

Title (max 150 char): Reactive bacterioplankton reveal assembly dynamics along a boreal terrestrial-aquatic continuum

Masumi Stadler^{1,*}, Paul A. del Giorgio¹

Affiliation:

(1) Groupe de Recherche Interuniversitaire en Limnologie (GRIL), Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC, Canada

Institutional e-mail addresses: MS (stadler.masumi@courrier.uqam.ca), PdG (del_giorgio.paul@uqam.ca)

Running title (max 50 char): Bacterioplankton along a boreal terrestrial-hydrological continuum.

*Contact corresponding author: Masumi Stadler

E-mail: m.stadler.jp.at@gmail.com

Phone: +1 (514)297-5330

Address: Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, Montréal, QC, H3C 3P8, Canada

Keywords: aquatic bacterial communities, bacterioplankton, terrestrial-aquatic continuum, ecosystem connectivity, boreal ecosystems, mass effects, environmental sorting, metacommunity, meta-ecosystem, coalescence, microbial assembly, rare biosphere, rank abundance

Author contributions: PdG designed the sampling, MS collected the data, MS analysed the data, MS and PdG discussed the results and wrote the manuscript.

Abstract

Inland waters form complex hydrological networks acting as a bridge of abiotic and biotic matter between the terrestrial milieu and ultimately the oceans. During transport through the aquatic network, microbial communities are modified and re-assembled, forming complex patterns of ecological successions. These processes have seldom been examined along a true land-estuary aquatic continuum. Here we reconstruct the microbial succession from soils through the aquatic network until the estuary within the Romaine River watershed in Eastern Québec for three years. Two discontinuities along the river (riverine lakes and reservoirs) create a unique opportunity to address shifts in residence times within an assembly context. In order to distinguish the total from the reactive fraction of microbial communities we sequenced the 16S rRNA amplicon DNA and RNA. Differences among microbial assemblages was mainly driven by habitat type and seasons. Different patterns in incidence and abundance based DNA and RNA dissimilarity revealed dominating mass effects whenever two contrasting communities mix, such as terrestrial-aquatic as well as reservoir hypolimnetic-riverine transitions. In general, increasing residence time favoured stronger selection over mass effects. Our findings highlight the importance of considering spatial history and distinguishing the total and active microbial fractions, especially in highly connected and dynamic systems.

Introduction

Since the 1880s, we are starting to grasp the intriguing variety of microbes and how they are involved in many elemental cycling processes on a global scale¹. Microbial communities found outside of the petri-dish are inherently taxonomically diverse, with richness estimations that defy imagination^{2,3}. Richness varies among ecosystems with highest taxonomic reservoirs found in soils and less so in comparably nutrient depleted aquatic ecosystems². While descriptive approaches of richness and taxonomic composition is common in almost any environment imaginable, little do we actually understand how microbes assemble into these diverse communities^{4,5}.

Most field studies that explore community dynamics examine one ecosystem at a time. Commonly seasonal and/or spatial fluctuations of microbial community composition within individual terrestrial, freshwater and marine ecosystems are investigated separately (hereafter ecosystem domains)^{6,7,8,9}. Even among freshwater studies, researchers tend to focus on either streams, lakes, or rivers¹⁰. While this approach gives an in-detail insight into ecosystem specific community dynamics, it neglects an inherent characteristic of nature. A characteristic so apparent but easily neglected - the connectivity beyond ecosystem boundaries. Since meta-community theory has been coined, our perception has changed from ecosystems viewed as isolated patches to an integrated, connected and dynamic network¹¹, where different ecosystem communities meet and mingle (i.e. community coalescence, Mansour et al.¹²). Nevertheless, lit-

erature is still rather scarce when searching for studies that cover multiple ecosystem domains within a study^{13,14} and it is even rarer to investigate multiple ecosystem domains that are physically connected^{15,16,17,18}.

How does ecosystem connectivity affect the smallest living forms on Earth? Microorganisms without motility machineries may seem stationary, although being dispersal limited to a certain degree¹⁹, the minuscule size of microorganisms generally promotes dispersal. Within a watershed, freshwaters act as a carrier of matter, from the terrestrial milieu over freshwater networks to ultimately the ocean¹². Especially during heavy rain events, freshwater flushes through the earth, extracting all soluble nutrients and capturing matter - dead or alive - along its hydrological evolution from a raindrop to collectively becoming a stream. On its further journey, the rushing water may be stopped by lakes and reservoirs²⁰ that temporarily give more time for some organisms to thrive²¹ and previously unavailable resources may be degraded²². Within the aquatic network, dynamics in hydrology have been determined to be one of the major drivers of microbial community composition²³ and evidence suggests that a large proportion of aquatic microbial communities can be traced back to soils^{15,16,24,25,26}. Thus, not every member within diverse communities at a single location shares a similar history in terms of origin^{23,27} or reacts similarly to changing environmental conditions²⁸. However, such passive dispersal implies that occurrence of a taxon in an unsuitable habitat merely by accident is also likely.

Death and dormancy^{29,30} questions the suitability of DNA approaches to study ecological 'communities'. In a strict ecological sense, a commu-

nity refers to an assemblage of various species that interact with each other, sharing niches in forms of available space or resources³¹. However, DNA methods are not able to distinguish the actively participating from the passive members. Since, biologists have expanded their molecular toolbox to RNA approaches to capture taxa that have invested in protein synthesis in the recent minutes to hours. Evidence of 16S rRNA has opened windows to study differential responses of the potentially active community to environmental variables^{32,26} and available resources³³. Being target of scrutiny and criticism, RNA may not be a good indicator of absolute growth or activity rates³⁴, however, changes in RNA levels of the same taxa along gradients does imply a certain reaction of a taxon to environmental or biological changes. Hence, to this date, the full potential of RNA approaches in the context of microbial community assembly has yet to be explored.

Inherited from community ecology in microbiology, microbial assembly processes have similarly been defined into four fundamental processes - selection, dispersal, diversification and drift -³⁵, which can be characterized by their degree of determinism and stochasticity³⁶. Recently, a statistical approach has gained popularity, which aims to quantify each assembly process^{37,38} and a plethora of studies have found interesting patterns since^{18,39,40}. While reported results have been insightful, the approach heavily relies on the assumption that taxonomic related organisms have similar functional traits (i.e. phylogenetic conservatism). This concept has been handled with care within the macro-biologies^{41,42}, but should be a target of further discussions in microorganisms, where species definitions remain unclear⁴³ and genetic exchanges are common⁴⁴. Indeed, many mi-

microbial traits have been shown to be phylogenetically conserved⁴⁵. It was, however, also noted that more complex traits encoded by many genes are more likely to be conserved while simpler traits that involve fewer genes are rather phylogenetically dispersed⁴⁶. As such, minuscule variation in genes⁴⁷ or expression¹ may have disproportionate effects on microbial phenotypes and thus, patterns based on phylogenetic conservatism shall be interpreted with care^{43,48,49}.

In this study, we followed the Romaine river in North-Eastern Québec, Canada, as an example watershed over several years and seasons. Our overall objective was to understand microbial succession and assembly processes along a terrestrial-hydrological continuum without any phylogenetic inference but rather simply by following and comparing the total DNA to the reactive RNA assemblages. Therefore, soils, soilwater, groundwater, streams, the river, lakes, reservoirs and the estuary were sampled for the 16S rRNA gene (DNA) and the 16S rRNA (RNA) to characterize the taxonomic composition of each total and active fraction, respectively. We specifically wanted to explore firstly, whether we have distinct microbial communities depending on habitat types and if we can find seasonality across all ecosystems. We hypothesized that firstly, seasonality will be reflected in all communities of the different ecosystems as boreal climate zones do exhibit strong seasonality that is accompanied by shifts in available light, temperature and thus affect vegetation and aquatic and terrestrial bodies substantially. Secondly, we expected a gradual change in community composition along the continuum rather than distinct separate clustering of terrestrial and aquatic communities as we expect headwa-

ter streams to represent sites of strong community coalescence between terrestrial and freshwater ecosystems¹². Thirdly, we hypothesized that dynamics in DNA and RNA assemblage divergence and convergence enables us to grasp what dominating assembly process (i.e. mass effect and selection) is governing the community at a certain point in the continuum. And lastly, we wanted to explore where in a typical rank abundance curve (e.g. rare vs abundant taxa) community reshuffling is happening as the communities change habitats along the continuum.

Material and methods

Catchment characteristics and sampling

To follow the movement of microbial communities within a watershed, samples were taken along the Romaine river (Côte-Nord region, Québec, Canada) (Fig. 1a-b) for three years from 2015-2017. The Romaine catchment belongs to the eastern black spruce-moss bioclimatic domain and drains an area (A) of approximately 14 500 km². For detailed catchment characteristics please refer to the supplementary methods (hereafter, SM) (SM1). In brief, the river flows through a series of large, shallow headwater lakes (hereafter, riverine lakes), and subsequently flows mainly south passing through three dams that were consecutively build in 2015 (RO2), 2016 (RO1), and 2017 (RO3). We partition the river in the fractions before and after the reservoir complex as upriver and downriver, respectively. The river has a maximum distance from the northern headwaters to the river mouth expanding to approximately 475.1 km.

