1 **Title:** Choice of capture and extraction methods affect detection of freshwater biodiversity from
2 environmental DNA

3 **Authors:** Kristy Deiner^{1,2}, Jean-Claude Walser², Elvira Mächler¹ and Florian Altermatt^{1,3}

4 ¹ Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O.
5 Box 611, 8600 Dübendorf, Switzerland

6 ² Swiss Federal Institute of Technology (ETH), Zürich, Genetic Diversity Centre, CHN E 55
7 Universitätstrasse 16, 8092 Zürich, Switzerland

8 ³Institute of Evolutionary Biology and Environmental Studies, University of Zurich
9 Winterthurerstr. 190, CH-8057 Zürich, Switzerland.

10

11 **Corresponding author:** Kristy Deiner, Eawag: Swiss Federal Institute of Aquatic Science and
12 Technology, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland. Phone: +41
13 (0)58 765 53 14, alpinedna@gmail.com

14 **Abstract:** Environmental DNA (eDNA) is used to detect biodiversity by the capture, extraction,
15 and identification of DNA shed to the environment. However, eDNA capture and extraction
16 protocols vary widely across studies. This use of different protocols potentially biases detection
17 results and could significantly hinder a reliable use of eDNA to detect biodiversity. We tested
18 whether choice of eDNA capture and extraction protocols significantly influenced biodiversity
19 detection in aquatic systems. We sampled lake and river water, captured and extracted eDNA
20 using six combinations of different protocols with replication, and tested for the detection of four
21 macroinvertebrate species. Additionally, using the same lake water technical replicates, we

compared the effect of capture and extraction protocols on metabarcoding of biodiversity using 16S for eubacteria and cytochrome c oxidase I (COI) for eukaryotes. Protocol combinations for capture and extraction of eDNA significantly influenced DNA yield and number of sequences obtained from next generation sequencing. We found significantly different detection rates of species ranging from zero percent to thirty-three percent. Differences in metabarcoded biodiversity were also detected between different protocols and resulted in many genera not detected between some protocol combinations. Our results highlight that the choice of molecular protocols used for capture and extraction of eDNA from water can strongly affect biodiversity detection. Consideration of biases caused by choice of protocols should lead to a more consistent and reliable molecular workflow for repeatable and increased detection of biodiversity in aquatic communities.

Keywords: 16S, cytochrome c oxidase I, environmental DNA, freshwater, molecular protocols, targeted species detection

1.1 Introduction

Biodiversity assessment is a main goal as well as a tool used in ecology and conservation biology (Vermeulen and Koziell 2002). Many different measuring approaches exist to assess biodiversity, and these various approaches are typically designed for specific groups of organisms. In recent years, the broadly applicable method of using environmental DNA (eDNA) as a tool to detect organisms in their environment has gained immense interest (Thomsen et al. 2014 this issue; Sutherland et al 2012). Assessment of biodiversity using eDNA relies on a molecular workflow comprising several steps including the capture, extraction and identification of an organism's DNA from environmental samples such as soil or water. The use of eDNA to detect species and measure biodiversity is now at the forefront of approaches in the toolbox for ecologists and conservation scientists (Goldberg et al. 2014 this issue; Yoccoz 2012). The rapid growth in its use, as well as an increased complexity and variation of molecular workflows used to detect eDNA (e.g., next generation sequencing technology (Shokralla et al. 2012)), make a consistent comparison of methodological procedures highly needed.

All molecular workflows used to analyze eDNA consist of capturing DNA from an environmental sample, followed by the extraction and purification of DNA. Purified DNA is then amplified for a specific gene target or the whole eDNA sample is made into a library which is then sequenced and identified or categorized into biodiversity units. For each one of these steps there are a multitude of possible protocols that can be used (Table 1). This heterogeneity in laboratory protocols, however, is likely to challenge comparisons across eDNA studies and to create uncertainty in its application for detecting biodiversity (Wang et al. 2013). The inconsistent use of different molecular protocols across studies is likely due to the fact that research conducted thus far has focused on whether or not a particular species or community

could be detected using DNA found in the chosen habitat (Taberlet et al. 2012), and less so on testing and describing how different laboratory protocols affect detection of a specific species or total biodiversity (Wang et al. 2013). When researchers use particular protocols for extraction of DNA, PCR and sequencing, the choice is often driven by personal preferences, costs, or locally available equipment. Furthermore, published studies typically do not detail the trial and error of testing laboratory protocols that likely took place (but see Goldberg et al. 2011). Therefore, a comparison of molecular protocols used in an eDNA molecular workflow is critically needed so that an understanding of any biases created by use of one protocol over another can be taken into account in future studies (Darling and Mahon 2011; Wang et al. 2013).

There are a large number of decisions that must be made when incorporating certain protocols into a molecular workflow for the identification of biodiversity using environmental DNA.

These decisions start with how to capture (where “capture” is defined as the concentration of cellular or extracellular DNA) genetic material found in the environmental sample (e.g., Pilliod et al. 2013), all the way to deciding what sequencing technology to use (Schloss et al. 2011; Zinger et al. 2012). In freshwater, different molecular protocols have been tested to understand potential biases associated with biodiversity detection and include comparisons of: extraction protocols (Lemarchand et al. 2005), different extraction protocols combined with different additives to alleviate PCR inhibitors (Jiang et al. 2005), different extraction protocols combined with different PCR protocols (Goldberg et al. 2011), and different extraction protocols combined with different sequencing protocols (Morgan et al. 2010). All of these studies found that detection of biodiversity in water depended on which protocols or combination of protocols were used. Shortcomings of many of these studies are that they had very little or no experimental replication of treatment groups (e.g., replicating the different combinations of extraction and PCR), and none

have tested whether protocols used for eDNA capture and eDNA extraction together bias biodiversity detection results (for exception see Piaggio et al. (2014)).

In this study, we tested the combined influence of different eDNA capture and eDNA extraction protocols on the detection of macroinvertebrate species and on the eDNA metabarcoded biodiversity described from freshwater samples. In a replicated experiment we used the same molecular workflow and varied only the laboratory protocols used for capture and extraction of eDNA (Figure 1). To do this, we took a sample of water from a lake and a river respectively, split them into 90 technical replicates each, and performed the six possible combinations under which eDNA was captured and extracted (Figure 1, Table 1). The chosen capture and extraction methods we compared are probably the most feasible with equipment already existing in aquatic/molecular biology laboratories, and also commonly used by many studies (see Table 1). We then targeted four macroinvertebrate species for detection. In the lake water, we sought to detect a water flea (*Daphnia longispina*) and a mussel (*Unio tumidus*). In the river water, we sought to detect a mayfly (*Baetis buceratus*) and an amphipod (*Gammarus pulex*). Additionally, using only the lake water technical replicates, we performed eDNA metabarcoding with the barcode regions of V2-V3 of 16S (approximately 500 bp) for eubacteria and cytochrome c oxidase subunit I (COI) (approximately 650 bp) for eukaryotes. Our goal was to determine the effect of varying laboratory protocols on the detectability of biodiversity. We demonstrate that protocol choice changes detection and we make recommendations for both future research to determine potential mechanisms, as well as suggest which of our tested protocols could enhance detection for freshwater biodiversity detection.

2.1 Methods

2.1.1 Study sites and samples

Water samples were collected from two sites, at the outflow of Lake Greifensee (latitude 47° 22' 21.40" N, longitude 8° 39' 20.67" E) and about 5.6 km away from the lake in its connected river Glatt (latitude 47° 24' 8.06" N, longitude 8° 36' 14.66" E) in Switzerland. Greifensee is a eutrophic, pre-alpine lake with a surface area of 8.5 km² and a maximum depth of 33 m. The outflowing Glatt is a channelized and human modified river. In both Lake Greifensee and river Glatt, diversity of invertebrates is well-known and has been monitored for more than two decades with sampling methods commonly used by aquatic ecologists (Altermatt 2013; AWEL 2012). In each of the two sites, water was collected by submerging two 1 L octagonal polyethylene terephthalate bottles (VWR International, Radnor, PA, USA) just below the surface near the shore of both study sites. Samples were stored on ice in the field, returned to the lab and stored in a –20 °C freezer until capture and extraction methods were performed. The transport time did not exceed 4.5 hours. Our goal was to compare capture and extraction methods; therefore, we used a single water sample from each site that was well-mixed before making 90 aliquots of 15 mL in their own individual tubes. We refer to these as the technical replicates. Fifteen of the 90 technical replicates were then randomly assigned to one of the six experimental treatments covering all possible combinations of two capture methods and three extraction methods (Figure 1).

