

# **Supplement to *Water tracing with environmental DNA in a high-Alpine catchment***

Elvira Mächler<sup>1, 2</sup>, Anham Salyani<sup>3</sup>, Jean-Claude Walser<sup>4</sup>, Annegret Larsen<sup>3, 5</sup>, Bettina Schaeffli<sup>3, 6</sup>, Florian Altermatt<sup>1, 2</sup>, and Natalie Ceperley<sup>3</sup>

<sup>1</sup>Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

<sup>2</sup>Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

<sup>3</sup>Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, University of Lausanne, 1015 Lausanne, Switzerland

<sup>4</sup>Federal Institute of Technology (ETH), Zürich, Genetic Diversity Centre, CHN E 55 Universitätstrasse 16, 8092 Zürich, Switzerland

<sup>5</sup>Soil Geography and Landscape Group, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands

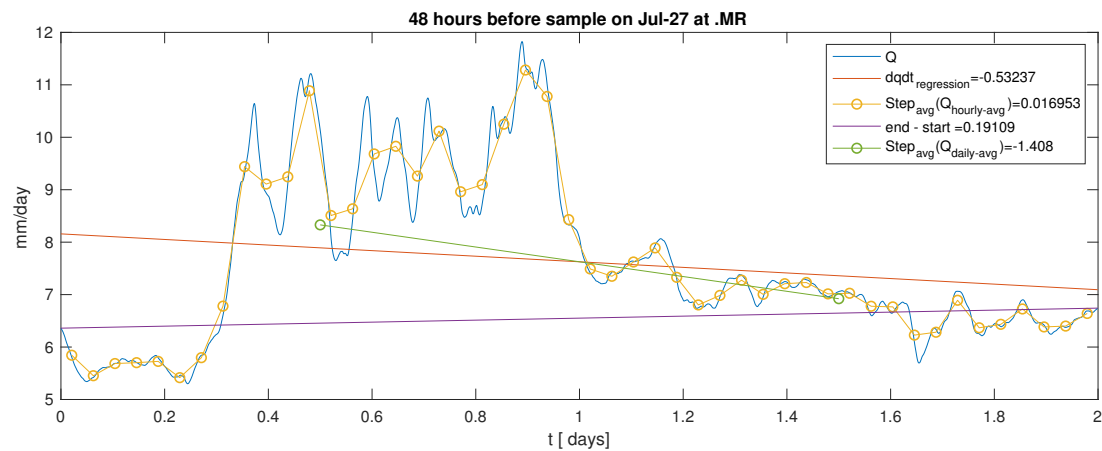
<sup>6</sup>Geography Institute, University of Bern, 3012 Bern, Switzerland

