Using unsorted sweep net samples to rapidly assess macroinvertebrate biodiversity

Melissa E. Carew[1], Rhys A. Coleman[2], Katie L. Robinson[1] and Ary A. Hoffmann[1]

[1] Pest and Environmental Adaptation Research Group (PEARG), School of BioSciences, Bio21 Institute, 30 Flemington Rd, The University of Melbourne 3010, Victoria, Australia.

[2] Applied Research, Melbourne Water, 990 La Trobe Street, Docklands 3008, Victoria, Australia.

Author for correspondence: Melissa Carew

Address: Pest and Environmental Adaptation Research Group (PEARG), Bio21 institute, 30 Flemington Rd, The University of Melbourne, Victoria, Australia, 3010.

Phone: +61 3 9035 3530

E-mail: mecarew@unimelb.edu.au

Running title: Assessing biodiversity with unsorted samples

Key words: DNA barcoding, DNA metabarcoding, macroinvertebrates, freshwater, bulk samples

Abstract

Macroinvertebrate biodiversity is routinely used to assess the ecological condition of freshwater environments. Macroinvertebrates are traditionally identified morphologically to mostly family level but using DNA metabarcoding they can be rapidly and reliably identified to species. Developing standardised, robust and cost-effective protocols would enable DNA metabarcoding to be broadly used for routine freshwater biological assessments leading to both potential cost savings as well as increased taxonomic resolution. To further reduce the cost and time it takes to process samples, we examine the feasibility of DNA metabarcoding unsorted macroinvertebrates (macroinvertebrates and debris) from sweep net samples collected from stream pool or edge habitats. We processed these unsorted samples with a standardised method and tested multiple primer sets widely used for invertebrate DNA metabarcoding. We found that currently available DNA metabarcoding primers are well suited for processing unsorted samples, though some performed better than others. However, macroinvertebrate density affected the number of species detected with DNA metabarcoding, particularly the detection of rare taxa. These findings show that DNA metabarcoding of unsorted net contents could streamline macroinvertebrate sample processing for bioassessment, but the optimal amounts of unsorted material for sub-sampling needs to be considered when assessing macroinvertebrate biodiversity.

41 **Introduction**

42 Biological monitoring with macroinvertebrates is routinely used to understand the condition
43 of freshwater ecosystems (Rosenberg & Resh 1993). Macroinvertebrate monitoring has
44 traditionally been done by collecting macroinvertebrates using a net, sorting them from
45 debris and then identifying them under a microscope (Lenat 1988; Chessman 1995).
46 However, advances in DNA sequencing have enabled macroinvertebrate samples to be
47 identified using a process known as DNA metabarcoding (Hajibabaei *et al.* 2011; Yu *et al.*
48 2012). DNA metabarcoding enables bulk processing of macroinvertebrate samples and
49 promises to provide cost-effective and rapid species identification for bioassessment
50 (Aylagas *et al.* 2016; Porter & Hajibabaei 2018).

51 DNA metabarcoding has successfully identified macroinvertebrate species in sorted
52 macroinvertebrate samples, where macroinvertebrates have been separated from debris
53 collected in sampling nets (e.g. Elbrecht *et al.* 2017b; Emilson *et al.* 2017; Carew *et al.*
54 2018b). While this process can provide identifications that strongly overlap with
55 morphologically identified specimens (Hajibabaei *et al.* 2011; Carew *et al.* 2018b), it misses
56 an opportunity to reduce the manual processing time associated with sorting
57 macroinvertebrates from debris. Elbrecht *et al.* (2017b) attribute two thirds of the cost of
58 processing macroinvertebrate samples to the collection and sorting of samples. Therefore,
59 protocols that enable more rapid sample processing could significantly decrease monitoring
60 costs.

61 One approach for reducing sample collection and sorting time is to assess
62 macroinvertebrate biodiversity using eDNA isolated from water samples. However, eDNA
63 can miss most macroinvertebrate biodiversity at sites when compared to
64 macroinvertebrates isolated from sampling nets and may be confounded by the detection of

non-target taxa (Macher *et al.* 2018; Hajibabaei *et al.* 2019a; Pereira-da-Conceicoa *et al.* 2019). Another option is to directly process unsorted macroinvertebrate samples (i.e. with debris) from sampling nets with DNA metabarcoding. Using this approach, Macher *et al.* (2018) found double the of number operational taxonomic units (OTUs) belonging to the Ephemeroptera, Plecoptera and Trichoptera when compared to eDNA sampling, suggesting substantially better detection of macroinvertebrate biodiversity. Majaneva et al. (2018) showed that the choice of DNA extraction method and primer selection with DNA metabarcoding could further improve the proportion of macroinvertebrate OTUs detected. Nichols *et al.* (2019) and Hajibabaei *et al.* (2019a) also detected a high proportion of macroinverterbrate DNA metabarcodes using unsorted net contents as a source of macroinverterbrate DNA with single and multi-amplicon approaches.

While these studies show the potential of DNA metabarcoding unsorted samples for biodiversity assessments, they have mostly focused on kick net samples, where less debris is typically collected compared to sweep net samples taken from pool or edge habitats. However, bioassessment is typically based on both kick and sweep net samples (e.g. Chessman 1995). To understand how effectively macroinvertebrate diversity can be identified from unsorted sweep net macroinvertebrate samples, we compare DNA metabarcoding identification of unsorted sweep net samples with sorted sweep net samples identified both morphologically and using DNA metabarcoding. We assess the performance of six DNA metabarcoding primer combinations and examine the composition and density of macroinvertebrates between samples. The family-level SIGNAL biotic index (Chessman 1995; Chessman 2003) and number of EPT (Ephemeroptera, Plecoptera and Trichoptera) taxa, which are commonly used in bioassessment in Australia, were also calculated for each

site using morphological and DNA metabarcoding identifications. We discuss the costs of processing unsorted samples compared to other processing methods.

Materials and Methods

Study design and unsorted sample processing

This study was designed to develop a standardised method for the high-throughput processing of unsorted macroinvertebrate sweep net samples from pool or edge habitats using DNA metabarcoding (Fig 1). Macroinvertebrate sweep net samples used in this study were collected from seven sites as part of routine biomonitoring surveys, according to rapid bioassessment protocols, (Chessman 1995) in autumn 2018 in the greater Melbourne area, Victoria, Australia (Table 1). A subset of macroinvertebrates was sorted from net collections by either a live pick or a laboratory sub-sampling method according to Walsh (1997). These 'sorted' samples were used to compare species compositions to 'unsorted' samples. The remaining residues or unsorted samples from each site were retained and preserved in 100% ethanol at 4°C for further processing. Sites sampled were selected so that they had varied macroinvertebrate diversity, density and debris compositions (Table 1).

Sorted macroinvertebrate samples were identified to family level morphologically and to species level with DNA metabarcoding (see below). The unsorted samples (which contained remaining macroinvertebrates) were sieved by thoroughly washing the residues with distilled water through three 250 mm entomological sieves with mesh sizes of 20 mm^2, 8 mm^2 and 250 μm^2. To reduce the volume of plant material, the material from the >20 mm^2 fraction was scanned for very large macroinvertebrate taxa which were retained, and the remaining material (of mostly leaves and twigs) were discarded. The 8 mm^2 fraction was also scanned for large taxa which were also retained, and the remaining material discarded.

112 The entire fraction of material between 8 mm^2 – 250 μm^2 was retained (as this fraction
113 contains the highest abundance of macroinvertebrates). For one sample this material was
114 evenly split into three portions and all taxa from the first portion were removed and
115 transferred to the third portion, to create a control with three different densities (no
116 invertebrates, standard macroinvertebrates and double macroinvertebrates). For all
117 samples the density of animals relative to debris was measured by retaining 5 ml of each
118 sample. For the 5 ml samples, all animals were sorted from the debris and identified to
119 family level, and dry weights for animals and debris were recorded to produce a density
120 ratio (Table 1).
121 The remaining material from the 8 mm^2 – 250 μm^2 fraction for all samples was dried
122 overnight at 40°C in deep petri dishes (Nunc A/S, Demark) in preparation for DNA
123 extraction. The total dry weight of each sample was recorded and 10% of the sample or 1 g
124 (if 10% of the sample dry weight was less than 10 g) were placed in 2 ml safe-lock tubes
125 (Eppendorf Pty. Ltd, Germany) along with a leg or a piece tissue from any large
126 macroinvertebrates retained from the larger fractions for DNA extraction. DNA
127 metabarcoding was used to identify species in unsorted samples. Cost differences for
128 standard methods compared to the bulk extraction were estimated based on laboratory
129 costs and the costs of consultants who collected the samples (Fig 1).
130
131 ***DNA extraction and DNA metabarcoding***
132 DNA was purified from sorted macroinvertebrate samples using a non-destructive DNA
133 extraction (Carew *et al.* 2018a), except soft bodied taxa and hard bodied taxa were placed in
134 separate tubes. Soft bodied taxa were incubated for 2 hrs and hard bodied taxa for 3 hrs at

56°C in 180 µL of T1 DNA extraction buffer and 25 µL of Proteinase K from the Nucleospin tissue kit (Macherey-Nagel Inc. Düren, Germany).

