Environmental DNA reveals that rivers are conveyer belts of biodiversity information

Journal Article

Author(s):
Deiner, Kristy (iD); Fronhofer, Emanuel A.; Mächler, Elvira; Walser, Jean-Claude; Altermatt, Florian

Environmental DNA reveals that rivers are conveyer belts of biodiversity information

Kristy Deiner[1,2], Emanuel A. Fronhofer[1,3], Elvira Mächler[1,3], Jean-Claude Walser[4] & Florian Altermatt[1,3]

DNA sampled from the environment (eDNA) is a useful way to uncover 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 broad spatial scales and across the land–water interface. Using an eDNA metabarcode approach, we detect 296 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 the classical community assessments by integrating biodiversity information over space. In addition, we demonstrate that many terrestrial species are detected; thus suggesting eDNA in river water also incorporates biodiversity information across terrestrial and aquatic biomes. Environmental DNA transported in river networks offers a novel and spatially integrated way to assess the total biodiversity for whole landscapes and will transform biodiversity data acquisition in ecology.

[1] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland. [2] Department of Biological Sciences, University of Notre Dame, 290B Galvin Life Sciences, Notre Dame, Indiana 46556, USA. [3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [4] Swiss Federal Institute of Technology (ETH), Zürich, Genetic Diversity Centre, CHN E 55 Universitätstrasse 16, 8092 Zürich, Switzerland. Correspondence and requests for materials should be addressed to K.D. (email: alpinedna@gmail.com) or to F.A. (email: florian.altermatt@eawag.ch).

While rivers cover <1% of the landmasses on earth, they are invaluable for biodiversity and ecosystem services, such as drinking water and energy production[1]. 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[2,3]. 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[4]. 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[5]. River networks additionally play an important role in shaping patterns of genetic and species diversity for many organisms across the landscape by dictating dispersal pathways[6,7].

Organic matter in the form of DNA is produced from organisms and is also transported through rivers via cells, tissues, gametes or organelles, and is termed environmental DNA (eDNA)[8–10]. 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[10,11]. This elegant process of collection and detection of a species DNA is becoming highly valuable for sampling biodiversity in ecology and conservation[10–17]. The spatial signal of eDNA, has only recently been explored and shows that in rivers eDNA can be transported over larger distances[8,18]. Therefore, we hypothesized that rivers, through the aggregation and transport of eDNA, act as conveyer belts of biodiversity information that can be used to estimate species richness over broad spatial scales and potentially across the land–water interface.

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[19]. Estimates of richness to establish a place as a hotspot or not have suffered from being under-sampled[20]. Under-sampling of biodiversity has many causes (and consequences) in conservation and ecology in general, but mainly comes from the sampling methods used for estimating richness in a way that is aggregated with respect to space[21]. For example, a classical method for estimating richness of aquatic macroinvertebrates in rivers is to use a kicknet method, where all individuals in a certain defined area of a stream are collected in a net[22]. 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 such as this causes an underestimation of biodiversity compared with if each species was independently sampled. Because it is typically infeasible to sample all species independently, statistical removal of the sampling artefact is recommended[21]. Estimating biodiversity through eDNA is a potential way to sample each species independent of space via their DNA becoming aggregated and transported through a river's network.

Second, an eDNA method of biodiversity monitoring in rivers has several advantages in that it is non-lethal for most classically sampled taxonomic groups, minimizes habitat disruption and can assess diversity across the tree of life with a single-field sampling protocol making it extremely cost effective. 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 eDNA has been called the 'game changer' for the biodiversity sampling[16], and in this study, we move this idea from theory into practice. We test the hypothesis that transported eDNA in rivers can be used in an unprecedented way to assess biodiversity of eukaryotes. We validate the ability of an eDNA metabarcoding method *in vitro* and *in situ* to assess globally important macroinvertebrate communities and produce taxonomic richness estimates of which reflect the biodiversity of a rivers' catchment. Lastly, we demonstrate that a large number of eukaryotic phyla from both aquatic and terrestrial taxa can be assessed from eDNA in river water and provide support for the hypothesis that rivers are conveyer belts of biodiversity information for landscapes.

Results

eDNA detection of metazoan eukaryotes. We detected a total of 296 families that span 19 eukaryotic phyla from the Glatt river catchment in Switzerland (Fig. 1). All families were independently geographically verified as known to occur in Switzerland or the four neighbouring countries (Fig. 2a; Supplementary Data 1). The majority of the families detected were Arthropoda ($N = 196$). Diversity in number of families detected was not proportional to read count and smaller organisms represented a much higher proportion of the sequences obtained (Rotifera; Fig. 2b). For example, two species in the phylum Rotifera accounted for 39% (92,907 sequences) of our data set. The majority of families were represented by >10 sequences ($N = 140$; Supplementary Data 1). The largest data reduction step in the bioinformatic workflow was in linking a taxonomic name with our sequences (Supplementary Fig. 1; step E), resulting in only 4% (240,340 sequences) of acquired sequences that could be used for inferences in our study (Table 1). Of the sequences that were identified to species and that were independently geographically verified as occurring in Switzerland, many are terrestrial ($N = 255$; Fig. 3; Supplementary Data 2).

