Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding?

Melissa E Carew Corresp., [1] , **Rhys A Coleman** [2] , **Ary A Hoffmann** [1]

[1] School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia

[2] Applied Research, Melbourne Water, Melbourne, Victoria, Australia

Corresponding Author: Melissa E Carew
Email address: mecarew@unimelb.edu.au

Background. High throughput DNA sequencing of bulk invertebrate samples or metabarcoding is becoming increasingly used to provide profiles of biological communities for environmental monitoring. As metabarcoding becomes more widely applied, new reference DNA barcodes linked to individual specimens identified by taxonomists are needed. This can be achieved through using DNA extraction methods that are not only suitable for metabarcoding but also for building reference DNA barcode libraries.

Methods. In this study, we test the suitability of a rapid non-destructive DNA extraction method for metabarcoding of freshwater invertebrate samples.

Results. This method resulted in detection of taxa from many taxonomic groups, comparable to results obtained with two other tissue-based extraction methods. Most taxa could also be successfully used for subsequent individual-based DNA barcoding and taxonomic identification. The method was successfully applied to field-collected invertebrate samples stored for taxonomic studies in 70% ethanol at room temperature, a commonly used storage method for freshwater samples.

Discussion. With further refinement and testing, non-destructive extraction has the potential to rapidly characterise species biodiversity in invertebrate samples, while preserving specimens for taxonomic investigation.

Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding?

Melissa E. Carew*, Rhys A. Coleman† and Ary A. Hoffmann*

*School of BioSciences, Bio21 Institute, 30 Flemington Rd, The University of Melbourne 3010, Victoria, Australia.

† Applied Research, Melbourne Water, PO Box 4342, Melbourne 3001 Victoria, Australia.

Author for correspondence: Melissa Carew

Address: Bio21 institute, 30 Flemington Rd, The University of Melbourne, Victoria, Australia, 3010.

Phone: +61 3 9035 3530

E-mail: mecarew@unimelb.edu.au

Abstract

Background. High throughput DNA sequencing of bulk invertebrate samples or metabarcoding is becoming increasingly used to provide profiles of biological communities for environmental monitoring. As metabarcoding becomes more widely applied, new reference DNA barcodes linked to individual specimens identified by taxonomists are needed. This can be achieved through using DNA extraction methods that are not only suitable for metabarcoding but also for building reference DNA barcode libraries.

Methods. In this study, we test the suitability of a rapid non-destructive DNA extraction method for metabarcoding of freshwater invertebrate samples.

Results. This method resulted in detection of taxa from many taxonomic groups, comparable to results obtained with two other tissue-based extraction methods. Most taxa could also be successfully used for subsequent individual-based DNA barcoding and taxonomic identification. The method was successfully applied to field-collected invertebrate samples stored for taxonomic studies in 70% ethanol at room temperature, a commonly used storage method for freshwater samples.

Discussion. With further refinement and testing, non-destructive extraction has the potential to rapidly characterize species biodiversity in invertebrate samples, while preserving specimens for taxonomic investigation.

Introduction

Species identification using high throughput DNA sequencing (HTS) of bulk samples containing multiple species (metabarcoding) is being increasingly applied in environmental monitoring, as it enables rapid identification of a wide range of taxa (Yu et al. 2012; Zhou et al. 2013; Gibson et al. 2014; Gibson et al. 2015; Elbrecht et al. 2017). However, associating sequences generated by metabarcoding with species names is dependent on comprehensive DNA barcoding libraries where individual specimens identified by taxonomists are DNA barcoded (Hajibabaei et al. 2011; Baird & Hajibabaei 2012; Carew et al. 2013; Zimmermann et al. 2014). Some taxonomic groups and geographic regions have comprehensive DNA barcode libraries (e.g. Hebert et al. 2013; Hendrich et al. 2015; Hebert et al. 2016), but many taxonomic groups or taxa from other geographical regions lack such coverage (see Carew et al. 2017; Kranzfelder, Ekrem & Stur 2017). This means that when environmental samples containing invertebrates are processed using metabarcoding, many sequences cannot be identified to species. This does not prevent the use of DNA barcodes for environmental monitoring, as sequences can be grouped into molecular operational taxonomic units (MOTUs) (Blaxter et al. 2005), which can be associated with various environmental parameters (i.e. Pawlowski et al. 2014; Lanzén et al. 2016). However, it does mean that pre-existing taxonomic and environmental information based on other approaches cannot be easily integrated (Carew et al. 2007; Schafer et al. 2011). Furthermore, an MOTU approach prevents the accuracy of DNA barcodes for species identifications to be tested (Lee 2004; Moritz & Cicero 2004), and it complicates the detection of PCR artifacts and chimeras when analyzing metabarcoding data (Creer et al. 2010; Carlsen et al. 2012).

To ensure that reference DNA barcode libraries are developed for invertebrate samples that are processed with HTS, field samples can be taken in duplicate, where one is used for

creating local reference DNA barcodes based on individual specimens and the other is homogenized and used for metabarcoding (e.g. Gibson et al. 2015). This approach can lead to a high species detection rate with HTS (Gibson et al. 2014; Gibson et al. 2015). However, it is possible that some taxa, particularly rare species, will not be found in both samples. If these species are only detected in the metabarcoded sample, there is not an opportunity to produce a reference DNA barcode when specimens are destroyed during DNA extraction.

Where only single samples are available, dissecting tissue or legs from individuals used for HTS keeps specimens largely intact for individual reference DNA barcoding. Detailed photographs of individual specimens can further assist identification, especially where damage occurs during dissection (e.g. Sweeney et al. 2011). However, this is a slow and difficult process, particularly for small taxa such as water mites, or for taxa with tissues inside shells, such as bivalves and gastropods. The preservation ethanol can also be used for non-destructive metabarcoding, leaving specimens undamaged (e.g. Hajibabaei et al. 2012). However, this approach can be less successful than tissue-based extraction for metabarcoding, depending on ethanol quality or re-use, storage time, presence of PCR inhibitors, as well as species biomass and composition (Hajibabaei et al. 2012; Carew et al. 2013).

A non-destructive DNA extraction method that enables reliable detection of taxa in bulk samples would help to overcome some of these challenges. Non-destructive methods involving temporarily immersing whole specimens in an extraction buffer have been successfully applied to individual specimens for reference DNA barcoding (Rowley et al. 2007; Castalanelli et al. 2010; Porco et al. 2010; Krosch & Cranston 2012;Wong et al. 2014; Cornils 2015). However, they have not been tested with bulk invertebrate samples, like those used for metabarcoding. It is not clear whether DNA from the same individuals extracted non-destructively for metabarcoding

can then be used for creating reference DNA barcodes. Also, it is not known whether combining species from different taxonomic groups with varying body forms in a non-destructive DNA extraction can be successfully metabarcoded to determine species biodiversity.

In this study, we aim to test whether it is possible to apply non-destructive DNA extractions to bulk invertebrate sample intended for metabarcoding. We first examine if individual invertebrates from common freshwater orders can be successfully extracted for metabarcoding and then reference DNA barcoding. We subjected taxa to a non-destructive total genomic extraction using a commercial kit, and followed with a Chelex extraction to test whether the same specimens can provide adequate DNA after multiple extractions. We then inspected the morphological integrity of specimens for taxonomic work after non-destructive extraction. Next, we constructed multiple bulk invertebrate samples for metabarcoding to compare species detection rates of samples processed using non-destructive DNA extractions to two different tissue- based DNA extraction methods. Finally, we tested the non-destructive DNA extraction method on samples collected for routine bioassessment, but stored in different ethanol concentration and temperatures, to determine if non-destructive extraction could be applied to routinely stored invertebrate collections.

Material & Methods

Component 1: Non-destructive DNA extractions on individuals.

The first component of this study aimed to determine whether a non-destructive DNA extraction can be used to obtain DNA from whole individuals from common freshwater macroinvertebrate orders/subclasses (Fig 1A). To test this, we included a small experiment (Experiment 1a) that examined the incubation time required for specimens to be immersed in extraction buffer to

obtain DNA without destruction of morphological characters, followed by a second broader experiment (Experiment 1b) to see how the method performed on individuals from multiple families covering 15 common orders/ classes/ subclasses.

A total genomic extraction (Extraction 1; Fig 1A), which is commonly used to prepare samples for HTS (e.g. Elbrecht, Peinert & Leese 2017), was used to obtain DNA from whole and intact invertebrate specimens. For extraction, ethanol was removed using a pipette from microcentrifuge tubes containing individual specimens, then 180 µL T1 buffer and 25 µL of proteinase K (extraction buffer) from the Nucleospin DNA extraction kit (Macherey-Nagel Inc. Bethlehem, PA, USA) were added. We tested the whether the presence of ethanol interfered with the DNA extraction by soaking several specimens overnight in TE buffer prior to extraction, but found no difference in successful amplification between soaked specimens and those with ethanol removed just prior to extraction (Table S1). Specimens were then incubated in the extraction buffer at 56°C. Based on other studies (Rowley et al. 2007; Castalanelli et al. 2010; Porco et al. 2010), we incubated specimens for 1 hr in the extraction buffer, but also tested a subset of taxa for 30 mins and/or 3 hrs depending on their levels of sclerotization (Table S1 and S2). The extraction buffer was removed and placed into a new 1.5 ml microcentrifuge tube for DNA extraction using a Nucleospin DNA extraction kit (Macherey-Nagel Inc. Bethlehem, PA, USA) following the manufacturer's instructions. After extraction buffer removal, absolute ethanol was added to microcentrifuge tubes containing whole specimens to preserve them for subsequent Chelex extraction and taxonomic identification. The condition of specimens was inspected under a dissecting microscope (Leica Microsystem and Instruments, Wetzlar, Germany).

