A peer-reviewed version of this preprint was published in PeerJ on 19 April 2016.

View the peer-reviewed version (peerj.com/articles/1966), which is the preferred citable publication unless you specifically need to cite this preprint.

Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J, Coissac E, Boyer F, Leese F. 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4:e1966
https://doi.org/10.7717/peerj.1966

PeerJ Preprints NOT PEER-REVIEWED

1 # Testing the potential of a ribosomal 16S marker for DNA
2 # metabarcoding of insects

3 Vasco Elbrecht[1,2], Pierre Taberlet[3,4], Tony Dejean[5], Alice Valentini[5], Philippe Usseglio-Polatera[6], Jean-Nicolas
4 Beisel[7,8], Eric Coissac[3,4], Fred Boyer[3,4], Florian Leese[1,9]
5
6 Affiliations:
7 1) Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen,
8 Germany
9 2) Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Universitätsstraße 150, 44801
10 Bochum, Germany
11 3) CNRS, Laboratoire d'Ecologie Alpine (LECA), 38000 Grenoble, France
12 4) Univ. Grenoble Alpes, Laboratoire d'Ecologie Alpine (LECA), 38000 Grenoble, France
13 5) SPYGEN, Savoie Technolac, 73370 Le Bourget du Lac, France
14 6) Univ Lorraine, CNRS, IBISE,UMR 7360, Lab Interdisciplinaire Environnements Continentaux, F-57070 Metz,
15 France
16 7) Ecole Natl Genie Eau & Environm Strasbourg, F-67070 Strasbourg, France
17 8) Univ Strasbourg, UMR CNRS 7362, LIVE, F-67083 Strasbourg, France
18 9) Centre for Water and Environmental Research (ZWU) Essen, University of Duisburg-Essen, Universitätsstraße 2,
19 45141 Essen,
20
21 Corresponding author: florian.leese@uni-due.de, +49.201-1834053, University of Duisburg-Essen, Faculty of Biology,
22 Aquatic Ecosystem Research, Universitätsstr. 5, D-45141 Essen, Germany

23 ## Abstract

24 Cytochrome c oxidase I (COI) is a powerful marker for DNA barcoding of animals, with good taxonomic resolution and
25 a large reference database. However, when used for DNA metabarcoding, estimation of taxa abundances and species
26 detection are limited due to primer bias caused by highly variable primer binding sites across the COI gene. Therefore,
27 we explored the ability of the 16S ribosomal DNA gene as an alternative metabarcoding marker for species level
28 assessments. Ten bulk samples, each containing equal amounts of tissue from 52 freshwater invertebrate taxa, were
29 sequenced with the Illumina NextSeq 500 system. In comparison to COI, the 16S marker amplified more insect species

PeerJ Preprints NOT PEER-REVIEWED

and amplified more equally, probably due to decreased primer bias. Rough estimation of biomass might thus be less biased with 16S than with COI. According to these results, the marker choice depends on the scientific question. If the goal is to obtain a taxonomic identification at the species level, then COI is more appropriate due to established reference databases and known taxonomic resolution of this marker, knowing that a greater proportion of species will be missed using COI Folmer primers. If the goal is to obtain a more comprehensive survey in a context where it is possible to build a local reference database, the 16S marker could be more appropriate.

Keywords: Biodiversity assessment, stream monitoring, small ribosomal subunit, high throughput sequencing

