

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 over 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 dawn of next-generation sequencing, microbial ecologists have documented a recurrent shape of rank abundance distributions across ecosystems. Rank abundance distributions with few abundant and a long tail of rare microorganisms have become a text-book characteristic of most environmental microbial communities, yet, we still know relatively little on how this distribution comes 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). However, these interpretations have seldom been explicitly confirmed. Inherited from macro-ecology, microbial assembly processes have similarly 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 within community assembly as local microbial communities reflect the balance between selection and dispersal processes from connected habitats that have occurred in the past (Fukami, 2004). Hence, accounting for community history is vital to understand community assembly.

Studies that investigate the relevance of history mainly followed microbes in time within an ecosystem (Shade and Gilbert, 2015). Temporal

history does shape local communities (i.e. legacy effects, Fukami (2015)), however, within the aquatic network, the uni-directional flow of water links temporal to spatial history. Evidenced by soil microbes being flushed into and occupying large proportions of aquatic communities (Ruiz-González et al., 2015a; Hauptmann et al., 2016; Crump et al., 2012; Besemer et al., 2013; Wisniski et al., 2020), hydrology is a major driver of aquatic microbial community composition (Niño-García et al., 2016a). As such, network connectivity is modulated by seasonal hydrological fluctuations (de Melo et al., 2019) and community structure at any given site within a fluvial network is the net result of upstream assembly processes. Therefore, spatial history is particularly relevant in highly inter-connected freshwater networks (Vass and Langenheder, 2017), and there have been various studies that investigate the spatial context of microbial community assembly. Stegen et al. (2013) quantify 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., 2016b). While the importance of mingling and interacting communities between different ecosystems is now amply recognized (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 of lakes (Logue and Lindström, 2010; Adams et al., 2014; Langen-

heder et al., 2017), along a single river (Winter et al., 2007; Savio et al., 2015; Hauptmann et al., 2016; Doherty et al., 2017; Gweon et al., 2020) or on interconnected stream networks (Besemer et al., 2013; Widder et al., 2014; Read et al., 2015; Hassell et al., 2018; Wisnoski and Lennon, 2020) and rarely have surrounding terrestrial ecosystems been considered as potential sources (Crump et al., 2012; Ruiz-González et al., 2015a; Wisnoski et al., 2020). Moreover, active and passive assembly processes are difficult to resolve as cell death and dormancy blur interpretations based on DNA (Cole, 1999; Jones and Lennon, 2010). Indicative of recent protein synthesis, RNA sequencing has helped to disentangle active from passive microbial members (Bowsher et al., 2019), however, only few freshwater studies have expanded their molecular tool-box (Logue and Lindström, 2010; Székely et al., 2013; Aanderud et al., 2016; Peter et al., 2018; Wisnoski et al., 2020). 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 these shifting importances of assembly processes and link those to the underlying rank abundance structure, we firstly 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 distinguished otherwise. In this study, we attempted a more holistic approach on aquatic community assembly by addressing the afore mentioned three critical steps.

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 have carried out this study within the La Romaine river watershed in the North-Eastern region of boreal Québec, Canada, over several years and seasons. Starting from upstream sources such as soils, soil-waters, and headwater streams, we continued to follow the extent river orders (Strahler order 0-8) until the estuary. Additionally, 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) community structure shifts along the terrestrial-freshwater-estuary continuum; we determine the general patterns of the spatial succession and its relation to different hydrologic conditions (i.e. seasons). We further differentiate the

reactive from the total bacterial assemblage by additionally sequencing RNA. We do not use RNA as an indication of activity per se (Blazewicz et al., 2013), rather we interpret the patterns of divergence and convergence of 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 community structures, we developed distance metrics with either incidence (presence-absence) or abundance-based dissimilarities. In a null scenario, DNA-/RNA-based assemblage structure remains 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, influencing mostly the abundance-based distance. The spatial patterns in the incidence- and abundance-based metrics therefore provide insight on how selection and mass effects vary along the continuum. And finally, we explore where taxa that exhibit strong selective dynamics along the continuum are located within the rank abundance curve.