Overall, 395 samples were collected for DNA (D) and 202 for RNA (R), covering spring (166-D, 69-R), summer (195-D, 99-R) and autumn (34-D,34-R). RNA samples were sampled from 2016 onwards. Following a terrestrial-aquatic continuum, various habitat types were sampled (Table S1). Due to the remoteness and inaccessibility of the northernmost headwaters, we sampled the Petite Romaine sub-catchment (PR, A: 310.73 km², elevation: 580 masl, Fig. 1c) for streams and headwater ponds. This sub-catchment represents a headwater stream network in our studied continuum. Samples were taken from soil, soilwater, groundwater, sediments, surface waters of streams, ponds, rivers, tributaries, reservoirs, lakes, and the estuary.

Samples were collected throughout the catchment (Fig. S1). 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 ± 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 were pooled for each site. Groundwater was directly collected from constructed wells with sub-

mersible 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 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, was filtered through 0.22 µm polycarbonate membrane filters (Merck Millipore, Darmstadt, Germany). Homogenized soil and sediment samples were transferred to aliquots of 0.25 g. 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.

Following the manufacturer's instructions, commercial DNA and RNA extraction kits were used (QIAGEN®, Hilden, Germany, details in SM2). 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 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 an Illumina MiSeq (PE250) platform.

Bioinformatic analysis

A detailed description of the bioinformatic treatment can be found in SM3. In brief, primers were removed from 16S rRNA DNA and cDNA (hereafter RNA) reads using *cutadapt* (Version 1.18, Martin⁵⁰). To identify amplicon sequence variants (ASVs), 16S rRNA amplicon reads were analysed

through the DADA2 (Divisive Amplicon Denoising Algorithm 2) pipeline (Version 1.14.1, Callahan et al.⁵¹). Identified ASVs that are identical in sequence but differ only by length were merged together, leading to units representing 100% similarity operational taxonomic units (OTUs) and chimeras were removed. Taxonomy was assigned with the *DECIPHER* package (Version 2.14.0, Wright⁵²) implementing the increased accuracy IDTAXA algorithm⁵³ and the provided trained classifier of the SILVA database (Version 138, Pruesse et al.⁵⁴). Several ASVs were found to be highly abundant only in RNA, and thus 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 threshold⁵⁵ with the *DECIPHER* package⁵².

All OTU observations with less than 10 reads per sample were removed. Furthermore, *metagenomeSeq* was used to transform and stabilize variation in library sizes with cumulative sum scaling (CSS)⁵⁶. CSS results were rounded to its integer to represent count data (hereafter: CSS reads). CSS results were compared with results achieved with various rarefaction thresholds (see Fig. S2).

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})^{57,58} based on all DNA samples with the function *pcoa* in the *ape* package⁵⁹ ($n = 372$, 20182 OTUs). The community matrix was Hellinger transformed to resolve a horse-shoe

effect⁶⁰. To correct any negative eigenvalues problematic for PERMANOVA analysis, the D_{BC} matrix was square-root transformed to Euclidean distance^{58,61}. To evaluate statistical differences in habitat type and season a PERMANOVA was computed with 9999 permutations with the *adonis* function. A PERMANOVA cannot distinguish among-group from within-group variation if data dispersion is variable among groups⁶², therefore, an analysis of multivariate homogeneity was computed with *betadisper*. Using *permute*, we finally tested whether dispersion is variable among groups. For all statistical analyses, an α level of 0.05 was chosen prior to analysis and all functions are part of the *vegan* package⁶³.

Secondly, to evaluate whether sampled RNA communities were further different from the DNA assemblages, we performed a second PCoA (D_{BC} with square-root transformation) with both DNA and RNA samples ($n = 572$, 20185 OTUs). Again, statistically different clusters were investigated with a PERMANOVA (9999 permutations), where habitat type, season and nucleic acid type (DNA vs. RNA) formed the clusters. The same framework explained above to check for dispersions was applied.

To examine how different DNA and RNA assemblages of the same sample are, the distance of each DNA-RNA sample pair within the PCoA ordination space was computed across n -dimensional space⁶⁴:

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

where p and q represent DNA and RNA site scores, respectively, of each sample and n is the used maximum number of dimensions. We fo-

cused on the first axes that cumulatively explain 75 % of the variation for each ordination ($n_{75\%}$) similar to Osterholz et al.³³. This approach was implemented, as it was evident from the PCoA that essential variation within non-aquatic samples were captured outside the first three axes. With this approach, we were able to narrow down the dimensions from 571 to 186 for the Bray-Curtis PCoA. The calculated distance extracts a proportion of the pair-wise dissimilarities on which the PCoA is based on (Fig. S3).

To further gain 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)^{58,65}(Fig. S4). By comparing incidence and abundance based dissimilarities, we can further distinguish in which samples' DNA-RNA 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 $n_{75\%}$ (236 of 571) axes (m_S). Finally, we calculated the residuals from each sample to the 1:1 line when the Sørensen and Bray-Curtis distance are plotted against each other (Fig.5a). The framework allows us to understand what shifts within the community is causing DNA and RNA assemblages to diverge. When the residuals are positive (>0), m_S is bigger than m_{BC} and thus there are fewer shared OTUs among DNA and RNA, richness differences are large and abundance differences are comparably smaller. In contrary, when the residuals are negative (<0) there is a higher number of shared OTUs but rather abundance differences are larger. These interpretations were statistically tested with Kruskal-Wallis tests (Fig. S5).

Abundance groups

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. S6). This approach classified all OTUs with ≥ 47 CSS reads as abundant, < 47 and ≥ 5 CSS reads as moderate, and < 5 CSS reads as rare (details in SM4).

As a second step, we defined OTUs into spatial abundance groups (spAGs) based on their mean DNA abundance in each habitat (Table 1). The approach is similar to the commonly used temporal abundance based groupings such as conditionally rare taxa⁶⁶, but rather applied in a spatial context with more explicit groupings to cover the whole community. In brief, we distinguish OTUs that are present everywhere (universal, cosmopolitan) from those that can be absent in certain habitat types. Additionally, there are categories for OTUs with differing abundance (abundant, moderate, rare) and whether they change AGs between habitats (shifters).

To evaluate how spAGs are distributed among the habitat types, we computed weighted averages scores of each OTU for the DNA-RNA PCoA with the *wascores* function in *vegan*⁶³. These species scores represent optima of each OTU within the PCoA ordination.

The packages *phyloseq*, *tidyverse*, *plyr* and *data.table* were used for data wrangling and transformation^{67,68,69,70}, and *doMC* and *parallel* enabled parallel processing^{71,72}. *ggplot2*, *ggsignif*, *ggnewscale* and *cowplot*

were used to visualize the results^{73,74,75,76}. For statistical analyses, *vegan* and *rstatix* were used^{63,77}. All analyses have been conducted in R v3.4.2⁷² and RStudio v1.3.1073⁷⁸. Maps were created with QGIS (version 3.12) and a digital elevation model provided by Natural Resources Canada. Watersheds were delineated with ArcMap (version 10.5.1, ESRI Inc., Redland, CA) and the Spatial Analyst Toolbox.

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 87 002 470 quality reads, with 156 347 identified ASVs. There were 56 107 unclassified ASVs that were removed in the downstream analyses. After sub-sampling only bacteria and 99 % similarity clustering, 47 402 282 reads and 20 220 OTUs were retained. The smallest and largest library size were both found in riverine lakes with 2851 and 1 333 391 reads, respectively. On average, the lowest library sizes were found in sediments, soil and soilwater with less than 30 000 reads. In contrast, most freshwater samples had a library size higher than 50 000 reads (Fig. S7). 33 phyla, 80 classes, 189 orders, 400 families, and 696 genera were represented in the dataset. Relative abundances of phyla varied across habitat types (Fig. S7) but on average, the meta-community across all ecosystems was composed of 52.11 % Proteobacteria, 17.17 % Actinobacteria, 8.06 % Verrucomicrobiota, 6.88% Bacteroidota, 2.91 % Acidobacteriota, 2.63 % Planctomycetota, 2.37 % Cyanobacteria, 1.43 % Myxococcota, 1.23 %

Desulfobacterota, 1.22 % Chloroflexi and 1.19 % Nitrospirota.

Gradual change of DNA assemblages along a terrestrial-aquatic continuum

Within system diversity (α) decreases along the continuum (Fig. S8) and β diversity conversely increases. We observed a gradual separation of terrestrially-influenced habitats such as soil, soilwater and sediment from freshwater and estuary samples along the first PCoA axis capturing 15.39 % of the variance (Fig. 2). This observation was supported by the PERMANOVA analysis ($F_{12} = 18.01$, $R^2 = 0.36$, $p = 0.0001$). Streams, groundwater, tributaries, headwater ponds, and lakes were creating a gradient between the two very distinct clusters of terrestrial and riverine/reservoir samples (Fig.2a).

The trajectory along the continuum has a striking seasonality where spring and summer/autumn samples cluster most pronounced in the reservoir samples along the second PCoA axis capturing 3.71 % of the variance (PERMANOVA: $F_2 = 11.01$, $R^2 = 0.037$, $p = 0.0001$). Soil, sediment, soilwater and groundwater samples, however, do not exhibit a clear seasonality (Fig.2b). Seasonality does not emerge as a strong driver even within a PCoA performed only with terrestrial samples (Fig. S9). Although PERMANOVA results strongly supported habitat type and seasonal clustering, the results could be affected by different dispersion of data. Differences in dispersion were found by habitat type alone (PERMDISP: $F_{12} = 7.76$, $p < 0.0001$), by solely season (PERMDISP: $F_2 = 40.62$, $p < 0.0001$) and by both habitat and season combined (PERMDISP: $F_{27} = 18.62$, $p < 0.0001$).

