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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

1 **Environmental DNA reveals that rivers are conveyer belts of biodiversity information**

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

3 [1] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic

4 Ecology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.

5 [2] Institute of Evolutionary Biology and Environmental Studies, University of Zurich,

6 Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

7 Running title: Rivers are conveyer belts of eDNA

8 Keywords: eDNA transport, metabarcoding, COI, river networks, community sampling, land

9 water interface, macroinvertebrates, biomonitoring

10 Type of Article: Letters

11 Abstract: **147**

12 Main text: **4879**

13 Text box: **360**

14 Number of references: **50**

15 Number of figures: **5**

16 Number of tables: **1**

17 *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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

18 **Abstract (150 max)**

19 DNA sampled from the environment (eDNA) is becoming a game changer for uncovering
20 biodiversity patterns. By combining a conceptual model and empirical data, we test whether
21 eDNA transported in river networks can be used as an integrative way to assess eukaryotic
22 biodiversity for large spatial scales and across the land-water interface. Using an eDNA
23 metabarcode approach we detected 300 families of eukaryotes, spanning 19 phyla across the
24 catchment of a river. We show for a subset of these families that eDNA samples overcome spatial
25 autocorrelation biases associated with classical community assessments integrating biodiversity
26 information over space. Additionally, we demonstrate that many terrestrial species can be
27 detected; thus revealing eDNA in river-water integrates biodiversity information across terrestrial
28 and aquatic biomes. Environmental DNA transported in river networks offers a novel and
29 spatially integrated way to assess total biodiversity for whole landscapes and will revolutionize
30 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.

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

33

34 ## Introduction

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

47 Organic matter in the form of DNA is produced from organisms and is also transported through
48 rivers via cells, tissues, gametes or organelles (termed environmental DNA 'eDNA'; Taberlet *et*
49 *al.* 2012; Deiner & Altermatt 2014; Turner *et al.* 2014). DNA can be isolated from these
50 organismal remains in the water, sequenced, and assigned back to the species of origin through
51 the method of eDNA metabarcoding (Taberlet *et al.* 2012; Ji *et al.* 2013). This elegant process of
52 collection and detection of a species' DNA is becoming highly valuable for biodiversity sampling
53 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.

54 2014; Kelly *et al.* 2014; Rees *et al.* 2014; Goldberg *et al.* 2015; Lawson Handley 2015). The
55 spatial signal of eDNA, however, has only recently been explored and shows that in rivers eDNA
56 can be transported over larger distances (Deiner & Altermatt 2014; Laramie *et al.* 2015).
57 Therefore, we hypothesized that rivers, through the aggregation and transport of eDNA, act as
58 conveyer belts of biodiversity information which can be used to estimate species richness over
59 large spatial scales and across the land-water interface (Fig. 1a, Box 1).

60 The relevance of biodiversity sampling with eDNA found in river water is twofold. First,
61 identifying biodiversity hotspots is invaluable for prioritizing global and regional conservation
62 efforts (Myers *et al.* 2000). Estimates of richness to establish a place as a hotspot or not have
63 suffered from being under-sampled (Noss *et al.* 2015). Under-sampling of biodiversity has many
64 causes (and consequences) in conservation and ecology in general, but mainly comes from
65 sampling methods used for estimating richness in a way that is aggregated with respect to space
66 (Gotelli & Colwell 2001). For example, a classical method for estimating richness of aquatic
67 macroinvertebrates is to use a kicknet method, where all individuals in a certain defined area of a
68 stream are sampled (Barbour *et al.* 1999). Many such samples are then taken and subsequently
69 pooled to represent richness for an entire river stretch or catchment. The pooling of spatially
70 autocorrelated samples causes an underestimation of biodiversity compared to if each species was
71 independently sampled. Because it is typically infeasible to sample all species independently,
72 statistical removal of the sampling artifact is recommended (Gotelli & Colwell 2001). Estimating
73 biodiversity through eDNA is, however, a potential way to sample each species independent of
74 space via their DNA becoming aggregated and transported through a river's network (Fig. 1, Box
75 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.

76 Second, river biodiversity is highly affected by environmental changes and tracking these
77 changes in space and time is of high interest (Heino *et al.* 2015). For example, the presence of
78 tolerant (or absence of sensitive) aquatic organisms is important for determining water quality
79 and has been used for over a century (Kolkwitz & Marsson 1909; Hilsenhoff 1988; Bonada *et al.*
80 2006). This valuable metric known as 'biomonitoring' is entering a new era and the demand in
81 its use has generated an undue burden on resource agencies. For example, the US, England, and
82 Switzerland combined spend approximately 117.4 to 206.6 million US dollars annually on
83 biomonitoring of aquatic systems (Table 1). This number represents only a small fraction of
84 what countries spend, but characterizes the value we place on using species in their environment
85 to monitor health of aquatic ecosystems. Biomonitoring is costly because of the different
86 methods and expertise required to collect information about each targeted taxonomic group
87 (e.g.,Table 1; Barbour *et al.* 1999; Stein *et al.* 2014). An eDNA method of biodiversity
88 monitoring in rivers has several advantages in that it is non-lethal, minimizes field time and can
89 sample diversity for all target groups across the tree of life with a single field sampling protocol.
90 Therefore, demonstrating the power of this tool to monitor biodiversity of important indicator
91 groups in rivers will provide a fast, non-lethal and inexpensive alternative tool compared with
92 classically used methods.

93 Whole community detection with environmental DNA has been called the 'game changer' for
94 biodiversity sampling (Lawson Handley 2015) and in this study we move this idea from theory
95 into practice. We developed a conceptual model (Fig. 1; Box 1) and test the hypothesis that
96 transported eDNA in rivers can be used in an unprecedented way to assess biodiversity of all
97 eukaryotes. We additionally validate the ability of this method to assess the globally important
98 macroinvertebrate communities and show that estimates of richness from eDNA compared with
99 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.

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

103

104 **Materials and Methods**

105 *eDNA sampling, amplification and next generation sequencing*

106 Water samples were collected from eight sites along the Glatt river network, a subcatchment of
107 the Rhine River in Switzerland (Fig. 2). The study sites were chosen because they represent
108 nodes in the river network where water from the major subcatchment tributaries combine and
109 flow into the mainstem of the river Glatt. They also have a known history of monitoring
110 macroinvertebrates for the past 15 years (AWEL 2012). At each site, DNA was isolated from
111 between 840 to 900 mL of river water sampled. Method for sampling, capture and extraction of
112 DNA followed that of Deiner *et al.* (2015), where the capture method of filtration was coupled
113 with a Phenol-Chloroform Isoamyl DNA extraction. Strict adherence to contamination control
114 was followed using a controlled lab for eDNA isolation and pre-PCR preparations (Deiner *et al.*
115 2015). Three independent extractions of 280 to 300 mL were carried out and then pooled to equal
116 DNA captured and purified from 840 to 900 mL of water. Total volume of water filtered for each
117 extraction replicate depended on the suspended solids in the sample of which clogged the filter.
118 Water for this study was collected minutes prior to collecting aquatic macroinvertebrates using a
119 classical sampling method (kicknet, for description see below and AWEL 2012; Altermatt *et al.*
120 2013) and therefore allowed for a comparison between the kicknet and eDNA methods for the
121 detection of aquatic macroinvertebrate communities within the same watershed at the same time
122 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.

