A peer-reviewed version of this preprint was published in PeerJ on 7 October 2019.

View the peer-reviewed version (peerj.com/articles/7745), which is the preferred citable publication unless you specifically need to cite this preprint.

Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7:e7745
https://doi.org/10.7717/peerj.7745

1 **Title: Validation of COI metabarcoding primers for terrestrial arthropods**

2

3 **Running Title: Primers for arthropod metabarcoding**

4 **Word count:** ____

5 **Authors:** Vasco Elbrecht[1*], Thomas WA Braukmann[1], Natalia V Ivanova[1,2], Sean WJ Prosser[1],
6 Mehrdad Hajibabaei[1,2], Michael Wright[1,2], Evgeny V Zakharov[1,2], Paul DN Hebert[1,2], Dirk Steinke[1,2]

7

8 Affiliations:

9 1) Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G
10 2W1, Canada

11 2) Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario,
12 N1G 2W1, Canada

13

14 *Corresponding author: Vasco Elbrecht (elbrecht@uoguelph.ca),

15

16 **Abstract:**

17 Metabarcoding can rapidly determine the species composition of bulk samples and thus aids
18 ecosystem assessment. However, it is essential to use primer sets that minimize amplification bias
19 among taxa to maximize species recovery. Despite this fact, the performance of primer sets employed
20 for metabarcoding terrestrial arthropods has not been sufficiently evaluated. Thus this study tests the
21 performance of 36 primer sets on a mock community containing 374 species. Amplification success
22 was assessed with gradient PCRs and the 21 most promising primer sets selected for metabarcoding.
23 These 21 primer sets where also tested by metabarcoding a Malaise trap sample. We identified eight
24 primer sets, mainly those including inosine and/or high degeneracy, that recovered more than 95% of
25 the species in the mock community. Results from the Malaise trap sample were congruent with the

26 mock community, but primer sets generating short amplicons produced potential false positives. Taxon
27 recovery from the 21 amplicon pools of the mock community and Malaise trap sample were used to
28 select four primer sets for metabarcoding evaluation at different annealing temperatures (40-60 C^{o})
29 using the mock community. Temperature did only have a minor effect on taxa recovery that varied with
30 the specific primer pair.

31 This study reveals the weak performance of some primer sets employed in past studies. It also
32 demonstrated that certain primer sets can recover most taxa in a diverse species assemblage. Thus there
33 is no need to employ several primer sets targeting the same amplicon. While we identified several
34 suited primer sets for arthropod metabarcoding, the primer selection depends on the targeted taxonomic
35 groups, as well as DNA quality, desired taxonomic resolution, and sequencing platform employed for
36 analysis.

37

38 **Key words:** DNA metabarcoding, primer bias, degeneracy, insects, biodiversity.

39

40

41 ## Introduction

42 Over the past decade, two methodological and technological advances have made it possible to
43 address the urgent need for the capacity to undertake large-scale surveys of biodiversity (Vörösmarty *et*
44 *al.* 2010; Dirzo *et al.* 2014; Steffen *et al.* 2015). First, the emergence of DNA barcoding which uses
45 sequence variation in short, standardized gene regions (i.e. DNA barcodes) to discriminate species, has
46 made it possible to quickly and reliably characterize species diversity (Hebert *et al.* 2003). Second,
47 high-throughput sequencers (HTS) permit the inexpensive acquisition of millions of sequence records
48 (Reuter *et al.* 2015). The coupling of HTS with DNA barcoding, commonly known as metabarcoding,
49 allows for characterization of biodiversity at unprecedented scales (Creer *et al.* 2016) as shown by
50 studies on terrestrial (Gibson *et al.* 2014; Beng *et al.* 2016), freshwater (Hajibabaei *et al.* 2011; Carew
51 *et al.* 2013; Andújar *et al.* 2017), and marine (Leray & Knowlton 2015) ecosystems.

52 Metabarcoding studies on bulk collections of animals usually targets a 658 bp region of the
53 cytochrome *c* oxidase subunit I (COI) (Folmer *et al.* 1994; Andújar *et al.* 2018). This gene region has
54 gained broad adoption because of the rapidly expanding reference database (Ratnasingham & Hebert
55 2007; Porter & Hajibabaei 2018b) and its good taxonomic resolution (Meusnier *et al.* 2008). Ribosomal
56 markers have been suggested as an alternative (Deagle *et al.* 2014; Marquina *et al.* 2018) because their
57 slower rate of evolution results in more conserved motifs/regions aiding the design of universal primer
58 sets. However, reference databases for ribosomal markers are very limited for most taxonomic groups
59 (Clarke *et al.* 2014) and ribosomal primer sets show no substantial improvement in taxon recovery over
60 well-designed COI primer sets (Elbrecht *et al.* 2016; Clarke *et al.* 2017; Elbrecht & Leese 2017;
61 Krehenwinkel *et al.* 2017).

62 An important consideration for metabarcoding studies is the primer combination used for
63 amplification of the target fragment. It is critical that primer sets optimally match the template

64 sequences of the target species. Mismatches between primer and template is skewing read abundance
65 and lead to a substantial bias in taxon detection (Piñol *et al.* 2014; Elbrecht & Leese 2015). Failure to
66 minimize amplification bias reduces the amount of taxa detected in a sample (Elbrecht & Leese 2017).
67 Furthermore, insufficient sequencing depth and/or low DNA concentration can introduce stochastic
68 effects that additionally bias taxon recovery (Barnes & Turner 2015; Leray & Knowlton 2017).
69 The effectiveness of primer sets can be evaluated by *in vitro* tests with mock communities
70 (Elbrecht & Leese 2015; Brandon-Mong *et al.* 2015; Leray & Knowlton 2017) or by *in silico* tests
71 (Clarke *et al.* 2014; Elbrecht & Leese 2016; Piñol *et al.* 2018; Bylemans *et al.* 2018b; Marquina *et al.*
72 2018). The failure to evaluate primers can seriously compromise data quality. For instance, a primer set
73 (Zeale *et al.* 2011) often employed for analyzing the gut contents of insect predators (see references in
74 (Jusino *et al.* 2018) lacks degeneracy, leading to poor taxon recovery (Brandon-Mong *et al.* 2015). The
75 use of multiple primer sets or even multiple marker genes was proposed to improve taxon recovery
76 (Alberdi *et al.* 2017; Zhang *et al.* 2018). This approach may be optimal for samples of very
77 phylogenetically divergent groups such as protists (Pawlowski *et al.* 2012) or marine benthic
78 communities (Cowart *et al.* 2015; Wangensteen *et al.* 2018; Drummond 2018). However, given the
79 increased cost and time (Bohmann *et al.* 2018; Zhang *et al.* 2018), the use of multiple primer sets is
80 unnecessary for taxonomic groups with limited diversity. We hypothesize that in the case of terrestrial
81 arthropods a single well-designed primer set can be sufficiently effective, and the use multiple primer
82 sets is not necessary.
83 This study compares the performance of commonly used and newly developed primer sets on
84 the recovery of species in a bulk DNA extract from 374 insect species (Braukmann *et al.* 2018) and
85 from a Malaise trap sample. Based on a hierarchical testing scheme (Figure 1) using gradient PCRs and
86 assessing species recovery with metabarcoding, we selected four primer pairs whose metabarcoding
87 performance was tested on a range of annealing temperatures.
88

89 **Material and Methods**

90 **Tested samples and experimental outline**

91 We used two samples to test a range of primer sets for metabarcoding: a mock community of
92 374 species (Braukmann *et al.* 2019) and a sample collected with a Malaise trap (Figure 1). The mock
93 community is comprised of 374 species (Figure 1A), each specimen represented by a individual BIN
94 (taxonomic breakdown shown in Figure S1A, (Ratnasingham & Hebert 2013)). A detailed list of
95 specimens and their Barcode of Life Datasystems process IDs (BOLD, Ratnasingham & Hebert 2007)
96 is given in Table S1. For most specimens, the full 658 bp barcode region was available through BOLD,
97 but we completed reads for three taxa with shorter sequences by extracting haplotypes from our
98 metabarcoding data using a denoising approach (Elbrecht *et al.* 2018b). The resulting reference library
99 is available as a fasta file (See Scripts S1 for the fasta file). To compare mock community results with a
100 field sample, we collected insects with a Townes-style Malaise trap (Bugdorm, Taiwan) deployed in a
101 grassland/forest area near Waterloo, Ontario, Canada (43°29'30.8"N 80°36'59.6"W). We selected a
102 single weekly sample (June 30 - July 7, 2018) and dried it for three days in a disposable grinding
103 chamber. The sample was ground to fine powder using an IKA Tube Mill control (IKA, Breisgau,
104 Germany) at 25,000 rpm for 2 x 3 minutes. DNA was extracted from 21 mg of ground tissue using the
105 DNeasy Blood & Tissue kit (Qiagen, Venlo, Netherlands).

106 These mock community DNA extracts were used to test 36 primer pairs by comparing amplification
107 success across a range of annealing temperatures. Twenty-one primers pairs whose amplicon
108 concentrations plateaued in amplicon concentration at lower annealing temperatures were selected for
109 metabarcoding both the mock community and the Malaise trap sample. Four representative primer sets
110 showing high success in species recovery were selected to determine the optimal annealing temperature
111 for maximizing species recovery from bulk samples (see Figure 1).

