–555. http://dx.doi.org/10.1038/ismej.2017.191

Peter, H., Singer, G., Ulseth, A. J., Dittmar, T., Prairie, Y. T. and Battin, T. J. (2020). Travel time and source variation explain the molecular transformation of dissolved organic matter in an alpine stream network. *Journal of Geophysical Research: Biogeosciences*, *125*, 1–17. http://dx.doi.org/10.1029/2019JG005616

Planas, D., Paquet, S. and Saint-Pierre, A. (2005). *Production-consumption of CO_2 in reservoirs and lakes in relation to plankton metabolism*, (p. 483–507). Environmental Science, Springer

Pollard, P. C. (2013). In situ rapid measures of total respiration rate capture the super labile DOC bacterial substrates of freshwater. *Limnology and Oceanography: Methods*, *11*, 584–593. http://dx.doi.org/10.4319/lom.2013.11.584

Pomeroy, L. R., Sheldon, J. E. and Sheldon, W. M. (1994). Changes in bacterial numbers and leucine assimilation during estimations of microbial respiratory rates in seawater by the precision winkler method.

Pradeep Ram, A. S., Billard, H., Perriere, F., Voldoire, O. and Colombet, J. (2024). Impact of top-down regulation on the growth efficiency of freshwater bacterioplankton. *Microorganisms*, *12*, 2061. http://dx.doi.org/10.3390/microorganisms12102061

Pradeep Ram, A. S., Chaibi-Slouma, S., Keshri, J., Colombet, J. and Sime-Ngando, T. (2016). Functional responses of bacterioplankton diversity and metabolism to experimental bottom-up and top-down forcings. *Microbial Ecology*, *72*, 347–358. http://dx.doi.org/10.1007/s00248-016-0782-0

Proia, L., Schiller, D. V., Gutierrez, C. and Marcé, R. (2016). Microbial carbon processing along a river discontinuum. *Freshwater Science*, *35*, 1133–1147. http://dx.doi.org/10.1086/689181.

Pruesse, E., Peplies, J. and Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*, 1823–1829. http://dx.doi.org/10.1093/bioinformatics/bts252

Pérez, M. T. and Sommaruga, R. (2006). Differential effect of algal- and soil-derived dissolved

organic matter on alpine lake bacterial community composition and activity. *Limnology and Oceanography*, *51*, 2527–2537. http://dx.doi.org/10.4319/lo.2006.51.6.2527

R Core Team (2024). R: A language and environment for statistical computing. URL https://www.r-project.org/

Rainey, P. B. and Travisano, M. (1998). Adaptive radiation in a hetergeneous environment. *Nature*, *394*, 69–72. http://dx.doi.org/10.1038/27900

Ramin, K. I. and Allison, S. D. (2019). Bacterial tradeoffs in growth rate and extracellular enzymes. *Frontiers in Microbiology*, *10*. http://dx.doi.org/10.3389/fmicb.2019.02956

Raymond, P. A., Saiers, J. E. and Sobczak, W. V. (2016). Hydrological and biogeochemical controls on watershed dissolved organic matter transport: Pulse-shunt concept. *Ecology*, *97*, 5–16. http://dx.doi.org/10.1890/14-1684.1

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J. and Griffiths, R. I. (2015). Catchment-scale biogeography of riverine bacterioplankton. *The ISME Journal*, *9*, 516–526. http://dx.doi.org/10.1038/ismej.2014.166

Reed, D. C., Algar, C. K., Huber, J. A. and Dick, G. J. (2014). Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 1879–1884. http://dx.doi.org/10.1073/pnas.1313713111

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., Laruelle, G. G., Lauerwald, R., Luyssaert, S., Andersson, A. J., Arndt, S., Arnosti, C., Borges, A. V., Dale, A. W., Gallego-Sala, A., Goddéris, Y., Goossens, N., Hartmann, J., Heinze, C., Ilyina, T., Joos, F., LaRowe, D. E., Leifeld, J., Meysman, F. J., Munhoven, G., Raymon, P. A., Spahni, R., Suntharalingam, P. and Thullner, M. (2013). Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nature Geoscience*, *6*, 597–607. athropogenic impacts on carbon flux in streams, http://dx.doi.org/10.1038/ngeo1830