eDNA detection of macroinvertebrates. Of the 296 families detected with eDNA for eukaryotes, 65 are used in the Swiss biomonitoring program[23]. Thirteen additional families were detected by kicknet samples only, totalling 78 macroinvertebrate families detected among our sampling sites of the river Glatt (Supplementary Fig. 2). From eDNA, we recovered between 23 and 40 families at each site (Supplementary Fig. 2). With the classical kicknet method, we sampled 17–24 families at each site (Supplementary Fig. 2). Of the total 78 families detected, 33 were detected by both methods, and often at the same location (Supplementary Fig. 2). Of the remaining 45 families, 32 were only detected with eDNA and 13 where only detected with the kicknet sample. Eleven of these 13 families only detected with the kicknet were detected in the eDNA data set, but did not meet bioinformatic thresholds used for filtering assignment values (for example, where below a 90% sequence similarity or an alignment length <100 base pairs (bp), Supplementary Table 1). The two undetected families (Potamanthidae and Aphelocheiridae) likely had insufficient sequence data on GenBank for the identification of their DNA sequence from eDNA (Supplementary Table 1). Of the 32 families only detected with eDNA, 8 have been found in previous sampling events over the 18 years of monitoring (Supplementary Table 2) and an additional 2 (Molannidae, Notonectidae) are known to occur in lake Greifensee, which feeds into the river Glatt, but are not known from the river Glatt (Supplementary Table 2).

Family richness (α-diversity) increased as a function of cumulative catchment area sampled for eDNA, whereas this was not observed for kicknet samples ($F_{1,6} = 5.45$, $P = 0.058$, $r^2 = 0.95$, eDNA; $F_{1,6} = 0.0001$, $P = 0.99$, $r^2 = 0.92$, kicknet; Fig. 4a). The slopes of the family–area relationship were different (slope$_{kicknet} = 0.0006$; slope$_{eDNA} = 0.1077$; $F_{1,12} = 29.87$, $P = 0.0001$), and the y intercept was higher for eDNA compared with kicknet ($F_{1,13} = 25.99$, $P = 0.0002$; Fig. 4a). β-diversity in the
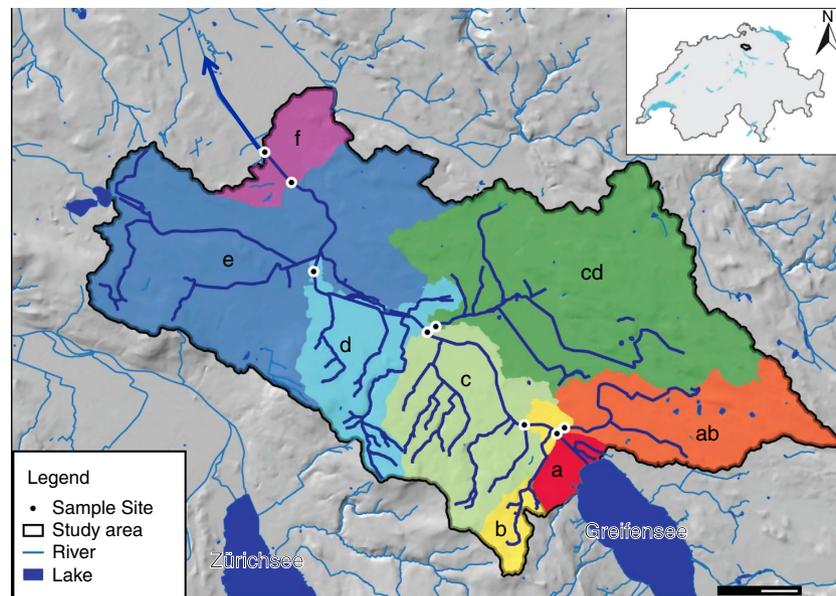
Figure 1 | Study area and location of sampling sites where environmental DNA samples and classical sampling methods were carried out. The direction of flow for the river Glatt is northwest (blue arrow). The main stem of the river originates from the outflow of lake Greifensee. Scale bar, 2 km. Coloured 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 'ab' and 'cd'. Sources for GIS data were from Swisstopo (DHM25, Gewässernetz Vector 25) and reprinted with permission.

form of community dissimilarity did not increase as a function of distance for eDNA ($r = 0.02$, $P = 0.44$), whereas for kicknet sampling, we observed an increase in dissimilarity (β-diversity) as a function of distance between sampling sites ($r = 0.52$, $P = 0.005$; Fig. 4b).

***In vitro* test using a mock community**. In total, we recovered 57,641 sequences from the mock community after the bioinformatic filtering and these sequences were identified to 25 of the 33 invertebrate taxa included in the mock community (Table 1; Supplementary Table 3). Of these sequences, 99.97% were correctly assigned to one of these taxa included in the mock community (Supplementary Table 4; Supplementary Table 5). The number of incorrectly assigned sequences was 0.03% (20/57,641) and all of these sequences belonged to two taxa (Tabanidae and Leuctridae; Supplementary Table 4). This resulted in a false-positive rate of 8% (2/25). Increasing the stringency of our bioinformatics thresholds set for accepting an assignment to a level that removes all false positives in the mock community (for example, increasing assignment similarity to >92%) introduces a false absences of 16% (4/25) in the mock community, that is, the exclusion of taxa that were present in the mock community, but had an assignment similarity <92% (Supplementary Table 4). Using a similarity threshold of 92% on our data derived from eDNA in water samples to account for possible false presences, however, did not change our main findings (Supplementary Fig. 3).

