Biological Conservation 183 (2015) 53–63

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Special Issue Article: Environmental DNA

Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA

Kristy Deiner [a,b,*], Jean-Claude Walser [b], Elvira Mächler [a], Florian Altermatt [a,c]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland
[b] Swiss Federal Institute of Technology (ETH), Zürich, Genetic Diversity Centre, CHN E 55 Universitätstrasse 16, 8092 Zürich, Switzerland
[c] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form 3 November 2014
Accepted 10 November 2014
Available online 3 December 2014

Keywords:
16S
Cytochrome c oxidase I
eDNA
Freshwater
Molecular protocols
Targeted species detection

ABSTRACT

Environmental DNA (eDNA) is used to detect biodiversity by the capture, extraction, and identification of DNA shed to the environment. However, eDNA capture and extraction protocols vary widely across studies. This use of different protocols potentially biases detection results and could significantly hinder a reliable use of eDNA to detect biodiversity. We tested whether choice of eDNA capture and extraction protocols significantly influenced biodiversity detection in aquatic systems. We sampled lake and river water, captured and extracted eDNA using six combinations of different protocols with replication, and tested for the detection of four macroinvertebrate species. Additionally, using the same lake water technical replicates, we compared the effect of capture and extraction protocols on metabarcode detections 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 which protocol combinations produced the highest metabarcoded biodiversity were detected and demonstrate that different protocols are required for different biodiversity targets. 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.

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 and Willerslev, 2015; 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 (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 currently used to analyze eDNA consist of capturing DNA from an environmental sample, followed by the extraction and purification of eDNA. Purified eDNA is then amplified for a specific gene target (e.g., metabarcode analysis) and 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

* Corresponding author at: Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland. Tel.: +41 (0)58 765 53 14.
E-mail address: alpinedna@gmail.com (K. Deiner).

http://dx.doi.org/10.1016/j.biocon.2014.11.018
0006-3207/

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 eDNA. These decisions start with how to capture genetic material found in the environmental sample; where "capture" is defined as the concentration of cellular or extracellular DNA (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 few 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 (Fig. 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 (Fig. 1, Table 1). The chosen capture and extraction methods we compared are perhaps the most feasible with equipment already existing in many aquatic/molecular biology laboratories, and are 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.

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

Capture method	Extraction method	Volume of water	Locus	Habitat	Targeted or metabarcode	Taxonomic group	Sequencing technology	Reference
Precipitation	Qiagen DNeasy	3 × 15 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	3 × 15 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	Metabarcode	Bacteria	Roche 454 GS-FLX-Ti	Ghai et al. (2011)
Filtration	Phenol–chloroform–isoamyl	10 L	16S rRNA	Lentic	Metabarcode	Bacteria	Roche 454 GS-FLX-Ti and Illumina GA II	Oh et al. (2011)
Filtration	Phenol–chloroform–isoamyl	45 L	16S rRNA	Lentic	Metabarcode	Bacteria	Sanger	Debroas et al. (2009)
Lyophilization	MO BIO ultraclean soil DNA kit	2 L	18S	Ground water	Metabarcode	Plant	Sanger	Poté et al. (2009)

