Author/s:
Nichols, SJ;Kefford, BJ;Campbell, CD;Bylemans, J;Chandler, E;Bray, JP;Shackleton, M;Robinson, KL;Carew, ME;Furlan, EM

Title:
Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa

Date:
2020-04-01

Citation:
Nichols, S. J., Kefford, B. J., Campbell, C. D., Bylemans, J., Chandler, E., Bray, J. P., Shackleton, M., Robinson, K. L., Carew, M. E. & Furlan, E. M. (2020). Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: Effects of debris and storage conditions on the recovery of target taxa. Freshwater Biology, 65 (4), pp.607-620. https://doi.org/10.1111/fwb.13443.

DR. SUSAN NICHOLS (Orcid ID : 0000-0002-3553-8009)

Article type : Standard Paper

Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: effects of debris and storage conditions on the recovery of target taxa

Nichols S.J.[1], Kefford B.J.[1], Campbell C.[1], Bylemans J.[1,5], Chandler E.[2], Bray J. P.[1], Shackleton M.[3], Robinson K.L.[4], Carew M.E.[4], Furlan E.M.[1]

1. Institute for Applied Ecology, University of Canberra, Bruce, ACT, Australia 2617
2. School of Environmental and Rural Science, University of New England, Armidale, NSW, 2350
3. School of Life Sciences, La Trobe University, Wodonga, Victoria, Australia 3690
4. School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia 3052
5. Research and Innovation Centre, Fondazione Edmund Mach, S. Michele all'Adige TN 38010, Italy

Correspondence

Susan Nichols, Institute for Applied Ecology, University of Canberra, Bruce, ACT, Australia 2617.

Email: sue.nichols@canberra.edu.au

Running Head- DNA metabarcoding of freshwater macroinvertebrates

SUMMARY

1. Macroinvertebrates are commonly sampled for bioassessment of freshwater ecosystems. However, current bioassessment protocols involve laborious sorting of the

animals from the debris (sample matrix) and morphological identification, where species level identifications are often difficult. DNA metabarcoding has the potential to improve bioassessment by reducing the time taken to process samples and improve the accuracy and speed of macroinvertebrate species identifications.

2. In this study, we evaluated DNA metabarcoding of macroinvertebrate samples, which include macroinvertebrates and the debris collected in the sample nets, to test if bulk, unsorted samples can be used to assess macroinvertebrate diversity. First, we tested if the sample matrix prevented the detection of six target macroinvertebrate taxa when DNA metabarcoding. Second, we tested if sample storage influenced the detection of the same six target macroinvertebrates. We also explored different levels of replication at the sample, sub-sample and PCR level and compared the overall macroinvertebrate families detected using DNA metabarcoding to those identified morphologically.

3. We found that the presence of the sample matrix did not interfere with or inhibit the detection of the six target macroinvertebrate taxa. Furthermore, we found that the various sample storage methods did not affect target macroinvertebrate detection. The reliability of detection of the target macroinvertebrates improved as hierarchical levels of replication were combined. We found strong overlap between the detection of overall macroinvertebrate family diversity when comparing DNA metabarcoding to morphological identification.

4. Extracting DNA from the bulk macroinvertebrate samples that included the sample matrix and using this for DNA metabarcoding could improve bioassessment by removing the need for laborious sorting of samples. Furthermore, DNA metabarcoding detection of the six target taxa was not dependent on sample storage of up to one year in 95% ethanol, at room temperature or after heating. DNA metabarcoding had the advantage of identifying macroinvertebrate species, but good DNA barcode libraries are needed for widespread species identifications. Further investigation should focus on including multiple samples with different macroinvertebrate composition and densities to refine and standardise bulk sample processing protocols, and on building comprehensive DNA barcode libraries for aquatic macroinvertebrates.

Keywords DNA barcoding, biological monitoring, freshwater invertebrates

Introduction

Large-scale freshwater bioassessment programs that sample macroinvertebrates are time-consuming. Macroinvertebrates need to be sorted from the sample matrix, which is collected

in sampling nets, so each individual can be morphologically examined and identified by a taxonomist (Haase *et al.*, 2010; Nichols *et al.*, 2017). This is made more problematic by the increasing difficulty in finding appropriately trained taxonomists to identify macroinvertebrates, especially to species level (Buyck, 1999). Furthermore, sorting and morphological identification can introduce errors that affect data quality and therefore the accuracy of bioassessments. Sources of error can include: the failure to remove representatives of all target taxa; preferential selection of large and conspicuous taxa; inability to correctly identify damaged and/or small taxa; incorrect identifications, especially with early instars where morphological characters are not fully developed; and some taxa are essentially impossible to identify to species level (Haase *et al.*, 2010). Most bioassessment programs have developed quality control procedures to mitigate these errors, but they cannot always be eliminated. Costs are usually balanced with the need for sufficient information to distinguish differences between sites (Jones, 2008). To reduce costs and improve speed, often only a sub-set of the collected sample is sorted in the laboratory or the sample is live-sorted in the field and, to reduce taxonomic error, samples are identified to mostly family (Nichols & Norris, 2006). However, the development of DNA-based identification, like DNA metabarcoding, has the potential to improve bioassessment by providing reliable and routine macroinvertebrate species identification, which could reduce time and errors associated with sorting and morphological identifications (Baird & Hajibabaei, 2012; Hebert *et al.*, 2003).

DNA barcoding is a well-established approach widely used for species identification, (Hebert *et al.*, 2003). It has been repeatedly shown to reveal more macroinvertebrate species than morphological identification (e.g. Jackson *et al.*, 2014; Sweeney *et al.*, 2011). Currently, over a million DNA barcodes are available to identify freshwater macroinvertebrates species through the Barcode of Life - BOLD systems 4 database (http://v4.boldsystems.org/) (Carew *et al.*, 2017). Using high throughput DNA sequencing (HTS) technology, it is now feasible and cost-effective to use species DNA barcodes for routine bioassessment through a process known as DNA metabarcoding (Yu *et al.*, 2012). DNA metabarcoding allows high-throughput and reliable identification of macroinvertebrate species biodiversity in multiple samples, simultaneously (e.g. Carew *et al.*, 2018; Elbrecht *et al.*, 2017; Porter & Hajibabaei, 2018). Unlike morphological identification, DNA metabarcoding has the benefit of being less susceptible to the suite of potential errors identified above and provides easy access to species level information (Carew *et al.*, 2017; Curry *et al.*, 2018). The species information obtained through DNA barcoding and

metabarcoding can increase the sensitivity of structural and functional assessments that rely on finer taxonomic resolution and potentially improve detection of small changes in stream condition (Stein *et al.*, 2014).

Current DNA metabarcoding protocols developed for assessing macroinvertebrate biodiversity often use sorted macroinvertebrates as a source of DNA (Carew, Coleman & Hoffman, 2018; Elbrecht & Steinke, 2019). However, if macroinvertebrates could be directly processed with the sample net contents for DNA metabarcoding this would allow considerable savings in time and effort and remove many of the errors introduced in the sorting process (Haase *et al.*, 2010; Macher *et al.*, 2018). Few studies have examined the feasibility of processing bulk samples where animals have not been sorted from sample matrix, but it appears to offer a viable means of speeding up routine DNA metabarcoding of macroinvertebrate samples (see Macher *et al.*, 2018; Majaneva *et al.*, 2018). Yet, this approach requires more evaluation before it could be successfully used for routine monitoring. Ideally, it needs development of consistent, standardized DNA metabarcoding protocols that examine the robustness of DNA metabarcoding at different level of replication, sample types (e.g. from different habitats) and likely field and laboratory sample storage scenarios (Pawlowski *et al.*, 2018; Pilgrim *et al.*, 2011).

The main aim of this study is to evaluate whether the presence of the sample matrix in unsorted samples interferes with detection of target macroinvertebrate taxa during DNA metabarcoding. We do this by first setting up different mock samples and examining the detection of six target macroinvertebrate taxa at different levels of replication (sample, sub-sample and PCR). We then use this same design to examine the effects of sample preservation conditions and time on detection by reproducing different sample storage conditions likely to be encountered in the field and laboratory. We then compare the overall detection of macroinvertebrate richness assessed through DNA metabarcoding to family-based morphological identification of samples. Finally, we discuss what is needed for DNA metabarcoding of unsorted macroinvertebrate samples to be used in routine freshwater bioassessment.

Methods

Macroinvertebrate collection and mock sample construction

Macroinvertebrate samples were collected with a 250 µm mesh kick-net from two locations on the Goodradigbee River, New South Wales (NSW), Australia (S35°25'28.95", E148°43'44.88") according to standard Australian River Assessment System (AUSRIVAS) bioassessment methods (Nichols *et al.*, 2000). The river has a predominately cobble substrate, and from previous surveys, is known to be in good condition with a diverse macroinvertebrate community. The entire contents from kick nets was placed in containers and preserved in the field with 95% ethanol and transported on ice to the laboratory. In the laboratory, each sample was first processed using a standard bioassessment protocol where the sample was sub-sampled (see Marchant, 1989; Nichols *et al.*, 2000) and approximately 200 animals removed. The whole of the sub-sampled material represented an estimated 1% of each collection. These sorted macroinvertebrates were then identified morphologically to family taxonomic level, except Chironomidae (to sub-family), Oligochaeta (to class) and Acarina (to order). The remaining sample residue was then visually scanned with the aid of a large magnifying lamp for 15 minutes and any additional taxa that were not found in the ~200 animal sub-sample were collected for identification. By conducting a visual scan, a more complete taxa list can be obtained by incorporating taxa that were not collected in the ~200 organism sub-sample (Table S1). The remaining 99% of sample residue that was not sub-sampled was used to construct the samples to test DNA metabarcoding (Figure 1, Table 1).