PeerJ Preprints NOT PEER-REVIEWED

38 ## Introduction

39 DNA metabarcoding is a novel and powerful method to assess biodiversity in ecosystems (Hajibabaei et al., 2011;
40 Taberlet et al., 2012; Yu et al., 2012; Carew et al., 2013; Gibson et al., 2014; Leray & Knowlton, 2015; Dowle et al.,
41 2015). Well-designed universal PCR primers are the most critical component when assessing species diversity in
42 ecosystems with DNA metabarcoding, because environmental samples typically contain hundreds of specimens of
43 phylogenetically different taxa. Substantial primer bias in commonly used DNA barcoding markers, such as the
44 Cytochrome c Oxidase subunit I (COI) gene for animals, prevents the detection of all taxa in a sample and thus the
45 estimation of taxa biomass is difficult (Deagle et al., 2014; Piñol et al., 2014; Elbrecht & Leese, 2015). However,
46 accurate and comprehensive taxa lists are critical for assessment of biodiversity and ecosystem health. Given the great
47 sequence variability of the COI marker, the use of alternative DNA metabarcoding markers has been suggested (Deagle
48 et al., 2014) and PCR-free metagenomics strategies are being tested for environmental assessment (Gómez-Rodríguez et
49 al., 2015; Tang et al., 2015). One marker with potential for species level resolution and more conserved regions is the
50 mitochondrial 16S rRNA gene (Deagle et al., 2014). In this study, we evaluate the performance of a short 16S region
51 which we compared with the standard COI Folmer (et al., 1994) marker for metabarcoding , using freshwater
52 invertebrates mock communities. The ten freshwater mock communities were each comprised of 52 morphologically
53 identified taxa and have been used in a previous study on COI primer bias (Elbrecht & Leese, 2015). Thus, they are
54 ideal to comparatively evaluate the success rate of a short 16S fragment for DNA-based monitoring.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1855v1 | CC-BY 4.0 Open Access | rec: 13 Mar 2016, publ: 13 Mar 2016

55 # Material and Methods

56 The same DNA aliquots as in Elbrecht & Leese (2015) were used to test the 16S marker to allow for a direct
57 comparison. Laboratory conditions and bioinformatic analyses were kept as similar as possible to the study by (Elbrecht
58 & Leese, 2015).
59
60 **DNA metabarcoding**
61 The used 16S markers ins_F / ins_R amplify a ~157 bp of the mitochondrial 16S gene. This marker was developed
62 using the ecoPrimers program (Riaz et al., 2011) and represents a variant of the Ins16S_1short primer pair (Clarke et al.,
63 2014). Fusion primers were used (Figure S1), allowing to load PCR amplicons directly onto the Illumina NextSeq 500
64 sequencer. The same tag shifting and simultaneous sequencing of forward and reverse primer and 10% PhiX spike in as
65 described by (Lundberg et al., 2013; Elbrecht & Leese, 2015) was used, to increase sequence diversity. Unique inline
66 barcodes on forward and reverse reads were used for sample indexing.
67 The same one-step PCR and library preparation conditions as in (Elbrecht & Leese, 2015) were used with the following
68 modifications: PCR extension time was reduced to 120 seconds and annealing temperature increased to 52.5°C to better
69 suit the fragment length and melting temperatures of the 16S Ins primers. Only one PCR replicate per sample was used
70 for sequencing. Amplicons were purified with magnetic beads, but only a left-sided size selection was carried out to
71 remove remaining primers and primer dimers (0.9x SPRIselect, Beckman Coulter, Bread, CA, USA). Concentrations
72 were quantified with the Qubit BR Kit (Thermofisher Scientific, Carlsbad, CA, USA) and the library for sequencing
73 was prepared by pooling 190 ng PCR product of all ten samples.
74 Paired-end Illumina sequencing was performed on a NextSeq 500 system using the mid output kit v2 kit with 300
75 cycles (150 bp PE sequencing) at the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research,
76 Bremerhaven Germany.
77
78 **Generation of 16S reference sequences**
79 Due to the limited availability of 16S reference sequences on GenBank (NCBI), we constructed a reference library for
80 the used 52 morphotaxa given tissue availability. Standard DNA salt extraction, PCR, PCR clean-up, and Sanger
81 sequencing were conducted as described in (Elbrecht et al., 2014), to amplify the 16S region with different primer sets
82 and combinations. Primers were newly developed or checked for mismatches to Ephemeroptera, Plecoptera and
83 Trichoptera using the PrimerMiner v0.2 R package (https://github.com/VascoElbrecht/PrimerMiner) and are available
84 together with the generated reference sequences on BOLDsystems (DS-TMIX16S). An annealing temperature of 52°C

Peer Preprints NOT PEER-REVIEWED

85 was used for all primer combinations using HotMaster Taq (5Prime; Gaithersburg, Maryland, USA) for amplification.

