bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Environmental DNA reveals that rivers are conveyer belts of biodiversity information

Kristy Deiner[1*], Emanuel A. Fronhofer[1,2], Elvira Mächler[1] and Florian Altermatt[1,2]

[1] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.

[2] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

Running title: Rivers are conveyer belts of eDNA

Keywords: eDNA transport, metabarcoding, COI, river networks, community sampling, land water interface, macroinvertebrates, biomonitoring

Type of Article: Letters

Abstract: **147**

Main text: **4879**

Text box: **360**

Number of references: **50**

Number of figures: **5**

Number of tables: **1**

*Corresponding author K. Deiner: Ph. +41 (0)58 765 67 35; alpinedna@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract (150 max)

DNA sampled from the environment (eDNA) is becoming a game changer for uncovering biodiversity patterns. By combining a conceptual model and empirical data, we test whether eDNA transported in river networks can be used as an integrative way to assess eukaryotic biodiversity for large spatial scales and across the land-water interface. Using an eDNA metabarcode approach we detected 300 families of eukaryotes, spanning 19 phyla across the catchment of a river. We show for a subset of these families that eDNA samples overcome spatial autocorrelation biases associated with classical community assessments integrating biodiversity information over space. Additionally, we demonstrate that many terrestrial species can be detected; thus revealing eDNA in river-water integrates biodiversity information across terrestrial and aquatic biomes. Environmental DNA transported in river networks offers a novel and spatially integrated way to assess total biodiversity for whole landscapes and will revolutionize biodiversity data acquisition in ecology.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

"Eventually, all things merge into one, and a river runs through it." – Norman Maclean & Richard Friedenberg

Introduction

While rivers cover just 1% of the landmasses on earth, they are invaluable for biodiversity and ecosystem services such as drinking water and energy production (Vörösmarty *et al.* 2010). Rivers, because of their characteristic dendritic network structure, also integrate information about the landscape through the collection and transport of sediments, organic matter, nutrients, chemicals and energy (Rodríguez-Iturbe & Rinaldo 1997; Willett *et al.* 2014). For example, information contained in sediments allows us to understand how river drainages form and change in time as a result of climate and tectonic forces (Clift & Blusztajn 2005). Rivers also act as the lung of the landscape by releasing large fluxes of CO_2 derived from terrestrial plant macromolecules, such as lignin and cellulose, through the breakdown and transport of coarse and fine particulate organic matter (Ward *et al.* 2013). River networks additionally play an important role by dictating dispersal pathways which drives patterns of genetic and species diversity for many organisms across the landscape (Altermatt 2013; Mari *et al.* 2014).

Organic matter in the form of DNA is produced from organisms and is also transported through rivers via cells, tissues, gametes or organelles (termed environmental DNA 'eDNA'; Taberlet *et al.* 2012; Deiner & Altermatt 2014; Turner *et al.* 2014). DNA can be isolated from these organismal remains in the water, sequenced, and assigned back to the species of origin through the method of eDNA metabarcoding (Taberlet *et al.* 2012; Ji *et al.* 2013). This elegant process of collection and detection of a species' DNA is becoming highly valuable for biodiversity sampling in ecology and conservation (Taberlet *et al.* 2012; Ji *et al.* 2013; Bohmann *et al.* 2014; Cristescu

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2014; Kelly *et al.* 2014; Rees *et al.* 2014; Goldberg *et al.* 2015; Lawson Handley 2015). The spatial signal of eDNA, however, has only recently been explored and shows that in rivers eDNA can be transported over larger distances (Deiner & Altermatt 2014; Laramie *et al.* 2015). Therefore, we hypothesized that rivers, through the aggregation and transport of eDNA, act as conveyer belts of biodiversity information which can be used to estimate species richness over large spatial scales and across the land-water interface (Fig. 1a, Box 1).

The relevance of biodiversity sampling with eDNA found in river water is twofold. First, identifying biodiversity hotspots is invaluable for prioritizing global and regional conservation efforts (Myers *et al.* 2000). Estimates of richness to establish a place as a hotspot or not have suffered from being under-sampled (Noss *et al.* 2015). Under-sampling of biodiversity has many causes (and consequences) in conservation and ecology in general, but mainly comes from sampling methods used for estimating richness in a way that is aggregated with respect to space (Gotelli & Colwell 2001). For example, a classical method for estimating richness of aquatic macroinvertebrates is to use a kicknet method, where all individuals in a certain defined area of a stream are sampled (Barbour *et al.* 1999). Many such samples are then taken and subsequently pooled to represent richness for an entire river stretch or catchment. The pooling of spatially autocorrelated samples causes an underestimation of biodiversity compared to if each species was independently sampled. Because it is typically infeasible to sample all species independently, statistical removal of the sampling artifact is recommended (Gotelli & Colwell 2001). Estimating biodiversity through eDNA is, however, a potential way to sample each species independent of space via their DNA becoming aggregated and transported through a river's network (Fig. 1, Box 1).

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Second, river biodiversity is highly affected by environmental changes and tracking these changes in space and time is of high interest (Heino *et al.* 2015). For example, the presence of tolerant (or absence of sensitive) aquatic organisms is important for determining water quality and has been used for over a century (Kolkwitz & Marsson 1909; Hilsenhoff 1988; Bonada *et al.* 2006). This valuable metric known as 'biomonitoring' is entering a new era and the demand in its use has generated an undue burden on resource agencies. For example, the US, England, and Switzerland combined spend approximately 117.4 to 206.6 million US dollars annually on biomonitoring of aquatic systems (Table 1). This number represents only a small fraction of what countries spend, but characterizes the value we place on using species in their environment to monitor health of aquatic ecosystems. Biomonitoring is costly because of the different methods and expertise required to collect information about each targeted taxonomic group (e.g.,Table 1; Barbour *et al.* 1999; Stein *et al.* 2014). An eDNA method of biodiversity monitoring in rivers has several advantages in that it is non-lethal, minimizes field time and can sample diversity for all target groups across the tree of life with a single field sampling protocol. Therefore, demonstrating the power of this tool to monitor biodiversity of important indicator groups in rivers will provide a fast, non-lethal and inexpensive alternative tool compared with classically used methods.

Whole community detection with environmental DNA has been called the 'game changer' for biodiversity sampling (Lawson Handley 2015) and in this study we move this idea from theory into practice. We developed a conceptual model (Fig. 1; Box 1) and test the hypothesis that transported eDNA in rivers can be used in an unprecedented way to assess biodiversity of all eukaryotes. We additionally validate the ability of this method to assess the globally important macroinvertebrate communities and show that estimates of richness from eDNA compared with classical methods more precise for estimating catchment area richness because of the removal of

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spatial autocorrelation. Lastly, we demonstrate that a large number of eukaryotic phyla from both aquatic and terrestrial taxa can be detected from eDNA in river water and confirm the hypothesis that rivers are conveyor belts of biodiversity information for landscapes.

Materials and Methods

eDNA sampling, amplification and next generation sequencing

Water samples were collected from eight sites along the Glatt river network, a subcatchment of the Rhine River in Switzerland (Fig. 2). The study sites were chosen because they represent nodes in the river network where water from the major subcatchment tributaries combine and flow into the mainstem of the river Glatt. They also have a known history of monitoring macroinvertebrates for the past 15 years (AWEL 2012). At each site, DNA was isolated from between 840 to 900 mL of river water sampled. Method for sampling, capture and extraction of DNA followed that of Deiner *et al.* (2015), where the capture method of filtration was coupled with a Phenol-Chloroform Isoamyl DNA extraction. Strict adherence to contamination control was followed using a controlled lab for eDNA isolation and pre-PCR preparations (Deiner *et al.* 2015). Three independent extractions of 280 to 300 mL were carried out and then pooled to equal DNA captured and purified from 840 to 900 mL of water. Total volume of water filtered for each extraction replicate depended on the suspended solids in the sample of which clogged the filter. Water for this study was collected minutes prior to collecting aquatic macroinvertebrates using a classical sampling method (kicknet, for description see below and AWEL 2012; Altermatt *et al.* 2013) and therefore allowed for a comparison between the kicknet and eDNA methods for the detection of aquatic macroinvertebrate communities within the same watershed at the same time point.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Polymerase chain reactions (PCRs) were carried out for the target gene, cytochrome oxidase I (COI), using the universal COI primers (Folmer *et al.* 1994) on pooled eDNA extractions for each of the eight sites and amplified a fragment of 658 base pairs (bp) excluding primer sequences. PCRs were carried out in 15 µL volumes with final concentrations of 1x supplied buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland), 1000 ng/µL bovine serum albumin (BSA; New England Biolabs, Inc., Ipswich, MA, USA), 0.2 mMol dNTPs, 2.0 mMol $MgCl_2$, 0.05 units per µL Taq DNA polymerase (Faststart TAQ, Roche, Inc., Basel, Switzerland), and 0.50 µMol of each forward and reverse primer (Folmer *et al.* 1994). 2 µL of the pooled extracted eDNA was added. The thermal-cycling regime was 95 °C for 4 minutes, followed by 35 cycles of 95 °C for 30 seconds, 48 °C for 30 seconds and 72 °C for 1 minute. A final extension of 72 °C for 5 minutes was carried out and the PCR was cooled to 10 °C until removed and stored at –20 °C until confirmation of products occurred. PCR products were confirmed by gel electrophoresis on a 1.4% agarose gel stained with GelRed (Biotium Inc., Hayward, CA USA). Three PCR replicates were performed on each of the eight eDNA samples from our study sites and products from the three replicates were pooled. Negative filtration, extraction and PCR controls were used to monitor any contamination during the molecular workflow and were also replicated three times. Reactions were then cleaned using AMPure XP beads following recommended manufacturer's protocol except 0.6 x bead concentration was used instead of 1.8 x based on recommended protocol for fragment size retention of >500 base pairs (p. 31, Nextera XT DNA 96 kit, Illumina, Inc., San Diego, CA, USA). We quantified each pooled reaction using the Qubit (1.0) fluorometer following recommended protocols for the dsDNA High Sensitivity DNA Assay which has an accuracy for double stranded DNA between 0.005-0.5 pg/µL (Agilent Technologies, Santa Clara, CA, USA).

