

Title (max 150 char): DNA-RNA dynamics reveal bacterioplankton assembly processes along a 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): Microbial communities 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, 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.

Introduction

Since the 1880s, we are starting to grasp the intriguing variety of microbial physiological pathways and how they are involved in many elemental cycling processes on a global scale (Caumette et al., 2015). The variety in physiological pathways has been hypothesised to arise from the vast taxonomic diversity that we observe around the globe. Microbial communities found outside of the petri-dish are inherently taxonomically diverse, with richness estimations that defy imagination (Thompson et al., 2017; Louca et al., 2019). Richness varies among ecosystems with highest taxonomic reservoirs found in soils and less so in comparably nutrient depleted aquatic ecosystems (Thompson et al., 2017). 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 (Shade, 2017, 2018). Inherited from community ecology in microbiology, microbial assembly processes have similarly been defined into four fundamental processes - selection, dispersal, diversification and drift - (Vellend, 2010), which can be characterised by their degree of determinism and stochasticity (Zhou and Ning, 2017). Recently, a statistical approach has gained popularity, which aims to quantify each assembly process (Stegen et al., 2013, 2015) and a plethora of studies have found interesting patterns since (Langenheder et al., 2017; Graham et al., 2017; Gweon et al., 2020). While reported results have been insightful, the approach heavily relies on the assumption that taxonomic related organisms have similar functional traits (i.e. phy-

logenetic conservatism). This concept has been handled with care within the macro-biologies (Losos, 2008; Warren et al., 2008), but should be a target of further discussions in microorganisms, where species definitions remain unclear (Achtman and Wagner, 2008) and genetic exchanges are common (Thomas and Nielsen, 2005). Indeed, many microbial traits have been shown to be phylogenetically conserved (Martiny et al., 2015) but it was also noted that more complex traits encoded by many genes are more likely to be conserved while simpler traits that involve fewer genes are rather phylogenetically dispersed (Martiny et al., 2013). As little as one mutation can change a microbe's phenotype (?), which exemplifies the extreme scenario of within-species versatility that may lead to niche segregation (Van Rossum et al., 2020). To complicate matters, phenotypic variation may arise from the same genotype due to differences in expression levels (e.g. epigenetics) (Caumette et al., 2015). As such, minuscule variation in genes or expression may have disproportionate effects on microbial phenotypes and thus, patterns based on phylogenetic conservatism shall be interpreted with care (Achtman and Wagner, 2008; Ackermann, 2015; Chase et al., 2018).

In concert with the complexity within microorganisms, ecosystems themselves are highly dynamic, complex and diverse. Most field studies, investigate the intriguing seasonal and spatial fluctuations of community composition within individual terrestrial, freshwater and marine ecosystems separately (hereafter ecosystem domains) (Shigyo et al., 2019; Jones et al., 2012; Hassell et al., 2018; Giovannoni and Vergin, 2012). Even among freshwater studies, researchers tend to focus on either streams,

lakes, or rivers (Logue and Lindström, 2008). While this approach gives an in-detail insight into ecosystem specific community dynamics, it neglects an inherent characteristic of nature. A characteristic so obvious but easily neglected - the connectivity beyond ecosystem boundaries.

Literature is scarce when searching for studies that cover multiple ecosystem domains within a study (Nemergut et al., 2011; Shade et al., 2013) and it is even rarer to investigate multiple ecosystem domains that are physically connected (Ruiz-González et al., 2015; Hauptmann et al., 2016; Doherty et al., 2017; Gweon et al., 2020). Microorganisms without motility machineries are seemingly stationary, although being dispersal limited to a certain degree (Hanson et al., 2012), 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 ocean. Especially during strong 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 reservoirs that temporarily give more time for some organisms to thrive and previously unavailable resources may be degraded with longer hydraulic residence times (Catalán et al., 2016). Within the aquatic network, dynamics in hydrology have been determined to be one of the major drivers of microbial community composition (Niño-García et al., 2016) and evidence suggests that a large proportion of aquatic microbial communities can be traced back to soils (Crump et al., 2012; Besemer et al., 2013; Ruiz-González et al., 2015;

Hauptmann et al., 2016). Different branches of the aquatic network collect taxa from all sub-catchments within a watershed, potentially creating sub-watershed specific unique communities at each sub-catchment's mouth. Thus, not every member within diverse communities at a single location shares a similar history in terms of origin (Niño-García et al., 2016; Comte et al., 2017) or reacts similarly to changing environmental conditions (Fierer et al., 2007). However, such passive dispersal implies that occurrence of a taxon in an unsuitable habitat merely by accident is also likely.