Reis, P. C., Thottathil, S. D., Ruiz-González, C. and Prairie, Y. T. (2020). Niche separation within aerobic methanotrophic bacteria across lakes and its link to methane oxidation rates. *Environmental Microbiology*, *22*, 738–751. http://dx.doi.org/10.1111/1462-2920.14877

Repeta, D. J. (2015). *Chemical Characterization and Cycling of Dissolved Organic Matter* (second edi ed.). Elsevier Inc. http://dx.doi.org/10.1016/B978-0-12-405940-5.00002-9

Repeta, D. J., Quan, T. M., Aluwihare, L. I. and Accardi, A. M. (2002). Chemical characterization of high molecular weight dissolved organic matter in fresh and marine waters. *Geochimica et Cosmochimica Acta*, *66*, 955–962. - [..] there is a component of DOM that is molecularly indistinguishable., http://dx.doi.org/10.1016/S0016-7037(01)00830-4

Revolution Analytics and Weston, S. (2019). doMC: Foreach parallel adaptor for 'parallel'. URL https://cran.r-project.org/package=doMC

Riedel, T., Biester, H. and Dittmar, T. (2012). Molecular fractionation of dissolved organic matter with metal salts. *Environmental Science and Technology*, *46*, 4419–4426. http://dx.doi.org/10.1021/es203901u

Ritz, C., Baty, F., Streibig, J. and Gerhard, D. (2015). Dose-response analysis using R. *PLOS One*, *10*, e0146021.

Rocca, J. D., Hall, E. K., Lennon, J. T., Evans, S. E., Waldrop, M. P., Cotner, J. B., Nemergut, D. R., Graham, E. B. and Wallenstein, M. D. (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. http://dx.doi.org/10.1038/ismej.2014.252

Rodibaugh, K. J., Becker, J. C., Timmins, G. and Nowlin, W. H. (2020). Physicochemical and carbon quantity–quality gradients equally influence bacterial carbon metabolism across an arid riverscape. *Aquatic Ecology*, *54*, 677–696. http://dx.doi.org/10.1007/s10452-020-09767-9

Roiha, T., Peura, S., Cusson, M. and Rautio, M. (2016). Allochthonous carbon is a major regulator to bacterial growth and community composition in subarctic freshwaters. *Scientific Reports*, *6*. http://dx.doi.org/10.1038/srep34456

Roland, F. and Cole, J. J. (2000). Prediction of bacterical growth efficiency in a riverine ecosystem. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, *27*, 2846–2847. http://dx.doi.org/10.1080/03680770.1998.11898188

RStudio Team (2024). Rstudio: Integrated development environment for R. URL http://www.posit.co/

Ruiz-González, C., Niño-García, J. P., Berggren, M. and Giorgio, P. A. D. (2017a). Contrasting dynamics and environmental controls of dispersed bacteria along a hydrologic gradient. *Advances in Oceanography and Limnology*, *8*, 222–234. http://dx.doi.org/10.4081/aiol.2017.7232

Ruiz-González, C., Niño-García, J. P. and del Giorgio, P. A. (2015a). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters*, *18*, 1198–1206. http://dx.doi.org/10.1111/ele.12499

Ruiz-González, C., Niño-García, J. P., Kembel, S. W. and del Giorgio, P. A. (2017b). Identifying the core seed bank of a complex boreal bacterial metacommunity. *ISME Journal*, *11*, 2012–2021. http://dx.doi.org/10.1038/ismej.2017.67

Ruiz-González, C., Niño-García, J. P., Lapierre, J.-F. and del Giorgio, P. A. (2015b). The quality of organic matter shapes the functional biogeography of bacterioplankton across boreal freshwater ecosystems. *Global Ecology and Biogeography*, *24*, 1487–1498. http://dx.doi.org/10.1111/geb.12356