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The eight reactions were then each diluted with molecular grade water (Sigma-Aldrich, Co. LLC. St. Lewis, MO., USA) to 0.2 ng/µL following the recommended protocol for library construction (Nextera XT DNA 96 kit, Illumina, Inc., San Diego, CA, USA). Libraries for the eight sites were prepared using the Nextera XT DNA kit following the manufacturer's recommended protocols and dual indexed using the Nextera XT index kit A (Illumina, Inc., San Diego, CA, USA). Briefly, this protocol uses a process called tagmentation whereby the amplicon is cleaved preferentially from the 5' and 3' ends and the index and adaptor are ligated onto the amplicon. The tagmentation process produces an amplicon pool for each site (i.e., library) with randomly cleaved fragments averaging 300 bp in length that are subsequently duel indexed. The library constructed for each site were then pooled and paired-end sequenced (2 x 250 bp) on an Illumina MiSeq at the Genomic Diversity Center at the ETH, Zurich, Switzerland following the manufacturer's run protocols (Illumina, Inc., San Diego, CA, USA). The MiSeq Control Software Version 2.2 including MiSeq Reporter 2.2 was used for the primary analysis and the de-multiplexing of the raw reads.

Bioinformatic analysis

Run quality was assessed using FastQC version 0.10.1. Forward and reverse sequences were merged with a minimum overlap of 25 bp and minimum length of 100 bp using SeqPrep (St. John 2011). Sequences that could not be merged were excluded from further analysis. Merged sequences with quality scores less than a mean of 25 where removed. Merged sequences were then de-replicated by removing exact duplicates, were de-noised using a sequence identity threshold of 99%, and were quality trimmed left and right by 28 bp using PrinSeq Lite version 0.20.3 (Schmieder & Edwards 2011). Subsequent sequences were then chimera checked using usearch version 6 (Edgar 2010). Remaining sequences larger than 100 bp in length were then

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

taxonomically identified using customized Blast searches against the NCBI non-redundant nucleotide database using the package blast 2.2.28, build on March 12, 2013 16:52:31 (Benson *et al.* 2012). Taxonomic assignment of a sequence was done using the best blast hit based on a bit score calculated using the default blastn search of a -3 penalty for a nucleotide mismatch and a reward of +1 for a nucleotide match. Sequences that did not match eukaryotes, were below 90.0% sequence similarity, had less than 100 bp overlap with query, had a taxonomic name not assigned below the level of family, matched best with unknown environmental samples and/or had a bit score less than 100 were excluded from biodiversity detection analysis for al sites. These parameters were used because they removed likely taxonomic identification errors or exclude data that was unidentified at the family level used for analysis (Deiner *et al.* 2013; Deiner *et al.* 2015). All raw sequences reads were deposited in NCBI's Sequence Read Archive (*accession numbers pending*).

After identification of sequences with the NCBI nucleotide sequence database, each uniquely identified taxon from any site (referred to as an operational taxonomic unit, OTU) was geographically verified as known to be present in Switzerland to the lowest level of taxonomy, or if no data was available for Switzerland, it was also considered present when the OTU was known to be present in Austria, France, Germany, and Italy. We excluded a few (and very rare cases) where it is known for sure that a species is not in Switzerland, but found in all four neighboring countries. Geographic verification was done in consultation with 24 expert taxonomists for various groups, primary literature and through database repositories as described in Table S1. If the species could be confidently confirmed as being present in Switzerland or in all four neighboring countries, their known habitat use was identified as being freshwater (defined as having at least one life stage inhabiting water), or terrestrial (which included species that inhabit riparian or wet habitats or typically feed in aquatic habitats, but do not have full life

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stages or reproduce in the water; Table S1). All taxa identified to a family level with a very restricted known geographic range outside of Europe (i.e., endemic to another continent and not known to be introduced) or known to be strictly marine were also excluded from all further analysis (Table S1). Additionally, because we used BSA as an additive in PCR, we cannot rule out that detections of *Bos taurus* or *Bos indicus* were due to this reagent and therefore excluded them from analysis.

Kicknet sampling and identification

Macroinvertebrates were detected using a standard kicknet sampling design described for federal and cantonal guidelines in Switzerland (Stucki 2010; Altermatt *et al.* 2013). Briefly, we took eight independent kicknet samples per site on October 29, 2012. Large inorganic and organic debris was removed and samples were pooled into a single collection jar with 70% EtOH. Jars were then stored at room temperature until morphological identification. This method and time of year has been shown to reflect the different microhabitats and provides a robust presence measure for many macroinvertebrates in Switzerland (Stucki 2010). Since eDNA has been shown to decay over short time periods (days to a few months; Strickler *et al.* 2015), using a single time point from a kicknet sample to compare with that of what is detected in the eDNA is valid. However, it is known that kicknet samples taken at different times of year, such as in the spring, can detect different species due to morphological constraints in the identification of specimens at young life stages or just their physical presence in the water due to timing for their life cycle (Stucki 2010). Specimens from each site were sorted to the lowest taxonomic level possible (family, genus or species level) using dichotomous keys agreed upon by the Swiss Federal Office of the Environment (Stucki 2010). Specimens that could not be identified to at least to the taxonomic rank of family were excluded from further analysis.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comparison of eDNA and kicknet macroinvertebrate detection

For each site, we summarized the number of eDNA detected families of macroinvertebrates and number of families observed for the classical kicknet method using the standardized list of macroinvertebrates for biomonitoring of Swiss waters by the Federal Office for the Environment (Stucki 2010). Using this standardized list we calculated each site's observed α-diversity (local richness) for macroinvertebrates and visualized it on a heatmap of incidence. The estimated catchment area sampled for each position in the network was calculated as the cumulative sum of the area of all subcatchments into which all surface waters (excluding the lake) drain above the sampling point (Fig. 2). Topological distance between sampling sites was calculated along the river's path. Catchment area and distance between sampling sites was calculated using Quantum Geographic Information System in version 2.8 (QGIS Core Development Team 2015). The number of families detected (considered here as α-diversity) by each sampling method (eDNA and kicknet) was $\log_{10}$ transformed and regressed against the $\log_{10}$ of the river area to test for the taxon area relationship. We were interested in whether or not the two sampling methods differ in the magnitude of diversity detected due to the transport of DNA (y-intercept of the taxon area relationship), but that the fundamental rate of increase in number of taxa for a given area was not changed (slope of the regression lines) as predicted from our conceptual model. Slopes and y-intercepts of the two regressions for the taxon area relationship were tested using an analysis of covariance (ANCOVA).

To test for a spatial autocorrelation in community dissimilarity (β-diversity, using the Jaccard dissimilarity index) and between sampling locations we used a Mantel's test with 9999 permutations. The Jaccard measure of β-diversity was used as it has been shown to estimate community dissimilarity for incidence data with less biases because of nestedness which is

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expected for the eDNA estimate of β-diversity due to transport (Cardoso *et al.* 2009). All statistical analyses were performed in R version 3.1.0 (R Development Core Team 2013).

Results

All Eukaryotes

We found a total of 1758 unique OTUs that could be assigned to independent taxa across all eukaryotes (Table S1). Of these, 1413 (80.3 %) could be confirmed to taxonomic families known to be present in Switzerland or in all four neighboring countries (Table S1). The total confirmed eukaryotic diversity sampled from the eight locations spanned 19 phyla of which 300 families of eukaryotes could be geographically verified as known to occur (Fig. 3). We could further geographically verify the presence of 472 genera and of these 260 species that are known to occur. Nearly half (45.7 %) of the species we detected from eDNA in water are known to be terrestrial (Fig. 4; N = 119).