Discussion

We demonstrate that rivers, through their collection and transport of eDNA, 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 sample location (Fig. 4a). This increased sensitivity is hypothesized to come from the process of transport of DNA through the network of a river.

Transport of DNA through a river network decreases the biases associated with the spatial autocorrelation (or limited scale of inference) inherent to the classical kicknet community sampling. The evidence from our work supports that eDNA found in rivers is a spatially integrated measure of biodiversity and this finding offers ecologists a new and unprecedented tool to sample landscape biodiversity with less sampling effort and potentially estimate richness of eukaryotic communities across biomes.

We hypothesize the following conceptual model as the explanation of our data. Typically, sampling methods for communities only capture a fraction of local α-diversity due to the imperfect detection and sampling bias (Fig. 5):

$$\alpha_{\text{classical}}^{x} = \alpha_{\text{real}}^{x} \cdot \delta^{\text{classical}} \qquad (1)$$

with $\alpha_{\text{classical}}^{x}$ representing the measured α-diversity at a spatial location x in a river network using classical sampling methods, α_{real}^{x} as the real α-diversity at this location and $\delta^{\text{classical}}$ as the detection rate of the sampling method. 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 underestimation of the real local richness[21].

Riverine networks have the potential to collect this information for us[2,3] 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[24] and transport of eDNA by the flow of water[8] are the mechanisms that enable an eDNA metabarcoding method to estimate the catchment-level biodiversity, while sampling at only one or very few locations:

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

with $\alpha_{\text{eDNA}}^{\text{catchment}}$ as the integrated measure of catchment α-diversity (Fig. 5). 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
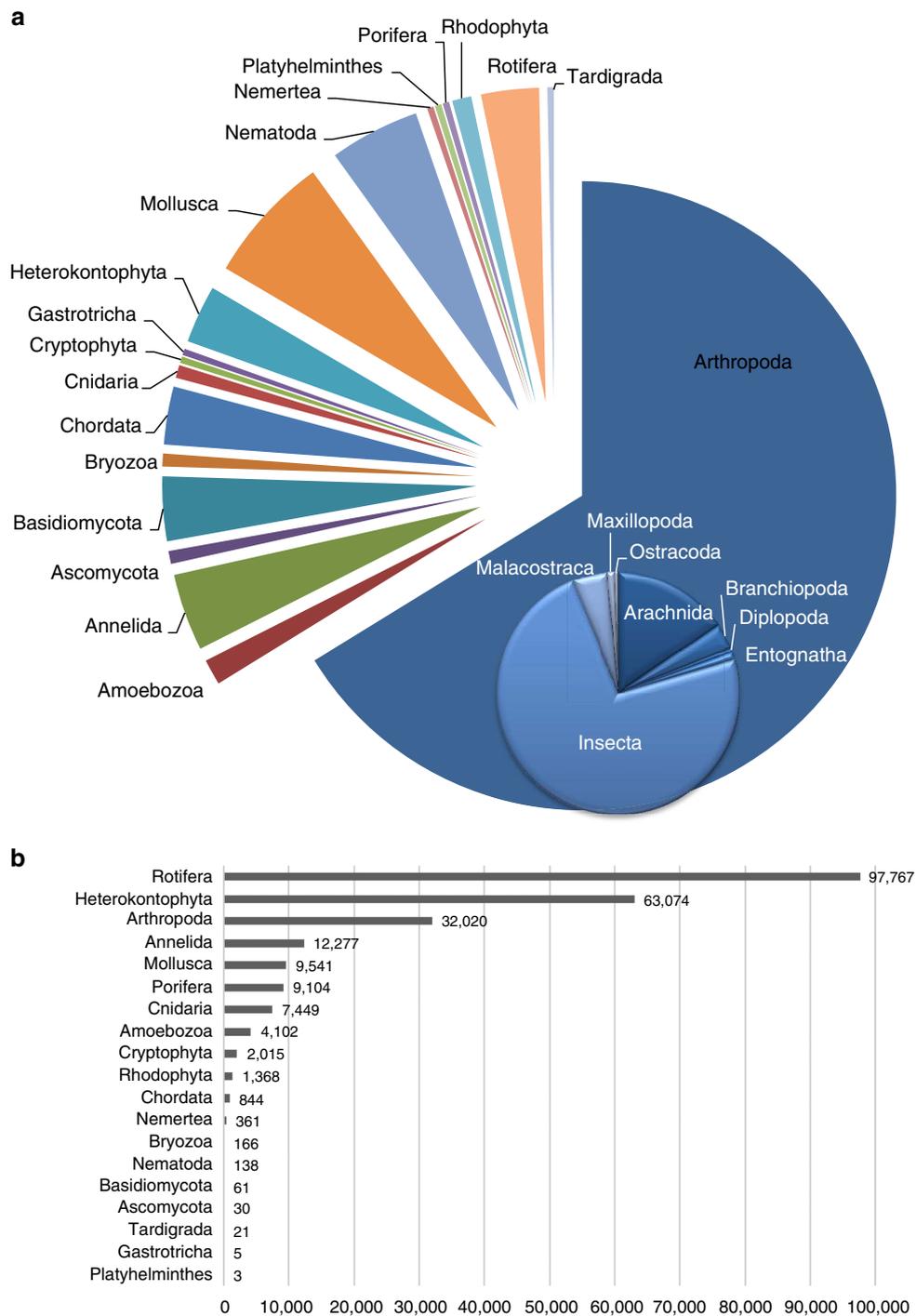
Figure 2 | Total eukaryotic diversity detected from the river Glatt using environmental DNA metabarcoding. (a) The number of families per phylum ($N = 296$) sampled and confirmed as known to be present in Switzerland or known from all four neighbouring countries (Austria, France, Germany and Italy). The inset further breaks the most abundant phylum (Arthropoda) into the number of families sampled by class ($N = 196$). **(b)** Grey bars depict the number of sequences receiving a taxonomic identification at the family level for each phylum.