2.1.2 Environmental DNA capture protocols

We conducted two DNA capture methods, namely filtration and precipitation (Table 1). Typically when filtration is used, much higher volumes of water are used (Table 1), but in order to treat technical replicates equally, and because the precipitation method is typically done with 15 mL of water, filtration technical replicates were also made with 15 mL of water. Filtration for the 45 technical replicates (Figure 1) for each site was carried out by first attaching a filter

housing (Swinnex ®, EMD Millipore Co., Billerica, MA, USA) containing a 0.22 µm glass fiber filter (GF/F, 25mm, Whatman International Ltd., England) to a 20 mL disposable syringe. Water from the 15 mL aliquot was poured into the syringe and the plunger was attached. Water was pushed through by hand at a flow rate of 1 mL per 10 seconds. Precipitation for the 45 technical replicates (Figure 1) for each site was carried out following Ficetola et al. (2008), with the exception that samples were centrifuged at 4 °C instead of centrifuging at room temperature. Six negative controls for each capture method were created by precipitating or filtrating 15 mL of molecular grade DNA free water (Sigma-Aldrich, Co. LLC. St. Lewis, MO., USA). Three negative controls from each capture method were then randomly assigned to each site and these three negative controls for each capture method were then randomly extracted with one of the three extraction protocols. Resulting in one negative control for each experimental unit and a total of six negative controls per site.

2.1.3 Environmental DNA extraction Protocols

We chose three DNA extraction protocols for both targeted and universal taxon detection of aquatic organisms (Table 1, Figure 1). The first method was Qiagen's DNeasy Blood & Tissue Kit (Qiagen GmbH, Hilden, Germany), for which we followed the manufacturer's protocol, except that lysis of precipitated material was carried out in the 50 mL tube used for the precipitation overnight and then transferred to a 1.5 mL tube for the remainder of the protocol. Additionally, for the final step we performed one elution with 100 µL with the provided AE buffer warmed to 55 °C. For filtration, the filter was soaked completely in the mixture of 200 µL of ATL lysis buffer and proteinase K and was incubated at 55 °C for 48 hours instead of 24 to allow for a more complete lysis. We again eluted with 100 µL of provided AE buffer warmed to 55 °C for the final step. The second method was MO BIO's PowerWater DNA Isolation Kit (MO

BIO Laboratories, Inc., Carlsbad, CA, USA). The method was used following the manufacturer's protocol for filtration. For the combination with precipitation as the capture method, the first buffer was added to the 50 mL tube and the pellet was re-suspended by pipetting up and down a few times and then transferred to the tube provided by the manufacture to be used for the bead beating step. The third extraction method was a modified phenol-chloroform-isoamyl extraction followed by an ethanol precipitation (Costas et al. 2007; with modifications listed in Deiner and Altermatt 2014, Online Appendix C). For filtration, the modified protocol was the same as in the cited study. For precipitation, no filter was involved, and the first lysis step was carried out in the 50 mL tube used for the precipitation and was then transferred to a new 1.5 mL for the remainder of the protocol. All technical replicates for both sites were quantified using the Qubit (1.0) fluorometer following recommended protocols for the dsDNA HS Assay which has a high accuracy for double stranded DNA between 1 ng/mL to 500 ng/mL (Life Technologies, Carlsbad, CA, USA).

2.1.4 PCR, Sanger and next generation sequencing

We sought to detect four species that have previously been detected using eDNA methods in this study system. For the lentic species, we chose a water flea (species complex of *Daphnia longispina*) and a mussel (*Unio tumidus*) (Deiner and Altermatt 2014). For the lotic species we chose a mayfly (*Baetis buceratus*) and an amphipod (*Gammarus pulex*) (Mächler et al. In press). These studies were conducted in the same system; we therefore had confidence in the primers used for detection of the chosen species (Deiner and Altermatt 2014; Mächler et al. In press). Additionally, traditional kicknet methods of sampling and morphological identification over the last twenty years have detected all chosen species in their respective lentic and lotic sites (AWEL 2012). For the eubacteria metabarcode approach, we chose the 16S primers B27F (Weisburg et

al. 1991) and B534R (Muyzer et al. 1993) to amplify a 500 bp region spanning the variable regions V2-V3 suitable for eubacterial taxon identification to the genus level (Chakravorty et al. 2007). For the eukaryotic metabarcode approach we applied the standard primers used in many barcoding studies (LCOI and HCOI, Folmer et al. 1994). These primers produce a PCR fragment of about 700 bp that spans the first part of the 5' end of the protein coding region of cytochrome c oxidase subunit I (COI). This region is typically used as the molecular barcode for animals (Hebert et al. 2003) and has been successful for identification of aquatic macroinvertebrates (Deiner et al. 2013; Hajibabaei et al. 2011).

For each of the targeted species, one PCR on each of the 90 technical replicates and 6 negative controls was performed for both lentic and lotic sites and the products were amplified with primers and following PCR protocols described in Deiner and Altermatt (2014) and Mächler et al. (In press). Given the scale of our replication among treatment groups, the choice not to additionally replicate PCRs was based on evidence that having high replication at the experimental treatment level maybe more important than additionally including PCR technical replicates (Kitchen et al. 2010; Tichopad et al. 2009). Additionally, to control for PCR inhibition, BSA (Bovine Serum Albumin) was added to all PCRs as it has been shown to prevent inhibition of PCR from eDNA samples (Jiang et al. 2005). PCR products were confirmed by gel electrophoresis on a 1.4% agarose gel stained with ethidium bromide or GelRed (Biotium Inc., Hayward, CA USA). A sample was considered positive if the PCR product was a single band of the predicted size. All positive PCR products for all species were cleaned using Exo I Nuclease (EXO I) and Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific Inc., Waltham, MD USA). EXO I-SAP reactions were carried out in 8.5 μ L volumes with a final concentration of 1.6 U/ μ L Exo I and 0.15 U/ μ L SAP. The thermal-cycling regime was 15 minutes at 37 °C followed by 15 minutes at 80 °C. All positive PCRs for all species were sequenced in both forward and

reverse directions using dideoxy chain termination chemistry with Big Dye v3.1 following recommended ABI protocols and run on an ABI3730 automated capillary sequencer (Applied Biosystems, Foster City, CA USA). Forward and reverse sequences were aligned using Sequencher 4.9 (Gene Codes, Ann Arbor, MI USA) and blasted against the NCBI's nucleotide database using default parameters to confirm species identity (Benson et al. 2012).

For each of the two metabarcodes (COI and 16S), PCR was performed once on each technical replicate for the lentic site only. PCRs on eDNA were carried out in 15 μ L volumes with final concentrations of 1x supplied buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland) 1x BSA, 0.2 mMol dNTPs, 2.0 mMol MgCl₂, 0.05 units per μ L Taq DNA polymerase (Faststart TAQ, Roche, Inc., Basel, Switzerland), and 0.54 μ Mol of each forward and reverse primer. BSA was added to all PCRs as it has been shown to prevent inhibition of PCR from eDNA samples (Jiang et al. 2005). 2 μ L of extracted eDNA was added that ranged in concentration from 0.05 to 0.55 ng/ μ L in the lentic site (Figure 2) and 0.05 to 0.61 ng/ μ L in the lotic site (not pictured). This range was the outcome of DNA that was extracted following each of the six molecular workflows. The thermal-cycling regime was 95 °C for 4 minutes, followed by 35 cycles of 95 °C for 30 seconds, either 55 °C or 48 °C (16S or COI respectively) for 30 seconds and 72 °C for 1 minute. A final extension of 72 °C for 7 minutes was carried out and the PCR was cooled to 10 °C until removed and stored at -20 °C until confirmation of products occurred. PCR products were confirmed by gel electrophoresis on a 1.4% agarose gel stained with ethidium bromide or GelRed (Biotium Inc., Hayward, CA USA). PCR products from the two genes (COI and 16S) for each technical replicate were pooled in equal proportions in a total of 20 μ L resulting in 90 reactions that were then used as input DNA for library construction. Each of the 90 pooled reactions were then cleaned using AMPure XP beads following the recommended manufacture's protocol, except 0.6 x reaction volume was used instead of 1.8 x based on recommended protocol

for fragment size retention of >500 base pairs (p. 31, Nextera XT DNA 96 kit, Illumina, Inc., San Diego, CA, USA).