Of the remaining 335 from the total of 1758 OTUs, 286 were identified to families not known to occur in Switzerland. More than half of these (175) were assigned to families known only to contain species living in the marine environment (e.g., sponges, red algae, worms and snails, etc.), 81 are known to be terrestrial, 24 are known to inhabit freshwater but not from Switzerland, and six could not have their habitat identified due to the identification only being at the level of family. Lastly, 59 OTUs were identified to families for which we could find no distribution data using our verification methods. Assignment statistics of sequence length and percent sequence similarity for all 335 OTUs tended to be at the threshold for accepting an assignment, but there

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were some outliers that had high assignment values indicating likely detections (Fig. S1). These outliers were, however, not included in the geographically confirmed family counts.

Macroinvertebrates

Of the 300 families considered detected for eukaryotes, 65 of them are used in the Swiss biomonitoring program (Fig. S2; Stucki 2010). Thirteen additional families were detected by kicknet samples only, totaling 78 macroinvertebrate families detected among our sampling sites of the river Glatt. From eDNA we recovered between 23 to 40 families at each site (Fig. S2). With the classical kicknet method we sampled between 17 to 24 families at each site (Fig. S2). Several families were detected by eDNA only have been previously sampled with kicknets at the same sites in the more than 15 years of monitoring (i.e., Coenagrionidae, Gyrinidae, Psychomyiidae, Rhyacophilidae; AWEL 2012). Of the total 78 families detected, 33 (42 %) were detected by both methods, and often at the same location (Fig. S2). Of the remaining 45 (58 %) of families, 32 were only detected with eDNA and 13 where only detected with the kicknet sample. Eight of these 13 families were detected in the eDNA dataset, but did not meet bioinformatics thresholds for filtering assignment values (e.g., where below a 90 % sequence similarity, data not shown). Of the remaining five families, three had sufficient sequence data on GenBank for identification (e.g., more than 100 sequences from COI across a broad array of genera within each family), but two had very low representation (Potamanthidae and Aphelocheiridae) with six and 22 sequences respectively. Of the 32 families detected with eDNA, three are known to occur in the lake Greifensee, which feeds into the river Glatt (i.e., Cambaridae, Notonectidae, and Sialidae; AWEL 2012), but are not known from the river Glatt. Local family richness (α-diversity) increased for sites sampled at more downstream positions in the river's network for eDNA, but did not for kicknet samples (Fig. 5a).

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cumulative family richness (γ-diversity) increased as a function of cumulative catchment area sampled for both sampling methods ($F_{1,6} = 114.1$, $r^2 = 0.95$, $p < 0.0001$, eDNA; $F_{1,6} = 69.66$, $r^2 = 0.92$, $p = 0.0001$ kicknet) (Fig. 5b). The slopes of the family-area relationship were not different ($F_{1,12} = 1.113$, $p = 0.312$), however the y-intercept was higher for eDNA compared with kicknet ($F_{1,13} = 41.61$, $p < 0.0001$) (Fig. 5b). β-diversity in the form of community dissimilarity did not increase as a function of distance for eDNA ($r = 0.02$, $p = 0.43$), whereas for kicknet sampling we observed a positive correlation in β-diversity as a function of distance between sampling sites ($r = 0.52$, $p = 0.003$) (Fig. 5c).

Discussion

We demonstrate that rivers, through their collection and transport of eDNA (Fig. 1), can be used to sample catchment-level biodiversity across the land-water interface. For aquatic macroinvertebrates, we found a greater richness in the number of families detected with eDNA compared with the classical kicknet method at the same time point (Fig. 5). This increased precision comes from the natural process of transport of DNA through the network of a river, which works to reduce the biases associated with spatial autocorrelation inherent to classical community sampling (Fig.1; Box 1). These novel results offer ecologists a new and unprecedented tool to sample landscape biodiversity and estimate richness of eukaryotic communities across biomes.

Our model (Box 1) identifies three important messages for the utility of eDNA as a genomic tool for biodiversity assessment. First, eDNA detection of species from river water decouples the presence of a species from its physical location in a habitat through the downstream transport. Transport distance has been shown to be limited to about 10 to 12 km (Deiner & Altermatt 2014),

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

thus allowing for increased precision in detection of patchy or elusively distributed species and allows for richness estimates with less sampling effort because of the integrated signal over space. Second, eDNA will likely sample a higher diversity compared with classical sampling methods at any given site, but this depends on the local distribution of species and factors affecting transport and degradation of eDNA. Third, the interpretation of the species presence inferred from an eDNA sample in a river is different from that of classical sampling methods. Namely, eDNA detection of species should be interpreted as an integrated signal of presence and the spatial scale is determined based on the potential transport distance for a system. Thus, our model suggests that eDNA in rivers is an efficient tool for large scale biodiversity assessments, and depending on the distance between water samples, less authoritative for very localized richness estimates.

Our data substantiate our conceptual predictions (Box 1), and by comparing eDNA with kicknet samples at each site, we highlight several important factors that illustrate both the power and current limitations of eDNA for biodiversity assessment. Many families of macroinvertebrates were detected at each site by both methods and have a great degree of overlap in which sites families were co-detected. For all sites, however, eDNA recovered more macroinvertebrate families compared to kicknet samples. We hypothesize this is likely due to the integrated signal from transported DNA which is evident by the fact that community composition does not change much (i.e., β-diversity remaining constant over distance), compared with kicknet estimated β-diversity which increased over the same river distance. This difference means that the two sampling methods give different information at the same site. Classical sampling methods give information that is very localized, whereas the eDNA metabarcode method in rivers measures presence of species for larger spatial scales. This novel finding is of great importance because in many cases diversity information for a large area is the goal for places such as biodiversity

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hotspots or conservation preserves (Noss *et al.* 2015), or entire river catchments (Sheldon *et al.* 2012). As scaling up of the classical community sampling method will likely always underestimate diversity (Gotelli & Colwell 2001), eDNA offers an empirical method to overcome this limitation and is an unparalleled way to estimate richness for larger areas.

Much of the current degradation of river habitat is at the catchment scale and cannot be attributed to a single point source (Vörösmarty *et al.* 2010). Biomonitoring currently relies on the costly sampling of macroinvertebrates across many sites to understand ecosystem health of rivers (Table 1; Stein *et al.* 2014). Therefore, an eDNA signal of macroinvertebrates can be used to measure more precisely diversity of a catchment with much less sampling effort. In contrast, understanding local changes in richness at a restoration site may still require classical sampling with kicknets. Interestingly, however, transport distances of eDNA match the scale at which local species' pools are recognized to be important for recolonization of restored patches in a river system (0-5 km) (Sundermann *et al.* 2011). Therefore, eDNA could be used as a way to measure the species' pool available for recolonization. A tool such as this can aid in prioritizing river restoration efforts by identifying regions which have high recolonization potential of target species and possibly set expectations for the magnitude of change expected for restoration sites already in recovery.

Our results also identify a way of empirically measuring transport of community eDNA in rivers. Our analysis of β-diversity in this study system shows that community eDNA is likely transported and detected over a scale greater than 12 km. This has been previously shown for two targeted species that live only in lake Greifensee (mussel and a waterflea), where their DNA could be detected at sites d and f respectively (Fig. 2; Deiner & Altermatt 2014). To determine the scale of transport for community eDNA in a river system, one subsequently needs to detect

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the scale where there is a positive spatial autocorrelation between eDNA and β-diversity. This empirical measure of transport is needed because, as shown by our conceptual model, eDNA detection of biodiversity is a function of the transport distance, but also a function of the distribution of species within the network. Transport itself is furthermore affected by local factors, such as degradation due to UVB, pH and temperature (Strickler *et al.* 2015), as well as discharge rates (Jane *et al.* 2015). Therefore, eDNA may not be necessarily transported and detected over the same distance for all river systems or consistently in time due to extreme events like heavy rainfall or drought. By using the correlation between β-diversity and river distance between sampling points, however, an *in situ* test can be performed and the scale of transport for community eDNA uncovered for any system and can be repeated across time to test if eDNA transport distance is stable in a system.

There are, however, important current limitations of the eDNA metabarcode method and are related to factors such as the importance of primer or marker choice and the biodiversity detected as a function of the reference data with its annotations available for identification of sequences (Deagle *et al.* 2014; Elbrecht & Leese 2015). For example: fish, flatworms, and diatoms in this dataset are underrepresented to what we know occurs in this system. This is most likely due to the choice of primers, the genetic maker and potentially the reference database used for identifying sequences. The primers used in this study are the universal Folmer primers for the 5' end of COI (Folmer *et al.* 1994) and it is known that these primers do not amplify DNA from fish and flatworms very well (Ivanova *et al.* 2007; Moszczynska *et al.* 2009; respectively). Additionally, for diatoms it is known that COI is not the best genetic marker suitable for species level identification (Zimmermann *et al.* 2015). Therefore, it is clear that more than one maker and/or primer set is needed to capture biodiversity for the tree of life adequately (Gibson *et al.* 2014). With an eDNA metabarcode sampling method this does not require additional sampling

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in the field, however, but creates a single sampling method whereby careful amplification of a genetic marker and primer choice will enable an integrated detection across taxonomic diversity from a single sample.