**Correspondence:** Elvira Mächler (elvira.maechler@eawag.ch)

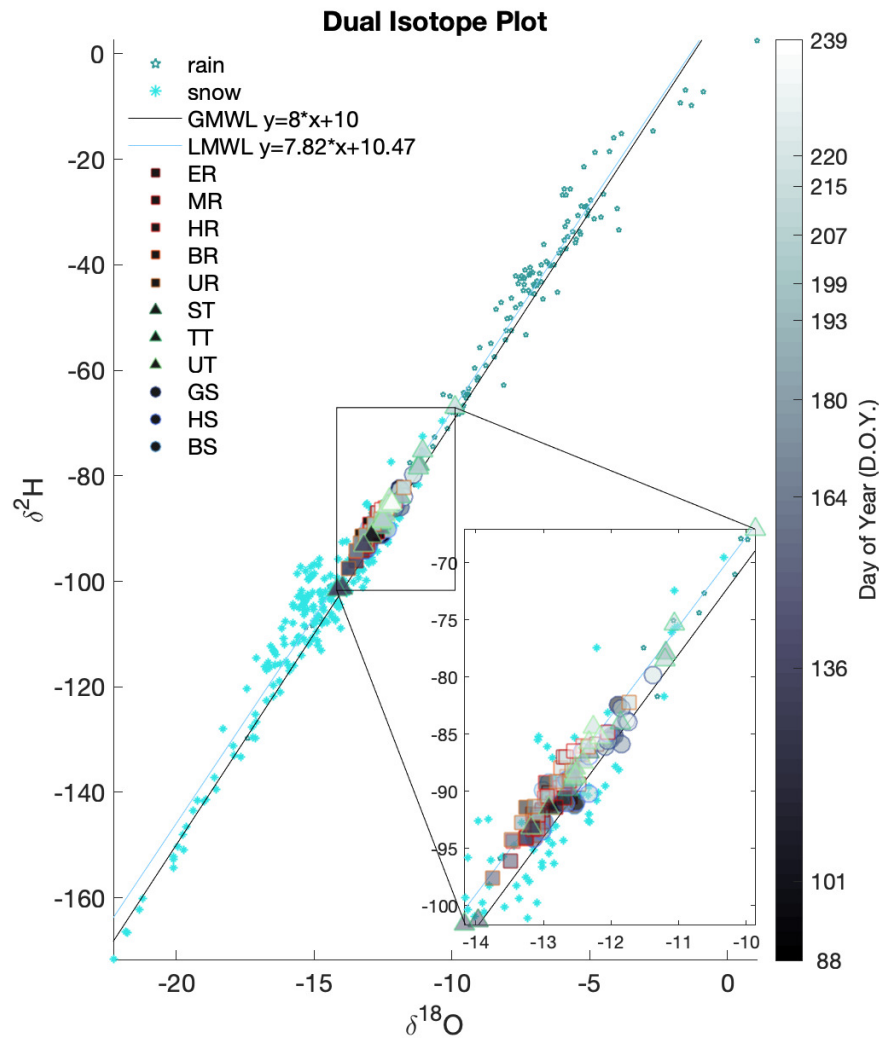
## **S0 Contents**

The supplement contains additional figures (S1), additional tables (S2), additional methods (S3), frequency thresholds for data cleaning (S4, separate file) and a species list (S5, separate file).