A Chelex extraction (Extraction 2; Fig 1A) was then performed to establish if two extractions of the same specimen was possible, i.e. an initial extraction intended for metabarcoding followed by an extraction intended for individual DNA barcoding. The material used for Chelex extraction varied among taxonomic groups, depending on how specimens needed to be prepared for identification (Table S2) but followed the methods of Carew et al. (2003)

The success of both DNA extractions was determined by amplification of the DNA barcode region (Hebert et al. 2003) using the primer set HCOI2198/LCOI1490 from (Folmer et al. 1994) according to Carew et al. (2015), as this primer set was known to amplify the majority of taxa considered in the study. Some DNA barcode amplicons isolated with DNA extraction methods were sequenced to ensure that the intended species were extracted. Sequencing reactions were performed in both directions using an ABI 3730XL capillary sequencer (Applied Biosystems, Foster City, California, USA) with sequencing reactions and runs performed by Macrogen (Seoul, Korea)

After DNA extraction, taxa were identified to the lowest possible level using relevant taxonomic keys (see Hawking 2000). In most instances this was species, but for some groups we had difficultly identifying species due to a lack of keys, expertise or because specimens were immature or damaged (e.g. most Ephemeroptera were missing legs/cerci prior to being processed).

All macroinvertebrate specimens used for these experiments were collected using a 250 μm net from riffle and pool habitats from several streams/rivers around the greater Melbourne area (Victoria, Australia) and stored in 100% ethanol at -20°C until required.

Component 2: Bulk DNA extraction for metabarcoding

In the second component of this study, we examined bulk DNA extraction using the non-destructive (ND) extraction protocol above (as used on individuals) and compared it to two commonly used tissue-based extraction methods based on total homogenization (H) and dissection of tissues (D) (Fig 1B). To allow a comparison between the three extraction methods (ND, H and D), we constructed three sets of invertebrate samples each containing three 'replicates'. We choose to use high numbers of individuals and a species composition representative of what we have previously found in field samples (e.g. Carew et al. 2018) to construct invertebrate samples. The first set contained 80 individuals (Sample A), the second set 100 individuals (Sample B) and third set contained 120 individuals (Sample C) (Fig 1B). One 'replicate' from each set was used to compare metabarcoding success of the ND, H and D based extraction methods. Technical replicates (i.e. metabarcoding the same DNA extraction twice) were performed on the third set containing 120 individuals to examine the robustness of species detection. Unfortunately, it was not possible to make identical replicates given that we could not mount or destroy samples to check species identification. Instead, the composition of each replicate was standardized as much as possible by selecting taxonomic groups that could be identified to a low taxonomic level under a dissecting microscope, and/or specimens that had been collected from the same sites. We also used similar sized individuals for each taxon to control for differences in biomass. Not all taxonomic groups considered individually in component 1 could be used in constructing samples due to limited availability of material, but we tried to represent animals of varying levels of sclerotization and size.

When constructing samples, we focused on testing how a non-destructive method could be applied to small taxa only ($<7mm^2$) in bulk DNA extractions; all large taxa ($>7mm^2$) were

dissected when comparing extraction methods to reduce the likelihood that larger species affected the detection of overall species diversity (Elbrecht, Peinert & Leese 2017). Furthermore, larger taxa tend to be less common and/or more easily dissected than smaller taxa in bulk samples, making this a practical approach.

When comparing the three extraction methods, smaller taxa were processed differently. For completely homogenized samples (H), dissected tissue of large taxa and complete specimens of smaller taxa were combined for DNA extraction. For dissected samples (D), all taxa (both large and small) were dissected and then legs/tissues were combined and homogenized for DNA extraction. Dissected taxa were returned to absolute ethanol. For the non-destructive (ND) approach, absolute ethanol was removed from tubes containing small taxa using a pipette and 180 µL of buffer T1 with 25 µL proteinase K from the Nucleospin DNA extraction kit (Macherey-Nagel Inc.) was added, ensuring specimens were completely immersed. Tubes were incubated for 1 hr at 56°C. After incubation, specimens were placed back into absolute ethanol. The tissues/legs dissected from the larger taxa were homogenized separately. All homogenization steps were performed in 1.5 ml microcentrifuge tubes. Tissues were snap frozen in liquid nitrogen and crushed with a sterile pestle. Then 180 µL of the T1 buffer and 25 µL proteinase K were added, and samples incubated for 3 hrs at 56°C. After incubation all samples were DNA extracted using a Nucleospin tissue kit (Macherey-Nagel Inc.) following the manufacturer's instructions. DNA was eluted in 100 µL of the elution buffer from the Nucleospin tissue kit. The time taken to prepare material for each DNA extraction method and the number tubes used in extractions was recorded. Where multiple extraction tubes were used for a sample, DNA extractions were combined prior to HTS.

198 A subset of the individuals, representing different taxonomic groups used in bulk non-
199 destructive extraction, was then subjected to individual DNA extraction (see Table S3), DNA
200 barcode amplification and sequencing according to the methods described for extraction 2
201 (above).

202 *Component 3: Non-destructive DNA extraction of field-collected samples*
203
204 In the final component, we examined whether the non-destructive DNA extraction method can
205 be used on samples collected for morphological identification as part of rapid bioassessment
206 surveys (Fig 1C). We obtained six macroinvertebrate samples collected on two occasions (spring
207 2014 and autumn 2016) from three sites along the Merri Creek, Melbourne, Australia (Table S4).
208 Prior to metabarcoding, the six samples were identified mostly to family level according to the
209 Rapid Bioassessment Protocols developed by EPA Victoria (available at:
210 http://www.epa.vic.gov.au/~/media/Publications/604%201.pdf). Therefore, family level
211 information was available for comparison with our species level DNA barcode identifications.
212 Sites were selected based on the diversity of macroinvertebrate families: Coburg Lake (MCL)
213 had low macroinvertebrate diversity, Rushwood Dr (MRD) had high diversity, while the third
214 site at O'Herns Rd (MOH) had an intermediate level. Prior to metabarcoding, spring 2014
215 samples were stored at room temperature in 70% ethanol, while the autumn 2016 samples were
216 stored in 100% ethanol at 4°C.
217 For metabarcoding, we followed the non-destructive DNA extraction workflow for bulk DNA
218 extraction outlined in component 2. Specimens in samples representing taxa not found with
219 metabarcoding or had not been previously DNA barcoded were removed from samples and
220 individually DNA barcoded.
221

High throughput DNA sequencing of bulk invertebrate samples

A two-step PCR process was used to obtain amplicons for Illumina MiSeq sequencing. The first PCR involved amplifying the DNA barcode region (Hebert et al. 2003) using three PCR primer sets. The primer sets include LCOI1490 (Folmer et al. 1994)/MLepR2 (Hebert et al. 2013); B (Hajibabaei et al. 2012)/COIBrev (5'-GATCARACAAAYARWGGYATWCGRTC-3') and miCOIintF (Leray et al. 2013) /HCOI2198 (Folmer et al. 1994). Primers were selected based on their ability to amplify a broad range of taxa. While some taxa may be missed in metabarcoding depending on the primers selected, using the same primer pairs across all extraction methods allowed a comparison of the extraction methods relative success for detecting species.

First round PCR reactions contained 2 µL of DNA template, 16.4 µL molecular biology grade water, 2.5 µL PCR buffer (Invitrogen, Carlsbad, CA, USA), 1 µL $MgCl_2$ (50 mM), 2 µL dNTPs mix (25 mM of each dNTP), 0.5 µL forward primer (10 µM), 0.5 µl reverse primer (10 µM), and 0.1 µL Platinum Taq polymerase (5 U/ml) (Invitrogen, Carlsbad, CA, USA) in a total volume of 25 µL, and were amplified in triplicate using the PCR conditions from Hajibabaei et al (2012). The PCR replicates were pooled and cleaned using a Mag-Sera magnetic beads (GE Healthcare Australia, Sydney, Australia). Cleaned amplicons were quantified with a Qubit Fluorometer using the dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA).

Template-specific primers had Illumina adaptors incorporated onto the 5' end for the attachment of Nextera-XT Illumina indexes (Illumina Corporation, San Diego, CA, USA) in the second round of PCR. The second round PCRs were performed after pooling three amplicons in approximately equal molar ratios. Reactions used 15 µL of the first-round amplicons, 25 µL BIO-X-ACT short mix (Bioline, London, England), 5 µL forward Nextera-XT index primer (10

246 µM), and 5 µL reverse Nextera-XT primer (10 µM). PCR conditions were as follows: 94°C for 5
247 mins followed by 12 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, then 1 cycle of
248 72°C for 5 mins. All amplicons were then cleaned again with Mag-Sera magnetic beads (as
249 above).