Horton's Law to capture the number of streams of this Strahler stream order $(N_y)^3$, 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 the eDNA specific detection probability (δ^{eDNA}) tends to be high as, in principle, only very few DNA molecules are needed for successful detection.

Our conceptual model identifies three important messages for the utility of eDNA as a genomic tool for the biodiversity assessment. First, eDNA detection of species from river water decouples the presence of a species from its physical location in a habitat through downstream transport. Transport distance in empirical systems has been measured between 240 m and 12 km (refs 8,25), and thus allows for the increased sensitivity in the detection of patchily or elusively distributed species. In addition, transport of eDNA allows for richness estimates with less

Table 1 | Sequences remaining after each bioinformatic filtering step and taxonomic assignment.

Sample name	Site name	Raw read count	(A) Merged	(B) Quality filter and trim	(C) Mapping	(D) Chimera removal	(E) Taxonomic assignment	Average % ± s.d. of identical matches	Average ± s.d. reference alignment length
02-01_S12	a	859,156	496,181	389,859	251,400	250,725	32,673	97.7 ± 2.7	204 ± 85
08-01_S18	ab	754,781	479,497	433,506	151,744	151,422	21,989	94.1 ± 2.7	195 ± 77
07-01_S17	b	587,445	368,095	304,991	168,274	167,810	38,316	98.1 ± 2.5	205 ± 81
03-01_S13	c	458,635	294,565	247,769	144,850	144,453	37,659	97.9 ± 2.6	213 ± 85
01-01_S11	cd	770,055	512,041	411,234	176,874	176,231	36,509	94.9 ± 2.8	195 ± 74
04-01_S14	d	664,475	444,870	392,865	187,002	186,445	21,783	96.6 ± 3.2	197 ± 77
05-01_S15	e	673,495	364,002	330,412	172,045	171,389	20,376	95.8 ± 3.4	192 ± 76
06-01_S16	f	1,238,917	740,208	621,210	311,638	310,375	31,035	96.1 ± 3.3	193 ± 76
Total for samples		6,006,959	3,699,459	3,131,846	1,563,827	1,558,850	240,340	96.6 ± 3.2	200 ± 80
Mock community		421,868	92,934	82,218	67,983	67,983	57,641	94.0 ± 3.6	300.8 ± 38

Letter in parentheses of columns refer to bioinformatic filtering step with details given in Methods and Supplementary Fig. 1. Assignment statistics are averages across all assignments for each site ± s.d. Alignment length is in base pairs.
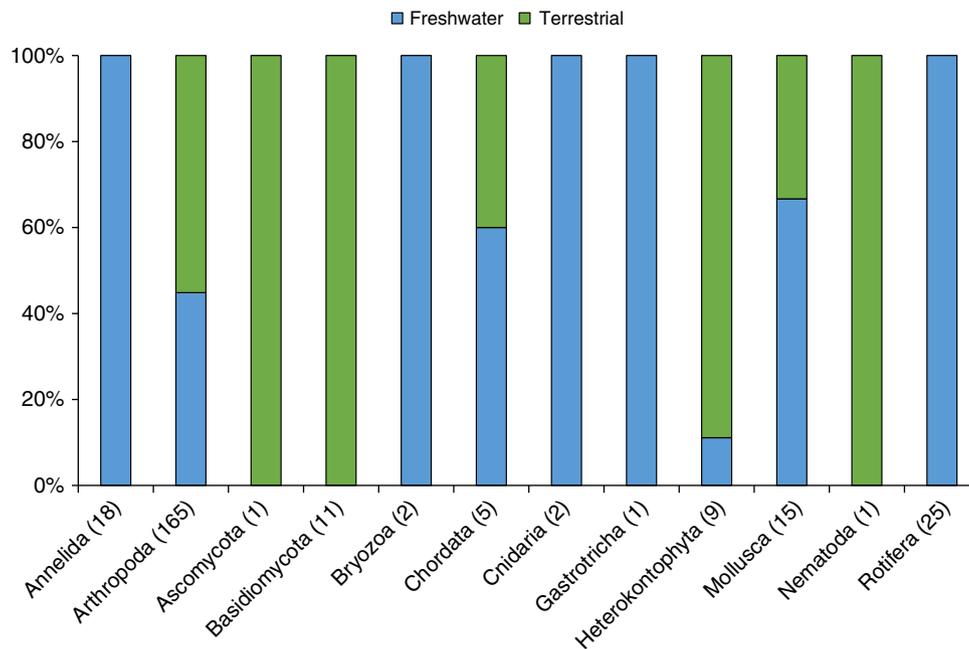

Figure 3 | Percent terrestrial or freshwater species for the subset of each phylum detected in eDNA. Species were confirmed as known to be present in Switzerland or known from all four neighbouring countries (Austria, France, Germany and Italy; $N = 255$). Number in brackets indicates the number of species confirmed for each phylum.

sampling effort because of the integrated signal over space. Second, eDNA will likely represent a sample of higher diversity compared with the 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 the classical sampling methods. Namely, eDNA detection of species should be interpreted as an integrated signal of presence and the spatial scale that is relevant is determined based on the potential transport distance for a system. Thus, our model suggests that eDNA in rivers is an efficient tool for broad scale biodiversity assessments, and depending on the distance between water samples, less authoritative for very localized richness estimates.