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 km 2 . 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 km 2 , 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 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 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, Martin (2013)). 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. (2017)). 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, Wright (2016)) implementing the increased accuracy IDTAXA algorithm (Murali et al., 2018) and the provided trained classifier of the SILVA database (Version 138, Pruesse et al. (2007)). 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 threshold (Větrovský and Baldrian, 2013) with the *DECIPHER* package (Wright, 2016).

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) (Paulson et al., 2013). 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}) (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) ($n = 372, 20182$ OTUs). 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 9999 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 *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* package (Oksanen et al., 2019).

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 space (Tabak, 2004):

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

where p and q represent DNA and RNA site scores, respectively, of each sample and n is the used maximum number of dimensions. We 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 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) (Legendre and Legendre, 1998; Sørensen, 1948)(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 taxa (Shade et al., 2014), 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* (Oksanen et al., 2019). 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 (McMurdie and Holmes, 2013; Wick-

ham et al., 2019; Wickham, 2011; Dowle and Srinivasan, 2019), and *doMC* and *parallel* enabled parallel processing (Revolution Analytics and Weston, 2019; R Core Team, 2020). *ggplot2*, *ggpubr*, *ggnewscale* and *cowplot* were used to visualize the results (Wickham, 2016; Kassambara, 2020a; Campitelli, 2020; Wilke, 2019). For statistical analyses, *vegan* and *rstatix* were used (Oksanen et al., 2019; Kassambara, 2020b). All analyses have been conducted in R v3.4.2 (R Core Team, 2020) and RStudio v1.3.1073 (RStudio Team, 2020). 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 repre-

sented 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 PER-

MANOVA 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 vast 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 (Ruiz-González et al., 2015a; Hauptmann et al., 2016), especially in systems with lower residence times and higher connectivity to the surrounding terrestrial systems (Crump et al., 2012; Besemer et al., 2013). 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 low-land lakes, and thus differences can arise through differences in network position (Carrara et al., 2013), residence times (Logue and Lindström, 2010) 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 (Shigyo et al., 2019; Rasche et al., 2011). Seasonality may take shape as different enzyme expression levels (Kaiser et al., 2010), which cannot be excluded.

Although seasonality within lakes has been widely reported (Crump et al., 2003; Kara et al., 2013; Niño-García et al., 2017), 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 hydrology (Niño-García et al., 2016a) 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 (Ruiz-González et al., 2013, 2015b; Reis et al., 2020), 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 effects (Hauptmann et al., 2016) that slowly disappear over a 25 km stretch as salinity increases (Bouvier and del Giorgio, 2002; Crump et al., 2004). In contrast to Doherty et al. (2017), 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 inci-

dence driven DNA-RNA differences are influxes of inactive bacteria (i.e mass effects) (Leibold et al., 2004). 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 OTUs (Campbell and Kirchman, 2013).

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 (Ruiz-González et al., 2015a; Hauptmann et al., 2016; Crump et al., 2012). 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 (Niño-García et al., 2016a; Read et al., 2015). 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 (Ruiz-González et al., 2013, 2015b) 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 community (Yu et al., 2014), and thus river specific selection (i.e. water residence time) continued in the downriver (Read et al., 2015). The estuary indicates a tendency towards selection likely to be driven by salinity gradients (Bouvier and del Giorgio, 2002; Crump et al., 2004), 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 OTUs (Campbell and Kirchman, 2013).

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 (Campbell and Kirchman, 2013; Campbell et al., 2011).

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 domains (Pandit et al., 2009). 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 strong (Monard et al., 2016). The previously observed numerical dominance of terrestrial taxa within freshwaters (Ruiz-González et al., 2015a), 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' proposes (Lennon and Jones, 2011).