86

87 **Bioinformatic analysis**

88 Figure S2B includes a flow chart of the data processing steps. All used custom R scripts are available in the
89 supplementary information (S1 scripts). First, reads were demultiplexed (R script splitreads_ins_v11.R) and paired end
90 reads merged using USEARCH v8.0.1623 -fastq_mergepairs with -fastq_merge_maxee 1.0 (Edgar & Flyvbjerg, 2015).
91 Primers were removed with cutadapt version 1.8.1 (Martin, 2011). Sequences from all ten replicates were pooled,
92 dereplicated, and singletons were removed to find operational taxonomic units (OTUs) using the UPARSE pipeline
93 (cluster_otus, 97% identity, Edgar, 2013). Chimeras were removed from the OTUs using uchime_denovo. The
94 remaining OTUs were identified by querying against all nucleotide records on NCBI using the Blast API and an R
95 script and our local 16S database using BLAST 2.2.31+ (Camacho et al., 2009). Taxonomy was assigned and checked
96 manually, and in rare cases matches of ~90% identity were accepted, if they matched the patterns which were
97 previously reported for COI (Elbrecht & Leese, 2015).
98 The ten samples were dereplicated using derep_fulllength, but singletons were included in the data set. Sequences of
99 each sample were matched against the OTUs with a minimum match of 97% using usearch_global. The hit tables were
100 imported and the sequence numbers were normalised to the total sequence abundance and tissue weight for the various
101 taxa. Only OTUs with a read abundance above 0.003% in at least one replicate were considered in downstream analysis.
102 Due to the exponential nature of PCR, statistical tests on weight adjusted relative read abundances were carried out on
103 decadic logarithm. Expected relative abundance was calculated by dividing 100% by the number of morphospecies
104 detected with each marker.

PeerJ Preprints NOT PEER-REVIEWED

105 # Results

106 ## Amplicon sequencing success and sequence processing

107 The NextSeq run generated 42.3 Gbp of raw sequencing data (NCBI SRA accession number SRR2217415). Cluster
108 density was 177 K/mm^2 and read quality good with Q30 ≥ 85.3%. Read abundance was 17% higher when sequencing
109 started with the P5_Ins_F primers (t-test, $p < 0.001$, Figure S2 A). This, however, did not introduce any significant
110 differences between forward and reverse primer in the bioinformatic processing downstream (t-test, Figure S2 B).

111 Initial OTU clustering generated 855 OTUs of which 22.5% were detected as chimeras. Sequences from each sample
112 were compared against the remaining 663 OTUs, but only 243 OTUs had at least one sample with > 0.003% sequence
113 abundance and were thus included in further analysis. Taxonomy could be assigned for most OTUs based on available
114 reference data and our own reference sequences. Reference data for the 16S marker could be generated for 42 of the 52
115 morphotaxa by Sanger sequencing. Together 16S sequences from NCBI we were able to obtain reference data for all
116 morphotaxa (Figure 1). However, in some cases the NCBI data was only obtained for taxa identified on Family or Order
117 level (e.g. Lymnaeidae, Nematoda, Acari) and might not be sufficient for reliable taxa identification. Supplementary
118 Table S1 gives an overview of assigned taxonomy for each OTU and table S2 shows the distribution of detected taxa
119 across the 10 replicates. Table S3 shows the sequence abundance for each morphotaxon from the 16S dataset as well as
120 the COI dataset from (Elbrecht & Leese, 2015), which was used for comparison of primer bias between both markers.

121 ## Taxon recovery with 16S

122 The taxonomic assignment was straigthforward for the COI marker, due to the availability of reliable reference
123 databases, which was not the case for the 16S marker. Forty-one out of 42 insect species were detected by the 16S. The
124 Sanger sequence generated for the Tipulidae ptesent in our mock samples showed mismatches at the 3" end of both the
125 forward and reverse 16S primers and was not detected in the data set.