Conclusions

We have demonstrated that rivers convey, through the collection and transport of environmental DNA, an unprecedented amount of information on biodiversity in landscapes. Our study demonstrates that eDNA can be used to sample community structure of river catchments and do so even across the land-water interface. As such, detection of eukaryotic fauna with DNA found and transported in rivers may unite historically separated research fields of aquatic and terrestrial ecology and provide an integrated measure of total biodiversity for rapid assessment for one of the most highly impacted biomes of the world.

Acknowledgements

We would like to thank Patrick Steinmann for assistance with kicknet samples and taxonomic identification, Katharina Kaelin for assistance with our study area map and geographic information used in this study, and Peter Penicka for contributing illustrations to Figure 1. We thank Florian Leese and Vasco Elbrecht for providing insightful feedback on a previous version of the manuscript. Data analysed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. This work was funded by Eawag: Swiss Federal Institute of Aquatic Science and Technology and the Swiss National Science Foundation (grant no. PP00P3_150698 to FA).

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Literature cited

1.

Altermatt, F. (2013). Diversity in riverine metacommunities: a network perspective. *Aquat Ecol*, 47, 365-377.

2.

Altermatt, F., Seymour, M. & Martinez, N. (2013). River network properties shape α-diversity and community similarity patterns of aquatic insect communities across major drainage basins. *Journal of Biogeography*, 40, 2249-2260.

3.

AWEL (2012). Zürcher Gewässer 2012: Entwicklung - Zustand - Ausblick. Amt für Abfall, Wasser, Energie und Luft, Zürich, Switzerland. (Available from: http://www.awel.zh.ch/internet/baudirektion/awel/de/wasserwirtschaft/veroeffentlichungen.html). Abfall Wasser, Energie, und Luft.

4.

Barbour, M.T., Gerritsen, J., Snyder, B. & Stribling, J. (1999). Rapid Bioassessment Protocols For Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish. (ed. USEPA) Washington, D.C., p. 339.

5.

Benson, D.A., Karsch-Mizrachi, I., Clark, K., Lipman, D.J., Ostell, J. & Sayers, E.W. (2012). GenBank. *Nucleic Acids Research*, 40, D48-D53.

6.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M. *et al*. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29, 358-367.

7.

Bonada, N., Prat, N., Resh, V.H. & Statzner, B. (2006). Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annu. Rev. Entomol.*, 51, 495-523.

8.

Cardoso, P., Borges, P.A. & Veech, J.A. (2009). Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Diversity and Distributions*, 15, 1081-1090.

9.

Clift, P.D. & Blusztajn, J. (2005). Reorganization of the western Himalayan river system after five million years ago. *Nature*, 438, 1001-1003.

10.

Cristescu, M.E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, 29, 566-571.

11.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10, 20140562.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12.

Deiner, K. & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. *PloS One*, 9, e88786.

13.

Deiner, K., Knapp, R.A., Boiano, D.M. & May, B. (2013). Increased accuracy of species lists developed for alpine lakes using morphology and cytochrome oxidase I for identification of specimens. *Molecular Ecology Resources*, 13, 820-831.

14.

Deiner, K., Walser, J.-C., Mächler, E. & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, 53-63.

15.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.

16.

Elbrecht, V. & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PeerJ PrePrints*, 3, e1258.

17.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294-299.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18.

Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H. *et al.* (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, 111, 8007-8012.

19.

Goldberg, C.S., Strickler, K.M. & Pilliod, D.S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, 183, 1-3.

20.

Gotelli, N.J. & Colwell, R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379-391.

21.

Heino, J., Melo, A.S., Bini, L.M., Altermatt, F., Al-Shami, S.A., Angeler, D.G. *et al.* (2015). A comparative analysis reveals weak relationships between ecological factors and beta diversity of stream insect metacommunities at two spatial levels. *Ecology and Evolution*, 5, 1235-1248.

22.

Hilsenhoff, W.L. (1988). Rapid field assessment of organic pollution with a family-level biotic index. *Journal of the North American Benthological Society*, 65-68.

23.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H. & Hebert, P.D. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 7, 544-548.

24.

Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H. *et al.* (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources*, 15, 216-227.

25.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A. *et al.* (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16, 1245-1257.

26.

Kelly, R.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F. *et al.* (2014). Harnessing DNA to improve environmental management. *Science*, 344, 1455-1456.

27.

Kolkwitz, R. & Marsson, M. (1909). Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung. *Internationale Revue der Gesamten Hydrobiologie und Hydrographie*, 2, 126-152.

28.

Laramie, M.B., Pilliod, D.S. & Goldberg, C.S. (2015). Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation*, 183, 29-37.

29.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lawson Handley, L. (2015). How will the 'molecular revolution'contribute to biological recording? *Biological Journal of the Linnean Society*, In press, 10.1111/bij.12516.

30.

Mari, L., Casagrandi, R., Bertuzzo, E., Rinaldo, A. & Gatto, M. (2014). Metapopulation persistence and species spread in river networks. *Ecology Letters*, 17, 426-434.

31.

Moszczynska, A., Locke, S.A., McLAUGHLIN, J.D., Marcogliese, D.J. & Crease, T.J. (2009). Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Molecular Ecology Resources*, 9, 75-82.

32.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.

33.

Noss, R.F., Platt, W.J., Sorrie, B.A., Weakley, A.S., Means, D.B., Costanza, J. *et al.* (2015). How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Diversity and Distributions*, 21, 236-244.

34.

QGIS Core Development Team (2015). QGIS Geographic Information System. Available from http://qgis.osgeo.org. Open Source Geospatial Foundation Project.

35.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

R Development Core Team (2013). R: A language and environment for statistical computing. Version 3.0.1. R Foundation for Statistical Computing Vienna, Austria.

36.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R. & Gough, K.C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51, 1450-1459.

37.

Rodríguez-Iturbe, I. & Rinaldo, A. (1997). *Fractal river basins: chance and self-organization*. Cambridge University Press, Cambridge, UK.

38.

Schmieder, R. & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27, 863-864.

39.

Sheldon, F., Peterson, E.E., Boone, E.L., Sippel, S., Bunn, S.E. & Harch, B.D. (2012). Identifying the spatial scale of land use that most strongly influences overall river ecosystem health score. *Ecological Applications*, 22, 2188-2203.

40.

St. John, J. (2011). SeqPrep. Available <https://github.com/jstjohn/SeqPrep>. Github.

41.

Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E. & Zakharov, E.V. (2014). Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? *PloS One*, 9, e95525.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

42.

Strickler, K.M., Fremier, A.K. & Goldberg, C.S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, 183, 85-92.

43.

Stucki, P. (2010). Methoden zur Untersuchung und Beurteilung der Fliessgewässer: Makrozoobenthos-Stufe F. Bundesamt für Umwelt-Vollzug, Bern, Switzerland. (Available from: http://www.bafu.admin.ch/publikationen/publikation/01575/index.html?lang=de).

44.

Sundermann, A., Stoll, S. & Haase, P. (2011). River restoration success depends on the species pool of the immediate surroundings. *Ecological Applications*, 21, 1962-1971.

45.

Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L.H. (2012). Environmental DNA. *Molecular Ecology*, 21, 1789-1793.

46.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. & Lodge, D.M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5, 676-684.

47.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vörösmarty, C.J., McIntyre, P., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P. *et al.* (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555-561.

48.

Ward, N.D., Keil, R.G., Medeiros, P.M., Brito, D.C., Cunha, A.C., Dittmar, T. *et al.* (2013). Degradation of terrestrially derived macromolecules in the Amazon River. *Nature Geoscience*, 6, 530-533.

49.

Willett, S.D., McCoy, S.W., Perron, J.T., Goren, L. & Chen, C.-Y. (2014). Dynamic reorganization of river basins. *Science*, 343, 1248765.

50.

Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. (2015). Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, 15, 526-542.

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Box 1: Conceptual model exemplifying integration of biodiversity information along a river network using eDNA

Classical sampling methods, such as kicknet sampling, in rivers are very time- and cost-intensive (Table 1; Stein *et al.* 2014). Typically, sample methods for communities only capture a fraction of local α-diversity due to imperfect detection and sampling bias (Fig. 1c):

$$\alpha^{x}_{classical} = \alpha^{x}_{real} \cdot \delta^{classical} \qquad (1)$$

with $\alpha^{x}_{classical}$ representing the measured α-diversity at a spatial location *x* in a river network using classical sampling methods, α^{x}_{real} is the real α-diversity at this location and $\delta^{classical}$ as the detection rate of the sampling method. In order to comprehensively estimate the biodiversity of a river catchment, a large number of such samples are required. If samples are spatially autocorrelated, pooling of community samples will result in an underestimate of the real local richness (Gotelli & Colwell 2001).