We quantified each pooled reaction using the Qubit (1.0) fluorometer following recommended protocols for the dsDNA BR Assay which has a high accuracy for double stranded DNA between 0.01 µg/mL to 5 µg/mL (Life Technologies, Carlsbad, CA, USA). The 90 pooled reactions were then each diluted with molecular grade DNA free water (Sigma-Aldrich, Co. LLC. St. Lewis, MO., USA) to 0.2 ng/µL following the recommended protocol for recommended concentration of DNA sample used for library construction (Nextera XT DNA 96 kit, Illumina, Inc., San Diego, CA, USA). This kit was used specifically because it requires only 1ng of starting DNA compared with TrueSeq which requires 1µg. Amplicons from the two metabarcode genes were pooled without any pre-labeling and were bioinformatically separated post-sequencing (see bioinformatics analysis section 2.3.1).

Metabarcoding libraries for the 90 technical replicates were prepared using the Nextera XT DNA 96 kit following manufacturer's recommended protocols (Illumina, Inc., San Diego, CA, USA) except the final mixture of the denatured and pooled 90 libraries, plus the PhiX control, were diluted to 1:40 instead of 1:25. Paired-end (2x250nt) sequencing was performed on an Illumina MiSeq (MiSeq Reagent kit v2, 250 cycles) at the Genomic Diversity Center at the ETH, Zurich, Switzerland following 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.

2.2 Contamination control procedures

For all pre-PCR procedures, samples were processed in a dedicated eDNA clean lab where no tissue extracted DNA or post PCR products are handled, and all equipment, such as the laminar

flow hood, pipettes and incubators, are dedicated for processing eDNA samples. The floor, walls and outside of all equipment in the room were cleaned with 10% bleach weekly, and the laminar flow hood and pipettes were decontaminated before each use with 10% bleach followed by a 30 minute ultraviolet light treatment. All consumables (e.g., filters, syringes, tips, tubes, gloves, etc.) used in the study were decontaminated by a 30 minute ultraviolet light treatment and filtered tips were used for all protocols pre-PCR. Additionally, all researchers were required to have not been near PCR products before entering the room, be showered, have on freshly washed clothes, and have all exposed skin and shoes covered with protective gear before entering the lab in order to minimize human DNA and other sources of contamination from face, hands, clothes, and shoes. Bottles used for sampling were purchased new for this study, had never come in contact with water from study locations before use, were additionally decontaminated with a 30 minute ultraviolet light treatment in the laminar flow hood, and sealed in the DNA clean lab before use. Bottles were opened only once at each site with gloved hands and new gloves were used for each site. The outside of the bottles may have come in contact with each other during transport. We therefore treated the outsides of all bottles with 10% bleach before they were brought back into the DNA clean lab and re-opened for processing.

All filtration steps were carried out in a laminar flow hood. The precipitation protocol, with the exception of centrifugation step of the 50 mL tubes due to logistical reasons with equipment, was carried out in the DNA clean lab. During the centrifugation step, tubes were always sealed and the outsides were decontaminated with 10% bleach before being brought back into the DNA clean lab. Filter housings were reused between technical replicates, but decontaminated with 10% bleach followed by a rinse with molecular grade DNA free water (Sigma-Aldrich Co. LLC. St. Lewis, MO, USA) and dried with a Kimwipe (Kimberly-Clark, Inc., Irving, TX, USA) before reuse. All filter housings were soaked for a minimum of 20 minutes in freshly mixed 10 %

bleach and treated with a 30 minute ultra-violet light in the laminar flow hood between experimental treatments. All extraction steps for the three methods were carried out in the DNA clean lab with the exception of the PCI extraction steps requiring use of phenol and chloroform. Due to the volatile and potential hazardous risk of breathing phenol and chloroform, this was performed in a standard flow hood in a lab where no PCR products are handled and was decontaminated with 10% bleach before each use. Researchers during this step of the PCI protocol followed procedures and use of protective gear as stated above. All PCRs were set up in the DNA clean lab laminar flow hood. Lastly, experimental treatment negative controls were used to monitor for any laboratory contamination.

2.3 Analysis

2.3.1 Bioinformatics Analysis

Raw reads from next generation sequencing were quality checked and filtered using PrinSeq Lite version 0.20.3 (Schmieder and Edwards 2011). Forward and reverse reads were merged (minimum overlap of 30 bp) using SeqPrep (St. John 2011). Non-overlapping reads were merged with two Ns using a customized python script to speed processing of libraries. The merged reads were de-replicated, de-noised (identity threshold 99%), and chimera checked using usearch version 6 (Edgar 2010). The merged and cleaned reads were binned into genes (16S and COI) using reference mapping applying the usearch option as part of QIIME version 1.7.0 (Caporaso et al. 2010). The reference database “gold_refdb” provided by Qiime was used for the binning of 16S amplicons (Online Appendix Table A1). A customized reference database was built from sequences obtained by using PrimerBLAST (Ye et al. 2012) with the COI primers used in our study. This custom database was then used for binning COI amplicon sequences. Sequences for both genes that did not match their reference databases were excluded from further

analysis. All non-overlapping merged sequences remaining after quality filtering were removed before taxonomic identification. For the taxonomic identification of the 16S amplicons with assembled reads greater than 350 bp, QIIME was used applying the RDP classifier option to obtain taxonomy assignments. Customized Blast searches against the NCBI non-redundant nucleotide database (Benson et al. 2012) were used for the taxonomic assignment for COI reads greater than 150 bp and was automated using Geneious version 6.0 (Biomatters Ltd., Auckland, New Zealand) (Online Appendix Table A2). For COI, sequences that did not match eukaryotes, were below 69.0% sequence similarity, or blasted to unknown environmental samples were excluded from taxonomic diversity analysis between technical replicates. Sequences used for biodiversity detection analysis were archived at datadryad.org (*DOI will be reported here upon acceptance of manuscript for publication*) and all raw sequences reads were deposited in GenBank's Sequence Read Archive (*accession numbers pending*).

2.3.2 Statistical Analysis

We used generalised linear models (glm) to analyse eDNA yield and number of next generation sequences after filtering data as a result of different capture and extraction methods. These variables were assessed to understand whether or not the different capture and extraction protocols or their interaction resulted in varying amounts and/or quality of eDNA. We used an inverse Gaussian error distribution for analyzing eDNA yield (Crawley 2013) and the p -values are based on F -significance tests. We used a Poisson error distribution for analyzing sequence number and p -values are based on Chi^2 -tests (Crawley 2013). Residual deviance of models was used as the goodness-of fit criterion in the model-evaluation. To test for consistency across capture and extraction methods in DNA yield from lentic and lotic waters we performed a correlation analysis. We also used a glm for analysing detection rates of the four targeted

macroinvertebrate species depending on the capture and extraction methods. The response variable was proportion of positive PCR reactions in the 15 technical replicates (i.e., detection rate). We used a quasibinomial link function, as we had some overdispersion in the data, given the model, and an *F*-significance test (Crawley 2013). All statistical analyses were done with the program R, version 3.0.1 (R Development Core Team 2013).

We used a bootstrap-approach to analyse the relationship of biodiversity detected for the different combinations of capture and extraction protocols. Both for the COI and 16S sequences, we bootstrapped the number of classes, orders, families and genera of organisms detected relative to the water volume analysed. For bootstrapping, we used the diversity data generated for each of the individual fifteen 15 mL technical replicates. We bootstrapped mean and 90% confidence intervals for volumes of $15 * n$ mL of water sampled, where *n* included all integers in the range of [1, 15]. Diversity values for each combination of capture/extraction method, gene (COI/16S), taxonomic level (class, order, family, genera) and volume are based on 999 bootstrap estimates. Bootstrap analysis was performed in R, version 3.0.1.

3.1 Results

3.1.1 Effect of methods on yield of extracted DNA and number of next generation sequences

We did not find a significant effect of capture protocols on the total amount of DNA recovered from lentic water as main effect alone. However, we found a highly significant main effect of extraction protocols on the total amount of DNA recovered (Figure 2A, Table 2). We also found a highly significant interaction of capture and extraction protocols on the total amount of DNA recovered from lentic water (Figure 2A, Table 2), showing that capture protocol had an effect on total eDNA recovery, but the effect was different for the three extraction protocols. Overall, the protocol combinations using filtration and PCI extraction yielded the highest total amount of

eDNA (average of 0.313 ± 0.03 ng/ μ L across 15 technical replicates). The protocol combination of precipitation and a PowerWater extraction yielded the lowest total amount of eDNA (average of 0.07 ± 0.01 ng/ μ L across 15 technical replicates). Similarly, we found a significant effect of capture method and a significant effect of extraction protocols on the total amount of DNA recovered from lotic water (glm, capture: $F(1,89)=9.6$, $p=0.003$, extraction: $F(2,89)=10.9$, $p<0.001$), while the interaction between capture and extraction method was not significant ($F(2,89)=1.9$, $p=0.14$). Except for the filtration-PCI approach, the total amount of DNA recovered from the lentic and lotic samples was highly consistent across the method combinations (correlation coefficient of 0.66 with all six combinations, $t(4)=1.7$, $p=0.15$, and correlation coefficient of 0.99 with the filtration-PCI approach excluded, $t(3)=12.9$, $p=0.001$).