126 ## COI versus 16S

127 Most insect taxa were amplified with both markers, (38 out of 42), no insect taxon was only detected by COI, while 16S
128 detected three more taxa (Ephemeridae, *Sercostoma personatum*, Rhyacophyla). The 16S primers worked very
129 effectively for insect taxa, specifically in the indicator taxa Ephemeroptera, Plecoptera and Trichoptera (100% detection
130 success, table 1). Of the ten other Metazoa, five were detected by COI, and only two by 16S. Variation in logarithmic
131 insect read abundance was much lower for the ribosomal 16S amplicons (SD=0.62%) than for the COI Folmer primers

PeerJ Preprints NOT PEER-REVIEWED

132 (SD=1.0%) used on the same samples as in (Elbrecht & Leese, 2015) (Fig 1). Logarithmic precision of read abundance
133 (distance to expected abundance) was significantly higher for 16S (SD=0.37) than COI (SD=0.72, paired Wilcoxon
134 signed-rank test, p = 0.002). Additionally, the COI primers showed more dropouts of a few specimens per taxa (orange
135 numbers, figure 1), while the 16S primer with the exception of three cases always amplified all 10 specimens of a taxon.
136 Table 1 compares the number of taxa recovered for the four most relevant orders for water quality assessment.

137 **Table 1.** Number of specimens recovered with the COI and 16S primers.

Taxonomic group	Recovered specimens			
	COI		16S	
Ephemeroptera	7/8	(88%)	8/8	(100%)
Plecoptera	4/4	(100%)	4/4	(100%)
Trichoptera	13/15	(86%)	15/15	(100%)
Diptera	7/8	(88%)	7/8	(88%)
Other insects	7/7	(100%)	7/7	(100%)
Other metazoa	5/10	(50%)	2/10	(20%)
Σ All insects	38/42	(91%)	41/42	(98%)
Σ All taxa	43/52	(83%)	43/52	(83%)

138
139
140
141
142
143

PeerJ Preprints NOT PEER-REVIEWED

144 ## Discussion

145 We successfully ported our DNA metabarcoding protocol from the MiSeq system (Elbrecht & Leese, 2015) onto the
146 NextSeq 500. As demonstrated in the previous study and also seen for the 16S Ins marker here, the use of fusion
147 primers with a dual sequencing strategy maximizes sequence diversity, but can lead to a slight bias in read abundance.
148 This however does not strongly affect read abundance of individual specimens between replicates (see Figure S5 & S7
149 in Elbrecht & Leese, 2015). Further, a slight bias on sequence abundance might be introduced between and within
150 samples by e.g. different amount of cuticula present when weighing tissue, tissue quality and variation in mitochondrial
151 copy number. However, these effects the same for both markers, so observed effects can be likely explained by primer
152 bias. Here, we focus on comparing the results obtained from sequencing the mock community of 52 taxa using the two
153 different markers and discuss their advantages and disadvantages.