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 summer (Ward and Stanford, 1983). 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 number SRxxxxx part of the BioProject PRJNA69302. 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

- 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.
- 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.
- 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(4):557–574.
- 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.
- Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. Springer, New York.
- Bouvier, T. C. and del Giorgio, P. A. (2002). Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnology and Oceanography*, 47(2):453–470.
- 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(2):e00003–19.
- Bray, J. R. and Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4):325–349.
- 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.
- Campbell, B. J. and Kirchman, D. L. (2013). Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *The ISME Journal*, 7:210–220.

- 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(31):12776–12781.
- Campitelli, E. (2020). ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'.
- 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(1):13–25.
- Cole, J. J. (1999). Aquatic microbiology for ecosystem scientists: New and recycled paradigms in ecological microbiology. *Ecosystems*, 2:215–225.
- 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.
- Crump, B. C., Hopkinson, C. S., Sogin, M. L., and Hobbie, J. E. (2004). Microbial Biogeography along an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time. *Applied and Environmental Microbiology*, 70(3):1494–1505.
- 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(4):2253–2268.
- 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(1):108–120.
- 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*.
- Dowle, M. and Srinivasan, A. (2019). data.table: Extension of 'data.frame'.
- Fukami, T. (2004). Assembly history interacts with ecosystem size to influence species diversity. *Ecology*, 85(12):3234–3242.

- 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.
- 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*.
- 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(10):3560–3572.
- 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(1474):1–16.
- Jones, S. E. and Lennon, J. T. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences*, 107(13):5881–5886.
- Kaiser, C., Koranda, M., Kitzler, B., Fuchslueger, L., Schnecker, J., Schweiger, P., Rasche, F., Zechmeister-Boltenstern, S., Sessitsch, A., and Richter, A. (2010). 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.
- Kara, E. L., Hanson, P. C., Hu, Y. H., Winslow, L., and McMahon, K. D. (2013). A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. *The ISME Journal*, 7:680–684.
- Kassambara, A. (2020a). ggpubr: 'ggplot2' Based Publication Ready Plots.
- Kassambara, A. (2020b). rstatix: Pipe-Friendly Framework for Basic Statistical Tests.

- 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(4):1–9.
- Legendre, P. and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2):271–280.
- Legendre, P. and Legendre, L. (1998). *Numerical Ecology*. Elsevier, Amsterdam, 2nd edition.
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., and Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7):601–613.
- Lennon, J. T. and Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, 9:119–130.
- 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.
- Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422(6933):714–716.
- 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.
- Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12.
- 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(10):995–1015.
- McMurdie, P. J. and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4):e61217.

- 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.
- Murali, A., Bhargava, A., and Wright, E. S. (2018). IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, 6(1):1–14.
- 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.
- 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(1):135–144.
- 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.
- 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(12):1506–1515.
- 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(APR):1–12.
- 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.
- Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Nigemann, 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.

- Pandit, S. N., Kolasa, J., and Cottenie, K. (2009). Contrasts between habitat generalists and specialists: An empirical extension to the basic meta-community framework. *Ecology*, 90(8):2253–2262.
- Paradis, E. and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(526-528).
- Paulson, J. N., Colin Stine, O., Bravo, H. C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, 10(12):1200–1202.
- Peter, H., Jeppesen, E., De Meester, L., 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.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., and Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21):7188–7196.
- R Core Team (2020). R: A language and environment for statistical computing.
- Rasche, F., Knapp, D., Kaiser, C., Koranda, M., Kitzler, B., Zechmeister-Boltenstern, S., Richter, A., and Sessitsch, A. (2011). Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *The ISME Journal*, 5:389–402.
- 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.
- Reis, P. C., Ruiz-González, C., Crevecoeur, S., Soued, C., and Prairie, Y. T. (2020). 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.
- Revolution Analytics and Weston, S. (2019). doMC: Foreach Parallel Adaptor for ‘parallel’.
- RStudio Team (2020). RStudio: Integrated Development for R.