For unsorted samples (macroinvertebrates and debris), a single clean ball-bearing was added to each tube, and samples were homogenised for 2 mins using a TissueLyser II (Qiagen, Germany). Then 400 µL of T1 buffer and 25 µL proteinase K were added from the Nucleospin tissue kit (Macherey-Nagel Inc.). Samples were incubated for 3 hrs at 56°C. After incubation, DNA was extracted from all samples using the Nucleospin tissue kit (Macherey-Nagel Inc. Düren, Germany) following the manufacturer's instructions. Extracted DNA was eluted in 100 µL of TE buffer. Eluted DNA from unsorted samples contained high amounts of contaminants that inhibited PCR, so DNA extractions for each sample were combined and further processed using a stool DNA extraction kit (Promega) to remove PCR inhibitors.

All samples were DNA metabarcoded as two technical PCR replicates (i.e. Elbrecht & Steinke 2018) using a two-step PCR process. DNA metabarcoding included negative controls from first round PCR amplifications; a macroinvertebrate control sample of known composition (to verify continuity of species detection between MiSeq runs); and a sample containing DNA from *Scaptodrosophila xanthorrhoeae* – an invertebrate restricted to north Queensland, Australia and not expected to occur in aquatic invertebrate samples – to check for sample cross-contamination and index switching.

The first round PCRs involved amplifying the 3' end of the mitochondrial cytochrome c oxidase subunit I (COI) DNA barcode region (Hebert *et al.* 2003) using six overlapping PCR primer sets previously used for DNA metabarcoding of macroinvertebrate samples (Table 2, Fig 2). First round PCR primers contained (5'-3') a universal adapter (Illumina Nextera transposase sequence; Illumina Corporation, San Diego, CA, USA) followed by the target-specific sequence. First round PCR reactions contained 2 µL of DNA template (1:10 dilution),

159 8.3 µL molecular biology grade water, 12.5 µL KAPA3G PCR with $MgCl_2$buffer (KAPA
160 Biosystems), 1 µL $MgCl_2$ (25 mM), 0.5 µL forward primer (10 µM), 0.5 µl reverse primer (10
161 µM), and 0.2 µL KAPA3G polymerase (5 U/ml) (KAPA Biosystems) in a total volume of 25 µL,
162 and were amplified using the PCR conditions in Table 2. The PCR amplicons for each of the
163 six primer sets were pooled in a ratio 1:1, except that 30% less of the short amplicon and
164 30% more of the long amplicon were added to allow for length-based biases in amplification
165 in second round PCRs. Pooled amplicons were cleaned using ExoSAP-IT (Applied Biosystems,
166 Waltham, MA) according to the manufacturer's instructions.
167 The cleaned pooled amplicons were then used as templates for second round PCRs.
168 Second round primers contained (5'-3') Illumina p5/p7 adapter sequence, a unique 8bp
169 index sequence and part of the universal adapter sequence (Illumina Corporation, San
170 Diego, CA, USA). Reactions used 3 µL of the pooled first-round amplicons, 12.5 µL MiFi mix
171 (Bioline, London, England), 2.5 µL forward p5 index primer (10 µM), and 2.5 µL reverse p7
172 primer (10 µM). PCR conditions were as follows: 94°C for 5 mins followed by 12 cycles of
173 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, then 1 cycle of 72°C for 5 mins. Amplicons
174 were pooled in equal amounts and the library was gel purified using a PureLink™ Quick Gel
175 Extraction and PCR Purification Combo Kit (Invitrogen, Waltham, MA). An Illumina MiSeq
176 run using a 600-cycle flow cell MiSeq sequencing kit V3 (300bp x 2) (Illumina, San Diego, CA)
177 was performed by the Australian Genome Research Facility Ltd (AGRF) to generate DNA
178 metabarcodes for each sample.
179
180 ***Bioinformatic processing***
181 Read pre-processing and sample demultiplexing was performed by AGRF. The DNA
182 metabarcoding data was then analysed using a custom pipeline. Reads were trimmed to

183 remove primer sequences and sorted into groups corresponding to the different amplicon
184 regions (see Fig 2) using Cutadapt v 1.16 (Martin 2011), standard Unix bash commands and
185 the filter_fasta.py script from QIIME (Caporaso *et al.* 2010). Quality filtering, chimera
186 removal and read clustering were performed with QIIME2 (Bolyen *et al.* 2019) using the
187 DADA2 plugin (Callahan *et al.* 2016). COI reference sequences were obtained from GenBank
188 (https://www.ncbi.nlm.nih.gov/genbank/), BOLD Systems (http://www.boldsystems.org/)
189 and additional unpublished DNA barcodes for freshwater macroinvertebrates from greater
190 Melbourne, Victoria, Australia. A custom reference library was constructed for each of the
191 amplicon regions with *in silico* PCR, using the OBITools package (Boyer *et al.* 2016).
192 Taxonomic classifications were assigned to OTU sequences with QIIME, using the uclust
193 algorithm (Edgar 2010). A step-down assignment procedure was used, in which taxonomic
194 assignment was first performed at a sequence identity threshold of 100%. Sequences that
195 were not assigned to a taxon at this step were passed to the next round of taxonomy
196 assignment, for which an identity threshold of 99% was used. This process was repeated
197 until an identity threshold of 85% was reached. For each OTU sequence, the most detailed
198 taxonomic lineage sharing at least 90% of matching reference sequences was returned,
199 except for the first round of taxonomy assignment, for which a consensus of at least 95%
200 was required. OTU sequences with reference sequence matches with 97-100% identity were
201 identified to species, while those with 95-96% were assigned to genus, 90-94% to family,
202 and 85-89% to order, where possible. A coarser identification was reported for OTU
203 sequences that matched reference sequences from multiple taxa in excess of the consensus
204 cut-offs. Some taxa contained incomplete taxonomic information in databases, so they were
205 identified to the highest taxonomic resolution possible. This included some species which
206 were identified using their BOLD BIN code or an interim code from our private DNA barcode

207 library. Generally, only identifications of >97% can be considered as accurate, but due to
208 'gaps' in reference libraries we used coarser identifications to enable better comparison to
209 morphological identifications.
210
211 ***DNA barcoding***
212 DNA metabarcoding data was checked for taxa that did not contain species level matches in
213 the DNA barcode reference libraries and were identified morphologically. As these taxa
214 were non-destructively DNA extracted for DNA metabarcoding, they could be retrieved from
215 sorted samples and targeted for individual DNA barcoding to improve the number of species
216 level detections. DNA barcoding was conducted using Illumina MiSeq sequencing according
217 to Shokralla et al. (2015). Samples for individual DNA barcoding were included in an Illumina
218 MiSeq run using a 600-cycle flow cell MiSeq sequencing kit V3 (300bp x 2) (Illumina, San
219 Diego, CA) which was performed by the Australian Genome Research Facility Ltd (AGRF).
220
221 **Results**
222 MiSeq sequencing yielded 3,038,516 high quality reads (Supplementary Table 1) with
223 greater than 20,000 reads per replicate for each sample and greater than 92,000 reads per
224 sample. Raw reads from DNA metabarcoding were uploaded to the National Centre for
225 Biotechnology's Short Read Archive (SAMN16274401- SAMN16274432) under BioProject
226 PRJNA665929. When sorting the DNA metabarcodes into separate amplicons, we observed
227 some amplicons with primer sequences other than the six gene-specific PCR primer
228 combinations used in the first round of PCR (Fig 2, Supplementary Table 2). We suspect that
229 this was due to small amounts of residual first round PCR primers being carried over into the
230 second round PCR reactions, in which all the amplicons for a sample were pooled together.

231 For all further analyses, amplicons were therefore categorised into four groups: 'left
232 amplicon' (corresponding to the B/R5 and BF2/BR1 region of the COI barcode), 'long
233 amplicon' (BF2/BR2 region), 'right amplicon' (miCOIintF/dgHCO2198/LepR1 and BF1/BF2
234 region) and 'short amplicon' (BF1/BR1 region).

235 We were able to assign greater than 90% of DNA metabarcodes to species level
236 across the seven unsorted and sorted samples (Fig 3) after individually DNA barcoding an
237 additional 35 specimens from the non-destructively extracted sorted sample
238 (Supplementary Table 3). We found no sequences from the *S. xanthorrhoeae* control sample
239 in the other samples, suggesting that sample cross-contamination and index switching were
240 non-existent or at a very low level. The detection of species in the control sample of known
241 composition was entirely consistent with previous MiSeq runs. We found a small number of
242 reads in the PCR controls. Species found in controls were removed from the final dataset if
243 they were present at a higher level in the PCR control than in a sample. This resulted in the
244 removal of one species - a detection of *Paratya australiensis* in sample CHA13 replicate 1.

245 Technical replicates showed little variability with 91% of taxa (families, genera and
246 species) found in both replicates. There was slightly less overlap between replicates for the
247 individual amplicons, with 86% of detections in both replicates for the left amplicon, 83% for
248 the long amplicon, 90% for the right amplicon and 88% for the short amplicon.