154 ### Power and limitations of 16S and COI markers in DNA metabarcoding

155 A key advantage of COI as a marker for DNA metabarcoding is that reference databases have been well established and
156 are actively developed and extended (Ratnasingham & Hebert, 2007). DNA barcoding and the COI gene has been
157 widely accepted by the scientific community as the barcoding marker of choice for animals (Ratnasingham & Hebert,
158 2013; Porter et al., 2014), despite some negative voices (Taylor & Harris, 2012). Additionally the taxonomic resolution
159 of the COI marker has been extensively tested, and its usefulness for identifying freshwater invertebrates on species
160 level demonstrated (Zhou et al., 2009; Pfrender et al., 2010; Zhou et al., 2010; Sweeney et al., 2011). However, a
161 documented concern of this marker is its large variability, which introduces primer bias due to mismatches at the primer
162 binding sites (Piñol et al., 2014), which creates the risk of losing some target taxa (Deagle et al., 2014). This large
163 variation makes estimating biomass from PCR-based DNA metabarcoding results difficult (Elbrecht & Leese 2015).
164 The results of this study show that the 16S ins primers show less amplification bias than the COI Folmer primers
165 previously tested. Specifically for the Ephemeroptera, Plecoptera and Trichoptera, the 16S results were very consistent
166 with variation in sequence abundance within these groups, with variation of only one order of magnitude magnitude for
167 most taxa. A further advantage is that the reduced primer bias in 16S could allow for lower sequencing depths and thus
168 a reduction in costs. Additionally the short length of the 16S marker used facilitates amplification, which is important
169 when dealing with degraded DNA or eDNA. The downside of using 16S as a marker at the present, however, is the
170 limited availability of reference sequences and the yet not fully explored taxonomic resolution on species level. We had
171 to establish our own 16S reference sequences for our mock communities *de novo* whenever tissue of the analysed
172 morphotaxa was still available. This created extra work and cost that was omitted when using COI..

PeerJ Preprints NOT PEER-REVIEWED

173 **Which marker to use?**

174 COI is the standard marker for barcoding of animals at the present and will typically yield the best resolved taxonomic
175 lists at the moment. Therefore, if the goal of a project is to obtain a taxonomic identification at the species level, COI is
176 most appropriate. However due to the codon degeneracy some taxa will likely not be amplified and thus missing in the
177 dataset, making the COI marker not ideal when complete taxon lists are required. However, the use of improved COI
178 primers with high degeneracy might lead to equally good detection and amplification consistency. If the project goal is
179 to obtain a more comprehensive survey and where it is possible to build a local reference database 16S can be a
180 versatile and even better alternative to COI, as this marker minimizes primer bias and provides more consistent PCR,
181 possibly allowing for rough biomass inferences. For species-level assignments, the potential of 16S remains largely
182 unexplored for assessment of relevant invertebrate indicator taxa such as Ephemeroptera, Plecoptera and Trichoptera.
183 Thus, prior to a routine application on 16S for species-level assessment we recommend reference sequencing of whole
184 mitochondrial genomes using high throughput sequencing (Tang et al., 2014), which not only allows for estimating
185 taxonomic resolution of the two different mitochondrial markers, but also build the backbone for future metagenomic
186 studies (Tang et al., 2015). Once comprehensive mitochondrial reference databases are established, also direct PCR-free
187 shotgun sequencing of bulk samples (metagenomics) becomes feasible. This approaches could further improve taxa
188 detection rates and potentially allow to estimate taxa abundance (Gómez-Rodríguez et al., 2015; Tang et al., 2015).
189 Using methods to enrich for mitochondrial reads we could further decrease sequencing costs for reference sequencing
190 and mitogenomics approaches alike (Zhou et al., 2013; Liu et al., 2015; Dowle et al., 2015).

191

192 Conclusions

193 In this study we show that the ribosomal 16S marker shows less primer bias than the COI barcoding marker, when
194 applied for DNA metabarcoding of freshwater invertebrates. Thus the 16S marker might allow to reduce sequencing
195 depth in DNA based stream assessment, which could reduce sequencing costs. The main drawback when compared to
196 COI is that little reference databases for stream invertebrates are available for the 16S marker and that taxonomic
197 resolution remains largely unknown. This, however, might change in the future when, especially when thanks to recent
198 developments in mitogenomics methods more complete mitochondrial reference genomes become available and 16S
199 could become an alternative marker to COI in cases where local reference databases can be easily generated and a more
200 comprehensive result, i.e. less affected by primer bias, is the goal.