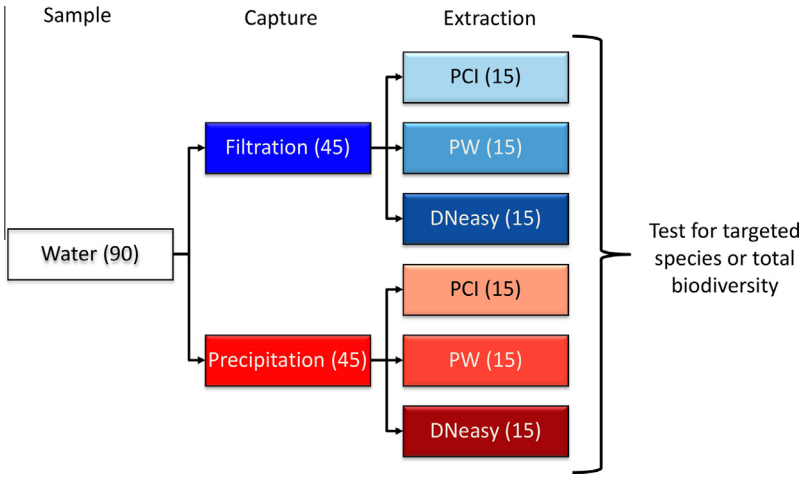

Fig. 1. Experimental design used to test for biases associated with combinations of two DNA-capture and three DNA-extraction protocols on eDNA yield, next generation sequence quantity, and biodiversity detected. Numbers of technical replicates are indicated in parentheses. Each technical replicate ($N = 90$) was a 15 mL aliquot from a single water sample, that was subjected to the six experimental treatments for both a lentic and lotic study site. PCI, PW and DNeasy are the different extraction methods performed (phenol–chloroform–isoamyl alcohol, MoBio PowerWater DNA Isolation Kit and Qiagen: DNeasy Blood & Tissue Kit respectively).

2. Methods

2.1. 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 km^2 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 h. Our goal was to compare capture and extraction methods; therefore, we used a single water sample from each site that was continuously mixed by inversion before making each of the 90 aliquots of 15 mL that represented our technical replicates. We refer to these as the technical replicates throughout the study. 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 (Fig. 1).

2.2. 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 between capture methods, 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 (Fig. 1) for each site was carried out by first attaching a filter housing (Swinnex®, EMD Millipore Co., Billerica, MA, USA) containing a glass fiber filter (GF/F, nominal pore size of 0.7 μm, 25 mm, 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 s. Precipitation for the 45 technical replicates (Fig. 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.

2.3. Environmental DNA extraction protocols

We chose three DNA extraction protocols for both targeted and universal taxon detection of aquatic organisms (Table 1, Fig. 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 h 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 Deiner and Altermatt (2014). 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).

2.4. 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 at al., 2014). 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 at al., 2014). 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 an approximate 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 650 bp that spans the first part of the 5' end of the protein coding region of 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 at al. (2014). 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). The 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 μL volumes with a final concentration of 1.6 U/μL Exo I and 0.15 U/μL SAP. The thermal-cycling regime was 15 min at 37 °C followed by 15 min at 80 °C. The cleaned PCR products 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). 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). A species was confirmed as being detected in a technical replicate if the PCR product had a single band of the expected size and could be directly sequenced with Sanger sequencing.

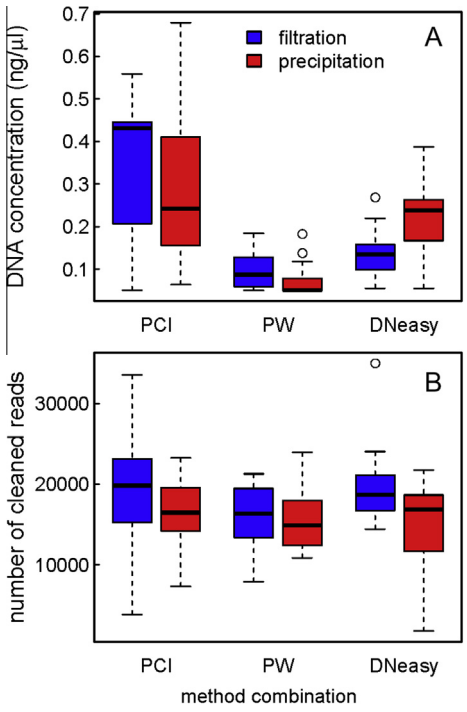

Fig. 2. Differences in DNA yield (A) and number of sequences after filtering (B) as a result of different capture and extraction methods of DNA from lentic water. Boxplots are based on 15 technical replicates. Black horizontal bars represent median values; boxes give 25% and 75% percentiles. Circles are values beyond interquartile ranges. Abbreviations are the same as in Fig. 1.