Before preparing the samples for the metabarcoding experiment, visible macroinvertebrates from the remaining sample residues were separated from the sample matrix and retained. The sample matrix of leaves (willow), sediment and detritus was mixed by stirring and then added to sterile 5 ml sample vials in approximate equal volumes taking care to add a similar sized leaf to each. These vials were used for the 'Matrix' only samples and also used to construct the mock bulk samples (Figure 1). The six macroinvertebrate target taxa were selected from the retained macroinvertebrates and combined with sample matrix to construct the 'Bulk' mock samples and also used to make a 'Sorted' sample type that contained only the six target taxa (i.e. no sample matrix added). To construct the Bulk and Sorted samples, we used three each of three smaller taxa (Simuliidae, Leptophlebiidae and Baetidae) and approximately 1/2 of an individual of three larger taxa (Coloburiscidae, Hydrobiosidae and Philopotamidae) for each sample, representing 12 individuals (or part of) per sample (Table 1). By adding the six target macroinvertebrate taxa we ensured their presence in all Sorted and Bulk mock samples and expected their detection with DNA metabarcoding. Approximately 5% by weight of the mock Bulk samples was made up of

target taxa. Although not specifically added to the Matrix samples, it was possible that DNA from the six macroinvertebrate target taxa was present through residual DNA in samples, such as eDNA, eggs or body parts. The samples were then stored under different conditions for the Time-1 and Time-2 experiments (described below).

Time-1 experiment: Detection of target macroinvertebrates in bulk mock field samples after short-term storage.

The Time-1 DNA metabarcoding experiment was used to determine if the presence of the sample matrix interferes with, or inhibits, the detection of the six target macroinvertebrates (Table 1). This was tested using three replicates of the Sorted, Bulk and Matrix samples, which each had DNA extracted (see below) in three replicates and were then each amplified using three PCR replicates. A single control sample of 95% ethanol was used to test for sample contamination. The Sorted samples of six macroinvertebrate taxa only were used as a positive control because no matrix related inhibition was expected; the Bulk mock sample, which contained the same six macroinvertebrate taxa plus the sample matrix was used to test the effect of the sample matrix on the detection of the six target taxa; while the Matrix sample was used to see if the target taxa would be detected from trace amounts of DNA contained in the matrix. The Time-1 experiment used samples stored at room temperature and processed within two months of collection. We also examined the total number of macroinvertebrate taxa detected by DNA metabarcoding from both the Bulk and Matrix samples (this did not include the Sorted samples because these only contained the six target macroinvertebrate taxa) for comparison to the morphological family identifications.

Time-2 experiment: Effect of storage conditions on the detection of target macroinvertebrates

If not kept on ice or refrigerated, the bulk samples collected from the field may be exposed to heat (in vehicles) during transportation to the laboratory and are often stored at room temperature for many months before processing. To test the effect of heat, storage time and temperature on the detection of taxa, the Sorted, Bulk and Matrix samples were subjected to three different treatments for 12 months. Treatments included: storing samples at room temperature (room temp samples) for direct comparison to the Time-1 experiment room-temperature samples; heat treating samples in a water bath for three hours at 50 ^{o}C before storing at room temperature (heated then room temp samples); and storing (in 95%) ethanol in a freezer at -20 ^{o}C after arrival in the laboratory (freezer samples) i.e. the conditions considered best for preserving DNA (Lear *et al.*, 2018). Data loggers recorded the ambient

room temperature in the laboratory, which averaged 21°C (range 15-28 °C). As in the Time-1 experiment, the same design was used with three sub-sample replicates of the Sorted, Bulk and Matrix samples that were DNA extracted in three replicates then amplified using three PCR replicates. This study design enabled comparison of Time-1 and Time-2 data to test for the effect of three possible sample storage practices on macroinvertebrate detection using DNA metabarcoding, as well as assessment of the level of replication (at various stages of the processing i.e. sub-sample, extraction, and PCR replicates) required for maximum detection of the six target macroinvertebrate taxa. Time-2 had a total of three control samples (one for heat treated, one for room temperature and one for freezer) of 95% ethanol to test for sample contamination.

Sample homogenization and DNA extraction

Most ethanol was removed from the samples and three 5 mm 304-stainless steel G100-grade ball bearings were placed into each of the 5 ml tubes containing each sample and homogenised on a bench-top Vortex Genie 2 homogeniser (GENEWORKS Pty Ltd) on the highest speed. After 10 minutes the samples were rotated and homogenised for a further 10 minutes to ensure they were well macerated. Samples were then centrifuged at 4700rpm for 60 secs and remaining ethanol removed. After sample homogenization, three sub-samples were collected from each sample for DNA extraction. All material from Sorted and ethanol control samples and ~10 mg of Bulk and Matrix samples were added to individual 1.5 ml Eppendorf tubes and DNA extracted using the Qiagen DNeasy blood and tissue extraction kit following the manufacturer's recommendations. DNA extracts were stored at 4 ºC until further processing. From each DNA extraction, three PCR replicates were used to obtain a total of 27 replicates (three samples, three sub-samples with one extraction per sub-sample, and three PCR replicates per sub-sample) per treatment and relevant storage period (**Figure captions**

Figure 1). The control samples were subjected to the same process.

Screening for PCR inhibition using real-time PCR

The potential effect of PCR inhibitors was first evaluated to determine the optimal working solution for HTS library preparation. The DNA extracts from all samples (i.e. excluding the negative extraction controls) were used to create a dilution series (i.e. neat DNA extracts, 10-fold and 100-fold DNA dilution). The obtained dilutions were subsequently screened using a SYBR® Green Real-Time PCR assay. A single PCR replicate was performed for each

dilution using the universal invertebrate primers BF1 and BR1 (Elbrecht & Leese, 2017). Individual PCR reactions contained 0.20 µL of AmpliTaq Gold DNA Polymerase (5 U/µL; Applied Biosystems, Foster City, CA, USA), 2.50 µL of GeneAmp Gold Buffer (10X; Applied Biosystems), 2.00 µL of $MgCl_2$ (25 nM), 0.65 µL of GeneAmp dNTP Blend (10 nM; Applied Biosystems), 0.20 µL UtraPure Bovine Serum Albumin (50 mg/mL; Invitrogen, Carlsbad, CA, USA), 0.60 SYBR® Green I Nucleic Acid Gel Stain (5X; Invitrogen), 1.00 µL of each primer (10 µM), 2.00 µL of template DNA and DEPC-treated water to a final volume of 25 µL. Real-time PCR reactions were run using the Viia7 Real-Time PCR system (Applied Biosystems). Thermal cycling conditions consisted of an initial activation of 5 min at 95 °C; 50 3-step cycles of 30 sec at 95 °C, 30 sec at 50 °C and 45 sec at 72 °C; a melting curve analyses consisting of a continuous increase of 0.2 °C/sec from 60 °C to 95 °C and a final extension of 10 minutes at 72 °C. The optimal template DNA concentration used for the preparation of the DNA metabarcoding libraries was determined for each sample based on the cycle threshold values (Ct values), the shape of the amplification curves and melt-curves. In addition, for all negative extraction controls, three PCR replicates were performed to determine the presence of invertebrate DNA and if this initial screen resulted in positive amplifications the extraction controls were also used in the DNA metabarcoding library construction.

Construction of DNA metabarcoding libraries

Separate DNA metabarcoding libraries were constructed for Time-1 and Time-2 experiments. Each DNA metabarcoding library was constructed using one-step PCR amplification with fusion-tagged primers (FTPs). While the use of FTPs may increase variability among replicates (Zizka *et al.*, 2019; O'Donnell *et al.*, 2016), a one-step amplification protocol was used here as it is less time and cost consuming and reduce the potential for cross-contamination and the formation of chimeric sequences. The FTPs contained Illumina MiSeq sequencing adaptor sequences and custom sequencing primers followed by a 7 bp molecular identification tag (MID-tag) and the invertebrate specific BF1or BR1 PCR primers (Elbrecht & Leese, 2017). Three PCR replicates were performed for each DNA extract with each PCR replicate containing a unique combination of MID-tags (Table S2). PCR conditions and thermal cycling profiles used for the library construction were as indicated above. Amplicons from individual PCR replicates were pooled based on DNA concentrations calculated from the Ct-values. The created amplicon pools were cleaned using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) in a 0.8 volume ratio relative to the pooled amplicon

product. Amplicon concentrations for each pool were determined using a Qubit HS Assay (Invitrogen) and equal amounts of the amplicon libraries were combined into a single tube. A final clean-up step was performed before Illumina MiSeq sequencing which was performed by the Ramaciotti Centre for Genomics (UNSW, Sydney) using a paired-end MiSeq run with the v3 2x300 bp sequencing kit.

Bioinformatic analysis

Trimmomatic version 0.36 (Bolger, Lohse & Usadel, 2014) was used to trim sequencing adaptors and primers from the sequencing reads. Simultaneously, low-quality bases (i.e. below quality 3) at the end of the sequencing reads were removed and a sliding window of 4-bases was used to trim reads when the average quality per base was below 15. Further bioinformatics filtering of the sequence reads followed the general workflow described in De Barba *et al.* (2014). The OBITOOLS (Boyer *et al.*, 2016) scripts were used to assign sequence reads to their respective samples and remove sequences less than 150 bp in length and sequences with an occurrence below a count of 10 sequences. PCR sequencing errors were removed, and identical sequences were clustered together as Molecular Operational Taxonomic Units (MOTUs) as described in De Barba *et al.* (2014). Low-abundance MOTUs were filtered from the dataset on a sample-wise basis; MOTUs were removed from a sample, using a custom Python script (Robbins-Pianka, 2013), if they made up <0.1% of the reads in that replicate, because low frequency reads largely represent PCR artefacts.

Taxonomic information was assigned to the sequence records in a two-step process. First, a taxonomic database was built using the standard invertebrate sequence records from the European Molecular Biology Laboratory data repository (release 137) using the ECOPCR scripts and the ECOTAG script to assign taxonomic data to the sequences. To improve on ambiguous taxonomic assignments, a second assignment step was included. All sequences with a score below 0.97 were extracted and re-evaluated using a custom reference database (i.e. a database of available DNA barcoding regions of aquatic invertebrates occurring in Australia – project codes AIA and AIC in the BOLD data repository). The second taxonomic assignment was performed in Geneious version 10.0.9 (Kearse *et al.*, 2012) and sequences were assigned to species (pairwise identity ≥ 97%), genus (97% > pairwise identity ≥ 95%) or family (95% > pairwise identity ≥ 90%). The data from both taxonomic assignments was imported into R version 3.4.1 (R Development Core Team, 2010) and combined for further analyses.