After the bioinformatic filtering of the data (Online appendix Table A1), we also found a highly significant effect of capture as a main effect and extraction as a main effect, as well as their interaction, on the number of sequences that were subsequently used for taxonomic diversity estimation (Figure 2B, Table 2). We did not perform next generation sequencing for lotic technical replicates and therefore cannot report patterns of read length and number of reads for this site.

3.1.2 Effect of methods on eDNA detection of targeted species

All four species were detected, but detection rates depended on the protocols used for capture and extraction of DNA (Figure 3, Table 3, Online Appendix B). The two main effects of capture and extraction were significantly different, but the interaction was not (Table 3). There was no significant effect of species identity, but there was a significant interaction of species identity and extraction protocol (Table 3). The protocol combination of filtration and DNeasy had the highest overall detection rate and detected all four species; however, precipitation combined with the PCI

protocol allowed for detection of three of the four species. We detected macroinvertebrate species less often with the PowerWater kit compared to DNeasy or PCI extraction protocols (Figure 3).

All negative controls for both sites showed no amplification for any targeted species. Lastly, some technical replicate PCRs from both the water flea (*Daphnia longispina*) and mayfly (*Baetis buceratus*) produced PCR products that could not be confirmed through sequencing. These failed sequencing reactions were likely due to the presence of a secondary band of a different size that co-sequenced for some technical replicates. The failed sequencing reactions were random with respect to capture and extraction protocols used for the water flea, but not for the mayfly. Specifically, all failed sequence reactions for the mayfly were from the combinations of filtration or precipitation with PCI or PowerWater.

3.1.3 Effect of methods on eDNA metabarcode detection of eubacteria and eukaryotes

Protocol combinations biased the levels of biodiversity detected for both eukaryotes (Figure 4A-D) and eubacteria (Figure 4E-H). The amount of biodiversity detected also varied strongly between technical replicates of the same protocol combination (Online appendix Table A2, A3), indicating that individual 15 mL technical replicates were not able to capture all sampled diversity. Our bootstrap analysis showed that the amount of diversity detected increased with the increasing number of samples pooled (i.e., increasing volume sampled), and the differences in actual number of taxa detected or missed per technical replicate became more obvious with more technical replicates sampled (Figure 4A-H). This pattern is illustrated by the different shapes of the saturation curves for each combination of protocols. For eukaryotes, filtration combined with the DNeasy or PCI extraction detected more diversity for all taxonomic ranks (Figure 4A-D). For eubacteria, there was a large overlap in diversity detected by all protocols at higher taxonomic ranks (Figure 4E-F), whereas at lower taxonomic ranks, the precipitation and

PowerWater protocol combination produced the highest amount of diversity detected (Figure 4G-H). At the genus level of diversity, when all technical replicates were taken into account, there was a difference of nearly 50 genera detected between the highest-yield and lowest-yield protocol combinations for both eukaryotes (Figure 4D) and eubacteria (Figure 4H). All negative controls showed no amplification for any of the two metabarcoding genes.

4.1 Discussion

We demonstrated through a replicated experiment that choice of eDNA capture and eDNA extraction protocols resulted in different detection rates of biodiversity in freshwater. Our results reveal that there is great potential to reduce false negative detections by using the appropriate combination of protocols, and that the protocols to use depend on what type of biodiversity is sought for detection. We make recommendations below for which protocol combinations we tested might decrease false negatives. In general, the finding that the choice of protocols affects diversity detected is especially important for rare or invasive species when the ability to detect both false negative and false positive detections is paramount for monitoring species extirpations or invasions (Darling and Mahon 2011). For example, filtration had the highest detection rate for macroinvertebrate species in both lentic and lotic waters, and the combination of filtration with the DNeasy Blood & Tissue Kit for extraction resulted in the highest detection rate for lotic species. It is therefore not a surprise that this combination of methods has resulted in the detection of rare or invasive eukaryotic species (Goldberg et al. 2011; Goldberg et al. 2013; Pilliod et al. 2013). This is not to say that the other combination of protocols, or those focused on different species will fail to detect species, as illustrated by the many studies that have detected species using precipitation (Ficetola et al. 2008; Thomsen et al. 2012). Rather that if no detection

is found and the species is documented through other means, perhaps the capture and extraction protocols can be modified or tested to improve detection (for example see Piaggio et al. (2014)).

For the eDNA metabarcoding, results are twofold: first, capture and extraction protocols yielded different metabarcoded diversity, and second, the protocol combination that detected the highest amount of biodiversity for eukaryotes was different from that of eubacteria. Therefore, when eDNA metabarcoding of multiple domains is sought from the same sample, our results indicate that multiple protocols of capture and extraction should be utilized to maximize the accuracy of the biodiversity estimates in aquatic habitats.

4.1.1 Differences between protocols in detection of biodiversity

The systematic bias in detected biodiversity when different capture protocols are used may be because i) capture protocols differ in what type of eDNA is captured, ii) an interaction of what is captured and how the DNA is isolated, and iii) whether there are differences in the purity of the isolated DNA. This is not an exhaustive list of potential mechanisms, but are some of the most parsimonious given our results. In terms of capture protocols, the precipitation method likely captures eDNA from both extracellular and intracellular DNA. Precipitation captures eDNA through the presence of salt and ethanol which precipitates extracellular DNA molecules from water, in combination with the centrifugal forces that cause whole cells or tissues to form a pellet from which DNA is then extracted (Maniatis et al. 1982). When DNA is captured on a filter, however, there is a size bias in the cellular material (e.g., organelles such as mitochondria) that is captured and which depends on what pore size of filter is used. In general, it is likely that only intracellular rather than extracellular DNA by itself is captured on a filter, because DNA molecules are too small and filtration is specifically used to separate intracellular DNA from extracellular DNA in water (Beebe 1991; DeFlaun et al. 1986). Some filters (e.g., cellulose

nitrate) have been demonstrated to bind extracellular DNA, but only when the DNA has first been denatured and when the filter has been presoaked at high salt conditions (Baker 1977). A recent study has also shown that small percentages (1% to 17%) of extracellular DNA do bind to many filter types (i.e., polyvinylidene fluoride, polyethersulfone, polycarbonate and mixed cellulose esters), but that this depended on pore size and the higher yield (17 %) was from pore sizes of 0.1 micron (Liang and Keeley 2013). Additionally, this study confirmed those of Baker (1977), albeit with different filter types, that under higher salt conditions more extracellular DNA can be bound (Liang and Keeley 2013). From this explanation we would expect the precipitation protocol could yield higher diversity, and/or an increase in detection of targeted species because captured eDNA comes from a higher percentage of both extra and intracellular DNA. We did find that greater eubacteria diversity was captured with the precipitation protocol; however, we found that for our targeted species (all of which are eukaryotes) and for metabarcodes detected eukaryotes, the filtration method of capture detected more diversity.

The higher diversity in eukaryotes detected with filtration, in combination with the DNeasy Blood & Tissue Kit, might be due to capture of cells in addition to an extraction protocol designed for eukaryotic cell lysis. The DNeasy Blood and Tissue Kit, as well as the PCI method of extraction, use a cell lysis (or biochemical) method, whereas, PowerWater uses a bead-beating (or mechanical) method to break open cells. The PowerWater extraction method, was designed for extraction of DNA from eubacteria (Callahan 2009) and the bead-beating method is necessary for breaking open gram negative and positive eubacteria cell walls (Rajendhran and Gunasekaran 2008; Tringe and Rubin 2005). Biochemical methods of lysis, compared to mechanical, are known to cause less DNA shearing forces and reduce fragmentation of DNA (Wintzingerode et al. 1997), potentially allowing greater detection of biodiversity. The mechanical method of extraction, in combination with precipitation, may explain why we detected a higher diversity for

eubacteria compared to the other methods of capture and extraction. There appears to be a trade-off between the gains in eubacteria diversity detected using the mechanical method of extraction and the reduction of eukaryotic diversity, which have cells that are more easily lysed. DNA from eukaryote cells, therefore, likely experienced higher shearing of the DNA, which may explain the reduced detection with the precipitation and PowerWater protocols.