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 μL volumes with final concentrations of 1x supplied buffer (Faststart TAQ, Roche, Inc., Basel, Switzerland), 1000 ng/μL BSA (New England Biolabs, Ipswich, MA, USA), 0.2 mMol dNTPs, 2.0 mMol MgCl2, 0.05 units per μL Taq DNA polymerase (Faststart TAQ, Roche, Inc., Basel, Switzerland), and 0.54 μMol of each forward and reverse prime. 2 μL of extracted eDNA was added that ranged in concentration from 0.05 to 0.55 ng/μL. This range was the outcome of DNA concentrations that were extracted following each of the six molecular workflows (Fig. 2A). The thermal-cycling regime was 95 °C for 4 min, followed by 35 cycles of 95 °C for 30 s, either 55 °C or 48 °C (16S or COI respectively) for 30 s and 72 °C for 1 min. A final extension of 72 °C for 7 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 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 μL 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 × reaction volume 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 cleaned and pooled reaction using the Qubit (1.0) fluorometer following recommended protocols for the dsDNA BR Assay. 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 concentration of a 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 1 ng of starting DNA compared with TrueSeq (Illumina, Inc., San Diego, CA, USA) which requires 1 μg and was unrealistic to achieve from PCR on our technical replicates. Amplicons from the two metabarcode genes were pooled without any pre-labeling and were bioinformatically separated post-sequencing (see bioinformatics analysis Section 4.1).

Metabarcode 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 (2×250 nt) sequencing was performed on an Illumina MiSeq (MiSeq Reagent kit v2, 250 cycles) at the Genomic Diversity Center at the ETH, Zurich, Switzerland following manufacture'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.

3. 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 min ultraviolet light (UV) treatment. All consumables (e.g., filters, syringes, tips, tubes, gloves, etc.) used in the study were decontaminated by a 30-min UV 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 min UV 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 min in freshly mixed 10% bleach and treated with a 30 min UV 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.

4. Analysis

4.1. Bioinformatics analysis

Raw reads from next generation sequencing were quality checked with PrinSeq Lite version 0.20.3 (Schmieder and Edwards, 2011) and filtered with ConDeTri version 2.1 (Smeds and Künstner, 2011) using the following parameters: rmN, hq = 20 lq = 15, lfrac = 0.1, frac = 0.7, minlen = 100. Forward and reverse reads were merged (minimum overlap of 30 bp and maximum mismatches of 2%) using SeqPrep (St. John, 2011). Reads that could not be merged were excluded from further analysis. The merged reads were further de-replicated, de-noised (identity threshold of 99%), and chimera checked using usearch version 6 (Edgar, 2010) and remaining reads represented the cleaned read dataset (Fig. 2B, Online Appendix Table A1). The cleaned reads were de-multiplexed into the two amplicons (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 identifying "16S like" amplicons (Online Appendix Table A2). A customized reference database was built from sequences obtained from genebank by using PrimerBLAST (Ye et al., 2012) with the COI primers used in our study for reference mapping of "COI like" amplicons in QIIME. Sequences for both genes that did not match their reference databases (an identity threshold of 70% was used) were excluded from further analysis.

Taxonomic identifications of the 16S sequences greater than 250 bp were conducted in QIIME applying the RDP classifier option. 16S sequences that did not receive an identification were excluded from further analysis. 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 A3). COI sequences that did not match eukaryotes, were below 69.0% sequence similarity, or blasted to unknown environmental samples, were excluded from further analysis. Sequence FASTA files used for biodiversity detection analysis were archived at datadryad.org (doi: http://dx.doi.org/10.5061/dryad.7q19c) and all raw sequences reads were deposited in Genbank's Sequence Read Archive (*SRP049043*).

4.2. Statistical analysis

We used generalised linear models (glm) to analyze eDNA yield and number of next generation sequences after cleaning 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 of potential DNA that could be used for detection of biodiversity. 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 to analyze detection rates of the four targeted macroinvertebrate species depending on the capture and extraction methods. The response variable was proportion of positive and sequence confirmed 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 analyzes were done with the program R, version 3.0.1 (R Development Core Team, 2013).

We used a bootstrap-approach to analyze 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 analyzed. For bootstrapping each gene, we used the diversity data generated for each of the individual fifteen 15 mL technical replicates (Online Appendix Tables A2 and A3). 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.

5. Results

5.1. 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 (Fig. 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 (Fig. 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/μL 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/μL 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 (Fig. 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.

5.2. 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 (Fig. 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 (Fig. 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 where from the combinations of filtration or precipitation with PCI or PowerWater. We have provided as supplement the results summarized in Fig. 3 with all PCR confirmations of each species for comparison (Online Appendix D).