To examine the relationship between sequencing depth and taxonomic detection, multiple rarefaction analyses were performed in QIIME2 (Bolyen *et al.*, 2018), using the diversity plugin and the alpha-rarefaction visualizer. For subsequent statistical analyses, sequencing depth was standardised at 1,000 reads per sample, using the single rarefaction script within QIIME1 (Caporaso *et al.*, 2010).

Statistical analysis

We evaluated DNA metabarcoding detection of target taxa using various analyses as follows, and performed in the R statistical package (R Development Core Team, 2010). Factorial Poisson models were used on Time-1 data to evaluate how efficient the DNA metabarcoding was in detecting the six target macroinvertebrate taxa in the Sorted, Bulk and Matrix samples. This analysis was repeated for the Time-2 experiment and included tests for the effects of storage method (frozen, heat treated, room temperature). To determine the effect of storage time (12 months), we compared Time-1 and Time-2 samples stored at room temperature. We used both Time-1 and Time-2 samples to determine what level of replication (at various stages of the processing i.e. sample, extraction and PCR replicates) was required for maximum detection of the target taxa. Sample type (Matrix, Sorted and Bulk), storage method (Frozen, Heated and Room temperature), and in one analysis of room temperature data in isolation, experiment factors (Time-1, Time-2) were treated as fixed factors. Random hierarchical components were included in all models (sub-replicates and extraction sub-replicates).

Comparison of morphological and metabarcoding detections

Using the same taxonomic level as the morphological data i.e. family level, except Chironomidae (sub-family), Oligochaeta (class) and Acarina (order), the results from DNA metabarcoding of all replicates of the Bulk and Matrix room temperature samples from Time-1 and Time-2 experiments (all storage treatment replicates combined) were compared with the macroinvertebrate list obtained from the morphologically identified sub-sample taken for the purposes of routine bioassessment at the beginning of the study (i.e. from the same kick-net collections). Euler diagrams and Venn diagrams (created using the R package VennDiagram) (Chen, 2018) were used for visual comparisons.

Results

The screening of the samples for PCR inhibitors showed that for all sample types most samples produced the best results when they were undiluted (i.e. 55.6%, 77.8% and 94.4% of the samples from the Sorted, Matrix and Bulk samples respectively). For the Matrix and Bulk samples 22.2% and 5.6% of the samples showed an increased amplification efficiency when samples were diluted 10-fold. Unexpectedly, the sorted samples had the highest percentage of samples that showed better amplification results when diluted 10 (19.4%) or 100-fold (25%) (see detailed qPCR results in Supplementary Material Table S3).

After filtering, a total of 12,952,370 sequences reads were obtained from both MiSeq runs combined and used for further analysis. Sequencing depth on a per replicate basis was highly variable with an average 42,273 reads per replicate (sd = 32,5256, range = 13-199,112; see reads per sample in Table S4). The percentage of reads assigned to invertebrate taxa was high with on average 99.78% (sd = 0.40), 98.42% (sd = 1.89) and 93.33% (sd = 15.06) for the Sorted, Bulk and Matrix samples respectively. The proportion of invertebrate reads assigned to the six target taxa was also high in both Time-1 and Time-2 (Table 2). Because metabarcoding can be highly sensitive, it is not unusual to observe some reads for negative control samples, despite the use of careful laboratory techniques. Contaminant DNA, if present, is likely to show disproportionately higher amplification in negative control samples, as they do not contain any added template DNA, which competes with potential contaminant DNA for PCR reagents. Read counts for the negative control samples were generally orders of magnitude lower than for the experimental samples following MOTU clustering (range = 13 – 656; median = 161), indicating only a low-level of background contamination was present. We therefore considered any potential background contamination to be successfully mitigated by our low-abundance read filtering strategy.

Overall (Time-1 and Time-2 data combined), 2561 unique invertebrate MOTUs were observed. The MOTUs were assigned to 76 unique taxa, of which 44 were identified to species level, and the rest at higher taxonomic levels (see Supplementary Material for morphological and DNA taxonomic lists, Table S1 and Table S5). All six macroinvertebrate target taxa were detected in both Time-1 and Time-2 samples. Time-1 samples detected a total of 52 unique taxa and Time-2 a total of 75 unique taxa.

For statistical analyses and for visualisation of the data with box plots, read counts were normalised across replicates. We found that a rarefaction level of 1,000 reads per replicate represented a good compromise between maximising taxonomic detection and minimising the number of low read count samples that had to be excluded from these

analyses (Figure S1). In total, 21 replicates (out of a total of 327) were excluded from the dataset because of low read counts (these included all eight negative control replicates, six Bulk replicates, three Matrix replicates, one Sorted replicate). At this sequencing depth, we were able to retain 90% of PCR replicates and still capture most of the taxonomic diversity (typically ≥ 90% of MOTUs per replicate), and thus obtained a good assessment of overall community composition for the comparison of treatment effects. However, our ability to detect low-abundance taxa was compromised somewhat, as evidenced by the relationship between non-normalised sequencing depth and the number of families detected per sample (Figure 2). Approximately 40,000 sequence reads per PCR replicate were required to reliably detect low-abundance taxa in the Bulk and Matrix samples. For the sorted samples, only 10,000 reads were needed to detect five taxa, but because of PCR bias, almost 40,000 reads were required to detect all six taxa (Figure 2).

Time-1 experiment: Detection of target macroinvertebrates in mock samples

All six target macroinvertebrate taxa known to be present in samples (Table 1) were detected in at least one of the Sorted, Bulk and Matrix PCR replicates (Figure 3). However, detection differed between target taxa and across sample types and storage treatments.

Leptophlebiidae were detected in all Time-1 PCR replicates, as were Simuliidae except for one Matrix sample (Figure 3). Baetidae and Coloburiscidae were also detected in most Time-1 samples. Hydrobiosidae were detected in 33-51% of replicates and Philopotamidae were detected in 18% of Sorted and Bulk replicates but in only one Matrix replicate at Time-1.

No statistical differences were found among sample types for the detection of target taxa at Time-1 (e.g. Matrix, z value = -0.687, p = 0.492; Sorted, z value = -0.563, p = 0.574; Figure 4). There was high variability in the detection of target taxa within sample replicates, sub-samples and PCR replicates for both Time-1 and Time-2 (Figure 5) and no significant effect of sample types (Matrix *z value* = -1.767, *P* = 0.0772).

Note that the statistical analysis and box plots are based on the rarefied read count data to normalise the data for statistical comparisons, which reduced the actual number of target taxa shown to be detected by DNA metabarcoding. Using all the data, the average number of target taxa detected in all sample and treatment types are shown in Table 3.

Time-2 experiment: Effect of storage conditions on the detection of macroinvertebrates

The pattern of detection for target taxa was similar for Time-2 compared to Time-1 samples stored at room temperature (Figure 3), except Philopotamidae were detected in more Time-2 replicates compared to Time-1, as were Hydrobiosidae (apart from in Matrix samples). Coloburiscidae were detected in more Time-1 Matrix replicates stored at room temperature than Time-2. Importantly, all the target taxa were consistently detected in the Bulk and Sorted samples. Thus, suggesting that including the sample matrix had not interfered with the detection of our target macroinvertebrate taxa.

Statistical analysis of storage time effects on the detection of the six target taxa (Time-1 and Time-2) identified no differences between sample type (Matrix, z value = -0.687, p = 0.492; Sorted, z value =-5.63, p = 0.574) or time period (Time-2, z value = 0.084, p = 0.933) for samples stored at room temperature (Figure 4; Time-1 and Time-2 Room temperature data only). No interactions were noted between time and sample type (Matrix: Time-2, z value = -0.744, P = 0.457, Sorted: Time-2, z value = 0.521, p = 0.602). However, we note that the current level of replication (three samples, three sub-samples with one extraction and three PCRs from each extraction) was required to detect all six taxa (Figure 3) because of the variability in detection at all stages of the metabarcoding process (Figure 5).

For Time-2, the pattern of detection for target taxa was generally similar for samples across storage type and comparable to Time-1 samples stored at room temperature (Figure 3, 4), except for a few notable differences. Philopotamidae were detected in more Time-2 replicates for all sample types compared to samples stored at room temperature in Time-1 and Hydrobiosidae were not detected in the frozen Matrix samples.

Statistically, the detection of target taxa was little affected by storage type (Frozen, Heated and Room temperature) within Time-2 samples (Figure 4b, Figure 5d,e,f). Target taxa detection in frozen samples did not differ from the heated treated (z value = -0.329, p = 0.742) or room temperature samples (z value = -0.007, p = 0.994). When sample type (i.e. Matrix, Sorted, Bulk) was considered as a crossed fixed effect with storage type, Time-2 Matrix samples had lower but non-significant detection of target taxa than the Bulk samples (z=-1.804, p=0.0713.). A near significant negative effect of heated Matrix samples was also apparent (z=-1.717, p=0.0860.; Figure 4b, Figure 5d).

Overall macroinvertebrate family diversity detected with DNA metabarcoding compared to morphological identification.

DNA metabarcoding results of all replicates (Room temp, Heated and Frozen) of the Bulk and Matrix samples was compared to morphological identification using mostly family identification. It is important to note that these DNA metabarcoding detections, with the exception of the six added taxa in the Bulk mock samples, are based on residual sources of DNA in the samples.

All taxa found to be abundant in the morphologically identified subsamples were also found using DNA metabarcoding. Overall, 25 families were identified morphologically from the sub-sample and scan taken for routine bioassessment (see taxa list in Supplementary material, Table S1). Taxa detected by metabarcoding were assigned to the same taxonomic resolution as the standard morphological bioassessment data to enable comparison: this resulted in 17 different taxa detected by metabarcoding at Time-1 and 20 taxa at Time-2 (Figure 6). While the data indicate some consistent overlap between the DNA metabarcoding and morphological identifications, we did find some differences. Fewer taxa were expected in Matrix samples since the matrix had all visible macroinvertebrates removed, but the results provide some interesting insights.