While dispersion between spring and summer was not statistically different (Tukey HSD: $p > 0.05$), they were always different when comparing autumn with other seasons (Tukey HSD: $p < 0.0001$). The average distance of samples within the autumn cluster to its median was smaller compared to other seasons (0.47 vs 0.62). Among habitat types, dispersion was larger in terrestrially influenced samples such as soil (Distance to median: 0.61), soilwater (0.63), streams (0.62) and tributaries (0.58) samples, compared to riverine (0.49), reservoir (0.51) and estuary (0.52) samples.

Clear divergence of aquatic RNA from DNA

When DNA and RNA samples were combined, the second PCoA analysis captured in summary 19.83 % of the dissimilarity variance in the first three axes, delineating habitat type (PC1), nucleic acid type (PC2) and seasons (PC3)(Fig. 3). Seasonality emerges the second strongest driver even in a PCoA only with RNA samples (Fig. S10). PERMANOVA analysis indicated significant clustering by habitat type ($F_{12} = 20.64$, $R^2 = 0.29$, $p = 0.0001$), season ($F_2 = 14.85$, $R^2 = 0.35$, $p = 0.0001$) and nucleic acid type ($F_1 = 25.99$, $R^2 = 0.03$, $p = 0.0001$). Similarly to the DNA PCoA, homogeneity of dispersion was mostly not fulfilled. According to PERMDISP, dispersion differed by all factorial combinations ($F_{48} = 10.92$, $p < 0.0001$), habitat type ($F_{12} = 31.61$, $p = 0.0001$) and season ($F_2 = 44.55$, $p = 0.0001$), however, not for nucleic acid type ($F_1 = 1.67$, $p = 0.19$). Dispersion patterns among seasons and habitats remain similar to the DNA PERMDISP results.

DNA-RNA pair-wise dissimilarity along the continuum

To further explore the patterns in DNA-RNA 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 $n_{75\%}$ axes as the depicted axes in Fig. 3 only capture aquatic DNA-RNA divergence. The distance metrics indicate that the largest dissimilarities among DNA and RNA are in spring, while summer and autumn DNA-RNA samples are in general more similar (Fig. 4a).

In order to examine, when dissimilarities are driven more strongly by a shift in shared taxa identities or abundance discrepancies, we calculated the residuals to the 1:1 line between m_{BC} and m_S (Fig. 4b). When examining the residuals along the continuum we can observe a gradual decline from soilwaters to the upriver. This decline is rather smooth in spring, and more abrupt in summer. Spring samples are generally positive, while summer and autumn samples tend to be negative after the continuum enters the river until the estuary. Only the river portion after the reservoir (down-river), turns positive in summer. Habitats not directly within the continuum such as sediments, tributaries, and lakes are generally positive except for summer samples of the riverine lakes and an autumn lake sample, which are negative.

Community shifts along the rank abundance curve

In general there were no OTUs that were always present and abundant, moderate or rare across all habitats, and the vasy majority of OTUs was

shifting between abundance groups (Tab.1). In order to understand where within a rank abundance curve the observed DNA-RNA abundance differences were happening, we calculated the mean DNA and RNA abundance difference of each spAG and examined their relationship to m_{BC} (Fig. 5a). A few abundance groups do contribute stronger than others with lower shifters being the most dynamic and large contributor to the observed DNA and RNA differences. Lower m_{BC} (e.g. < 0.6) seem to have minor DNA:RNA differences in rare and specialist OTUs, while moderate, lower shifters, cosmopolitans, and general shifters seem to contribute more substantially. In the extremest cases of $m_{BC} > 0.8$, the prior mentioned slight contributors, which are rare, specialist and upper shifters, increase their discrepancies as well. Overall, it seems that moderate and lower shifters have the highest variations in terms of their DNA and RNA abundances due to their small sample sizes. Additionally, shifters and cosmopolitans steadily contribute to drifting DNA and RNA abundances.

Furthermore, to understand the distribution of the spAGs among the sampled habitats, the weighted average species scores (i.e. optima) of each OTU within the first three PCoA axes was explored (Fig. 5b). It becomes evident that overall, many rare and specialist taxa have their optima in terrestrial samples, while depending on the OTU, moderate, cosmopolitans and shifters can have optima in either aquatic or terrestrial habitats. While lower shifters do not necessarily have a strong pattern in ecosystem preference, they do have a tendency towards an optimum in RNA, compared to all other spatial abundance groups that lean towards DNA. In terms of seasons, only cosmopolitans have optima in spring, while most

other spAGs lean towards summer and autumn.

Discussion

As we followed the changes of microbial communities along a terrestrial-aquatic continuum, the results indicate a gradual shift in total as well as the reactive microbial community assemblages with pronounced seasonality solely in aquatic systems. Differences in abundance and incidence based dissimilarities between DNA and RNA of the same sample indicate a shift in assembly forces dominated by mass effects from soilwaters to streams, while selective activation of a smaller pool of taxa seem to unravel in the reservoirs and the estuary. Riverine and reservoirs exhibit a stronger seasonal pattern regarding their dominating assembly process. Overall, abundance differences between DNA and RNA seem to emerge across the entire rank abundance curve with disproportionately high dynamics in microbes that shift among abundance groups across habitat types (e.g. cosmopolitans, shifters).

Spatio-temporal differences in microbial assemblages

When solely focusing on the total microbial assemblage, the strongest compositional gradient followed the terrestrial-aquatic continuum. Streams, headwater ponds, tributaries and lakes seem to represent intermediate states of community composition between soils and larger aquatic water bodies (river, reservoirs). This finding supports the emerging evidence of soil microorganisms in aquatic systems^{15,16}, especially in systems with

lower residence times and higher connectivity to the surrounding terrestrial systems^{24,25}. However, community composition does not seem to gradually diverge with shifts in residence time. Lakes with comparably higher residence times than streams spread similarly between terrestrial ecosystems and reservoirs. The sampled lakes ranged from mountainous to lowland lakes, and thus differences can arise through differences in network position⁷⁹, residence times²¹ or potentially lake area:volume ratios, which imply their degree of available processing time and connectivity to the surrounding land, respectively. Evidence that the large riverine lakes located closest to the headwaters cluster with the river and reservoir assemblages hint that network position may play a rather minor role in aquatic community composition.

The divergence in habitats are accompanied by a strong seasonal separation in DNA composition between spring and summer/autumn as early as streams within the aquatic network. Such seasonality seems of minor importance in the sampled soil, sediment and soilwater although other studies have found distinct seasonal patterns in soils^{6,80}. Seasonality may take shape as different enzyme expression levels⁸¹, which cannot be excluded. Although seasonality within lakes has been widely reported^{82,83,84}, the above discussed variety of sampled lakes may affect responses to seasonal fluctuations (e.g. elevation levels) and thus confounds a clear conclusion on seasonal effects on boreal lentic systems in this study.

The river itself not only exhibited a clear seasonal divergence, but showed signs of inter-annual differences only in spring. It seems that in

2015, the river portions after and before the reservoir are more dissimilar compared to 2016, however, the downriver samples furthest from the reservoir converge back to an upriver community. In comparison, 2016 upriver and downriver sites are in general more similar, and the downstream sites remain clustered separately. It remains unclear what the cause of the inter-annual difference is. Previously reported relevance of hydrology²³ had a minor impact, as precipitation and discharge were only marginally different among the two years (data not shown). Lasting effects of reservoir discharges feeding downstream rivers have previously been observed^{85,86,87}, and thus the addition of another reservoir in 2016 may have strengthened the reservoir impact and also shortened the distance from the last reservoir to the estuary. Hence, the shortening of time the community has downstream to converge back to an upriver assemblage could explain our observations.

At the end of the continuum, the estuary samples initially cluster with the river and reservoirs, but converge back towards the centre and further to the stream assemblages. This indicates that estuaries can be impacted by community dynamics unravelling in the freshwater network, with lasting effects¹⁶ that slowly disappear over a 25 km stretch as salinity increases^{88,89}. In contrast to Doherty et al.¹⁷, seasonality was reflected in the estuary close to the river delta as well, however, as the reservoir footprint fades, so does the seasonality. Hence, seasonality may be reflected in the freshwater microbes but not in the marine specific members.

Not only did community composition differ but habitat variance was substantially dissimilar among habitats as well, with high dispersion in soil,

soilwater, streams and comparably low dispersions in rivers and reservoirs. It is beyond the scope of this study to evaluate whether high dispersion within a habitat indeed corresponds to varying local environmental conditions. However, in order to further disentangle assembly dynamics that may explain some differences in habitat variance, community assembly was explored by comparing DNA-RNA dynamics to focus on differences in the patterns of inactive and active bacteria.

Community assembly shifts along the continuum

By examining different levels of dissimilarity among incidence and abundance based metrics, we observe habitat as well as seasonal assembly shifts along the continuum. We interpret differences in abundance and incidence based dissimilarity as a result of different governing assembly processes at play. Such that higher m_S is indicative of fewer shared OTUs, smaller abundance differences, and higher richness differences among DNA-RNA. Assembly processes that can contribute to dominantly incidence driven DNA-RNA differences are influxes of inactive bacteria (i.e mass effects)¹¹. On the other hand, higher m_{BC} suggests higher number of shared OTUs and rather large abundance differences between DNA-RNA. More shared OTUs between DNA-RNA can arise through selective pressure through the environment as well as biological interactions. Larger abundance differences may result from high activity of a few selected OTUs³².