PCR inhibition is another possible mechanism by which different detections were observed for targeted species and metabarcoded diversity. PCR inhibition could be caused by either the presence of chemical inhibitors such as hemin or tannic acids or the ratio of total DNA to that of the targeted DNA. In our study we controlled for PCR inhibition of unknown chemicals that may co-extract or be carried over from the extraction themselves (e.g., proteinase K, phenol, etc.) through the addition of BSA to all technical replicate PCRs. The addition of BSA has been demonstrated to enhance detection of targeted species eDNA extracted from water samples (Jiang et al. 2005), as well as enhance PCR for metabarcoded eubacteria when both known and unknown inhibitors are present in water (Kreader 1996). We also observed that even though the method combination of filtration and PCI produced the highest yield of total DNA, this method only detected three of the four species and suggests that quantity of total DNA may not necessarily be an indication of targeted DNA in a sample. Extraction protocols that have high yields of non-target DNA from a sample have been shown to produce PCR inhibition (Thompson et al. 2006). The targeted species detection rates for all methods were low in our study (less than 40%, Figure 3), given that the four species are known to be present at their respective sites. There was, however, likely some stochasticity affecting detection from the same 2 L of water given that there are likely few DNA molecules from each of our given species. We dealt with this stochasticity of target DNA at low concentration by randomizing the technical replicates before assigning them to each treatment group. When a low number of DNA fragments from a

targeted species is expected and background inhibitors or DNA are high in concentration, dilution of the DNA extraction can be conducted. While this is not recommended for environmental DNA studies due to decreased sensitivity in detection (Wintzingerode et al. 1997), this decreased ability to detect any given target can be counter balanced by increasing the sensitivity of the PCR assay (for an example of how to increase sensitivity see Wright et al. 2014).

4.1.2 Taxonomic identification and metabarcoding eDNA

Our goal was not to describe the species diversity of these sites *per se* (as is the goal of many diversity studies in aquatic systems, see for example Altermatt et al. (2013) and Besemer et al. (2013)), but rather to quantify the difference in taxonomic diversity sampled from each technical replicate as it was subjected to the different molecular protocols. We do not place high certainty on our taxonomic assignments as they are based on BLAST searches against what is currently in searchable databases (QIIME's gold_refdb for 16S and GenBank for COI), and therefore refrain from discussing in absolute terms what species were detected across this large amount of diversity. These databases, however, are the most complete databases for which to do taxonomic assignment of unknown sequences for the genes used for metabarcoding. Combined across all technical replicates we observed detection of 16 phyla of eukaryotes and 16 phyla of eubacteria common to aquatic systems (Online Appendix Table A2, A3). We found cases of both accurate and interesting taxonomic assignments at lower levels of taxonomy, but also identifications of genera not expected to be in our study area. An exciting finding was that we identified a genus of freshwater sponges, *Eunapius* (Online Appendix Table A3), which has only one known species (*Eunapius fragilis*) documented in Switzerland. We recovered 36 sequences of *E. fragilis* in different technical replicates and they ranged in size from 154 to 469 bp. We were conservative in our analysis and treated sequences assigned to the same taxon as equal between technical

replicates as long as they met all stringent bioinformatic thresholds used for including sequences in the identification process. Choice to treat the taxonomic assignment data in this manner means that we likely underestimated diversity at the different taxonomic levels, but because we applied the same criteria to all treatment groups it is an unlikely cause of our detected differences. We were confined in using a taxonomic approach for analysis because sequences produced with the Illumina Nextera XT DNA kit do not always cover the same region of the amplified gene (due to the random fragmentation process used by the Nextera XT library construction protocol) and this prevents accurate assignments of operational taxonomic units via an alignment (Jones et al. 2011). Spurious taxonomic identifications are most likely due to the fact that the reference database used does not include the taxon sequenced. With continued sequencing efforts to refine and curate sequences in databases used for identification, the confidence in and ability to accurately assign sequences to taxonomic groups will only increase in the future (Deiner et al. 2013; Hajibabaei et al. 2011; Kim et al. 2012).

4.2 Recommended protocols for biodiversity detection

Within the context of the protocols tested in this study, we advocate different approaches for targeting eukaryotes or eubacteria through eDNA metabarcoding. For eukaryotes using the COI gene, we recommend the use of filtration and Qiagen DNeasy extraction kit. When seeking to metabarcode eubacteria in water, precipitation and PowerWater detected more genera than any other method combination. One caveat is that many eubacteria eDNA studies want to exclude the capture of diversity from dead organisms (Liang and Keeley 2013; Zinger et al. 2012), so precipitation is typically not used to avoid inadvertently including DNA from lysed cells that came from dead or dying eubacteria. Therefore, most studies of eubacteria in freshwater (e.g., Lemarchand et al. 2005) have used filtration and PCI, or filtration and the PowerWater kit, and in

our study, these combinations resulted in the second and third highest detection of genera for eubacteria biodiversity. When starting a new project and developing a molecular workflow for eDNA detection we further recommend doing a pilot study to compare protocol effect on detection, as well as surveying the literature for new technological advances given the rapidly changing nature of this field.

4.4 Additional considerations and conclusions

We tested the differences caused by capture and extraction methods on the amount and identity of biodiversity detected in aquatic samples. However, there are many additional steps in the molecular workflow that may further lead to synergistic biased results in detection. We recommend future method development studies comparing additional protocol combinations focus on: sampling and field preservation techniques (Pilliod et al. 2013), PCR protocols (e.g., Chandler et al. 1997; Goldberg et al. 2011; Schmidt et al. 2013), primer choice (e.g., Tang et al. 2012; Zinger et al. 2012), inhibitors (e.g., Jiang et al. 2005) McKee et al. 2014 this issue), sequencing platforms (e.g., Claesson et al. 2010) and bioinformatic pipelines (Schloss et al. 2011). All steps of which have been shown individually to cause biases in biodiversity detection using eDNA. Additionally, in this study we varied two steps in the whole molecular workflow while holding the above mentioned steps equal for all technical replicates. We do not know to what extent these decisions may have biased the outcome between experimental treatments. For example, we chose for logistical reasons to freeze our water samples after collection. The freeze-thaw process can be used in DNA extraction during the cell lysis stage. Freezing our sample post collection may have lysed some cells resulting in this DNA not being captured by filtration.

The use of eDNA is increasing as a broadly applied method to monitor biodiversity, making the use of comparable and appropriate methodologies crucial. A general agreement on the use of

certain protocols is not necessarily needed nor possible given that rapid change in genetic technology. The acknowledgement, however, of known differences in what biodiversity is detected with current protocols is necessary as it will help to justify and establish the use of eDNA methods for robust detection of biodiversity.

Acknowledgements

We thank MoBio, Inc. for providing a PowerWater DNA Isolation Kit and vortex adaptor for testing in this study. Data analysed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. Funding is from the Swiss National Science Foundation Grant 31003A_135622 (to F.A.) and Eawag discretionary funds (to K.D.).

References

Altermatt, F., 2013. Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology* 47, 365-377.

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

AWEL, 2012. Zürcher Gewässer 2012: Entwicklung–Zustand–Ausblick, ed. W. AWEL Amt für Abfall, Energie und Luft, Zürich, Switzerland.

Baker, R.F., 1977. Binding of DNA to cellulose nitrate filters under denaturing conditions. *Analytical biochemistry* 78, 569-571.

Beebee, T.J., 1991. Analysis, purification and quantification of extracellular DNA from aquatic environments. *Freshwater Biology* 25, 525-532.

566 Benson, D., Karsch-Mizrachi, I., Clark, K., Lipman, D., Ostell, J., Sayers, E., 2012. GenBank.
567 Nucleic Acids Research 40, D48-D53.

568 Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., Battin, T.J., 2013. Headwaters are
569 critical reservoirs of microbial diversity for fluvial networks. Proceedings of the Royal Society B:
570 Biological Sciences 280.

571 Callahan, H., 2009. Inhibitor-free DNA purification from water samples. BioTechniques 46.

572 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K.,
573 Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-
574 throughput community sequencing data. Nature methods 7, 335-336.

575 Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S
576 ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of
577 Microbiological Methods 69, 330-339.

578 Chandler, D.P., Fredrickson, J.K., Brockman, F.J., 1997. Effect of PCR template concentration
579 on the composition and distribution of total community 16S rDNA clone libraries. Molecular
580 Ecology 6, 475-482.

581 Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., O'Toole,
582 P.W., 2010. Comparison of two next-generation sequencing technologies for resolving highly
583 complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids
584 Research 38, e200.