249

250 ***Detection of macroinvertebrates in sorted and unsorted samples***

251 All four amplicons were successful in targeting invertebrate DNA in samples. In the sorted
252 macroinvertebrate samples (i.e. samples containing only freshwater macroinvertebrate
253 specimens and no debris), greater than 97% of DNA metabarcodes were from
254 macroinvertebrates for all amplicons (data not shown).

In the unsorted samples (i.e. macroinvertebrate specimens and debris), this was reduced but the majority of DNA metabarcodes were still from macroinvertebrates (Fig 4). The proportion of DNA metabarcodes from macroinvertebrates was highest in the right and short amplicon datasets, with 85% and 82% of DNA metabarcodes from macroinvertebrates, respectively. The next most commonly detected DNA metabarcodes were from terrestrial invertebrates, particularly for the left and long amplicons where they comprised 16% of DNA metabarcodes. DNA metabarcodes from invertebrates that could not be identified as terrestrial or aquatic represented 4 - 8% of DNA metabarcodes for all amplicons. Microinvertebrates including Branchiopoda, Cyclopoida, Calanoida, Podocopida, and Macrodasyida, were often detected in unsorted samples. For the left and short amplicons, 5% of DNA metabarcodes were from microinvertebrates.

Composition of macroinvertebrate families and species identified in sorted and unsorted samples

Across the seven sites, 70 families and 247 macroinvertebrate species were identified (Supplementary Table 5). This corresponded to a total of 171 independent macroinvertebrate family identifications from the 14 samples (seven sorted and seven unsorted). Macroinvertebrate families identified from samples from the same location substantially overlapped between sample types (sorted or unsorted) and identification method (morphology or DNA metabarcoding) (Fig 5a). Family level identifications overlapped by 56% for all sample types and identification methods. For the sorted samples, family identifications from DNA metabarcodes and morphology overlapped by 96%, showing almost all families identified morphologically were also detected with DNA metabarcoding. Differences in family level identifications between sampling method and sample type could

largely be attributed to the detection in unsorted samples of small or rare taxa, such as Hydridae and Acarina, that were not found in sorted samples. We also found large conspicuous rare taxa, such as Atyidae and some Odonata, that were entirely removed during sorting, were often not detected in unsorted samples. Differences in the detection of taxa between morphological identification and DNA metabarcoding of sorted samples resulted from DNA metabarcoding missing small uncommon taxa or heavily sclerotised taxa, like adult beetles. Many of these taxa were detected in unsorted samples which were homogenised rather than non-destructively extracted like the sorted samples.

When comparing the family-level SIGNAL biotic index scores and the number of EPT taxa (Table 4), we found little difference in the values assigned to the sorted morphological, sorted DNA metabarcoding and unsorted DNA metabarcoding identified subsamples. The SIGNAL scored varied by less than 0.71 and the number of EPT taxa varied by less than two families between sample processing methods at each site.

A total of 455 independent macroinvertebrate species identifications were made from the seven sites. The proportion of species that were found in both the sorted and unsorted DNA metabarcoding samples from the same site was low, with only 36% of species detected in both sample types (Fig 5b). When comparing unsorted and sorted pairs, we typically found more species in unsorted samples compared to their sorted counterparts (Table 1) with five out of the seven sites showing more species present in the unsorted samples. We found 41% of species were only found in unsorted samples and 22% of species were only found in sorted samples (Fig 5b).

Amplification of macroinvertebrate taxa

302 The detection of macroinvertebrate species varied between the four amplicons (Table 5).
303 The right amplicon was the best at detecting macroinvertebrate species with 91.4% of the
304 taxa detected by DNA metabarcoding detected with this amplicon alone. The worst
305 performing amplicon was the long amplicon which only detected 66.9% of species.
306 However, the long amplicon was also represented by fewer DNA metabarcodes than the
307 other amplicons (Supplementary Table 4). Using multiple amplicons improved the detection
308 of macroinvertebrate species, with the right and short amplicon detecting 98.5% of
309 macroinvertebrate species found with DNA metabarcoding.
310 Amplification biases were observed when comparing the number of DNA
311 metabarcodes (or sequence reads) recovered from different macroinvertebrate orders for
312 each amplicon (Fig 6). Macroinvertebrate insects accounted for most of the DNA
313 metabarcodes and were also the most abundant based on enumeration of individuals in
314 sorted samples. For all amplicons, there was a bias towards detection of Ephemeroptera,
315 but this was most extreme for the long amplicon where a third of DNA barcodes recovered
316 were from Ephemeroptera. The right and short amplicons were more effective for detecting
317 Gastropoda from the Hygrophila including the Ancylidae, Lymnaeidae, Physidae and
318 Planorbidae in this study. Species from the Hygrophila were poorly detected with the left
319 and long amplicons. The left amplicon tended to be more biased towards amplifying insects
320 over non insect groups compared to other amplicons.
321
322 ***Density of macroinvertebrates relative to debris***
323 Macroinvertebrates were detected in all density controls (Table 1, Fig 7) including the 'no
324 invertebrates' control (Fig 7a). The density controls showed more species were detected as
325 the density of animals increased with the 'double invertebrates' control detecting 85

species, the standard density control detecting 71 species and the 'no invertebrates' control detecting 55 species (Table 1). The detection of species in density controls overlapped by 35% for all controls and 56% for species shared between the two control samples containing animals (Fig 7a). However, 40% of species were only found in one on the three controls. This was largely driven by the detection of rare taxa (Fig 7b). When rare taxa occurring at <0.01% of reads were excluded, the overlap of taxa detection increased substantially with 67% of taxa detected overlapping between all controls and 74% found in both density controls containing animals.

The macroinvertebrate to debris density ratio varied between the seven unsorted samples (Table 3), but the number of taxa also varied, making comparisons of detection compared to density difficult.

Discussion

Our analyses indicate that DNA metabarcoding of freshwater macroinvertebrates in unsorted net samples offers a viable means of assessing species diversity. Despite macroinvertebrates being mixed with debris, we were able to link the majority of DNA metabarcodes to freshwater macroinvertebrate species. Furthermore, currently-used family level biotic metrics, namely the SIGNAL biotic index and the number of EPT taxa, calculated from the unsorted DNA metabarcoded material, varied little from those calculated from traditional sorted and morphologically identified specimens. This study therefore shows the feasibility of using this approach in macroinvertebrate biodiversity assessments, consistent with previous studies (Macher *et al.* 2018; Majaneva *et al.* 2018; Hajibabaei *et al.* 2019a; Pereira-da-Conceicoa *et al.* 2019), and now validated for sampling where debris issues are particularly problematic.

350 We found that greater than 65% of DNA metabarcodes in unsorted samples for the four
351 amplicon groups analysed were of macroinvertebrate origin. The right amplicon (generated
352 with the miCOIintF/dgHCO2198/LepR1 and BF1/BF2 primer combinations) and short
353 amplicon (generated with the BF1/BR1 primer combination) were the most effective, with
354 88% and 86% of DNA metabarcodes in our unsorted sample dataset corresponding to
355 macroinvertebrates, respectively. The BF1/BR1 primer combination was also found to be
356 effective for isolating macroinvertebrate DNA metabarcodes in unsorted samples by Nichols
357 *et al.* (2019), where they found greater than 95% of DNA metabarcodes were from
358 invertebrates. Moreover, when the DNA metabarcodes were combined from the right and
359 short amplicons, they detected almost all macroinvertebrate species found in the DNA
360 metabarcoding dataset. This suggests that for unsorted samples, combined analyses using
361 these two amplicons offer the best opportunity for detecting macroinvertebrate species
362 diversity. While we could not quantify the effectiveness of individual primer sets, using
363 primer cocktails of the miCOIintF/BF1 primers with dgHCO2198/LepR1/BF2 to generate the
364 right amplicon and miCOIintF/BF1 primers with the BR1/R5 primer to generate the short
365 amplicon would likely replicate the high detection rate of these amplicons. We would
366 recommend mixing these amplicons in a 2: 1 ratio of right to short prior to second round
367 PCRs to compensate for length based amplification and sequencing bias which is commonly
368 found when sequencing amplicons with large length differences (Hajibabaei *et al.* 2019b).
369 Detection of macroinvertebrate DNA metabarcodes was lower using the left amplicon
370 (generated with the B/R5 and BF2/BR1 primer combinations) and long amplicon (generated
371 with the BF2/BR2 primer combination) in the unsorted samples. The left and long amplicons
372 also detected 10% more terrestrial invertebrate DNA, suggesting that they were somewhat
373 less effective for processing unsorted samples where macroinvertebrates are the target.