112

113 **Gradient PCRs**

114 Thirty-six primer combinations commonly used for metabarcoding were selected for gradient
115 PCR tests (Figure 2). Some of these primers represent new combinations, as well as new variants of
116 primers by shifting the primer binding site by 3 bp, by incorporating degeneracy, or by replacing
117 inosine "I" with "N" and vice versa. PrimerMiner v0.18 was used to generate an alignment
118 visualization (Elbrecht & Leese 2016) using reference sequences for 31 arthropod orders downloaded
119 and aligned as part of an earlier study ((Vamos *et al.* 2017). The plot of the full alignment with binding
120 sites for all primers used in this study is available in Figure S2.

121 Mock community gradient PCRs for 36 primer combinations were run on an Eppendorf
122 Mastercycler pro (Hamburg, Germany). PCRs were set up with 2× Multiplex PCR Master Mix Plus
123 (Qiagen, Hilden, Germany), 0.5 µM of each primer (IDT, Skokie, Illinois), 12.5 ng DNA, and
124 molecular grade water (HyPure, GE, Utha, USA) for a total volume of 25 µL. One positive control and
125 one negative control using the BF2 + BR2 primer set (Elbrecht & Leese 2017) were included with each
126 primer set.

127 The following thermocycling protocol was used: initial denaturation at 95°C for 5 min then 29
128 cycles of denaturation at 95°C for 30 s followed by a gradient of annealing temperatures from 44.5 –
129 64.5 °C for 30 s with extension at 72°C for 50 s, and a final extensions of 5 min at 72°C. PCR success
130 and fragment length were determined by visualizing amplicons on a 1% agarose gel. Amplicon
131 concentration was quantified without prior cleanup using a High Sensitivity dsDNA Kit on a Qubit
132 fluorometer (Thermo Fisher Scientific, MA, USA).

133

134 **Primer selection for metabarcoding**

135 Based on the results of gradient PCR (Figure 1B), we selected 21 primer sets for metabarcoding
136 that showed strong, consistent amplification and reached plateau in amplicon concentration at lower
137 annealing temperatures (Figure 1C, Figure S8). A few primer sets generated amplicons at 46°C, but

138 were excluded because they failed to reach an asymptote in concentration at lower annealing
139 temperatures.
140

141 **Metabarcoding (mock community and malaise trap)**

142 21 primer sets for both the mock community and the Malaise trap sample where selected for
143 DNA metabarcoding and Illumina MiSeq sequencing. We employed a fusion primer based two-step
144 PCR protocol that amplifies target fragments in the first step and attaches in-line tags and Illumina
145 TruSeq library sequence tails during the second PCR (Elbrecht & Steinke 2018). We used in-line tags
146 of different length and sequenced half the samples reverse orientation as well as amplicons shifted
147 against each other to ensure sufficient sequence diversity for sequencing (Elbrecht & Leese 2015). The
148 7 bp tags with different insert lengths were randomly generated using R scripts (Elbrecht & Steinke
149 2018), but were subsequently manually edited to maximize the Levenshtein distance between tags
150 (Figure S3). Figure S4 shows the fusion primer sequences used for library preparation. For the first
151 PCR step, we used the same protocol as for the gradient PCR, but used a fixed annealing temperature
152 of 46°C and 24 cycles of amplification. One negative control containing the BF2 + BR2 primer
153 combination and one containing no primers were included in the PCR (see Table S2 for primer list).
154

155 1 µL of the PCR product generated by each primer set was used as template for the second PCR
156 step (with no quantification or reaction cleanup) under similar PCR conditions except the extension
157 time was increased to 2 minutes while the number of cycles was reduced to 14. PCR products were
158 cleaned using SPRIselect (Beckman Coulter, CA, USA) with a sample to volume ratio of 0.76x. DNA
159 concentration was quantified using a Qubit fluorometer, High Sensitivity dsDNA Kit (Thermo Fisher
160 Scientific, MA, USA). Subsequently, individual libraries were equimolar pooled following adjustment
161 for amplicon length (Table S1). The mock community library was sequenced on an Illumina MiSeq
162 with 300 bp paired end sequencing (v3 chemistry) with a 5% PhiX spike in. Amplicons for the Malaise

163 sample were generated with half the DNA amount (6.25 ng) and 29 cycles for the first PCR step.
164 Individual libraries were pooled equimolar, but we factored in the preference for shorter reads by
165 Illumina sequencing using the mock community sequencing results (Figure S9, Table S1). The Malaise
166 sample was also sequenced on a Illumina MiSeq with 300 bp paired end sequencing (v3 chemistry)
167 with a 5% PhiX spike in.
168

169 **Bioinformatic processing**

170 Quality control of raw sequence data was done with FastQC v0.11.7 and multiQC v1.4 (Ewels
171 *et al.* 2016). Sequence data were first demultiplexed and processed with the R wrapper script JAMP
172 v0.68 (https://github.com/VascoElbrecht/JAMP). Reads were paired-end merged using Usearch
173 v11.0.667 (Edgar 2010), allowing for more relaxed settings with respect to mismatches between reads
174 (fastq_maxdiffs = 99, fastq_pctid = 75). Primer sequences were subsequently trimmed using cutadapt
175 v1.18 with default settings (Martin 2011). Reads deviating by more than 10 bp from the expected
176 amplicon length were discarded. Usearch (Edgar & Flyvbjerg 2015) was used to remove reads with an
177 expected error probability of 1 or higher, and to dereplicate and map reads against the 374 reference
178 sequences of the mock community (usearch_global with minimum 97% identity). Resulting tables were
179 automatically summarized into a hit table of all samples using the function map2ref implemented in
180 JAMP. The hit table was subsampled using a custom R script (Scripts S1) to determine the number of
181 taxa detected at different sequencing depths. Figure 3 overviews the processing steps and all scripts are
182 available in Scripts S1.

183 Data for the Malaise sample was processed using the same pipeline but mapped against a
184 reference database consisting of public sequence records for arthropods found in Ontario (downloaded
185 from BOLD December 2018). Gaps and terminal Ns were removed from all sequences. Sequences
186 outside the length range of 648-668 bp were discarded (Scripts S1). Reads were mapped against this
187 reference database using map2ref, but singletons in each sample were discarded and mapping required

188 a 99% match and maxaccepts=0, maxrejects=0, to reduce the number of false positives. Reads
189 matching to the same Barcode Index Number (BIN, Ratnasignham & Hebert 2013) were collapsed and
190 reads that matched reference sequences that lacked a BIN assignment were merged based on taxonomy,
191 and combined into 11 MOTUs.
192
193 **Gradient metabarcoding**
194 Out of the 21 metabarcoded primer sets, we selected four primer sets that recovered most of the
195 mock community (ArF5 + Fol-degen-rev, BF3 + BR2, mlCOIintF + Fol-degen-rev and fwhF2 +
196 fwhR2n, Figure 1D) to evaluate the impact of nine annealing temperatures (40.0, 41.6, 43.7, 46.0, 48.5,
197 50.8, 53.0, 54.7 and 56.0 °C) on taxon recovery. Temperatures below 46°C were specifically chosen to
198 explore the impact of non-specific amplification. Other than running the first and second PCR step as
199 gradient PCRs, all laboratory conditions and bioinformatic steps were identical to the prior mock
200 community metabarcoding run. For tagging samples in the second PCR step, additional fusion primers
201 were developed (Fig S5) and checked for sufficient Levenshtein distance (Fig S6, (Elbrecht & Steinke
202 2018)). Individual samples were equimolar pooled, and the library sequenced using an Illumina MiSeq
203 with 300 bp paired end sequencing (v3 chemistry) and a 5% PhiX spike in. Bioinformatic analysis was
204 identical to the previous mock community MiSeq run at 46 °C annealing temperature.
205
206 **Statistical analysis**
207 For statistical analysis R v3.5.0 was used - all scripts to generate figures are available in Scripts
208 S1. The relative abundance of reads per taxon (above 0.001%) for each of the 21 primer sets (Table S1)
209 tested with the mock community was analysed using a Principal Component Analysis implemented in
210 the R package FactoMineR v1.34. The same data was used to visualize the similarity between
211 communities recovered with each primer set, using the R package vegan v2.5-2. A dendrogram was
212 generated using both Jaccard similarity and Bray–Curtis dissimilarity.

213 **Results**

214 **Gradient PCR results and primer set selection**

215 All primer sets generated amplicons with the expected length (Fig. S7) although a few
216 amplicons showed faint secondary bands after gradient PCR. Amplicon concentrations reached an
217 asymptote for 21 of the 36 primer sets (58%) at < 50°C and they were selected for sequencing (Fig.
218 S8). While some other primers showed clear bands in the agarose gel (Fig. S7), they were excluded
219 from sequencing because of their limited annealing temperature range.

220 Amplification success for newly designed primers was mixed (Fig. S7 & Fig. S8). A more
221 degenerate version of ZBJ−ArtF1c + ZBJ−ArtR2c had decreased amplification efficiency. Substituting
222 N for inosine led to increased amplification efficiency for BF1i+BR1i, while replacing inosine with N
223 reduced amplification efficiency for Bn+En. The binding site of the BF2 primer was shifted 3 bp
224 forward (BF3) to reduce slippage effects (Elbrecht *et al.* 2018a). In combination with the BR2 primer
225 set, both versions showed similar amplification efficiency.