Our data comparing eDNA with kicknet samples at each site highlights several important factors that illustrate both the power
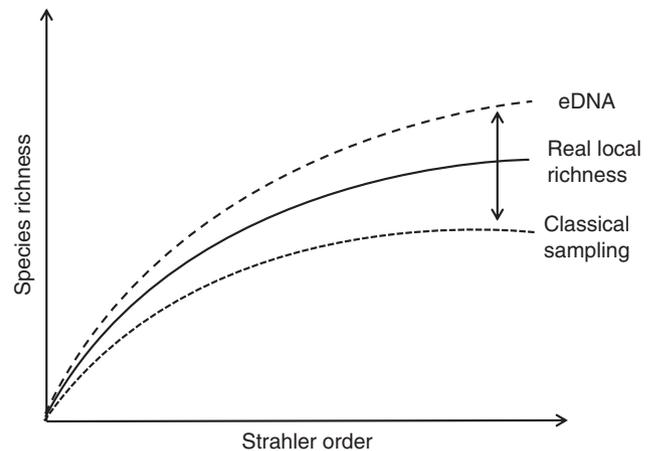
and current limitations of using eDNA for the biodiversity assessment. Many families of macroinvertebrates were detected at each site by both methods and have a great degree of overlap, in which sites families were co-detected. For all sites, however, eDNA recovered more macroinvertebrate families compared with 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 (that is, β-diversity remaining constant over distance), compared with kicknet estimated β-diversity that increased over the same river distance in our study area. This difference means that the two sampling methods give different information at the same site. Classical sampling methods give information that is localized, whereas the eDNA metabarcoding method in rivers measures presence of species on broader spatial scales. Scaling up of the classical community sampling method will likely always
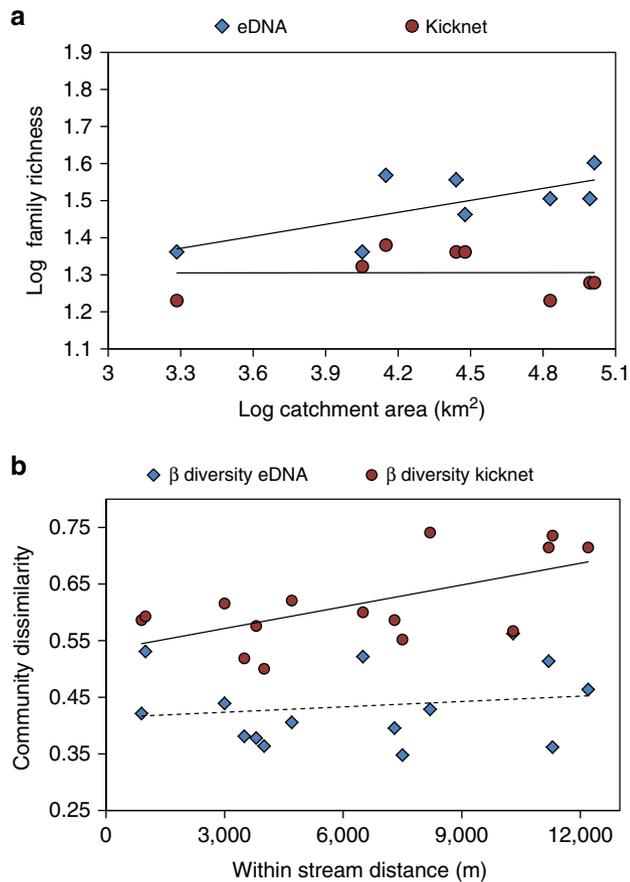
Figure 4 | Difference of benthic macroinvertebrate family richness and community dissimilarity estimated between the environmental DNA and kicknet sampling. (**a**) α-diversity measured at each site and $\log_{10}$ transformed taxon–area relationship for eDNA and kicknet samples. Slopes of line and the y intercept are significantly higher for eDNA compared with kicknet ($P < 0.0001$, $P = 0.0002$, respectively), indicating that eDNA samples a greater amount of diversity over a greater area compared with kicknets. (**b**) Correlation of community dissimilarity with along-stream geographic distances between sample sites. Solid line for kicknet β-diversity indicates a significant positive relationship with stream distance ($P = 0.005$). Dashed line for eDNA β-diversity indicates no significant relationship with distance ($P = 0.44$).

Figure 5 | Conceptual model of environmental DNA dynamics in a hypothetical river network. While classical sampling only detects a fraction of real local diversity due in part to spatial autocorrelation, 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.

underestimate diversity[21], eDNA offers an empirical method to overcome this limitation and is an unparalleled way to estimate richness for larger areas. This novel finding is of great importance because in many cases estimating diversity for a large area is the goal, such as that for biodiversity hotspots[19], conservation preserves or entire river catchments[26].