374 Furthermore, we found more concerning biases with these two amplicons. Both amplicons
375 were poor at detecting Gastropoda from the Hygrophila. This included the families
376 Ancylidae, Lymnaeidae, Physidae and Planorbidae in our study. Species belonging to these
377 families are commonly collected in temperate Australia (Chessman 1995; Chessman *et al.*
378 1997) and include some common introduced species, such as *Physella acuta* and
379 *Pseudosuccinea columella.* The B/R5 and BF2/BR2 primer combinations are favoured in
380 northern hemisphere studies, where Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa
381 are key indicator groups (Macher *et al.* 2018; Hajibabaei *et al.* 2019a; Leese *et al.* 2020).
382 However, these taxa were less prevalent in our samples where they represented only 45%
383 of the total number of individuals in samples and 25% of the species diversity. In the greater
384 Melbourne area (Victoria, Australia), Diptera, particularly from the Chironomidae, account
385 for the greatest species diversity in pool or edge samples (e.g.Carew *et al.* 2018b). Hence,
386 testing primer combinations on locally collected samples remains important for selecting
387 the optimal primer sets for assessing region-specific biodiversity (Blackman *et al.* 2019;
388 Hajibabaei *et al.* 2019b).
389 We developed a simple processing method based on sieving samples in the laboratory to
390 reduce sample volume and increase the proportion of macroinvertebrates. We then sub-
391 sampled the material for DNA metabarcoding. We found it took less than 15 minutes per
392 sample to sieve and prepare material for drying in the oven. This was less time than
393 methods that sort macroinvertebrates from debris, that can take 30-45 minutes for a live
394 field pick or 60 minutes or more for a lab sort (Nichols & Norris 2006). While our method
395 was based in part on Majaneva *et al.* (2018) and Macher *et al.* (2018), we were restricted by
396 the equipment available to homogenise samples. Rather than homogenise the entire
397 sample and subsample homogenised material for DNA extraction, we subsampled the

material prior to homogenisation. While processing material in this way would be unlikely to affect the detection of common or abundant taxa, it is expected to be less effective for detection of rare taxa. Rare taxa are more likely to be missed if animals remain whole. However, Pereira-da-Conceicoa *et al.* (2019) have since shown that a domestic hand blender could offer a low cost alternative to providing initial homogenisation of unsorted samples.

Despite differences in sample processing, patterns of species detection between sorted unsorted samples were similar to that observed in other studies (Elbrecht *et al.* 2017b; Macher *et al.* 2018; Majaneva *et al.* 2018; Nichols *et al.* 2019; Pereira-da-Conceicoa *et al.* 2019). This included high overlap in the families identified using morphological versus DNA methods with sorted samples (i.e. Elbrecht *et al.* 2017a), and an overall higher macroinvertebrate species diversity found in unsorted samples (i.e. Pereira-da-Conceicoa *et al.* 2019), but this was not universal across all of our samples. We could attribute some differences in detection to the live pick or lab sort method, in that there was no opportunity for DNA fragmentation or leaching into the preservation ethanol for species entirely removed during the live pick, such as large and highly visible Atyidae and Odonata. We also observed that the non-destructive DNA extraction method used to isolate DNA from sorted macroinvertebrate samples caused some taxa to be missed by DNA barcoding, such as adult Hydraenidae beetles in the Charlie's Creek sample. Non-destructive DNA extraction can be less reliable for detecting heavily sclerotised taxa (Carew *et al.* 2018a).

Unlike using sorted macroinvertebrate samples as a source of DNA for metabarcoding, with unsorted samples there is less control over the composition and density of macroinvertebrate DNA, which can impact the detection of species (Elbrecht *et al.* 2017a). Sorting macroinvertebrates increases their concentration within a DNA extraction sample and enables subsampling to be adjusted to allow more time or ensure a greater proportion

422 of a sample is sorted if the environmental sample contains a low density of
423 macroinvertebrates or high biodiversity (Walsh 1997; Nichols & Norris 2006). Our study
424 showed that density may be a factor in the detection of macroinvertebrate diversity from
425 unsorted material. Ideally, incremental subsampling of unsorted material from a variety of
426 sites with differing density and biodiversity would be performed to determine a level of
427 subsampling that captures representative biodiversity across all sites. Furthermore, storing
428 unsorted material can facilitate the leaching of DNA from animals in samples (Nichols *et al.*
429 2019) and may lessen the impact of density, which explains why we found many
430 macroinvertebrate DNA metabarcodes in our 'no invertebrate' control. However,
431 subsampling of unsorted material would need to be balanced against the cost of processing
432 larger amounts of material, the time required for assessments and the need to detect rare
433 taxa.

434 Processing unsorted pool or edge samples also leads to sampling of non-target
435 invertebrates, such as terrestrial invertebrates and microinvertebrates. While this
436 information may provide insight into non-target invertebrate taxa present at sites and has
437 the potential to broaden the taxa used in bioassessment, it does highlight the need for
438 comprehensive macroinvertebrate DNA barcode reference libraries. These libraries are
439 important for determining which invertebrate DNA metabarcodes in unsorted samples are
440 from freshwater or terrestrial origin (Weigand *et al.* 2019).

441 We estimated the cost of processing unsorted macroinvertebrate samples to species
442 level using DNA metabarcoding to be ~AU$200 per sample, while the cost of sorting and
443 morphologically identifying macroinvertebrate samples to family level only ranges from
444 AU$200-AU$600 depending on the sorting method used. Furthermore, Marshall et al.
445 (2006) found using morphological species identification cost six times more than family

identification. Therefore, DNA metabarcoding of unsorted samples can substantially reduce the cost of processing (based on using 10% of the sample dry weight for DNA metabarcoding) and provides greater opportunity for routine and cost-effective species level identifications in biological monitoring programs. This increased taxonomic resolution can give important insights into species-based distributions, biodiversity and responses to stressors or management interventions. However, the cost of processing unsorted samples would increase, due to the time and cost of including additional DNA extractions, if higher volumes of unsorted material were required for assessing biodiversity.

DNA metabarcoding of unsorted aquatic macroinvertebrate edge or pool samples offers a means of rapidly assessing species diversity with the benefit of reducing the time and expense in processing and identifying macroinvertebrate specimens. However, complete DNA barcode reference libraries are needed to ensure that the taxa found are freshwater macroinvertebrates. Furthermore, primer selection should be carefully considered so that macroinvertebrate taxa are best targeted and detected. This is best achieved with multiple amplicons. The impact of diversity and density in samples should be further investigated to ensure that macroinvertebrate diversity between sites is described in a standardised way. Additional research should enable standardised protocols to be further developed for use on unsorted net edge or pool samples to assess macroinvertebrate biodiversity and for bioassessment**.**

Acknowledgements

The authors would like to thank Claudette Kellar, Zac Billingham, Genevieve Hehir and Vin Pettigrove for providing macroinvertebrate samples for this study and Eddie Tsyrlin for

photographs he supplied for Figure 1. This study was funded primarily by the Australian Research Council through a Linkage grant (LP150100876) and Fellowship (FL100100066), with additional support from Melbourne Water Corporation.

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**. https://doi.org/10.3389/fmars.2016.00096

Blackman R, Mächler E, Altermatt F*, et al.* (2019) Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. *Metabarcoding and Metagenomics* **3**. https://doi.org/10.3897/mbmg.3.34735

Bolyen E, Rideout JR, Dillon MR*, et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* **37**, 852-857. https://doi.org/10.1038/s41587-019-0209-9

Boyer F, Mercier C, Bonin A*, et al.* (2016) OBITools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* **16**, 176-182. https://doi.org/10.1111/1755-0998.12428

Bush A, Compson ZG, Monk WA*, et al.* (2019) Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution* **7**, 434. https://doi.org/10.3389/fevo.2019.00434

Callahan BJ, McMurdie PJ, Rosen MJ*, et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581-583. https://doi.org/10.1038/nmeth.3869

Caporaso JG, Kuczynski J, Stombaugh J*, et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335-336. https://doi.org/10.1038/nmeth.f.303

Carew ME, Coleman RA, Hoffmann AA (2018a) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ* **6**, e4980. https://doi.org/10.7717/peerj.4980

Carew ME, Kellar CR, Pettigrove VJ, Hoffmann AA (2018b) Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological Indicators* **85**, 440-450. https://doi.org/10.1016/j.ecolind.2017.11.002

Chessman BC (1995) Rapid assessment of rivers using macroinvertebrates: A procedure based on habitat-specific sampling, family level identification and a biotic index. *Austral Ecology* **20**, 122-129. https://doi.org/10.1111/j.1442-9993.1995.tb00526.x

Chessman BC (2003) New sensitivity grades for Australian river macroinvertebrates. *Marine and Freshwater Research* **54**, 95-103. https://doi.org/10.1071/MF02114

Chessman BC, Growns JE, Kotlash AR (1997) Objective derivation of macroinvertebrate family sensitivity grade numbers for the SIGNAL biotic index: Application to the Hunter River system, New South Wales. *Marine and Freshwater Research* **48**, 159-172. https://doi.org/10.1071/MF96058

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461. https://doi.org/10.1093/bioinformatics/btq461