- 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.
- 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(2):316–331.
- Ruiz-González, C., Salazar, G., Logares, R., Proia, L., Gasol, J. M., and Sabater, S. (2015b). Weak coherence in abundance patterns between bacterial classes and their constituent OTUs along a regulated river. *Frontiers in Microbiology*, 6(NOV):1–13.
- 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(12):4994–5007.
- Shade, A. and Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, 23(6):335–340.
- Shade, A., Gregory Caporaso, J., 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.
- Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, A. (2014). Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio*, 5(4):e01371–14.
- Shigyo, N., Umeki, K., and Hirao, T. (2019). Seasonal dynamics of soil fungal and bacterial communities in cool-temperate montane forests. *Frontiers in Microbiology*, 10(AUG):1–14.
- 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.
- 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.

- 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.
- Tabak, J. (2004). Differential Geometry. In *Geometry: The language of Space and Form*, page 150. Facts on File, Inc, New York.
- Vass, M. and Langenheder, S. (2017). The legacy of the past: Effects of historical processes on microbial metacommunities. *Aquatic Microbial Ecology*, 79:13–19.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85(2):183–206.
- 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(2):1–10.
- Ward, J. V. and Stanford, J. A. (1983). The Serial Discontinuity Concept of Lotic Ecosystems. In Fontaine III, T. D. and Bartell, S. M., editors, *Dynamics of Lotic Ecosystems*, pages 29–42. Ann Arbor Science, Ann Arbor.
- Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1):1–29.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Gromelund, 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(43):1686.
- 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(35):12799–12804.
- Wilke, C. O. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.

- 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(2):421–431.
- Wisnoski, N. I. and Lennon, J. T. (2020). Microbial community assembly in a multi-layer dendritic metacommunity. *Oecologia*.
- 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(4):e02968.
- Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, 8(1):352–359.
- Yu, Z., Yang, J., Amalfitano, S., Yu, X., and Liu, L. (2014). Effects of water stratification and mixing on microbial community structure in a subtropical deep reservoir. *Scientific Reports*, 4:5821.
- Zhou, J. and Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbial and Molecular Biology Reviews*, 81(4):1–32.