Along the continuum, we observed a consistent relatively high dissimilarity among DNA-RNA in soilwater and streams in both m_S and m_{BC} ,

and the residuals indicate that these habitats are largely positive and thus dominated by mass effects. These results highlight the high connectivity and influxes of passive taxa into these ecosystems^{15,16,24}. Along the subsequent continuum, almost all habitats are mainly positive during the spring freshet, and thus indicate the tight linkage of assembly dynamics to hydrological differences in these ecosystems^{23,90}. In support of this observation, we also found a decoupling between DNA and RNA α -diversity in spring, indicating that there is a large influx of inactive bacteria (Fig. S11).

Assembly flips towards selection in the upriver in summer, and remains selection dominated in the reservoirs. Remarkably, assembly shifts again and mass effects dominate in the downriver portion. This assembly shift is dependent on the season, as autumn downriver samples remain selection dominated. This difference among seasons may be explained by the weaker and absence of stratification in the reservoirs RO2 and RO1 in autumn, respectively (data not shown). Thus, released hypolimnion specific members^{85,86} may have largely become inactive along the downriver, which resulted in a prevailing mass effect in summer. In autumn, the lack of stratification may not have allowed establishment of a hypolimnion specific community⁹¹, and thus river specific selection (i.e. water residence time) continued in the downriver⁹⁰. The estuary indicates a tendency towards selection likely to be driven by salinity gradients^{88,89}, however, seasonal patterns cannot be evaluated due to the lack of RNA samples. As we observed freshwater footprints in DNA fading along the estuary, RNA indicates that there is a selective force that favours activity differences of different OTUs³².

Who drives DNA-RNA differences?

In order to further address the question of where within rank abundance curves DNA-RNA discrepancies are happening, we examined different spatial abundance groups and their DNA-RNA abundance differences. It seems that OTUs that shift between abundant, moderate and rare abundances among habitats, are consistently those that have higher DNA-RNA discrepancies (e.g. cosmopolitan, shifters). When comparing shifters that only switch between abundant and moderate (upper shifter) and those that alternate between moderate and rare (lower shifter), it seems that the lower shifters have larger RNA fluctuations compared to their DNA. Furthermore, from the analysis of the OTUs' optima within the DNA-RNA PCoA ordination it emerged that lower shifters have in general a tendency towards high abundances in RNA. This observation is in line with the recent evidence that a few rare bacteria can have disproportionately high activity levels^{32,92}.

Overall, taxa across the rank abundance curve contribute to the observed abundance differences among DNA-RNA. It is noteworthy that along the sampled terrestrial-aquatic continuum there was not a single OTU that was abundant nor moderate everywhere. This observation indicates that there is not a single generalist that can occupy all ecosystem domains⁹³. Given the evidence that there are strong mass effects in almost all lower residence time aquatic systems, dispersal limitation may play a minor role, but selection on microorganisms is indeed rather strong⁹⁴. The previously observed numerical dominance of terrestrial taxa within freshwa-

ters¹⁵, thus is rather likely to arise from rare terrestrial bacteria that thrive to higher abundances in aquatic habitats as the concept of microbial 'seed banks' proposes⁹⁵.

In summary, our study explored microbial compositional differences and DNA-RNA discrepancies along a boreal terrestrial-hydrological continuum. The results indicate that indeed, differences in DNA and RNA can inform community dynamics and that the dominating assembly process shifts from mass effects to selection as the influence of hydrology and thus terrestrial imprint on aquatic ecosystems fades. However, mass effects emerge in the downriver in summer as OTUs not adapted to the local environment are flushed into the river from the reservoir. Similarly, the seasonal patterns in upriver assembly are very similar to the riverine lakes, which feed into the river. These findings highlight the importance of considering network history when studying rivers that flow through higher residence time water bodies as they tend to have stronger selective pressures with lasting downstream effects, especially in summer²⁰. Although not quantitative, the relative change of DNA-RNA assemblages along the continuum enabled us to study community scale assembly dynamics without phylogenetic inference. We also gave first insights into the complex in- and reactivation processes that individual OTUs experience along a terrestrial-aquatic continuum. Details of the contribution of different phylogenetic groups as well as population-scale patterns are future avenues to be unravelled. Furthermore, it remains unclear how the strong differences

in assembly processes may affect local microbial processes and ecosystem functioning, which is another facet to be explored.

Data accessibility

The raw 16S rRNA gene sequences, both DNA and cDNA are available at the public NCBI Sequence Read Archive (SRA) under the accession numbers SRxxxxx. The code and meta data that were used to produce this manuscript are available at <https://github.com/CarBBAS/xxx> (Zenodo DOI). (Links, accession numbers and DOI will be updated upon acceptance)

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

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-González, 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, Mario Muscarella for insightful discussions and Yves Prairie for statistical advice. We also thank xx anonymous reviewers for constructive comments that 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.

References

- [1] Pierre Caumette, Jean-Claude Bertrand, and Philippe Normand. *Some Historical Elements of Microbial Ecology*. Springer Science and Business Media, Dordrecht, 2015. ISBN 978-94-017-9118-2. doi: 10.1007/978-94-017-9118-2_2.
- [2] Luke R. Thompson, Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J. Locey, Robert J. Prill, Anupriya Tripathi, Sean M. Gibbons, Gail Ackermann, Jose A. Navas-Molina, Stefan Janssen, Evguenia Kopylova, Yoshiaki Vázquez-Baeza, Antonio González, James T. Morton, Siavash Mirarab, Zhenjiang Zech Xu, Lingjing Jiang, Mohamed F. Haroon, Jad Kanbar, Qiyun Zhu, Se Jin Song, Tomasz Kosciolek, Nicholas A. Bokulich, Joshua Lefler, Colin J. Brislawn, Gregory Humphrey, Sarah M. Owens, Jarrad Hampton-Marcell, Donna Berg-Lyons, Valerie McKenzie, Noah Fierer, Jed A. Fuhrman, Aaron Clauzet, Rick L. Stevens, Ashley Shade, Katherine S. Pollard, Kelly D. Goodwin, Janet K. Jansson, Jack A. Gilbert, Rob Knight, and The Earth Microbiome Project Consortium. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681):457–463, 2017. ISSN 14764687. doi: 10.1038/nature24621.
- [3] Stilianos Louca, Florent Mazel, Michael Doebeli, and Laura Wegener Parfrey. A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS biology*, 17(2):e3000106, 2019. ISSN 15457885. doi: 10.1371/journal.pbio.3000106.
- [4] Ashley Shade. Diversity is the question, not the answer. *ISME Journal*, 11:1–6, 2017. ISSN 17517370. doi: 10.1038/ismej.2016.118. URL <http://dx.doi.org/10.1038/ismej.2016.118>.
- [5] Ashley Shade. Understanding Microbiome Stability in a Changing World. *mSystems*, 3(e00157-17):1–4, 2018. ISSN 2379-5077. doi: 10.1128/mSystems.00157-17. URL <http://msystems.asm.org/lookup/doi/10.1128/mSystems.00157-17>.
- [6] Nobuhiko Shigyo, Kiyoshi Umeki, and Toshihide Hirao. Seasonal dynamics of soil fungal and bacterial communities in cool-temperate montane forests. *Frontiers in Microbiology*, 10(AUG):1–14, 2019. ISSN 1664302X. doi: 10.3389/fmicb.2019.01944.

- [7] Stuart E Jones, Tracey A Cadkin, Ryan J Newton, and Katherine D McMahon. Spatial and temporal scales of aquatic bacterial beta diversity. *Frontiers in Microbiology*, 3(318):1–10, 2012. doi: 10.3389/fmicb.2012.00318.
- [8] Norman Hassell, Kara A. Tinker, Thomas Moore, and Elizabeth A. Ottesen. Temporal and spatial dynamics in microbial community composition within a temperate stream network. *Environmental Microbiology*, 20(10):3560–3572, 2018. ISSN 14622920. doi: 10.1111/1462-2920.14311.
- [9] Stephen J. Giovannoni and Kevin L. Vergin. Seasonality in ocean microbial communities. *Science*, 335(6069):671–676, 2012. ISSN 10959203. doi: 10.1126/science.1198078.
- [10] Jürg Brendan Logue and Eva S Lindström. Biogeography of bacterioplankton in inland waters. *Freshwater Reviews*, 1:99–114, 2008. doi: 10.1608/FRJ-1.1.9.
- [11] M. A. Leibold, M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7):601–613, 2004. ISSN 1461023X. doi: 10.1111/j.1461-0248.2004.00608.x.
- [12] India Mansour, Catherine M. Heppell, Masahiro Ryo, and Matthias C. Rillig. Application of the microbial community coalescence concept to riverine networks. *Biological Reviews*, 93:1832–1845, 2018. ISSN 1469185X. doi: 10.1111/brv.12422.
- [13] Diana R. Nemergut, Elizabeth K. Costello, Micah Hamady, Catherine Lozupone, Lin Jiang, Steven K. Schmidt, Noah Fierer, Alan R. Townsend, Cory C. Cleveland, Lee Stanish, and Rob Knight. Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology*, 13(1):135–144, 2011. ISSN 14622912. doi: 10.1111/j.1462-2920.2010.02315.x.
- [14] Ashley Shade, J. Gregory Caporaso, Jo Handelsman, Rob Knight, and Noah Fierer. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME Journal*, 7(8):1493–1506, 2013. ISSN 17517362. doi: 10.1038/ismej.2013.54. URL <http://dx.doi.org/10.1038/ismej.2013.54>.

