

Title (max 150 char): Diverging DNA and RNA communities along a boreal terrestrial-hydrological 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, terrestrial-aquatic continuum, ecosystem connectivity, residence time, boreal ecosystems, mass effects, environmental sorting, metacommunity, rare biosphere

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

(The Introduction should assume that the reader is knowledgeable in the field and should therefore be as brief as possible but can include a short historical review where desirable.)

Tralala

The consecutive flooding of the reservoirs creates a unique opportunity to study the transformation process of a running system into a standing water body. The shifts mainly imposed due to the increased residence times are likely to influence not only biogeochemical properties but also their microbial assemblages.

Material and methods

Catchment characteristics and sampling design

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 . The catchment was glaciated 7 000 – 10 000 years ago and left mostly a till blanket and veneer as surficial material. It is mainly dominated by acid rocks (e.g. granodiorite, granite, quartz diorite) with granitised sedimentary and volcanic rock, and has isolated patches of permafrost (0-10%)(Natural Resources Canada). The soil is composed of roughly 61.4% sand, 31.9% silt, 6.7% clay and

stores approximately 140.4 t ha^{-1} of organic carbon (in top 5 cm; given are catchment averages)(Lehner and Grill, 2013; Hengl et al., 2014).

The river springs between the Atlantic and Saint Lawrence watersheds ($52^{\circ}52'20''\text{N}$ $63^{\circ}36'55''\text{W}$; 702 masl), and consequently flows through a series of lakes (hereafter riverine lakes) including the biggest lake in the catchment – Lake Brûlé ($A: 127.11 \text{ km}^2$, elevation: 470 masl). The river mainly flows towards the South with a maximum distance from the northern headwaters to the river mouth expanding to approximately 475.1 km. Nearly half of the catchment is covered by coniferous forests (*P.mariana*-moss), with mixed forests being rather minor (11%) and deciduous stands with white birch (*Betula papyrifera*) and trembling aspen (*Populus tremuloides*) are even more rare (2%)(d'audiences publiques sur l'environnement and Canadian Environmental Assessment Agency, 2009). The northern part of the catchment is characterised by a flat open black spruce (*Picea mariana*)-lichen forest with shrubs and moss-lichen (Fig. S1a). As one follows the river downstream, the relief changes drastically to a steep mountainous stretch that forms sections of canyons (Fig. S1b). The river loses 330 m of elevation from the mountainous section until it makes a sharp turn to the west into the lower coastal plain. The coastal plain is characterised by peatland areas with swamps and shallow waters that are completely permafrost free (Fig. S1c). There are two larger tributaries in the coastal plain that flow through the lakes Puyjalon ($A: 13.10 \text{ km}^2$) and Allard ($A: 19.24 \text{ km}^2$). A weather station located in the lower coastal plain ($50^{\circ} 16'55.000'' \text{ N}$, $63^{\circ} 36'41.000'' \text{ W}$, Havre-Saint-Pierre Airport, Natural Resources Canada) recorded an annual precipitation of 810.77 ± 35.25

mm and 1.18 ± 0.73 °C, -32.63 ± 1.36 °C, and 25.8 ± 0.66 °C for mean, minimum and maximum temperature over the sampled years.

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

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. 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 have 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.

Surface soil samples were collected by mixing three randomly selected cores (30 cm) that were taken in proximity of installed piezometers to sample soil-water. 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 randomly installed piezometers were pooled together. Groundwater was directly collected from constructed wells with submersible pumps. 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. 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.

Sample processing and sequencing

Homogenised soil and sediment samples were transferred to aliquots of 0.25 g and frozen immediately on the same day of sampling. A minimum of 25 mL and 250 mL of soil-/hyporheic-water and surface water was filtered through 0.22 µm polycarbonate membrane filters (Merck Millipore, Darmstadt, Germany), respectively, and subsequently stored frozen. All DNA and RNA samples were frozen at -20 °C at the field station and further stored at -80 °C at the university laboratory until extraction. Samples for RNA extraction were submerged in RNAlater® and LifeGuard® Soil Preservation solution (QIAGEN®, Hilden, Germany) for water and humic samples (soil, soilwater, hyporheic water), respectively. To allow stabilisation, samples were left at 4 °C overnight and were subsequently stored frozen.