Death and dormancy (Cole, 1999; Jones and Lennon, 2010) questions the suitability of DNA approaches to study ecological 'communities'. In a strict ecological sense, a community refers to an assemblage of various species that interact with each other, sharing niches in forms of available space or resources (Konopka, 2009). 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 (Blazewicz et al., 2013). Evidence of 16S rRNA has opened windows to study differential responses of the potentially active community to environmental variables (?) and available resources (Osterholz et al., 2016). Being target of scrutiny and criticism, RNA may not be a good indicator of absolute growth or activity rates (Blazewicz et al., 2013), 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.

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 bulk DNA to the reactive RNA assemblages, respectively. Therefore, soils, soil-water, groundwater, streams, the river, lakes, reservoirs and the estuary were sampled for the 16S rRNA gene (DNA) and the 16S rRNA (RNA) to characterise 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 hypothesised 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 headwater streams to represent "ecotones" between terrestrial and larger aquatic water bodies due to their high connectivity to the surrounding terrestrial environment. Thirdly, we hypothesised 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 ex-

plore 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 km 2 . For detailed catchment characteristics please refer to the supplementary methods (hereafter, SM) (SM1). In brief, the river flows through a series of larger headwater lakes (hereafter, riverine lakes), and subsequently flows mainly south passing through three dams that were consecutively built in 2015 (RO2), 2016 (RO1), and 2017 (RO3). 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 and 202 for RNA, covering spring (166:69, DNA:RNA), summer (195:99) and autumn (34:34). RNA samples were sampled from 2016 onwards. To follow a terrestrial-aquatic continuum various habitat types have been sampled (Table S1). Due to the remoteness and inaccessibility of the northern most 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 (detailed sampling methods in SM2).

Following the manufacturers instructions, commercial DNA and RNA extraction kits were used (details in SM2). RNA extracts were synthesized to cDNA and all samples were sent to Génome Québec Innovation Centre (Montréal, QC, Canada) for paired-end metabarcoding of the 16S rRNA V4 region using the primers 515F (5'-GTGCCAGCMGCCGCGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') on a 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)). All 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 also to merge potential differences among 16S rRNA copies within a single genome, ASVs were merged into OTUs by a 99% similarity threshold (Větrovský and Baldrian, 2013) with the *DECIPHER* package (Wright, 2016).

All OTU observations with less than 10 reads per sample were removed. Furthermore, *metagenomeSeq* was used to transform and stabilise 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).

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 log transformed, ($\log_2(x + 1)$) to resolve a strong horse-shoe effect. To make the distance matrix Euclidean, the D_{BC} matrix was square-root transformed (Legendre and Legendre, 1998; Borcard et al., 2011). A PERMANOVA was computed with 9999 permutations to evaluate statistical differences in habitat type and season 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):

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

where p and q represent DNA and RNA site scores, respectively, of each sample and n is the 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) (Sørensen, 1948; Legendre and Legendre, 1998). 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 a ratio between the Sørensen and Bray-Curtis distance to see when incidence based dissimilarity is greater than abundance and vice versa ($m_S:m_{BC}$). This framework allows us to understand what shifts within the community is causing DNA and RNA assemblages to diverge. When m_{BC} turns smaller than m_S (high $m_S:m_{BC}$), we have fewer shared OTUs among DNA and RNA, and abundance differences are comparably smaller (more similar in m_{BC}). In contrary, when there is a higher number of shared OTUs and abundance differences are larger, m_S will be smaller than m_{BC} (low $m_S:m_{BC}$). These interpretations were statistically tested with Kruskal-Wallis tests (Fig. S4).