123 Polymerase chain reactions (PCRs) were carried out for the target gene, cytochrome oxidase I
124 (COI), using the universal COI primers (Folmer *et al.* 1994) on pooled eDNA extractions for
125 each of the eight sites and amplified a fragment of 658 base pairs (bp) excluding primer
126 sequences. PCRs were carried out in 15 µL volumes with final concentrations of 1x supplied
127 buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland), 1000 ng/µL bovine serum albumin
128 (BSA; New England Biolabs, Inc., Ipswich, MA, USA), 0.2 mMol dNTPs, 2.0 mMol $MgCl_2$,
129 0.05 units per µL Taq DNA polymerase (Faststart TAQ, Roche, Inc., Basel, Switzerland), and
130 0.50 µMol of each forward and reverse primer (Folmer *et al.* 1994). 2 µL of the pooled extracted
131 eDNA was added. The thermal-cycling regime was 95 °C for 4 minutes, followed by 35 cycles of
132 95 °C for 30 seconds, 48 °C for 30 seconds and 72 °C for 1 minute. A final extension of 72 °C
133 for 5 minutes was carried out and the PCR was cooled to 10 °C until removed and stored at –20
134 °C until confirmation of products occurred. PCR products were confirmed by gel electrophoresis
135 on a 1.4% agarose gel stained with GelRed (Biotium Inc., Hayward, CA USA). Three PCR
136 replicates were performed on each of the eight eDNA samples from our study sites and products
137 from the three replicates were pooled. Negative filtration, extraction and PCR controls were used
138 to monitor any contamination during the molecular workflow and were also replicated three
139 times. Reactions were then cleaned using AMPure XP beads following recommended
140 manufacturer's protocol except 0.6 x bead concentration was used instead of 1.8 x based on
141 recommended protocol for fragment size retention of >500 base pairs (p. 31, Nextera XT DNA
142 96 kit, Illumina, Inc., San Diego, CA, USA). We quantified each pooled reaction using the Qubit
143 (1.0) fluorometer following recommended protocols for the dsDNA High Sensitivity DNA Assay
144 which has an accuracy for double stranded DNA between 0.005-0.5 pg/µL (Agilent
145 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.

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

160 *Bioinformatic analysis*

161 Run quality was assessed using FastQC version 0.10.1. Forward and reverse sequences were
162 merged with a minimum overlap of 25 bp and minimum length of 100 bp using SeqPrep (St. John
163 2011). Sequences that could not be merged were excluded from further analysis. Merged
164 sequences with quality scores less than a mean of 25 where removed. Merged sequences were
165 then de-replicated by removing exact duplicates, were de-noised using a sequence identity
166 threshold of 99%, and were quality trimmed left and right by 28 bp using PrinSeq Lite version
167 0.20.3 (Schmieder & Edwards 2011). Subsequent sequences were then chimera checked using
168 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.

169 taxonomically identified using customized Blast searches against the NCBI non-redundant
170 nucleotide database using the package blast 2.2.28, build on March 12, 2013 16:52:31 (Benson *et*
171 *al.* 2012). Taxonomic assignment of a sequence was done using the best blast hit based on a bit
172 score calculated using the default blastn search of a -3 penalty for a nucleotide mismatch and a
173 reward of +1 for a nucleotide match. Sequences that did not match eukaryotes, were below
174 90.0% sequence similarity, had less than 100 bp overlap with query, had a taxonomic name not
175 assigned below the level of family, matched best with unknown environmental samples and/or
176 had a bit score less than 100 were excluded from biodiversity detection analysis for al sites.
177 These parameters were used because they removed likely taxonomic identification errors or
178 exclude data that was unidentified at the family level used for analysis (Deiner *et al.* 2013; Deiner
179 *et al.* 2015). All raw sequences reads were deposited in NCBI's Sequence Read Archive
180 (*accession numbers pending*).

181 After identification of sequences with the NCBI nucleotide sequence database, each uniquely
182 identified taxon from any site (referred to as an operational taxonomic unit, OTU) was
183 geographically verified as known to be present in Switzerland to the lowest level of taxonomy, or
184 if no data was available for Switzerland, it was also considered present when the OTU was
185 known to be present in Austria, France, Germany, and Italy. We excluded a few (and very rare
186 cases) where it is known for sure that a species is not in Switzerland, but found in all four
187 neighboring countries. Geographic verification was done in consultation with 24 expert
188 taxonomists for various groups, primary literature and through database repositories as described
189 in Table S1. If the species could be confidently confirmed as being present in Switzerland or in
190 all four neighboring countries, their known habitat use was identified as being freshwater
191 (defined as having at least one life stage inhabiting water), or terrestrial (which included species
192 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.

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

199 *Kicknet sampling and identification*

200 Macroinvertebrates were detected using a standard kicknet sampling design described for federal
201 and cantonal guidelines in Switzerland (Stucki 2010; Altermatt *et al.* 2013). Briefly, we took
202 eight independent kicknet samples per site on October 29, 2012. Large inorganic and organic
203 debris was removed and samples were pooled into a single collection jar with 70% EtOH. Jars
204 were then stored at room temperature until morphological identification. This method and time
205 of year has been shown to reflect the different microhabitats and provides a robust presence
206 measure for many macroinvertebrates in Switzerland (Stucki 2010). Since eDNA has been
207 shown to decay over short time periods (days to a few months; Strickler *et al.* 2015), using a
208 single time point from a kicknet sample to compare with that of what is detected in the eDNA is
209 valid. However, it is known that kicknet samples taken at different times of year, such as in the
210 spring, can detect different species due to morphological constraints in the identification of
211 specimens at young life stages or just their physical presence in the water due to timing for their
212 life cycle (Stucki 2010). Specimens from each site were sorted to the lowest taxonomic level
213 possible (family, genus or species level) using dichotomous keys agreed upon by the Swiss
214 Federal Office of the Environment (Stucki 2010). Specimens that could not be identified to at
215 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.

216 *Comparison of eDNA and kicknet macroinvertebrate detection*

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

235 To test for a spatial autocorrelation in community dissimilarity (β-diversity, using the Jaccard
236 dissimilarity index) and between sampling locations we used a Mantel's test with 9999
237 permutations. The Jaccard measure of β-diversity was used as it has been shown to estimate
238 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.

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

241

242 **Results**

243 *All Eukaryotes*

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

252 Of the remaining 335 from the total of 1758 OTUs, 286 were identified to families not known to
253 occur in Switzerland. More than half of these (175) were assigned to families known only to
254 contain species living in the marine environment (e.g., sponges, red algae, worms and snails,
255 etc.), 81 are known to be terrestrial, 24 are known to inhabit freshwater but not from Switzerland,
256 and six could not have their habitat identified due to the identification only being at the level of
257 family. Lastly, 59 OTUs were identified to families for which we could find no distribution data
258 using our verification methods. Assignment statistics of sequence length and percent sequence
259 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.