Following the manufacturers instructions, PowerWater® and PowerSoil® DNA and RNA extraction kits (MoBio, Carlsbad, CA, USA) were used to extract water and soil/soil-/hyporheic-water/sediment samples, respec-

tively. In 2017, the equivalent DNeasy® and RNeasy® PowerWater® Kits (QIAGEN®) were used for DNA and RNA samples, respectively. cDNA was synthesised from the extracted RNA with a high capacity cDNA Reverse Transcription Kit (Applied Biosystems™, Foster City, CA, USA). Successful extraction was evaluated via PCR amplification of the 515F-806R primers (IDT Technologies, Coralville, IA, USA) and DNA concentration was measured with a NanoDrop 2000c (Thermo Fisher Scientific Inc., Waltham, MA, USA). Extracts 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

Primers were removed using the software *cutadapt* (Version 1.18, Martin (2013)), which allows for the removal of the primer sequence and its variants in their true and complement orientations. Additionally, all reads shorter than 125 nucleotides were removed as they cannot achieve a minimum overlap necessary for paired-end merging in downstream processing.

To identify amplicon sequence variants (ASVs), 16S rRNA amplicon reads were analysed through the DADA2 (Divisive Amplicon Denoising Algorithm 2) pipeline (Version 1.14.1, Callahan et al. (2017)) on R Version 3.6.3 (R Core Team, 2020)). Read qualities were evaluated for each sequencing plate separately and read length was trimmed according to their

quality scores. The most conservative trimming criteria still allowed for an overlap of forward and reverse reads by 60 bp. Samples were pooled by plate, season and sequencing depth for learning the error rates. DADA2 runs on a sample by sample basis, and thus removes observed singletons by sample to avoid inclusion of false-positive sequencing errors. To retain more rare taxa within a sampling campaign (year-season combinations) along the continuum, samples were 'pseudo'-pooled for the *dada()* step. This step enables the removal of singletons by pool but retains singletons within a sample. Paired-ends were merged after successful inference of amplicon variants. Subsequently, chimeras were removed (*removeBimer-aDenovo()*) and all ASVs that are identical in sequence but differ only by length were merged together with *collapseNoMismatch()*, leading to ASVs that represent a unit similar to 100% similarity operational taxonomix units. Finally, 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)).

All ASV 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). For alpha diversity estimations, CSS results were rounded to its integer to represent count data necessary for alpha-diversity calculations. CSS results were compared with results achieved with various rarefaction thresholds. To account for random sampling effects, rarefaction was conducted with 100 random iterations. Similarly to the CSS data treat-

ments, the mean of all iterations was rounded to its integer to be used for subsequent analysis. There were no substantial differences in the results between CSS and various rarefaction thresholds (Fig. S2). Consequently, CSS results were used for downstream analyses for consistency.

Data handling and statistical analyses

To explore differences in microbial community composition across habitat types, a Principal Coordinates Analysis (PCoA) was conducted with Bray-Curtis dissimilarities on all DNA samples. Prior to the PCoA analysis, the community matrix was log transformed, ($\log_2(x + 1)$) as no transformation resulted in a strong horse-shoe effect (Fig. S3). The Bray-Curtis dissimilarity matrix was square-root transformed to approximate Euclidean distances (Legendre and Legendre, 1998; Borcard et al., 2011). A PERMANOVA was computed with 999 permutations to evaluate statistical differences in habitat type and season.