Ruiz-González, C., Proia, L., Ferrera, I., Gasol, J. M. and Sabater, S. (2013). Effects of large river dam regulation on bacterioplankton community structure. *FEMS Microbiology Ecology*, *84*, 316–331. http://dx.doi.org/10.1111/1574-6941.12063

Ruiz-González, C., Salazar, G., Logares, R., Proia, L., Gasol, J. M. and Sabater, S. (2015c). Weak coherence in abundance patterns between bacterial classes and their constituent otus along a regulated river. *Frontiers in Microbiology*, *6*, 1–13. http://dx.doi.org/10.3389/fmicb.2015.01293

Rust, F., Bodmer, P. and del Giorgio, P. (2022). Modeling the spatial and temporal variability in surface water CO_2 and CH_4 concentrations in a newly created complex of boreal hydroelectric reservoirs. *Science of the Total Environment*, *815*, 152459. http://dx.doi.org/10.1016/j.scitotenv.2021.152459

Ryan, K. A., Garayburu-Caruso, V. A., Crump, B. C., Bambakidis, T., Raymond, P. A., Liu, S. and Stegen, J. C. (2024). Riverine dissolved organic matter transformations increase with watershed area, water residence time, and damköhler numbers in nested watersheds. *Biogeochemistry*. http://dx.doi.org/10.1007/s10533-024-01169-5

Savio, D., Sinclair, L., Ijaz, U. Z., Parajka, J., Reischer, G. H., Stadler, P., Blaschke, A. P., Blöschl, G., Mach, R. L., Kirschner, A. K. T., Farnleitner, A. H. and Eiler, A. (2015). Bacterial diversity along a 2600 km river continuum. *Environmental Microbiology*, *17*, 4994–5007. http://dx.doi.org/10.1111/1462-2920.12886

Savory, J. J., Kaiser, N. K., McKenna, A. M., Xian, F., Blakney, G. T., Rodgers, R. P., Hendrickson, C. L. and Marshall, A. G. (2011). Parts-per-billion fourier transform ion cyclotron resonance mass measurement accuracy with a walking calibration equation. *Analytical Chemistry*, *83*, 1732–1736. http://dx.doi.org/10.1021/ac102943z

Schimel, J. P., Bennett, J. and Fierer, N. (2005). *Microbial community composition and soil nitrogen cycling: is there really a connection?*, (p. 171–188). Cambridge University Press

Schmidt, M. L., White, J. D. and Denef, V. J. (2016). Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. *Environmental Microbiology*, *18*, 1212–1226. http://dx.doi.org/10.1111/1462-2920.13143

Seidel, M., Yager, P. L., Ward, N. D., Carpenter, E. J., Gomes, H. R., Krusche, A. V., Richey, J. E., Dittmar, T. and Medeiros, P. M. (2015). Molecular-level changes of dissolved organic matter along the Amazon river-to-ocean continuum. *Marine Chemistry*. http://dx.doi.org/10.1016/j.marchem.2015.06.019

Severin, I., Lindström, E. S. and Örjan Östman (2014). Relationships between bacterial community composition, functional trait composition and functioning are context dependent -

but what is the context? *PLoS ONE*, *9*, 1–10. http://dx.doi.org/10.1371/journal.pone.0112409

Shade, A., Caporaso, J. G., Handelsman, J., Knight, R. and Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME Journal*, *7*, 1493–1506. http://dx.doi.org/10.1038/ismej.2013.54

Shade, A. and Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, *23*, 335–340. http://dx.doi.org/10.1016/j.tim.2015.01.007

Sievert, C. (2020). Interactive web-based data visualization with r, plotly, and shiny. URL https://plotly-r.com

Singer, E., Wagner, M. and Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. *ISME Journal*, *11*, 1949–1963. http://dx.doi.org/10.1038/ismej.2017.59

Singh, S., Inamdar, S., Mitchell, M. and McHale, P. (2014). Seasonal pattern of dissolved organic matter (DOM) in watershed sources: Influence of hydrologic flow paths and autumn leaf fall. *Biogeochemistry*, *118*, 321–337. http://dx.doi.org/10.1007/s10533-013-9934-1