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. 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 DNA abundance in each sample (Table 1). In brief, we distinguish OTUs that are present everywhere (universal [AG], 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; Wickham 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 visualise the results (Wickham, 2016; Kassambara, 2020a; Campitelli, 2020; Wilke, 2019). For statistical analyses, *vegan* and *rstatix* were used (Oksanen et al., 2019; Kassambara, 2020b). 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 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. S6). 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. S6) 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. Abiotic results (Fig. S7)

Gradual change of DNA assemblages along a terrestrial-aquatic continuum

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 14.62 % of the variance (Fig.2a). This observation was supported by the PERMANOVA analysis ($F_{12} = 17.07$, $R^2 = 0.35$, $p = 0.0001$). 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 3.61 % of the variance (PERMANOVA: $F_2 = 11.09$, $R^2 = 0.038$, $p = 0.0001$). 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. S8). Interestingly, lakes are located in general more within the summer/autumn trajectory. These lakes, however, are not part of the continuum but were sampled across the watershed and thus are rather a part of the extended meta-community. Similarly, sediment samples cluster in the summer/autumn path without a single case lying in the spring trajectory.

Although PERMANOVA results strongly supported habitat type and seasonal clustering, the results could be affected by different dispersion of data. Differential levels of dispersion was found by habitat type alone

(PERMDISP: $F_{12} = 32.32$, $p < 0.0001$), by solely season (PERMDISP: $F_2 = 42.77$, $p < 0.0001$) and by both habitat and season combined (PERMDISP: $F_{27} = 17.81$, $p < 0.0001$). Non-homogeneity of dispersion does confound a clear conclusion based on statistical evidence that habitat type and seasons are major drivers, ordination results indicate a very apparent difference in habitat type and seasonal microbial communities.

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). 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 ($F_1 = 1.67$, $p = 0.19$).

Overall, the PCoA's first three axes indicate that there is a very substantial seasonality in freshwater samples in both DNA and RNA. A similar seasonal and nucleic acid type dissimilarity to freshwater samples cannot be observed in terrestrially influenced samples (Fig. S8) and the strong seasonality in aquatic samples is captured in a PCoA only with RNA samples as well (Fig. S8). Visual inspection indicated that a lot of the ter-

restrially influenced samples' dissimilarity was indeed split upon additional axes, a clearer separation of soil and soilwater samples for example was observed in axes 6 (1.75%), 10 (1.12%), 11 (1.06%) and 12 (0.99%)(data not shown). However, apparent seasonality and large differences in nucleic acid type was not observed.

DNA and RNA mismatch patterns along the continuum

To further explore the patterns in DNA-RNA differences along the continuum, we calculated the distance in ordination space between DNA and RNA of each sample (m_{BC}) along $n_{75\%}$ axes as the depicted axes in Fig. 3 only capture aquatic DNA-RNA divergence. In spring, we can observe the largest differences in DNA-RNA in soilwater, followed by a gradual convergence between DNA and RNA along the continuum (Fig. 4a). This pattern is not reflected in summer and autumn, where the overall dissimilarity between DNA and RNA remains relatively similar across ecosystems. It is noteworthy, that outside of the direct continuum, tributaries and the riverine lakes located closer to the headwaters have a similar range in DNA-RNA divergence to soilwaters in spring. This DNA-RNA divergence in tributaries and riverine lakes cannot be observed in other seasons.

In order to focus on solely species identity differences between DNA and RNA, we further examined the distance in PCoA ordination computed with the Sørensen dissimilarity (m_S), which is solely based on incidences (i.e. presence-absence). m_S indicates that the fraction of shared taxa between DNA and RNA increases along the continuum with the highest discrepancies in soilwater and streams (Fig. 4b). The decrease in m_S

along the continuum is more steep compared to m_{BC} . Similarly to m_{BC} , spring tributary and riverine lakes seem to exhibit similar magnitudes of DNA-RNA discrepancies.

In order to examine, when dissimilarities are driven more strongly by a shift in shared taxa identities or abundance discrepancies, we compared m_{BC} with m_S . When examining the $m_S:m_{BC}$ ratio along the continuum we can observe a gradual decline in $m_S:m_{BC}$ from soilwaters to the upriver indicating that there are in general fewer shared OTUs among DNA and RNA, and furthermore a higher similarity in DNA and RNA abundance. Subsequently, starting with the upriver summer, the following downstream ecosystems have a tendency to converge towards 1 and even fall below the 1:1 threshold. Within the extended meta-community, most ecosystems have a tendency of higher $m_S:m_{BC}$ ratios, with only riverine lakes in summer and other watershed lakes in autumn to have a small $m_S:m_{BC}$.

Community shifts along the rank abundance curve

In order to understand where within a rank abundance curve, the observed community shifts are happening we calculated the DNA and RNA abundance difference and examined their relationship to m_{BC} (Fig. 5a). It seems that in general community reshuffling is happening across all abundance groups, while 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. Interestingly, 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 extredest 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, which contributes to the divergence of DNA and RNA assemblages. Additionally, shifters and cosmopolitans steadily contribute to drifting DNA and RNA abundances.