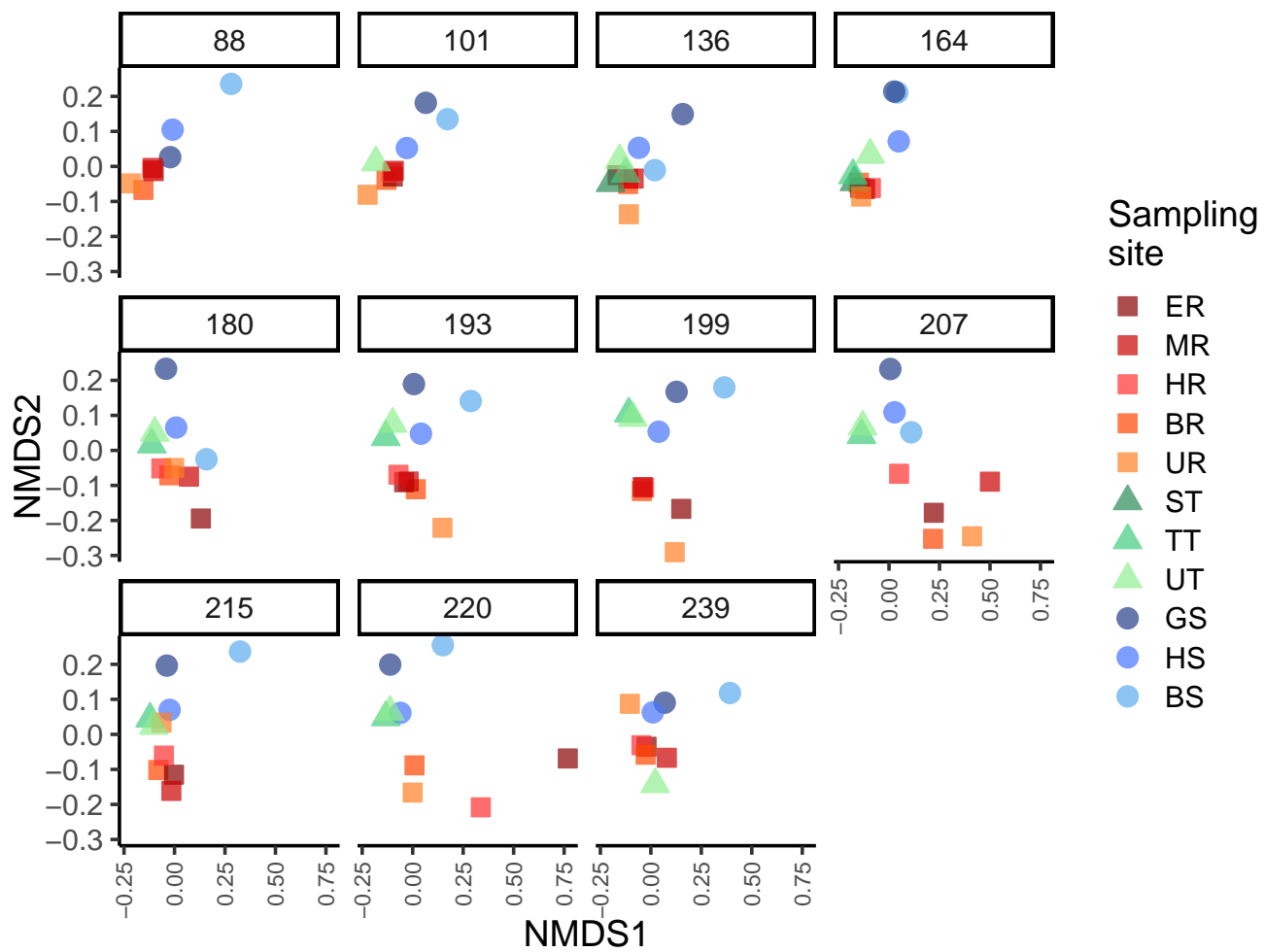
S1 Additional Figures



**Figure S1.** Example determination of  $dq/dt$  at the morning river "MR" on July 27, 2017, showing four different methods as they relate to discharge. Red line shows the regression of all measurements against time, which was retained. Yellow points show the hourly average of  $Q$  which was used to determine  $dq/dt$  by hourly time-step accordingly. Purple line shows the  $dq/dt$  as determined by the single, 2-day time step. And green dots show the daily average and the  $dq/dt$  determine by their difference.



**Figure S2.** Stable isotopes from precipitation (rain in cyan stars, snow in cyan asterisks) and eDNA sample sites (red squares are main channel, green triangles are tributaries, and blue circles are springs). Color shading shows date of sample. Local and global meteoric water lines are shown (LMWL, GMWL). The LMWL is calculated with all precipitation samples shown and used to calculate the line conditioned excess, lc-ex. A magnifying inset zooms onto all the samples.



**Figure S3.** Community composition in the main river channel, tributaries and springs over the sampling season. Facets give the day of the year.

S2 Additional Tables

Table S1. Forward and reverse primer sequence for tailed PCR reactions.

| Primer name   | Primer sequence (5'-3')  |
|---------------|--|
| mlCOIintF-FS0 | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGWACWGGWTGAACWGTWTAYCCYC     |
| mlCOIintF-FS1 | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGWACWGGWTGAACWGTWTAYCCYCC   |
| mlCOIintF-FS3 | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGGGWACWGGWTGAACWGTWTAYCCYCC |
| jgHCO2198-FS0 | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAIACYTCIGGRTGICCRAARAAYCA   |
| jgHCO2198-FS1 | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTAIACYTCIGGRTGICCRAARAAYCA  |
| jgHCO2198-FS2 | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATAIACYTCIGGRTGICCRAARAAYCA |

Table S2. Implemented data filtering steps and the remaining numbers of ZOTUs and reads in the data set.

| Filtering step   | ZOTUs | Number of reads |
|------------------|-------|-----------------|
| Raw data         | 9858  | 12,548,868      |
| Samples only     | 9760  | 9,859,536       |
| Cleaned samples  | 9693  | 9,728,558       |
| Rarefied samples | 9622  | 2,863,213       |

Table S3. Results of the GLMM model selection based on a  $\chi^2$ -test. DF = degrees of freedom, AIC = Akaike information criterion, BIC = Bayesian's information criterion, logLik = log-likelihood.

| Model                                    | DF | AIC   | BIC   | logLik  | Deviance | $\chi^2$ | $\chi^2$ DF | P-value |
|--|----|-------|-------|---------|----------|----------|-------------|---------|
| dq/dt + water source + (1 Sampling Site) | 5  | 13308 | 13322 | -6649.1 | 13298    |          |             |         |
| dq/dt * water source + (1 Sampling Site) | 7  | 11608 | 11608 | -5787.4 | 11575    | 1723.4   | 2           | < 0.001 |
| E.C. + water source + (1 Sampling Site)  | 5  | 15713 | 15727 | -7851.6 | 15703    |          |             |         |
| E.C. * water source+ (1 Sampling Site)   | 7  | 15669 | 15687 | -7827.3 | 15655    | 48.629   | 2           | < 0.001 |

**Table S4.** Generalized linear mixed effect model results for the random effects. Std. Dev = standard deviation.

| Model | Random effect             | Variance | Std. Dev. |
|-------|---------------------------|----------|-----------|
| dq/dt | Sampling site (intercept) | 0.043    | 0.208     |
| E.C.  | Sampling site (intercept) | 0.040    | 0.201     |