250 Library quantification, normalization, pooling and the Illumina MiSeq run using a 600-
251 cycle flow cell MiSeq sequencing kit V3 (300bp x 2) (Illumina, San Diego, CA) were performed
252 by Australian Genome Research Facility Ltd (AGRF) according to the manufacturer's protocols.
253 Raw sequence data generated with HTS was deposited in the National Center for Biotechnology
254 Information (NCBI) Short Read Archive (SRA) database under BioProject PRJNA413851.
255

256 *HTS bioinformatics analysis*
257
258 Sequences from HTS were imported into Geneious version R10 (Kearse et al. 2012) and were
259 trimmed to <197 bp. FLASH version 1.2.9 (Magoc & Salzberg 2011) was used to merge set
260 paired reads with default settings. Merged reads were annotated with each set of forward and
261 reverse primer pairs and then extracted to isolate each of the three amplicons. Only amplicons of
262 appropriate size (+/- 6 bp of the amplicons expected size) with both forward and reverse primers
263 on each end were retained. Primers were then trimmed and a custom *de novo* assembly at 98%
264 similarity was used to reduce redundancy in the dataset. The number of reads that contributed to
265 each contig (group of sequences with > 98% sequence similarity) from the *de novo* assembly was
266 recorded. Singletons and contigs containing ten or fewer reads were discarded (see Bokulich et
267 al. 2013). The remaining contigs were aligned, edited and checked for an open reading frame.
268 Contigs were then BLAST searched against a DNA barcoding reference database of freshwater
269 macroinvertebrates, and species with > 97.5% match were identified. The reference database of

freshwater macroinvertebrates was constructed by downloading and combining DNA barcodes from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and BOLD systems v3 (http://www.boldsystems.org/) databases from families or orders known to have aquatic life stages. The ten best matches were also checked to ensure that species matches were specific. Chimeric sequences were identified and removed with USEARCH version 9 (Edgar et al. 2011) using the freshwater macroinvertebrates as a reference database. Specimens from any families not detected at >97.5% with HTS were subjected to individual DNA barcoding (see above).

Results

Component 1: Individual non-destructive DNA extraction

We obtained DNA suitable for amplifying DNA barcodes from individual freshwater macroinvertebrates following a non-destructive DNA extraction (Table 1 and 2). Sanger sequencing on amplified DNA barcodes revealed the expected species DNA. We obtained DNA barcodes for all specimens in experiment 1a (Table 2), but some Acarina and Oligochaeta in experiment 1b that amplified did not produce amplicons that could be clearly sequenced (Table 3).

For most taxa, whole specimens only required immersion, and subsequent incubation for one hour at 56°C in the T1 buffer with proteinase K from the Nucleospin kit was sufficient to extract DNA for PCR amplification. However, we found that complete digestion of tissue in species from the Gastropoda, Oligochaeta and Hirudinea (Table 2 and 3) was an issue. The 30 min and/or 1 hr incubation completely digested Oligochaeta and Hirudinea specimens, suggesting that these orders are unlikely to be suitable for non-destructive extraction when whole animals are immersed in extraction buffer for 30 mins or longer. The longer incubation of 3 hrs

completely digested tissues in the Gastropoda, but some tissues remained after 1 hr (or 30 min), making a second extraction possible. The presence of a shell meant that taxonomic examination was also possible. We failed to obtain DNA for amplification from an adult Coleopteran after 30 min immersed in the extraction buffer but did after an incubation time of ≥ 1 hr (Table 2) suggesting these specimens require longer immersion in extraction buffer for DNA extraction.

Using an one hour incubation as a standard, it was possible to obtain DNA with a non-destructive total genomic extraction (Extraction 1) from a taxonomically wide range of species (Table 3). However, we found some specimens from the Acarina, Bivalvia as well as a Ceratopogonidae specimen did not amplify. Given that other specimens from these groups did amplify, it is possible that the undetected taxa were extracted successfully but were not amplified by the standard Folmer DNA barcoding primers (LCOI1490/HCOI2198).

We were also able to perform a second Chelex extraction (Extraction 2) on most specimens, including those where DNA was extracted from legs (Amphipoda, Trichoptera, Ephemeroptera, Plecoptera, Hemiptera, Coleoptera (larvae), Megaloptera and Decapoda). However, only some Diptera and Acarina were successfully amplified in the second Chelex extraction. Dissected Diptera tissues tended to stick to forceps during dissection, and this may have resulted in some material being lost. For the Acarina, all tissue with DNA may have been completely digested in the first extraction because of the small size of specimens.

Most specimens were sufficiently intact for morphological identification (Fig S1A). Some specimens had a coating of white residue from the T1 extraction buffer, but this could be removed by adding more ethanol or by manual removal using forceps. Many of the less sclerotized specimens had a hyaline appearance, but this assisted identification when specimens were mounted, particularly for Diptera (i.e. Chironomidae). However, the Amphipoda were

318 substantially damaged during the extraction process and were often missing antennae and legs
319 which are needed for taxonomic identification.
320
321 *Component 2: Bulk DNA extraction for metabarcoding*
322
323 The three extraction methods based on homogenization, dissection and non-destructive
324 extraction of small taxa varied in the amount of time and number of tubes required (Table 1 and
325 4). The most rapid method involved completely homogenizing small taxa in samples (H) along
326 with dissected large taxa, but this method also used the most microcentrifuge tubes for DNA
327 extraction, increasing the cost of extraction per sample. The slowest method was sample
328 dissection (D) of all taxa, which took 2-3 times longer than the homogenization (H) protocol, but
329 this method meant all DNA could be extracted in a single microcentrifuge tube. Non-destructive
330 extraction (ND) only took 1.5 times longer than the homogenization (H) protocol to complete,
331 but used fewer microcentrifuge tubes.
332 All DNA extraction methods produced amplicons for each of the three HTS primer sets,
333 with greater than 84% of the expected taxa detected in samples of known taxon composition
334 (Table 5). The DNA extraction based on dissection of specimens (D) produced the most reliable
335 detection of taxa, with only a single taxon, *Micronecta* sp., not detected in the sample containing
336 80 individuals (Fig 2). Samples that were homogenized or subjected to non-destructive DNA
337 extraction were more variably detected. Generally, homogenized samples with ≥100 individuals
338 detected most taxa (Fig 2). However, six species, *Cosmioperla* sp. ABX7338 (Eustheniidae),
339 *Anisocentropus latifascia* (Calamoceratidae), *Paratanytarsus grimmii* (Chironomidae),
340 *Micronecta* sp. (Corixidae) *Tasmanophlebia* sp. ACM3395 (Oniscigastridae) and *Diplacodes*
341 *haematodes* (Libellulidae), were not detected in the sample containing 80 individuals. Similarly,

up to four taxa were not detected in the samples that underwent non-destructive extraction. These often represented the more sclerotized taxa. For example, heavily sclerotized Coleopterans from the Dytiscidae, Ptilodactylidae and Elmidae were not detected in some samples (Fig 2). We also found that some chironomids were not detected in the sample containing 100 individuals; and the Calocidae (Trichoptera), which can retract into a stone case, were not detected in samples containing 80 and 100 individuals.

Particular extraction methods performed better for some taxa in terms of the number of reads (or sequences) produced and consistency of detection (Fig 2). For example, adult Dytiscidae (*Necterosoma* sp.) consistently produced >1800 reads when whole adult beetles were homogenized for DNA extraction, but when legs were dissected or whole specimens were immersed for DNA extraction, <20 reads were produced. In contrast, the ND method consistently detected Corixidae (*Micronecta* sp.) with >70 reads, whereas H and D consistently produced a low number of reads, and sometimes failed to detect this taxon. Most of the remaining taxa were readily detected with all three extraction methods. Technical replication of sample C with all three extraction methods showed that metabarcoding method was robust (Fig 2). There were only two instances of where taxon represented with low reads was detected in one replicate but not the other.

Overall, when multiple specimens were combined for non-destructive DNA extraction in the T1 buffer with proteinase K and incubated for 1 hr, there was less digestion of tissues in specimens when compared to digestion in the individual DNA extraction trials (Fig S2). In particular, amphipod specimens were largely unaltered by bulk non-destructive extraction. After non-destructive DNA extraction, several taxa were successfully amplified and DNA sequenced following a second Chelex extraction (Table S3).