When examining the optima of each OTU within the first three PCoA axes (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

Combine DNA and RNA results in one section

Microbial journey from headwaters to the estuary

Terrestrial-aquatic

Within aquatic

Residence time Reservoir

Estuary

Taxa

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 Robert Ptacnik 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

- Achtman, M. and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6(6):431–440.
- Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, 13(8):497–508.
- 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.
- Blazewicz, S. J., Barnard, R. L., Daly, R. A., and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities : limitations and uses. *The ISME Journal*, 7:2061–2068.
- Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. Springer, New York.
- 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. *ISME Journal*, 11(12):2639–2643.
- Campitelli, E. (2020). ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'.
- Catalán, N., Marcé, R., Kothawala, D. N., and Tranvik, L. J. (2016). Organic carbon decomposition rates controlled by water retention time across inland waters. *Nature Geoscience*, 9(May):1–7.
- Caumette, P., Bertrand, J.-C., and Normand, P. (2015). *Some Historical Elements of Microbial Ecology*. Springer Science and Business Media, Dordrecht.

- Chase, A. B., Gomez-Lunar, Z., Lopez, A. E., Li, J., Allison, S. D., Martiny, A. C., and Martiny, J. B. (2018). Emergence of soil bacterial ecotypes along a climate gradient. *Environmental Microbiology*, 20(11):4112–4126.
- Cole, J. J. (1999). Aquatic microbiology for ecosystem scientists: New and recycled paradigms in ecological microbiology. *Ecosystems*, 2:215–225.
- Comte, J., Berga, M., Severin, I., Logue, J. B., and Lindström, E. S. (2017). Contribution of different bacterial dispersal sources to lakes: Population and community effects in different seasons. *Environmental Microbiology*, 19(6):2391–2404.
- 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(9):1629–1639.
- 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’.
- Fierer, N., Bradford, M. A., and Jackson, R. B. (2007). Toward an Ecological Classification of Soil Bacteria. *Ecology*, 88(6):1354–1364.
- Giovannoni, S. J. and Vergin, K. L. (2012). Seasonality in ocean microbial communities. *Science*, 335(6069):671–676.
- Graham, E. B., Crump, A. R., Resch, C. T., Fansler, S., Arntzen, E., Kennedy, D. W., Fredrickson, J. K., and Stegen, J. C. (2017). Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. *Environmental Microbiology*, 19(4):1552–1567.
- 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*.
- Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10:497–506.

- 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., Cadkin, T. A., Newton, R. J., and McMahon, K. D. (2012). Spatial and temporal scales of aquatic bacterial beta diversity. *Frontiers in Microbiology*, 3(318):1–10.
- 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.
- Kassambara, A. (2020a). ggpubr: 'ggplot2' Based Publication Ready Plots.
- Kassambara, A. (2020b). rstatix: Pipe-Friendly Framework for Basic Statistical Tests.
- Konopka, A. (2009). What is microbial community ecology ? *The ISME Journal*, 3:1223–1230.
- 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 Legendre, L. (1998). *Numerical Ecology*. Elsevier, Amsterdam, 2nd edition.
- Logue, J. B. and Lindström, E. S. (2008). Biogeography of bacterioplankton in inland waters. *Freshwater Reviews*, 1:99–114.
- Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11(10):995–1003.

- Louca, S., Mazel, F., Doebeli, M., and Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS biology*, 17(2):e3000106.
- Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12.
- Martiny, A. C., Treseder, K., and Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, 7:830–838.
- Martiny, J. B. H., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science*, 350(6261):aac9323.
- 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.
- 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.
- 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. (2016). Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal*, 10:1755–1766.
- 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., Niggemann, J., and Dittmar, T. (2016). Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *The ISME Journal*, 10:1717–1730.