Sleighter, R. L. and Hatcher, P. G. (2008). Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower chesapeake bay by ultrahigh resolution electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry*, *110*, 140–152. http://dx.doi.org/10.1016/j.marchem.2008.04.008

Smith, D. P. and Peay, K. G. (2014). Sequence depth, not pcr replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, *9*. http://dx.doi.org/10.1371/journal.pone.0090234

Smith, E. M. and Prairie, Y. T. (2004). Bacterial metabolism and growth efficiency in lakes: The importance of phosphorus availability. *Limnology and Oceanography*, *49*, 137–147. http://dx.doi.org/10.4319/lo.2004.49.1.0137

Smith, S. V. and Hollibaugh, J. T. (1993). Coastal metabolism and the oceanic organic carbon balance. *Reviews of Geophysics*, *31*, 75–89. http://dx.doi.org/10.1029/92RG02584

Soininen, J. (2010). Species turnover along abiotic and biotic gradients: Patterns in space equal patterns in time? *BioScience*, *60*, 433–439. http://dx.doi.org/10.1525/bio.2010.60.6.7

Soininen, J., Korhonen, J. J., Karhu, J. and Vetterli, A. (2011). Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnology and Oceanography*, *56*, 508–520. http://dx.doi.org/10.4319/lo.2011.56.2.0508

Spencer, R. G. M., Aiken, G. R., Wickland, K. P., Striegl, R. G. and Hernes, P. J. (2008). Seasonal and spatial variability in dissolved organic matter quantity and composition from the yukon river basin , alaska. *Global Biogeochemical Cycles*, *22*, 1–13. DOM quantity and chemical composition, CDOM, dissolved lignin phenols, http://dx.doi.org/10.1029/2008GB003231

Stadler, M. (2021). Carbbas/paper_stadler-delgiorgio_ismej_2021. http://dx.doi.org/10.5281/zenodo.5567270

Stadler, M., Barnard, M. A., Bice, K., Melo, M. L. D., Dwivedi, D., Freeman, E. C., Garayburu-caruso, V. A., Linkhorst, A., Mateus-barros, E., Shi, C., Tanentzap, A. J. and Meile, C. (2023). Applying the core-satellite species concept : Characteristics of rare and common riverine dissolved organic matter. *Frontiers in Water*, p. 1156042. http://dx.doi.org/10.3389/frwa.2023.1156042

Stadler, M. and del Giorgio, P. A. (2022). Terrestrial connectivity , upstream aquatic history and seasonality shape bacterial community assembly within a large boreal aquatic network. *The ISME Journal*, *16*, 937–947. http://dx.doi.org/10.1038/s41396-021-01146-y

Stadler, M., Ejarque, E. and Kainz, M. J. (2020). In-lake transformations of dissolved organic matter composition in a subalpine lake do not change its biodegradability. *Limnology and Oceanography*, *65*, 1554–1572. http://dx.doi.org/10.1002/lno.11406

Stadler, M., Ruiz-González, C., Vick-Majors, T. J. and del Giorgio, P. A. (2021). Microbial 16S rRNA gene (DNA) and transcripts (cDNA) along a boreal soil-freshwater-estuary continuum. http://dx.doi.org/10.5281/zenodo.4611420

Staley, J. T. and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*, 321–346.

Stedmon, C. A. and Markager, S. (2005). Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using parafac analysis. *Limnology and Oceanography*, *50*, 686–697.

Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., Rockhold, M. L. and Konopka, A. (2013). Quantifying community assembly processes and identifying features that impose them. *The ISME Journal*, *7*, 2069–2079. http://dx.doi.org/10.1038/ismej.2013.93

Stegen, J. C., Lin, X., Fredrickson, J. K. and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology*, *6*, 1–15. http://dx.doi.org/10.3389/fmicb.2015.00370

Steiner, P. A., Corte, D. D., Geijo, J., Mena, C., Yokokawa, T., Rattei, T., Herndl, G. J. and Sintes, E. (2019). Highly variable mRNA half-life time within marine bacterial taxa and functional genes. *Environmental Microbiology*, *21*, 3873–3884. http://dx.doi.org/10.1111/1462-2920.14737