514 Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for
515 freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science* **5**.
516 https://doi.org/10.3389/fenvs.2017.00011
517 Elbrecht V, Peinert B, Leese F (2017a) Sorting things out: Assessing effects of unequal
518 specimen biomass on DNA metabarcoding. *Ecology and Evolution* **7**, 6918-6926.
519 https://doi.org/10.1002/ece3.3192
520 Elbrecht V, Steinke D (2018) Scaling up DNA metabarcoding for freshwater
521 macrozoobenthos monitoring. *Freshwater Biology* **64**, 380-387.
522 https://doi.org/10.1111/fwb.13220
523 Elbrecht V, Vamos EE, Meissner K*, et al.* (2017b) Assessing strengths and weaknesses of DNA
524 metabarcoding-based macroinvertebrate identification for routine stream
525 monitoring. *Methods in Ecology and Evolution* **8**, 1265-1275.
526 https://doi.org/10.1111/2041-210x.12789
527 Emilson CE, Thompson DG, Venier LA*, et al.* (2017) DNA metabarcoding and morphological
528 macroinvertebrate metrics reveal the same changes in boreal watersheds across an
529 environmental gradient. *Scientific Reports* **7**, 12777. https://doi.org/10.1038/s41598-
530 017-13157-x
531 Hajibabaei M, Porter TM, Robinson CV*, et al.* (2019a) Watered-down biodiversity? A
532 comparison of metabarcoding results from DNA extracted from matched water and
533 bulk tissue biomonitoring samples. *PLoS ONE* **14**, e0225409.
534 https://doi.org/10.1371/journal.pone.0225409
535 Hajibabaei M, Porter TM, Wright M, Rudar J (2019b) COI metabarcoding primer choice
536 affects richness and recovery of indicator taxa in freshwater systems. *PLoS ONE* **14**,
537 e0220953. https://doi.org/10.1371/journal.pone.0220953
538 Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental Barcoding: A
539 Next-Generation Sequencing Approach for Biomonitoring Applications Using River
540 Benthos. *PLoS ONE* **6**, e17497. https://doi.org/10.1371/journal.pone.0017497
541 Hajibabaei M, Spall J, Shokralla S, van Konynenburg S (2012) Assessing biodiversity of a
542 freshwater benthic macroinvertebrate community through non-destructive
543 environmental barcoding of DNA from preservative ethanol. *BMC Ecology* **12**, 28.
544 https://doi.org/10.1186/1472-6785-12-28
545 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identification through DNA
546 barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*
547 **270**, 313-321. https://doi.org/10.1098/rspb.2002.2218
548 Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA
549 barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes*
550 *fulgerator*. *Proceedings of the National Academy of Sciences of the United States of*
551 *America* **101**, 14812-14817. https://doi.org/10.1073/pnas.0406166101
552 Leese F, Sander M, Buchner D*, et al.* (2020) Improved freshwater macroinvertebrate
553 detection from eDNA through minimized non-target amplification. *bioRxiv*,
554 2020.2004.2027.063545. https://doi.org/10.1101/2020.04.27.063545
555 Lenat DR (1988) Water quality assessment of streams using a qualitative collection method
556 for benthic macroinvertebrates. *Journal of the North American Benthological Society*
557 **7**, 222-233. https://doi.org/10.2307/1467422
558 Leray M, Yang J, Meyer C*, et al.* (2013) A new versatile primer set targeting a short fragment
559 of the mitochondrial COI region for metabarcoding metazoan diversity: Application

for characterizing coral reef fish gut contents. *Frontiers in Zoology* **10**, 34. https://doi.org/10.1186/1742-9994-10-34

Macher JN, Vivancos A, Piggott JJ*, et al.* (2018) Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resourses* **18**, 1456-1468. https://doi.org/10.1111/1755-0998.12940

Majaneva M, Diserud OH, Eagle SHC, Hajibabaei M, Ekrem T (2018) Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding and Metagenomics* **2**, e26664. https://doi.org/10.3897/mbmg.2.26664

Marshall JC, Steward AL, Harch BD (2006) Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: The benefits and costs of using species abundance data. *Hydrobiologia* **572**, 171-194. https://doi.org/10.1007/s10750-005-9007-0

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**. https://doi.org/10.14806/ej.17.1.200

Nichols SJ, Kefford BJ, Campbell CD*, et al.* (2019) 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**, 607-620. https://doi.org/10.1111/fwb.13443

Nichols SJ, Norris RH (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. https://doi.org/10.1007/s10750-006-0253-6

Pereira-da-Conceicoa L, Elbrecht V, Hall A*, et al.* (2019) Metabarcoding unsorted kick-samples facilitates macroinvertebrate-based biomonitoring with increased taxonomic resolution, while outperforming environmental DNA. *bioRxiv*, 792333. https://doi.org/10.1101/792333

Porter TM, Hajibabaei M (2018) Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology* **27**, 313-338. https://doi.org/10.1111/mec.14478

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

Shokralla S, Porter TM, Gibson JF*, et al.* (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports* **5**, 9687. https://doi.org/10.1038/srep09687

Walsh CJ (1997) A multivariate method for determining optimal subsample size in the analysis of macroinvertebrate samples. *Marine and Freshwater Research* **48**, 241-248. https://doi.org/https://doi.org/10.1071/MF96096

Weigand H, Beermann AJ, Ciampor F*, et al.* (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. https://doi.org/10.1016/j.scitotenv.2019.04.247

Yu DW, Ji YQ, Emerson BC*, et al.* (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* **3**, 613-623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

606 Walsh CJ (1997) A multivariate method for determining optimal subsample size in the
607 analysis of macroinvertebrate samples. *Marine and Freshwater Research* **48**, 241-
608 248. https://doi.org/https://doi.org/10.1071/MF96096
609 Weigand H, Beermann AJ, Ciampor F*, et al.* (2019) DNA barcode reference libraries for the
610 monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future
611 work. *Science of the Total Environment* **678**, 499-524.
612 https://doi.org/10.1016/j.scitotenv.2019.04.247
613 Yu DW, Ji YQ, Emerson BC*, et al.* (2012) Biodiversity soup: Metabarcoding of arthropods for
614 rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*
615 **3**, 613-623. https://doi.org/10.1111/j.2041-210X.2012.00198.x
616
617

Data Acessability

DNA sequences from DNA metabarcoding of field samples: NCBI SRA accession no. SAMN16274401 - SAMN16274432 under NCBI BioProject: PRJNA665929. DNA sequences from DNA barcoding: GenBank accession no. MW051365-MW051403.

Author Contribution

MEC performed the molecular lab work, carried out the data analysis, participated in the study design and drafted the manuscript. KLR designed the bioinformatic pipeline used to conduct the analysis. RAC assisted in sourcing samples, participated in the study design and helped to draft the manuscript. AAH refined the study design and helped to draft the manuscript. All authors read and approved the final manuscript.

629 **Table 1.** Site location, collection date and the number of freshwater macroinvertebrate families and species in samples in this study.

Location						Number of macroinvertebrate families (species)		
						Sorted		Unsorted residue
Site code	Site	Longitude	Latitude	Collection method	Collection date	Morphologically identified	DNA metabarcoding	DNA metabarcoding
BAR11	Barringo Creek	144.61606	-37.42256	Lab sort	2-Jul-18	20	17 (37)	14 (35)
BOY134	Boyd Creek	144.89422	-37.38911	Lab sort	3-Jul-18	22	20 (46)	19 (39)
CHA13	Charlie's Creek	144.71938	-37.42815	Lab sort	2-Jul-18	22	20 (42)	20 (50)
DPW226	Deep Creek site 1	144.77790	-37.28618	Lab sort	3-Jul-18	10	10 (19)	9 (20)
DPW673	Deep Creek site 2	144.80049	-37.63069	Lab sort	3-Jul-18	23	18 (36)	24 (55)
LY13	Little Yarra River site 1	145.60467	-37.77680	Live pick	30-Apr-18	25	24 (45)	26 (75)
LY1616	Little Yarra River site 2	145.62040	-37.78060	Live pick	24-Apr-18	21	19 (42)	25 (76)

630

631 **Table 2.** Primer pairs, PCR conditions and amplicon sequence lengths (without template
632 primers included) for the primer combinations used in this study.

Primer pair	Amplicon sequence length (bps)	PCR conditions	References
B/ R5	316	94°C 3 min, followed by 35 cycles of 94°C 40 sec, 46°C 60 sec, 72°C 30 sec, followed by 72°C 5 min	(Hajibabaei *et al.* 2012)
miCOIintF/ dgHCO2198/ LepR1	313	94°C 3 min, followed by 5 cycles of 94°C 30 sec, 45°C 40 sec, 72°C 60 sec followed by 30 cycles of 94°C 30 sec, 51°C 40 sec, 72°C 60 sec, followed by 72°C 5 min	(Hebert *et al.* 2004; Leray *et al.* 2013)
BF1/ BR1 BF1/ BR2 BF2/ BR1 BF2/ BR2	217 313 322 421	94°C 3 min, followed by 35 cycles of 94°C 40 sec, 50°C 30 sec, 72°C 60 sec, followed by 72°C 5 min	(Elbrecht & Leese 2017)

633

634 **Table 3.** Composition and density of invertebrates in unsorted samples.