- [15] Clara Ruiz-González, Juan Pablo Niño-García, and Paul A del Giorgio. Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters*, 18:1198–1206, 2015. doi: 10.1111/ele.12499.
- [16] Aviaja L. Hauptmann, Thor N. Markussen, Marek Stibal, Nikoline S. Olsen, Bo Elberling, Jacob Bælum, Thomas Sicheritz-Pontén, and Carsten S. Jacobsen. 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(1474): 1–16, 2016. ISSN 1664302X. doi: 10.3389/fmicb.2016.01474.
- [17] Mary Doherty, Patricia L Yager, Mary Ann Moran, Victoria J Coles, Caroline S Fortunato, Alex V Krusche, Patricia M Medeiros, Jérôme P Payet, Jeffrey E Richey, Brandon M Satinsky, Henrique O Sawakuchi, Nicholas D Ward, Byron C Crump, and Justin Robert Seymour. Bacterial Biogeography across the Amazon River-Ocean Continuum. *Frontiers in Microbiology*, 2017. doi: 10.3389/fmicb.2017.00882.
- [18] Hyun S. Gweon, Michael J. Bowes, Heather L. Moorhouse, Anna E. Oliver, Mark J. Bailey, Michael C. Acreman, and Daniel S. Read. Contrasting community assembly processes structure lotic bacteria meta-communities along the river continuum. *Environmental Microbiology*, 2020. ISSN 1462-2912. doi: 10.1111/1462-2920.15337.
- [19] China A Hanson, Jed A Fuhrman, M Claire Horner-Devine, and Jennifer B H Martiny. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10: 497–506, 2012. ISSN 1740-1526. doi: 10.1038/nrmicro2795. URL <http://dx.doi.org/10.1038/nrmicro2795>.
- [20] James V Ward and Jack A Stanford. The Serial Discontinuity Concept of Lotic Ecosystems. In T. D. Fontaine III and S. M. Bartell, editors, *Dynamics of Lotic Ecosystems*, pages 29–42. Ann Arbor Science, Ann Arbor, 1983. ISBN 9780250406128.
- [21] Jürg B Logue and Eva S Lindström. Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *The ISME Journal*, 4:729–738, 2010. doi: 10.1038/ismej.2009.156.
- [22] Núria Catalán, Rafael Marcé, Dolly N. Kothawala, and Lars J. Tranvik. Organic carbon decomposition rates controlled by water re-

- tention time across inland waters. *Nature Geoscience*, 9(May):1–7, 2016. ISSN 1752-0894. doi: 10.1038/ngeo2720. URL <http://dx.doi.org/10.1038/ngeo2720>.
- [23] Juan Pablo Niño-García, Clara Ruiz-González, and Paul A del Giorgio. Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal*, 10:1755–1766, 2016. doi: 10.1038/ismej.2015.226.
 - [24] Byron C Crump, Linda A Amaral-Zettler, and George W Kling. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *The ISME Journal*, 6(9):1629–1639, 2012. ISSN 1751-7370. doi: 10.1038/ismej.2012.9. URL <http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=3498914&tool=pmcentrez&rendertype=abstract>.
 - [25] Katharina Besemer, Gabriel Singer, Christopher Quince, Enrico Bertuzzo, William Sloan, and Tom J Battin. Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proceedings of the Royal Society B*, 280:20131760, 2013. doi: <http://dx.doi.org/10.1098/rspb.2013.1760>.
 - [26] Nathan I. Wisnoski, Mario E. Muscarella, Megan L. Larsen, Ariane L. Peralta, and Jay T. Lennon. Metabolic insight into bacterial community assembly across ecosystem boundaries. *Ecology*, 101(4):e02968, 2020. ISSN 00129658. doi: 10.1002/ecy.2968.
 - [27] Jérôme Comte, Mercè Berga, Ina Severin, Jürg Brendan Logue, and Eva S. Lindström. Contribution of different bacterial dispersal sources to lakes: Population and community effects in different seasons. *Environmental Microbiology*, 19(6):2391–2404, 2017. ISSN 14622920. doi: 10.1111/1462-2920.13749.
 - [28] Noah Fierer, Mark A Bradford, and Robert B Jackson. Toward an Ecological Classification of Soil Bacteria. *Ecology*, 88(6):1354–1364, 2007.
 - [29] Jonathan J. Cole. Aquatic microbiology for ecosystem scientists: New and recycled paradigms in ecological microbiology. *Ecosystems*, 2:215–225, 1999. ISSN 14329840. doi: 10.1007/s100219900069.
 - [30] S. E. Jones and J. T. Lennon. Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy*

of Sciences, 107(13):5881–5886, 2010. ISSN 0027-8424. doi: 10.1073/pnas.0912765107. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0912765107>.

- [31] Allan Konopka. What is microbial community ecology ? *The ISME Journal*, 3:1223–1230, 2009. doi: 10.1038/ismej.2009.88.
- [32] Barbara J Campbell and David L Kirchman. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *The ISME Journal*, 7(1):210–220, 2013. ISSN 1751-7362. doi: 10.1038/ismej.2012.93. URL <http://www.nature.com/doifinder/10.1038/ismej.2012.93>.
- [33] Helena Osterholz, Gabriel Singer, Bernd Wemheuer, Rolf Daniel, Meinhard Simon, Jutta Niggemann, and Thorsten Dittmar. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *The ISME Journal*, 10: 1717–1730, 2016. ISSN 1751-7362. doi: 10.1038/ismej.2015.231. URL <http://dx.doi.org/10.1038/ismej.2015.231>.
- [34] Steven J Blazewicz, Romain L Barnard, Rebecca A Daly, and Mary K Firestone. Evaluating rRNA as an indicator of microbial activity in environmental communities : limitations and uses. *The ISME Journal*, 7:2061–2068, 2013. ISSN 1751-7362. doi: 10.1038/ismej.2013.102. URL <http://dx.doi.org/10.1038/ismej.2013.102>.
- [35] Mark Vellend. Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85(2):183–206, 2010.
- [36] Jizhong Zhou and Daliang Ning. Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbial and Molecular Biology Reviews*, 81(4):1–32, 2017.
- [37] James C Stegen, Xueju Lin, Jim K Fredrickson, Xingyuan Chen, David W Kennedy, Christopher J Murray, Mark L Rockhold, and Allan Konopka. Quantifying community assembly processes and identifying features that impose them. *The ISME Journal*, 7:2069–2079, 2013. ISSN 1751-7362. doi: 10.1038/ismej.2013.93. URL <http://dx.doi.org/10.1038/ismej.2013.93>.
- [38] James C Stegen, Xueju Lin, Jim K Fredrickson, and Allan E Konopka. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology*, 6(370):1–15, 2015. doi: 10.3389/fmicb.2015.00370.

- [39] Silke Langenheder, Jianjun Wang, Satu Maaria Karjalainen, Tiina M. Laamanen, Kimmo T. Tolonen, Annika Vilmi, and Jani Heino. Bacterial metacommunity organization in a highly connected aquatic system. *FEMS Microbiology Ecology*, 93(4):1–9, 2017. ISSN 15746941. doi: 10.1093/femsec/fiw225.
- [40] Emily B. Graham, Alex R. Crump, Charles T. Resch, Sarah Fansler, Evan Arntzen, David W. Kennedy, Jim K. Fredrickson, and James C. Stegen. Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. *Environmental Microbiology*, 19(4):1552–1567, 2017. ISSN 14622920. doi: 10.1111/1462-2920.13720.
- [41] Jonathan B. Losos. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11(10):995–1003, 2008. ISSN 1461023X. doi: 10.1111/j.1461-0248.2008.01229.x.
- [42] Dan L. Warren, Richard E. Glor, and Michael Turelli. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62(11):2868–2883, 2008. ISSN 00143820. doi: 10.1111/j.1558-5646.2008.00482.x.
- [43] Mark Achtman and Michael Wagner. Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6(6):431–440, 2008. ISSN 17401526. doi: 10.1038/nrmicro1872.
- [44] Christopher M. Thomas and Kaare M. Nielsen. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3(9):711–721, 2005. ISSN 17401526. doi: 10.1038/nrmicro1234.
- [45] Jennifer B H Martiny, Stuart E Jones, Jay T Lennon, and Adam C Martiny. Microbiomes in light of traits: A phylogenetic perspective. *Science*, 350(6261):aac9323, 2015. ISSN 10959203. doi: 10.1126/science.aac9323.
- [46] Adam C Martiny, Kathleen Treseder, and Gordon Pusch. Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, 7:830–838, 2013. ISSN 1751-7362. doi: 10.1038/ismej.2012.160. URL <http://dx.doi.org/10.1038/ismej.2012.160>.
- [47] Kristoffer Forslund, Shinichi Sunagawa, Luis P. Coelho, and Peer Bork. Metagenomic insights into the human gut resistome and

the forces that shape it. *BioEssays*, 36(3):316–329, 2014. ISSN 02659247. doi: 10.1002/bies.201300143.