Steven, B., Hesse, C., Soghigian, J., Gallegos-Graves, L. V. and Dunbar, J. (2017). Simulated rRNA/DNA ratios show potential to misclassify active populations as dormant. *Applied and Environmental Microbiology*, *83*, 1–11. http://dx.doi.org/10.1128/AEM.00696-17

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. and Schmidt, M. (2017). rrnDB : improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *43*, 593–598. http://dx.doi.org/10.1093/nar/gku1201

Sukenik, A., Kaplan-Levy, R. N., Welch, J. M. and Post, A. F. (2012). Massive multiplication of genome and ribosomes in dormant cells (akinetes) of aphanizomenon ovalisporum (cyanobacteria). *ISME Journal*, *6*, 670–679. http://dx.doi.org/10.1038/ismej.2011.128

Székely, A. J., Berga, M. and Langenheder, S. (2013). Mechanisms determining the fate of dispersed bacterial communities in new environments. *The ISME Journal*, *7*, 61–71. http://dx.doi.org/10.1038/ismej.2012.80

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analysis of the vegetation on danish commons. *Biologiske Skrifter Kongelige Danske Videnskabernes Selskab*, *5*, 1–34. http://dx.doi.org/10.1007/BF02852438

Tabak, J. (2004). *Differential Geometry*, (p. 150). Facts on File, Inc.

Tanentzap, A. J., Fitch, A., Orland, C., Emilson, E. J., Yakimovich, K. M., Osterholz, H. and Dittmar, T. (2019). Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 24689–24695. http://dx.doi.org/10.1073/pnas.1904896116

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Xu, Z. Z., Jiang, L., Haroon, M. F., Kanbar, J., Zhu, Q., Song, S. J., Kosciolek, T., Bokulich, N. A., Lefler, J., Brislawn, C. J., Humphrey, G., Owens, S. M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J. A., Clauset, A., Stevens, R. L., Shade, A., Pollard, K. S., Goodwin, K. D., Jansson, J. K., Gilbert, J. A., Knight, R. and Consortium, T. E. M. P. (2017). A communal catalogue reveals earth's multiscale microbial diversity. *Nature*, *551*, 457–463. http://dx.doi.org/10.1038/nature24621

Tranvik, L. J. (1992). Allochthonous dissolved organic matter as an energy source for pelagic bacteria and the concept of the microbial loop. *Hydrobiologia*, *229*, 107–114. http://dx.doi.org/10.1007/BF00006994

Tranvik, L. J. (1998). *Degradation of Dissolved Organic Matter in Humic Waters by Bacteria*, volume 133. Springer

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., Dillon, P., Finlay, K., Fortino, K., Knoll, L. B., Kortelainen, P. L., Kutser, T., Larsen, S., Laurion, I., Leech, D. M., McCallister, S. L., McKnight, D. M., Melack, J. M., Overholt, E., Porter, J. A., Prairie, Y., Renwick, W. H., Roland, F., Sherman, B. S., Schindler, D. W., Sobek, S., Tremblay, A., Vanni, M. J., Verschoor, A. M., von Wachenfeldt, E. and Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, *54*, 2298–2314. http://dx.doi.org/10.4319/lo.2009.54.6_part_2.2298

Tranvik, L. J. and Höfle, M. G. (1987). Bacterial growth in mixed cultures on dissolved organic carbon from humic and clear waters. *Applied and Environmental Microbiology*, *53*, 482–488.

Turner, K. W., Edwards, T. W. and Wolfe, B. B. (2014). Characterising runoff generation processes in a lake-rich thermokarst landscape (Old Crow Flats, Yukon, Canada) using $\delta\,^{18}$O, $\delta\,^{2}$H and d-excess measurements. *Permafrost and Periglacial Processes*, *25*, 53–59. http://dx.doi.org/10.1002/ppp.1802

Vannote, R. L., Minshall, W. G., Cummins, K. W., Sedell, J. R. and Cushing, C. E. (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*, 130–137.