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

262 *Macroinvertebrates*

263 Of the 300 families considered detected for eukaryotes, 65 of them are used in the Swiss
264 biomonitoring program (Fig. S2; Stucki 2010). Thirteen additional families were detected by
265 kicknet samples only, totaling 78 macroinvertebrate families detected among our sampling sites
266 of the river Glatt. From eDNA we recovered between 23 to 40 families at each site (Fig. S2).
267 With the classical kicknet method we sampled between 17 to 24 families at each site (Fig. S2).
268 Several families were detected by eDNA only have been previously sampled with kicknets at the
269 same sites in the more than 15 years of monitoring (i.e., Coenagrionidae, Gyrinidae,
270 Psychomyiidae, Rhyacophilidae; AWEL 2012). Of the total 78 families detected, 33 (42 %) were
271 detected by both methods, and often at the same location (Fig. S2). Of the remaining 45 (58 %)
272 of families, 32 were only detected with eDNA and 13 where only detected with the kicknet
273 sample. Eight of these 13 families were detected in the eDNA dataset, but did not meet
274 bioinformatics thresholds for filtering assignment values (e.g., where below a 90 % sequence
275 similarity, data not shown). Of the remaining five families, three had sufficient sequence data on
276 GenBank for identification (e.g., more than 100 sequences from COI across a broad array of
277 genera within each family), but two had very low representation (Potamanthidae and
278 Aphelocheiridae) with six and 22 sequences respectively. Of the 32 families detected with
279 eDNA, three are known to occur in the lake Greifensee, which feeds into the river Glatt (i.e.,
280 Cambaridae, Notonectidae, and Sialidae; AWEL 2012), but are not known from the river Glatt.
281 Local family richness (α-diversity) increased for sites sampled at more downstream positions in
282 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.

283 Cumulative family richness (γ-diversity) increased as a function of cumulative catchment area
284 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 =
285 0.92, p = 0.0001 kicknet) (Fig. 5b). The slopes of the family-area relationship were not different
286 ($F_{1,12}$ = 1.113, p = 0.312), however the y-intercept was higher for eDNA compared with kicknet
287 ($F_{1,13}$ = 41.61, p < 0.0001) (Fig. 5b). β-diversity in the form of community dissimilarity did not
288 increase as a function of distance for eDNA (r = 0.02, p = 0.43), whereas for kicknet sampling we
289 observed a positive correlation in β-diversity as a function of distance between sampling sites (r =
290 0.52, p = 0.003) (Fig. 5c).

291

292 **Discussion**

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

302 Our model (Box 1) identifies three important messages for the utility of eDNA as a genomic tool
303 for biodiversity assessment. First, eDNA detection of species from river water decouples the
304 presence of a species from its physical location in a habitat through the downstream transport.
305 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.

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

317 Our data substantiate our conceptual predictions (Box 1), and by comparing eDNA with kicknet
318 samples at each site, we highlight several important factors that illustrate both the power and
319 current limitations of eDNA for biodiversity assessment. Many families of macroinvertebrates
320 were detected at each site by both methods and have a great degree of overlap in which sites
321 families were co-detected. For all sites, however, eDNA recovered more macroinvertebrate
322 families compared to kicknet samples. We hypothesize this is likely due to the integrated signal
323 from transported DNA which is evident by the fact that community composition does not change
324 much (i.e., β-diversity remaining constant over distance), compared with kicknet estimated β-
325 diversity which increased over the same river distance. This difference means that the two
326 sampling methods give different information at the same site. Classical sampling methods give
327 information that is very localized, whereas the eDNA metabarcode method in rivers measures
328 presence of species for larger spatial scales. This novel finding is of great importance because in
329 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.

330 hotspots or conservation preserves (Noss *et al.* 2015), or entire river catchments (Sheldon *et al.*
331 2012). As scaling up of the classical community sampling method will likely always
332 underestimate diversity (Gotelli & Colwell 2001), eDNA offers an empirical method to overcome
333 this limitation and is an unparalleled way to estimate richness for larger areas.

334 Much of the current degradation of river habitat is at the catchment scale and cannot be attributed
335 to a single point source (Vörösmarty *et al.* 2010). Biomonitoring currently relies on the costly
336 sampling of macroinvertebrates across many sites to understand ecosystem health of rivers (Table
337 1; Stein *et al.* 2014). Therefore, an eDNA signal of macroinvertebrates can be used to measure
338 more precisely diversity of a catchment with much less sampling effort. In contrast,
339 understanding local changes in richness at a restoration site may still require classical sampling
340 with kicknets. Interestingly, however, transport distances of eDNA match the scale at which local
341 species' pools are recognized to be important for recolonization of restored patches in a river
342 system (0-5 km) (Sundermann *et al.* 2011). Therefore, eDNA could be used as a way to measure
343 the species' pool available for recolonization. A tool such as this can aid in prioritizing river
344 restoration efforts by identifying regions which have high recolonization potential of target
345 species and possibly set expectations for the magnitude of change expected for restoration sites
346 already in recovery.

347 Our results also identify a way of empirically measuring transport of community eDNA in rivers.
348 Our analysis of β-diversity in this study system shows that community eDNA is likely
349 transported and detected over a scale greater than 12 km. This has been previously shown for
350 two targeted species that live only in lake Greifensee (mussel and a waterflea), where their DNA
351 could be detected at sites d and f respectively (Fig. 2; Deiner & Altermatt 2014). To determine
352 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.

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

364 There are, however, important current limitations of the eDNA metabarcode method and are
365 related to factors such as the importance of primer or marker choice and the biodiversity detected
366 as a function of the reference data with its annotations available for identification of sequences
367 (Deagle *et al.* 2014; Elbrecht & Leese 2015). For example: fish, flatworms, and diatoms in this
368 dataset are underrepresented to what we know occurs in this system. This is most likely due to
369 the choice of primers, the genetic maker and potentially the reference database used for
370 identifying sequences. The primers used in this study are the universal Folmer primers for the 5'
371 end of COI (Folmer *et al.* 1994) and it is known that these primers do not amplify DNA from fish
372 and flatworms very well (Ivanova *et al.* 2007; Moszczynska *et al.* 2009; respectively).
373 Additionally, for diatoms it is known that COI is not the best genetic marker suitable for species
374 level identification (Zimmermann *et al.* 2015). Therefore, it is clear that more than one maker
375 and/or primer set is needed to capture biodiversity for the tree of life adequately (Gibson *et al.*
376 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.

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

380 **Conclusions**

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

388

389 **Acknowledgements**

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

398

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.

399 **Literature cited**

400 1.

401 Altermatt, F. (2013). Diversity in riverine metacommunities: a network perspective. *Aquat Ecol*,
402 47, 365-377.
403 2.

404 Altermatt, F., Seymour, M. & Martinez, N. (2013). River network properties shape α-diversity
405 and community similarity patterns of aquatic insect communities across major drainage
406 basins. *Journal of Biogeography*, 40, 2249-2260.
407 3.

408 AWEL (2012). Zürcher Gewässer 2012: Entwicklung - Zustand - Ausblick. Amt für Abfall,
409 Wasser, Energie und Luft, Zürich, Switzerland. (Available from:
410 http://www.awel.zh.ch/internet/baudirektion/awel/de/wasserwirtschaft/veroeffentlichunge
411 n.html). Abfall Wasser, Energie, und Luft.
412 4.

413 Barbour, M.T., Gerritsen, J., Snyder, B. & Stribling, J. (1999). Rapid Bioassessment Protocols
414 For Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and
415 Fish. (ed. USEPA) Washington, D.C., p. 339.
416 5.