226

227 **Metabarcoding and bioinformatic processing**

228 Sequencing of the mock community tested with 21 primer sets on MiSeq (300 bp PE) produced
229 24,348,000 reads of good quality (Q30=<85.8% of reads). Raw sequence data is available on NCBI
230 SRA via accession number SRX4908948. Sequencing depth was negatively correlated with amplicon
231 length (Fig. S9, linear regression, $p < 0.0001$), with at least 0.28 million sequences per sample. The
232 number of discarded sequences after data processing varied among primer sets (Fig. 3A); on average
233 80.61% (SD = 9.84%) of the reads were mapped to the 374 reference sequences. For the primer sets
234 MZplankF2 + C_LepFolR, BF3 + BR2, BF2 + BR2 and AncientLepF3 + C_LepFolR more than 3% of
235 the amplicons deviated by more than 10 bp from the expected amplicon length (Fig. 3, Fig. S10).
236 Primer combinations involving mlCOIintF, BF1, BF2 and fwhF2 showed length variation of 1-2 bp

base pairs (Fig. S10). Additionally, an average of 12.03% (SD = 8.07%) of all reads were discarded through expected error quality filtering (min ee = 1, Fig. 3A). In particular, longer amplicons with little or no overlap in paired end sequencing were affected (Fig. 3B). Raw read data mapped against reference sequences is depicted in Table S1.

The Malaise sample yielded 16,629,020 reads of good quality (Q30=<92.59% of reads). Raw sequence data is available on NCBI SRA via accession number SRX5175597. Sequencing depth was positively correlated with amplicon length (linear regression, p = 0.0004, Fig S9), but there was a reduced length bias in comparison to the mock community sequencing run (Fig. S9).

Primer performance and BIN / species recovery with metabarcoding

Recovery of the mock community was high for most primer sets with an average of 91% of the 374 species recovered (SD = 0.64%, subsampling to 100,000 reads, Figure 4A). With decreasing sequencing depth, recovery diminished, as shown by rarefaction curves (Fig. S11). The primer sets ZBJ−ArtF1c + ZBJ−ArtR2c, LepF1 + MLepF1−Rev and LCO1490 + Ill_C_R showed poor species recovery in comparison with the other primers. Interestingly, rarefaction analysis showed no strict relationship between recovery and primer degeneracy. For example, LCO1490 + HCO2198 had no degenerate sites but had good recovery (90% of taxa). However, primers that lacked degeneracy often had low amplification success and detected fewer species than primer sets with degeneracy (Fig. S12). The primer combinations fwhF2 + fwhR2n, BF2/BF3 + BR2, ArF5 + Fol-degen-rev and mlCOIintF showed the best performance with similar recovery rates (recovery =< 95% of the community, Fig.4, Fig. S12). Taxa recovery was consistent across orders, except for Hymenoptera which where often recovered with lower read counts (Fig. S1A). A Principal Component Analysis (PCA) of relative taxon recovery shows that primer combinations with similar taxon recovery tended to cluster together (Fig. S13), although only 29.36% of variability can be explained by both components. Jaccard similarity and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27801v2 | CC BY 4.0 Open Access | rec: 15 Jul 2019, publ: 15 Jul 2019

261 Bray-Curtis based dendograms (Fig. S14) illustrate that recovery is generally similar among primers,
262 but that combinations with poor species recovery tend to cluster together.
263
264 Sequencing of the Malaise sample confirmed the strong performance of some primer sets, but
265 others showed lower species recovery (Figure 4B). As the species composition of the Malaise sample
266 was unknown, BIN counts at different sequencing depths were used to estimate taxon recovery for all
267 sets of primers. Heat maps for both the Malaise sample (Fig. S18) and the mock community (Fig. S12)
268 were generally congruent but short amplicons from the Malaise sample detected more taxa present in
269 very low abundance. This trend was also reflected in the number of taxa detected with each primer set
270 (Figure 4B) because longer amplicons such as Ill_B_F + HCO2198, AcientLepF3 + C_LepFolR or
271 LCO1490 + HCO2198 exhibited lower taxon recovery than shorter fragments. Most primers that
272 performed well for the mock community also did so for the Malaise sample (highlighted in green in
273 Figure 4B), except the ArF5 + Fol−degen−rev primer set. These patterns were consistent with varying
274 sequencing depths with no asymptote reached in the rarefaction analysis (Fig. S19). Additionally, the
275 rarefaction analysis shows a greater range in the number of taxa detected with different primer sets than
276 for the mock community (Fig S11). Detection across orders was very consistent for primer sets that
277 show good taxa recovery, while especially Hymenoptera and Hemiptera were underrepresented with
278 primer sets recovering fewer taxa (Fig S1B).
279
280 **Gradient PCR metabarcoding**
281 When the performance of the four good performing primer pairs was analyzed at nine annealing
282 temperatures, 23,770,810 sequences (NCBI SRA; ID: SRX4908947) were obtained with good read
283 quality (Q30 =< 82.9% of reads). Sequence coverage averaged 0.58 million (SD = 0.1 million) per
284 sample with a lowest value of 0.38 million reads. Results at 46°C were very similar to the prior
285 metabarcoding run with abundance differences mostly affecting low abundant OTUs (Figure S15,

286 linear regression adj. $R^2 > 0.97$). Changes in annealing temperature from 40 - 56°C only had minor
287 effects on species recovery (Figure 5). In particular, two primer sets (mlCOIintF + Fol−degen−rev;
288 fwhF2 + fwhR2n) showed little variation in species recovery across the range of annealing
289 temperatures. By comparison, recovery rates decreased at temperatures above ~53°C for both BF3 +
290 BR2 and ArF5 + Fol−degen−rev (Figure 5, Figure S16). Length variation in amplicons as a result of
291 primer slippage was not temperature dependent, but the BF3+BR2 primer set generated more short
292 non-target amplicons at lower temperatures (over 1/4 of sequences, Figure S17).
293

294 **Discussion**

295 Using a mock community, we tested a total of 36 different primer combinations, 21 of which
296 were selected for a more detailed metabarcoding analysis. While we did not run replicates for most
297 primer sets, results at 46 °C for gradient metabarcoding and the mock community run were similar.
298 This result is consistent with previous studies which indicated that replicates typically produce similar
299 results (Elbrecht *et al.* 2017; Braukmann *et al.* 2018), particularly when the variation of low abundant
300 OTUs (i.e. < 0.001 %) introduced by stochastic effects is ignored (Leray & Knowlton 2017).
301 Consequently, for metabarcoding of bulk samples, replication should be done at the sampling level
302 (Hurlbert 1984) rather than using DNA extracts or replicate PCRs. While technical replicates do
303 increase confidence in experimental outcomes (Zepeda-Mendoza *et al.* 2016; Elbrecht & Steinke 2018;
304 Macher & Weigand 2018), they deliver limited information given the substantial increase in cost and
305 laboratory workload. If the detection of rare taxa is important for a project, an increase in sequencing
306 depth (Smith & Peay 2014; Braukmann *et al.* 2018) and use of a tagging system resistant to tag
307 switching (e.g. fusion primers (Elbrecht *et al.* 2017)) is a good alternative to replication. Even with the
308 shallow sequencing depth (100,000 reads) used in this study, most primer sets recovered a majority of
309 the taxa in the mock community. This was not necessarily the case for the Malaise trap sample (Figure

310 S19) which is more diverse than the mock community tested (Steinke et al. In Prep). However, the
311 comparison of taxon recovery at different sequencing depth by the tested primer sets allowed for good
312 benchmarking, without capturing the full community. We were also able to characterize the positive
313 bias of the Illumina MiSeq towards shorter fragments (Figure S9), which can be off set by adjusted
314 amplicon concentrations when running fragments of different length in the same run (Fig. S9).
315
316 *Primer performance*
317 As several primer sets recovered most of the taxa in the mock community in similar proportions
318 (Figure 4A), our study has identified several suitable primer sets for metabarcoding terrestrial
319 arthropods communities. The exact choice of primer set will depend on the context of a study, required
320 amplicon length and desired taxonomic resolution (Meusnier *et al.* 2008; Porter & Hajibabaei 2018a).
321 For instance, the fwhF2 + fwhR2n primer set produces a 205 bp amplicon that is ideal when targeting
322 degraded DNA in eDNA or gut contents (Bylemans *et al.* 2018a). The BF1 + BR2 and all three
323 mlCOIintF-based primer sets generate slightly longer fragments (316/313 bp), but they are prone to
324 slippage (Elbrecht *et al.* 2018a) which can cause problems with sequence denoising (Callahan *et al.*
325 2017) during data analysis. We overcame this problem for the longer BF2+BR2 fragment (421 bp) by
326 moving the BF2 primer 3 bp forward (BF3). The BF3 + BR2 combination as well as the ArF5 +
327 Fol−degen−rev primer set represent good choices for long (>400 bp) COI fragment amplification. The
328 ArF5 + Fol−degen−rev primer set appears to be less affected by non-specific amplification at lower
329 annealing temperatures than the BF3 + BR2 primer pair. Although these longer fragments improve
330 taxonomic resolution, they show less overlap in Illumina paired end sequencing leading to more reads
331 being excluded during quality filtering (Figure 3).
332 We observed an increase in rare taxa detected with short amplicons in the Malaise sample, but
333 these are likely false positives due to the decreased taxonomic resolution of shorter amplicons
334 (Meusnier *et al.* 2008; Porter & Hajibabaei 2018a). Even though the ZBJ−ArtF1c + ZBJ−ArtR2c