Varela, M. M., Rodríguez-Ramos, T., Guerrero-Feijóo, E. and Nieto-Cid, M. (2020). Changes in activity and community composition shape bacterial responses to size-fractionated marine DOM. *Frontiers in Microbiology*, *11*. http://dx.doi.org/10.3389/fmicb.2020.586148

Vass, M. and Langenheder, S. (2017). The legacy of the past: Effects of historical processes on microbial metacommunities. *Aquatic Microbial Ecology*, *79*, 13–19. http://dx.doi.org/10.3354/ame01816

Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, *85*, 183–206.

Vellend, M. (2016). *Theory of ecological communities*. Princeton University Press.

Voss, B. M., Peucker-Ehrenbrink, B., Eglinton, T. I., Spencer, R. G., Bulygina, E., Galy, V., Lamborg, C. H., Ganguli, P. M., Montluçon, D. B., Marsh, S., Gillies, S. L., Fanslau, J., Epp, A. and Luymes, R. (2015). Seasonal hydrology drives rapid shifts in the flux and composition of dissolved and particulate organic carbon and major and trace ions in the Fraser River, Canada. *Biogeosciences*, *12*, 5597–5618. http://dx.doi.org/10.5194/bg-12-5597-2015

Větrovský, T. and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, *8*, 1–10. http://dx.doi.org/10.1371/journal.pone.0057923

Wang, Y., Ye, F., Wu, S., Wu, J., Yan, J., Xu, K. and Hong, Y. (2020). Biogeographic pattern of bacterioplanktonic community and potential function in the Yangtze river: Roles of abundant and rare taxa. *Science of the Total Environment*, *747*, 141335. http://dx.doi.org/10.1016/j.scitotenv.2020.141335

Wetzel, R. and Likens, G. (2000). *Limnological Analyses* (3rd edt. ed.). Springer Press.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*, 1–29. URL http://www.jstatsoft.org/v40/i01/

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. URL https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, 1686. http://dx.doi.org/10.21105/joss.01686

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W., Kettle, H., Flint, H. J., Haas, A. F., Laroche, B., Kreft, J.-U., Rainey, P. B., Freilich, S., Schuster, S., Milferstedt, K., van der Meer, J. R., Großkopf, T., Huisman, J., Free, A., Picioreanu, C., Quince, C., Klapper, I., Labarthe, S., Smets, B. F., Wang, H. and Soyer, O. S. (2016). Challenges in microbial ecology: building predictive understanding of

community function and dynamics. *The ISME Journal*, *10*, 2557–2568. http://dx.doi.org/10.1038/ismej.2016.45

Widder, S., Besemer, K., Singer, G. A., Ceola, S., Bertuzzo, E., Quince, C., Sloan, W. T., Rinaldo, A. and Battin, T. J. (2014). Fluvial network organization imprints on microbial co-occurrence networks. *Proceedings of the National Academy of Sciences*, *111*, 12799–12804. http://dx.doi.org/10.1073/pnas.1411723111

Wiegner, T., Kaplan, L., Ziegler, S. and Findlay, R. (2015). Consumption of terrestrial dissolved organic carbon by stream microorganisms. *Aquatic Microbial Ecology*, *75*, 225–237. http://dx.doi.org/10.3354/ame01761

Wiegner, T. N. and Seitzinger, S. P. (2001). Photochemical and microbial degradation of external dissolved organic matter inputs to rivers. *Aquatic Microbial Ecology*, *24*, 27–40.

Wilhelm, L., Besemer, K., Fasching, C., Urich, T., Singer, G. A., Quince, C. and Battin, T. J. (2014). Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environmental Microbiology*, *16*, 2514–2524. http://dx.doi.org/10.1111/1462-2920.12392

Wilke, C. O. (2019). cowplot: Streamlined plot theme and plot annotations for 'ggplot2'. URL https://cran.r-project.org/package=cowplot

Wilkinson, G. M., Pace, M. L. and Cole, J. J. (2013). Terrestrial dominance of organic matter in north temperate lakes. *Global Biogeochemical Cycles*, *27*, 43–51. http://dx.doi.org/10.1029/2012GB004453