- 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.
- 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.
- 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. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecology Letters*, 18:1198–1206.
- Shade, A. (2017). Diversity is the question, not the answer. *ISME Journal*, 11:1–6.
- Shade, A. (2018). Understanding Microbiome Stability in a Changing World. *mSystems*, 3(e00157-17):1–4.
- 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. *ISME Journal*, 7(8):1493–1506.
- 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.
- Stegen, J. C., Lin, X., Fredrickson, J. K., and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology*, 6(370):1–15.
- Tabak, J. (2004). Differential Geometry. In *Geomtry: The language of Space and Form*, page 150. Facts on File, Inc, New York.
- Thomas, C. M. and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3(9):711–721.
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Xu, Z. Z., Jiang, L., Haroon, M. F., Kanbar, J., Zhu, Q., Song, S. J., Koscioletk, T., Bokulich, N. A., Lefler, J., Brislawn, C. J., Humphrey, G., Owens, S. M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J. A., Clauset, A., Stevens, R. L., Shade, A., Pollard, K. S., Goodwin, K. D., Jansson, J. K., Gilbert, J. A., Knight, R., and The Earth Microbiome Project Consortium (2017). A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature*, 551(7681):457–463.
- Van Rossum, T., Ferretti, P., Maistrenko, O. M., and Bork, P. (2020). Diversity within species: interpreting strains in microbiomes. *Nature Reviews Microbiology*, 18:491–506.
- 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.
- Warren, D. L., Glor, R. E., and Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62(11):2868–2883.

- 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., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Wilke, C. O. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.
- Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, 8(1):352–359.
- 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. Habitat types are indicated in colours, and shapes differentiate seasons. Percentage of variance explained are given in square brackets for the first and second axis.

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. Seasons are given in different point shapes, colouring visualises different habitat types, opacity indicates nucleic acid type. Percentage of variance explained are given in brackets for the first, second and third axes.

Figure 4. 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. Seasons are given in different point shapes, colouring visualises different habitat types, opacity indicates nucleic acid type. Percentage of variance explained by the corresponding axes are given in square brackets.

Figure. 5 Distance between DNA and RNA samples within PCoA ordination space reveals sections within the continuum that are dominated by species sorting or mass effects. a) Distance based on Bray-

Curtis dissimilarity (m_{BC}) among DNA and RNA assemblages of the same sample within the axes capturing 75 % of the variance in PCoA. b) Ratio between distance calculated based on PCoAs computed with Sørensen (m_S) and Bray-Curtis dissimilarity (m_{BC}). Points indicate the arithmetic mean and error bars represent standard deviations from the mean. Sample sizes of each point are indicated above in the corresponding colours. Sample sizes are equivalent in panels a and b.

Figure 6. DNA and RNA dissimilarities arise through abundance differences across the rank abundance curve. a) Difference in DNA and RNA abundance against m_{BC} . 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.