Much of the current degradation of river habitat is at the catchment scale and cannot be attributed to a single point or source[1]. Biomonitoring currently relies on the costly and lethal sampling of macroinvertebrates across many sites to understand the ecosystem health of rivers[27] and tracking these changes in space and time is of high interest[28]. Biomonitoring is entering a new era and the demand in its use has generated an undue burden on resource agencies. For example, the United States, England and Switzerland combined spend ~117.4–206.6 million US dollars annually on biomonitoring of aquatic systems (Supplementary Table 6). This number represents only a small fraction of what countries spend on biomonitoring at more local levels, but characterizes the value we place on using species in their environment to monitor the health of aquatic ecosystems.

Biomonitoring is costly because of the different methods and expertise required to collect information about each targeted taxonomic group (for example, Supplementary Table 6)[22,27]. An eDNA signal of macroinvertebrates can be used to estimate more accurately diversity of a catchment with much less sampling effort and would therefore decrease the cost associated with biomonitoring when the goal is to measure the ecosystem health on large scales in river systems.

By contrast, understanding local changes in richness at a restoration site, for example, may still require classical sampling with kicknets. Interestingly, however, transport distances of eDNA are on a similar scale at which local species' pools are recognized to be important for recolonization of restored patches in a river system (0–5 km)[29]. Therefore, eDNA could be used as a way to measure the species' pool available for recolonization. The scale of inference for eDNA, however, can be >5 km due to long-distance transport within basins and between basins due to other vectors such as faeces from predators. The complementarity between methods will aid in prioritizing river restoration efforts by identifying regions that 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 <12 km. To determine the scale of transport for community eDNA in a river system, one subsequently needs to detect the scale at which there is a positive spatial autocorrelation with β-diversity (for example, Fig. 4b). 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 of eDNA due to ultraviolet, pH and temperature[30], as well as discharge rates[25]. 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 an eDNA estimate of β-diversity and river distance between sampling points; however, an *in situ* test can be performed and the scale of transport for community eDNA can be uncovered for any system,

and can be repeatedly measured across time to test if eDNA transport distance is stable in a system.

There are still important current limitations of the eDNA metabarcoding method. These challenges are related to factors, such as the importance of primer or marker choice, the amplicon sequence length and the biodiversity detected, as a function of the reference data available for identification of sequences[31,32]. For example: fish, flatworms and diatoms in our data set are underrepresented to what we know occurs in the studied system. This is most likely due to the choice of primers, the genetic marker and the reference database. The primers used in this study are the universal Folmer primers for the 5′ end of cytochrome *c* oxidase I (COI)[33], and it is known that these primers do not amplify DNA from fish and flatworms very well[34,35]; respectively. In addition, for diatoms it is known that COI is not the best genetic marker suitable for species level identification[36]. Therefore, it is clear that more than one marker and/or primer set is needed to adequately assess biodiversity for the tree of life[37]. However, use of an eDNA metabarcode method does not require additional sampling in the field. Rather it creates a single-field sampling method, whereby careful amplification of many genetic markers in the laboratory will enable an integrated detection for total biodiversity from a single sample[38].

An additional challenge faced by the further application of this approach is the need for continued development of diverse, but curated databases with taxonomically classified sequences. Our mock community analysis corroborated that we had a high accuracy in assignment of sequences, when compared with the reference sequence generated from the DNA used for the mock community (96.4–99.9% similarity). The variance in assignment accuracy increased to (90.1–99.8%) when compared with NCBI's nucleotide database. The gaps in NCBI's nucleotide database for targeted groups, such as macroinvertebrates used for biomonitoring will need to be augmented and assessed before the tool can be more widely applied in management. Because of uncertainly in the database, we removed many sequences that could not be confidently assigned to the family taxonomic level. At the current filtering level, we are already accepting a false-absence rate of 14% (Supplementary Table 1). Reducing our data set further using more stringent criteria increased type II error by creating many more false absences for taxa we actually collected in our kicknet samples at the time of sampling (Supplementary Note 1). Therefore, at this stage in deployment of an eDNA metabarcoding approach, researchers need to strive to reduce false absences and false presences, while understanding that the tool is in rapid development and false error rates for macroinvertebrates related to this method are still unknown beyond the estimates given here. In comparison with morpho-logical assessments of macroinvertebrates at the family level, however, identification error is reported to range between 22.1% (ref. 39) and 33.8% (ref. 40), suggesting that the only alternative used in regulatory monitoring settings already has a high false-positive-presence/absence rate. Most of the sequences from our data set were removed because the taxonomic assignment failed. The solution for this is to increase the deposition of sequences in curated databases such as The Barcode of Life Database[41] through continued collaboration between molecular ecologists and taxonomists. Digitizing specimens in the form of sequences is an essential step that will vastly improve our ability to accurately identify DNA found in the environment.

We have demonstrated that rivers convey, through the collection and transport of eDNA, an unprecedented amount of information on biodiversity in landscapes. Our study shows 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.

Methods

eDNA sampling and library preparation for next generation sequencing. Water samples were collected from eight sites along the Glatt river network, a subcatchment of the Rhine river in Switzerland (Fig. 1). 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 main stem of the river Glatt. They also have a known history of monitoring macroinvertebrates for the past 15 years[42]. At each site, DNA was isolated from between 840 and 900 ml of river water sampled. Method for sampling, capture and extraction of DNA followed that of Deiner *et al.*[43], 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[43]. Three independent extractions of 280–300 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 before collecting aquatic macroinvertebrates, using a classical sampling method kicknet, for description see below and (refs 24,42), 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.