Eight of the morphologically identified taxa were consistently not detected with DNA metabarcoding (Acarina, Hydrophilidae, Tipulidae, Athericidae, Ecnomidae, Tanypodinae, Caenidae and Calamoceratidae) (Figure 6). However, note that Diptera (other than those assigned to families) and Chironomidae (other than those assigned to sub-family) were detected in either Bulk or Matrix samples and these groups did contain numerous MOTUs that could not be assigned at a lower level of taxonomy. The morphologically identified taxa that failed to be detected by metabarcoding were small animals and/or present at low abundance in the collections as shown by enumeration of morphologically identified specimens i.e. Acarina (1 individual), Hydrophilidae (1 individual), Tipulidae, Arthericidae (1 individual), Ecnomidae (4 individuals), Tanypodinae (only seen in the visual scan), Caenidae (9 individuals) and Calamoceridae (1 individual) (Table S1). However, all of the more abundant taxa found morphologically were also found using DNA metabarcoding of the residual DNA contained within the sample matrix.

When comparing different sample types, we found that in the Time-1 Matrix samples, DNA metabarcoding identified 14 taxa compared with 17 at Time-2. Time-1 samples had two unique detections, Chironomidae (other than those assigned to sub-family) and Corydalidae, and two detections shared only with morphologically identified taxa list (Gomphidae and Leptoceridae). For Time-2, the Matrix samples had two unique detections, Diamesinae (a

Chironmidae sub-family) and Diptera (other), the Glossosomatidae detection was shared with Bulk samples and Psephenidae was detected as with the morphologically identified list of taxa (Figure 6). In the Bulk mock samples, DNA metabarcoding identified 13 taxa in Time-1 samples and 17 taxa in Time-2 samples. Elmidae and Glossosomatidae were additional taxa detected in Time-2 but not at Time-1 (Figure 6). Similar patterns of detection were observed when comparing samples stored at room temperature with those that were heated and frozen (see Supplementary Material Figure S2).

Discussion

In this study, we show that directly extracting DNA from macroinvertebrates within the sample matrix collected in sample nets, and using this for DNA metabarcoding, could effectively detect macroinvertebrates even if samples were stored under sub-optimal conditions. We found no evidence of inhibition or interference from the sample matrix with few differences in detection of the six target macroinvertebrate taxa between the different sample types. The detection of all six target macroinvertebrates suggest the feasibility of using DNA metabarcoding of bulk net contents for bioassessment, which would remove the need for laborious (and costly) sorting of samples.

PCR inhibition has reportedly been problematic for samples containing a high organic content (Andersen *et al.*, 2012), as represented by our Matrix and Bulk samples. However, we showed that sample type did not have a significant effect on the taxa recovered, thus suggesting that negative effects of PCR inhibitors may be minimal for bulk samples such as those collected using kick-nets. Furthermore, the amplification of non-target sequences can also reduce the recovery and detection of sequences from the taxa of interest (Collins *et al.*, 2019). Previous research has found that less than 50% of the OTUs and approximately 60% of the total sequences recovered from bulk kick-net samples could be assigned to metazoan taxa (Macher *et al.*, 2018). Our results indicate only a minimal decrease in the percentage of reads assigned to invertebrate taxa when the sample matrix is included in the analyses. The higher recovery of target taxa may indicate that the primer set used here (BF1/BR1) (Elbrecht & Leese, 2017) is more suitable for DNA metabarcoding analyses of bulk macroinvertebrate samples compared to the primers (BF2/BR2) used by Macher *et al.* (2018). Nonetheless, other variations in the metabarcoding workflow (i.e. one-step vs two-step library preparation protocol and the amplicon size selection protocol) could also contribute to the observed differences.

The DNA metabarcoding success was not dependent on sample storage with storage up to one year in 95% ethanol at room temperature, or after heating, not affecting the detection of the six target macroinvertebrate taxa. Stein *et al.* (2013) showed that initial fixation of macroinvertebrates in >95% ethanol preserves DNA for DNA barcoding for up to six months even if samples are then stored under sub-optimal conditions. However, our study suggests that storage of bulk samples with 95% ethanol for up to a year means samples will remain suitable for DNA metabarcoding even if subjected to heat in vehicles or storage at room temperature. However, freezer storage may present a better storage option for samples that require storage for longer than a year or for samples to be used for standard DNA barcoding of large amplicons (which require better preserved DNA) to build DNA barcode reference libraries (Lear *et al.*, 2018; Baird *et al.*, 2011).

We found little difference in the detection of the target macroinvertebrate taxa in the different sample types (Bulk, Matrix and Sorted) when considering different levels of replication i.e. sample, sub-sample or PCR. This result is consistent with Weigand and Macher (2018) who found taxon detection among PCR replicates mostly overlapped. However, we did find that detection of taxa increased slightly as levels of replication were combined. This is likely attributed to increased sampling effort and greater sequencing depth. We found at our PCR replication level, that at least 40,000 reads were needed to reliably identify all six macroinvertebrate target taxa in the sorted samples, as two of our target taxa, the Hydrobiosidae and Philopotamidae, appeared to amplify poorly compared to the other taxa. This may have been a result of PCR bias against these taxa or may reflect the reduced abundance of these taxa in the sample (i.e. only half an individual was present in each sample). However, only 10,000 reads were required to reliably detect five of the six target taxa.

Our study was primarily designed to determine a) if the presence of the sample matrix impaired the detection of specific macroinvertebrate taxa, b) the possible influence of sub-optimal sample storage, and c) appropriate levels of replication at the various stages of metabarcoding. To evaluate how much of a bulk sample is required to get a representative sample of overall macroinvertebrate richness would have required an experimental design examining different proportions of the entire bulk samples from multiple sites with differing macroinvertebrate community composition and densities. However, our results do provide some important insights that should be considered when using bulk samples directly for DNA metabarcoding in bioassessment. We found that the overall macroinvertebrate composition in

the mock Bulk and Matrix samples for each treatment (Time-1 room temp, Time-2 room temp, Time-2 heated and Time-2 frozen) substantially overlapped, with all of the more abundant macroinvertebrate families found morphologically also detected through metabarcoding. However, like the detection of the six macroinvertebrate target taxa, we found more taxa were detected when treatments replicates were combined.

Although sequencing depth across samples is always variable, the observed variation in sequencing depth per replicate was relatively high, despite laboratory protocols designed to normalize the sequencing depth as much as possible. Variations in DNA quality, DNA extraction efficiency and PCR amplification efficiency might have contributed to the high variation in sequencing depth. Variation in the proportion of mis-amplified or low-quality reads may also have contributed to variable sequencing depth post-filtering, but no clear patterns were found in the data to determine the precise cause. While a greater and more uniform sequencing depth could improve the detection of taxa, extra families were detected by DNA metabarcoding when replicates and treatments were combined. This is likely to be attributed to a greater proportion of the of the bulk collection being processed for metabarcoding resulting in more rare or infrequent taxa being available for detection. This appears to be the case in the Venn Diagrams that show less frequent taxa appearing only in the Matrix or Bulk mock samples randomly across treatments. As with any sampling effort, this suggest that a large proportion of the bulk collection should be processed if needing to detect the rarer taxa present in collections. Options to enable greater volumes of bulk samples could include homogenization of whole samples either wet or dry, and isolation of the organic fraction from the sediment (Aylagas *et al.*, 2016), or using a DNA extraction procedure as done for large amounts of soil material (Taberlet *et al.*, 2012) or even extracting the DNA from the ethanol commonly used to preserve the samples (Hajibabaei *et al.*, 2012; Zizka *et al.*, 2019). Another, option would be sieving the bulk sample to remove large debris, which could reduce overall sample volume then homogenizing.

DNA metabarcoding two riffle kick-net collections from the Goodrabigbee River revealed high overall macroinvertebrate diversity, but we were unable to assign all macroinvertebrates to species because of the lack of reference DNA barcodes for freshwater macroinvertebrates, suggesting more individual DNA barcoding is required to fill gaps in DNA barcode libraries (see Carew *et al.*, 2017; Weigand *et al.*, 2019). Thus, even with DNA metabarcoding we were only able to identify some macroinvertebrate taxa to courser taxonomic levels. Nonetheless, although we found more macroinvertebrate families with the

morphological identification, DNA metabarcoding found 76 unique taxa (44 identified to species) from the Bulk and Matrix samples. Admittedly a larger proportion of the collection was used to obtain all the Bulk and Matrix samples used for DNA metabarcoding, but conversely, the morphologically derived taxa list resulted from the addition of a visual scan of the entire collection to pick up extra taxa not observed in the sub-sample obtained for routine bioassessment. However, we demonstrated we could identify substantially more taxa at greater resolution using DNA metabarcoding than with morphological identification (which was mostly at family level).

Macroinvertebrate DNA metabarcoding identification, while showing considerable overlap with morphological identification, often differs in the families recovered (e.g. Carew *et al.*, 2018; Elbrecht & Leese, 2017; Hajibabaei *et al.*, 2012). Eight taxa identified morphologically were not detected with DNA metabarcoding. While it is possible that the DNA of these taxa was not in the remaining sample used for metabarcoding, other factors such as primer bias, where certain taxa are not amplified (or poorly amplified), inadequate read-depth, or relatively low DNA abundance, can all affect the ability of DNA metabarcoding to detect taxa in a sample (Elbrecht & Steinke, 2019; Elbrecht *et al.*, 2017). These factors likely affected the DNA metabarcoding detection of some taxa in our study. However, including multiple DNA barcode amplicons and sequencing depth of 100,000 reads per sample can improve the recovery of taxa (Carew, Coleman & Hoffman, 2018; Clarke *et al.*, 2014; Elbrecht & Steinke, 2019). Our study only used a single amplicon, but we found the detection of taxa plateaued at 40,000 reads showing that, at least for this study, a lower sequencing depth was sufficient, but sample volume was more important. However, despite the Sorted samples containing only the six target macroinvertebrate taxa, the variation in their detectability suggests potential primer bias. Therefore, multi-primer approaches (e.g. Carew, Coleman & Hoffman, 2018; Hajibabaei *et al.*, 2012) are likely to improve the overall detection of taxa in bulk samples.