Riverine networks, however, have the potential to collect this information for us (Rodríguez-Iturbe & Rinaldo 1997; Willett *et al.* 2014) if we use an appropriate sampling method not biased by spatial autocorrelation for the area under study. Characteristic properties of rivers, such as the specific distribution of biodiversity (Altermatt *et al.* 2013) and transport of eDNA by the flow of water (Deiner & Altermatt 2014) make eDNA a promising method to estimate catchment level biodiversity while sampling at only one or very few locations

$$\alpha^{catchment}_{eDNA} = \left(\alpha^{x}_{real} + \sum \alpha^{y}_{real} \cdot N_y \cdot \beta_y \cdot \beta_{x,y} \cdot \tau_{x,y}\right) \cdot \delta^{eDNA} \qquad (2)$$

with $\alpha^{catchment}_{eDNA}$ as the integrated measure of catchment α-diversity (see also Fig. 1). The sum captures the information integrated by the riverine system for all locations *y* (Strahler stream order) upstream of the sampling location *x*. The local diversity at a site of Strahler stream order *y* has to be weighted according to Horton's Law to capture the number of streams of this Strahler stream order (N_y; Rodríguez-Iturbe & Rinaldo 1997) as well as by the Strahler stream order-characteristic β-diversity (β_y). The estimate of catchment-level biodiversity increases with increasing β-diversity between the sampling point and all upstream locations ($\beta_{x,y}$) as well as with increasing transport distance ($\tau_{x,y}$; net rate including shedding and degradation). Note that

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the eDNA specific detection probability (δ^{eDNA}) tends to be high as, in principle, only very few DNA molecules are needed for successful detection.

Figure legends

Figure 1

Conceptual model of environmental DNA spatial dynamics in a hypothetical river network. a) Visualization of species distribution in a landscape illustrating the release and accumulation of DNA in river water throughout its catchment. b) Characteristically high between-community diversity among headwaters (Strahler stream order 1; thinnest lines) is indicated by different colors representing local richness (α-diversity). Increasing size of pie chart indicates change in abundance. Flow direction is indicated with an arrow. Strahler stream order is indicated by the increasing width of river lines. c) While classical sampling only detects a fraction of real local diversity, eDNA sampling allows an estimate of catchment level diversity including both aquatic and terrestrial taxa and integrates this information across space due to downstream transport of eDNA.

Figure 2

Study area and location of sampling sites where environmental DNA samples and classical sampling methods were carried out. The direction of flow for the river Glatt is northwest (blue arrow). The main stem of the river originates from the outflow of lake Greifensee. Colored regions represent the catchment upstream of each sampling point. Letters are used to indicate the position in the river network starting from the outflow ‘a’ to ‘f’ and the two sampled tributaries

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

'ab' and 'cd'. Sources for GIS data were from Swisstopo (DHM25, Gewässernetz Vector 25) and reprinted with permission.

Figure 3

Total eukaryotic diversity detected from the river Glatt using environmental DNA metabarcoding. The number of families per phyla (N = 300) sampled and confirmed as being present in Switzerland or known from all four neighboring countries (Austria, France, Germany, and Italy). The inset further breaks the most abundant phylum (Arthropoda) into the number of families sampled by class (N = 200).

Figure 4

Percent terrestrial or freshwater species for the subset of each phylum that could be confirmed as known to be present in Switzerland or known from all four neighboring countries (Austria, France, Germany, and Italy; N=260). Number in brackets indicates the number of species confirmed for each phylum.

Figure 5

Difference of benthic macroinvertebrate family richness and community dissimilarity estimated between environmental DNA and kicknet sampling. a) α-diversity measured at each sample site in the river network for macroinvertebrate families. b) Log_{10} transformed taxon-area relationship for eDNA and kicknet samples. Slopes of lines are not significantly different ($p = 0.312$), however the y-intercept is significantly higher for eDNA compared with kicknet ($p < 0.001$) indicating that eDNA samples a greater amount of diversity. c) Correlation of community dissimilarity with along-stream geographic distances between sample sites. Solid line for kicknet

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

β-diversity indicates a significant positive relationship with stream distance ($p = 0.003$). Dashed line for eDNA β-diversity indicates no significant relationship with distance ($p = 0.43$).

Table

Table 1: Annual bioassessment costs in millions of dollars for freshwater resources. Information not available is abbreviated as NA. Original currency in brackets.

Target Group	USA[1]	England[2]	Switzerland[3]
Fish	31.4 - 58.2	NA	0.35 (0.33 CHF)
Benthic invertebrates	38.1 - 70.7	NA	0.61 (0.58 CHF)
Algae (diatoms)	34.7 - 64.5	NA	0.32 (0.3 CHF)
Macrophytes	NA	NA	0.3 (0.29 CHF)
Total	104.2 - 193.4	11.6 (7.3£)	1.58 (1.5 CHF)

Sources: [1] Stein *et al.* 2014; [2] Richard Walmsley, Forestry Commission personal communication; [3] Markus Wüest, Federal Office for the Environment FOEN personal communication.

Supplementary material

Figure S1

Length of alignment and similarity between sequence and reference data for operational taxonomic units at the family level of taxonomy. (a) Confirmed known geographic presence of families in Switzerland or from all four neighboring countries (Austria, France, Germany, and Italy) (b) confirmed not to be present. Most unconfirmed taxonomic sequence assignments at the family level fell at two of the many thresholds set for accepting an assignment as valid (i.e., 100 bp in length and 90% identical). The outliers are color coded as follows: Euphausiidae (blue), Erobdellidae (orange), Naccariaceae (red), Rytididae, Calyptraeidae, Naticidae, Ampullariidae, Aeolidiidae (all belonging to Gastropoda, black).