365

366 *Component 3: Detection of macroinvertebrate taxa in field samples.*
367
368 A high number of sequences were obtained for each of the field-collected samples from Merri
369 Creek, with the exception of the 2014 Coburg Lake (MCL 2014) sample (Fig 3). Overall, most
370 taxa were detected in samples, including the 2014 Coburg Lake sample (Table 5). Species from
371 the orders Gastropoda, Diptera, Coleoptera, Ephemeroptera, Hemiptera and Trichoptera were
372 detected through the ND protocol (Table 1; Fig 3). However, we found that gastropods from the
373 family Hydrobiidae (which have an operculum) were consistently missed with this protocol. We
374 also encountered some individual taxa that were not detected in samples. These tended to be
375 small specimens, such as the Hydroptilidae in the 2016 Coburg Lake sample (MCL) and the
376 Tanypodinae in the 2016 O'Herns Rd (MOH 2016) sample, or more sclerotized specimens such
377 as the Corixidae in the 2014 MCL sample and the Dytiscidae in the 2014 MOH sample. Some
378 non-detections likely resulted from amplification failure rather than extraction failure; e.g. some
379 non-insect taxa, involving species of Oligochaeta, Dugesiidae and Hirudinea, were not detected.
380 There were no obvious differences in the likelihood of detection of species in the 2014 samples
381 compared to 2016 samples (Fig 3), even though the latter were better preserved for DNA
382 isolation after being stored in 100% ethanol at 4°C after collection rather than in 70% ethanol at
383 room temperature.
384 Specimens subjected to the non-destructive extraction protocol remained suitable for
385 subsequent DNA barcoding and taxonomic examination, as established through DNA
386 amplification and sequencing of the DNA barcode region and the presence of taxonomic
387 characters.
388

Discussion

Our study showed that non-destructive DNA extraction could be used for preparing invertebrate samples for metabarcoding. It can allow many different taxa from samples used for metabarcoding to also be used for individual DNA barcoding and taxonomic examination, particularly at the larval/nymph stage. This can help reference DNA barcoding for species level identification, an important step given that incomplete DNA barcoding reference libraries remain a key limitation to identifying species when conducting metabarcoding of environmental samples (Aylagas, Borja & Rodriguez-Ezpeleta 2014; Cristescu 2014; Carew et al. 2017; Elbrecht et al. 2017). In addition, using a non-destructive extraction method can facilitate 'more targeted' production of DNA barcoding reference libraries, particularly where samples have first been identified to coarser taxonomic levels (i.e. family/genus), such as those from rapid bioassessment surveys (i.e. Carew et al. 2016; Elbrecht et al. 2017) as it enables specimens to be examined after they are metabarcoded. In this case, metabarcoding can be conducted first and specimens without DNA barcodes can be recognized later based on morphological examination. These taxa can be removed from samples and be targeted for individual DNA barcoding and further taxonomic investigation to build DNA barcode library coverage. This process is particularly useful for areas where routine surveys are conducted or for sites which are repeatedly surveyed, leading to locally comprehensive DNA reference libraries for determining species diversity. With further testing, non-destructive extraction may become particularly useful when invertebrate samples cannot be destroyed for metabarcoding, such as those in archived or museum collections (i.e. Carew et al. 2016).

While more reliable detection of biodiversity (see Elbrecht et al. 2017) may be possible with DNA extraction protocols using complete homogenization and replication of the extraction

step, we did not to compare the effect of replication at the extraction step in this study. Therefore, it was unclear whether replication would improve the detection of taxa using non-destructive extraction. However, we did complete technical replication of one sample for each extraction method when metabarcoding. This revealed that our metabarcoding was highly reproducible, with same taxa detected in each replicate when using non-destructive extraction. While the other two extraction methods failed to detect one taxon represented by a low number of reads between replicates. This showed that for this sample replication of the metabarcoding step was not important for increasing detection of biodiversity. However, replication of more samples would be useful to determine if this is always the case.

Overall, species detection success for the extraction methods trialled in this study were similar to those employed by other metabarcoding studies using invertebrate samples (Gibson et al. 2014; Elbrecht & Leese 2015; Carew et al. 2016; Elbrecht, Peinert & Leese 2017; Elbrecht et al. 2017). We found the number of species detected with metabarcoding in samples prepared using non-destructive DNA extraction were similar to those where small taxa were dissected or whole individuals homogenized. Moreover, large taxa which were processed via homogenization in all extraction methods were mostly detected at a similar rate to those prepared using non-destructive DNA extraction.

However, where we did not detect taxa, there was bias towards particular groups for the three extraction methods tested here. For example, heavily sclerotized taxa such as Coleoptera were readily detected when whole individuals were homogenized, but often failed to be detected when subjected to non-destructive DNA extraction. It is likely more DNA is released by homogenization of the whole animal, resulting in more DNA being available for PCR. Typically, species biomass is linked to detection success when metabarcoding, and large or more common

436 species can affect the detection of smaller rarer species that contribute less DNA (Elbrecht &
437 Leese 2015; Downie et al. 2016; Elbrecht, Peinert & Leese 2017). PCR biases, due to PCR
438 primer selection and the number of primer sets used, can also lead to metabarcoding failing to
439 detect some taxa from particular orders or families (Clarke et al. 2014; Brandon-Mong et al.
440 2015; Aylagas et al. 2016; Elbrecht & Leese 2017). Our lower detection rates in field samples
441 probably occurred because of a lack of degeneracy in metabarcoding DNA primers, especially in
442 the case of non-insect taxa, such as the Oligochaeta, Hirudinea and Dugesiidae. Additional
443 primer sets, such as those suggested in Elbrecht et al. (2017) which are specifically designed for
444 freshwater invertebrates, could improve the detection of these taxa.
445 While non-destructive DNA extraction did create some amplification bias when
446 metabarcoding, the method could be modified to improve detection of certain taxa. For example,
447 more sclerotized taxa (adult and some larval Coleoptera), or taxa retracted into a stone case
448 (some Trichoptera e.g. Calocidae) or having shell with an operculum (Gastropoda e.g.
449 Hydrobiidae), were less likely to be detected. To counter this, a longer incubation time in the T1
450 extraction buffer with proteinase K may be needed during bulk non-destructive extraction to
451 release sufficient DNA for metabarcoding. Therefore, it may be necessary to separate these taxa
452 from less sclerotized specimens and then use an incubation time >1 hr for non-destructive DNA
453 extraction. Based on initial trials, up to 3 hr may be suitable for these animals, as DNA barcodes
454 were successfully amplified when an individual non-destructive DNA extraction was performed.
455 While this would increase the time taken to conduct DNA extractions due to the requirement for
456 coarse sorting, it would likely be more rapid than dissecting specimens to preserve taxonomic
457 features. Further testing and replication considering a broader range of taxa would be useful to

refine the non-destructive DNA extraction protocol for routine use, especially where validation of metabarcoding and individual DNA barcoding for reference libraries are still required.

There was a substantial difference in appearance of specimens extracted non-destructively as individuals compared to those extracted in bulk. For example, most Diptera and Amphipoda were morphologically unaltered after bulk extraction but had a strong hyaline appearance after individual-based extraction. The hyaline appearance likely reflected a high degree of tissue digestion after 1 hr of incubation in the extraction buffer. In contrast, in most other sclerotized taxa, bulk versus individual extraction had little impact on detection and the former approach often left specimens in better condition for individual DNA barcoding and taxonomic examination. While dissection of all taxa in samples also allowed for taxa to be re-examined, this took substantial time and often led to some taxa being damaged. In particular, small Diptera and Gastropoda were not easily dissected, often resulting in damaged taxonomic characters needed for identification. The morphological integrity of the smaller taxa was better preserved using non-destructive DNA extraction.

Non-destructive DNA extraction of small taxa in combination with dissection of large taxa, and those unsuited to non-destructive DNA extraction (e.g. Oligochaeta, Hirudinea and Dugesiidae), enabled DNA barcoding of taxa with varying levels of sclerotization from over 15 different macroinvertebrate orders. This included invertebrate samples that were stored for at least three years in 70% ethanol at room temperature which is a common practice when morphological examination is needed (Rosenberg & Resh 1993; Haase et al. 2004), and samples stored in 100% ethanol at 4°C which provides better preservation of DNA (Baird et al. 2011; Stein et al. 2014). This means the non-destructive approach could be trialed for extracting DNA from archived or museum collections, even if invertebrates have been stored in 70% ethanol for

up to 3 years. However, taxon detection becomes problematic in samples stored for greater than five year in sub-optimal conditions for DNA preservation (i.e. Carew et al. 2016). Therefore, further testing of non-destructive DNA extraction would be useful to determine how it performs on older material.

Conclusions

In summary, we show that non-destructive DNA extraction protocols can be used for preparing a variety of freshwater invertebrate species for bulk DNA extraction and subsequent metabarcoding. When non-destructive DNA extraction of small taxa is combined with dissection of large taxa, detection of species diversity is comparable to other DNA extraction methods. With further refinement the approach offers means to increase the speed of bulk DNA extraction of invertebrate samples for metabarcoding, while enabling the same samples to be used for individual DNA barcoding and taxonomic identification. The approach also appears suitable for samples not specifically stored for DNA -based approaches.

Acknowledgements

The authors would like to thank Claudette Kellar, David Sharley, Steve Marshall, Eddie Tsyrlin and Cameron Amos for providing macroinvertebrate samples for this study.

References

Aylagas E, Borja Á, Irigoien X, Rodríguez-Ezpeleta N. 2016. Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science* 3. DOI:10.3389/fmars.2016.00096.