Species level information accessible through DNA metabarcoding has enormous potential to improve bioassessment. Species identification can be integrated with data on species-level sensitivities to stressors (Baird & Hajibabaei, 2012; Bray *et al.*, 2019) to provide bioassessment with increased diagnostic capability. Bulk DNA extraction and the simultaneous processing of multiple biomonitoring samples offers great change in the immediacy, accuracy and quantity of information that can be obtained, without sacrificing current biomonitoring infrastructure investment. How to deal with this new information

within the bioassessment context presents a major challenge (Keck *et al.*, 2017). What is needed are the tools to enable the mainstream use of molecular techniques in freshwater bioassessment and to make better use of the species level data we collect (Pawlowski *et al.*, 2018). However, integrating these new approaches with existing bioassessment frameworks and applying those in large-scale monitoring requires fundamental research to improve and streamline processes.

Conclusions

Advances in high-throughput DNA sequencing promise fast and effective molecular bioassessment, which could aid routine and large-scale biomonitoring. Reliance on traditional freshwater bioassessment protocols that use sorting of target taxa from the sample matrix and then conducting morphological identification could be greatly improved if DNA metabarcoding could directly process macroinvertebrates within the sample matrix. Our study showed it is feasible to process samples for DNA metabarcoding without separating the target specimens from the sample matrix. However, to integrate the findings of this study into bioassessment will require further investigation using multiple samples with different macroinvertebrate composition and densities and building comprehensive DNA barcode libraries for freshwater macroinvertebrates.

Acknowledgements

This research was supported by a grant from the NSW Environmental Trust (project title: DNA-based Identification for Routine Aquatic Bio-assessment; http://www.environment.nsw.gov.au/grants/2015-research.htm#dna, accessed 1 May 2019). We acknowledge the contributions from Anthony Chariton, Donald Baird, Trefor Reynoldson, Dianne Gleeson, Ross Thompson and Gavin Rees who provided advice through a project advisory committee. We thank Andrew O'Reilly Nugent and Bernd Gruber for bioinformatic and statistical assistance. We are grateful to anonymous reviewers for their constructive comments. The authors declare no conflicts of interest.

Data accessibility

Data from the manuscript are available through the following citation:

Nichols *et al.* [date to be added]: Towards routine DNA metabarcoding of macroinvertebrates using bulk samples for freshwater bioassessment: effects of debris and storage conditions on the recovery of target taxa. https ://doi.org/[details to be added]

References

Andersen K., Bird K.L., Rasmussen M., Haile J., Breuning-Madsen H., Kjaer K.H., Orlando L., Gilbert M.T.P. & Willerslev E. (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology,* **21,** 1966-1979.

Aylagas E., Mendibil I., Borja Á. & Rodríguez-Ezpeleta N. (2016) Marine Sediment Sample Pre-processing for Macroinvertebrates Metabarcoding: Mechanical Enrichment and Homogenization. *Frontiers in Marine Science,* **3**.

Baird D.J. & Hajibabaei M. (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology,* **21,** 2039-2044.

Baird D.J., Pascoe T.J., Zhou X. & Hajibabaei M. (2011) Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs specimen age. *Journal of the North American Benthological Society,* **30,** 125-130.

Bolger A.M., Lohse M. & Usadel B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics,* **30,** 2114-2120.

Bolyen E. & Rideout J.R. & Dillon M.R. & Bokulich N.A. & Abnet C. & Al-Ghalith G.A. & Alexander H. & Alm E.J. & Arumugam M. & Asnicar F. & Bai Y. & Bisanz J.E. & Bittinger K. & Brejnrod A. & Brislawn C.J. & Brown C.T. & Callahan B.J. & Caraballo-Rodríguez A.M. & Chase J. & Cope E. & Da Silva R. & Dorrestein P.C. & Douglas G.M. & Durall D.M. & Duvallet C. & Edwardson C.F. & Ernst M. & Estaki M. & Fouquier J. & Gauglitz J.M. & Gibson D.L. & Gonzalez A. & Gorlick K. & Guo J. & Hillmann B. & Holmes S. & Holste H. & Huttenhower C. & Huttley G. & Janssen S. & Jarmusch A.K. & Jiang L. & Kaehler B. & Kang K.B. & Keefe C.R. & Keim P. & Kelley S.T. & Knights D. & Koester I. & Kosciolek T. & Kreps J. & Langille M.G.I. & Lee J. & Ley R. & Liu Y.-X. & Loftfield E. & Lozupone C. & Maher M. & Marotz C. & Martin B.D. & Mcdonald D. & Mciver L.J. & Melnik A.V. & Metcalf J.L. & Morgan S.C. & Morton J. & Naimey A.T. & Navas-Molina J.A. &

Nothias L.F. & Orchanian S.B. & Pearson T. & Peoples S.L. & Petras D. & Preuss M.L. & Pruesse E. & Rasmussen L.B. & Rivers A. & Robeson I.I.M.S. & Rosenthal P. & Segata N. & Shaffer M. & Shiffer A. & Sinha R. & Song S.J. & Spear J.R. & Swafford A.D. & Thompson L.R. & Torres P.J. & Trinh P. & Tripathi A. & Turnbaugh P.J. & Ul-Hasan S. & Van Der Hooft J.J.J. & Vargas F. & Vázquez-Baeza Y. & Vogtmann E. & Von Hippel M. & Walters W. & Wan Y. & Wang M., et al. (2018) QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints,* **6,** e27295v27292.

Boyer F., Mercier C., Bonin A., Le Bras Y., Taberlet P. & Coissac E. (2016) OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources,* **16,** 176-182.

Bray J.P., Reich J., Nichols S.J., King G.K.K., Mac Nally R., Thompson R., O'reilly-Nugent A. & Kefford B.J. (2019) Biological interactions mediate context and species-specific sensitivities to salinity. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **374**.

Buyck B. (1999) Taxonomists are an endangered species in Europe. *Nature,* **401,** 321-321.

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., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., Mcdonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Tumbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J. & Knight R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods,* **7,** 335-336.

Carew M.E., Coleman R.A. & Hoffman A.A. (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *Peerj,* **6**.

Carew M.E., Kellar C.R., Pettigrove V.J. & Hoffmann A.A. (2018) Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological Indicators,* **85,** 440-450.

Carew M.E., Nichols S.J., Batovska J., St Clair R., Murphy N.P., Blacket M.J. & Shackleton M.E. (2017) A DNA barcode database of Australia's freshwater macroinvertebrate fauna. *Marine and Freshwater Research,* **68,** 1788-1802.

Chen H. (2018) VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. https://CRAN.R-project.org/package=VennDiagram.

Clarke L.J., Soubrier J., Weyrich L.S. & Cooper A. (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources,* **14,** 1160-1170.

Collins R.A., Bakker J., Wangensteen O.S., Soto A.Z., Corrigan L., Sims D.W., Genner M.J. & Mariani S. (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution,* **Accepted Author Manuscript. doi:10.1111/2041-210X.13276**.

Curry C.J., Gibson J.F., Shokralla S., Hajibabaei M. & Baird D.J. (2018) Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshwater Science,* **37,** 178-189.

De Barba M., Miquel C., Boyer F., Mercier C., Rioux D., Coissac E. & Taberlet P. (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources,* **14,** 306-323.

Elbrecht V. & Leese F. (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science,* **5**.

Elbrecht V. & Steinke D. (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. *Freshwater Biology,* **64,** 380– 387.

Elbrecht V., Vamos E.E., Meissner K., Aroviita J. & Leese F. (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution,* **8,** 1265-1275.

Haase P., Pauls S.U., Schindehutte K. & Sundermann A. (2010) First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. *Journal of the North American Benthological Society,* **29,** 1279-1291.

Hajibabaei M., Spall J.L., Shokralla S. & Van Konynenburg S. (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *Bmc Ecology,* **12**.

Hebert P.D.N., Cywinska A., Ball S.L. & Dewaard J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences,* **270,** 313-321.

Jackson J.K., Battle J.M., White B.P., Pilgrim E.M., Stein E.D., Miller P.E. & Sweeney B.W. (2014) Cryptic biodiversity in streams: a comparison of macroinvertebrate

communities based on morphological and DNA barcode identifications. *Freshwater Science,* **33,** 000.

Jones F.C. (2008) Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. *Environmental Reviews,* **16,** 45-69.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P. & Drummond A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England),* **28,** 1647-1649.

Keck F., Vasselon V., Tapolczai K., Rimet F. & Bouchez A. (2017) Freshwater biomonitoring in the Information Age. *Frontiers in Ecology and the Environment,* **15,** 266-274.

Lear G., Dickie I., Banks J., Boyer S., Buckley H.L., Buckley T.R., Cruickshank R., Dopheide A., Handley K.M., Hermans S., Kamke J., Lee C.K., Macdiarmid R., Morales S.E., Orlovich D.A., Smissen R., Wood J. & Holdaway R. (2018) Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *New Zealand Journal of Ecology,* **42,** 10-+.

Macher J.N., Vivancos A., Piggott J.J., Centeno F.C., Matthaei C.D. & Leese F. (2018) Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources,* **18,** 1456-1468.

Majaneva M., Diserud O.H., Eagle S.H.C., Hajibabaei M. & Ekrem T. (2018) Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding and Metagenomics,* **2,** e26664.

Marchant R. (1989) A sub-sampler for samples of benthic invertebrates. *Bulletin of the Australian Society of Limnology,* **12,** 49–52.

Nichols S.J., Barmuta L.A., Chessman B.C., Davies P.E., Dyer F.J., Harrison E.T., Hawkins C.P., Jones I., Kefford B.J., Linke S., Marchant R., Metzeling L., Moon K., Ogden R., Peat M., Reynoldson T.B. & Thompson R.M. (2017) The imperative need for nationally coordinated bioassessment of rivers and streams. *Marine and Freshwater Research,* **68,** 599-613.

Nichols S.J. & Norris R.H. (2006) River condition assessment may depend on the sub-sampling method: field live-sort versus laboratory sub-sampling of invertebrates for bioassessment. *Hydrobiologia,* **572,** 195-213.

Nichols S.J., Sloane P., Coysh J., Williams C. & Norris R. (2000) *Australian Capital Territory, Australian River Assessment System (AUSRIVAS), Sampling and Processing Manual,* Cooperative Research Centre for Freshwater Ecology University of Canberra, Canberra [available online at https://ausrivas.ewater.org.au/index.php/manuals-a-datasheets].

O'donnell J.L., Kelly R.P., Lowell N.C. & Port J.A. (2016) Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. *Plos One,* **11**.