335 primer set detected over 700 taxa in the Malaise sample, a value comparable to other well performing
336 primer pairs, it failed to detect abundant BINs that most of the other primer pairs recovered (Fig. S18).
337 Well performing primers showed no bias against specific orders, while less suitable primers did
338 struggle with detection of Hymenoptera and Hemiptera. The decreased detection of Hymenoptera in the
339 mock community can likely be attributed to the lysis protocol used for DNA extraction from the insect
340 abdomens (Braukmann *et al.* 2019). This was not the case for the malaise sample, where the bulk
341 sample was ground to a fine powder, making the tissue more accessible to the lysis buffer.
342
343 *Primer design*
344 Primer sets with differing degeneracy, inosine inclusion, and differences in the primer binding
345 region showed variable taxonomic recovery making it difficult to establish clear predictors for primer
346 performance. While degeneracy generally improves the universality of a primer (Krehenwinkel *et al.*
347 2017), some highly degenerate primers performed poorly in our tests (Figure 4). Additionally, even if a
348 primer set shows good taxon recovery, it can still be susceptible to dimerization, to non-specific
349 amplification, or to primer slippage (Elbrecht *et al.* 2018a) (Figure S10). Until these complexities are
350 difficult to predicted *in silico*, it is important to validate metabarcoding primer sets *in vivo* using taxa
351 and samples from the targeted ecosystems. For example, the BF2+BR2 primer set generated non-
352 specific amplicons (often bacterial), which can become a serious complication for eDNA studies where
353 target DNA is scarce (Macher *et al.* 2018; Hajibabaei *et al.* 2019b). High primer degeneracy will likely
354 increase primer universality but decrease specificity. This is less problematic when metabarcoding
355 DNA extracts from bulk specimen samples where target DNA predominates, but can be different for
356 environmental DNA samples.
357 The present study did not reveal if the use of inosine can reduce problems created by high primer
358 degeneracy. Some primers modified with inosine performed well, but others did not. The same was true
359 for highly degenerate primers. However, we did show that for the fusion primer system (Elbrecht &

360 Steinke 2018), primers employed in the second PCR step can be designed with "N" instead of inosine
361 (Figure S5). This substantially reduces costs when large fusion primer quantities are needed for reliably
362 tagging and sample multiplexing. Primer performance could be further improved by adding degeneracy
363 and / or using inosine, but performance will suffer if too much degeneracy is added. Despite careful
364 primer design following best practices (Abd-Elsalam 2003), primer performance can still vary in its
365 suitability for the primer binding site. A primer that works well on paper, might still not work *in vivo*
366 and we strongly recommend testing primers with a mock community or field sample.

367

368 *Annealing temperature*

369 While primer choice is critical for metabarcoding projects, PCR is also biased by the polymerase
370 used (Nichols *et al.* 2018), cycle number (Vierna *et al.* 2017; Krehenwinkel *et al.* 2017), GC content
371 (Braukmann *et al.* 2019), inhibitors (Demeke & Jenkins 2009; Sellers *et al.* 2018), and annealing
372 temperature (Aylagas *et al.* 2016; Clarke *et al.* 2017; Krehenwinkel *et al.* 2018). It is generally assumed
373 that primers bind better at lower annealing temperatures as potential mismatches between template and
374 primer have less influence. While touchdown PCR does not improve species recovery (Clarke *et al.*
375 2017), lower annealing temperatures slightly increase it (Aylagas *et al.* 2016). Although it seems
376 intuitive that lower annealing temperatures lead to better taxonomic recovery, previous studies
377 explored only a limited temperature range never going below 46°C, likely due to the increased risk of
378 non-specific amplification. We studied four representative primer pairs at 9 different annealing
379 temperatures across a wider range (gradient PCR from 40 - 56°C) and were unable to find a universal
380 effect of annealing temperature. BF3 + BR2, mlCOIintF + Fol−degen−rev and fwhF2 + fwhR2n primer
381 sets are largely unaffected by changes in annealing temperature. On the other hand, recovery peaks at
382 48.5 °C for the ArF5 + Fol−degen−rev primer set. For all four primer pairs, annealing temperatures
383 between 46 - 50 °C are probably good choices for metabarcoding. However, this highly depends on
384 melting temperature (T_m). It is advisable to test newly designed metabarcoding primer across an

385 annealing temperature gradient. However, given that of most tested primers did perform similarly well
386 at temperatures usually used for metabarcoding, sequencing gradient PCRs might not always be
387 necessary. Running the four primer pairs at temperatures below 46 °C did not substantially increase
388 taxa recovery, while for some primers it also increased the risk of dimer amplification and occurrence
389 of non-target DNA.
390
391 *No need for multiple primer sets*
392 Eight primer combinations (Figure 4, highlighted in green) each detected 95% or more of the taxa
393 present in the mock community, and most of them could therefore be suitable choices for
394 metabarcoding studies targeting terrestrial arthropods. Of these, seven showed very good performance
395 with the malaise trap sample. This is in stark contrast to earlier studies (Alberdi *et al.* 2017; Zhang *et*
396 *al.* 2018) recommending the use of multiple primer sets to increase coverage. This discrepancy can be
397 explained by primer choice, because (Zhang *et al.* 2018) used LCO1490 and HCO2198 primers which
398 lack degeneracy, and (Alberdi *et al.* 2017) worked with gut content samples, thus replicates might be
399 substantially affected by stochastic effects resulting from low DNA yield. Additionally, the primers
400 used (ZBJ−ArtF1c + ZBJ−ArtR2c) by Alberdi et al. (2017) performed poorly in our study. This
401 particular primer combination (Zeale *et al.* 2011) is widely used for metabarcoding studies (Jusino *et*
402 *al.* 2018) but our results show substantial amplification bias, confirming the low taxon recovery
403 observed before for this primer pair (Brandon-Mong *et al.* 2015). An alternative primer pair to analyze
404 gut content from predators consuming insects could be the pair fwhF2 + fwhR2n because it shows
405 better taxonomic recovery.
406 The use of COI primer sets with limited or no degeneracy such as in (Zhang *et al.* 2018; Jusino *et*
407 *al.* 2018) is not recommended. In general, careful primer design and validation (ideally using mock
408 communities) cannot be replaced by the use of multiple COI primer sets (Alberdi *et al.* 2017; Zhang *et*
409 *al.* 2018) or ribosomal markers (Deagle *et al.* 2014), given the increased workload of a multi

410 marker/primer approach and the limited taxonomic resolution of ribosomal markers (Clarke *et al.* 2017;
411 Marquina *et al.* 2018). These results were also recently confirmed by (Hajibabaei *et al.* 2019a), which
412 showed that the use of multiple primer sets did not substantially improve taxa detection.
413
414
415 **Conclusions**
416 Our study demonstrates that the fwh2, BF1/2/3 + BR2 and mlCOIintF based primer sets all
417 perform well when metabarcoding terrestrial arthropod samples. For most of these primer sets,
418 annealing temperatures of 46-50°C are ideal. When data is analyzed for Exact Sequence Variants
419 (ESVs), the BF3 + BR2 primer set is recommended as it is not affected by primer slippage. The present
420 study also reinforces the importance of careful primer validation using mock and field samples,
421 especially when primer performance has not yet been evaluated for the taxonomic group under study.
422 As a general rule, the use of multiple primer sets seems rarely justified as it increases laboratory effort
423 without substantially improving taxon recovery.
424
425
426 **Data availability**
427 Raw sequence data is available on the NCBI SRA archive; accession SRX4908948 provides data for
428 the mock community, accession SRX5175597 for the Malaise sample, and accession SRX4908947 for
429 the gradient PCR experiment. Demultiplexed read 1 and read 2 files are available for all sequencing
430 runs under the accessions listed in Table S3. The JAMP bioinformatics pipeline is available on GitHub
431 https://github.com/VascoElbrecht/JAMP with the used settings detailed in Scripts S1. Sequence
432 alignments generated with PrimerMiner are available at Dryad DOI.
433
434 **Authors' contributions**

Project design VE, DS; mock community assembly NI, SP; laboratory work VE, TWAB; Illumina sequencing MW, MH; bioinformatic analysis and statistics VE; funding PDNH, manuscript draft VE, DS, PDNH, TWAB . All authors developed the manuscript.

Supporting Information

Figure S1: Mock community composition.

Figure S2: Sequence alignment for 29 insect orders, including primer binding annotations. The alignment was used for primer development.

Figure S3: Evaluation of Levenshtein distances for fusion primers used to metabarcode the 21 primer sets.

Figure S4: Fusion primers used to metabarcode the 21 primer sets.

Figure S5: Fusion primers used for gradient metabarcoding.

Figure S6: Evaluation of Levenshtein distances for fusion primers used in gradient PCR.

Figure S7: Gradient PCR gels for the initial 36 primer combinations.

Figure S8: Amplicon concentration of the 36 primer sets after the first gradient PCR test.

Figure S9: Sequencing depth for the mock community metabarcoding run.

Figure S10: Distribution of read lengths after paired end merging for the mock community metabarcoding run.

Figure S11: Rarefaction curves showing taxon recovery for the mock sample with different primer sets.

Figure S12: Heat map showing taxon recovery for the mock sample with different primer sets.

Figure S13: Principal component analysis of the metabarcoding OTU table for the mock community metabarcoding run.

459 **Figure S14**: Jaccard similarity and Bray-Curtis distance based on taxa recovered from the mock
460 community metabarcoding run

461 **Figure S15**: Plot showing the similarity between taxon recovery at 46 °C with primers of both the
462 mock community metabarcoding and the final gradient metabarcoding run.

463 **Figure S16**: Heat map showing taxon recovery with four primer sets at different annealing
464 temperatures (40 - 56 °C).

465 **Figure S17**: Distribution of read length after paired end merging for the final gradient run.