417 Benson, D.A., Karsch-Mizrachi, I., Clark, K., Lipman, D.J., Ostell, J. & Sayers, E.W. (2012).
418 GenBank. *Nucleic Acids Research*, 40, D48-D53.
419 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.

420 Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M. *et al.* (2014).
421 Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology*
422 *& Evolution*, 29, 358-367.
423 7.

424 Bonada, N., Prat, N., Resh, V.H. & Statzner, B. (2006). Developments in aquatic insect
425 biomonitoring: a comparative analysis of recent approaches. *Annu. Rev. Entomol.*, 51,
426 495-523.
427 8.

428 Cardoso, P., Borges, P.A. & Veech, J.A. (2009). Testing the performance of beta diversity
429 measures based on incidence data: the robustness to undersampling. *Diversity and*
430 *Distributions*, 15, 1081-1090.
431 9.

432 Clift, P.D. & Blusztajn, J. (2005). Reorganization of the western Himalayan river system after
433 five million years ago. *Nature*, 438, 1001-1003.
434 10.

435 Cristescu, M.E. (2014). From barcoding single individuals to metabarcoding biological
436 communities: towards an integrative approach to the study of global biodiversity. *Trends*
437 *in Ecology & Evolution*, 29, 566-571.
438 11.

439 Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014). DNA
440 metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match.
441 *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.

442 12.

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

445 13.

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

449 14.

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

453 15.

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

456 16.

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

460 17.

461 Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for
462 amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan
463 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.

464 18.

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

469 19.

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

473 20.

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

476 21.

477 Heino, J., Melo, A.S., Bini, L.M., Altermatt, F., Al-Shami, S.A., Angeler, D.G. *et al.* (2015). A
478 comparative analysis reveals weak relationships between ecological factors and beta
479 diversity of stream insect metacommunities at two spatial levels. *Ecology and Evolution*,
480 5, 1235-1248.

481 22.

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

484 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.

485 Ivanova, N.V., Zemlak, T.S., Hanner, R.H. & Hebert, P.D. (2007). Universal primer cocktails for
486 fish DNA barcoding. *Molecular Ecology Notes*, 7, 544-548.
487 24.

488 Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H. *et al.*
489 (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams.
490 *Molecular Ecology Resources*, 15, 216-227.
491 25.

492 Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A. *et al.* (2013). Reliable,
493 verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16,
494 1245-1257.
495 26.

496 Kelly, R.P., Port, J.A., Yamahara, K.M., Martone, R.G., Lowell, N., Thomsen, P.F. *et al.* (2014).
497 Harnessing DNA to improve environmental management. *Science*, 344, 1455-1456.
498 27.

499 Kolkwitz, R. & Marsson, M. (1909). Ökologie der tierischen Saprobien. Beiträge zur Lehre von
500 der biologischen Gewässerbeurteilung. *Internationale Revue der Gesamten Hydrobiologie*
501 *und Hydrographie*, 2, 126-152.
502 28.

503 Laramie, M.B., Pilliod, D.S. & Goldberg, C.S. (2015). Characterizing the distribution of an
504 endangered salmonid using environmental DNA analysis. *Biological Conservation*, 183,
505 29-37.
506 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.

507 Lawson Handley, L. (2015). How will the 'molecular revolution'contribute to biological
508 recording? *Biological Journal of the Linnean Society*, In press, 10.1111/bij.12516.
509 30.

510 Mari, L., Casagrandi, R., Bertuzzo, E., Rinaldo, A. & Gatto, M. (2014). Metapopulation
511 persistence and species spread in river networks. *Ecology Letters*, 17, 426-434.
512 31.

513 Moszczynska, A., Locke, S.A., McLAUGHLIN, J.D., Marcogliese, D.J. & Crease, T.J. (2009).
514 Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic
515 trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths.
516 *Molecular Ecology Resources*, 9, 75-82.
517 32.

518 Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A. & Kent, J. (2000).
519 Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.
520 33.

521 Noss, R.F., Platt, W.J., Sorrie, B.A., Weakley, A.S., Means, D.B., Costanza, J. *et al.* (2015). How
522 global biodiversity hotspots may go unrecognized: lessons from the North American
523 Coastal Plain. *Diversity and Distributions*, 21, 236-244.
524 34.

525 QGIS Core Development Team (2015). QGIS Geographic Information System. Available from
526 http://qgis.osgeo.org. Open Source Geospatial Foundation Project.
527 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.

528 R Development Core Team (2013). R: A language and environment for statistical computing.
529 Version 3.0.1. R Foundation for Statistical Computing Vienna, Austria.
530 36.

531 Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R. & Gough, K.C. (2014). REVIEW:
532 The detection of aquatic animal species using environmental DNA–a review of eDNA as
533 a survey tool in ecology. *Journal of Applied Ecology*, 51, 1450-1459.
534 37.

535 Rodríguez-Iturbe, I. & Rinaldo, A. (1997). *Fractal river basins: chance and self-organization*.
536 Cambridge University Press, Cambridge, UK.
537 38.

538 Schmieder, R. & Edwards, R. (2011). Quality control and preprocessing of metagenomic
539 datasets. *Bioinformatics*, 27, 863-864.
540 39.

541 Sheldon, F., Peterson, E.E., Boone, E.L., Sippel, S., Bunn, S.E. & Harch, B.D. (2012).
542 Identifying the spatial scale of land use that most strongly influences overall river
543 ecosystem health score. *Ecological Applications*, 22, 2188-2203.
544 40.

545 St. John, J. (2011). SeqPrep. Available <https://github.com/jstjohn/SeqPrep>. Github.
546 41.

547 Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E. & Zakharov, E.V. (2014). Is DNA barcoding
548 actually cheaper and faster than traditional morphological methods: results from a survey
549 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.

550 42.

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

554 43.

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

559 44.

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

562 45.

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

565 46.

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

569 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.

570 Vörösmarty, C.J., McIntyre, P., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P. *et al.*
571 (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555-
572 561.
573 48.

574 Ward, N.D., Keil, R.G., Medeiros, P.M., Brito, D.C., Cunha, A.C., Dittmar, T. *et al.* (2013).
575 Degradation of terrestrially derived macromolecules in the Amazon River. *Nature*
576 *Geoscience*, 6, 530-533.
577 49.

578 Willett, S.D., McCoy, S.W., Perron, J.T., Goren, L. & Chen, C.-Y. (2014). Dynamic
579 reorganization of river basins. *Science*, 343, 1248765.
580 50.

581 Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. (2015). Metabarcoding
582 vs. morphological identification to assess diatom diversity in environmental studies.
583 *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.

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

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

589 $$\alpha_{classical}^{x} = \alpha_{real}^{x} \cdot \delta^{classical} \quad (1)$$

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

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

602 $$\alpha_{eDNA}^{catchment} = \left(\alpha_{real}^{x} + \sum \alpha_{real}^{y} \cdot N_{y} \cdot \beta_{y} \cdot \beta_{x,y} \cdot \tau_{x,y}\right) \cdot \delta^{eDNA} \quad (2)$$

603 with $\alpha_{eDNA}^{catchment}$ as the integrated measure of catchment α-diversity (see also Fig. 1). The sum
604 captures the information integrated by the riverine system for all locations *y* (Strahler stream
605 order) upstream of the sampling location *x*. The local diversity at a site of Strahler stream order *y*
606 has to be weighted according to Horton's Law to capture the number of streams of this Strahler
607 stream order (N_{y}; Rodríguez-Iturbe & Rinaldo 1997) as well as by the Strahler stream order-
608 characteristic β-diversity (β_{y}). The estimate of catchment-level biodiversity increases with
609 increasing β-diversity between the sampling point and all upstream locations ($\beta_{x,y}$) as well as
610 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.