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

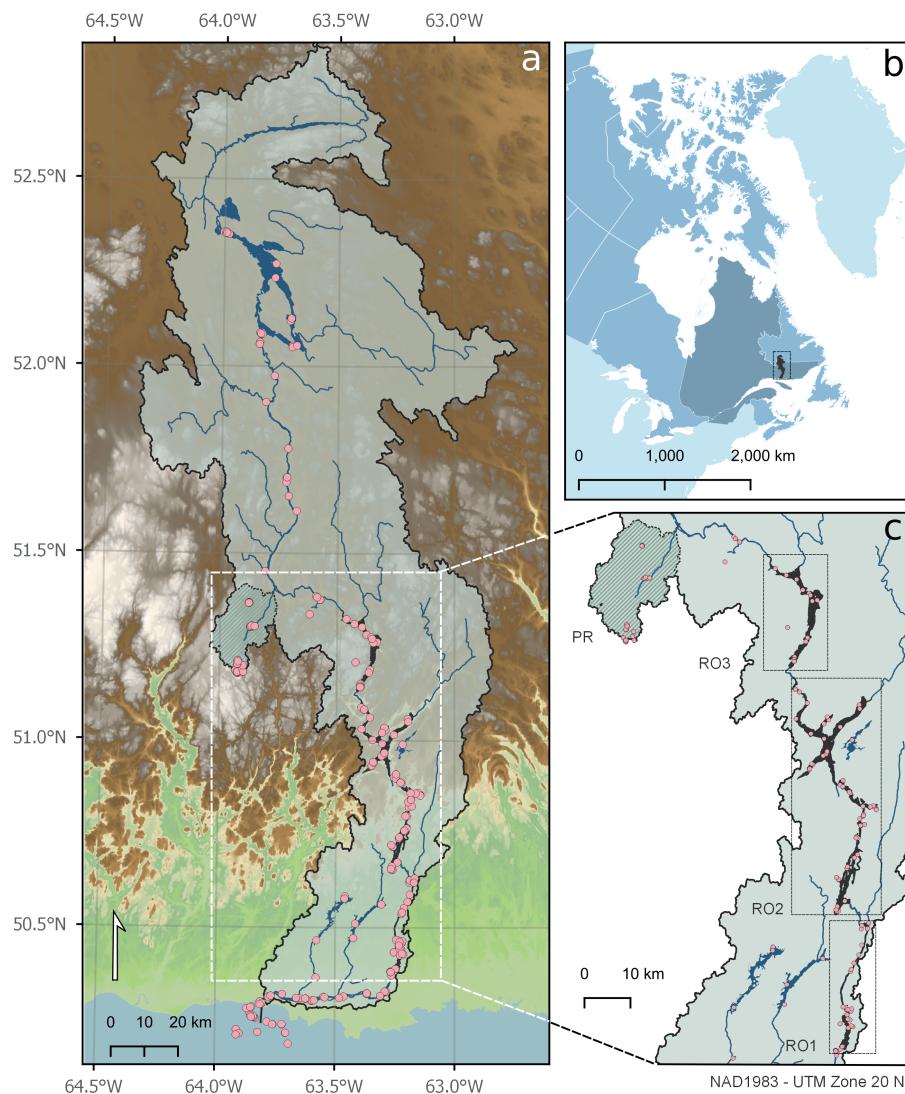


Figure 1: Location and overview of the La Romaine catchment. a) Scale and overview of the whole La Romaine catchment. Samples are represented as points. b) Location of the catchment within Canada and Québec. c) Focus on all built reservoirs RO1 (2015), RO2 (2014) and RO3 (2017) and the headwater stream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