Aylagas E, Borja A, Rodriguez-Ezpeleta N. 2014. Environmental status assessment using DNA metabarcoding: Towards a genetics based Marine Biotic Index (gAMBI). *PLoS ONE* 9:e90529. DOI:10.1371/journal.pone.0090529.

Baird DJ, Hajibabaei M. 2012. Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology* 21:2039-2044. DOI:10.1111/j.1365-294X.2012.05519.x.

Baird DJ, Pascoe TJ, 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. DOI:10.1899/10-013.1.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 360:1935-1943. DOI:10.1098/rstb.2005.1725.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods* 10:57-59. DOI:10.1038/nmeth.2276.

524 Brandon-Mong GJ, Gan HM, Sing KW, Lee PS, Lim PE, Wilson JJ. 2015. DNA metabarcoding
525 of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological*
526 *Research* 105:717-727. DOI:10.1017/S0007485315000681.
527 Carew ME, Nichols SJ, Batovska J, St Clair R, Murphy NP, Blacket MJ, Shackleton ME. 2017.
528 A DNA barcode database of Australia's freshwater macroinvertebrate fauna. *Marine and*
529 *Freshwater Research* 68: 1788-1802. DOI:10.1071/MF16304.
530 Carew ME, Hoffmann AA. 2015. Delineating closely related species with DNA barcodes for
531 routine biological monitoring. *Freshwater Biology* 60:1545-1560.
532 DOI:10.1111/fwb.12587.
533 Carew ME, Kellar CR, Pettigrove VJ, Hoffmann AA. 2018. Can high-throughput sequencing
534 detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological*
535 *Indicators* 85:440-450. DOI:https://doi.org/10.1016/j.ecolind.2017.11.002.
536 Carew ME, Metzeling L, St Clair R, Hoffmann AA. 2016. Detecting invertebrate species in
537 archived collections using next-generation sequencing. *Molecular Ecology Resources* 7:
538 915–930. DOI:10.1111/1755-0998.12644.
539 Carew ME, Pettigrove V, Cox RL, Hoffmann AA. 2007. The response of Chironomidae to
540 sediment pollution and other environmental characteristics in urban wetlands. *Freshwater*
541 *Biology* 52:2444-2462. DOI:10.1111/j.1365-2427.2007.01840.x.
542 Carew ME, Pettigrove V, Hoffmann AA. 2003. Identifying chironomids (Diptera:
543 Chironomidae) for biological monitoring with PCR-RFLP. *Bulletin of Entomological*
544 *Research* 93:483-490. DOI:10.1079/BER2003268.

545 Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA. 2013. Environmental monitoring using
546 next generation sequencing: Rapid identification of macroinvertebrate bioindicator
547 species. *Frontiers in Zoology* 10:45. DOI:10.1186/1742-9994-10-45.
548 Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H. 2012. Don't make a
549 mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing
550 studies? *Fungal Ecology* 5:747-749. DOI:10.1016/j.funeco.2012.06.003.
551 Castalanelli MA, Severtson DL, Brumley CJ, Szito A, Foottit RG, Grimm M, Munyard K, Groth
552 DM. 2010. A rapid non-destructive DNA extraction method for insects and other
553 arthropods. *Journal of Asia-Pacific Entomology* 13:243-248.
554 DOI:10.1016/j.aspen.2010.04.003.
555 Clarke LJ, Soubrier J, Weyrich LS, Cooper A. 2014. Environmental metabarcodes for insects: In
556 silicoPCR reveals potential for taxonomic bias. *Molecular Ecology Resources* 14:1160-
557 1170. DOI:10.1111/1755-0998.12265.
558 Cornils A. 2015. Non-destructive DNA extraction for small pelagic copepods to perform
559 integrative taxonomy. *Journal of Plankton Research* 37:6-10.
560 DOI:10.1093/plankt/fbu105.
561 Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M,
562 Carvalho GR, Blaxter ML, Lambshead PJD, Thomas WK. 2010. Ultrasequencing of the
563 meiofaunal biosphere: Practice, pitfalls and promises. *Molecular Ecology* 19:4-20.
564 DOI:10.1111/j.1365-294X.2009.04473.x.
565 Cristescu ME. 2014. From barcoding single individuals to metabarcoding biological
566 communities: towards an integrative approach to the study of global biodiversity. *Trends*
567 *in Ecology and Evolution* 29:566-571. DOI:10.1016/j.tree.2014.08.001.

568 Dowle E, Pochon X, Banks J, Shearer K, Wood SA. (2016) Targeted gene enrichment and high
569 throughput sequencing for environmental biomonitoring: A case study using freshwater
570 macroinvertebrates. *Molecular Ecology Resources* 16:1240-1254. DOI: 10.1111/1755-
571 0998.12488.
572 Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and
573 speed of chimera detection. *Bioinformatics* 27:2194-2200.
574 DOI:10.1093/bioinformatics/btr381.
575 Elbrecht V, Leese F. 2015. Can DNA-Based ecosystem assessments quantify species abundance?
576 Testing primer bias and biomass--sequence relationships with an innovative
577 metabarcoding protocol. *PLoS ONE* 10:e0130324. DOI:10.1371/journal.pone.0130324.
578 Elbrecht V, Leese F. 2017. Validation and development of COI metabarcoding primers for
579 freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science* 5.
580 DOI:10.3389/fenvs.2017.00011.
581 Elbrecht V, Peinert B, Leese F. 2017. Sorting things out: Assessing effects of unequal specimen
582 biomass on DNA metabarcoding. *Ecology and Evolution* 7:6918-6926.
583 DOI:10.1002/ece3.3192.
584 Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F. 2017. Assessing strengths and
585 weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine
586 stream monitoring. *Methods in Ecology and Evolution* 8: 1265–1275. DOI:10.1111/2041-
587 210X.12789.
588 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of
589 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
590 *Molecular Marine Biology and Biotechnology* 3:294-299.

591 Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, Hallwachs W,
592 Hajibabaei M. 2014. Simultaneous assessment of the macrobiome and microbiome in a
593 bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the*
594 *National Academy of Sciences* 111:8007-8012. DOI:10.1073/pnas.1406468111.
595 Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M. 2015. Large-scale
596 biomonitoring of remote and threatened ecosystems via high-throughput sequencing.
597 *PLoS ONE* 10:e0138432. DOI:10.1371/journal.pone.0138432.
598 Haase P, Lohse S, Pauls S, Schindehuette K, Sundermann A, Rolauffs P, Hering D. 2004.
599 Assessing streams in Germany with benthic invertebrates: Development of a practical
600 standardised protocol for macroinvertebrate sampling and sorting. *Limnologica* 34:349-
601 365. DOI:10.1016/S0075-9511(04)80005-7.
602 Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011. Environmental Barcoding: A
603 next-generation sequencing approach for biomonitoring applications using river benthos.
604 *PLoS ONE* 6:e17497. DOI:10.1371/journal.pone.0017497.
605 Hajibabaei M, Spall J, Shokralla S, van Konynenburg S. 2012. Assessing biodiversity of a
606 freshwater benthic macroinvertebrate community through non-destructive environmental
607 barcoding of DNA from preservative ethanol. *BMC Ecology* 12:28. DOI:10.1186/1472-
608 6785-12-28.
609 Hawking JH. 2000. A preliminary guide to keys and zoological information to identify
610 invertebrates form Australian freshwaters. Identification Guide No 2 (2nd Edition),
611 Cooperative Research Centre for Freshwater Ecology: Albury, Australia..

612 Hebert PD, Zakharov EV, Prosser SW, Sones JE, McKeown JT, Mantle B, La Salle J. 2013. A
613 DNA 'Barcode Blitz': Rapid digitization and sequencing of a natural history collection.
614 *PLoS ONE* 8:e68535. DOI:10.1371/journal.pone.0068535.
615 Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identification through DNA
616 barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences*
617 270:313-321. DOI:10.1098/rspb.2002.2218.
618 Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA,
619 Pedersen S, Jannetta P, deWaard JR. 2016. Counting animal species with DNA barcodes:
620 Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*
621 371. DOI:10.1098/rstb.2015.0333.
622 Hendrich L, Morinière J, Haszprunar G, Hebert PDN, Hausmann A, Köhler F, Balke M. 2015. A
623 comprehensive DNA barcode database for Central European beetles with a focus on
624 Germany: Adding more than 3500 identified species to BOLD. *Molecular Ecology*
625 *Resources* 15:795-818. DOI:10.1111/1755-0998.12354.
626 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
627 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious
628 Basic: An integrated and extendable desktop software platform for the organization and
629 analysis of sequence data. *Bioinformatics* 28:1647-1649.
630 DOI:10.1093/bioinformatics/bts199.
631 Kranzfelder P, Ekrem T, Stur E. 2017. DNA Barcoding for Species Identification of Insect
632 Skins: A test on Chironomidae (Diptera) pupal exuviae. *Journal of Insect Science*
633 17:111-111. DOI:10.1093/jisesa/iex075.