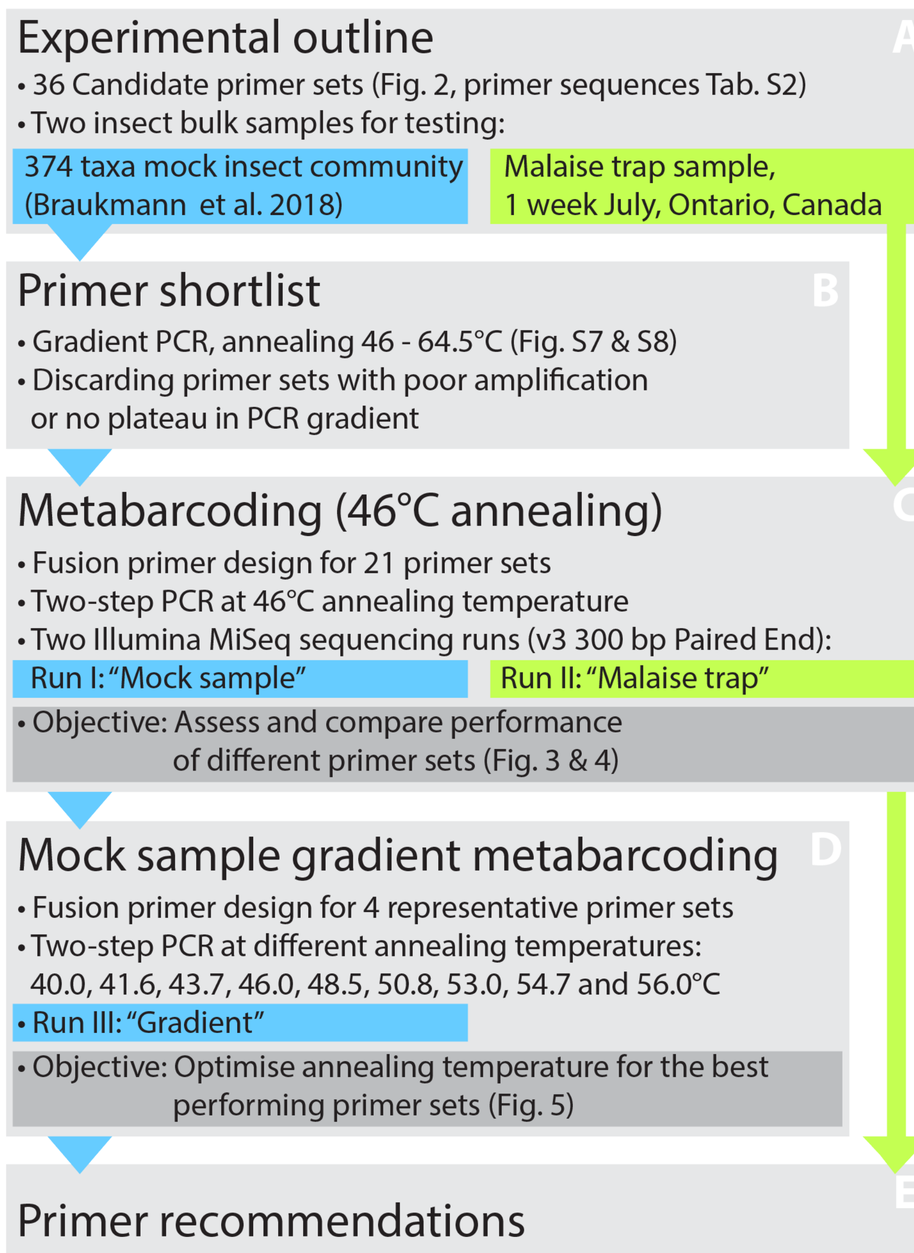
466 **Figure S18**: Heat map showing taxon recovery for the Malaise trap metabarcoding run with 21 primer
467 sets.

468 **Figure S19**: Rarefaction curves showing taxon recovery for the Malaise trap metabarcoding run with
469 different primer sets.

470 **Scripts S1**: R scripts used for bioinformatics processing, figure generation and statistical analysis.

471 **Table S1**: Raw OTU table for both the 21 primer and the gradient metabarcoding run, as well as details
472 on mock sample composition.

473 **Table S2**: Primer sequences and primer combinations evaluated in this study.



474

475 **Figure 1**: Overview of the experimental design. The performance of 36 primer pairs was tested via
476 gradient PCRs with a mock community of insects (A). The 21 pairs that showed best amplification
477 results (B) were selected for further DNA metabarcoding runs utilizing both the mock community and
478 a Malaise trap sample (C). Based on the metabarcoding results, four primer sets showing the good
479 performance were selected for a third test that examined the effects of varying annealing temperatures
480 on taxon recovery and non-specific amplification (D). Based on all results, the optimal primer sets were
481 designated (E).

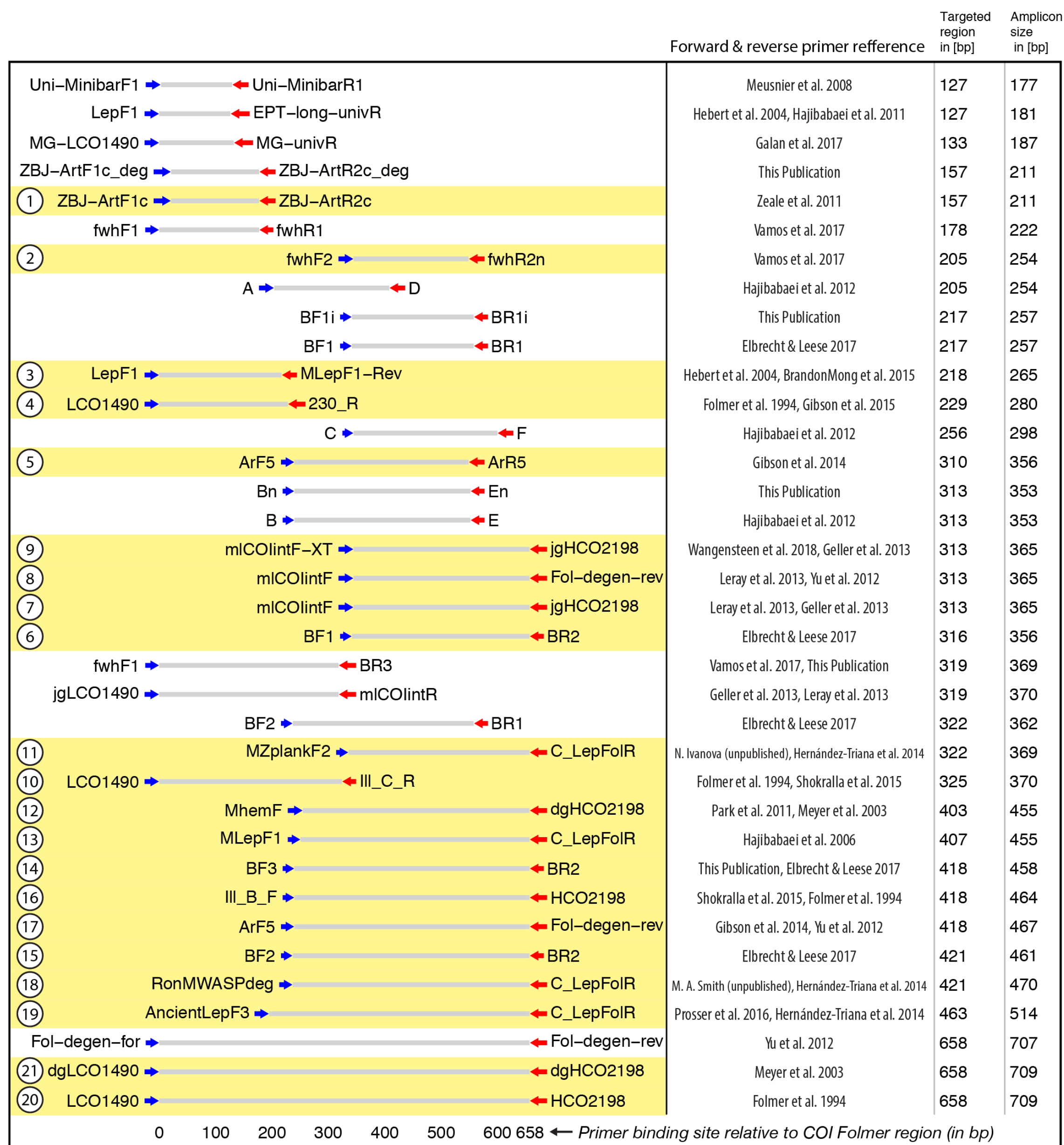
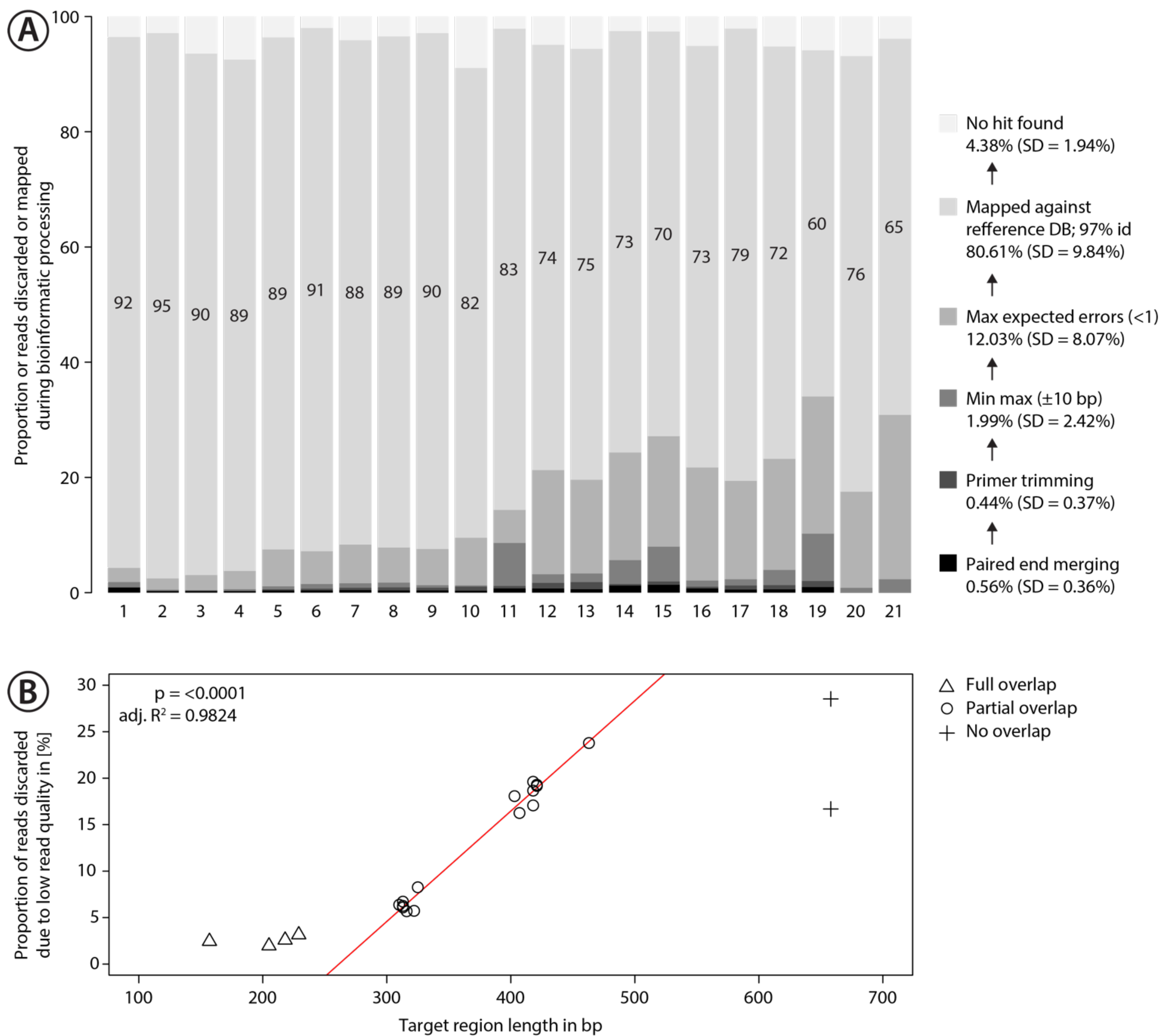