Winter, C., Hein, T., Kavka, G., Mach, R. L. and Farnleitner, A. H. (2007). Longitudinal changes in the bacterial community composition of the Danube river: A whole-river approach. *Applied and Environmental Microbiology*, *73*, 421–431. http://dx.doi.org/10.1128/AEM.01849-06

Wisnoski, N. I. and Lennon, J. T. (2021). Microbial community assembly in a multi-layer dendritic metacommunity. *Oecologia*, *195*, 13–24. http://dx.doi.org/10.1007/s00442-020-04767-w

Wisnoski, N. I., Muscarella, M. E., Larsen, M. L., Peralta, A. L. and Lennon, J. T. (2020). Metabolic insight into bacterial community assembly across ecosystem boundaries. *Ecology*, *101*, e02968. http://dx.doi.org/10.1002/ecy.2968

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*, 3–36.

Wright, E. S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *The R Journal*, *8*, 352–359. http://dx.doi.org/V12242009

Xenopoulos, M. A., Barnes, R. T., Boodoo, K. S., Butman, D., Catalán, N., D'Amario, S. C., Fasching, C., Kothawala, D. N., Pisani, O., Solomon, C. T., Spencer, R. G., Williams, C. J. and Wilson, H. F. (2021). How humans alter dissolved organic matter composition in freshwater: relevance for the earth's biogeochemistry. *Biogeochemistry*, *154*, 323–348. http://dx.doi.org/10.1007/s10533-021-00753-3

Yannarell, A. C. and Triplett, E. W. (2004). Within- and between-lake variability in the composition of bacterioplankton communities: Investigations using multiple spatial scales. *Applied and Environmental Microbiology*, *70*, 214–223. http://dx.doi.org/10.1128/AEM.70.1.214-223.2004

Zark, M. and Dittmar, T. (2018). Universal molecular structures in natural dissolved organic matter. *Nature Communications*, *9*, 1–8. http://dx.doi.org/10.1038/s41467-018-05665-9

Zeglin, L. H. (2015). Stream microbial diversity in response to environmental changes: Review and synthesis of existing research. *Frontiers in Microbiology*, *6*, 454. http://dx.doi.org/10.3389/fmicb.2015.00454

Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Nostrand, J. D. V., Yang, Y. and He, Z. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *PNAS*, E835–E845. http://dx.doi.org/10.1073/pnas.1324044111

Zhou, J. and Ning, D. (2017). Stochastic community assembly: Does it matter in microbial ecology? *Microbial and Molecular Biology Reviews*, *81*, 1–32.

Zhou, L., Wu, Y., Zhou, Y., Zhang, Y., Xu, H., Jang, K. S., Dolfing, J., Spencer, R. G. and Jeppesen, E. (2024). Terrestrial dissolved organic matter inputs drive the temporal dynamics of riverine bacterial ecological networks and assembly processes. *Water Research*, *249*. http://dx.doi.org/10.1016/j.watres.2023.120955

Zhou, L., Zhou, Y., Tang, X., Zhang, Y., Zhu, G., Székely, A. J. and Jeppesen, E. (2021). Eutrophication alters bacterial co-occurrence networks and increases the importance of chromophoric dissolved organic matter composition. *Limnology and Oceanography*, *66*, 2319–2332. http://dx.doi.org/10.1002/lno.11756

Zigah, P. K., McNichol, A. P., Xu, L., Johnson, C., Santinelli, C., Karl, D. M. and Repeta, D. J. (2017). Allochthonous sources and dynamic cycling of ocean dissolved organic carbon revealed by carbon isotopes. *Geophysical Research Letters*, *44*, 2407–2415. http://dx.doi.org/10.1002/2016GL071348

Zimmerman, A. E., Martiny, A. C. and Allison, S. D. (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME Journal*, *7*, 1187–1199. http://dx.doi.org/10.1038/ismej.2012.176

Zinger, L., Gobet, A. and Pommier, T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, *21*, 1878–1896. http://dx.doi.org/10.1111/j.1365-294X.2011.05362.x