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

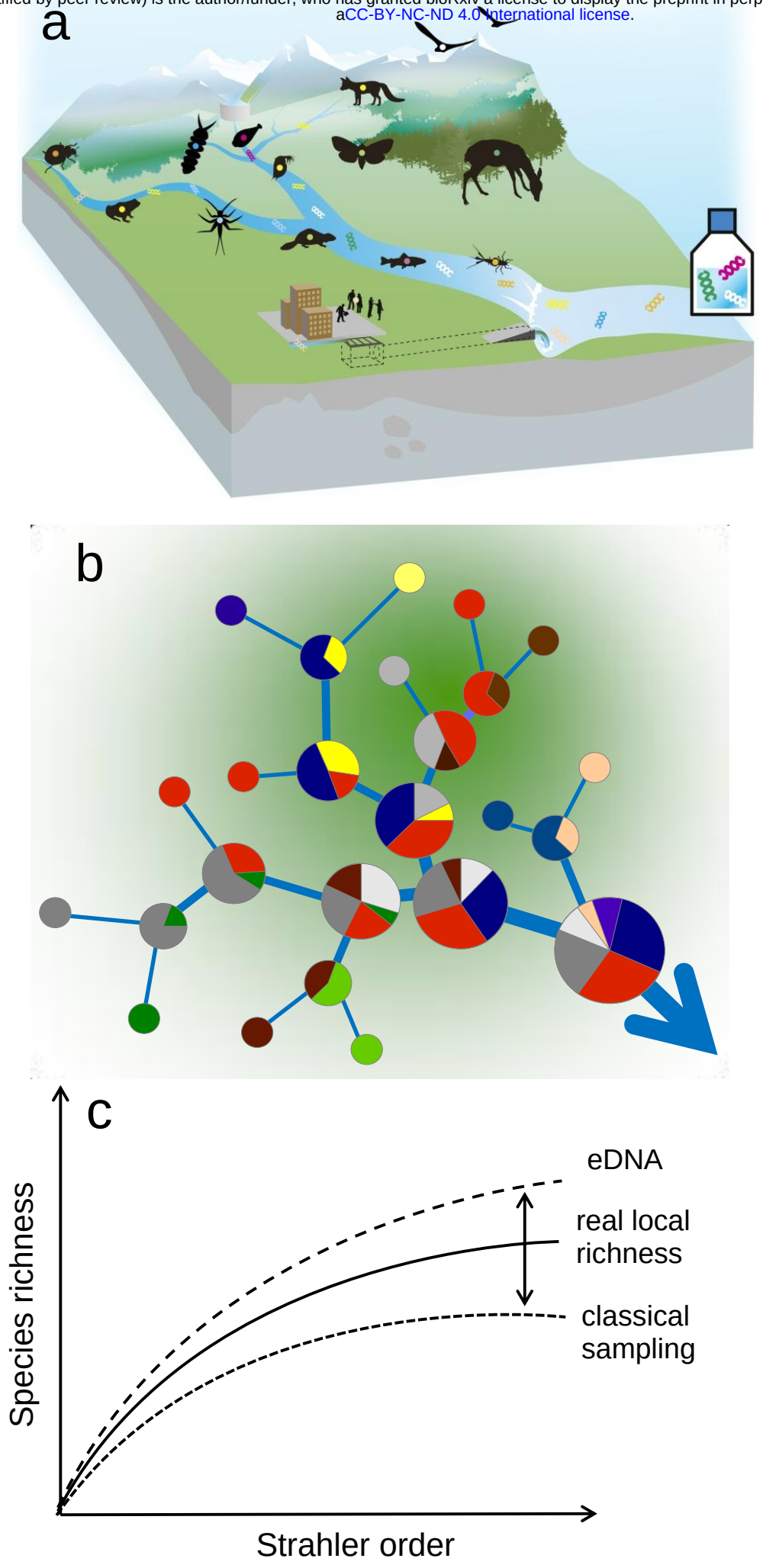
Figure S2

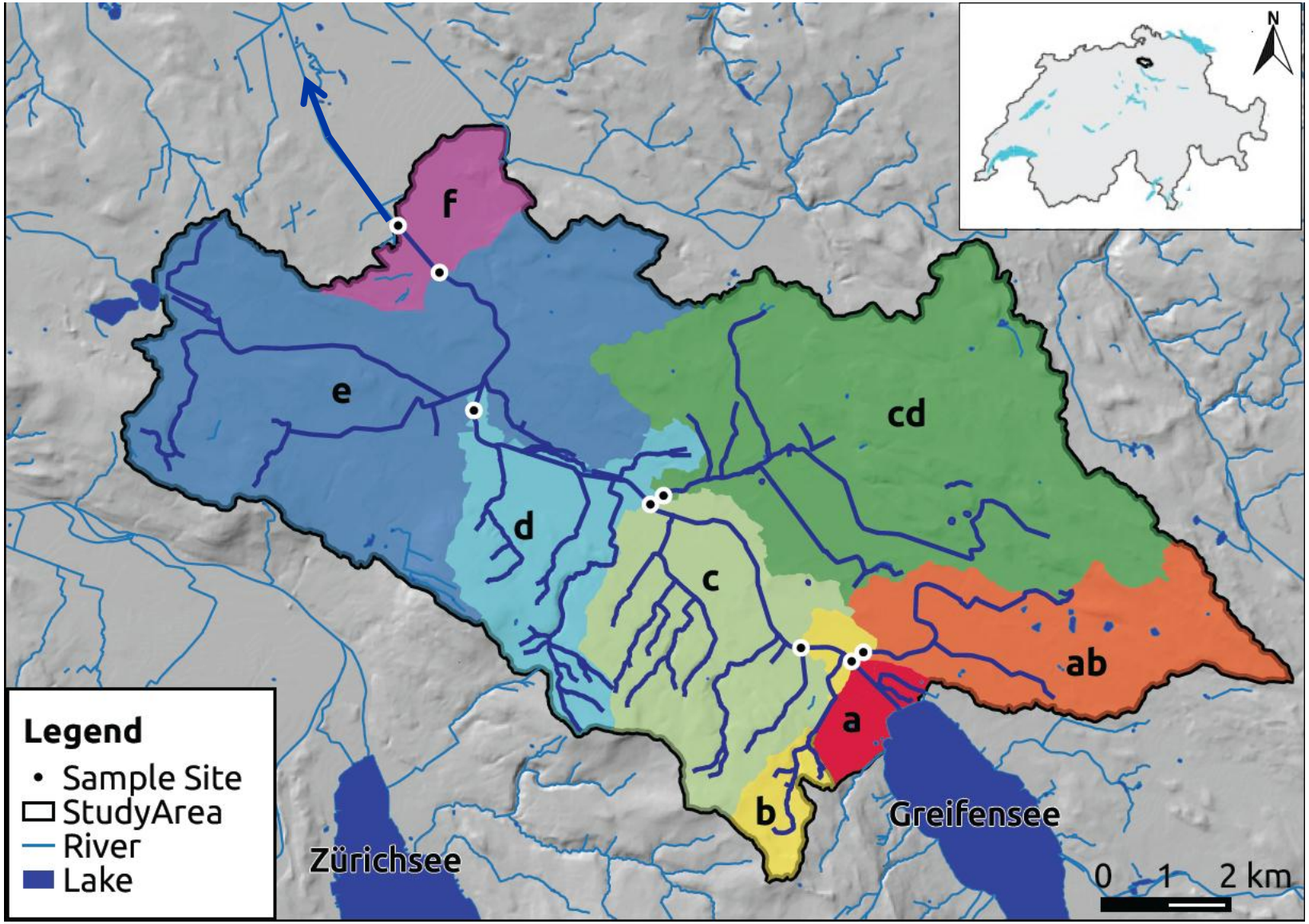
Heatmap illustrating each macroinvertebrate family, its detection across each site and the equivalency in detection for environmental DNA and kicknet sampling methods. Blue indicates presence for eDNA, red indicates presence for kicknet and white indicates not detected. For the equivalency, black indicates both were or were not detected at the same site, white indicates the family was detected by only one of the sampling methods.