Secondly, to evaluate whether sampled RNA communities were further different from the DNA assemblages, we computed a second PCoA (Bray-Curtis dissimilarities with square-root transformation) with both DNA and RNA samples. Again, statistically different clusters were investigated with a PERMANOVA (999 permutations), where habitat type, season and nucleic acid type (DNA-RNA) were independent factors.

PCoA distances....

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). Distance, diversity metrics and PERMANOVA were calculated using *vegan*, and *ape* was used to compute the PCoA (Oksanen et al., 2019; Paradis and Schliep, 2018). *ggplot2*, *ggnpubr* and *ggnewscale* were used to visualise the results (Wickham, 2016; Kasambara, 2020; Campitelli, 2020). All analyses have been conducted in R (R Core Team, 2020) and RStudio (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) and the Spatial Analyst Toolbox.

Results

Discussion

Data accessibility

DNA and RNA sequences have been archived (database name and number). The code and meta data to produce this manuscript are available at <https://github.com/masumistadler/xxx>. (Links will be updated and published 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 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 and Mario Muscarella for insightful discussions on the results. 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

- Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. Springer, New York.
- 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'.
- d'audiences publiques sur l'environnement, B. and Canadian Environmental Assessment Agency (2009). Romaine Hydroelectric Complex Project: Investigation and Public Hearing Report. Technical report.
- Dowle, M. and Srinivasan, A. (2019). data.table: Extension of 'data.frame'.
- Hengl, T., de Jesus, J., MacMillan, R., Batjes, N., Heuvelink, G., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J., Walsh, M., and Gonzalez, M. (2014). SoilGrids1km—global soil information based on automated mapping. *PLoS ONE*, 9(8):e105992.
- Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots.
- Legendre, P. and Legendre, L. (1998). *Numerical Ecology*. Elsevier, Amsterdam, 2nd edition.
- Lehner, B. and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15):2171–2186.
- Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12.
- 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.

- 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.
- 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.
- 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.
- Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, 8(1):352–359.

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. PCoA analysis reveals microbial community shifts from terrestrial to freshwater samples. Spring and summer-autumn show distinct paths in multivariate space. Seasons are given in different point shapes, colouring visualises different habitat types. Percentage of variance explained are given in 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. 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.