Figure 2: Target and amplicon length for the 36 primer sets evaluated via gradient PCR. The 21 primer sets selected for sequencing are highlighted in yellow while an ID for each pair is shown on the left. Primer references: (Folmer *et al.* 1994; Meyer 2003; Hebert *et al.* 2004; Hajibabaei *et al.* 2006; Meusnier *et al.* 2008; Hajibabaei *et al.* 2011; Zeale *et al.* 2011; Park *et al.* 2011; Yu *et al.* 2012; Hajibabaei *et al.* 2012; Leray *et al.* 2013; Geller *et al.* 2013; Gibson *et al.* 2014; Hernández-Triana *et al.* 2014; Shokralla *et al.* 2015; Brandon-Mong *et al.* 2015; Gibson *et al.* 2015; Prosser *et al.* 2015; Elbrecht & Leese 2017; Vamos *et al.* 2017; Wangensteen *et al.* 2018; Galan *et al.* 2018).

PeerJ Preprints NOT PEER-REVIEWED



490

491 **Figure 3**: Proportion of sequences discarded or mapped to reference sequences in the mock

492 community. **A**: Bar plots show the relative proportion of reads that were discarded or mapped.

493 Numbers in bars indicate the proportion of reads that matched one of the 374 species in the mock

494 community. The number for each primer pair on the x-axis corresponds with that in Figure 2. **B**:

495 Proportion of sequences discarded by max expected errors = 1 filtering using Usearch, plotted against

496 the length of the target region (in bp). Red line indicates linear regression.

497

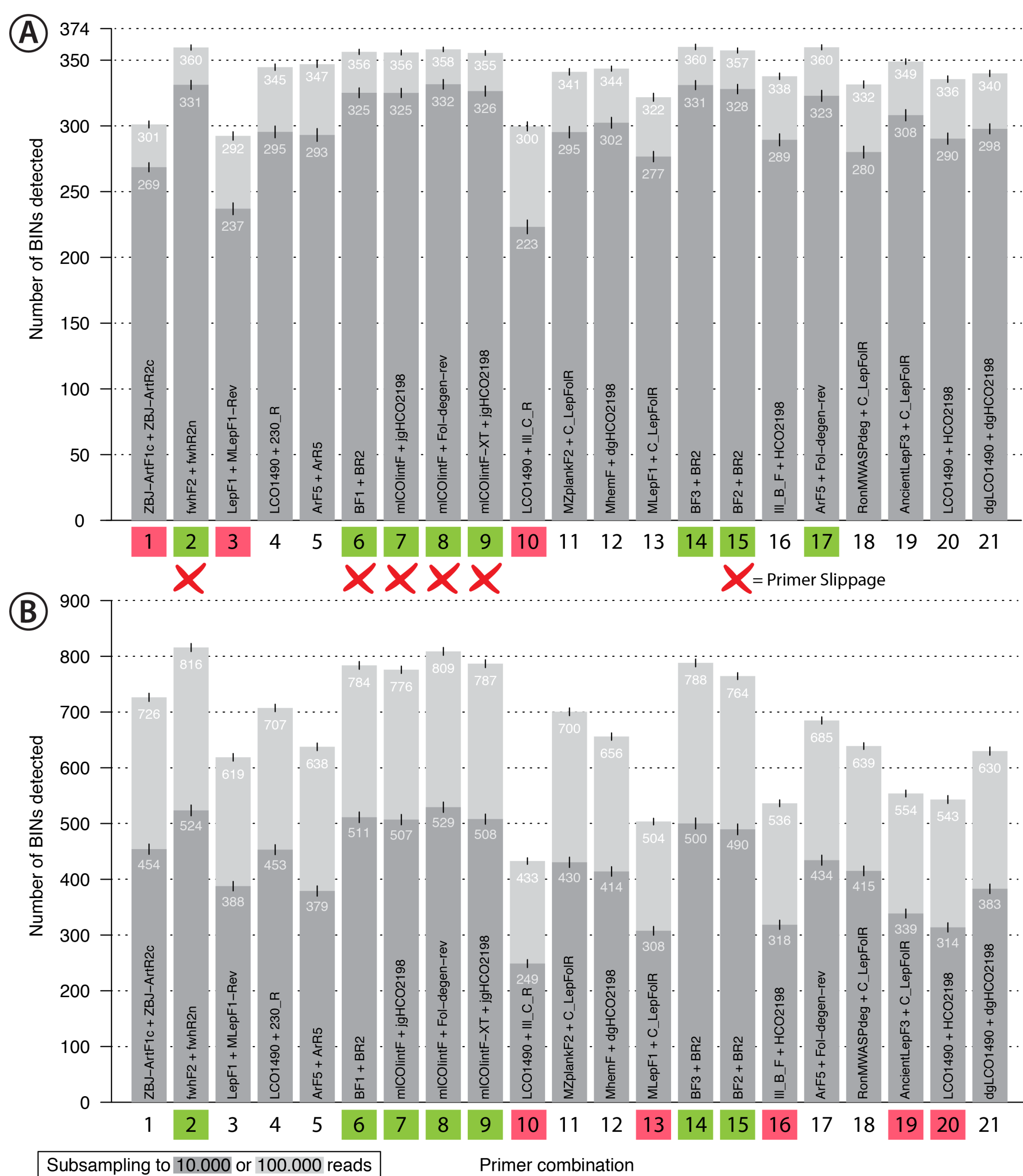
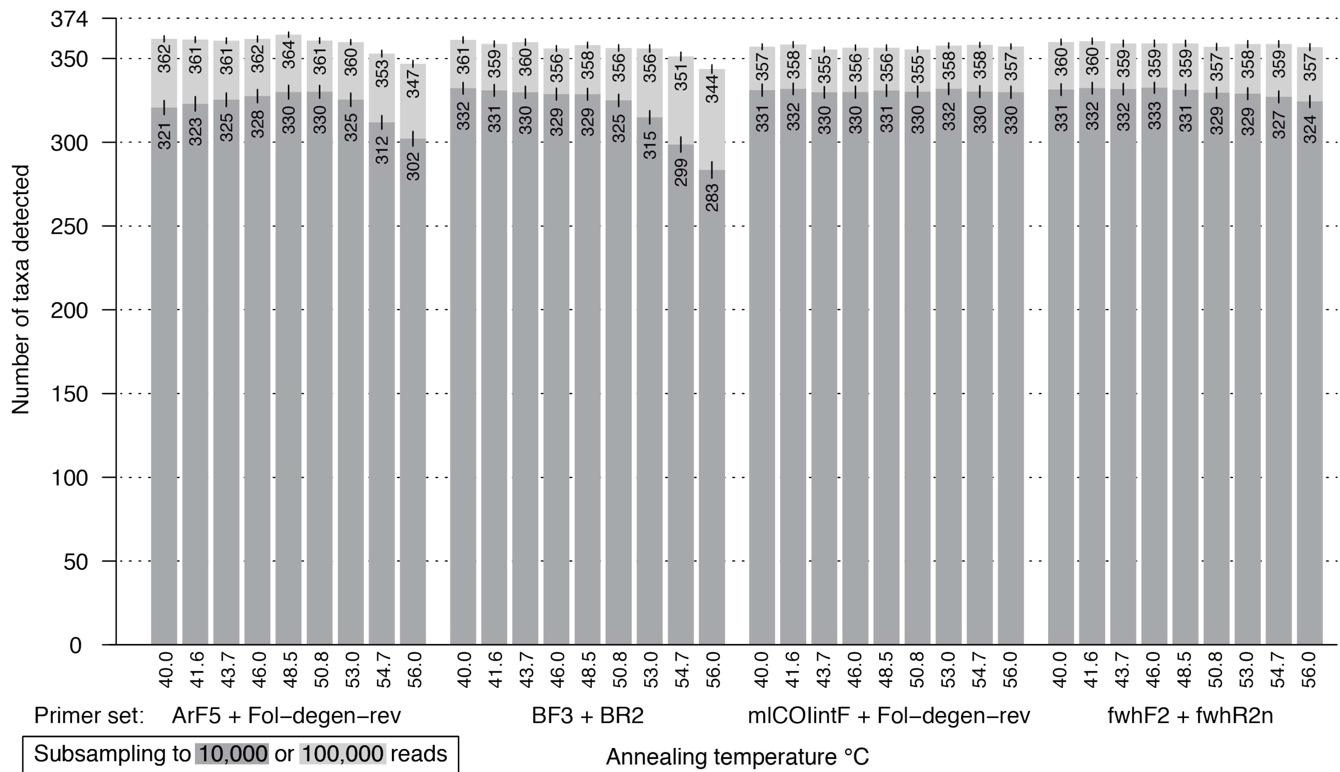