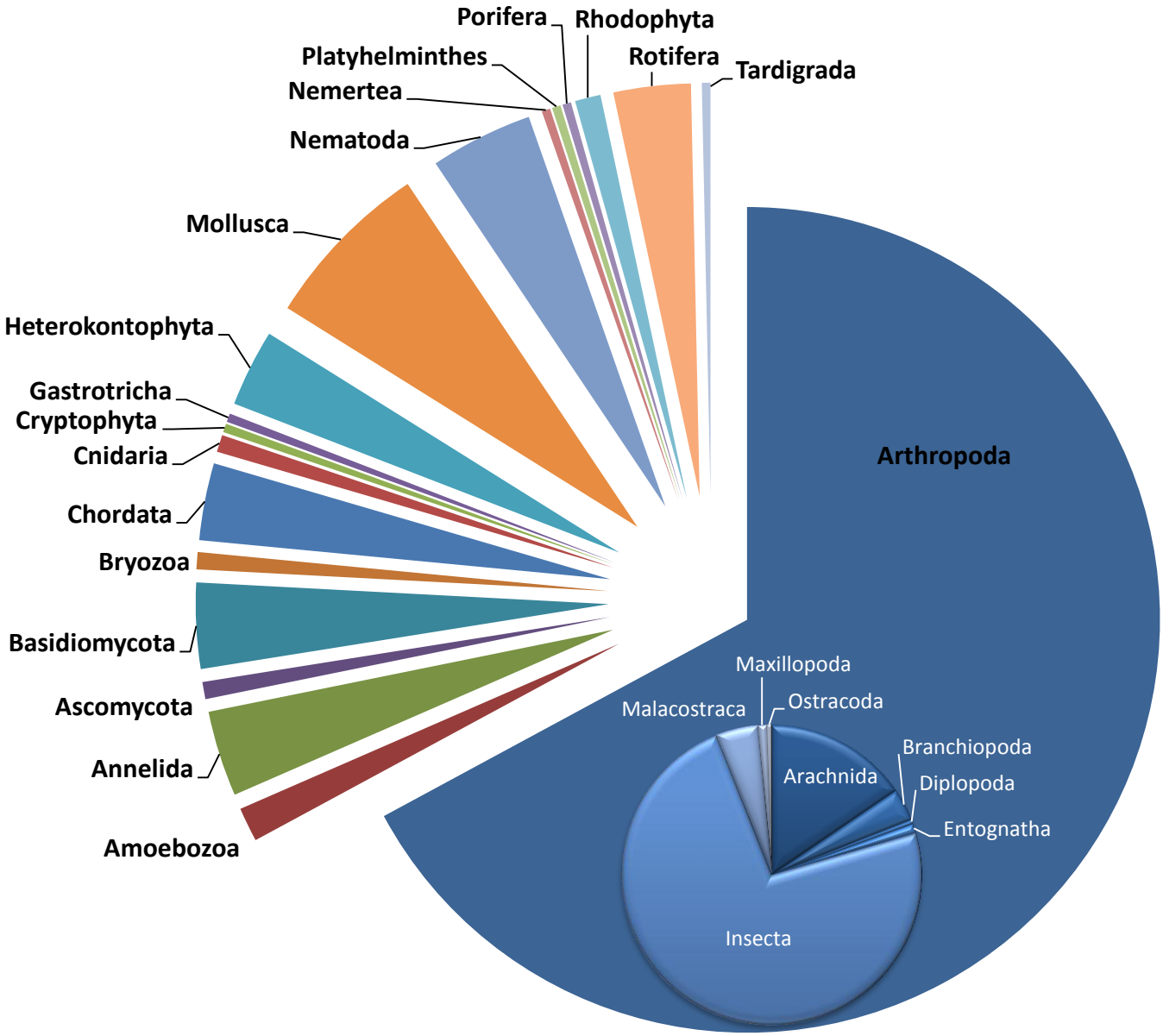
Table S1

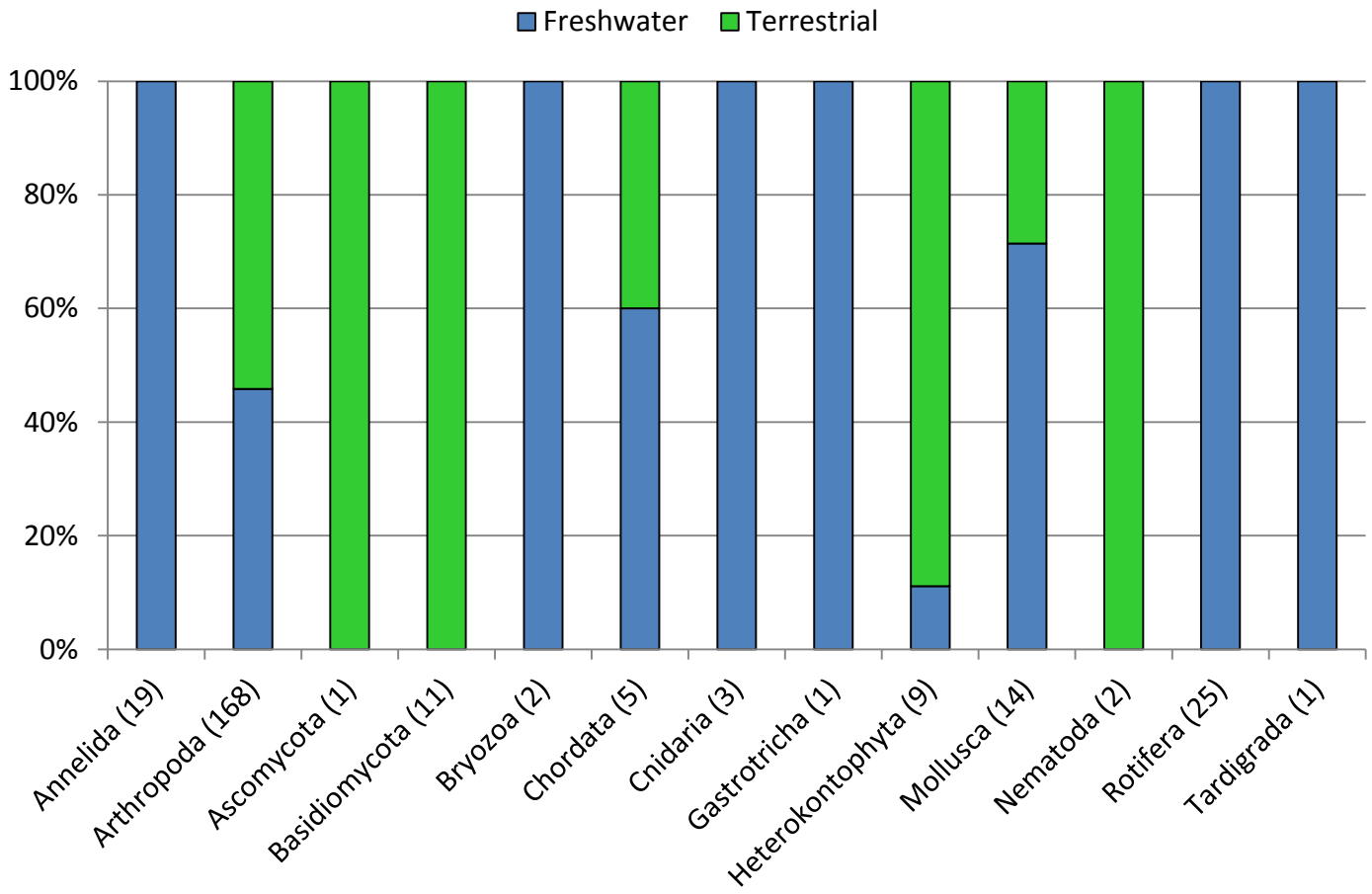
Assignment statistics and taxonomic assignments for sampled operational taxonomic units from environmental DNA in the river Glatt, Switzerland. At the family level and below, taxonomic assignment of cells were color coded for the geographic verification process. Green colored cells were confirmed as known to be present, red colored cells were confirmed as not known to be present, and white colored cells had no data available for verification. Not available is abbreviated as “na”. Geographic verification of each taxa was done by personal communication with expert taxonomists, primary literature, and database repositories as indicated in confirmation resources of the second worksheet.

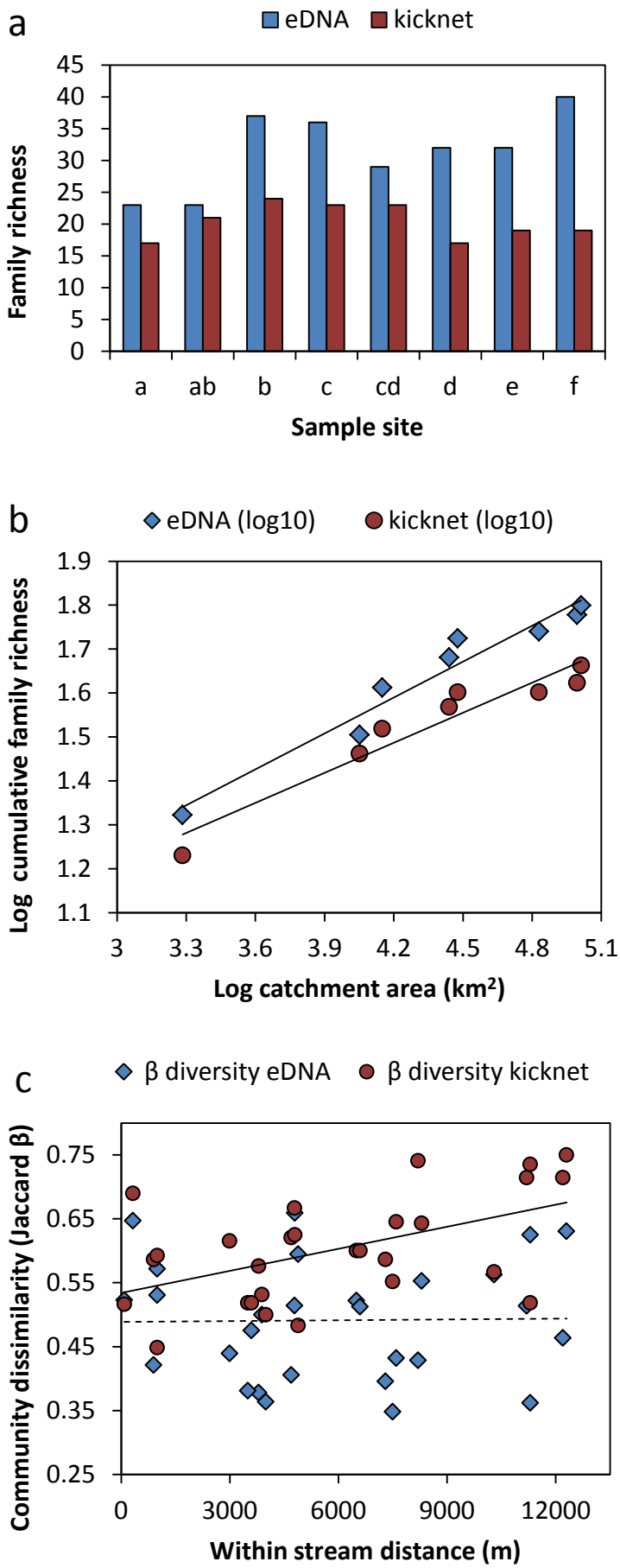
bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