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

613

614 **Figure legends**

615 **Figure 1**

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

626 **Figure 2**

627 Study area and location of sampling sites where environmental DNA samples and classical
628 sampling methods were carried out. The direction of flow for the river Glatt is northwest (blue
629 arrow). The main stem of the river originates from the outflow of lake Greifensee. Colored
630 regions represent the catchment upstream of each sampling point. Letters are used to indicate the
631 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.

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

634 **Figure 3**

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

640 **Figure 4**

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

645 **Figure 5**

646 Difference of benthic macroinvertebrate family richness and community dissimilarity estimated
647 between environmental DNA and kicknet sampling. a) α-diversity measured at each sample site
648 in the river network for macroinvertebrate families. b) Log_{10} transformed taxon-area relationship
649 for eDNA and kicknet samples. Slopes of lines are not significantly different ($p = 0.312$),
650 however the y-intercept is significantly higher for eDNA compared with kicknet ($p < 0.001$)
651 indicating that eDNA samples a greater amount of diversity. c) Correlation of community
652 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.

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

655 **Table**

656 Table 1: Annual bioassessment costs in millions of dollars for freshwater resources. Information
657 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)

658 Sources: [1] Stein *et al.* 2014; [2] Richard Walmsley, Forestry Commission personal communication;
659 [3] Markus Wüest, Federal Office for the Environment FOEN personal communication.

660

661 **Supplementary material**

662 Figure S1

663 Length of alignment and similarity between sequence and reference data for operational
664 taxonomic units at the family level of taxonomy. (a) Confirmed known geographic presence of
665 families in Switzerland or from all four neighboring countries (Austria, France, Germany, and
666 Italy) (b) confirmed not to be present. Most unconfirmed taxonomic sequence assignments at the
667 family level fell at two of the many thresholds set for accepting an assignment as valid (i.e., 100
668 bp in length and 90% identical). The outliers are color coded as follows: Euphausiidae (blue),
669 Erobdellidae (orange), Naccariaceae (red), Rytididae, Calyptraeidae, Naticidae, Ampullariidae,
670 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.

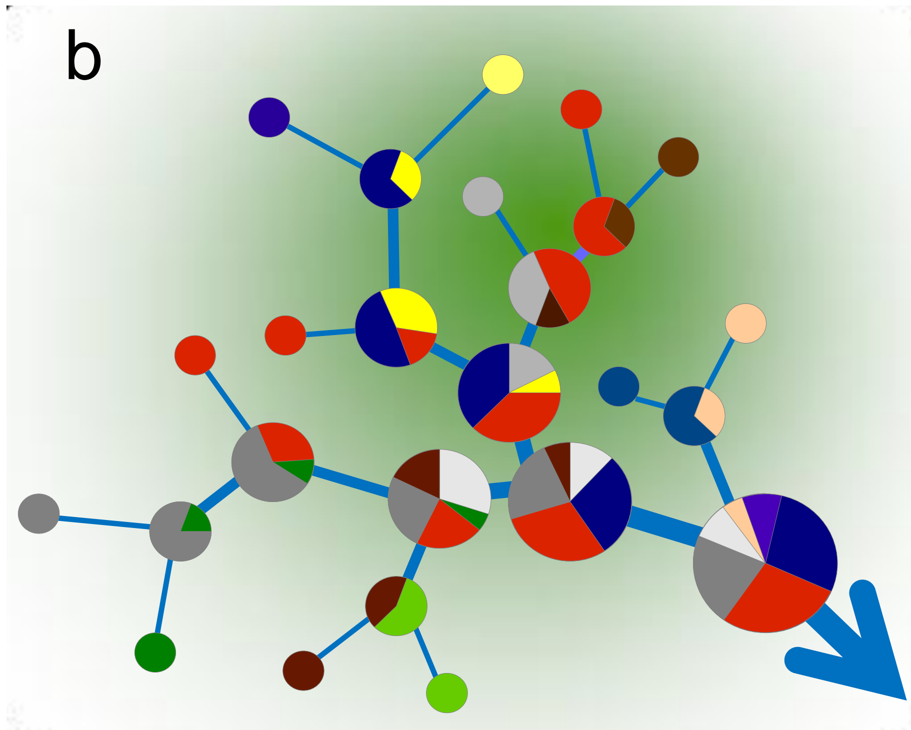
671 Figure S2

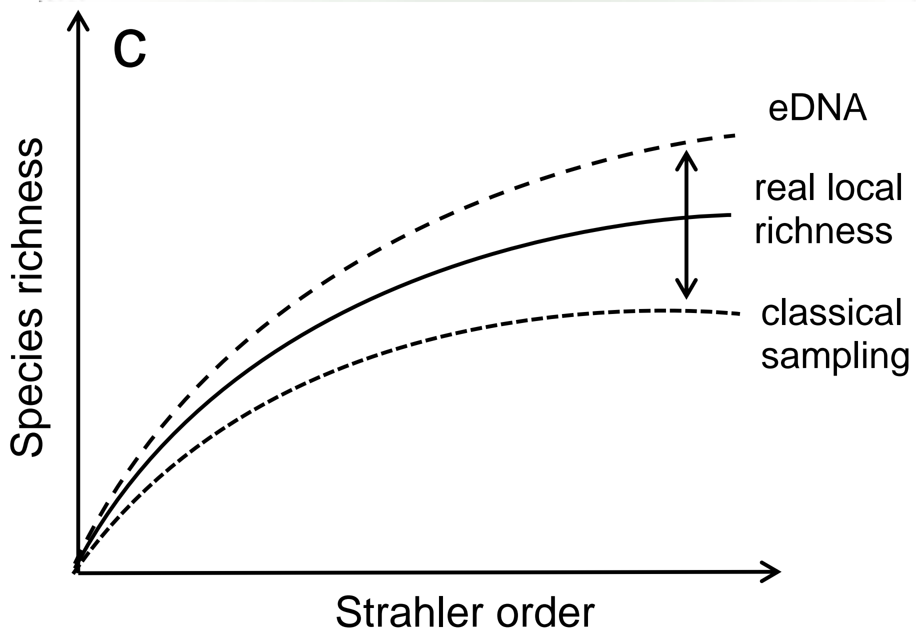
672 Heatmap illustrating each macroinvertebrate family, its detection across each site and the
673 equivalency in detection for environmental DNA and kicknet sampling methods. Blue indicates
674 presence for eDNA, red indicates presence for kicknet and white indicates not detected. For the
675 equivalency,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