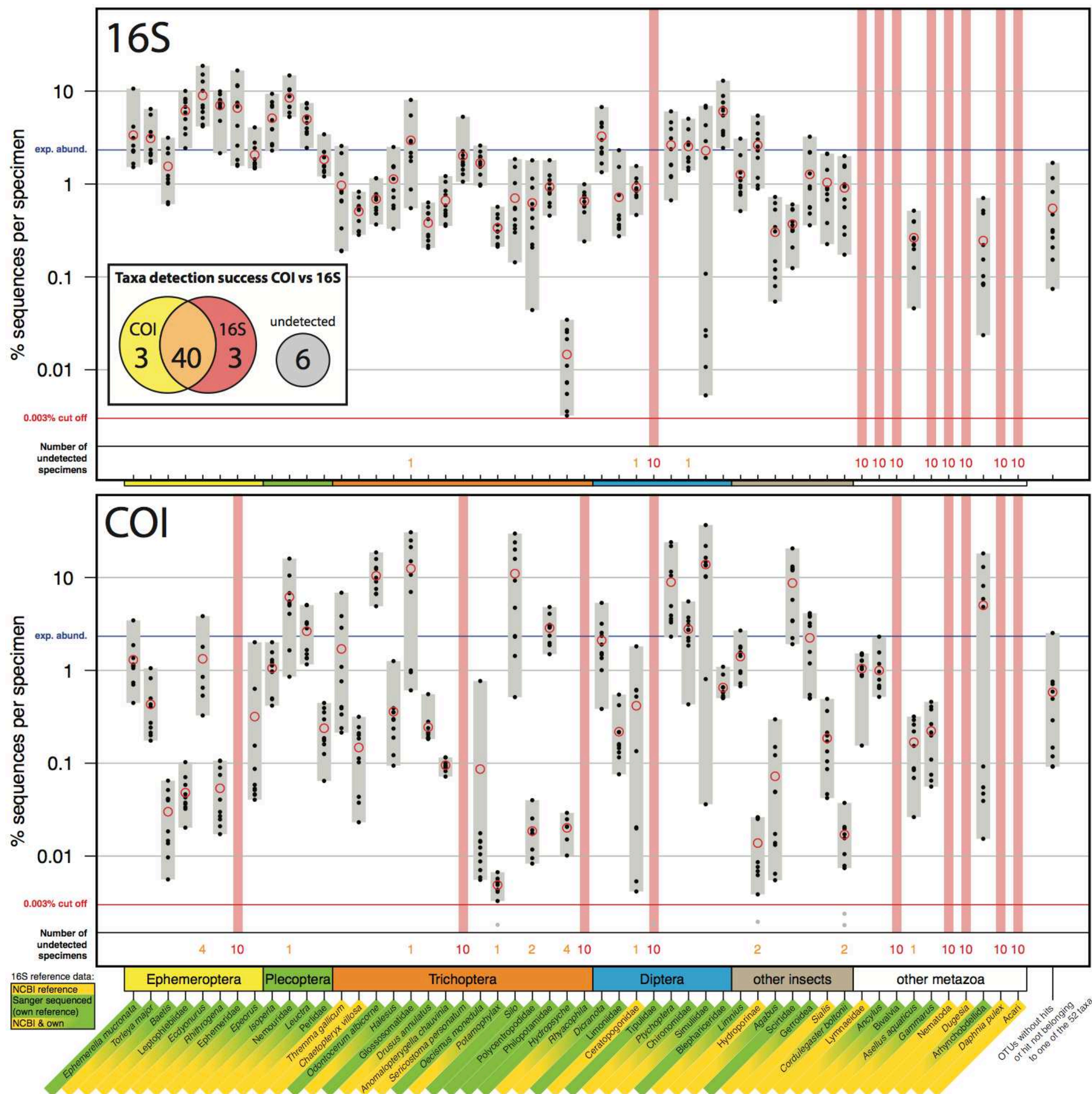
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1855v1 | CC-BY 4.0 Open Access | rec: 13 Mar 2016, publ: 13 Mar 2016