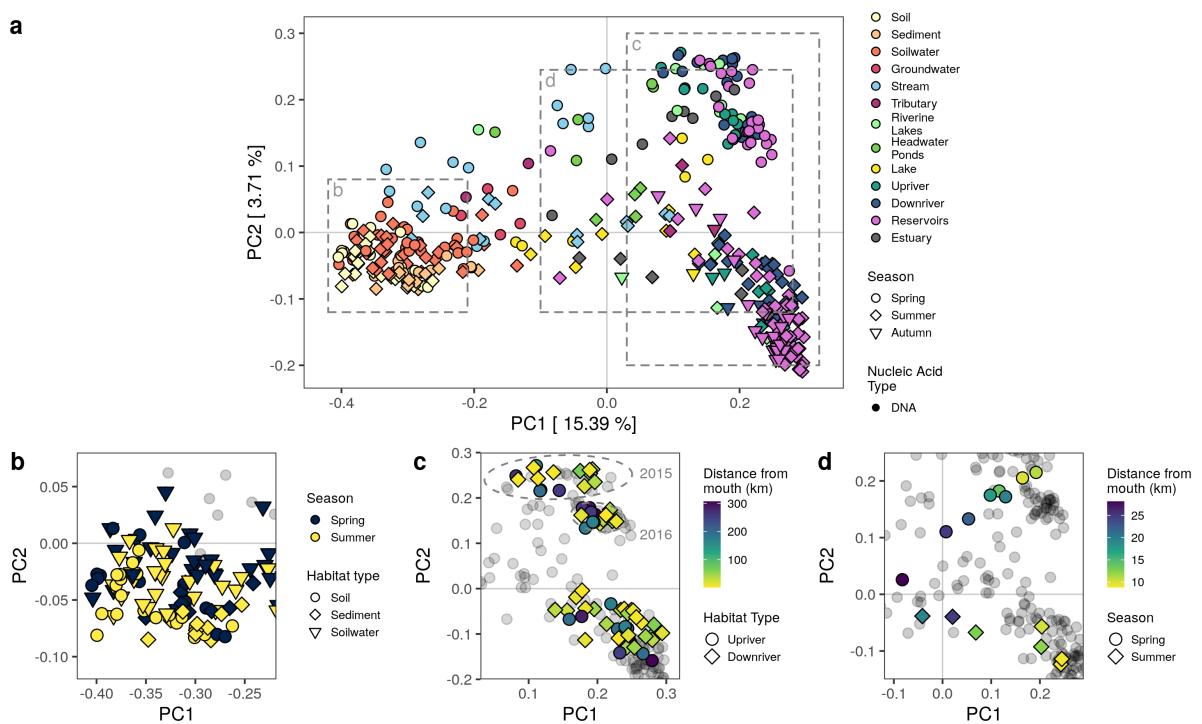


Figure 2: Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons. Overall PCoA analysis of DNA samples (a) has been further explored with focus plots on terrestrial, riverine and estuary samples in panels b, c, d, respectively. Overall, 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.

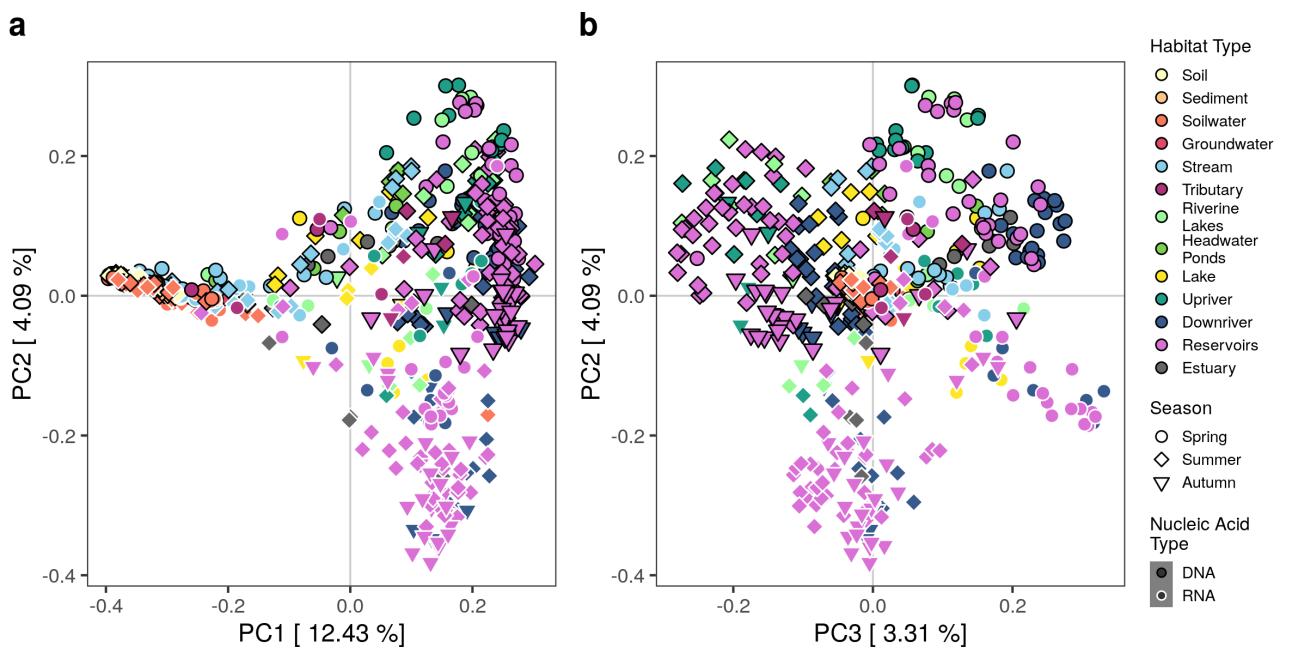


Figure 3: RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially influenced habitats. PCoA analysis including RNA samples. a) Visualisation of first and second axis of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis 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.