	Visual observations		Total dry weight of sieved sample	DNA extraction			Dry weight density (grams)		
Unsorted sample code	**Relative macroinvertebrate density**	**Debris**		**weight used (grams)**	**% of sample**	**tubes required**	**Debris**	**macroinvertebrates**	**density ratio**
BAR11	medium	Herbaceous plant and leaf debris	17.7	1.8	10	5	0.402	0.053	0.132
BOY134	high	Herbaceous plant debris	1.7	1.0	60	1	0.247	0.016	0.065
CHA13	medium	Herbaceous plant debris	23.0	2.3	10	6	0.892	0.002	0.002
DPW226	low density	Herbaceous plant debris, leaves and stones	19.4	1.9	10	6	0.404	0.026	0.064
DPW673	High, many copepods	Herbaceous and woody plant debris, sand and stones	23.3	2.3	10	6	0.353	0.078	0.221
LY13	medium-low	Herbaceous and woody plant debris, sand and stones	38.0	3.8	10	5	1.883	0.02	0.011
LY1616	medium-low	Herbaceous and woody plant debris, sand and stones	6.47	1.3	10*	5	0.206	0.004	0.019
LY1616 -double invertebrates	medium	Herbaceous and woody plant debris, sand and stones	4.74	1.4	10*	4	0.301	0.014	0.047
LY1616 -no invertebrates	No animals	Herbaceous and woody plant debris, sand and stones	4.61	1.4	10*	3			

635 *****As a percentage of the total sample weight

636

637

638 **Table 4.** SIGNAL biotic index scores and number of EPT taxa calculated from families identified using morphology and DNA
639 metabarcoding for sorted and unsorted samples from the seven study sites.

	SIGNAL biotic index			Number of EPT taxa		
	sorted		bulk residue	sorted		bulk residue
Site code	morphologically identified	DNA metabarcoding	DNA metabarcoding	morphologically identified	DNA metabarcoding	DNA metabarcoding
BAR11	4.65	4.53	4.79	5	5	4
BOY134	3.73	3.95	3.79	7	7	5
CHA13	3.82	3.55	4.10	5	6	5
DPW226	3.40	3.40	4.11	1	1	1
DPW673	3.67	3.94	3.21	6	5	7
LY13	4.85	4.75	4.96	12	10	11
LY1616	5.14	5.05	5.08	9	8	9

640

641 **Table 5.** Percentage of macroinvertebrate species detected with each amplicon and paired
642 combinations of amplicons using DNA metabarcoding.
643

		Amplicon			
		Left	**Long**	**Right**	**Short**
Amplicon	**Left**	80.8			
	Long	86.8	66.9		
	Right	96.2	93.2	91.4	
	Short	94.4	92.1	98.5	88.0

Figure 1. Workflow for collecting and processing samples used in this evaluation. Blue boxes refer to the work plan and sampling design while green boxes refer to the DNA metabarcoding component. Sampling and sorting components that cover 2/3 of total costs (taken from Elbrecht *et al.* (2017b)) are in the red dotted box.

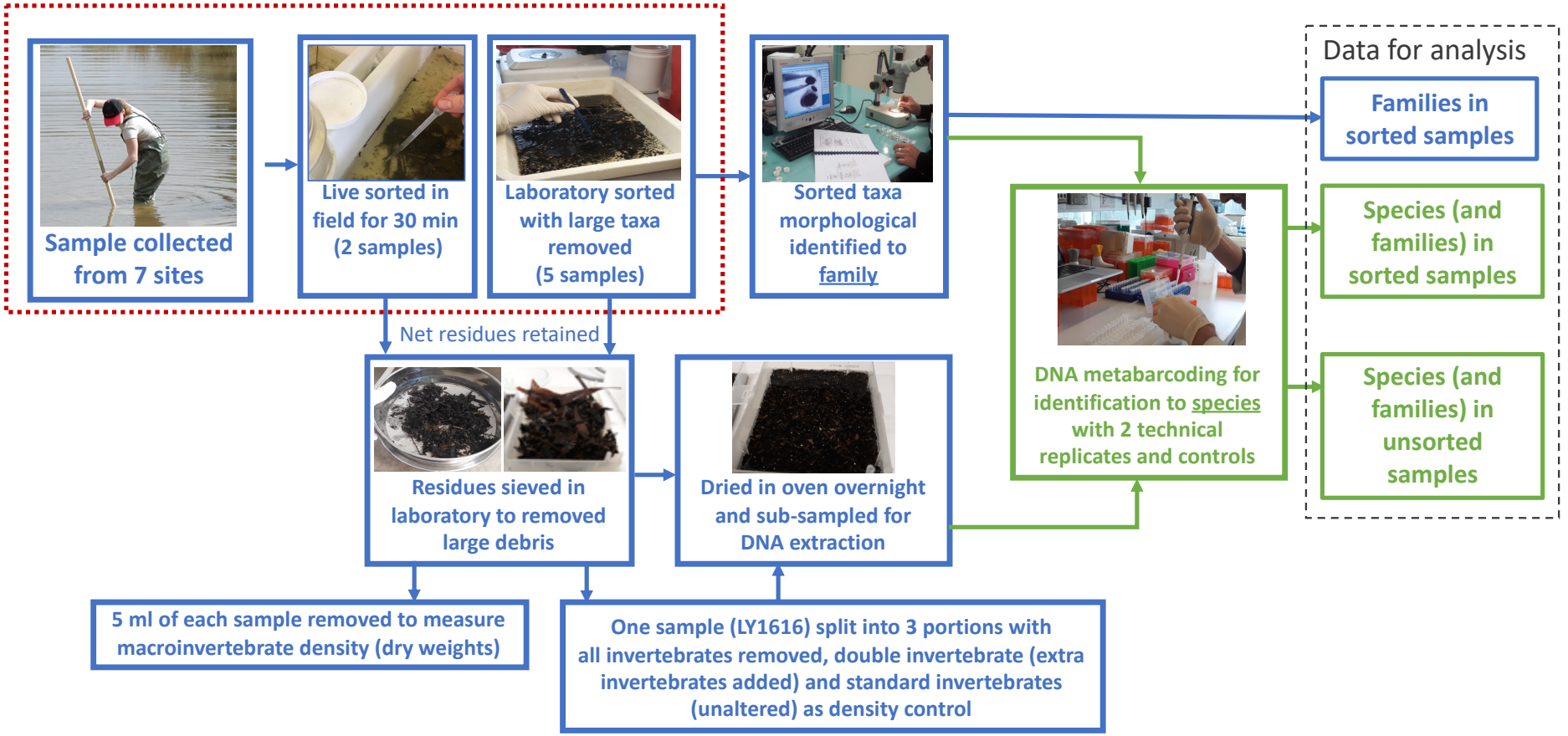


Figure 2. Overview of primer combinations and amplicons generated from sorted macroinvertebrate samples and unsorted net contents (PCR amplification box). Primer combinations and amplicons are indicated by colours. Some amplicons were combined for analysis as indicated amplicon analysis box. Amplicon lengths are given from insects.

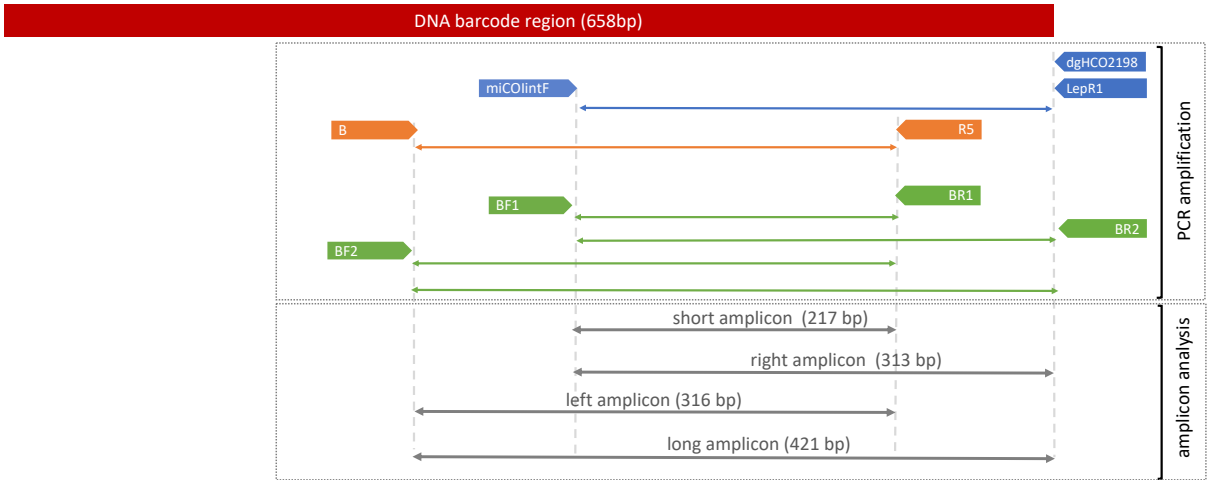


Figure 3. The proportion of DNA metabarcodes identified to the taxonomic level of species, genus, family, order and class and, sequences that could not be assigned a taxonomic rank (unassigned) for unsorted and sorted samples. Note: not all species identifications have species names attached but are linked to voucher specimens with individual DNA barcodes (see Supplementary Table 4).