Tables

Figures

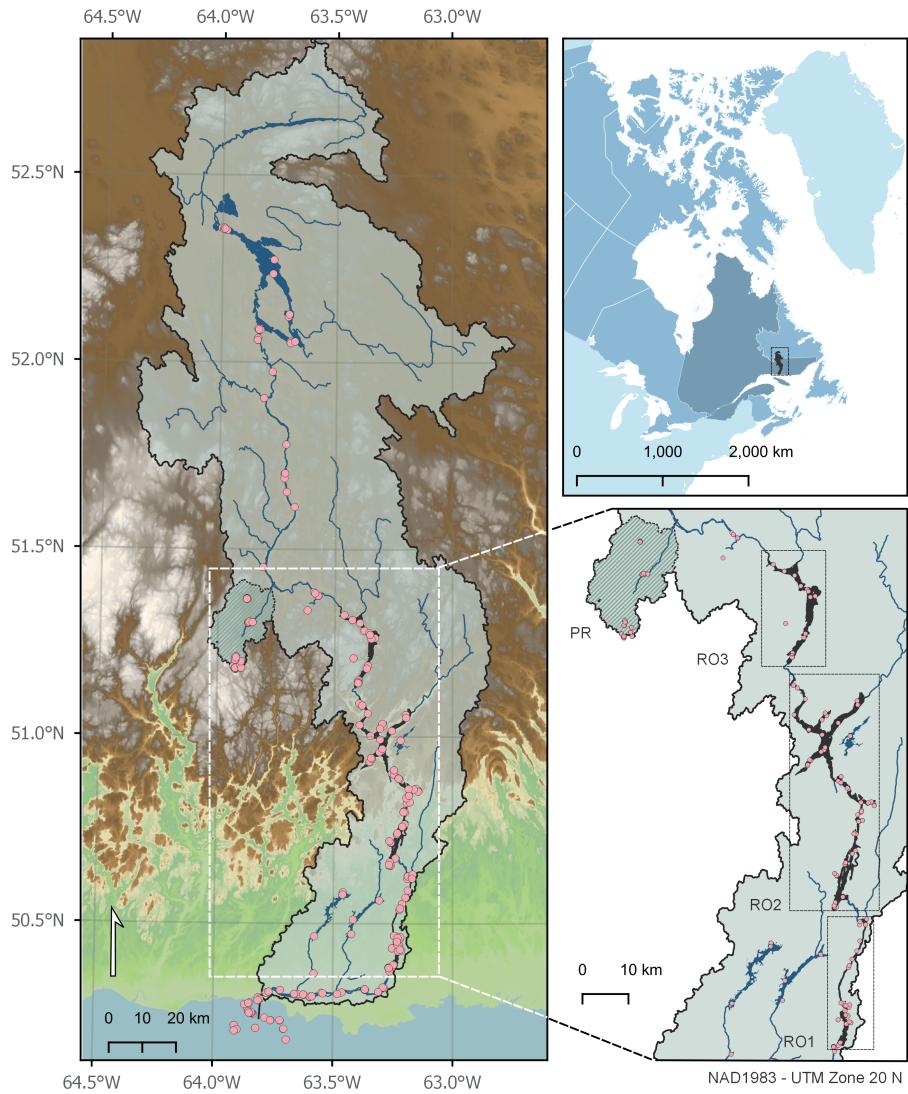


Figure 1: Location and overview of the La Romaine catchment. a) Scale and overview of the whole La Romaine catchment. Samples are represented as points. b) Location of the catchment within Canada and Québec. c) Focus on all built reservoirs RO1 (2015), RO2 (2014) and RO3 (2017) and the headwater stream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