- [48] Martin Ackermann. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, 13(8):497–508, 2015. ISSN 17401534. doi: 10.1038/nrmicro3491. URL <http://dx.doi.org/10.1038/nrmicro3491>.
- [49] Alexander B. Chase, Zulema Gomez-Lunar, Alberto E. Lopez, Jun-hui Li, Steven D. Allison, Adam C. Martiny, and Jennifer B.H. Martiny. Emergence of soil bacterial ecotypes along a climate gradient. *Environmental Microbiology*, 20(11):4112–4126, 2018. ISSN 14622920. doi: 10.1111/1462-2920.14405.
- [50] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12, 2013. ISSN 2226-6089. doi: 10.14806/ej.17.1.200. URL <http://journal.embnet.org/index.php/embnetjournal/article/view/200/479>.
- [51] Benjamin J. Callahan, Paul J. McMurdie, and Susan P. Holmes. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, 11(12):2639–2643, 2017. ISSN 17517370. doi: 10.1038/ismej.2017.119. URL <http://dx.doi.org/10.1038/ismej.2017.119>.
- [52] Erik S Wright. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, 8(1):352–359, 2016. ISSN 2073-4859. doi: V12242009.
- [53] Adithya Murali, Aniruddha Bhargava, and Erik S. Wright. IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, 6(1):1–14, 2018. ISSN 20492618. doi: 10.1186/s40168-018-0521-5.
- [54] Elmar Pruesse, Christian Quast, Katrin Knittel, Bernhard M Fuchs, Wolfgang Ludwig, Jörg Peplies, and Fank Oliver Glöckner. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21):7188–7196, 2007. ISSN 03051048. doi: 10.1093/nar/gkm864.

- [55] Tomáš Větrovský and Petr Baldrian. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, 8(2):1–10, 2013. ISSN 19326203. doi: 10.1371/journal.pone.0057923.
- [56] Joseph N. Paulson, O. Colin Stine, Héctor Corrada Bravo, and Mihai Pop. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, 10(12):1200–1202, 2013. ISSN 15487091. doi: 10.1038/nmeth.2658.
- [57] J. Roger Bray and J. T. Curtis. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4): 325–349, 1957. doi: 10.2307/1942268.
- [58] Pierre Legendre and Louis Legendre. *Numerical Ecology*. Elsevier, Amsterdam, 2nd edition, 1998.
- [59] E Paradis and K Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(526-528), 2018.
- [60] Pierre Legendre and Eugene D. Gallagher. Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2): 271–280, 2001. ISSN 00298549. doi: 10.1007/s004420100716.
- [61] Daniel Borcard, Francois Gillet, and Pierre Legendre. *Numerical Ecology with R*. Springer, New York, 2011. ISBN 9781441979759. doi: 10.1007/978-1-4419-7976-6.
- [62] Marti J. Anderson and Daniel C.I. Walsh. 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(4):557–574, 2013. doi: 10.1890. URL <http://www.esajournals.org/doi/abs/10.1890/12-2010.1>.
- [63] Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L Simpson, Peter Solymos, M Henry H Stevens, Eduard Szoecs, and Helene Wagner. vegan: Community Ecology Package, 2019. URL <https://cran.r-project.org/package=vegan>.
- [64] John Tabak. Differential Geometry. In *Geometry: The language of Space and Form*, page 150. Facts on File, Inc, New York, 2004. ISBN 9780816068760.

- [65] Thorvald Sørensen. 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. *Biologicaliske Skrifter Kongelige Danske Videnskabernes Selskab*, 5:1–34, 1948. ISSN 00155551. doi: 10.1007/BF02852438.
- [66] Ashley Shade, Stuart E Jones, J Gregory Caporaso, Jo Handelsman, Rob Knight, Noah Fierer, and A Gilbert. Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio*, 5(4):e01371–14, 2014. doi: 10.1128/mBio.01371-14.Editor.
- [67] Paul J. McMurdie and Susan Holmes. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4):e61217, 2013.
- [68] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- [69] Hadley Wickham. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1):1–29, 2011. URL <http://www.jstatsoft.org/v40/i01/>.
- [70] Matt Dowle and Arun Srinivasan. data.table: Extension of ‘data.frame’, 2019. URL <https://cran.r-project.org/package=data.table>.
- [71] Revolution Analytics and Steve Weston. doMC: Foreach Parallel Adaptor for ‘parallel’, 2019. URL <https://cran.r-project.org/package=doMC>.
- [72] R Core Team. R: A language and environment for statistical computing, 2020. URL <https://www.r-project.org/>.
- [73] H Wickham. ggplot2: Elegant Graphics for Data Analysis, 2016. URL <https://ggplot2.tidyverse.org>.
- [74] A Kassambara. ggpubr: ‘ggplot2’ Based Publication Ready Plots, 2020.

- [75] Elio Campitelli. *ggnewscale*: Multiple Fill and Colour Scales in 'ggplot2', 2020. URL <https://cran.r-project.org/package=ggnewscale>.
- [76] Claus O. Wilke. *cowplot*: Streamlined Plot Theme and Plot Annotations for 'ggplot2', 2019. URL <https://cran.r-project.org/package=cowplot>.
- [77] Alboukadel Kassambara. *rstatix*: Pipe-Friendly Framework for Basic Statistical Tests, 2020. URL <https://cran.r-project.org/package=rstatix>.
- [78] RStudio Team. RStudio: Integrated Development for R, 2020. URL <http://www.rstudio.com/>.
- [79] Francesco Carrara, Andrea Rinaldo, Andrea Giometto, and Florian Altermatt. Complex Interaction of Dendritic Connectivity and Hierarchical Patch Size on Biodiversity in River-Like Landscapes. *The American Naturalist*, 183(1):13–25, 2013. ISSN 0003-0147. doi: 10.1086/674009.
- [80] Frank Rasche, Daniela Knapp, Christina Kaiser, Marianne Koranda, Barbara Kitzler, Sophie Zechmeister-Boltenstern, Andreas Richter, and Angela Sessitsch. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *ISME Journal*, 5:389–402, 2011. ISSN 17517362. doi: 10.1038/ismej.2010.138. URL <http://dx.doi.org/10.1038/ismej.2010.138>.
- [81] Christina Kaiser, Marianne Koranda, Barbara Kitzler, Lucia Fuchsleger, Jörg Schnecker, Peter Schweiger, Frank Rasche, Sophie Zechmeister-Boltenstern, Angela Sessitsch, and Andreas Richter. Belowground carbon allocation by trees drives seasonal patterns of extracellular enzyme activities by altering microbial community composition in a beech forest soil. *New Phytologist*, 187(3):843–858, 2010. ISSN 0028646X. doi: 10.1111/j.1469-8137.2010.03321.x.
- [82] Byron C Crump, George W Kling, Michele Bahr, and John E Hobbie. Bacterioplankton Community Shifts in an Arctic Lake Correlate with Seasonal Changes in Organic Matter Source. *Applied and environmental microbiology*, 69(4):2253–2268, 2003. ISSN 00992240. doi: 10.1128/AEM.69.4.2253.

- [83] Emily L. Kara, Paul C. Hanson, Yu Hen Hu, Luke Winslow, and Katherine D. McMahon. A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. *ISME Journal*, 7(3):680–684, 2013. ISSN 17517362. doi: 10.1038/ismej.2012.118. URL <http://dx.doi.org/10.1038/ismej.2012.118>.
- [84] Juan Pablo Niño-García, Clara Ruiz-Gonzalez, Paul A. del Giorgio, Clara Ruiz-González, and Paul A. del Giorgio. Exploring the Ecological Coherence between the Spatial and Temporal Patterns of Bacterioplankton in Boreal Lakes. *Frontiers in Microbiology*, 8(APR):1–12, 2017. ISSN 1664302X. doi: 10.3389/fmicb.2017.00636.
- [85] Clara Ruiz-González, Lorenzo Proia, Isabel Ferrera, Josep M. Gasol, and Sergi Sabater. Effects of large river dam regulation on bacterioplankton community structure. *FEMS Microbiology Ecology*, 84(2):316–331, 2013. ISSN 01686496. doi: 10.1111/1574-6941.12063.
- [86] Clara Ruiz-González, Guillem Salazar, Ramiro Logares, Lorenzo Proia, Josep M. Gasol, and Sergi Sabater. Weak coherence in abundance patterns between bacterial classes and their constituent OTUs along a regulated river. *Frontiers in Microbiology*, 6(NOV):1–13, 2015. ISSN 1664302X. doi: 10.3389/fmicb.2015.01293.
- [87] Paula C.J. Reis, Clara Ruiz-González, Sophie Crevecoeur, Cynthia Soued, and Yves T. Prairie. Rapid shifts in methanotrophic bacterial communities mitigate methane emissions from a tropical hydropower reservoir and its downstream river. *Science of the Total Environment*, 748:141374, 2020. ISSN 18791026. doi: 10.1016/j.scitotenv.2020.141374. URL <https://doi.org/10.1016/j.scitotenv.2020.141374>.
- [88] Thierry C. Bouvier and Paul A. del Giorgio. Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnology and Oceanography*, 47(2):453–470, 2002.
- [89] Byron C. Crump, Charles S. Hopkinson, Mitchell L. Sogin, and John E. Hobbie. Microbial Biogeography along an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time. *Applied and Environmental Microbiology*, 70(3):1494–1505, 2004. ISSN 00992240. doi: 10.1128/AEM.70.3.1494-1505.2004.