PCRs were carried out for the target gene, COI, using the standard COI primers[33] on pooled eDNA extractions for each of the eight sites and amplified a fragment of 658 bp excluding primer sequences. PCRs were carried out in 15 µl volumes with final concentrations of 1 × supplied buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland), 1,000 ng µl^{-1} bovine serum albumin (New England Biolabs, Inc., Ipswich, MA, USA), 0.2 mMol dNTPs, 2.0 mMol MgCl$_2$, 0.05 U µl^{-1} Taq DNA polymerase (Faststart TAQ, Roche, Inc., Basel, Switzerland) and 0.50 µMol of each forward and reverse primer[33]. A measure of 2 µl of the pooled extracted eDNA was added. The thermal-cycling regime was 95 °C for 4 min, followed by 35 cycles of 95 °C for 30 s, 48 °C for 30 s and 72 °C for 1 min. A final extension of 72 °C for 5 min 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 × bead concentration was used instead of 1.8 × based on recommended protocol for fragment size retention of >500 bp (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 that has an accuracy for double stranded DNA between 0.005 and 0.5 pg µl^{-1} (Agilent Technologies, Santa Clara, CA, USA). At this step negative controls showed no quantifiable DNA and we therefore did not process them further.

The eight reactions were then each diluted with molecular grade water (Sigma-Aldrich, Co. LLC. St. Lewis, MO, USA) to 0.2 ng µl^{-1} 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). In brief, 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 (that is, library) with randomly cleaved fragments averaging 300 bp in length that are subsequently duel indexed. The library constructed for each site was then pooled and paired-end sequenced (2 × 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. Workflow of process is presented in Supplementary Fig. 1. 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[44]. 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 to remove any primer sequence[45]. Sequences were then mapped to the COI Barcode of Life Database (iBOL phase 4.00)[41] using a map_reads_reference.py script with the minimum per cent identity to consider a match as 50% and the minimum sequence

length match to a reference of 50% to remove any sequences not likely of COI origin. Subsequent sequences were then chimera checked using usearch version 6 (ref. 46). Remaining sequences <100 bp in length were then taxonomically identified using customized Blast searches against the NCBI non-redundant nucleotide database using the package blast 2.2.28, build on 12 March 2013 16:52:31 (ref. 47). 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 <90.0% sequence similarity, had <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 <100 were excluded from biodiversity detection analysis for all 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[43,48].

After identification of sequences with the NCBI nucleotide sequence database, each uniquely identified taxon from any site 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 taxon was known to be present in Austria, France, Germany and Italy. We excluded the one and very rare case (that is, *Culicoides fascipennis*), where it is known for sure that a species is not in Switzerland, but found in all four neighbouring countries. Geographic verification was done in consultation with 25 expert taxonomists for various groups, primary literature and through database repositories as described in Supplementary Tables 1 and 2. If the species could be confidently confirmed as being present in Switzerland or in all four neighbouring 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 stages or reproduce in the water; Supplementary Table 2). In addition, because we used bovine serum albumin 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.

Mock community analysis. A mock community approach was used to verify that our laboratory methods and bioinformatics pipeline were capable of correctly detecting the taxa of interest. We composed a mock community of invertebrate taxa from 33 different families spanning three phyla (all known to be present in our study area, Supplementary Table 3). We individually extracted their DNA, pooled and sequenced the mock community in accordance with the same methods used for analysis of eDNA samples from the river Glatt (see Supplementary Note 1 for complete methods). We additionally Sanger sequenced all 33 DNA extractions from taxa following that of Mächler et al.[49] to generate a sequence reference database to assess the assignment errors when using NCBI's nucleotide database[47].

Kicknet sampling and identification. Macroinvertebrates were detected using a standard kicknet sampling design described for federal and cantonal guidelines in Switzerland[23,24] and represent our positive control for each site. In brief, we took eight independent kicknet samples per site on 29 October 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[23]. Since eDNA has been shown to decay over short time periods of a few days to a few months;[30] 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 the morphological constraints in the identification of specimens at young life stages or that their physical presence in the water is limited due to timing for their life cycle[23]. 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[23]. Specimens that could not be identified to at least to the taxonomic rank of family were excluded from further analysis.

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, including only aquatic taxa on the standardized list of macroinvertebrates for biomonitoring of Swiss waters by the Federal Office for the Environment[23]. 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. 1). Topological distance between the sampling sites was calculated along the river's path. Catchment area and distance between the sampling sites were calculated using Quantum Geographic Information System in version 2.8 (ref. 50). 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), and that the rate of increase in number of taxa for a given area was faster for eDNA compared with the kicknet (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.

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 9,999 permutations. Here we exclude the tributaries as it is not possible for eDNA to flow into these locations (for example, cd into a). 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 that is expected for the eDNA estimate of β-diversity due to transport[51]. All statistical analyses were performed in R version 3.1.0 (ref. 52).