Tables

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

Figures

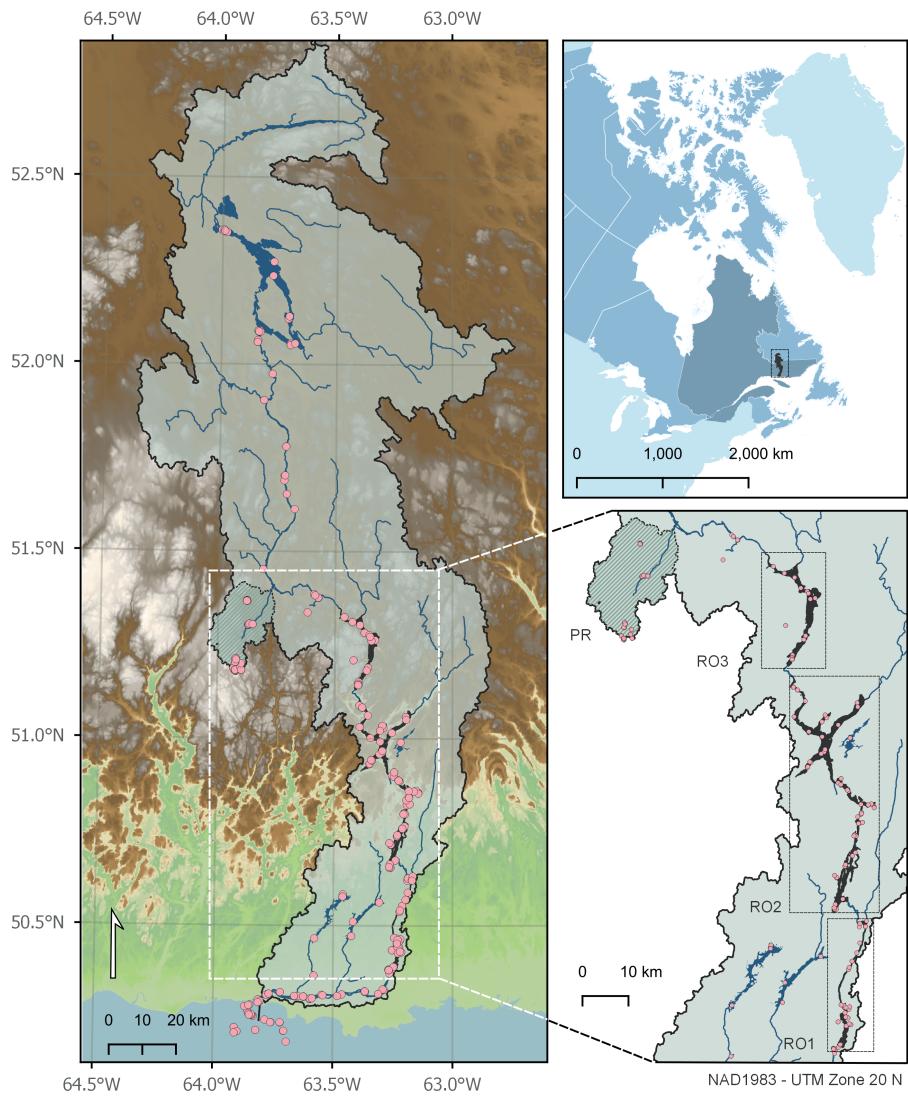


Figure 1: Location and overview of the La Romaine catchment. a) Scale and overview of the whole La Romaine catchment. Samples are represented as points. b) Location of the catchment within Canada and Québec. c) Focus on all built reservoirs RO1 (2015), RO2 (2014) and RO3 (2017) and the headwater stream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

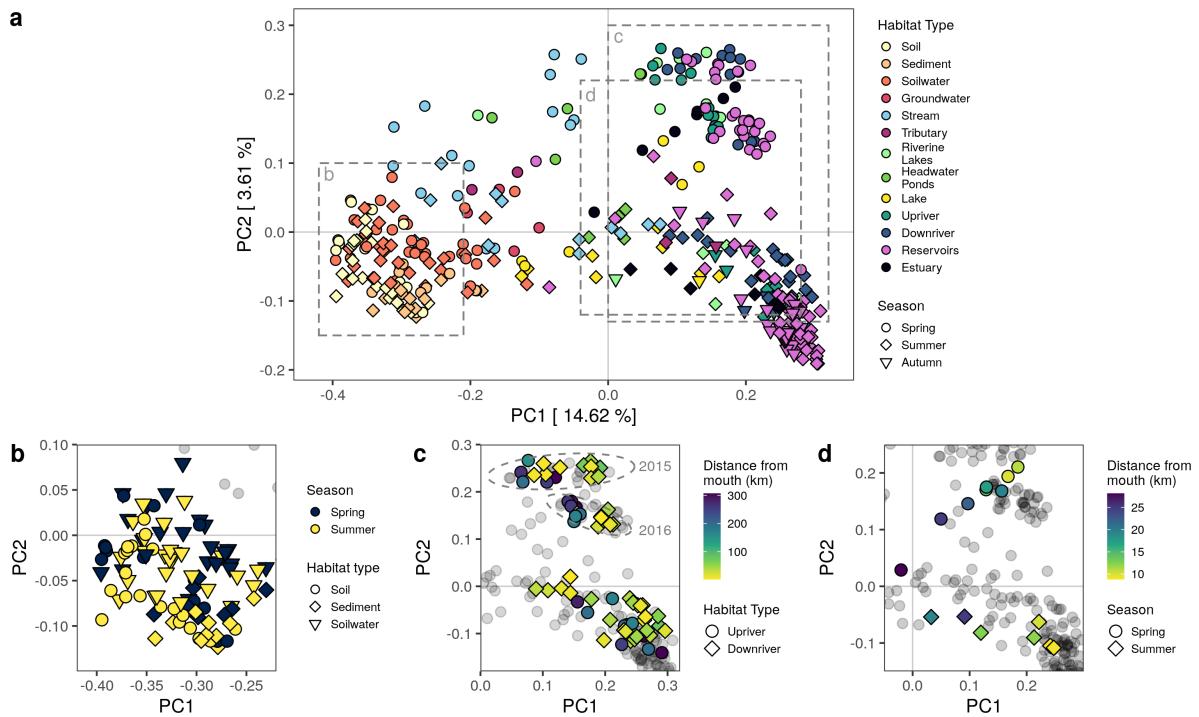


Figure 2: Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons.
 Overall PCoA analysis of DNA samples (a) has been further explored with focus plots on terrestrial, riverine and estuary samples in panels b, c, d, respectively. Overall,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