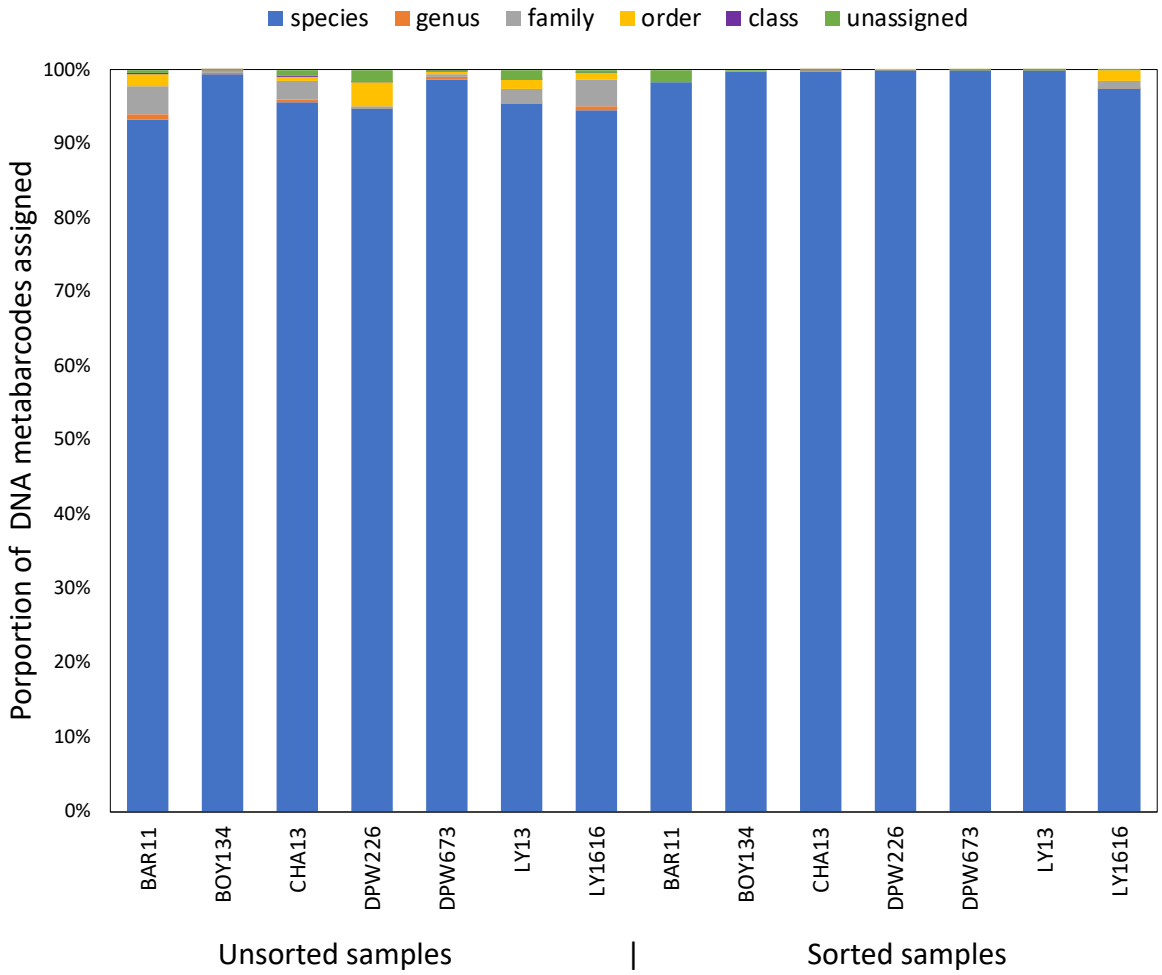


Figure 4. Proportion of DNA metabarcodes assigned to major taxonomic groups in unsorted samples containing sampling debris for the left, long, right and short amplicons. Percentages less than 1% are not shown.

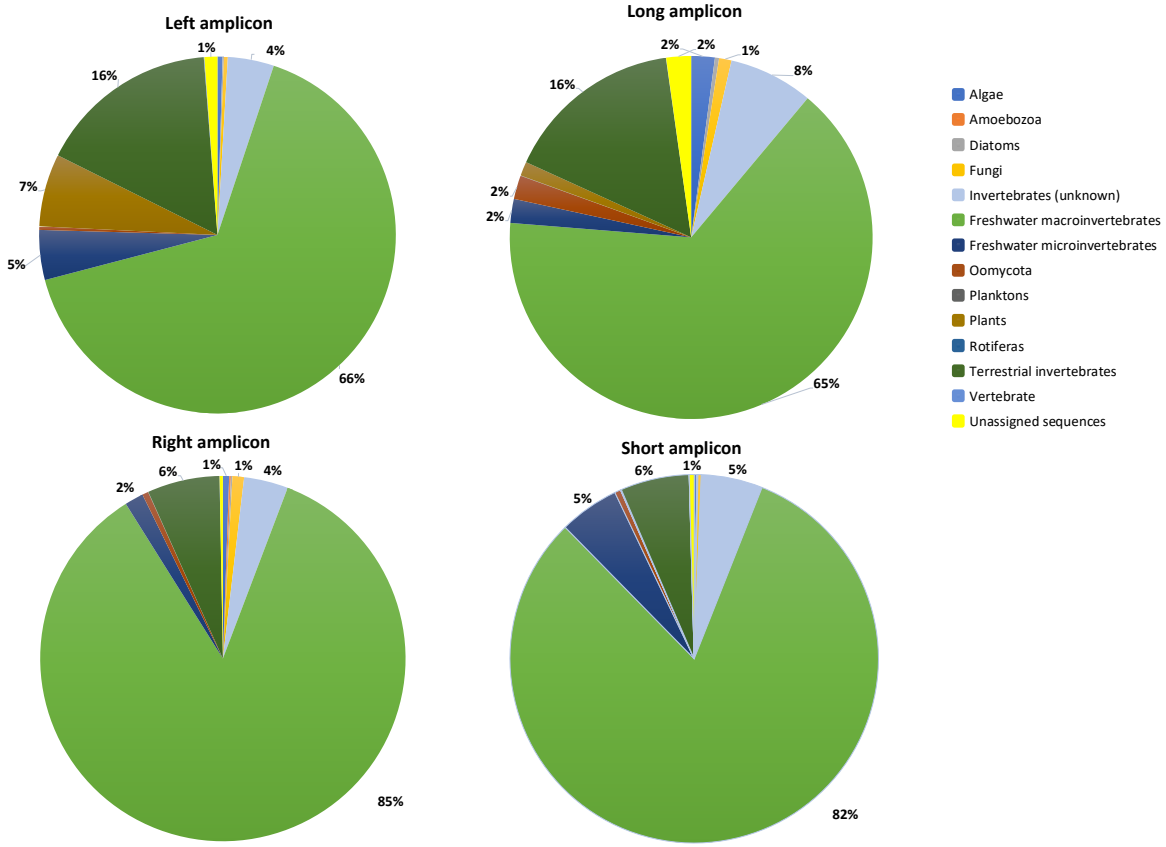
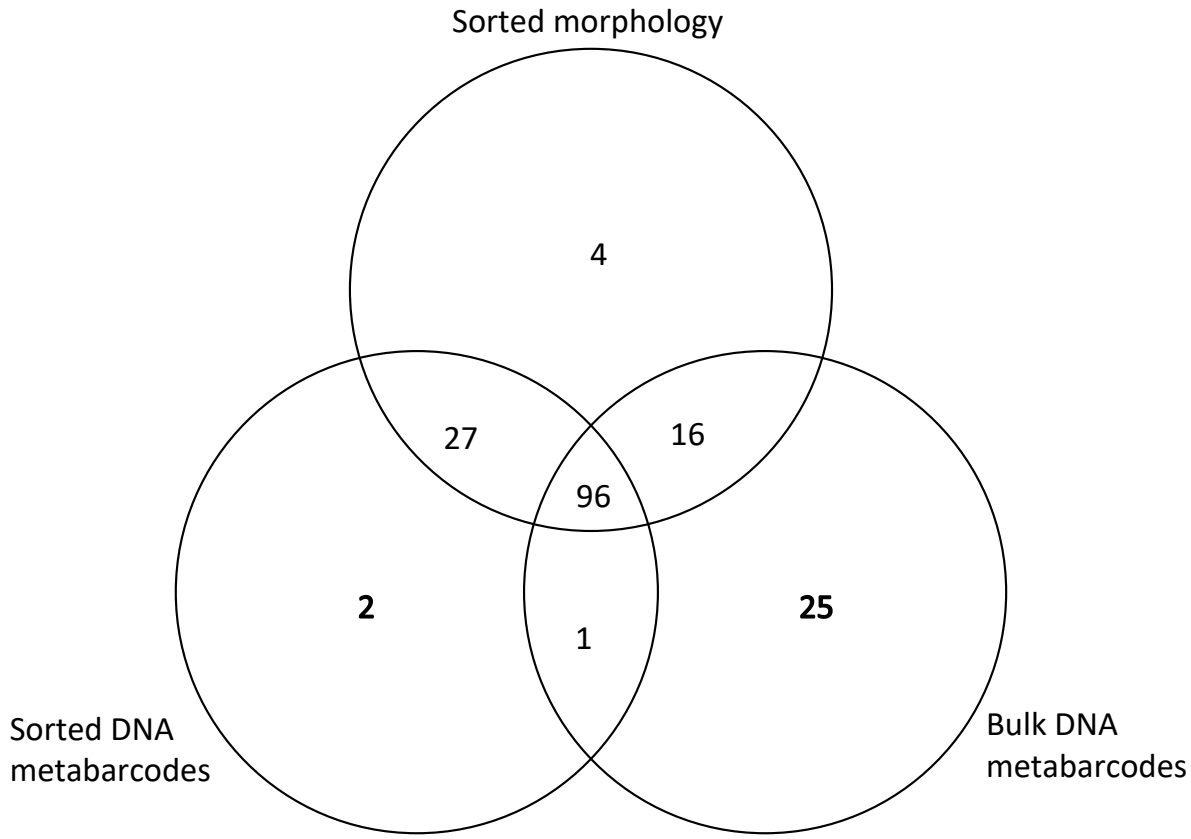