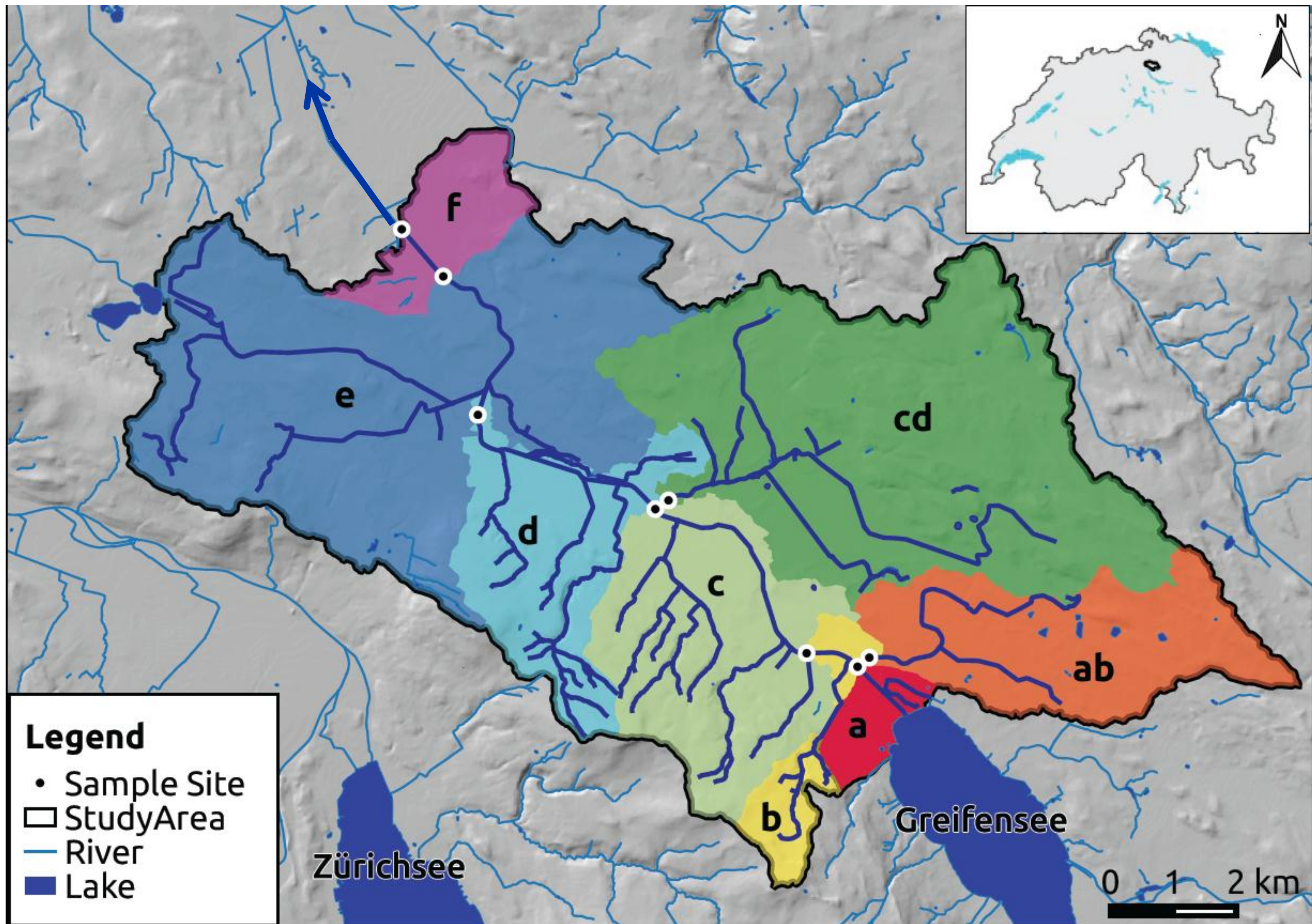
677 Table S1

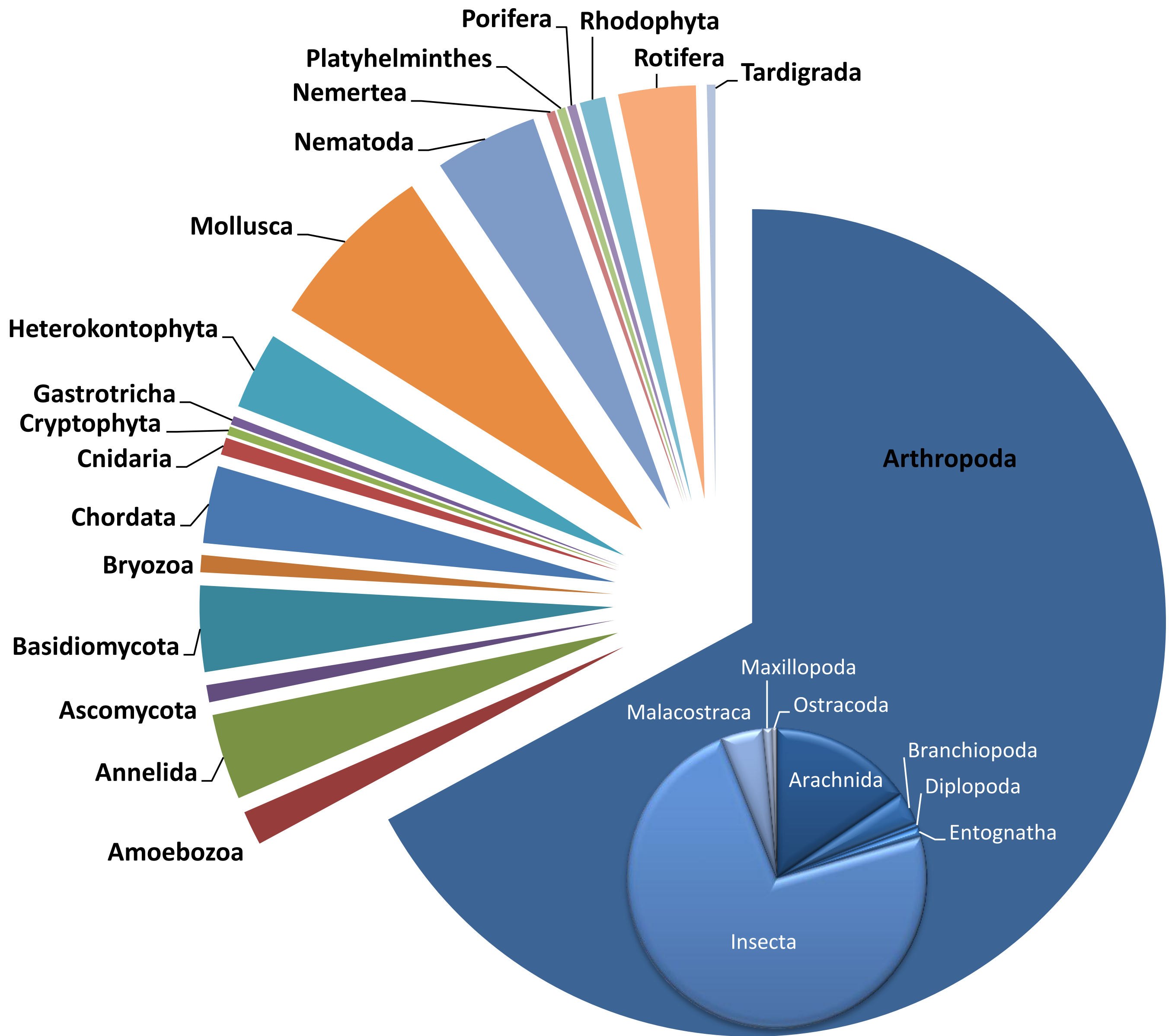
678 Assignment statistics and taxonomic assignments for sampled operational taxonomic units from
679 environmental DNA in the river Glatt, Switzerland. At the family level and below, taxonomic
680 assignment of cells were color coded for the geographic verification process. Green colored cells
681 were confirmed as known to be present, red colored cells were confirmed as not known to be
682 present, and white colored cells had no data available for verification. Not available is
683 abbreviated as “na”. Geographic verification of each taxa was done by personal communication
684 with expert taxonomists, primary literature, and database repositories as indicated in confirmation
685 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.