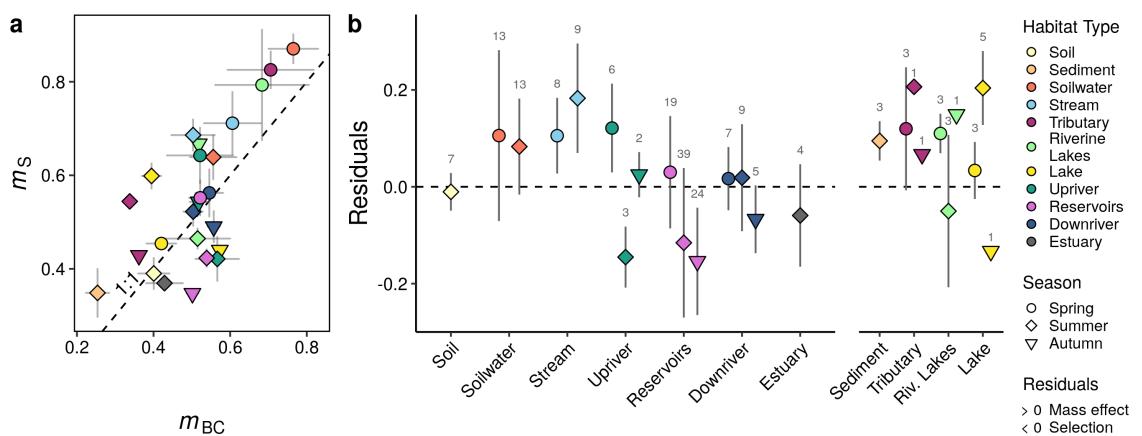


Figure 4: Residuals to m_S and m_{BC} 1:1 line reveal transition from mass effects to species sorting along the continuum. a) Distance between DNA and RNA of the same sample based on Sørensen dissimilarity (m_S) against distance calculated with Bray-Curtis dissimilarity (m_{BC}). Distance was calculated 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.

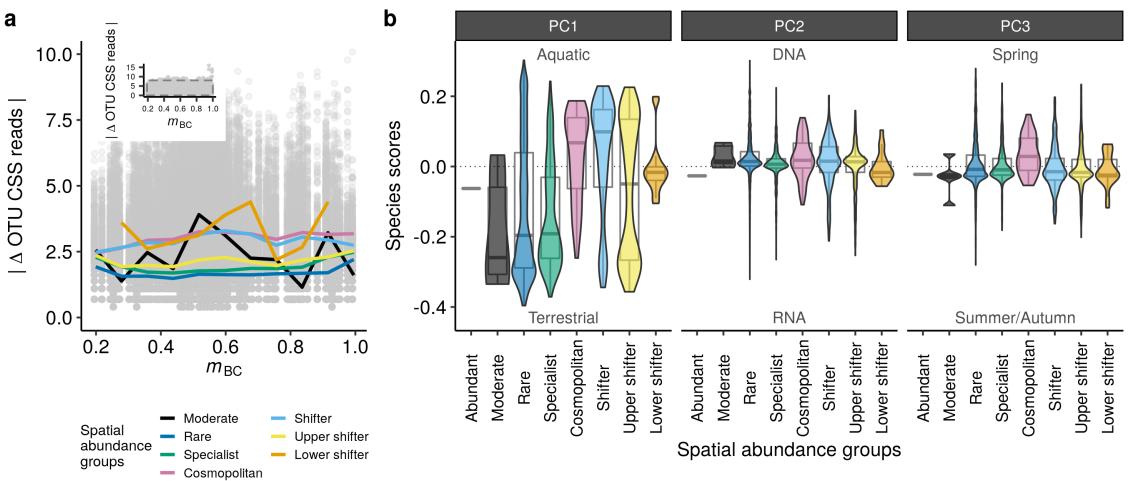


Figure 5: DNA and RNA dissimilarities arise through abundance differences across the rank abundance curve. a) Difference in DNA and RNA abundance (CSS reads) against m_{BC} . Given are absolute values. Lines represent rolling means with bin size = 10. b) Boxplots and violin plots showing the distribution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