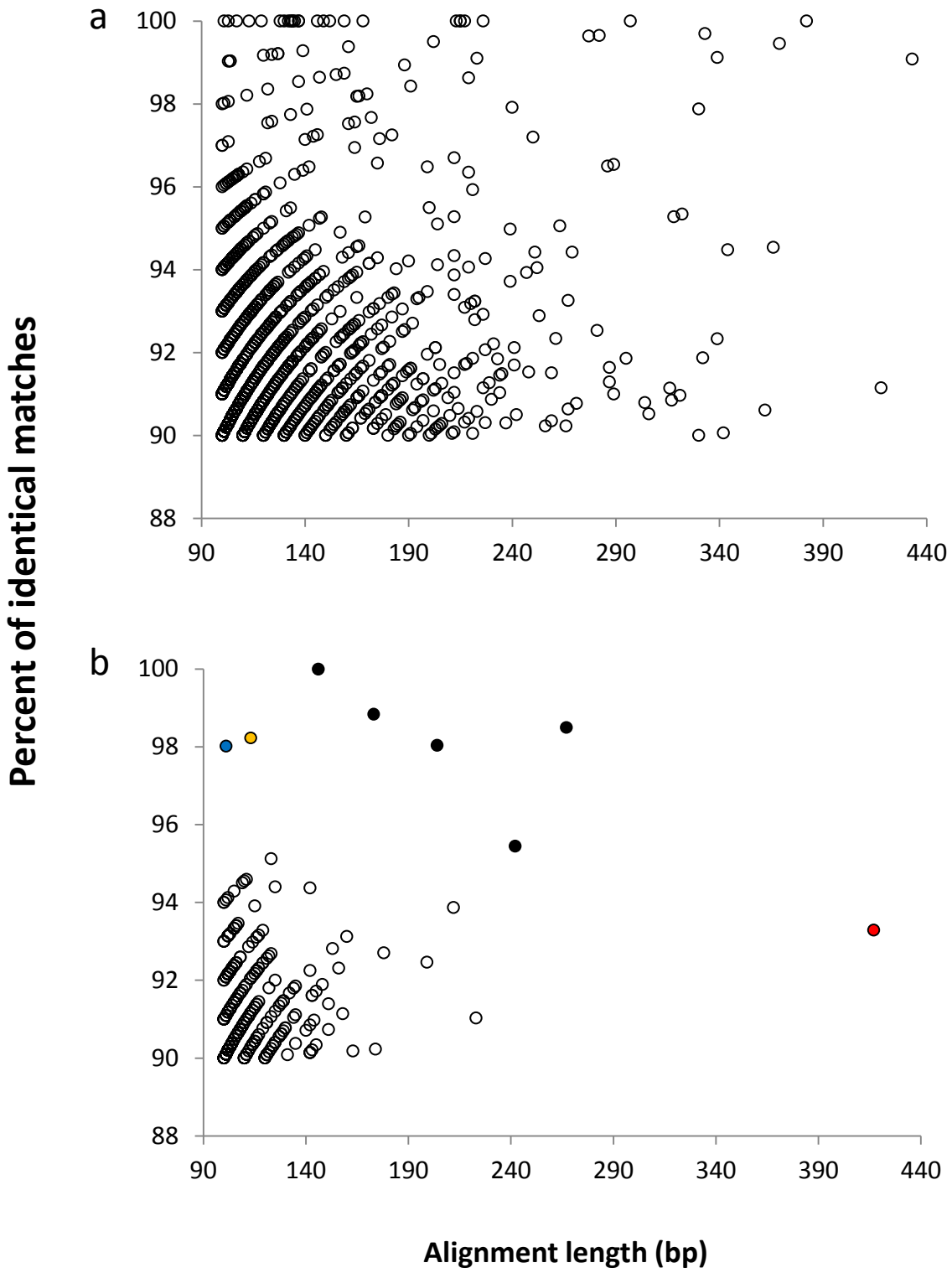



N
f
e
d
c
cd
ab
a
b
Legend
Sample Site
StudyArea
River
Lake
Zürichsee
Greifensee
0 1 2 km


Porifera
Rhodophyta
Platyhelminthes
Rotifera
Tardigrada
Nemertea
Nematoda
Mollusca
Heterokontophyta
Gastrotricha
Cryptophyta
Cnidaria
Chordata
Bryozoa
Basidiomycota
Ascomycota
Annelida
Amoebozoa
Arthropoda
Maxillopoda
Malacostraca
Ostracoda
Branchiopoda
Arachnida
Diplopoda
Entognatha
Insecta


Freshwater
Terrestrial
100%
80%
60%
40%
20%
0%
Annelida (19)
Arthropoda (168)
Ascomycota (1)
Basidiomycota (11)
Bryozoa (2)
Chordata (5)
Cnidaria (3)
Gastrotricha (1)
Heterokontophyta (9)
Mollusca (14)
Nematoda (2)
Rotifera (25)
Tardigrada (1)


a
eDNA
kicknet
Family richness
45
40
35
30
25
20
15
10
5
0
a
ab
b
c
cd
d
e
f
Sample site
b
eDNA (log10)
kicknet (log10)
Log cumulative family richness
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
3
3.3
3.6
3.9
4.2
4.5
4.8
5.1
Log catchment area (km²)
c
β diversity eDNA
β diversity kicknet
Community dissimilarity (Jaccard β)
0.75
0.65
0.55
0.45
0.35
0.25
0
3000
6000
9000
12000
Within stream distance (m)


a
b
100
98
96
94
92
90
88
90
140
190
240
290
340
390
440
Percent of identical matches
Alignment length (bp)

bioRxiv preprint doi: https://doi.org/10.1101/020800; this version posted June 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

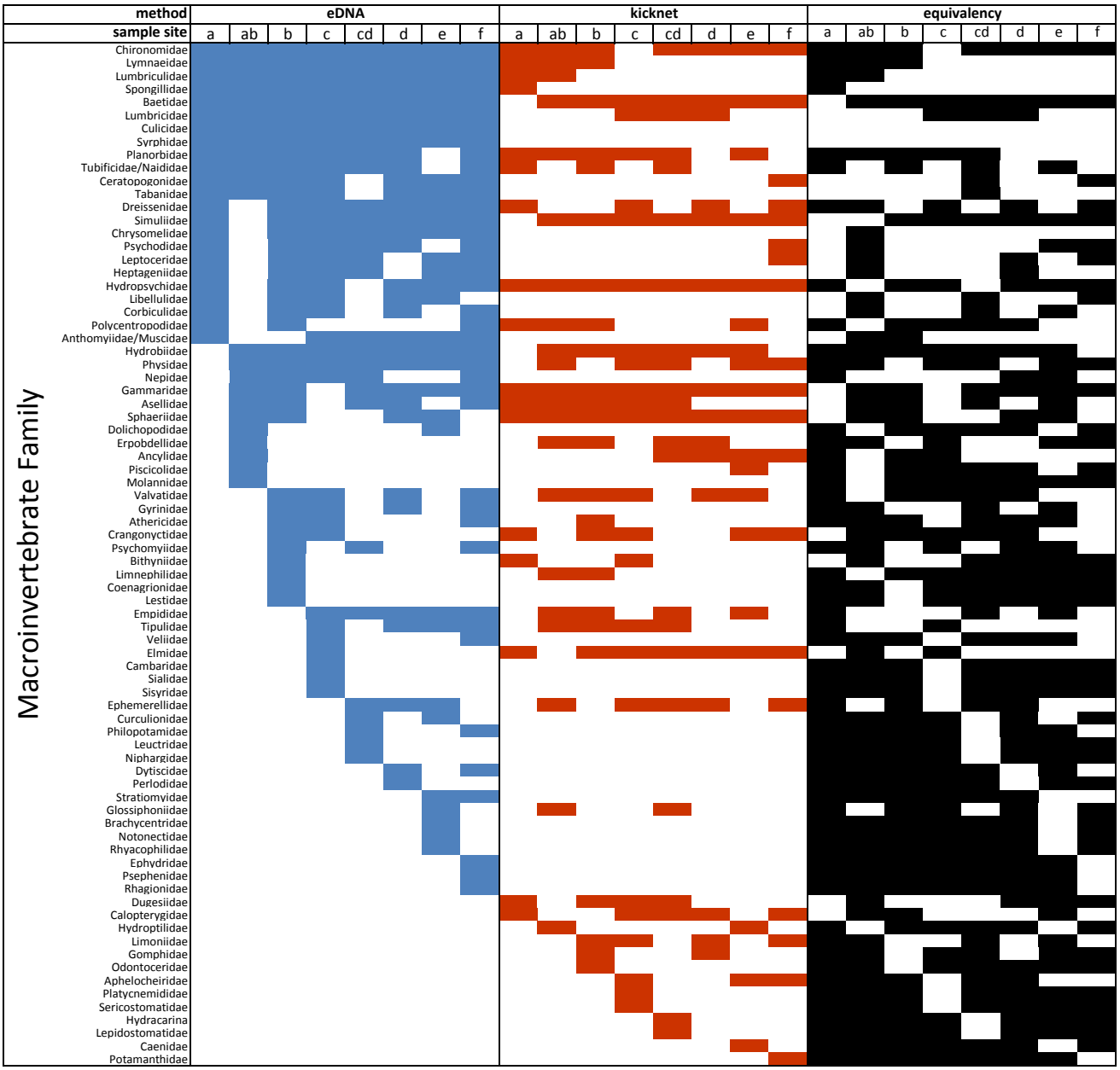



Figure S1. Heatmap illustrating each macroinvertebrate family, its detection across each site and the equivalency in detection for environmental DNA and kicknet sampling methods. Blue indicates presence for eDNA and red indicates presence for kicknet and white indicates not detected. For the equivalency, black indicates both were or were not detected at the same site, white indicates the family was detected by only one of the sampling methods.