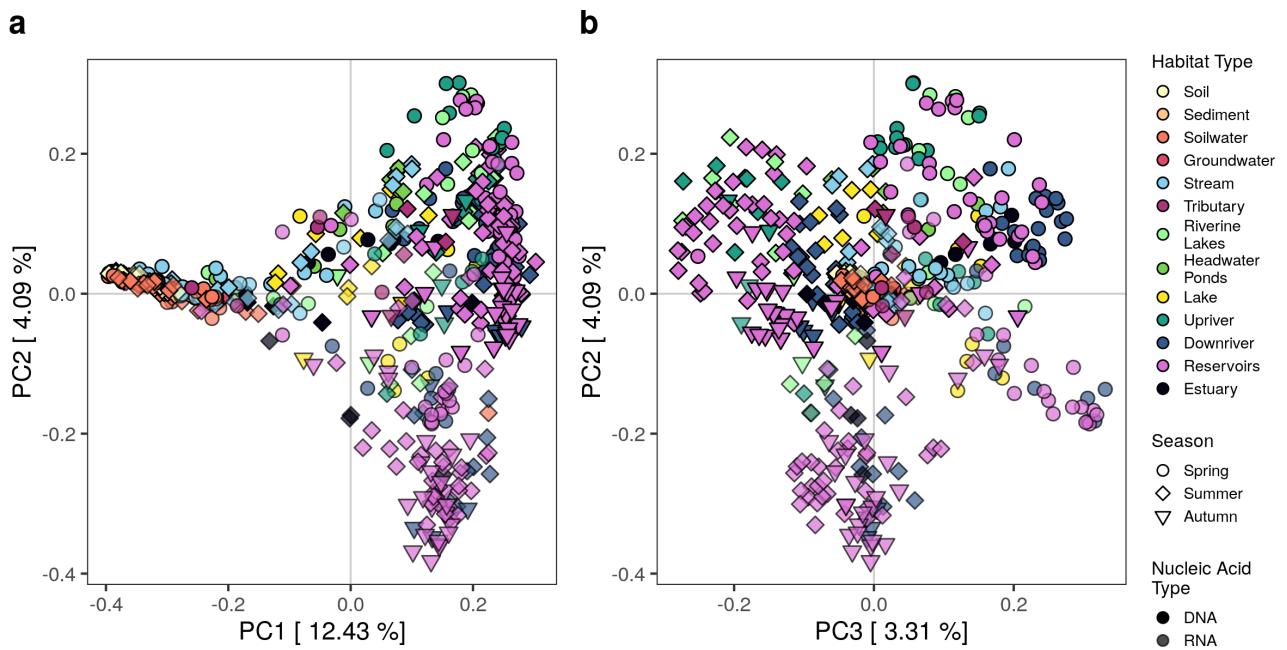


Figure 3: RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially influenced habitats. PCoA analysis including RNA samples. a) Visualisation of first and second axis of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis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