Krosch MN, and Cranston PS. 2012. Non-destructive DNA extraction from Chironomidae, including of fragile pupal exuviae, extends analysable collections and enhances vouchering. *Chironomus Newsletter on Chironomidae Research* 25:22-27. DOI:10.5324/cjcr.v0i25.1532

Lanzén A, Lekang K, Jonassen I, Thompson EM, Troedsson C. 2016. High-throughput metabarcoding of eukaryotic diversity for environmental monitoring of offshore oil-drilling activities. *Molecular Ecology* 25:4392-4406. DOI:10.1111/mec.13761.

Lee MSY. 2004. The molecularisation of taxonomy. *Invertebrate Systematics* 18:1-6. DOI:10.1071/Is03021.

Leray M, Yang J, Meyer C, Mills S, Agudelo N, Ranwez V, Boehm J, Machida R. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.

Magoc T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957-2963. DOI:10.1093/bioinformatics/btr507.

Moritz C, Cicero C. 2004. DNA barcoding: Promise and pitfalls. *PLoS Biology* 2:1529-1531. DOI:10.1371/journal.pbio.0020354.

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources* 14:1129-1140. DOI:10.1111/1755-0998.12261.

Porco D, Rougerie R, Deharveng L, Hebert P. 2010. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high-throughput context: The

example of Collembola. *Molecular Ecology Resources* 10:942-945. DOI:10.1111/j.1755-0998.2010.2839.x.

Rosenberg DM, Resh VH. 1993. *Freshwater biomonitoring and benthic macroinvertebrates*: Chapman & Hall.

Rowley DL, Coddington JA, Gates MW, Norrbom AL, Ochoa RA, Vandenberg NJ, Greenstone MH. 2007. Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes* 7:915-924. DOI:10.1111/j.1471-8286.2007.01905.x.

Schafer RB, Kefford B, Metzeling L, Liess M, Burgert S, Marchant R, Pettigrove V, Goonan P, Nugegoda D. 2011. A trait database of stream invertebrates for the ecological risk assessment of single and combined effects of salinity and pesticides in South-East Australia. *Science of the Total Environment* 409:2055-2063. DOI:10.1016/j.scitotenv.2011.01.053.

Stein ED, Martinez MC, Stiles S, Miller PE, Zakharov EV. 2014. Is DNA barcoding actually cheaper and faster than traditional morphological methods: Results from a survey of freshwater bioassessment efforts in the United States? *PLoS ONE* 9:e95525. DOI:10.1371/journal.pone.0095525.

Sweeney BW, Battle JM, Jackson JK, 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. DOI:10.1899/10-016.1.

Wong WH, Tay YC, Puniamoorthy J, Balke M, Cranston PS, Meier R. 2014. 'Direct PCR' optimization yields a rapid, cost-effective, nondestructive and efficient method for

obtaining DNA barcodes without DNA extraction. *Molecular Ecology Resources* 14:1271-1280. DOI:10.1111/1755-0998.12275.

Yu DW, Ji YQ, Emerson BC, Wang XY, Ye CX, Yang CY, Ding ZL. 2012. Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3:613-623. DOI:10.1111/j.2041-210X.2012.00198.x.

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2:4. DOI:10.1186/2047-217x-2-4.

Zimmermann J, Abarca N, Enke N, Skibbe O, Kusber WH, Jahn R. 2014. Taxonomic reference libraries for environmental barcoding: A best practice example from diatom research. *PLoS ONE* 9:e108793. DOI:10.1371/journal.pone.0108793.

691

Table 1(on next page)

Outline for the experimental workflow for testing non-destructive DNA extraction

1
2

Component	Questions	Answers
1. Non-destructive DNA extraction of single individuals	Can non-destructive DNA extraction be used to obtain DNA from macroinvertebrates?	Yes, but not from taxa with little sclerotization, such as the Oligochaeta and Hirudinea.
	What is a suitable incubation time for immersion in extraction buffer that does not destroy morphological traits but yields DNA?	An hour works best for most taxa, but for Amphipoda less time would be needed to avoid damage to taxonomic characters.
	Can DNA be extracted from a specimen used for non-destructive DNA extraction?	Yes, for most taxa this is possible. It can be difficult for small taxa like Acarina and some Diptera
2. Comparison of bulk non-destructive extraction to tissue-based DNA extraction for metabarcoding samples of known composition – multiple individuals	How does bulk non-destructive (ND) extraction compares to total homogenization (H) and tissue dissection (D) based extractions when detecting taxa with metabarcoding?	Detection of taxa was similar across all three extraction methods with >84% of taxa detected. However, some more sclerotized taxa (Coleoptera) were often missed by the ND method compared to other methods.
	Can specimens from bulk ND extractions be used for DNA barcoding and taxonomic identification?	Yes, unlike individual ND extraction, specimens subject to bulk ND were largely unaltered and were easily used for DNA barcoding and taxonomic identification
3. Non-destructive DNA extraction of rapid bioassessment samples stored for morphological identification – multiple individuals	Can taxa be detected in field samples when ND extraction is used?	Yes, most taxa were detected using ND extraction protocol except Hydrobiidae snail and some small taxa.
	Do storage conditions (100% ethanol with refrigeration) and sub-optimally (70% ethanol at room temperature) affect taxa detection success when using non-destructive DNA extraction?	There were no obvious difference in the number of taxa detected between samples stored under different conditions.

Table 2(on next page)

Results of experiment 1a examining the effect of incubation time in the T1 buffer with proteinase K (from the Nucleospin DNA extraction kit) on non-destructive total genomic extractions of species from nine invertebrate groups

The success of non-destructive (ND) extraction (Extraction 1) for standard DNA barcoding PCR is indicated by the '+' symbol ('-' for no PCR product), while the success of the second Chelex extraction (Extraction 2) on the same material is indicated the adjacent column. GenBank accession numbers for sequenced specimens are given in parenthesis.

Order/Subclass	Species (Family)	Sclerotization	Life stage	Incubation time for Extraction 1					
				30 min		1 hr		3 hr	
				Extraction 1 (1st ND)	Extraction 2 (2nd Chelex)	Extraction 1 (1st ND)	Extraction 2 (2nd Chelex)	Extraction 1 (1st ND)	Extraction 2 (2nd Chelex)
Oligochaeta	*Lumbriculus variegatus* (Lumbriculidae)	Soft bodied	Adult	+ (MG976202)	-	+	-		
Gastropoda	*Physa acuta* (Physidae)	soft bodied/shell	Adult	+ (MG976201)	+ (MG976114)	+	+	+	-
Amphipoda	*Austrochiltonia subtenuis* (Chiltoniidae)	light sclerotized	Adult	+ (MG976104)	+ (MG976105)	+	+		
Diptera	*Procladius villosimanus* (Chironomidae)	sclerotized head	Larvae			+ (MG976143)	+ (MG976144)	+	+
Trichoptera	*Hellyethira simplex* (Hydroptilidae)	sclerotized head and thorax	Larvae	+ (MG976170)	+ (MG976171)	+ (+)			
Ephemeroptera	*Offadens* sp. (Baetidae)	sclerotized	Nymph	+ (MG976108)	+ (MG976109)	+ (+)			
Plecoptera	*Dinotoperla thwaites* (Gripopterygidae)	sclerotized	Nymph	+ (MG976167)	+ (MG976167)	+ (+)			
Hemiptera	*Micronecta* sp. (Corixidae)	sclerotized	Adult	+ (MG976178)	+ (MG976178)	+ (+)			
Coleoptera	*Necterosoma* sp. (Dytiscidae)	heavily sclerotized	Adult	-	+	+ (MH000193)	+* (MH000194)	+	+*

* Extraction 2 using crushed legs failed, but 2 hr incubation of whole animal in Chelex (with proteinase K) was successful.

Table 3(on next page)

Results of experiment 1b examining non-destructive DNA extraction trial on individuals from multiple macroinvertebrate groups.

A breakdown of the species in each family can be found in Table S1.

1

Higher Taxonomic rank	Sclerotization	Families	Material for extraction		Individuals tested	Number of individuals with successful extractions		
			Extraction 1	Extraction 2		Extraction 1 (1st ND)	Extraction 2 (2nd Chelex)	Both extractions
Acarina	moderate	2*	whole animal	whole animal	6	4	1	1
Oligochaeta	soft bodied	3*	whole animal	no material	5	5	0	0
Hirudinea	soft bodied	1	whole animal	no material	1	1	0	0
Bivalvia	soft bodied/ shell	1	whole animal	no material	3	2	0	0
Gastropoda	soft bodied/ shell	2	whole animal	Whole animal	5	5	4	4
Diptera	Little - moderate	6	whole animal	tissue sample	12	11	5	5
Trichoptera	moderate	5	whole animal	whole animal/leg	8	8	8	8
Ephemeroptera	moderate	2	whole animal	leg	4	4	4	4
Plecoptera	moderate	2	whole animal	leg	4	4	4	4
Hemiptera	moderate - heavy	3	whole animal	leg	4	4	4	4
Coleoptera	moderate - heavy	5	whole animal	leg	8	8	8	8
Amphipoda	moderate	2	whole animal	leg	4	4	4	4
Megaloptera	moderate	1	whole animal	leg	1	1	1	1
Decapoda	moderate - heavy	2	whole animal	leg	2	2	2	2
Odonata	moderate	1	whole animal	leg	1	1	1	1

2 *All families in these orders were not identified. Lowest possible identifications are given.