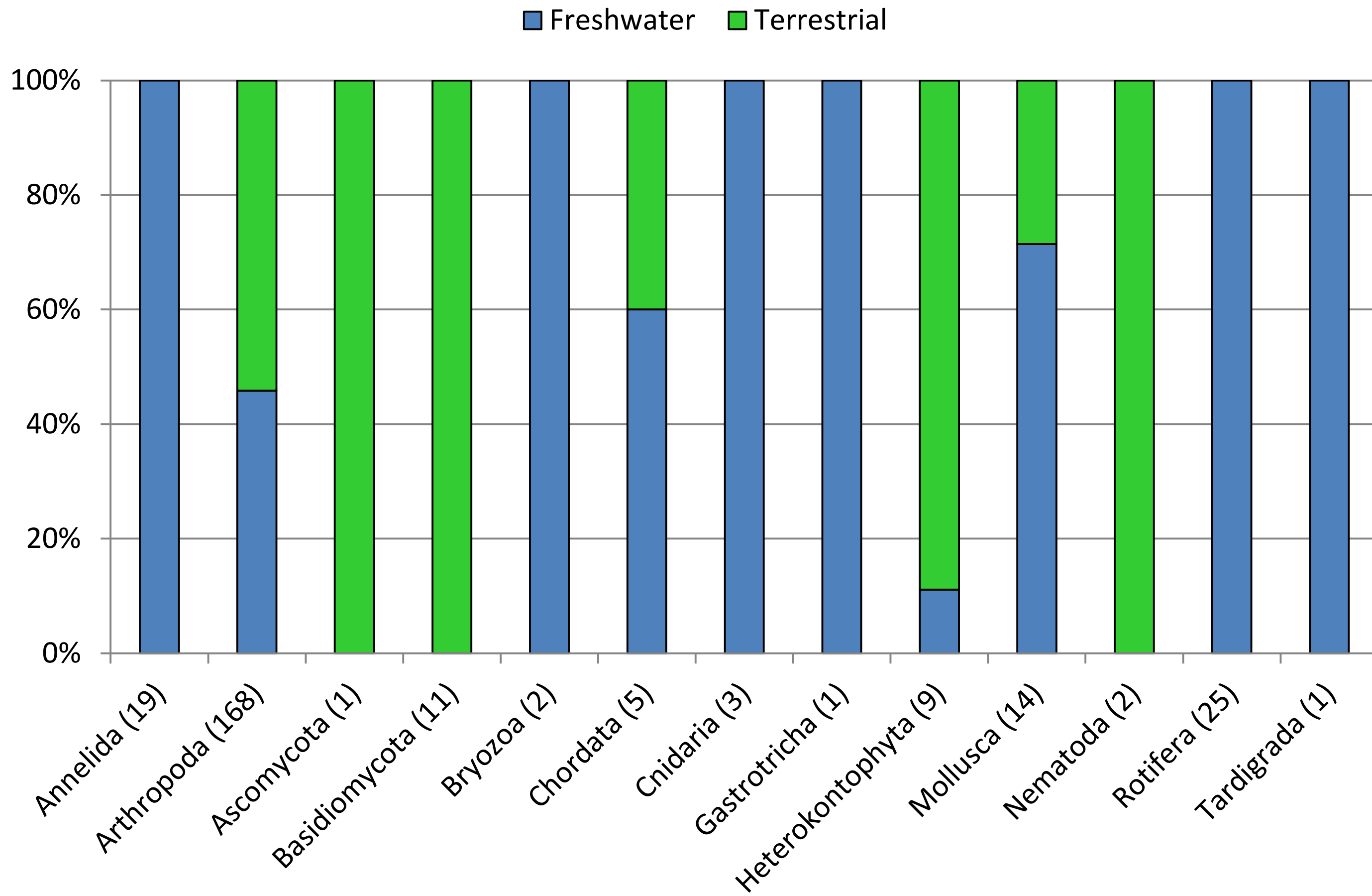




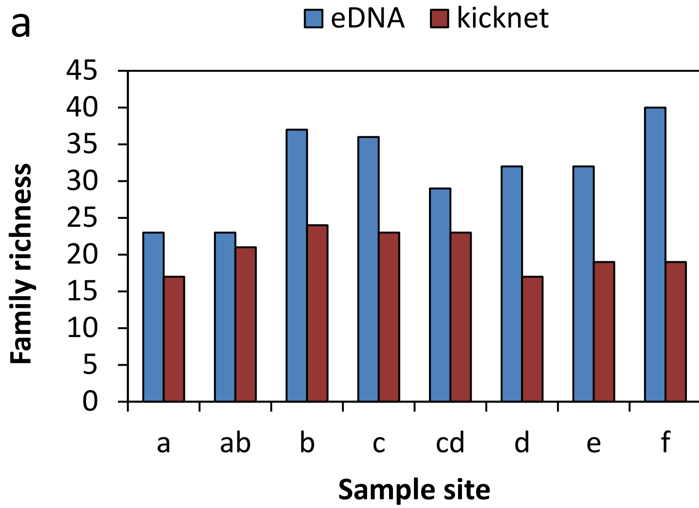

f
e
cd
d
c
ab
a
b
Greifensee
Zürichsee
N
0
1
2 km
Legend
Sample Site
StudyArea
River
Lake