5.3. Effect of methods on eDNA metabarcode detection of eubacteria and eukaryotes

Protocol combinations biased the levels of biodiversity detected for both eukaryotes (Fig. 4A–D) and eubacteria (Fig. 4E–H). The amount of biodiversity detected also varied between technical replicates of the same protocol combination (Online Appendix

Table 2
Results of two generalized linear models (glms), explaining differences in DNA yield (A) and number of cleaned reads after filtering 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 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 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
(A) DNA yield in lentic water						
Capture	1	0.19	88	360.76	0.11	0.74
Extraction	2	147.54	86	213.23	40.89	<0.0001
Capture × extraction	2	15.91	84	197.32	4.41	0.015
Null			89	360.96		
(B) Number of cleaned reads						
Capture	1	11045.6	88	176,143		<0.0001
Extraction	2	4145.3	86	171,997		<0.0001
Capture × extraction	2	5305.1	84	166,692		<0.0001
Null			89	187,188		

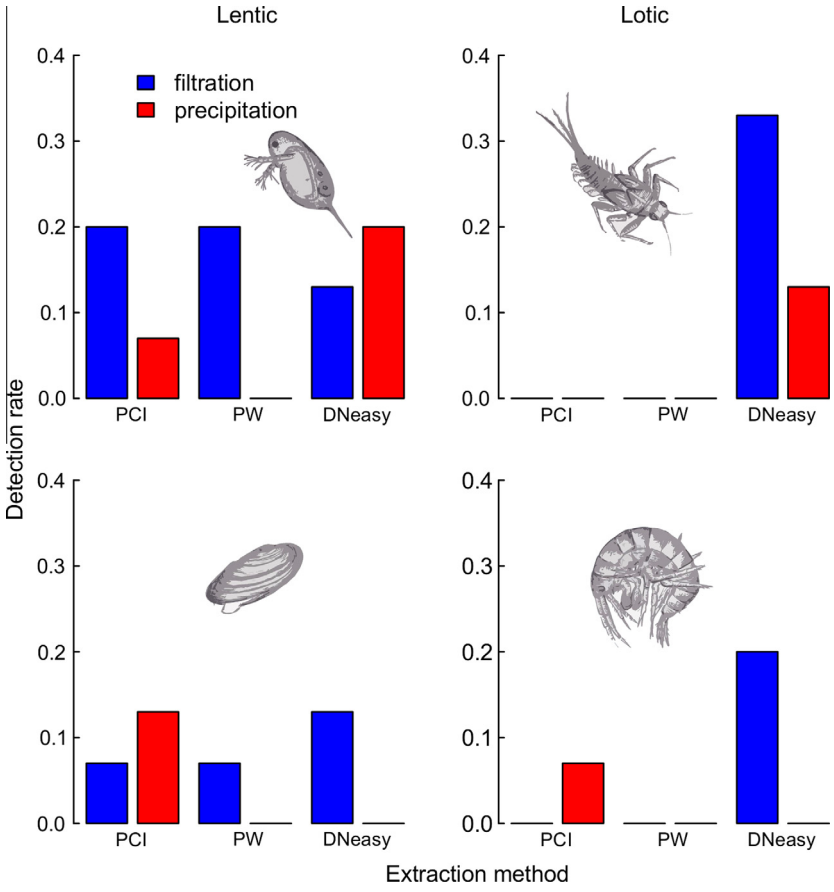

Fig. 3. Detection difference of four macroinvertebrate species as a result of capture and extraction methods. Lentic species included a water flea (top left, *Daphnia longispina*) and a mussel (bottom left, *Unio tumidus*), and lotic species included a mayfly (top right, *Baetis buceratus*) and an amphipod (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.