Figure 4: Bar plot showing the number of BINs recovered using metabarcoding with 21 primer pairs. The dark grey bar indicates subsampling at 10,000 reads while the light grey bar indicates subsampling at 100,000 reads per sample, each run with 1,000 replicates. Error bars show the standard deviation. Primer combinations affected by primer slippage (Elbrecht & Steinke 2018) are marked with a red X. **A**: Mock sample data, with primer combinations highlighted in green that detected more than 350 of the 374 BINs, while those that recovered fewer than 310 BINS are highlighted in red. **B**: Malaise trap data - primer combinations highlighted in green detected more than 750 BINs while those highlighted in red detected less than 600 BINs.



508

509 **Figure 5**: Bar plot showing the number of BINs recovered from the mock community at different

510 annealing temperatures. The dark grey bar indicates subsampling at 10,000 reads, while the light grey

511 bar depicts subsampling at 100,000 reads per samples; both were run with 1,000 replicates. Error bars

512 show the standard deviation.

513

514

515

516 **Acknowledgements**

517 We thank all staff at the CBG who collected the samples employed to assemble the mock community,
518 and Al Woodhouse from the Waterloo District School Board for his aid with Malaise trap sampling.
519 We also thank the Optimist Club of Kitchener-Waterloo for access to their site at Camp Heidelberg.
520 This study was supported by funding through the Canada First Research Excellence Fund. It represents
521 a contribution to the Food From Thought research program.

522

523

524 Abd-Elsalam, K.A. (2003). Bioinformatic tools and guideline for PCR primer design. *african Journal*
525 *of biotechnology*, **2**, 91–95–95.

526 Alberdi, A., Aizpurua, O., Gilbert, M.T.P. & Bohmann, K. (2017). Scrutinizing key steps for reliable
527 metabarcoding of environmental samples (A. Mahon, Ed.). *Methods in Ecology and Evolution*, **17**,
528 730–14.

529 Andújar, C., Arribas, P., Gray, C., Bruce, K., Woodward, G., Yu, D.W. & Vogler, A.P. (2017).
530 Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular*
531 *Ecology*.

532 Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P. & Emerson, B.C. (2018). Why the COI barcode
533 should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, **27**, 3968–3975.

534 Aylagas, E., Borja, A., Irigoien, X. & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA
535 Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Environmental*
536 *Science*, **3**, 1809–12.

537 Barnes, M.A. & Turner, C.R. (2015). The ecology of environmental DNA and implications for
538 conservation genetics. *Conservation Genetics*, 1–17.

539 Beng, K.C., Tomlinson, K.W., Shen, X.H., Surget-Groba, Y., Hughes, A.C., Corlett, R.T. & Slik,
540 J.W.F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity
541 and composition to land-use change in the tropics. *Scientific Reports*, 1–13.

542 Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L.D.S.B., Jones, G., Streicker, D.G. & Gilbert,
543 M.T.P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet
544 and population structure. *Molecular ecology resources*, **18**, 1050–1063.

545 Brandon-Mong, G.J., Gan, H.M., Sing, K.W., Lee, P.S., Lim, P.E. & Wilson, J.J. (2015). DNA
546 metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of*
547 *Entomological Research*, **105**, 717–727.

548 Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Elbrecht, V., Steinke, D., Ratnasingham, S.,
549 deWaard, J.R., Sones, J.E., Zakharov, E.V. & Hebert, P.D.N. (2019). Metabarcoding a Diverse
550 Arthropod Mock Community. *Molecular ecology resources*, 1–41.

551 Braukmann, T.W., Ivanova, N.V., Prosser, S.W., Elbrecht, V., Steinke, D., Ratnasingham, S., deWaard,
552 J.R., Sones, J.E., Zakhariv, E.V. & Hebert, P.D. (2018). Revealing the Complexities of
553 Metabarcoding with a Diverse Arthropod Mock Community. *bioRxiv*, 1–40.

554 Bylemans, J., Furlan, E.M., Gleeson, D.M., Hardy, C.M. & Duncan, R.P. (2018a). Does Size Matter?
555 An Experimental Evaluation of the Relative Abundance and Decay Rates of Aquatic
556 Environmental DNA. *Environmental science & technology*, **52**, 6408–6416.

557 Bylemans, J., Gleeson, D.M., Hardy, C.M. & Furlan, E. (2018b). Toward an ecoregion scale evaluation
558 of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray-
559 Darling Basin (Australia). *Ecology and Evolution*, **58**, 403–16.

560 Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017). Exact sequence variants should replace
561 operational taxonomic units in marker-gene data analysis. 1–5.

562 Carew, M.E., Pettigrove, V.J., Metzeling, L. & Hoffmann, A.A. (2013). Environmental monitoring
563 using next generation sequencing: rapid identification of macroinvertebrate bioindicator species.
564 *Frontiers in zoology*, **10**, 1–1.

565 Clarke, L.J., Beard, J.M., Swadling, K.M. & Deagle, B.E. (2017). Effect of marker choice and thermal
566 cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, **7**, 873–883.

567 Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. (2014). Environmental metabarcodes for insects:
568 in silicoPCR reveals potential for taxonomic bias. *Molecular ecology resources*, **14**, 1160–1170.

569 Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J. & Arnaud-Haond, S. (2015).
570 Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and
571 Molecular Surveys of Seagrass Communities (S. Mazzuca, Ed.). *PloS one*, **10**, e0117562–26.

572 Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C. & Bik, H.M.
573 (2016). The ecologist's field guide to sequence-based identification of biodiversity (R. Freckleton,
574 Ed.). *Methods in Ecology and Evolution*, **7**, 1008–1018.

575 Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014). DNA metabarcoding and
576 the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, 1–4.

577 Demeke, T. & Jenkins, G.R. (2009). Influence of DNA extraction methods, PCR inhibitors and
578 quantification methods on real-time PCR assay of biotechnology-derived traits. *Analytical and*
579 *Bioanalytical Chemistry*, **396**, 1977–1990.

580 Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B. & Collen, B. (2014). Defaunation in the
581 Anthropocene. *Science (New York, NY)*, **345**, 401–406.

582 Drummond, A.J. (2018). Evaluating a multigene environmental DNA approach for biodiversity
583 assessment. 1–19.

584 Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**,

585 2460–2461.

586 Edgar, R.C. & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-
587 generation sequencing reads. *Bioinformatics*, **31**, 3476–3482.

588 Elbrecht, V. & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species
589 Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative
590 Metabarcoding Protocol (M. Hajibabaei, Ed.). *PloS one*, **10**, e0130324–16.

591 Elbrecht, V. & Leese, F. (2016). PrimerMiner: an R package for development and in silico validation
592 of DNA metabarcoding primers. *Methods in Ecology and Evolution*, **8**, 622–626.

593 Elbrecht, V. & Leese, F. (2017). Validation and development of freshwater invertebrate metabarcoding
594 COI primers for Environmental Impact Assessment. *Frontiers in Environmental Science*, **5**, 1–11.

595 Elbrecht, V. & Steinke, D. (2018). Scaling up DNA metabarcoding for freshwater macrozoobenthos
596 monitoring. *Freshwater biology*.

597 Elbrecht, V., Hebert, P.D.N. & Steinke, D. (2018a). Slippage of degenerate primers can cause variation
598 in amplicon length. *Scientific Reports*, 1–5.

599 Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.-N., Coissac, E.,
600 Boyer, F. & Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA
601 metabarcoding of insects. *PeerJ*, **4**, e1966–12.

602 Elbrecht, V., Vamos, E.E., Steinke, D. & Leese, F. (2018b). Estimating intraspecific genetic diversity
603 from community DNA metabarcoding data. *PeerJ*, **6**, e4644–13.

604 Elbrecht, V., Vamos, E., Meissner, K., Aroviita, J. & Leese, F. (2017). Assessing strengths and
605 weaknesses of DNA metabarcoding based macroinvertebrate identification for routine stream
606 monitoring. *Methods in Ecology and Evolution*, 1–21.