Pawlowski J., Kelly-Quinn M., Altermatt F., Apotheloz-Perret-Gentil L., Beja P., Boggero A., Borja A., Bouchez A., Cordier T., Domaizon I., Feio M.J., Filipe A.F., Fornaroli R., Graf W., Herder J., Van Der Hoorn B., Jones J.I., Sagova-Mareckova M., Moritz C., Barquin J., Piggott J.J., Pinna M., Rimet F., Rinkevich B., Sousa-Santos C., Specchia V., Trobajo R., Vasselon V., Vitecek S., Zimmerman J., Weigand A., Leese F. & Kahlert M. (2018) The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment,* **637,** 1295-1310.

Pilgrim E.M., Jackson S.A., Swenson S., Turcsanyi I., Friedman E., Weigt L. & Bagley M.J. (2011) Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society,* **30,** 217-231.

Porter T.M. & Hajibabaei M. (2018) Automated high throughput animal CO1 metabarcode classification. *Scientific Reports,* **8**.

R Development Core Team. (2010) R : a language and environment for statistical computing. R Foundation for Statistical Computing Vienna (Austria): ISBN 3-900051-07-0.

Robbins-Pianka. (2013) https://gist.github.com/adamrp/7591573.

Stein E.D., White B.P., Mazor R.D., Jackson J.K., Battle J.M., Miller P.E., Pilgrim E.M. & Sweeney B.W. (2014) Does DNA barcoding improve performance of traditional stream bioassessment metrics? *Freshwater Science,* **33,** 302–311.

Stein E.D., White B.P., Mazor R.D., Miller P.E. & Pilgrim E.M. (2013) Evaluating ethanol-based sample preservation to facilitate Use of DNA barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. *Plos One,* **8**.

Sweeney B.W., Battle J.M., Jackson J.K. & Dapkey T. (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? *Journal of the North American Benthological Society,* **30,** 195-216.

Taberlet P., Prud'homme S.M., Campione E., Roy J., Miquel C., Shehzad W., Gielly L., Rioux D., Choler P., Clement J.C., Melodelima C., Pompanon F. & Coissac E. (2012) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology,* **21,** 1816-1820.

Weigand A.M. & Macher J.-N. (2018) A DNA metabarcoding protocol for hyporheic freshwater meiofauna: Evaluating highly degenerate COI primers and replication strategy. *Metabarcoding and Metagenomics,* **2,** e26869.

Weigand H., Beermann A.J., Čiampor F., Costa F.O., Csabai Z., Duarte S., Geiger M.F., Grabowski M., Rimet F., Rulik B., Strand M., Szucsich N., Weigand A.M., Willassen E., Wyler S.A., Bouchez A., Borja A., Čiamporová-Zaťovičová Z., Ferreira S., Dijkstra K.-D.B., Eisendle U., Freyhof J., Gadawski P., Graf W., Haegerbaeumer A., Van Der Hoorn B.B., Japoshvili B., Keresztes L., Keskin E., Leese F., Macher J.N., Mamos T., Paz G., Pešić V., Pfannkuchen D.M., Pfannkuchen M.A., Price B.W., Rinkevich B., Teixeira M.a.L., Várbíró G. & Ekrem T. (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment,* **678,** 499-524.

Yu D.W., Ji Y.Q., Emerson B.C., Wang X.Y., Ye C.X., Yang C.Y. & Ding Z.L. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution,* **3,** 613-623.

Zizka V.M.A., Leese F., Peinert B. & Geiger M.F. (2019) DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome,* **62,** 122-136.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table 1. The six target macroinvertebrates taxa from the Goodradigbee River, NSW used to construct each Bulk and Sorted samples.

Number/amount of individual added per sample	Order	Family
3	Diptera	Simuliidae
3	Ephemeroptera	Leptophlebiidae
3	Ephemeroptera	Baetidae
0.5	Ephemeroptera	Coloburiscidae
0.5	Trichoptera	Hydrobiosidae
0.5	Trichoptera	Philopotamidae

Table 2. Proportion of invertebrate reads assigned to target taxa in Time-I and Time-2 for all sample types for all treatments combined.

	Sample type		
Proportion of invertebrate reads assigned to target taxa	Sorted	Matrix	Bulk
Time-1	0.999	0.956	0.974
Time-2	0.964	0.941	0.977

Table 3. The average number of target taxa detected in all sample and treatment types using the non-normalised data.

	Average no of target taxa detected			
Replicates	Overall	Bulk	Matrix	Sorted
All Time-1 and 2 combined	5.92	6.00	5.75	6.00
All samples combined	5.22	5.67	5.09	5.42
All sub-sample combined	5.06	5.14	4.58	4.97

Figure captions

Figure 1. Study design for metabarcoding experiment. Sample types included macroinvertebrates only samples termed 'Sorted' samples (i.e. six taxa stored together); sample matrix only, which included leaves, detritus and sediment with visible whole animals removed termed 'Matrix' samples; and mock bulk samples, termed 'Bulk' samples, which were six macroinvertebrate taxa combined with sample matrix. All samples were stored in 95% ethanol.

Figure 2. The relationship between non-normalised sequencing depth and the number of macroinvertebrate families detected per PCR replicate for the Bulk, Matrix and Sorted samples from the Goodradigbee River, New South Wales (NSW), Australia. Bulk and Matrix samples contained the sample matrix while the Sorted samples contained only the six target macroinvertebrate families.

Figure 3. The number of samples in which each target macroinvertebrate taxon were detected in Time-1 (stored at room temperature) and Time 2 samples for all treatments (stored at room temperature for 12 months, heat treated and stored at room temperature for 12 months and frozen stored at -20 ^{o}C for 12 months).

Figure 4. Comparative detection of target macroinvertebrate taxa between sample types stored at room temperature from Time-1 and all treatments from Time-2 based on the rarefied data. Boxes show medians and 25th and 75th quartiles, the whiskers are 95th quantiles, jittered scatter plots present individual data points.

Figure 5. Variation in target taxa within Time-1 sample types (Bulk, Matrix, Sorted) a) samples, b) sample sub-replicates, and c) extraction (PCR) sub-replicates. Time-2 data variability (d, e, f) between sample types further examined between storage types (Frozen, Heated and Room Temperature samples) based on the rarefied data. Boxes show medians and 25th and 75th quartiles, the whiskers are 95th quantiles, jittered scatter plots present individual

data points. Note that one set of Time-1 Bulk extraction sub-replicates all had less than 1,000 sequence reads per replicate and were thus filtered from the dataset.

Figure 6. Comparison of macroinvertebrate taxa morphologically identified (grey circle) in the bioassessment sub-sample and taxa (at the same taxonomic resolution) detected with DNA metabarcoding of the Bulk (red circle) and Matrix (green circle) samples at Time-1 (all room temp replicates) and Time-2 (all room temp, heated and frozen replicates combined). The six families known to be present in the Bulk samples are shown with *.

Table and Figure captions for supporting information

Table S1. Taxa collected for standard bioassessment from Goodradigbee River, NSW

Table S2. Fusion-tagged primer combinations for each PCR replicate. Forward and reverse primers are the universal invertebrate primers BF1 and BR1.

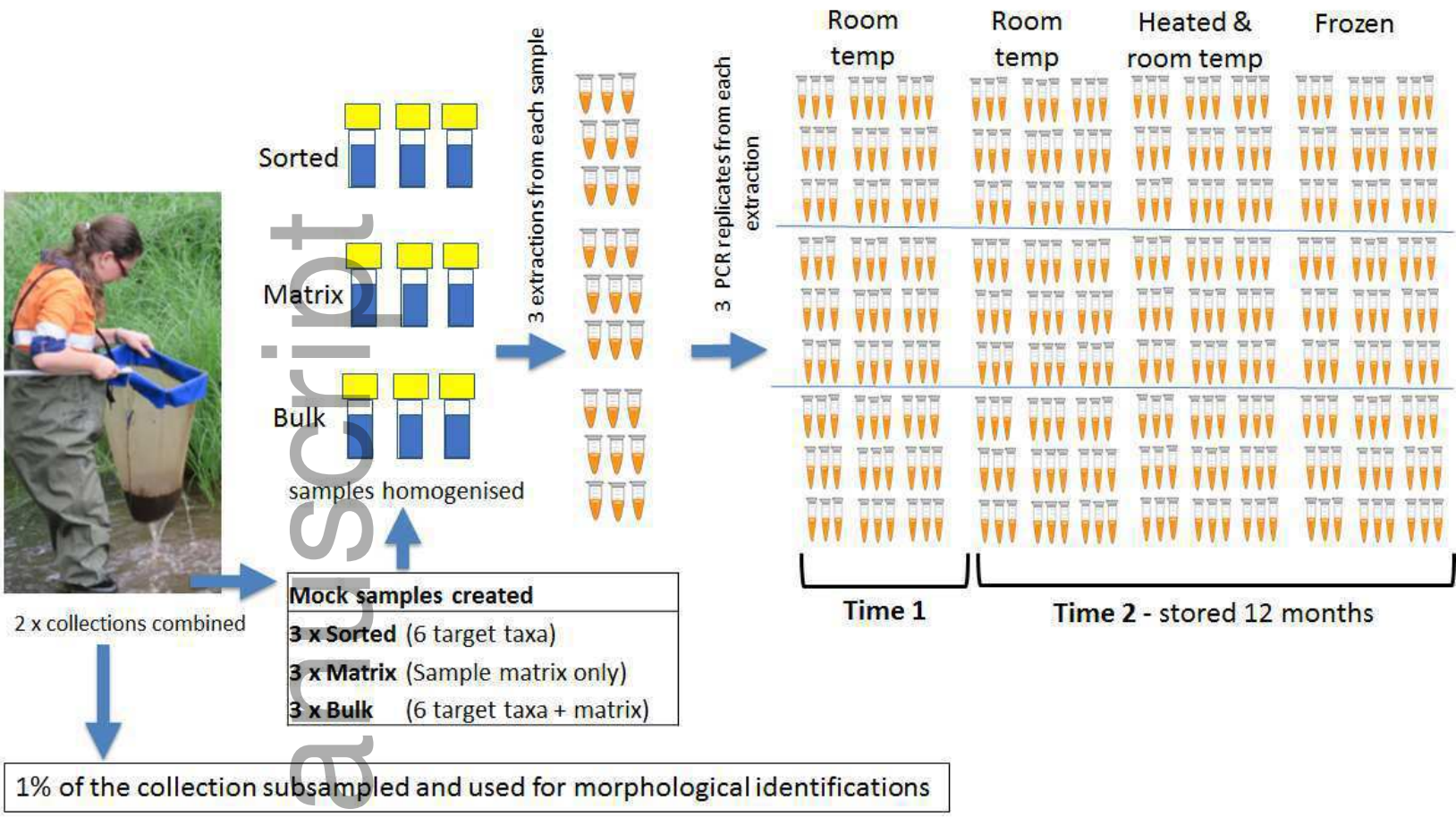
Table S3. Dilution and Ct value for each PRC replicate for Time-1 and Time-2 experiments.