Table 3
Generalized linear model explaining detection rate of four targeted macroinvertebrate species as a result of species identity, capture and extraction methods, as well as their interaction. A binomial error distribution was used, and the *p*-values are based on Chi^2-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 × extraction	6	12.79	11	13.31	0.047
Null			23	46.03	

Tables A2 and 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 (Fig. 4A–H). This pattern is illustrated by the different shapes of the saturation curves for each combination of protocols (Fig. 4A–H). For eukaryotes, filtration combined with the DNeasy or PCI extraction detected more diversity for all taxonomic ranks (Fig. 4A–D). For eubacteria, there was a large overlap in diversity detected by all protocols at higher taxonomic ranks (Fig. 4E–F), whereas at lower taxonomic ranks, the precipitation and PowerWater protocol combination produced the highest amount of diversity detected (Fig. 4G–H). At the genus level of diversity, when all technical replicates were taken into account, there was a difference of more than 50 genera detected between the highest-yield and lowest-yield protocol combinations for both eukaryotes (Fig. 4D) and eubacteria (Fig. 4H). All negative controls showed no amplification for either of the two metabarcoded genes.

6. 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 in Section 7 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

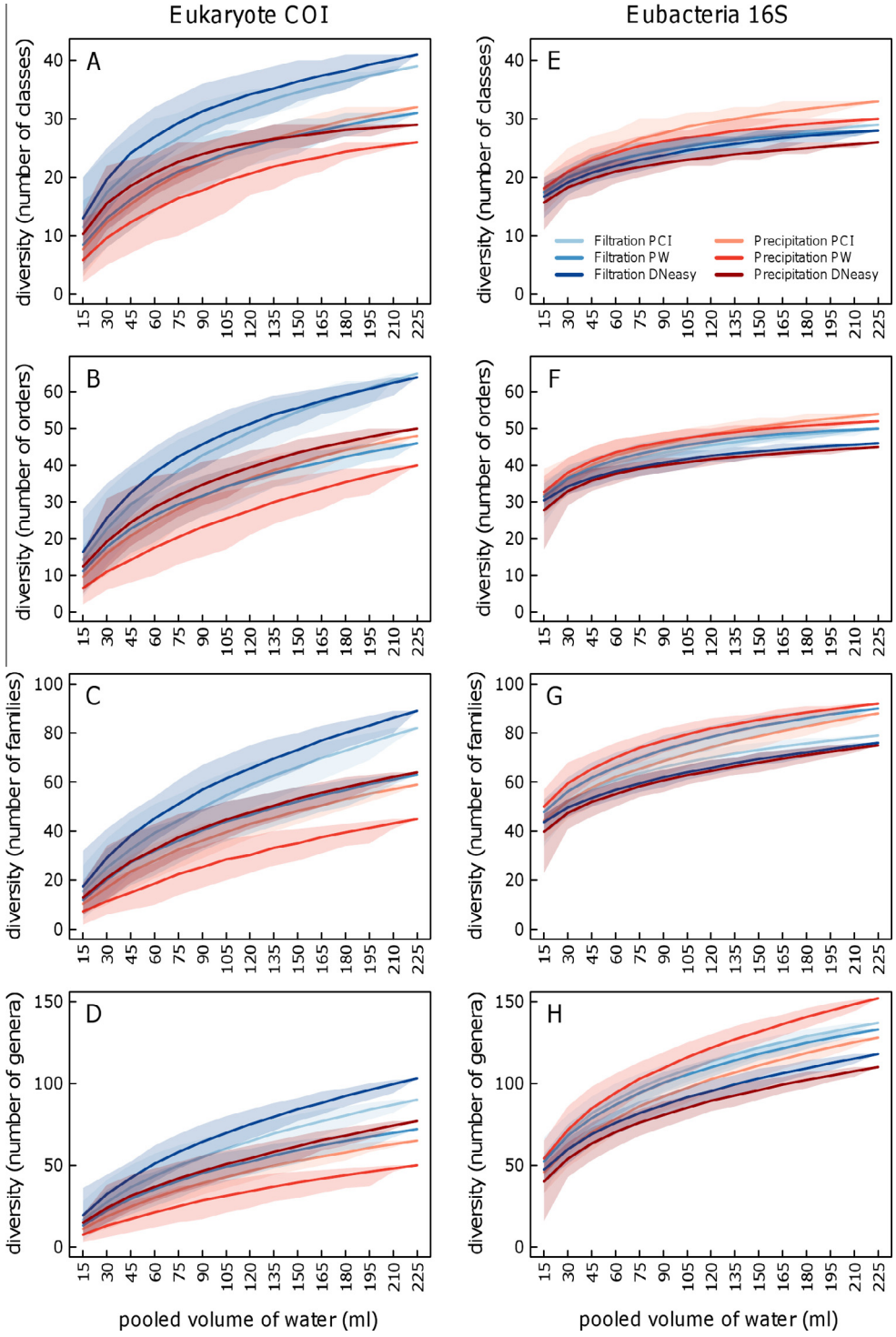

Fig. 4. Detected diversity of eukaryotes (A–D) and eubacteria (E–H) based on Illumina MiSeq data and as a result of capture and extraction methods (red and blue color, see legend) and volume of lentic water sampled. Diversity of eukaryotes and eubacteria was estimated from taxonomic identifications of COI and 16S genes respectively, and was estimated from 15 technical replicates for each of the six combinations of capture and extraction methods. Cumulative diversity curves with increasing volume of water sampled are based on bootstrapped values, giving the mean (line) and 90% confidence interval (light shaded area). Diversity is reported at the level of classes (A and E), orders (B and F), families (C and G) and genera (D and H).

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, 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 eDNA detection (for example see Piaggio et al. (2014)).

For biodiversity detected through 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 water sample, our results indicate that multiple protocols of capture and extraction can be utilized to maximize the accuracy of the biodiversity estimates from aquatic habitats.

6.1. 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 (Turner et al., 2014). 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 (Beebee, 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 (1–17%) 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 (Liang and Keeley, 2013). The higher yield of 17% of extracellular DNA was from pore sizes of 0.1 μm (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 metabarcode 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 the capture of cells or organelles 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 resulting in greater detection of eukaryotic 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 precipitation and PowerWater molecular workflow.

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 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 therefore do not think that chemical inhibition is a parsimonious explanation for our low detection patterns of targeted species.