3

Table 4(on next page)

Sample preparation and comparison of species detection using metabarcoding based on three DNA extraction methods.

Extraction methods are based on non-destructive (ND), homogenization (H) and dissection (D) in samples of known taxonomic composition containing 80, 100 or 120 individuals.

1

	Extraction method	Sample A 80 individuals	Sample B 100 individuals	Sample C 120 individuals (technical replicate)
Time taken to prepare sample for DNA extraction (min)	ND	10	15	10
	H	7	10	8
	D	21	20	27
Number of DNA extraction tubes required	ND	2	2	2
	H	3	3	3
	D	1	1	1
Number of species detected with HTS	ND	46	33	40 (40)
	H	38	45	37(38)
	D	47	41	41 (42)
% of expected species or genera detected with metabarcoding compared to samples of known taxonomic composition	ND	93	89	98 (98)
	H	88	100	98 (100)
	D	98	100	98 (100)

2

Table 5(on next page)

Amplicon size and number of reads (sequences) obtained for each of the three amplicons used for HTS.

Samples include three sets of constructed samples containing 3 replicates extracted using non-destructive (ND), complete homogenization (H) and dissection (D) based DNA extraction protocols, and six field-collected macroinvertebrate samples from three sites along Merri Creek (Melbourne, Australia) extracted using a non-destructive extraction protocol.

1

Sample (number of individuals)	Extraction method	LCOI/MLepR2 amplicon	B/COIBrev amplicon	mtCOIintF/HCOI amplicon		Overall species detection success (%)
		280bp	293bp	313bp	Total	
Samples of known composition:						
Sample A (80)	ND	48288	56031	44645	148964	92
	H	117667	116313	61850	295830	84
	D	79356	86299	38716	204371	97
Sample B (100)	ND	67448	75433	19463	162344	89
	H	44783	49313	14826	108922	100
	D	111207	124103	68281	303591	100
Sample C (120)	ND	76164	54706	32963	163833	95
	H	39418	65326	24561	129305	95
	D	79524	109656	26218	215398	95
Sample C (120) technical replicate	ND	70875	65871	20231	156977	95
	H	86622	134934	24233	245789	100
	D	50590	93821	26826	171237	100
Field collected samples:						
MRD 2016	ND	75891	44593	69640	190124	88
MRD 2014	ND	69445	43188	57083	169716	73
MOH 2016	ND	47943	41761	57124	146828	83
MOH 2014	ND	67644	49726	39482	156852	80
MCL 2016	ND	81108	15177	60074	156359	73
MCL 2014	ND	727	687	164	1578	55 (low reads)

2

3

Figure 1(on next page)

Diagram of the workflow for testing the success of non-destructive DNA extraction.

(A) Individual non-destructive DNA extraction (Extraction 1) followed by a Chelex extraction (Extraction 2) performed on single specimens from 15 different macroinvertebrate orders (Component 1). (B) Comparison of the detection of taxa via metabarcoding by using three DNA extraction based on non-destructive processing, homogenization of whole small taxa and dissection of all taxa (Component 2). (C) Comparison of the detection of taxa via metabarcoding from 6 field collected samples stored in different ethanol concentration, and temperatures an with varying levels of diversity. Drawings by Melissa Carew.

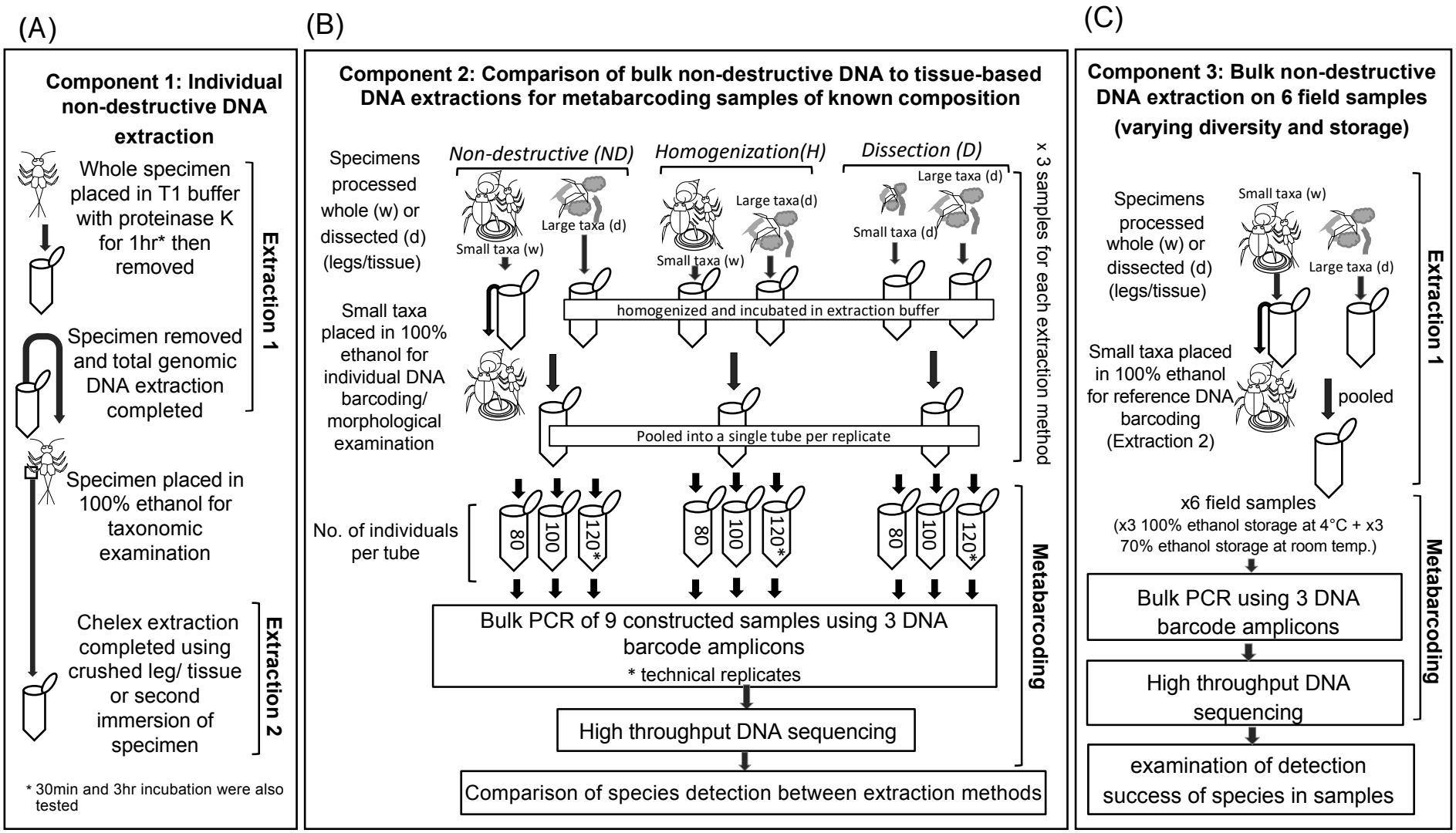

(A)
Component 1: Individual non-destructive DNA extraction
Whole specimen placed in T1 buffer with proteinase K for 1hr* then removed
Specimen removed and total genomic DNA extraction completed
Extraction 1
Specimen placed in 100% ethanol for taxonomic examination
Chelex extraction completed using crushed leg/ tissue or second immersion of specimen
Extraction 2
* 30min and 3hr incubation were also tested
(B)
Component 2: Comparison of bulk non-destructive DNA to tissue-based DNA extractions for metabarcoding samples of known composition
Specimens processed whole (w) or dissected (d) (legs/tissue)
Non-destructive (ND)
Homogenization(H)
Dissection (D)
Small taxa (w)
Large taxa (d)
Large taxa(d)
Small taxa placed in 100% ethanol for individual DNA barcoding/ morphological examination
homogenized and incubated in extraction buffer
Pooled into a single tube per replicate
x 3 samples for each extraction method
No. of individuals per tube
80
100
120*
Bulk PCR of 9 constructed samples using 3 DNA barcode amplicons
* technical replicates
High throughput DNA sequencing
Comparison of species detection between extraction methods
Metabarcoding
(C)
Component 3: Bulk non-destructive DNA extraction on 6 field samples (varying diversity and storage)
Specimens processed whole (w) or dissected (d) (legs/tissue)
Small taxa placed in 100% ethanol for reference DNA barcoding (Extraction 2)
pooled
Extraction 1
x6 field samples
(x3 100% ethanol storage at 4°C + x3 70% ethanol storage at room temp.)
Bulk PCR using 3 DNA barcode amplicons
High throughput DNA sequencing
examination of detection success of species in samples
Metabarcoding

Figure 2(on next page)

High throughput DNA sequencing of nine constructed samples of known taxonomic composition extracted using non-destructive (ND), complete homogenization (H) and dissection (D) based DNA extraction protocols.