Table S4. Number of sequence reads per PCR replicate.

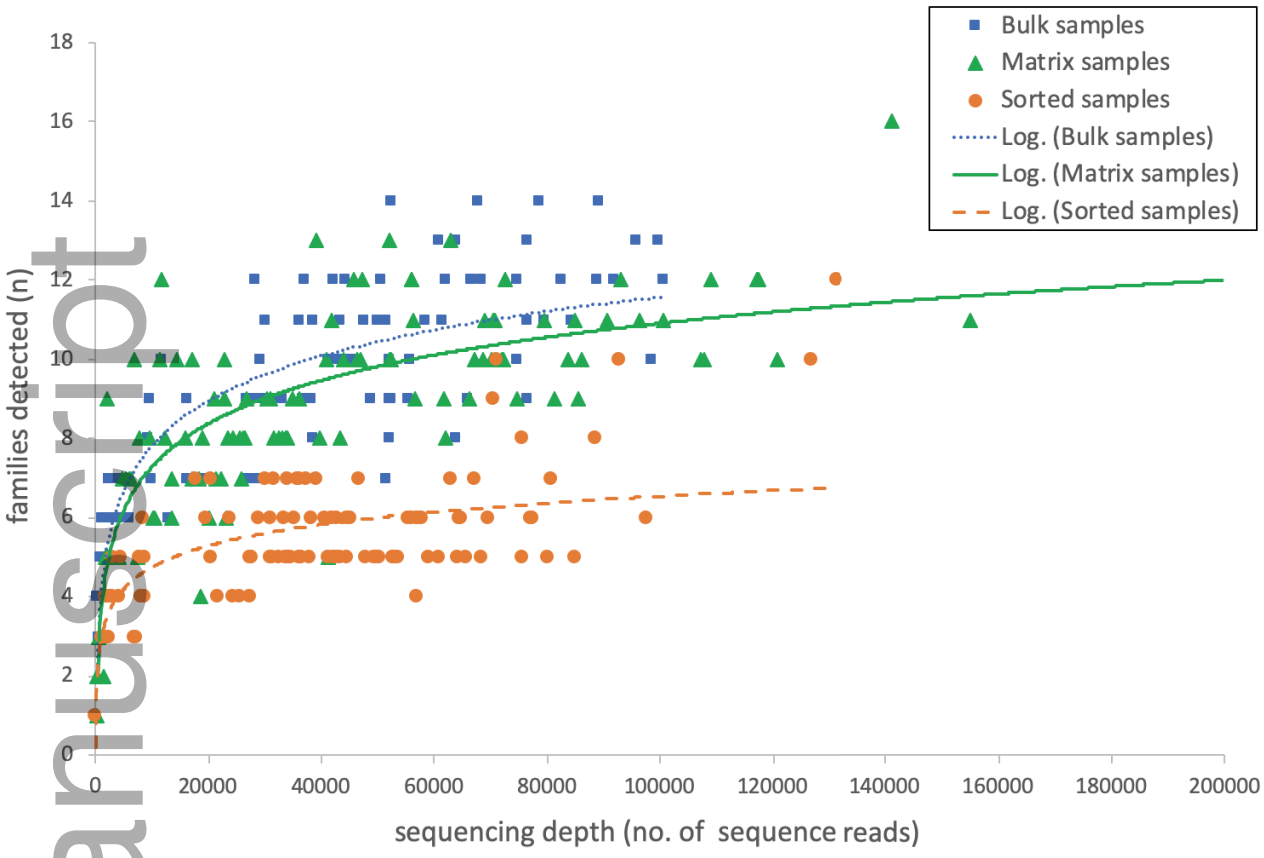
Table S5. Taxa detected by metabarcoding in all replicates combined for Time-1 and Time-2 samples.

Figure S1. Rarefaction curves. The sequencing reads for each PCR replicate were subject to rarefaction analysis for the a) bulk, b) matrix and c) sorted samples. Read datasets were subsampled at intervals of 1,000 reads, at sequencing depths ranging from 1,000 – 30,000 reads. Ten replicates were performed at each subsampling depth, the median value of which is plotted. For PCR replicates with less than 30,000 reads, rarefaction was only performed at sequencing depths less than the total read count.

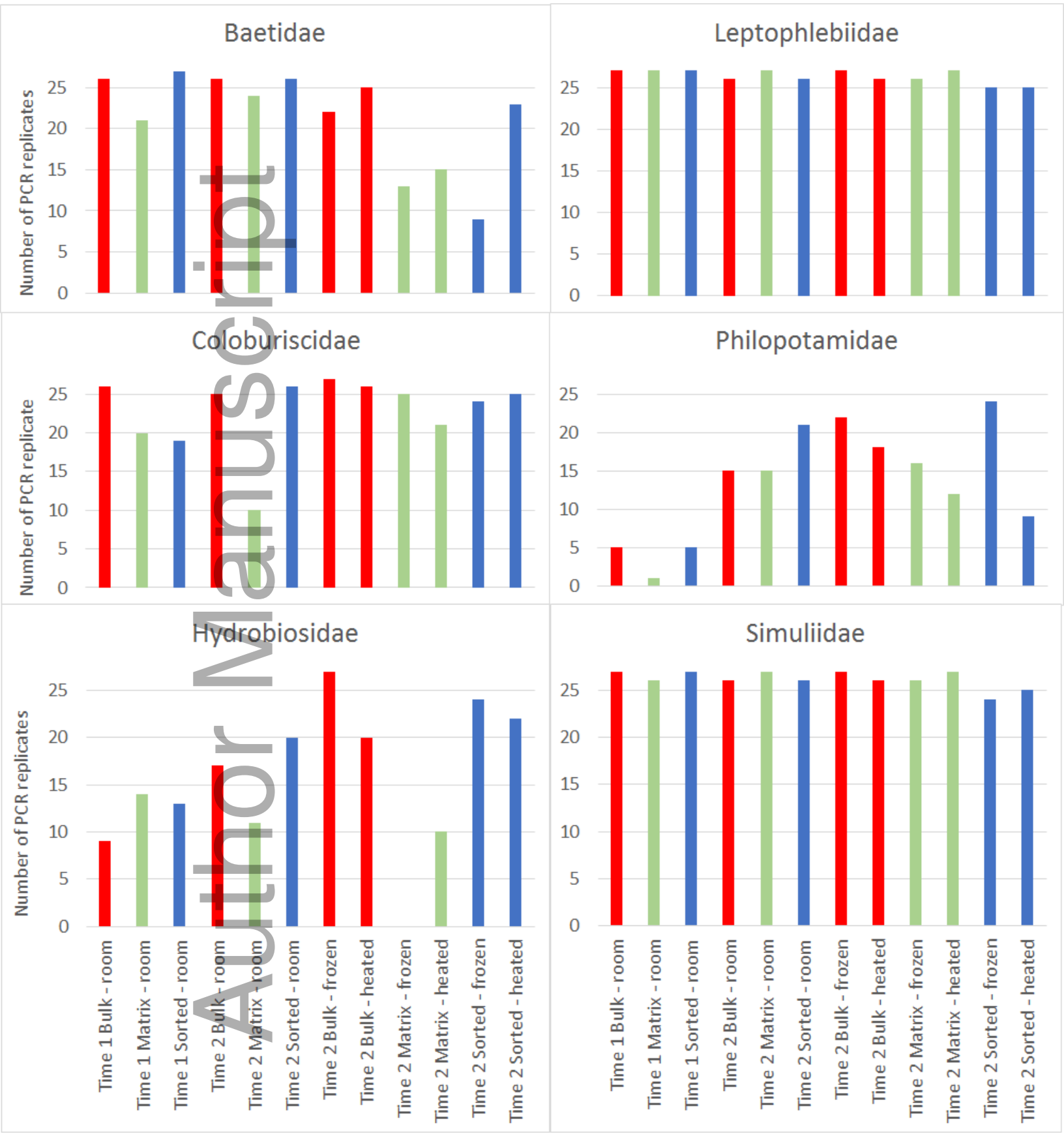
Figure S2. Comparison of macroinvertebrate taxa detected in room temp (A), frozen (B) and heated (C) samples at Time-2 Room temperature Bulk, Matrix and Morphologically identified samples. The six families known to be present in the bulk samples are shown with *.