607 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. (2016). MultiQC: summarize analysis results for
608 multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.

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

612 Galan, M., Pons, J.-B., Tournayre, O., Pierre, É., Leuchtmann, M., Pontier, D. & Charbonnel, N.
613 (2018). Metabarcoding for the parallel identification of several hundred predators and their prey:
614 Application to bat species diet analysis. *Molecular ecology resources*, **14**, 1–16.

615 Geller, J., Meyer, C., Parker, M. & Hawk, H. (2013). Redesign of PCR primers for mitochondrial
616 cytochrome coxidase subunit I for marine invertebrates and application in all-taxa biotic surveys.
617 *Molecular ecology resources*, **13**, 851–861.

618 Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I. & Hajibabaei, M. (2015).
619 Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput
620 Sequencing (D. Fontaneto, Ed.). *PloS one*, **10**, e0138432–15.

621 Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W. &
622 Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk
623 sample of tropical arthropods through DNA metasystematics. *Proceedings of the National*
624 *Academy of Sciences*, **111**, 8007–8012.

625 Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. & Hebert, P.D.N. (2006). DNA barcodes
626 distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the*
627 *United States of America*, **103**, 968–971.

628 Hajibabaei, M., Porter, T.M., Wright, M. & Rudar, J. (2019a). COI metabarcoding primer choice
629 affects richness and recovery of indicator taxa in freshwater systems. *bioRxiv*, **11**, 2639–30.

630 Hajibabaei, M., Porter, T., Robinson, C. & BioRxiv, D.B. (2019b). Watered-down biodiversity? A
631 comparison of metabarcoding results from DNA extracted from matched water and bulk tissue
632 biomonitoring samples. *biorxiv.org*

633 .

634 Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. & Baird, D.J. (2011). Environmental Barcoding: A
635 Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PloS*
636 *one*, **6**, 1–7.

637 Hajibabaei, M., Spall, J.L., Shokralla, S. & van Konynenburg, S. (2012). Assessing biodiversity of a
638 freshwater benthic macroinvertebrate community through non-destructive environmental barcoding
639 of DNA from preservative ethanol. *BMC ecology*, **12**, 1–1.

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

642 Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004). Ten species in one:
643 DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator.
644 *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–
645 14817.

646 Hernández-Triana, L.M., Prosser, S.W., Rodríguez-Perez, M.A., Chaverri, L.G., Hebert, P.D.N. &
647 Ryan Gregory, T. (2014). Recovery of DNA barcodes from blackfly museum specimens (Diptera:
648 Simuliidae) using primer sets that target a variety of sequence lengths. *Molecular ecology*
649 *resources*, **14**, 508–518.

650 Hurlbert, S.H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological*
651 *Monographs*, **54**, 187–211.

652 Jusino, M.A., Banik, M.T., Palmer, J.M., Wray, A.K., Xiao, L., Pelton, E., Barber, J.R., Kawahara,
653 A.Y., Gratton, C., Zachariah Peery, M. & Lindner, D.L. (2018). An improved method for utilizing
654 high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular*
655 *ecology resources*.

656 Krehenwinkel, H., Fong, M., Kennedy, S., Huang, E.G., Noriyuki, S., Cayetano, L. & Gillespie, R.
657 (2018). The effect of DNA degradation bias in passive sampling devices on metabarcoding studies
658 of arthropod communities and their associated microbiota (M. Kuntner, Ed.). *PloS one*, **13**,

659 e0189188–14.

660 Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B. & Gillespie, R.G. (2017).
661 Estimating and mitigating amplification bias in qualitative and quantitative arthropod
662 metabarcoding. *Scientific Reports*, 1–12.

663 Leray, M. & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal
664 patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, **112**, 2076–
665 2081.

666 Leray, M. & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic
667 OTUs in Illumina COI metabarcoding. *PeerJ*, **5**, e3006–27.

668 Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & Machida,
669 R.J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region
670 for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents.
671 *Frontiers in zoology*, **10**, 1–14.

672 Macher, J.N. & Weigand, A.M. (2018). A DNA metabarcoding protocol for hyporheic freshwater
673 meiofauna: Evaluating highly degenerate COI primers and replication strategy. *Metabarcoding and*
674 *Metagenomics*, **2**, 63–15.

675 Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. & Leese, F. (2018).
676 Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate
677 cytochrome coxidase I primers. *Molecular ecology resources*, **17**, 1231–43.

678 Marquina, D., Andersson, A.F. & Ronquist, F. (2018). New mitochondrial primers for metabarcoding
679 of insects, designed and evaluated using in silicomethods. *Molecular ecology resources*, 1–39.

680 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
681 *EMBnet journal*, **17**, 10–12.

682 Meusnier, I., Singer, G.A., Landry, J.-F., Hickey, D.A., Hebert, P.D. & Hajibabaei, M. (2008). A
683 universal DNA mini-barcode for biodiversity analysis. *BMC genomics*, **9**, 214.

684 Meyer, C.P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification
685 patterns in the tropics. *Biological Journal of the Linnean Society*, **79**, 401–459.

686 Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. &
687 Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular ecology resources*,
688 **12**, R18–13.

689 Park, D.-S., Foottit, R., Maw, E. & Hebert, P.D.N. (2011). Barcoding Bugs: DNA-Based Identification
690 of the True Bugs (Insecta: Hemiptera: Heteroptera) (M.T.P. Gilbert, Ed.). *PloS one*, **6**, e18749–9.

691 Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. (2014). Universal and blocking primer mismatches
692 limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods.
693 *Molecular ecology resources*, **15**, 1–12.

694 Piñol, J., Senar, M.A. & Symondson, W.O.C. (2018). The choice of universal primers and the
695 characteristics of the species mixture determine when DNA metabarcoding can be quantitative.

696 *Molecular Ecology*, **6**, 1809–13.

697 Porter, T.M. & Hajibabaei, M. (2018a). Automated high throughput animal CO1 metabarcode
698 classification. *Scientific Reports*, 1–10.

699 Porter, T.M. & Hajibabaei, M. (2018b). Over 2.5 million COI sequences in GenBank and growing (W.
700 Arthofer, Ed.). *PloS one*, **13**, e0200177–16.

701 Prosser, S.W.J., deWaard, J.R., Miller, S.E. & Hebert, P.D.N. (2015). DNA barcodes from century-old
702 type specimens using next-generation sequencing. *Molecular ecology resources*, **16**, 487–497.

703 Ratnasingham, S. & Hebert, P. (2007). BOLD: The Barcode of Life Data System (http://www.
704 barcodinglife. org). *Molecular Ecology Notes*, **7**, 355–364.

705 Ratnasingham, S. & Hebert, P.D.N. (2013). A DNA-Based Registry for All Animal Species: The
706 Barcode Index Number (BIN) System (D. Fontaneto, Ed.). *PloS one*, **8**, e66213–16.

707 Reuter, J.A., Spacek, D.V. & Snyder, M.P. (2015). High-Throughput Sequencing Technologies.
708 *Molecular Cell*, **58**, 586–597.

709 Sellers, G.S., Di Muri, C., Gómez, A. & Hänfling, B. (2018). Mu-DNA: a modular universal DNA
710 extraction method adaptable for a wide range of sample types. *Metabarcoding and Metagenomics*,
711 **2**, e24556–11.

712 Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B. &
713 Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification
714 using an Illumina MiSeq platform. *Scientific Reports*, **5**, 9687.

715 Smith, D.P. & Peay, K.G. (2014). Sequence Depth, Not PCR Replication, Improves Ecological
716 Inference from Next Generation DNA Sequencing (C.A. Kellogg, Ed.). *PloS one*, **9**, e90234.

717 Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R.,
718 Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M.,
719 Persson, L.M., Ramanathan, V., Reyers, B. & Sörlin, S. (2015). Planetary boundaries: guiding
720 human development on a changing planet. *Science (New York, NY)*, **347**, 1259855.

721 Vamos, E.E., Elbrecht, V. & Leese, F. (2017). Short COI markers for freshwater macroinvertebrate
722 metabarcoding. *Metabarcoding and Metagenomics*, 1–20.

723 Vierna, J., Doña, J., Vizcaíno, A., Serrano, D. & Jovani, R. (2017). PCR cycles above routine numbers
724 do not compromise high-throughput DNA barcoding results. (F. Chain, Ed.). *Genome*, **60**, 868–
725 873.

726 Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S.,
727 Bunn, S.E., Sullivan, C.A., Liermann, C.R. & Davies, P.M. (2010). Global threats to human water
728 security and river biodiversity. *Nature*, **467**, 555–561.

729 Wangensteen, O.S., Palacín, C., Guardiola, M. & Turon, X. (2018). DNA metabarcoding of littoral
730 hard-bottom communities: high diversity and database gaps revealed by two molecular markers.
731 *PeerJ*, **6**, e4705–30.

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

735 Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C. & Jones, G. (2011). Taxon-specific PCR for
736 DNA barcoding arthropod prey in bat faeces. *Molecular ecology resources*, **11**, 236–244.

737 Zepeda-Mendoza, M.L., Bohmann, K., Baez, A.C. & Gilbert, M.T.P. (2016). DAMe: a toolkit for the
738 initial processing of datasets with PCR replicates of double-tagged amplicons for DNA
739 metabarcoding analyses. *BMC Research Notes*, 1–13.

740 Zhang, G.K., Chain, F.J.J., Abbott, C.L. & Cristescu, M.E. (2018). Metabarcoding using multiplexed
741 markers increases species detection in complex zooplankton communities. *Evolutionary*
742 *Applications*, **215**, 403–40.

743