We 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%, Fig. 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 in combination with co-extracted inhibitors or non-target 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).

6.2. 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 the exact taxonomic names assigned to every sequences as they are based on BLAST searches against what is currently in searchable databases (QIIME's gold_refdb for 16S and GenBank for COI). We therefore refrain from discussing in absolute terms what species were detected. The databases we used, however, are the most complete

databases for which to do taxonomic assignment of unknown sequences for the genes we used for metabarcoding. 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 equally, it is an unlikely cause of our detected differences. We were confined in using a taxonomic approach for analysis of our sequence data 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). With continued sequencing efforts to refine and curate sequences in databases used for identification of standard metabarcode genes, 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).

7. 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 combinations. 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 it is hoped that a single water sample can be used for eDNA metabarcode detection of both eubacteria and eukaryotes, we recommend splitting the water sample and use the specific capture and extraction protocols that performed best for each biodiversity target. Or a trade-off can be made by applying the filtration and PCI molecular workflow, as this method performed almost as well for both groups of biodiversity (Fig. 4D and H). Lastly, given the rapidly changing nature of this field, 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 and including from the literature any new technological advances.

8. Additional considerations and conclusions

We tested the differences caused by capture and extraction methods on biodiversity detected from water 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., 2015), sequencing platforms (e.g., Claesson et al., 2010) and bioinformatic pipelines (Schloss et al., 2011). All of which have been shown individually to cause biases in biodiversity detection using eDNA. Additionally, in this study we varied two steps in the 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 and thawing our sample post collection may have changed what biodiversity could be detected from the different combination of protocols (Takahara et al., 2015). However, this would have been expected to decrease diversity detected with filtration, but as this method performed well compared to precipitation, not choosing this field preservation method would likely only increase performance of detections with 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 exact 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. We thank Heather Callahan, Caren Goldberg, Mehrdad Hajibabaei and Sarah Owens for providing helpful guidance at the experimental design stage of the project. Caren Goldberg, David Pilliod, Phillip Thomsen, Cameron Turner, and two anonymous reviewers who provided helpful comments on previous drafts for this manuscript. Data analyzed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. Funding was from the Swiss National Science Foundation Grant 31003A_135622 (to F.A.) and Eawag discretionary funds (to K.D.).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biocon.2014.11.018.