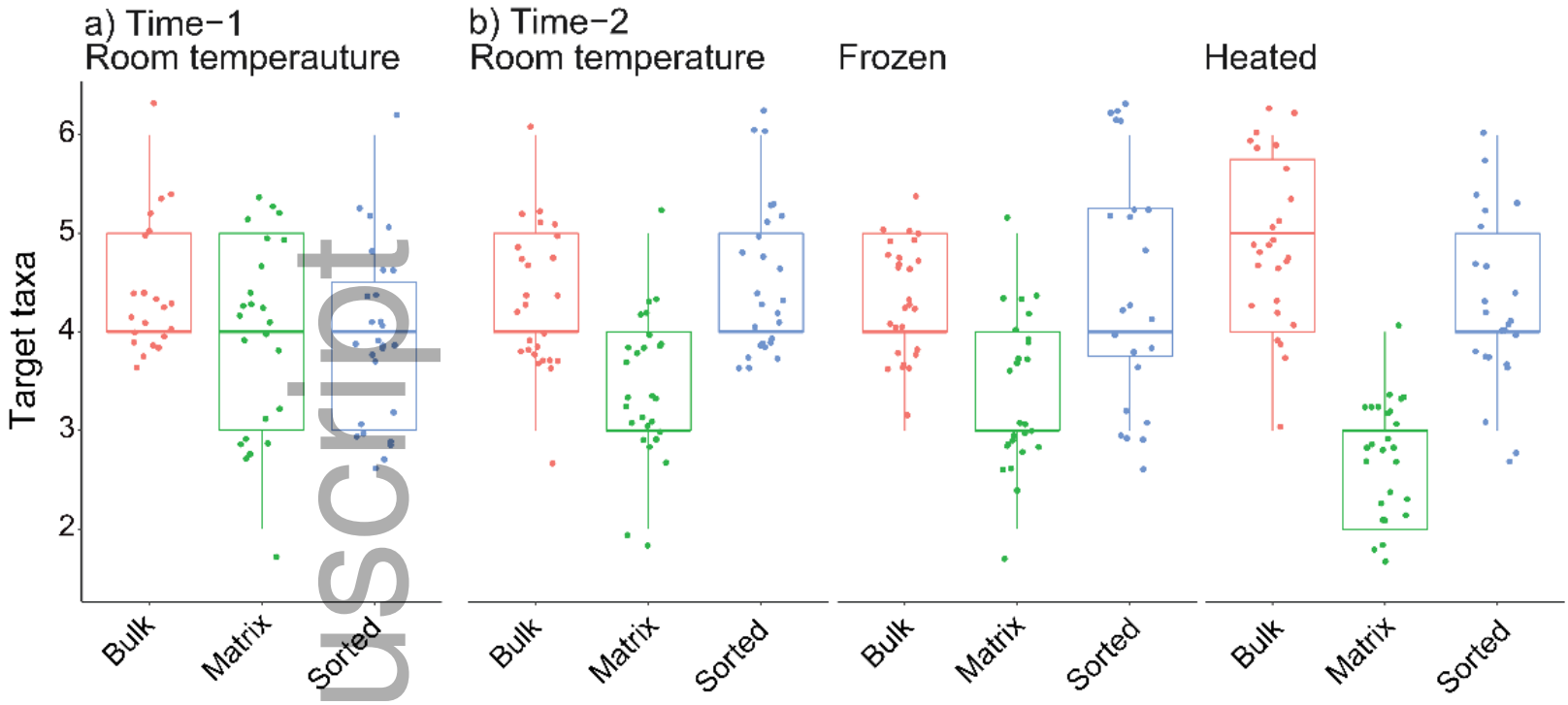
fwb_13443_f1.eps



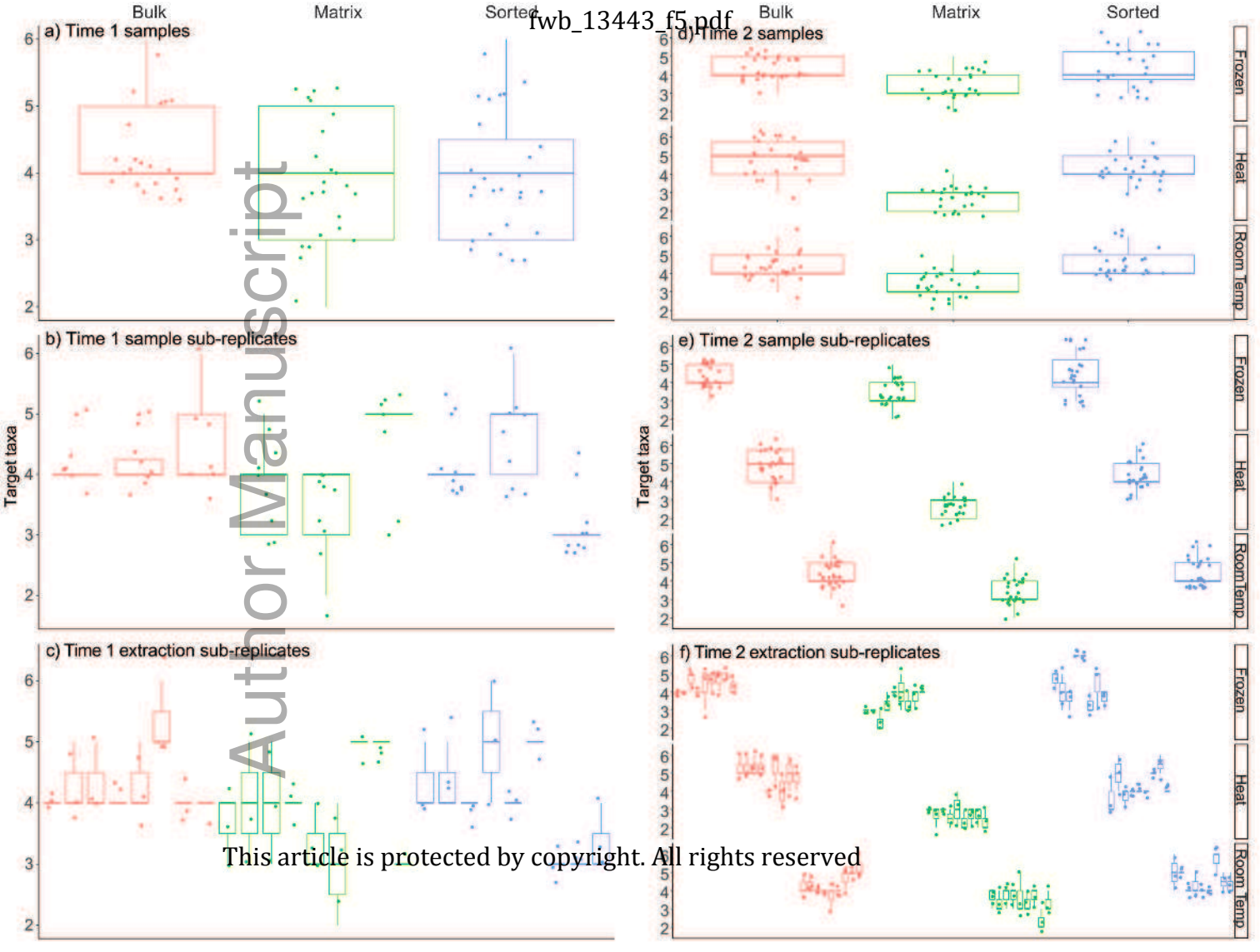
fwb_13443_f2.eps



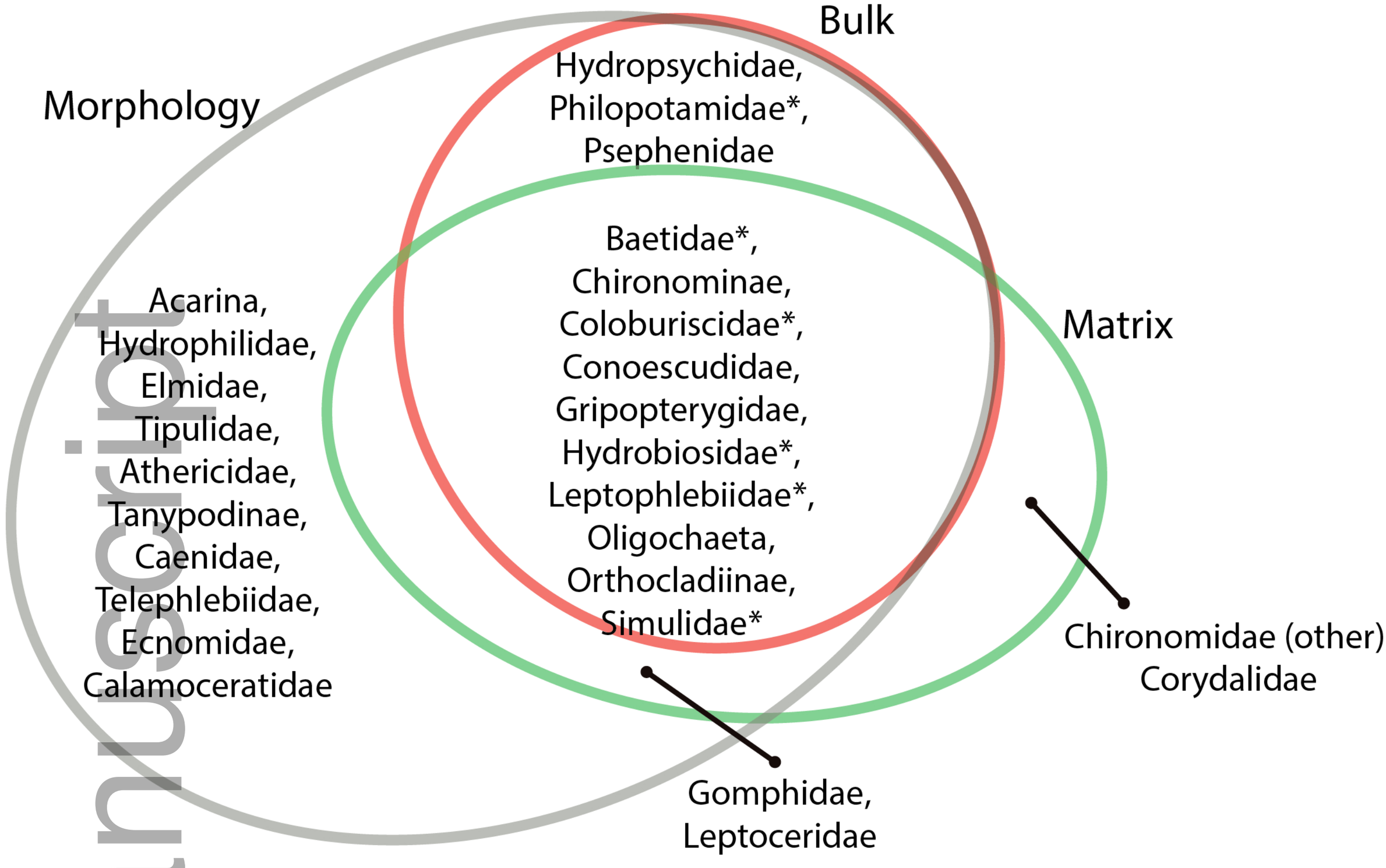
fwb_13443_f3.eps



fwb_13443_f4.eps



a) Time-1: Room temperature



b) Time-2: Room temperature, Heated and Frozen