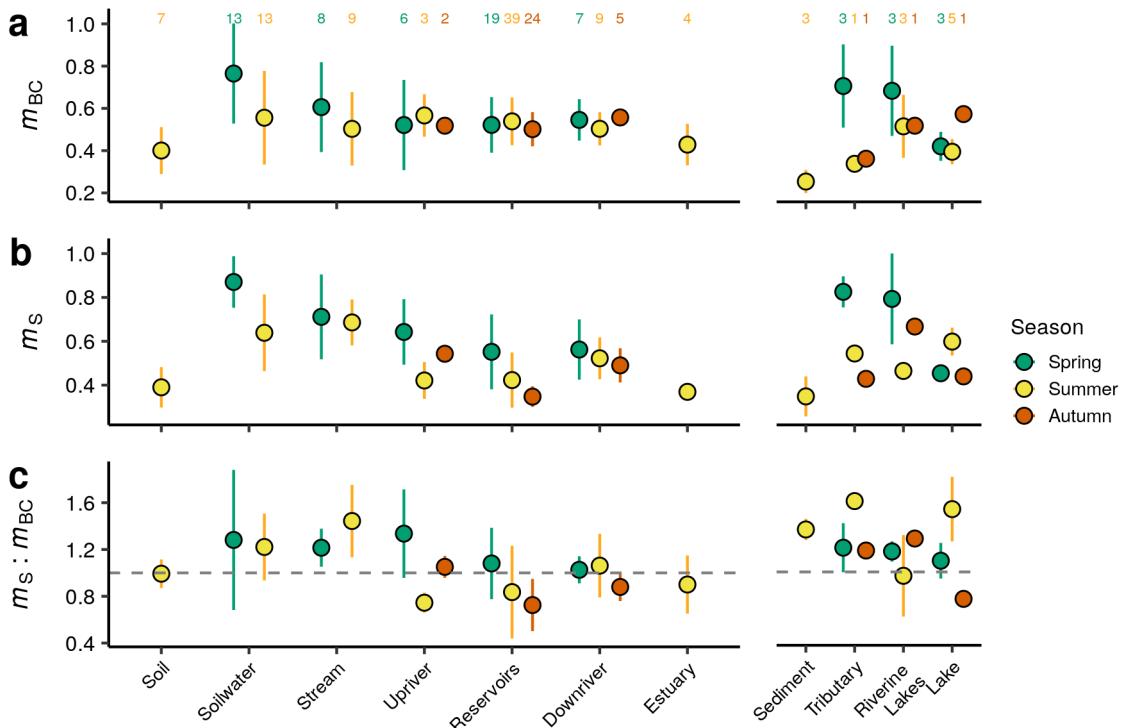


Figure 4: Distance between DNA and RNA samples within PCoA ordination space reveals sections within the continuum that are dominated either by species sorting or mass effects. a) Distance based on Bray-Curtis dissimilarity (m_{BC}) among DNA and RNA assemblages of the same sample within the axes capturing 75 % of the variance in PCoA. b) Ratio between distance calculated based on PCoAs computed with Sørensen (m_S) and Bray-Curtis dissimilarity (m_{BC}). Points indicate the arithmetic mean and error bars represent standard deviations from the mean. Sample sizes of each point are indicated above in the corresponding colours. Sample sizes are equivalent in panels a and b.

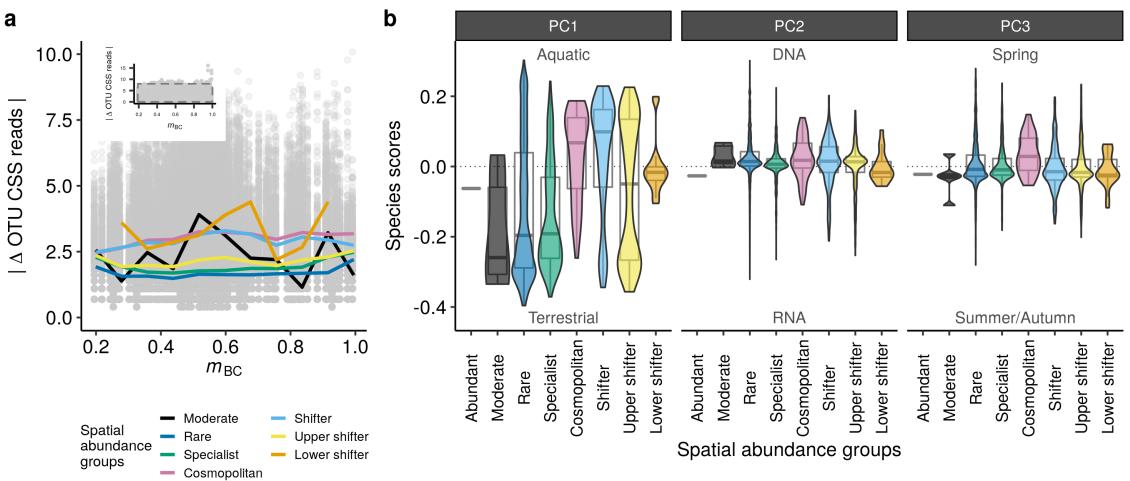


Figure 5: DNA and RNA dissimilarities arise through abundance differences across the rank abundance curve. a) Difference in DNA and RNA abundance against m_{BC} . Lines represent rolling means with bin size = 10. b) Boxplots and violin plots showing the distribution of species scores and thus OTU optima 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.