References

Altermatt, F., 2013. Diversity in riverine metacommunities: a network perspective. Aquat. Ecol. 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. J. Biogeogr. 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. Anal. Biochem. 78, 569–571.
Beebee, T.J., 1991. Analysis, purification and quantification of extracellular DNA from aquatic environments. Freshw. Biol. 25, 525–532.
Benson, D., Karsch-Mizrachi, I., Clark, K., Lipman, D., Ostell, J., Sayers, E., 2012. GenBank. Nucleic Acids Res. 40, D48–D53.
Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., Battin, T.J., 2013. Headwaters are critical reservoirs of microbial diversity for fluvial networks. Proc. R. Soc. B: Biol. Sci., 280.
Caldwell, J.M., Raley, M.E., Levine, J.F., 2007. Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. Environ. Sci. Technol. 41, 3277–3283.
Callahan, H., 2009. Inhibitor-free DNA purification from water samples. Biotechniques, 46.
Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336.

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods 69, 330–339.
Chandler, D.P., Fredrickson, J.K., Brockman, F.J., 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Mol. Ecol. 6, 475–482.
Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., O'Toole, P.W., 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 38, e200.
Costas, B.A., McManus, G., Doherty, M., Katz, L.A., 2007. Use of species-specific primers and PCR to measure the distributions of planktonic ciliates in coastal waters. Limnol. Oceanogr. Methods 5, 163–173.
Crawley, M.J., 2013. The R Book, Second ed. John Wiley & Sons, West Sussex, United Kingdom.
Darling, J.A., Mahon, A.R., 2011. From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. Environ. Res. 111, 978–988.
Debroas, D., Humbert, J.F., Enault, F., Bronner, G., Faubladier, M., Cornillot, E., 2009. Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget–France). Environ. Microbiol. 11, 2412–2424.
DeFlaun, M.F., Paul, J.H., Davis, D., 1986. Simplified method for dissolved DNA determination in aquatic environments. Appl. Environ. Microbiol. 52, 654–659.
Deiner, K., Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river. PLoS One 9, e88786.
Deiner, K., Knapp, R.A., Boiano, D.M., May, B., 2013. Increased accuracy of species lists developed for alpine lakes using morphology and cytochrome oxidase I for identification of specimens. Mol. Ecol. Resour. 13, 820–831.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.
Ficetola, G.F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using environmental DNA from water samples. Biol. Lett. 4, 423–425.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech. 3, 294–299.
Ghai, R., Rodríguez-Valera, F., McMahon, K.D., Toyama, D., Rinke, R., de Oliveira, T.C.S., Garcia, J.W., de Miranda, F.P., Henrique-Silva, F., 2011. Metagenomics of the water column in the pristine upper course of the Amazon River. PLoS One 6, e23785.
Goldberg, C.S., Pilliod, D.S., Arkle, R.S., Waits, L.P., 2011. Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS One 6, e22746.
Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J., Waits, L.P., 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). Freshw. Sci. 32, 792–800.
Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One 6, e17497.
Hebert, P.D., Cywinska, A., Ball, S.L., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London, B 270, 313–321.
Jerde, C.L., Mahon, A.R., Chadderton, W.L., Lodge, D.M., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conservation Lett. 4, 150–157.
Jiang, J., Alderisio, K.A., Singh, A., Xiao, L., 2005. Development of procedures for direct extraction of cryptosporidium DNA from water concentrates and for relief of PCR inhibitors. Appl. Environ. Microbiol. 71, 1135–1141.
Jones, M., Ghoorah, A., Blaxter, M., 2011. JMOTU and Taxonerator: turning DNA Barcode sequences into annotated operational taxonomic units. PLoS One 6, e19259.
Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon, Y.S., Lee, J.-H., Yi, H., Won, S., Chun, J., 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int. J. Syst. Evol. Microbiol. 62, 716–721.
Kitchen, R.R., Kubista, M., Tichopad, A., 2010. Statistical aspects of quantitative real-time PCR experiment design. Methods 50, 231–236.
Kreader, C.A., 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl. Environ. Microbiol. 62, 1102–1106.
Lemarchand, K., Berthiaume, F., Maynard, C., Harel, J., Payment, P., Bayardelle, P., Masson, L., Brousseau, R., 2005. Optimization of microbial DNA extraction and purification from raw wastewater samples for downstream pathogen detection by microarrays. J. Microbiol. Methods 63, 115–126.