PeerJ Preprints | NOT PEER-REVIEWED

201 **Acknowledgements:** FL and VE are supported by a grant of the Kurt Eberhard Bode foundation to FL.
202 We thank Uwe John and Nancy Kuehne (Alfred Wegener Institute Helmholtz Centre for Marine and Polar
203 Research, Bermerhaven) for running the library on the MiSeq sequencer.
204
205 **Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the
206 content and writing of the paper.
207
208 Authors contributions:
209
210 Experimental design: VE and FL
211 Primer design: Pierre and coauthors
212 Laboratory work: FL and VE
213 Bioinformatics and statistics: VE, PT, EC
214 Written the paper: VE, FL and PT
215

PeerJ Preprints
NOT PEER-REVIEWED

216 # Figures



217

218 **Figure 1**: Comparison of 16S Ins (A) and COI Folmer (B) primer performance, both tested with the same 10 bulk

219 samples each containing 52 morphologically distinct macroinvertebrate taxa. The 52 taxa are shown on the x-axis with

220 the number of reads obtained with 16S and COI for each sample indicated by black dots on the logarithmic y-axis

221 (mean abundance of detected morphotaxa is indicated by red circles). Sequence abundance was normalized across the

222 ten replicates and the amount of tissue used in each DNA extraction. Only OTUs which had minimum abundance of

223 0.003% in at least one of the 10 samples were included in the analysis. Number of morphotaxa which were not detected

224 is indicated by orange and red numbers in each plot. A thick vertical line in light red indicates if a morphotaxa was not

225 detected. Detection rates between 16S and COI marker are summarized in a venn diagram. The availability of 16S

226 reference data from NCBI and own Sanger sequences is indicated by yellow and green background colour behind the

227 taxon names on the x-axis.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1855v1 | CC-BY 4.0 Open Access | rec: 13 Mar 2016, publ: 13 Mar 2016

228

Peer Preprints NOT PEER-REVIEWED

229 **Supplementary information**

230 **Figure S1:** 16S fusion primers used in this study

231 **Figure S2:** Distribution of reads obtained by NextSeq and number of reads discarded throughout the different

232 bioinformatics processing steps.

233

234 **Table S1**: Sequence of each OTU with abundance of assigned reads and assigned taxonomy.

235 **Table S2:** Distribution of OTUs across the 52 taxa.

236 **Table S3:** Raw number of reads assigned to each of the 52 taxa for 16S and COI across the 10 replicates.

237

238 **Scripts S1:** R scripts used in this study to process sequence data and create plots