585 Costas, B.A., McManus, G., Doherty, M., Katz, L.A., 2007. Use of species-specific primers and
586 PCR to measure the distributions of planktonic ciliates in coastal waters. *Limnol. Oceanogr.*
587 *Methods* 5, 163-173.

588 Crawley, M.J., 2013. *The R book*, Second edn. John Wiley & Sons, West Sussex, United
589 Kingdom.

590 Darling, J.A., Mahon, A.R., 2011. From molecules to management: adopting DNA-based
591 methods for monitoring biological invasions in aquatic environments. *Environmental Research*
592 111, 978-988.

593 Debroas, D., Humbert, J.F., Enault, F., Bronner, G., Faubladier, M., Cornillot, E., 2009.
594 Metagenomic approach studying the taxonomic and functional diversity of the bacterial
595 community in a mesotrophic lake (Lac du Bourget–France). *Environmental microbiology* 11,
596 2412-2424.

597 DeFlaun, M.F., Paul, J.H., Davis, D., 1986. Simplified method for dissolved DNA determination
598 in aquatic environments. *Applied and environmental microbiology* 52, 654-659.

599 Deiner, K., Altermatt, F., 2014. Transport Distance of Invertebrate Environmental DNA in a
600 Natural River. *PLoS ONE* 9, e88786.

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

604 Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
605 26, 2460-2461.

606 Ficetola, G.F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using
607 environmental DNA from water samples. *Biology Letters* 4, 423-425.

608 Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification
609 of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular*
610 *marine biology and biotechnology* 3, 294-299.

611 Ghai, R., Rodríguez-Valera, F., McMahon, K.D., Toyama, D., Rinke, R., de Oliveira, T.C.S.,
612 Garcia, J.W., de Miranda, F.P., Henrique-Silva, F., 2011. Metagenomics of the water column in
613 the pristine upper course of the Amazon river. *PLoS ONE* 6, e23785.

614 Goldberg, C.S., Pilliod, D.S., Arkle, R.S., Waits, L.P., 2011. Molecular detection of vertebrates
615 in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant
616 salamanders. *PLoS ONE* 6, e22746.

617 Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J., Waits, L.P., 2013. Environmental DNA
618 as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*).
619 *Freshwater Science* 32, 792-800.

620 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., Baird, D.J., 2011. Environmental
621 barcoding: a next-generation sequencing approach for biomonitoring applications using river
622 benthos. *PLoS ONE* 6, e17497.

623 Hebert, P.D., Cywinska, A., Ball, S.L., 2003. Biological identifications through DNA barcodes.
624 *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, 313-321.

625 Jerde, C.L., Mahon, A.R., Chadderton, W.L., Lodge, D.M., 2011. "Sight-unseen" detection of
626 rare aquatic species using environmental DNA. *Conservation Letters* 4, 150-157.

627 Jiang, J., Alderisio, K.A., Singh, A., Xiao, L., 2005. Development of procedures for direct
628 extraction of *Cryptosporidium* DNA from water concentrates and for relief of PCR inhibitors.
629 *Applied and environmental microbiology* 71, 1135-1141.

630 Jones, M., Ghoorah, A., Blaxter, M., 2011. jMOTU and Taxonerator: turning DNA Barcode
631 sequences into annotated operational taxonomic units. *PLoS ONE* 6, e19259.

632 Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon, Y.S., Lee, J.-H.,
633 Yi, H., Won, S., Chun, J., 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence
634 database with phylotypes that represent uncultured species. *International Journal of Systematic
635 and Evolutionary Microbiology* 62, 716-721.

636 Kitchen, R.R., Kubista, M., Tichopad, A., 2010. Statistical aspects of quantitative real-time PCR
637 experiment design. *Methods* 50, 231-236.

638 Kreader, C.A., 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4
639 gene 32 protein. *Applied and environmental microbiology* 62, 1102-1106.

640 Lemarchand, K., Berthiaume, F., Maynard, C., Harel, J., Payment, P., Bayardelle, P., Masson, L.,
641 Brousseau, R., 2005. Optimization of microbial DNA extraction and purification from raw
642 wastewater samples for downstream pathogen detection by microarrays. *Journal of
643 Microbiological Methods* 63, 115-126.

644 Liang, Z., Keeley, A., 2013. Filtration Recovery of Extracellular DNA from Environmental
645 Water Samples. *Environmental Science & Technology* 47, 9324-9331.

646 Mächler, E., Deiner, K., Steinmann, P., Altermatt, F., In press. Utility of environmental DNA for
647 monitoring rare and indicator macroinvertebrate species. *Freshwater Science*.

648 Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular cloning: a laboratory manual. Cold
649 Spring Harbor Laboratory Cold Spring Harbor, NY.

650 Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N., Kawabata, Z.i., 2012. Surveillance of
651 fish species composition using environmental DNA. *Limnology* 13, 193-197.

652 Morgan, J.L., Darling, A.E., Eisen, J.A., 2010. Metagenomic sequencing of an in vitro-simulated
653 microbial community. *PLoS ONE* 5, e10209.

654 Muyzer, G., De Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations
655 by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes
656 coding for 16S rRNA. *Applied and environmental microbiology* 59, 695-700.

657 Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R.,
658 Konstantinidis, K.T., 2011. Metagenomic insights into the evolution, function, and complexity of
659 the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Applied*
660 *and environmental microbiology* 77, 6000-6011.

661 Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E.,
662 Avery, M.L., 2014. Detecting an elusive invasive species: a diagnostic PCR to detect Burmese
663 python in Florida waters and an assessment of persistence of environmental DNA. *Molecular*
664 *Ecology Resources* 14, 374-380.

665 Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P., Richardson, J., 2013. Estimating
666 occupancy and abundance of stream amphibians using environmental DNA from filtered water
667 samples. *Canadian Journal of Fisheries and Aquatic Sciences* 70, 1123-1130.

668 Poté, J., Mavingui, P., Navarro, E., Rosselli, W., Wildi, W., Simonet, P., Vogel, T.M., 2009.
669 Extracellular plant DNA in Geneva groundwater and traditional artesian drinking water
670 fountains. *Chemosphere* 75, 498-504.

671 Rajendhran, J., Gunasekaran, P., 2008. Strategies for accessing soil metagenome for desired
672 applications. *Biotechnology Advances* 26, 576-590.

673 Schloss, P.D., Gevers, D., Westcott, S.L., 2011. Reducing the Effects of PCR Amplification and
674 Sequencing Artifacts on 16S rRNA-Based Studies. *PLoS ONE* 6, e27310.

675 Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina
676 metabarcoding of a soil fungal community. *Soil Biology and Biochemistry* 65, 128-132.

677 Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets.
678 *Bioinformatics* 27, 863-864.

679 Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing
680 technologies for environmental DNA research. *Molecular Ecology* 21, 1794-1805.

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

682 Sutherland, W.J., Bardsley, S., Clout, M., Depledge, M.H., Dicks, L.V., Fellman, L., Fleishman,
683 E., Gibbons, D.W., Keim, B., Lickorish, F., 2012. A horizon scan of global conservation issues
684 for 2013. *Trends in ecology & evolution*.

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

687 Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The
688 widely used small subunit 18S rDNA molecule greatly underestimates true diversity in
689 biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences* 109,
690 16208-16212.

691 Team, R.D.C., 2013. R: A language and environment for statistical computing. R Project for
692 Statistical Computing Vienna, Austria.

693 Thompson, D., Rajal, V., Batz, S., Wuertz, S., 2006. Detection of *Salmonella* spp. in water using
694 magnetic capture hybridization combined with PCR or real-time PCR. *Journal of Water and*
695 *Health* 4, 67-75.

696 Thomsen, P., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L.,
697 Willerslev, E., 2012. Monitoring endangered freshwater biodiversity using environmental DNA.
698 *Molecular Ecology* 21, 2565-2573.

699 Tichopad, A., Kitchen, R., Riedmaier, I., Becker, C., Ståhlberg, A., Kubista, M., 2009. Design
700 and Optimization of Reverse-Transcription Quantitative PCR Experiments. *Clinical Chemistry*
701 55, 1816-1823.

702 Tringe, S.G., Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental samples.
703 *Nature Reviews Genetics* 6, 805-814.

704 Vermeulen, S., Koziell, I., 2002. Integrating global and local values: a review of biodiversity
705 assessment. IIED.