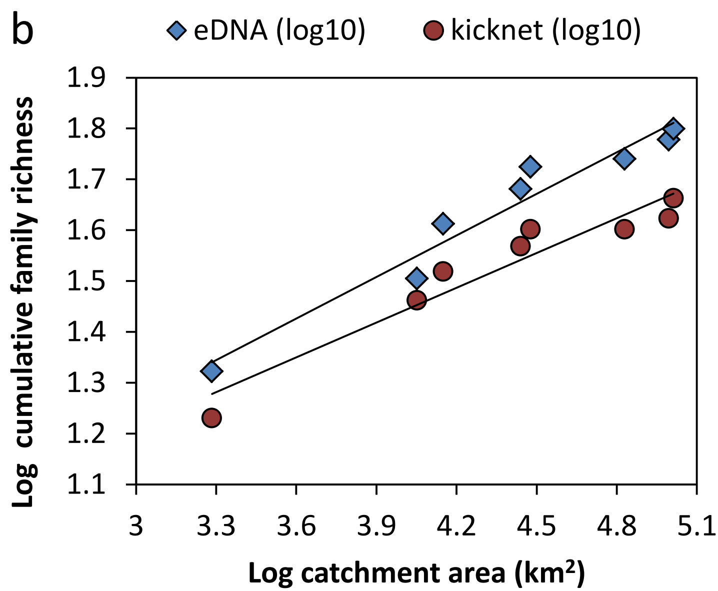


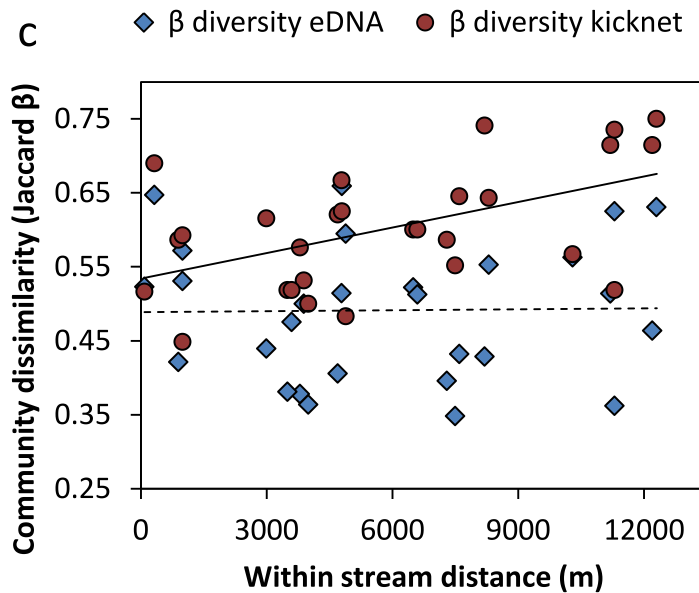
Porifera
Rhodophyta
Platyhelminthes
Rotifera
Tardigrada
Nemertea
Nematoda
Mollusca
Heterokontophyta
Gastrotricha
Cryptophyta
Cnidaria
Chordata
Bryozoa
Basidiomycota
Ascomycota
Annelida
Amoebozoa
Arthropoda
Maxillopoda
Malacostraca
Ostracoda
Branchiopoda
Arachnida
Diplopoda
Entognatha
Insecta

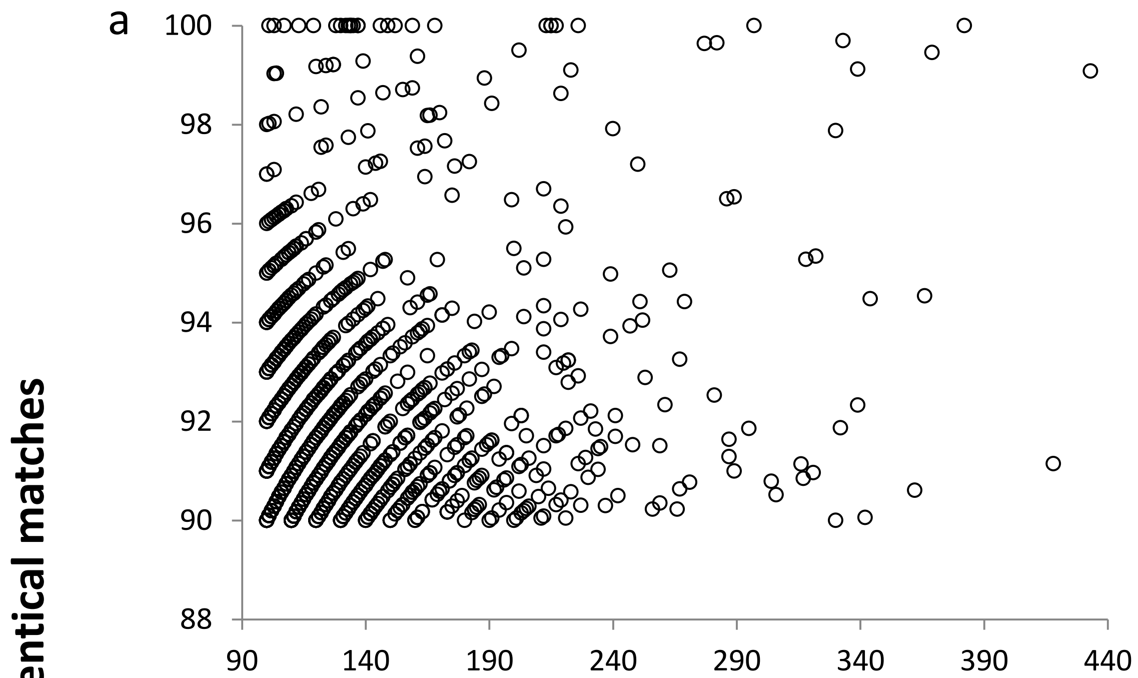


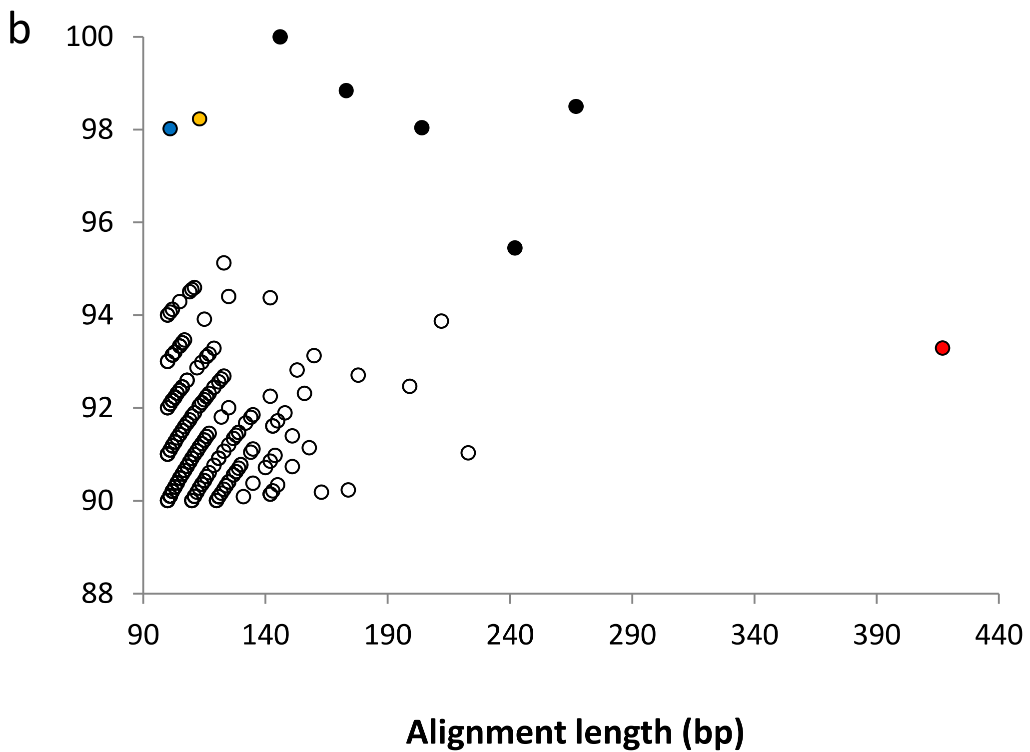
Freshwater
Terrestrial
100%
80%
60%
40%
20%
0%
Annelida (19)
Arthropoda (168)
Ascomycota (1)
Basidiomycota (11)
Bryozoa (2)
Chordata (5)
Cnidaria (3)
Gastrotricha (1)
Heterokontophyta (9)
Mollusca (14)
Nematoda (2)
Rotifera (25)
Tardigrada (1)


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
100
98
96
94
92
90
88
90
140
190
240
290
340
390
440
Percent of identical matches


b
100
98
96
94
92
90
88
90
140
190
240
290
340
390
440
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 aCC-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, 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.