- [90] Daniel S Read, Hyun S Gweon, Michael J Bowes, Lindsay K Newbold, Dawn Field, Mark J Bailey, and Robert I Griffiths. Catchment-scale biogeography of riverine bacterioplankton. *The ISME Journal*, 9:516–526, 2015. ISSN 1751-7362. doi: 10.1038/ismej.2014.166. URL <http://dx.doi.org/10.1038/ismej.2014.166>.
- [91] Zheng Yu, Jun Yang, Stefano Amalfitano, Xiaoqing Yu, and Lemian Liu. Effects of water stratification and mixing on microbial community structure in a subtropical deep reservoir. *Scientific Reports*, 4:5821, 2014. ISSN 20452322. doi: 10.1038/srep05821.
- [92] B. J. Campbell, L. Yu, J. F. Heidelberg, and D. L. Kirchman. Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences*, 108(31):12776–12781, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1101405108. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1101405108>.
- [93] Shubha N. Pandit, Jurek Kolasa, and Karl Cottenie. Contrasts between habitat generalists and specialists: An empirical extension to the basic metacommunity framework. *Ecology*, 90(8):2253–2262, 2009. ISSN 00129658. doi: 10.1890/08-0851.1.
- [94] C Monard, S Gantner, S Bertilsson, S Hallin, and J Stenlid. Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports*, 6:37719, 2016. doi: 10.1038/srep37719. URL <http://dx.doi.org/10.1038/srep37719>.
- [95] Jay T Lennon and Stuart E Jones. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, 9:119–130, 2011. ISSN 1740-1526. doi: 10.1038/nrmicro2504. URL <http://dx.doi.org/10.1038/nrmicro2504>.

Figure legends

Figure 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).

Figure 2. Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons. Overall PCoA analysis of DNA samples (a) has been further explored with focus plots on terrestrial, riverine and estuary samples in panels b, c, d, respectively. Overall, the PCoA reveals microbial community shifts from terrestrial to freshwater samples. Spring and summer-autumn show distinct paths in multivariate space. Percentage of variance explained are given in square brackets for the first and second axes.

Figure 3. RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially influenced habitats. PCoA analysis including RNA samples. a) Visualization of first and second axis of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis using the second and third axis, differentiating nucleic acid type and seasons, respectively. Percentage of variance explained by the corresponding axes are given in square brackets.

Figure 4. Residuals to m_S and m_{BC} 1:1 line reveal transition from mass effects to species sorting along the continuum. a) Distance between DNA and RNA of the same sample based on Sørensen dissimilarity (m_S) against distance calculated with Bray-Curtis dissimilarity (m_{BC}). Distance was calculated within the axes capturing 75 % of the variance in both PCoAs using different dissimilarity measures. b) Residuals to 1:1 line between m_S and m_{BC} along the terrestrial-aquatic continuum. Habitats outside the direct continuum are given separately. Points represent the arithmetic mean and error bars represent the standard error (a) and standard deviations from the mean (b). Sample sizes of each point are indicated above the points. Sample sizes are equivalent in panels a and b.

Figure 5. DNA and RNA dissimilarities arise through abundance differences across the rank abundance curve. a) Difference in DNA and RNA abundance (CSS reads) against m_{BC} . Given are absolute values. Lines represent rolling means with bin size = 10. b) Boxplots and vio-

lin plots showing the distribution of species scores and thus OTU optima within PCoA space. The middle line represents the median, lower and upper hinges of boxplots correspond to the 25th and 75th percentiles. Upper and lower whiskers expand to the largest and smallest value, respectively, but no further than 1.5 times the inter-quartile range from the hinge. There were a plethora of outliers that lie beyond the whiskers for all boxes and thus, were removed for visualization purposes. Violin plots visualize the probability density distribution smoothed by a kernel density estimator.

Table 1: Abundance groups

Abundance groups	Criteria	Categorized OTUs
Universal abundant	Abundant* in all habitats, never absent	0
Universal moderate	Moderate* in all habitats, never absent	0
Universal rare	Rare* in all habitats, never absent	0
Abundant	Only abundant observations, can be absent	1
Moderate	Only moderate observations, can be absent	4
Rare	Only rare observations, can be absent	17459
Specialist	Only abundant in one habitat type, never abundant in any other habitats	1561
Cosmopolitan	Shifts between abundant, moderate and rare, but present in all habitats	46
Shifter	Shifts between abundant, moderate and rare [†]	342
Upper shifter	Shifts in the upper fraction (abundant - moderate) of the rank abundance curve [†]	755
Lower shifter	Shifts in the lower fraction (moderate - rare) of the rank abundance curve [†]	17

* Abundant: $47 \geq \text{CSS reads}$, Moderate: $5 \geq \text{CSS reads} < 47$, Rare: $0 > \text{CSS reads} < 5$

[†] Does not have to be present in all habitats

Tables

Figures

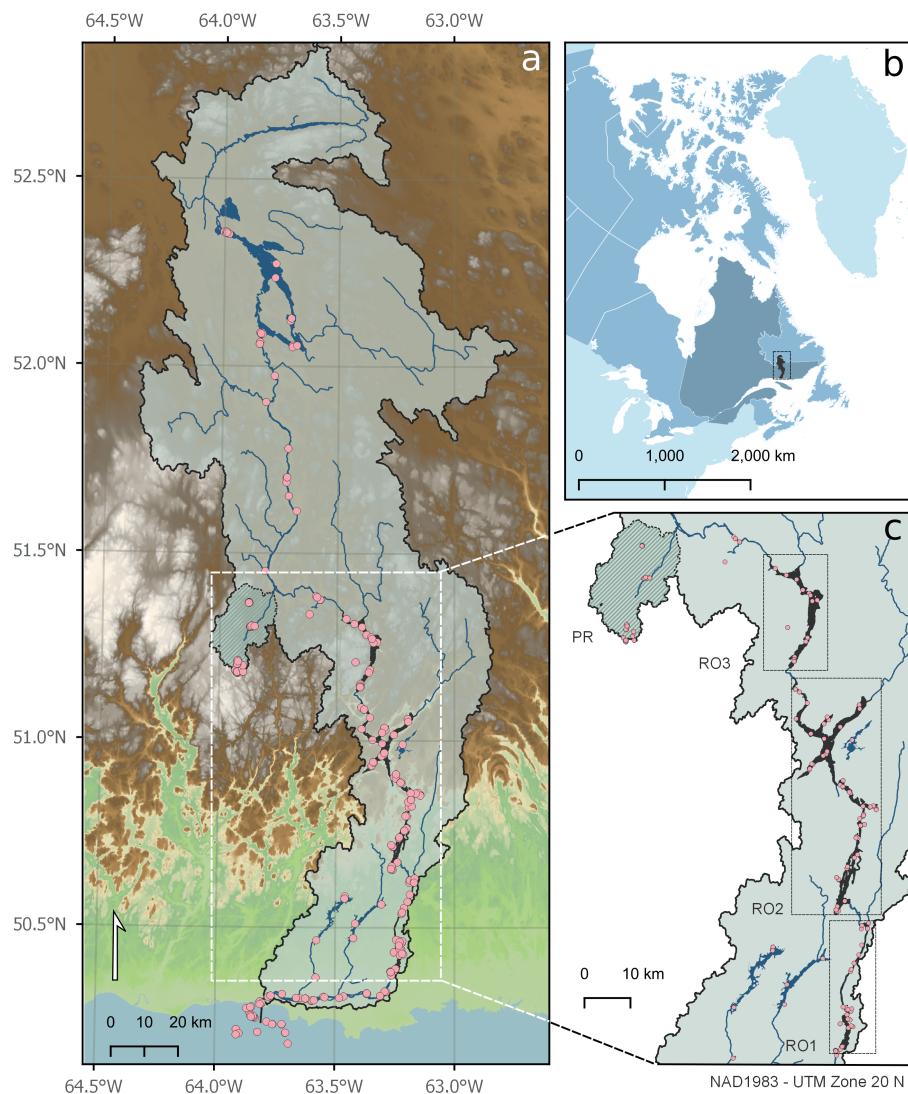


Figure 1: Location and overview of the La Romaine catchment. a) Scale and overview of the whole La Romaine catchment. Samples are represented as points. b) Location of the catchment within Canada and Québec. c) Focus on all built reservoirs RO1 (2015), RO2 (2014) and RO3 (2017) and the headwater stream sub-catchment Petite Romaine (PR).