**Table S5.** Results of the generalized linear model selection based on a F-tests. DF = degrees of freedom.

| Model          | Residual DF | Residual Deviance | DF | Deviance  | F      | P-value |
|----------------|-------------|-------------------|----|-----------|--------|---------|
| dq/dt + origin | 105         | 1.00250           |    |           |        |         |
| dq/dt * origin | 104         | 0.90183           | 1  | 0.10067   | 11.677 | < 0.001 |
| E.C. + origin  | 105         | 1.0421            |    |           |        |         |
| E.C. * origin  | 104         | 1.0359            | 1  | 0.0062649 | 0.6468 | 0.4231  |

**Table S6.** Results for vector fitting onto NMDS ordination. NMDS1 stands for the first NMDS axis and NMDS2 for the second axis.

| Environmental variable | NMDS1  | NMDS2  | R <sup>2</sup> | P-value |
|------------------------|--------|--------|----------------|---------|
| Elevation              | -0.886 | 0.464  | 0.080          | 0.071   |
| E.C.                   | 0.152  | 0.988  | 0.421          | 0.001   |
| Temperature            | -0.927 | -0.376 | 0.208          | 0.002   |
| $\delta^{18}\text{O}$  | -0.613 | 0.790  | 0.148          | 0.015   |
| lc-ex                  | -0.027 | -1.000 | 0.173          | 0.005   |
| Total solar radiation  | 0.019  | -1.000 | 0.011          | 0.731   |
| Baseflow               | 0.820  | -0.571 | 0.003          | 0.920   |
| Snow cover area        | -0.842 | -0.540 | 0.006          | 0.872   |
| dq/dt                  | -0.777 | -0.629 | 0.003          | 0.905   |

### S3 Metabarcoding laboratory procedure

#### S3.1 Extraction of eDNA

After the collection of all the eDNA samples, we extracted the samples in a randomized order. We used the DNeasy® Blood and Tissue kit (Qiagen, Hilden, Germany) following the protocol for animal tissue besides a few changes due to the pooling of  
5 two filters per site for the extraction. We incubated the first filter with 360 µL ATL buffer and 40 µL Proteinase K for 24 hours. Afterward, we transferred the incubated filter into a new 1.5 mL tube with a hole on the bottom. This tube was then placed into the tube with the remaining incubated buffer mixture. The tube-tube compound was centrifuged for 1 minute at 6000 g in order to remove all the liquid from the filter in the upper tube. The upper tube containing the filter was then discharged. We added the second filter to the lower tube containing the buffer mixture and added another 40 µL of Proteinase K. After an incubation  
10 of 24 hours, we used the same method as described above to remove all liquid from the second filter. From this step onwards, we followed the protocol with the only change of using the doubled volumes in order to keep the volume:volume ratios equal until the liquid was pipetted on the spin columns in step 4 of the provided protocol. Finally, the eDNA was eluted in 75 µL AE buffer and stored at –20 °C until further processing.

#### S3.2 Library preparation

15 We target a 313 bp fragment of the cytochrome oxidase I (COI, Geller et al. 2013, Leray et al. 2013) for amplicon sequencing. We used an Illumina MiSeq dual-barcoded two-step PCR amplicon sequencing protocol. First, we performed a PCR with modified primers that contained an adaptor-specific tail, a heterogeneity spacer, and the amplicon target site (see Table S1). The extracted eDNA samples were randomized over four 96 well PCR plates. A single PCR reaction consisted of 1X Buffer I (ThermoFisher Scientific, Illkirch Cedex, France), BSA (0.1 mg/ µL, GeneON, Ludwigshafen am Rhein, Germany), dNTP  
20 (0.2 mM), MgCl<sub>2</sub> (1 mM), mICOIntF and jgHCO2198 primer mixes (0.16 µM each, see Table S1), AmpliTaq Gold (1.25 U/ µL) and 3 µL of extracted eDNA in a total reaction volume of 30 µL. The PCR regime consisted of 95 °C for ten minutes, followed by 50 cycles of denaturation at 95 °C for 15 seconds, annealing at 62 °C for 30 seconds and elongation at 72 °C for 45 seconds, ending the PCR with a final hold of 72 °C for 5 minutes. Amplification success was verified with the QiAexcel Screening Cartridge by using the AL420 method (Qiagen, Hilden, Germany). Samples that did not amplify were rerun with 1  
25 µL of DNA template instead of 3 µL due to previous indication of inhibited PCR. Per PCR plate we run a positive control (PC) consisting of 2 µL sample and 1 µL artificial DNA (0.01 ng/ µL, see Mächler et al., 2019).