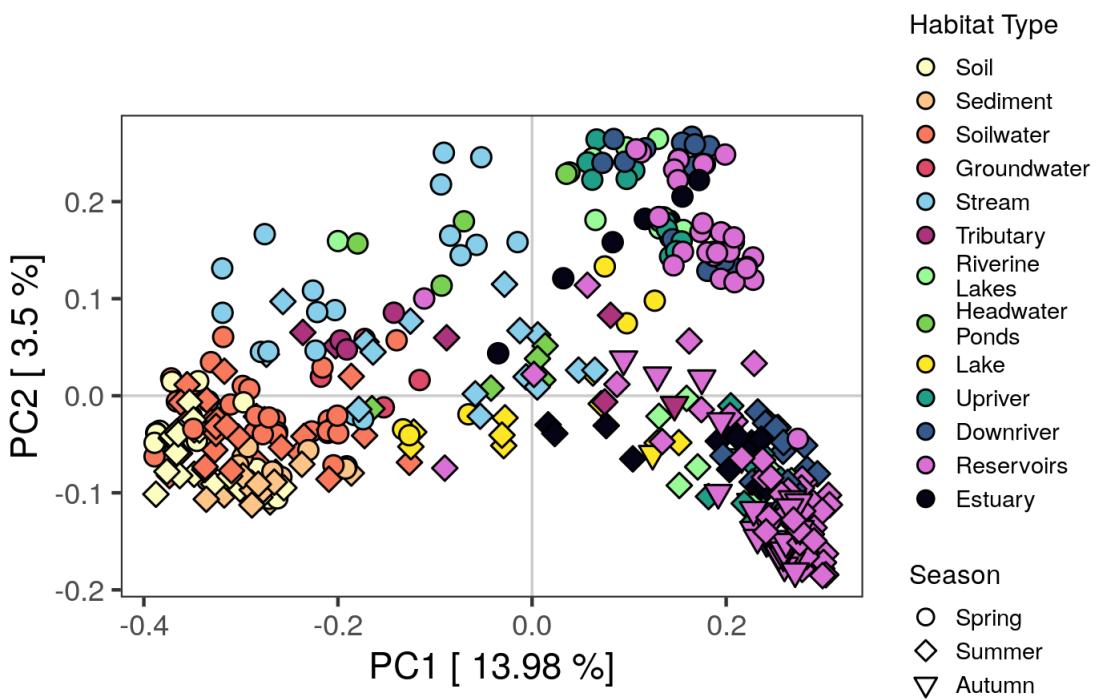


Figure 2: Microbial community composition gradually changes along a terrestrial-hydrological continuum and diverges between seasons. PCoA analysis reveals microbial community shifts from terrestrial to freshwater samples. Spring and summer-autumn show distinct paths in multivariate space. Seasons are given in different point shapes, colouring visualises different habitat types. Percentage of variance explained are given in brackets for the first and second axis.

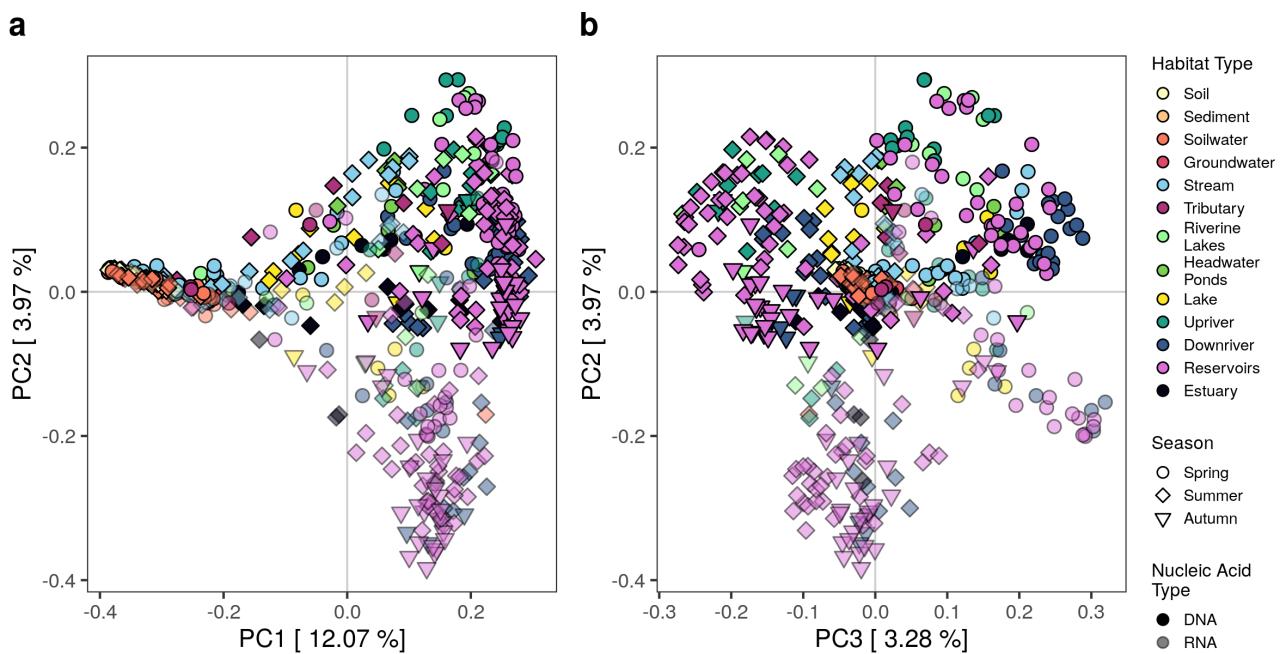


Figure 3: RNA assemblages diverge from DNA within aquatic habitats, less so in terrestrially influenced habitats. PCoA analysis including RNA samples. a) Visualisation of first and second axis of PCoA, differentiating habitat type and nucleic acid type, respectively. b) Different view on PCoA analysis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