706 Wang, J., McLenachan, P.A., Biggs, P.J., Winder, L.H., Schoenfeld, B.I., Narayan, V.V., Phiri,
707 B.J., Lockhart, P.J., 2013. Environmental bio-monitoring with high-throughput sequencing.
708 Briefings in bioinformatics 14, 575-588.

709 Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA
710 amplification for phylogenetic study. Journal of bacteriology 173, 697-703.

711 Wintzingerode, F.V., Göbel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity
712 in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS microbiology reviews 21,
713 213-229.

714 Wright, E.S., Yilmaz, L.S., Ram, S., Gasser, J.M., Harrington, G.W., Noguera, D.R., 2014.
715 Exploiting extension bias in polymerase chain reaction to improve primer specificity in
716 ensembles of nearly identical DNA templates. Environmental microbiology 16, 1354-1365.

717 Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-
718 BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC
719 bioinformatics 13, 134.

720 Yoccoz, N.G., 2012. The future of environmental DNA in ecology. Molecular Ecology 21, 2031-
721 2038.

722 Zinger, L., Gobet, A., Pommier, T., 2012. Two decades of describing the unseen majority of
723 aquatic microbial diversity. Molecular Ecology 21, 1878-1896.

724

725 **Figure legends**

726
 727 Figure 1: Experimental design used to test for biases associated with combinations of two DNA-
 728 capture and three DNA-extraction protocols on eDNA yield, next generation sequence data
 729 quality, and biodiversity detected. Numbers of technical replicates are indicated in parentheses.
 730 Each technical replicate (N=90) was a 15 mL aliquot from a single water sample, with the six
 731 experimental treatments completed for both a lentic and lotic study site. PCI, PW and DNeasy
 732 are the different extraction methods performed (Phenol-Chloroform-Isoamyl alcohol, MoBio
 733 PowerWater DNA Isolation Kit it and Qiagen: DNeasy Blood & Tissue Kit respectively).

734

735 Figure 2: Differences in DNA yield (A), raw average next generation sequence length (B), and
 736 number of reads after filtering data (C) as a result of different capture and extraction methods of
 737 DNA from lentic water. Boxplots are based on 15 technical replicates. Black horizontal bars
 738 represent median values; boxes give 25% and 75% percentiles. Circles are values beyond
 739 interquartile ranges. Abbreviations are the same as in Fig.1.

Figure 3: Detection difference of four macroinvertebrate species as a result of capture and extraction methods. Lentic species included a water flea (top left, *Daphnia longispina*) and a mussel (bottom left, *Unio tumidus*), and lotic species tested were a mayfly (top right, *Baetis buceratus*) and an amphipod (bottom right, *Gammarus pulex*). Detection rate was calculated as the proportion of positive and sequence-confirmed amplifications from 15 technical replicates. Abbreviations are as in Fig. 1.

Figure 4: Detected diversity of eukaryotes

(A–D) and eubacteria (E–H) based on Illumina MiSeq data and as a result of capture and extraction methods (red and blue color, see legend) and volume of lentic water sampled.

Diversity of eukaryotes and eubacteria was estimated from taxonomic identifications of COI and 16S genes respectively, and was measured in 15 independent 15 mL technical replicates for each of the six combinations of capture and extraction methods. Cumulative diversity curves with increasing volume sampled are based on bootstrapped values, giving the mean (line) and 90% confidence interval (light shaded area). Diversity was measured at the level of classes (A and E), orders (B and F), families (C and G) and genera (D and H).

757 **Tables**

758 Table 1: Comparison of capture and extraction methods used for detecting biodiversity in water with eDNA. This selection is not
 759 exhaustive, but rather exemplifies the variability in capture methods, extraction methods, sample effort (i.e., water volume), sequencing
 760 approach, and combinations thereof across different taxa and freshwater environments.

Capture Method	Extraction Method	Volume of water	Locus	Habitat	Targeted or Metabarcoding	Taxonomic group	Sequencing technology	Reference
Precipitation	Qiagen DNeasy	3x15 mL	cyt b	lentic	Targeted	amphibian	Sanger	Ficetola et al. 2008
Filtration	Qiagen DNeasy	5 L	cyt b	lotic	Targeted	amphibian	Sanger	Goldberg et al. 2011
Filtration	MO BIO PowerWater	2 L	d-loop	lotic	Targeted	fish	Sanger	Jerde et al. 2011
Precipitation and Filtration	Qiagen DNeasy	2 L	cyt b	lotic	Targeted	fish	Sanger	Minamoto et al. 2012
Precipitation	Qiagen DNeasy	3x15 mL	cyt b, COI	lentic and lotic	Targeted	fish, amphibian, crustacean, insect, mammal	Sanger	Thomsen et al. 2012
Centrifugation	QIAamp DNA Stool Mini Kit	(250 or 500 mL)	NADH5	lentic	Targeted	mammal	Sanger	Caldwell et al. 2007
Filtration	EPICENTRE	4 L	16S rRNA	lotic	Metabarcoding	bacteria	Roche 454 GS-FLX-Ti	Ghai et al. 2011
Filtration	Phenol-Chloroform-Isoamyl	10 L	16S rRNA	lentic	Metabarcoding	bacteria	Roche 454 GS-FLX-Ti and Illumina GA II	Oh et al. 2011
Filtration	Phenol-Chloroform-Isoamyl	45 L	16S rRNA	lentic	Metabarcoding	bacteria	Sanger	Debroas et al. 2009
Lyophilization	MO BIO Ultraclean soil DNA Kit	2 L	18S	ground water	Metabarcoding	plant	Sanger	Poté et al. 2009

761 Table 2: Results of three generalized linear models (glms), explaining differences in DNA yield (A) and number of reads after filtering
762 data (B) as a result of different capture and extraction methods, as well as their interaction. In the first glm (A), an inverse Gaussian error
763 distribution was used, and the p -values are based on F-significance tests, with F -values given. In the second glm (B), a Poisson error
764 distribution was used, and p -values are based on Chi^2 -tests. Df = Degrees of freedom.

Estimate	Df	Deviance	Residual Df	Residual deviance	F-value	p-value
<i>A) DNA yield in lentic water</i>						
Capture	1	0.19	88	360.76	0.11	0.74
Extraction	2	147.536	86	213.23	40.89	<0.0001
Capture x Extraction	2	15.91	84	197.32	4.41	0.015
Null			89	360.96		
<i>B) Number of reads</i>						
Capture	1	11045.6	88	176143		<0.0001
Extraction	2	4145.3	86	171997		<0.0001
Capture x Extraction	2	5305.1	84	166692		<0.0001
Null			89	187188		

765 Table 3. Generalized linear model explaining detection rate of four targeted macroinvertebrate species as a result of species identity,
 766 capture and extraction methods, as well as their interaction. A binomial error distribution was used, and the p-values are based on *Chi*²-
 767 significance tests. *Df* = Degrees of freedom.

Estimate	Df	Deviance	Residual Df	Residual deviance	p-value
Species identity	3	4.99	20	41.04	0.17
Capture	1	4.71	19	36.32	0.03
Extraction	2	10.22	17	26.11	0.006
Species identity x Extraction	6	12.79	11	13.31	0.047
Null			23	46.03	

768

769 Online Appendixes

770 Online Appendix Table A1: Eubacteria taxonomic assignments of sequences used for bootstrap
771 analysis to determine biodiversity detection differences as a result of different capture. NA
772 denotes a taxonomic name that could not be assigned and these cells were excluded in bootstrap
773 analysis. Species level is presented here, but was excluded from all analysis.

774 Online Appendix Table A2: Eukaryote taxonomic assignments of sequences used for bootstrap
775 analysis to determine biodiversity detection differences as a result of different capture and
776 extraction protocols. NA denotes a taxonomic name that could not be assigned and these cells
777 were excluded in bootstrap analysis. Species level is presented here, but was excluded from all
778 analysis.

779 Online Appendix B: Online Appendix B: Sequences obtained from each technical replicate to
780 confirm detection of species. F = filtration, P= precipitation, PCI = Phenol chloroform Isoamyl,
781 PW = PowerWater, number in name for each sequence is the technical replicate number.