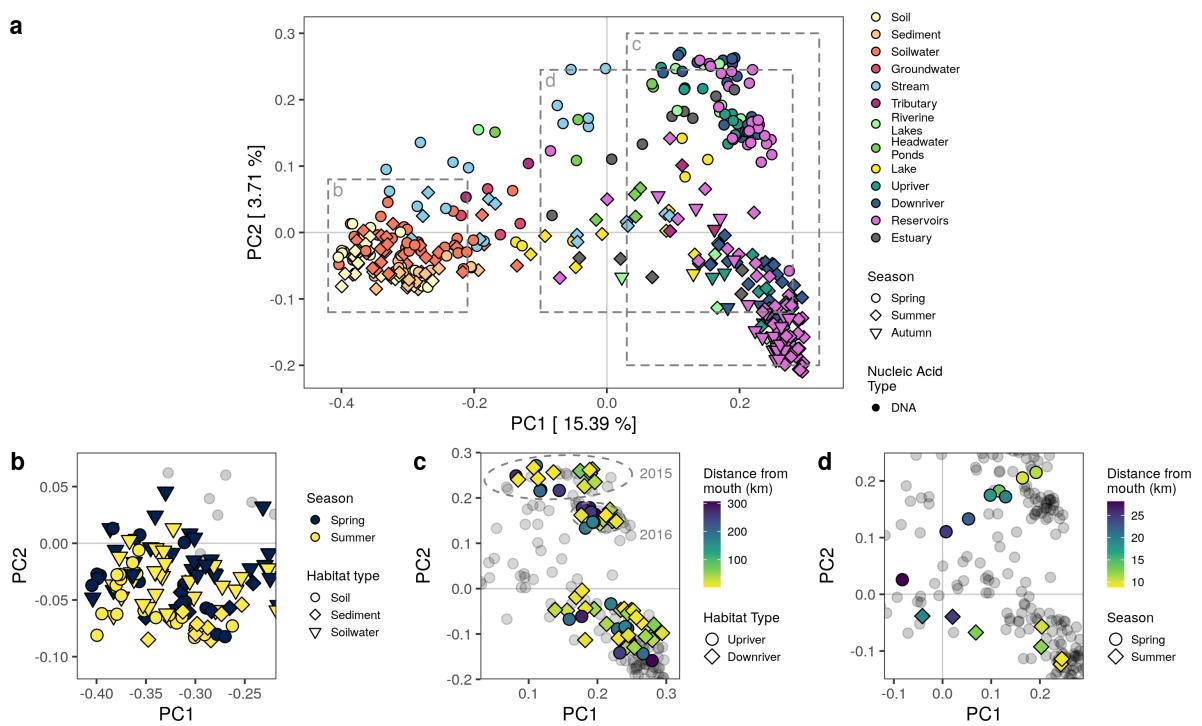


Figure 2: Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons. Overall PCoA analysis of DNA samples (a) has been further explored with focus plots on terrestrial, riverine and estuary samples in panels b, c, d, respectively. Overall, the PCoA reveals microbial community shifts from terrestrial to freshwater samples. Spring and summer-autumn show distinct paths in multivariate space. Percentage of variance explained are given in square brackets for the first and second axes.

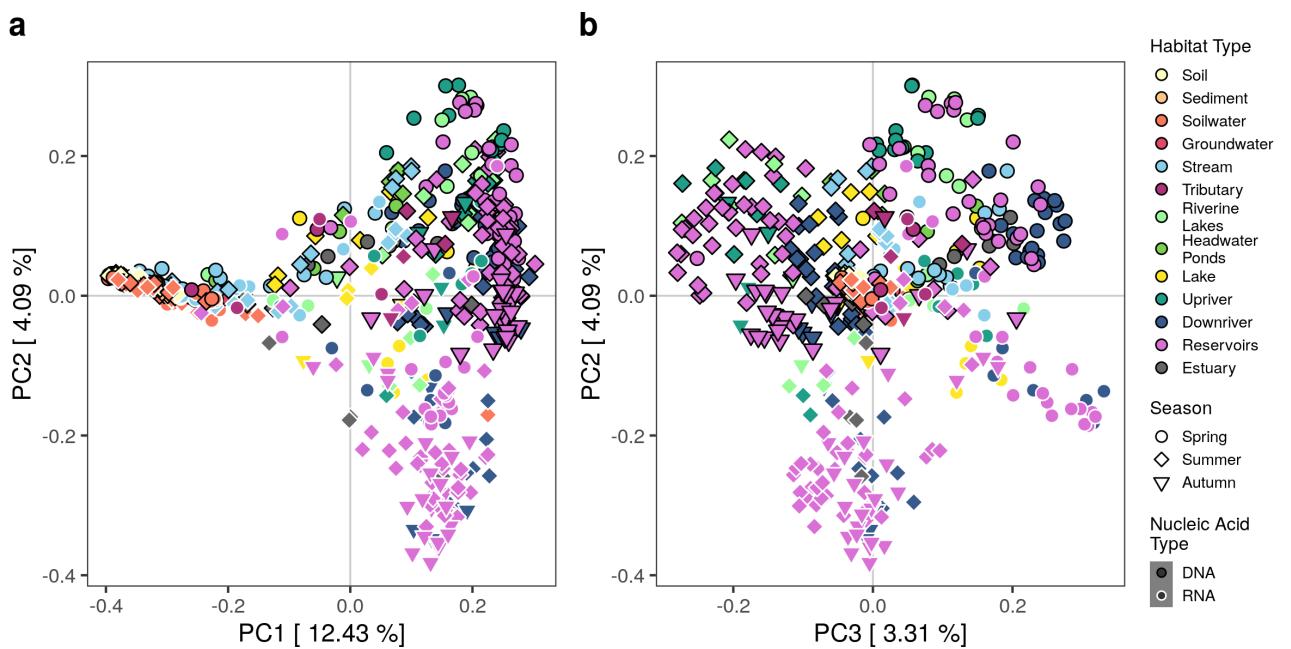


Figure 3: RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially influenced habitats. PCoA analysis including RNA samples. a) Visualisation of first and second axis of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis using the second and third axis, differentiating nucleic acid type and seasons, respectively. Percentage of variance explained by the corresponding axes are given in square brackets.

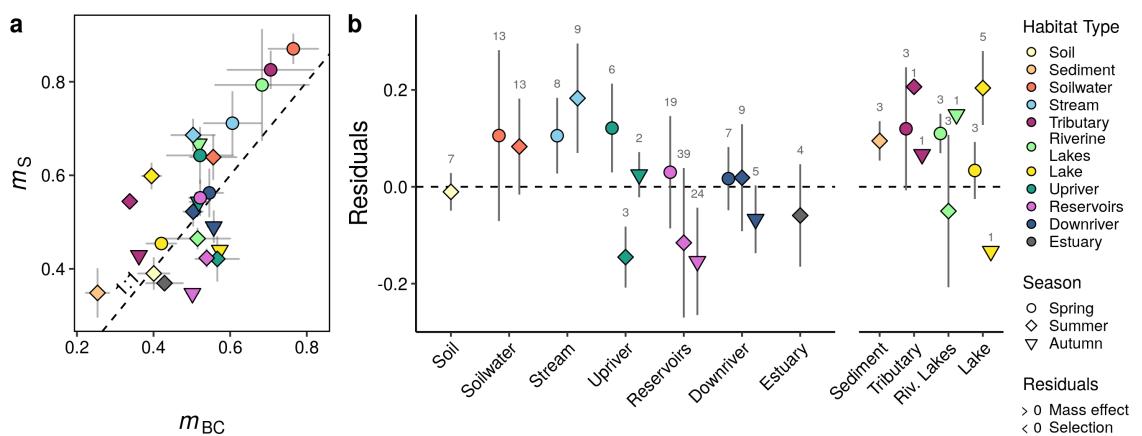


Figure 4: Residuals to m_S and m_{BC} 1:1 line reveal transition from mass effects to species sorting along the continuum. a) Distance between DNA and RNA of the same sample based on Sørensen dissimilarity (m_S) against distance calculated with Bray-Curtis dissimilarity (m_{BC}). Distance was calculated within the axes capturing 75 % of the variance in both PCoAs using different dissimilarity measures. b) Residuals to 1:1 line between m_S and m_{BC} along the terrestrial-aquatic continuum. Habitats outside of the direct continuum are given separately. Points represent the arithmetic mean and error bars represent the standard error (a) and standard deviations from the mean (b). Sample sizes of each point are indicated above the points. Sample sizes are equivalent in panels a and b.

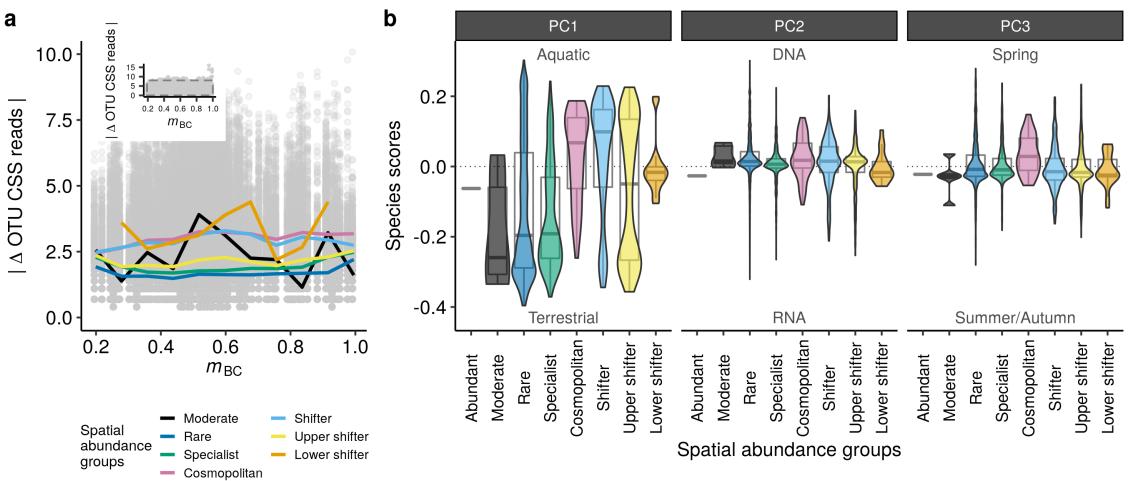


Figure 5: DNA and RNA dissimilarities arise through abundance differences across the rank abundance curve. a) Difference in DNA and RNA abundance (CSS reads) against m_{BC} . Given are absolute values. Lines represent rolling means with bin size = 10. b) Boxplots and violin plots showing the distribution of species scores and thus OTU optima within PCoA space. The middle line represents the median, lower and upper hinges of boxplots correspond to the 25th and 75th percentiles. Upper and lower whiskers expand to the largest and smallest value, respectively, but no further than 1.5 times the inter-quartile range from the hinge. There were a plethora of outliers that lie beyond the whiskers for all boxes and thus, were removed for visualisation purposes. Violin plots visualise the probability density distribution smoothed by a kernel density estimator.