239

240 # References

241 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL 2009. BLAST+: architecture and
242 applications. *BMC bioinformatics* 10:421.
243 Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA 2013. Environmental monitoring using next generation
244 sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in zoology* 10:1–1.
245 Clarke LJ, Soubrier J, Weyrich LS, Cooper A 2014. Environmental metabarcodes for insects: in silicoPCR reveals
246 potential for taxonomic bias. *Molecular ecology resources* 14:1160–1170.
247 Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P 2014. DNA metabarcoding and the cytochrome c oxidase
248 subunit I marker: not a perfect match. *Biology Letters* 10:20140562–20140562.
249 Dowle EJ, Pochon X, Banks J, Shearer K, Wood SA 2015. Targeted gene enrichment and high throughput sequencing
250 for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular ecology*
251 *resources*:n/a–n/a.
252 Edgar RC 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 10:996–
253 998.
254 Edgar RC, Flyvbjerg H 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads.
255 *Bioinformatics* 31:3476–3482.
256 Elbrecht V, Leese F 2015. Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias
257 and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PloS one* 10:e0130324–16.
258 Elbrecht V, Feld CK, Gies M, Hering D, Sondermann M 2014. Genetic diversity and dispersal potential of the stonefly
259 Dinocras cephalotes in a central European low mountain range. *Freshwater Science* 33:181–192.
260 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R 1994. DNA primers for amplification of mitochondrial cytochrome
261 c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 3:294–299.
262 Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, Hallwachs W, Hajibabaei M 2014.
263 Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through
264 DNA metasystematics. *Proceedings of the National Academy of Sciences* 111:8007–8012.
265 Gómez-Rodríguez C, Crampton-Platt A, Timmermans MJTN, Baselga A, Vogler AP 2015. Validating the power of
266 mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in*
267 *Ecology and Evolution* 6:883–894.
268 Hajibabaei M, Shokralla S, Zhou X, Singer G, Baird DJ 2011. Environmental Barcoding: A Next-Generation
269 Sequencing Approach for Biomonitoring Applications Using River Benthos. *PloS one*.
270 Leray M, Knowlton N 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine
271 benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*:201424997–6.
272 Liu S, Wang X, Xie L, Tan M, Li Z, Su X, Zhang H, Misof B, Kjer KM, Tang M, Niehuis O, Jiang H, Zhou X 2015.
273 Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis.
274 *Molecular ecology resources*.
275 Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL 2013. Practical innovations for high-throughput
276 amplicon sequencing. *Nature Methods* 10:999–1002.
277 Martin M 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17:10–
278 12.
279 Pfrender ME, Hawkins CP, Bagley M, Courtney GW, Creutzburg BR, Epler JH, Fend S, Schindel D, Ferrington LC,
280 Hartzell PL, Jackson S, Larsen DP, Lévesque A, Morse JC, Petersen MJ, Ruiter D, Whiting M 2010. Assessing
281 macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The*
282 *Quarterly review of biology* 85:319–340.
283 Piñol J, Mir G, Gomez-Polo P, Agustí N 2014. Universal and blocking primer mismatches limit the use of high-
284 throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular ecology resources*:1–12.
285 Porter TM, Gibson JF, Shokralla S, Baird DJ, Golding GB, Hajibabaei M 2014. Rapid and accurate taxonomic
286 classification of insect (class Insecta) cytochrome coxidase subunit 1 (COI) DNA barcode sequences using a naïve
287 Bayesian classifier. *Molecular ecology resources* 14:n/a–n/a.
288 Ratnasingham S, Hebert P 2007. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular*
289 *Ecology Notes* 7:355–364.
290 Ratnasingham S, Hebert PDN 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN)
291 System. *PloS one* 8:e66213–16.
292 Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E 2011. ecoPrimers: inference of new DNA barcode
293 markers from whole genome sequence analysis. *Nucleic acids research* 39:e145–e145.
294 Sweeney BW, Battle JM, Jackson JK, Dapkey T 2011. Can DNA barcodes of stream macroinvertebrates improve
295 descriptions of community structure and water quality? *Journal of the North American Benthological Society*
296 30:195–216.
297 Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH 2012. Environmental DNA. *Molecular Ecology* 21:1789–1793.
298 Tang M, Hardman CJ, Ji Y, Meng G, Liu S, Tan M, Yang S, Moss ED, Wang J, Yang C, Bruce C, Nevard T, Potts SG,
299 Zhou X, Yu DW 2015. High-throughput monitoring of wild bee diversity and abundance via mitogenomics.

PeerJ Preprints NOT PEER-REVIEWED
300 *Methods in Ecology and Evolution* 6:1034–1043.
301 Tang M, Tan M, Meng G, Yang S, Su X, Liu S, Song W, Li Y, Wu Q, Zhang A, Zhou X 2014. Multiplex sequencing of
302 pooled mitochondrial genomes-a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic acids*
303 *research* 42:gku917–e166.
304 Taylor HR, Harris WE 2012. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA
305 barcoding. *Molecular ecology resources*:no–no.
306 Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z 2012. Biodiversity soup: metabarcoding of arthropods for
307 rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3:613–623.
308 Zhou X, Adamowicz SJ, Jacobus LM, DeWalt RE, Hebert PD 2009. Towards a comprehensive barcode library for
309 arctic life - Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in zoology*
310 6:30.
311 Zhou X, Jacobus LM, DeWalt RE, Adamowicz SJ, Hebert PDN 2010. Ephemeroptera, Plecoptera, and Trichoptera
312 fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. *Journal of the*
313 *North American Benthological Society* 29:814–837.
314 Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q 2013. Ultra-deep sequencing enables high-
315 fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2:4.
316