782 *Daphnia longispina*

783 F04PCI

784 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
785 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
786 C

787 F22PCI

788 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
789 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
790 C

791 F32PCI

792 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
793 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
794 C

795 F05PW

796 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG

797 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
798 C

799 F02PW
800 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
801 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
802 C

803 F08PW
804 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
805 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
806 C

807 F15DNeasy
808 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
809 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
810 C

811 F21DNeasy
812 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
813 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
814 C

815 P44PCI
816 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
817 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
818 C

819 P06DNeasy
820 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
821 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
822 C

823 P34DNeasy
824 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
825 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
826 C

827 P37DNeasy
828 GTACTTCTAGAGAACGTTTTCAAATTAAAAGATTTTTTTTACTTCAGGTCAAGGTGCAG
829 TTTATACAGAGGTGGAGGTGAGCTACAATTTTCTTTACAAATATGGATACTAAGCCT
830 C

831 *Unio tumidus*

832 F10PCI
833 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
834 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
835 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

836 F36PW
837 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
838 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
839 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

840 F12DNeasy
841 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
842 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
843 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

844 F37DNeasy
845 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
846 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
847 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

848 P41PCI
849 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
850 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
851 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

852 P44PCI
853 TTACTGGTTGGACAGTATACCCACCTTTGTCTGGAAATGTTGCTCATTCTGGAGCTTC
854 AGTGGATTTGGCTATTTTTTCTTTGCATCTTGCAGGTGCTTCTTCTATTTTGGGTGCTA
855 TTAACTTTATTTCTACTGTTGGTAATATGCGATCTCCTGGTTTGGTTGCTGAACGGAT

856 *Baetis buceratus*

857 F03DNeasy
858 ACTCTATTGATCTCTAGAAGAATTGTTGATGTGGGTGCAGGTACTGGTTGAACCGTTT
859 ATCCACCCCTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
860 TTCTCTGCATTTAGCGGGGGTTTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
861 GTGGTAAACATGCGTAGTCCTGG

862 F09DNeasy
863 ACTCTATTGATCTCTAGAAGAATTGTTGATGTGGGTGCAGGTACTGGTTGAACCGTTT
864 ATCCACCCCTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
865 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
866 GTGGTAAACATGCGTAGTCCTGG

867 F12DNeasy
868 ACTCTATTGATCTCTAGAGAATTGTTGATGTGGGTGCAGGTACTGGTTGAACCGTTT
869 ATCCACCCTTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
870 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
871 GTGGTAAACATGCGTAGTCCTGG

872 F28DNeasy
873 ACTCTATTGATCTCTAGAGAATTGTTGATGTGGGCGCAGGTACTGGTTGAACCGTTT
874 ATCCACCCTTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
875 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
876 GTGGTAAACATGCGTAGTCCTGG

877 F34QDNeasy
878 ACTCTATTGATCTCTAGAGAATTGTTGATGTGGGCGCAGGTACTGGTTGAACCGTTT
879 ATCCACCCTTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
880 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
881 GTGGTAAACATGCGTAGTCCTGG

882 P06DNeasy
883 ACTCTATTGATCTCTAGAGAATTGTTGATGTGGGCGCAGGTACTGGTTGAACCGTTT
884 ATCCACCCTTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
885 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
886 GTGGTAAACATGCGTAGTCCTGG

887 P31DNeasy
888 ACTCTATTGATCTCTAGAGAATTGTTGATGTGGGCGCAGGTACTGGTTGAACCGTTT
889 ATCCACCCTTGGCTGCTAATATCGCTCATGGAGGATCTTCAGTTGATTTTGCTATCTT
890 TTCTCTGCATTTAGCGGGGGTCTCTTCAATTTAGGGGCTGTGAATTTTATTACAACC
891 GTGGTAAACATGCGTAGTCCTGG

892

893 *Gammarus pulex*

894 F25DNeasy
895 TAACCCTTCTACTTATAAGTAGTATAGTAGAAAGAGGCGTAGGAACGGGTTGGACGG
896 TATACCCACCGTTGGCGGGTATCTCAGCTCATGGTGGTGGGGCGGTAGATCTAGCTA
897 TTTTTTCACTACATTTAGCAGGGGCCTCCTCTATCCTCGGCGCTATTAATTTTATCTCT
898 ACTGTAGT

899 F34DNeasy
900 TAACCCTTCTACTTATAAGTAGTATAGTAGAAAGAGGCGTAGGAACGGGTTGGACGG
901 TATACCCACCGTTGGCGGGTATCTCAGCTCATGGTGGTGGGGCGGTAGATCTAGCTA

902 TTTTTTCACTACATTTAGCAGGGGCCTCCTCTATCCTCGGCGCTATTAATTTTATCTCT
903 ACTGTAGT

904 F40DNeasy

905 TAACCCTTCTACTTATAAGTAGTATAGTAGAAAGAGGCGTAGGAACGGGTTGGACGG
906 TATACCCACCGTTGGCGGGTATCTCAGCTCATGGTGGTGGGGCGGTAGATCTAGCTA
907 TTTTTTCACTACATTTAGCAGGGGCCTCCTCTATCCTCGGCGCTATTAATTTTATCTCT
908 ACTGTAGT

909 P19PCI

910 TAACCCTTCTACTTATAAGTAGTATAGTAGAAAGAGGCGTAGGAACGGGTTGGACGG
911 TATACCCACCGTTGGCGGGTATCTCAGCTCATGGTGGTGGGGCGGTAGATCTAGCTA
912 TTTTTTCACTACATTTAGCAGGGGCCTCCTCTATCCTCGGCGCTATTAATTTTATCTCT
913 ACTGTAGT

914 Online Appendix C: Phenol Chloroform DNA Extraction Protocol used for *environmental DNA*.

915 Part One:

- 916 1. Filter or precipitate water sample.
- 917 2. Make a master mix of tissue lysis buffer and Proteinase K by adding 500 ul of Tissue Lysis
918 Buffer (TLB, 100 mM Tris.Cl pH8.0, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) to 20 ul of
919 Proteinase K (4mg/mL). multiply these numbers by the number of samples you want to
920 process and add 10% for error (TLS buffer is soapy).
- 921 3. mix by inverting several times, aliquot 520 uL of mixture to each tube containing your filter
922 or precipitated material.
- 923 4. Incubate overnight at 55 °C
- 924 5. Label and UV three 1.5 mL tubes for each sample (for use in part two, steps 0, 4 and 8).

925 Part Two:

926 *When working with PCI and CI a lab coat, gloves and eyewear should be worn. Part two should
927 also be done in a standard hood to prevent breathing of chemicals. Also- if PCI is pinkish in
928 color it is too old and should not be used.

- 929 0. Remove samples from incubator and remove filter from each sample by using a new 1000uL
930 pipet tip for each sample to slide the filter up the side of tube. Make sure to push hard against
931 the filter to get the tissue lysis buffer to come out of the filter. If you can fold it over once or
932 twice and squeeze solution out do so.
- 933 1. Add an equal volume of PCI (phenol chloroform isoamyl alcohol), which in this case should
934 be ~450 uL, to each tube (a little less to account for loss of TLB in filter). PCI has two
935 layers, withdraw PCI from the bottom layer (organic phase).
- 936 2. Shake manually for 5 minutes
- 937 3. Centrifuge samples for 5 minutes at 10,000 rpm
- 938 4. Pipette off supernatant (top layer) and place each sample in a clean microfuge tube, which has
939 been numbered.
- 940 5. Now add 500 ul of CI to each tube
- 941 6. Shake manually for 5 minutes
- 942 7. Centrifuge for 5 minutes at 10,000 rpm
- 943 8. Pipette off supernatant and place in a new set of microfuge tubes (if you feel that all the
944 Phenol is not removed you can repeat steps 5-8)

- 945 9. Add 40 ul of 5M NaCl to each tube
946 10. Add 900 ul (this is two equal volumes) of 100% ETOH to each tube
947 11. Freeze overnight, or for a minimum of 1 hour.
948
949 Part 3:
950 1. Centrifuge samples for 30 minutes at 10,000 rpm at 4 °C
951 2. Pour off ETOH, being careful not to lose pellet (might not be visible)
952 3. Add 900 ul of 70% ETOH (this is two equal volumes)
953 4. Centrifuge again for 30 minutes at 10,000 rpm at 4 °C
954 5. Pour off ETOH, again being careful not to lose pellet (but remove as much ETOH as
955 possible)
956 6. Let tubes air dry with lid open in laminar flow hood for 15 min to allow all ETOH to
957 evaporate
958 7. Resuspend DNA in 50-100 ul of AE buffer from qiagen kit that has been UVd.
959 8. Place in incubator at 55 °C for ten min to redissolve DNA, remove from incubator vortex
960 samples for 10 seconds and store at -20 °C.
961

962 **To make tissue lysis buffer:**

- 963 50 ml 1 M Tris.Cl pH8.0
964 5 ml 500 mM EDTA pH8.0
965 10 ml 10% SDS
966 20 ml 5 M NaCl
967 Add dH2O to 500 ml
968 Store at room temperature