Liang, Z., Keeley, A., 2013. Filtration recovery of extracellular DNA from environmental water samples. Environ. Sci. Technol. 47, 9324–9331.
Mächler, E., Deiner, K., Steinmann, P., Altermatt, F., 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshw. Sci. 33, 11174–11183.
Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Cold Spring Harbor, NY.
Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N., Kawabata, Z.I., 2012. Surveillance of fish species composition using environmental DNA. Limnology 13, 193–197.
Morgan, J.L., Darling, A.E., Eisen, J.A., 2010. Metagenomic sequencing of an in vitro-simulated microbial community. PLoS One 5, e10209.
Muyzer, G., De Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700.
Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R., Konstantinidis, K.T., 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. Appl. Environ. Microbiol. 77, 6000–6011.
Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E., Avery, M.L., 2014. Detecting an elusive invasive species: a diagnostic PCR to detect burmese python in Florida waters and an assessment of persistence of environmental DNA. Mol. Ecol. Resour. 14, 374–380.
Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P., Richardson, J., 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. Can. J. Fish. Aquat. Sci. 70, 1123–1130.
Poté, J., Mavingui, P., Navarro, E., Rosselli, W., Wildi, W., Simonet, P., Vogel, T.M., 2009. Extracellular plant DNA in Geneva groundwater and traditional artesian drinking water fountains. Chemosphere 75, 498–504.
Rajendhran, J., Gunasekaran, P., 2008. Strategies for accessing soil metagenome for desired applications. Biotechnol. Adv. 26, 576–590.
Schloss, P.D., Gevers, D., Westcott, S.L., 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6, e27310.
Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina metabarcoding of a soil fungal community. Soil Biol. Biochem. 65, 128–132.
Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.
Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794–1805.
Smeds, L., Künstner, A., 2011. ConDeTri-a content dependent read trimmer for Illumina data. PLoS One 6, e26314.
St. John, J., 2011. SeqPrep. Available <https://github.com/jstjohn/SeqPrep>.
Sutherland, W.J., Bardsley, S., Clout, M., Depledge, M.H., Dicks, L.V., Fellman, L., Fleishman, E., Gibbons, D.W., Keim, B., Lickorish, F., 2012. A horizon scan of global conservation issues for 2013. Trends Ecol. Evol..
Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. Mol. Ecol. 21, 1789–1793.
Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc. Natl. Acad. Sci. 109, 16208–16212.
Team, R.D.C., 2013. R: A Language and Environment for Statistical Computing. R Project for Statistical Computing Vienna, Austria.
Thompson, D., Rajal, V., Batz, S., Wuertz, S., 2006. Detection of Salmonella spp. in water using magnetic capture hybridization combined with PCR or real-time PCR. J. Water Health 4, 67–75.
Thomsen, P., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Willerslev, E., 2012. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 21, 2565–2573.
Tichopad, A., Kitchen, R., Riedmaier, I., Becker, C., Ståhlberg, A., Kubista, M., 2009. Design and optimization of reverse-transcription quantitative PCR experiments. Clin. Chem. 55, 1816–1823.
Tringe, S.G., Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental samples. Nat. Rev. Genet. 6, 805–814.
Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L., Lodge, D.M., 2014. Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods Ecol. Evol. (In press).
Vermeulen, S., Koziell, I., 2002. Integrating global and local values: a review of biodiversity assessment. IIED.
Wang, J., McLenachan, P.A., Biggs, P.J., Winder, L.H., Schoenfeld, B.I., Narayan, V.V., Phiri, B.J., Lockhart, P.J., 2013. Environmental bio-monitoring with high-throughput sequencing. Briefings Bioinformatics 14, 575–588.
Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173, 697–703.
Wintzingerode, F.V., Göbel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21, 213–229.
Wright, E.S., Yilmaz, L.S., Ram, S., Gasser, J.M., Harrington, G.W., Noguera, D.R., 2014. Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. Environ. Microbiol. 16, 1354–1365.
Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134.
Yoccoz, N.G., 2012. The future of environmental DNA in ecology. Mol. Ecol. 21, 2031–2038.
Zinger, L., Gobet, A., Pommier, T., 2012. Two decades of describing the unseen majority of aquatic microbial diversity. Mol. Ecol. 21, 1878–1896.