Figure 5. Venn diagrams showing a) the number of macroinvertebrate families identified from sorted samples using morphological examination and DNA metabarcodes, and unsorted samples using DNA metabarcodes; and b) the number of species detected using DNA metabarcoding in sorted and unsorted samples.

a)



b)

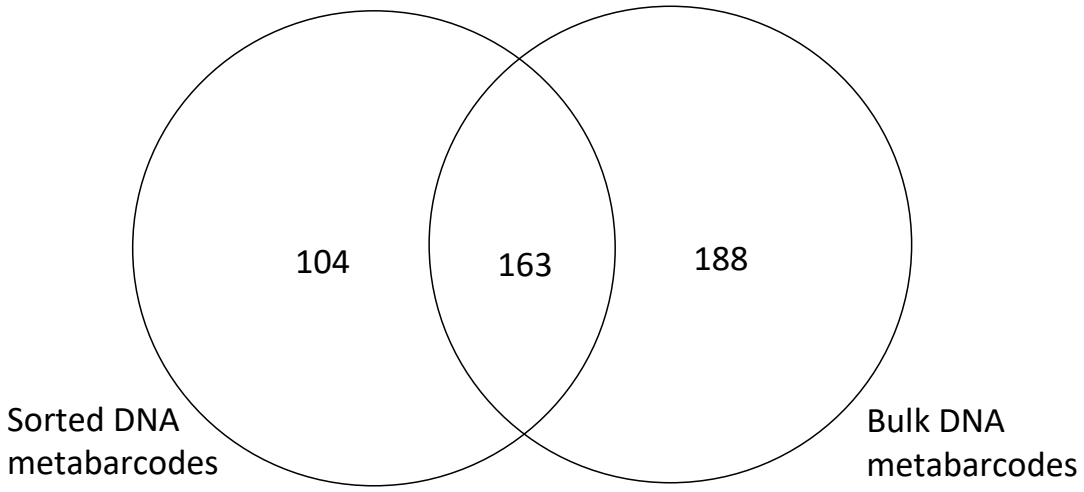


Figure 6. The proportion of DNA metabarcodes (sequence reads) assigned to macroinvertebrate orders and classes for each amplicon (Left, Long, Right and Short). The abundance of individuals from the different groups is also shown.

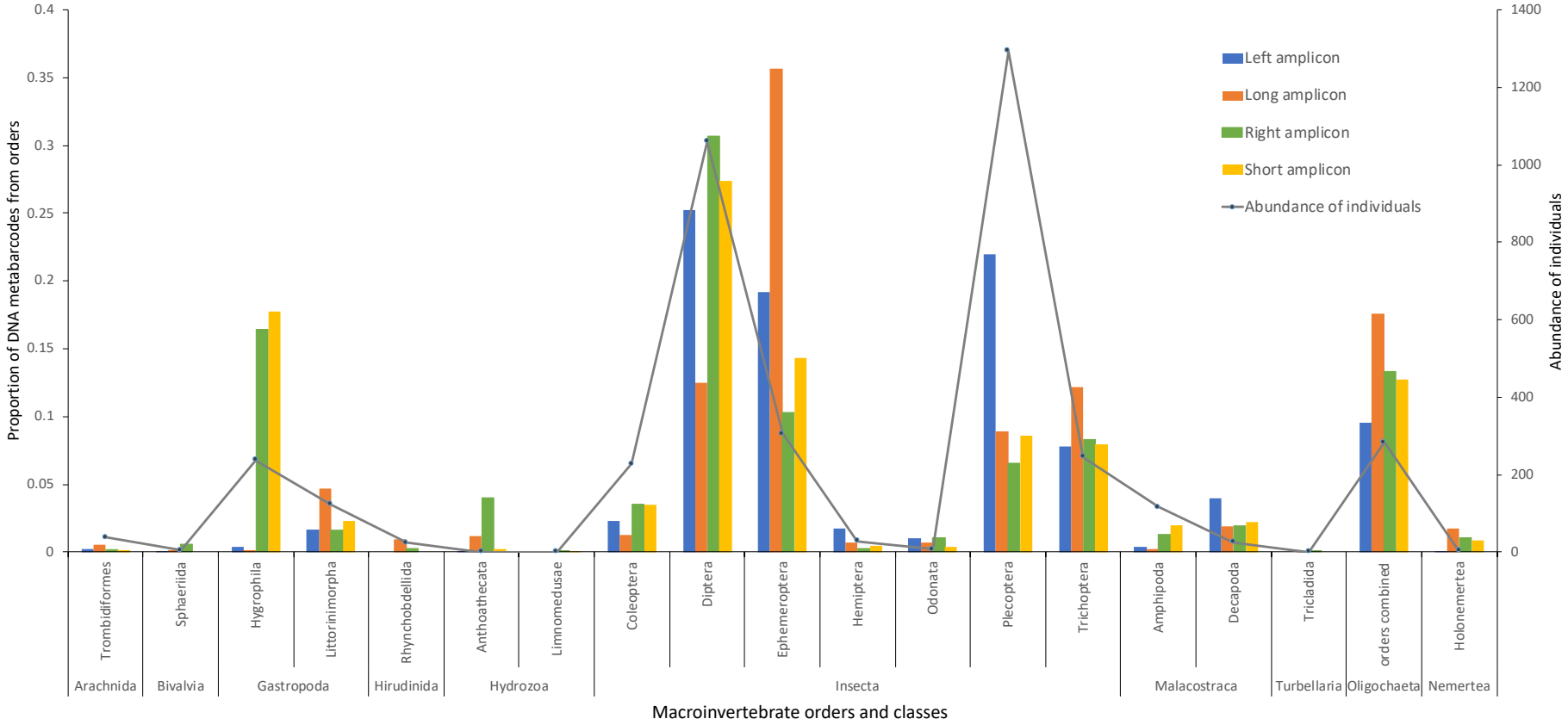
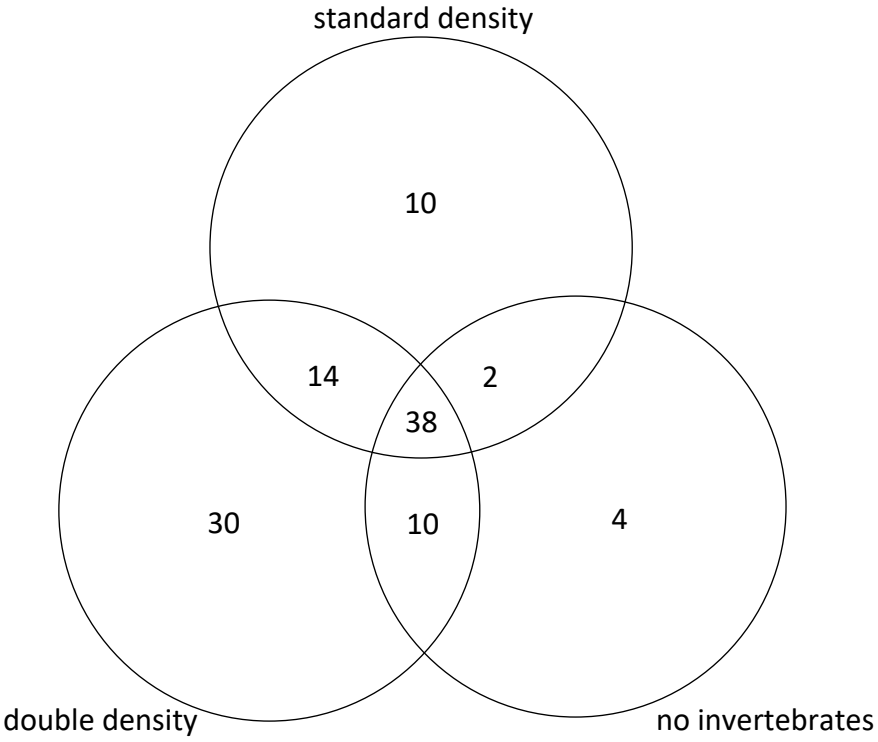


Figure 7. Detection of species in three way partitioned LY1616 unsorted sample density control samples. Density control samples include an unaltered sample (standard density); a sample with all visible invertebrates removed (no invertebrates); and a sample where all animals removed from the no invertebrate sample were added to increase the density of macroinvertebrates in the sample (double density). a) all detections, b) detections with taxa present at <0.01% in a sample removed.

a)



b)