References

1. Vörösmarty, C. J. et al. Global threats to human water security and river biodiversity. *Nature* **467**, 555–561 (2010).
2. Willett, S. D., McCoy, S. W., Perron, J. T., Goren, L. & Chen, C.-Y. Dynamic reorganization of river basins. *Science* **343**, 1248765 (2014).
3. Rodríguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: chance and Self-Organization* (Cambridge University Press, 1997).
4. Clift, P. D. & Blusztajn, J. Reorganization of the western Himalayan river system after five million years ago. *Nature* **438**, 1001–1003 (2005).
5. Ward, N. D. et al. Degradation of terrestrially derived macromolecules in the Amazon River. *Nat. Geosci.* **6**, 530–533 (2013).
6. Altermatt, F. Diversity in riverine metacommunities: a network perspective. *Aquat. Ecol.* **47**, 365–377 (2013).
7. Mari, L., Casagrandi, R., Bertuzzo, E., Rinaldo, A. & Gatto, M. Metapopulation persistence and species spread in river networks. *Ecol. Lett.* **17**, 426–434 (2014).
8. Deiner, K. & Altermatt, F. Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE* **9**, e88786 (2014).
9. Turner, C. R. et al. Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods Ecol. Evol.* **5**, 676–684 (2014).
10. Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* **21**, 1789–1793 (2012).
11. Ji, Y. et al. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* **16**, 1245–1257 (2013).
12. Bohmann, K. et al. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358–367 (2014).
13. Cristescu, M. E. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* **29**, 566–571 (2014).
14. Goldberg, C. S., Strickler, K. M. & Pilliod, D. S. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biol. Conserv.* **183**, 1–3 (2015).
15. Kelly, R. P. et al. Harnessing DNA to improve environmental management. *Science* **344**, 1455–1456 (2014).
16. Lawson Handley, L. How will the 'molecular revolution' contribute to biological recording? *Biol. J. Linn. Soc.* **115**, 750–766 (2015).
17. Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. & Gough, K. C. REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* **51**, 1450–1459 (2014).
18. Laramie, M. B., Pilliod, D. S. & Goldberg, C. S. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biol. Conserv.* **183**, 29–37 (2015).
19. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
20. Noss, R. F. et al. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Divers. Distrib.* **21**, 236–244 (2015).
21. Gotelli, N. J. & Colwell, R. K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391 (2001).
22. Barbour, M. T., Gerritsen, J., Snyder, B. & Stribling, J. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. *USEPA* **339**, 1–340 (1999).
23. Stucki, P. *Methoden zur Untersuchung und Beurteilung der Fliessgewässer: Makrozoobenthos-Stufe F* (Bundesamt für Umwelt-Vollzug, 2010).
24. Altermatt, F., Seymour, M. & Martinez, N. River network properties shape α-diversity and community similarity patterns of aquatic insect communities across major drainage basins. *J. Biogeogr.* **40**, 2249–2260 (2013).
25. Jane, S. F. et al. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **15**, 216–227 (2015).

26. Sheldon, F. *et al.* Identifying the spatial scale of land use that most strongly influences overall river ecosystem health score. *Ecol. Appl.* **22,** 2188–2203 (2012).

27. Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E. & Zakharov, E. V. 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 (2014).

28. Heino, J. *et al.* A comparative analysis reveals weak relationships between ecological factors and beta diversity of stream insect metacommunities at two spatial levels. *Ecol. Evol.* **5,** 1235–1248 (2015).

29. Sundermann, A., Stoll, S. & Haase, P. River restoration success depends on the species pool of the immediate surroundings. *Ecol. Appl.* **21,** 1962–1971 (2011).

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

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

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

33. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3,** 294–299 (1994).

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

35. Moszczynska, A., Locke, S. A., McLaughlin, J. D., Marcogliese, D. J. & Crease, T. J. Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths. *Mol. Ecol. Resour.* **9,** 75–82 (2009).

36. Zimmermann, J., Glöckner, G., Jahn, R., Enke, N. & Gemeinholzer, B. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Mol. Ecol. Resour.* **15,** 526–542 (2015).

37. Gibson, J. *et al.* Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proc. Natl Acad. Sci.* **111,** 8007–8012 (2014).

38. Drummond, A. J. *et al.* Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* **4,** 1 (2015).

39. Stribling, J. B., Pavlik, K. L., Holdsworth, S. M. & Leppo, E. W. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *J. N. Am. Benthol. Soc.* **27,** 906–919 (2008).

40. Haase, P., Pauls, S. U., Schindehütte, K. & Sundermann, A. First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. *J. N. Am. Benthol. Soc.* **29,** 1279–1291 (2010).

41. Ratnasingham, S. & Hebert, P. D.N. bold: the barcode of life data system. *Mol. Ecol. Notes* **7,** 355–364 (2007).

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

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

44. SeqPrep. Available at https://github.com/jstjohn/SeqPrep (2011).

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

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

47. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **40,** D48–D53 (2012).

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

49. Mächler, E., Deiner, K., Steinmann, P. & Altermatt, F. Utility of Environmental DNA for Monitoring Rare and Indicator Macroinvertebrate Species. *Freshw. Sci.* **33,** 1174–1183 (2014).

50. QGIS Geographic Information System. Available at http://qgis.osgeo.org (Open Source Geospatial Foundation Project, 2015).

51. Cardoso, P., Borges, P. A. & Veech, J. A. Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Divers. Distrib.* **15,** 1081–1090 (2009).

52. R: A language and environment for statistical computing. Version 3.0.1 (R Foundation for Statistical Computing, 2013).

Acknowledgements

Author contributions

K.D. and F.A. designed the study, K.D. and E.M. collected the data, J.-C.W. developed the pipeline and conducted the bioinformatics analysis, E.F., F.A. and K.D. developed the conceptual model, and all authors contributed to the analysis and writing of the manuscript.

Additional information