Samples contain 80, 100 or 120 individuals. Sample C was run with a technical replicate. The greyscale indicates the number of reads (sequences) returned that match each taxon and 'n' the number of specimens used from each taxonomic group to compose samples. Species underlined were 'small taxa' extracted using the non-destructive method. Note: the Leptophlebiidae could not be easily distinguished as many specimens were missing legs and were early instars, so detection within samples with the same number of individuals varied.

PeerJ

Manuscript to be reviewed

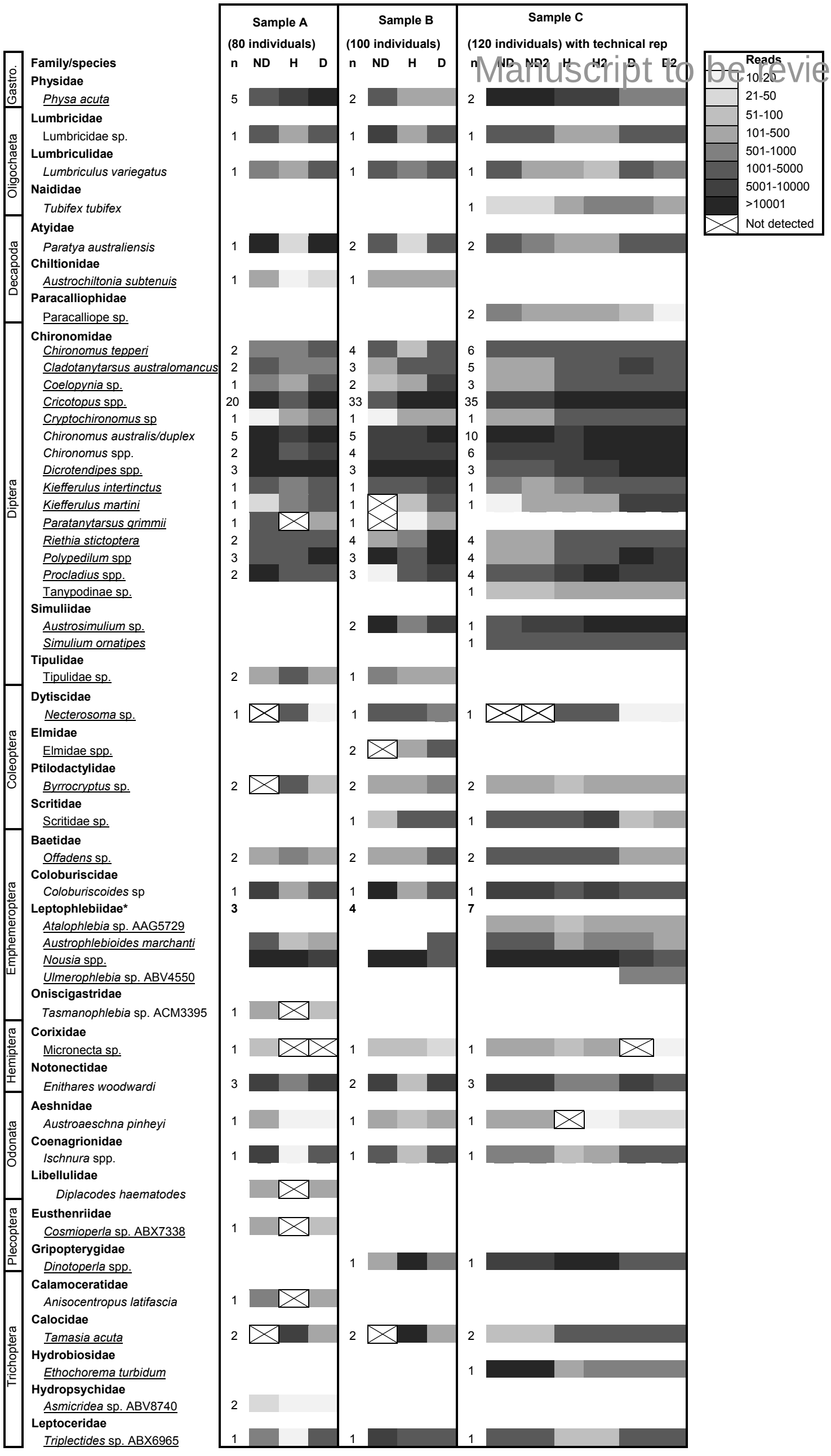
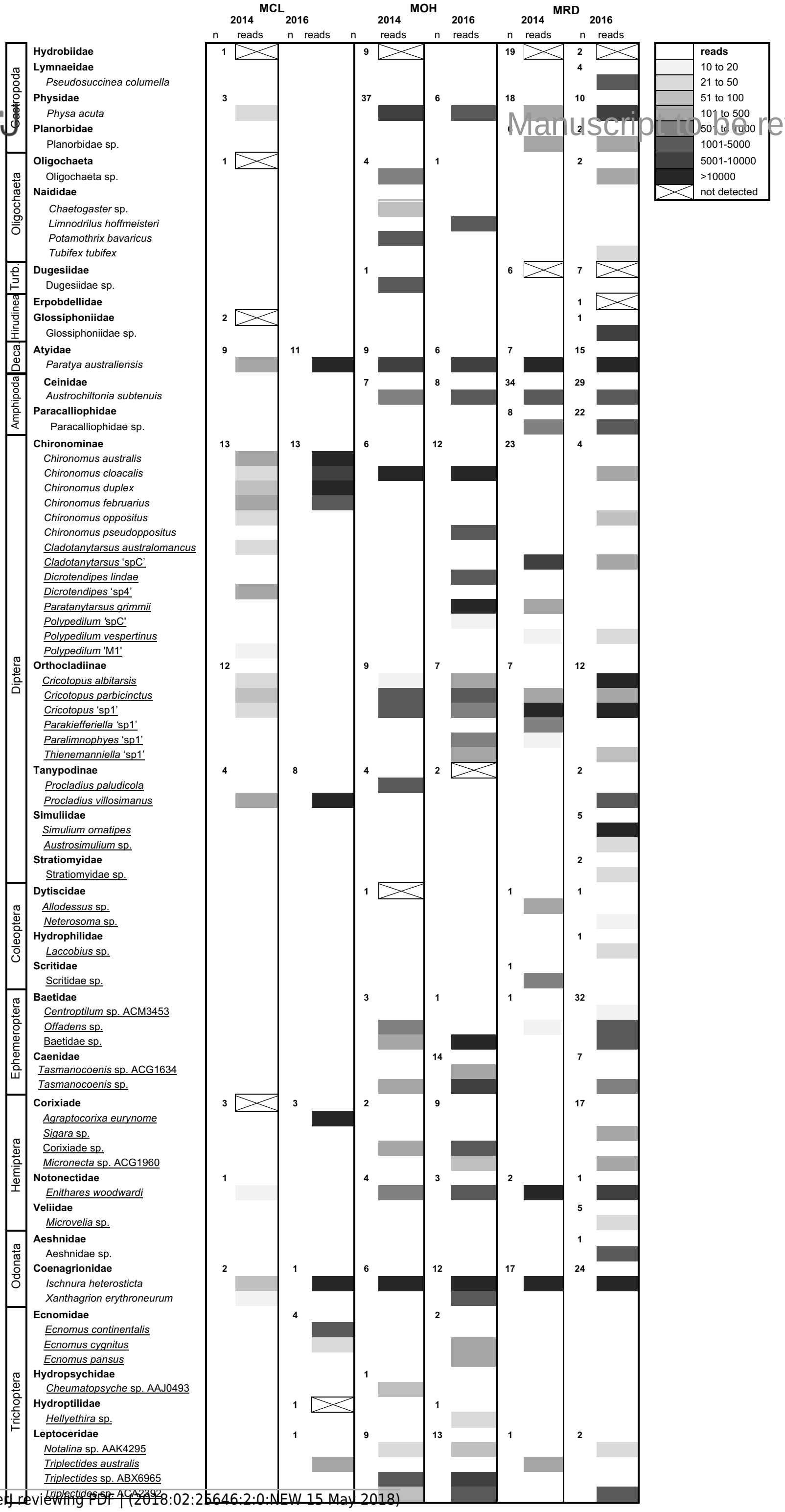


Figure 3(on next page)

High-throughput DNA sequencing of six field-collected samples from Merri Creek 2014-2016.

Sites in Melbourne, Australia include Merri Creek at Coburg Lake, Coburg (MCL); Merri Creek at O'Herns Road, Broadmeadows (MOH) and Merri Creek at Rushwood Drive, Craigieburn (MRD). The greyscale indicates the number of reads (sequences) returned that match each taxon and ‘n’ the number of individuals morphologically identified from each family. Species underlined were ‘small taxa’ extracted using the non-destructive method. Some species with unclear taxonomy are listed by their BOLD BIN (three letters and four numbers) which can be used to find the sequence data and taxonomic information on the taxon on the BOLD version 3 website at www.http://v3.boldsystems.org

PeerJ Manuscript to be reviewed