For each eDNA sample we run five replicates of the tailed PCR reaction, pooled them and cleaned the pool with the illustra  
GFX 96 PCR Purification Kit (GE Healthcare, Glattbrugg, Switzerland). Afterwards, we performed a second PCR where we add an index to each of the samples by using the Nextera XT Index kit v2 (Illumina, Zurich, Switzerland). The indexed reactions  
30 were cleaned up with SPRI beads (Beckman Coulter, Germany). The indexed reactions were quantified with the Spark® 10M Multimode Microplate Reader (Tecan Group Ltd., Männedorf, Switzerland) and pooled them in equimolar parts into a final pool that we cleaned with SPRI beads. All controls (FC, EC, PC, NC) were run alongside the samples and were pooled according to their concentrations. Controls that were too low to quantify were pooled into the second lowest concentrated pool

with 10  $\mu$ L, equal to the volume of the lowest sample in the respective pool. The libraries were added at 16 pM concentration and PhiX control was added at a 10% 160 concentration. A paired-end (2x300 nt) sequencing was performed on an Illumina MiSeq (MiSeq Reagent kit v3, 300 cycles) following the manufacture's run protocols (Illumina, California, USA).

### **S3.3 Laboratory conditions and negative controls**

- 5 We followed the previously described measures for work with eDNA (Deiner and Altermatt, 2014; Deiner et al., 2015; Mächler et al., 2015) in order to minimize contamination. Reused field material like filter housings and syringes were soaked 40 minutes in 2.5 % sodium hypochlorite (i.e. bleach), rinsed with deionized water and treated with UV light prior to the reuse in the field. Further, we implemented three different level of negative controls: First, at the beginning of each field day we implemented a negative filter control (FC) consisting of 1 L MilliQ water that was previously treated with UV-C light and brought to the
- 10 field to check if reused material was clean. Second, we included for each batch of extractions a negative extraction control (EC) which contained a previously UV treated GF/F filter resulting in 8 extraction controls. Third, we included a negative PCR control (NC) in each PCR plate containing sigma water instead of DNA. All negative controls were run alongside the eDNA samples during the laboratory workflow.

### **S3.4 Bioinformatic data processing**

- 15 The data was demultiplexed and the quality of the reads was checked with FastQC (Andrews et al., 2010). Raw reads were end-trimmed (usearch v10.0.240, R1:30nt, R2:50nt) and merged with an overlap of min 15 bp max 300 bp (Flash, v1.2.11). Next, the primer sites were removed (full length, no mismatch allowed (cutadapt v1.12) and thereafter, the data was quality filtered (prinseq-lite v0.20.4) using the following parameters: size range (100–500), GC range (30–70), mean quality (20), and low complexity filter dust (30). In a next step, UNOISE3 (usearch v10.0.240) was used to determine amplicon sequence variants.
- 20 UNOISE3 has a build-in error-correction to reduce the influence of sequencing errors (Edgar, 2016). An additional clustering at 99% sequence identity was performed to reduce sequence diversity and to account for possible amplification errors in the first PCR resulting in ZOTUs (zero-radius OTUs, thereafter called ZOTUs). This resulted in 12.5 M reads corresponding to 9858 ZOTUS. As a final step, the ZOTUs were assigned to taxa (blast 2.3.0 and usearch v10.0.240, tax filter = 0.9).



## References

- Andrews, S. et al.: FastQC: a quality control tool for high throughput sequence data, 2010.
- Deiner, K. and Altermatt, F.: Transport distance of invertebrate environmental DNA in a natural river, *PloS one*, 9, e88 786, 2014.
- Deiner, K., Walser, J.-C., Mächler, E., and Altermatt, F.: Choice of capture and extraction methods affect detection of freshwater biodiversity  
5 from environmental DNA, *Biological Conservation*, 183, 53–63, 2015.
- Edgar, R. C.: UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing, *BioRxiv*, p. 081257, 2016.
- Mächler, E., Deiner, K., Spahn, F., and Altermatt, F.: Fishing in the water: effect of sampled water volume on environmental DNA-based  
detection of macroinvertebrates, *Environmental science & technology*, 50, 305–312, 2015.
- Mächler, E., Little, C. J., Wuethrich, R., Alther, R., Fronhofer, E. A., Gounand, I., Harvey, E., Huerlemann, S., Walser, J.-C., and Altermatt,  
10 F.: Assessing different components of biodiversity across a river network using eDNA, *bioRxiv*, p. 546549, 2019.