



HAL
open science

Molecular biosurveillance of wood-boring cerambycid beetles using DNA metabarcoding

Lois Veillat, Stephane Boyer, Marina Querejeta Coma, Emmanuelle Magnoux,
Alain Roques, Carlos Lopez-Vaamonde, Géraldine Roux

► **To cite this version:**

Lois Veillat, Stephane Boyer, Marina Querejeta Coma, Emmanuelle Magnoux, Alain Roques, et al..
Molecular biosurveillance of wood-boring cerambycid beetles using DNA metabarcoding. 2023. hal-
04169148

HAL Id: hal-04169148

<https://hal.inrae.fr/hal-04169148>

Preprint submitted on 23 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PREPRINT

Author-formatted, not peer-reviewed document posted on 12/07/2023

DOI: <https://doi.org/10.3897/arphapreprints.e109313>

Molecular biosurveillance of wood-boring cerambycid beetles using DNA metabarcoding

Lois Veillat,  Stephane Boyer, Marina Querejeta Coma, Emmanuelle Magnoux,  Alain Roques, 
Carlos Lopez-Vaamonde, Géraldine Roux

1 To be submitted to Metabarcoding & Metagenomics

2

3 **Molecular biosurveillance of wood-boring cerambycid beetles using** 4 **DNA metabarcoding**

5

6

7 Lois Veillat¹, Stéphane Boyer², Marina Querejeta^{2,4}, Emmanuelle Magnoux¹, Alain Roques¹,
8 Carlos Lopez-Vaamonde^{1,2*}, Geraldine Roux^{1,3*}

9

10 ¹INRAE, UR633, Zoologie Forestière, F-45075 Orléans, France

11 ²Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours,
12 France

13 ³Université d'Orléans, rue de Chartres, B.P. 6759, 45067 Orléans cedex 2, France

14 ⁴Department of Functional Biology, University of Oviedo, Oviedo, Asturias, Spain

15

16 Correspondance: Lois Veillat and Géraldine Roux, INRAE, UR633, Zoologie Forestière, F-
17 45075 Orléans, France. lois.veillat@inrae.fr ; geraldine.roux@univ-orleans.fr

18

19 *These authors contributed equally to this work

20

21 **Abstract**

22

23 Individual sorting and identification of thousands of insects collected in mass trapping
24 biosurveillance programs is a labor intensive and time-consuming process. Metabarcoding,
25 which allows for the simultaneous identification of multiple individuals in a single mixed
26 sample, has the potential to expedite this process. However, detecting all the species present
27 in a bulk sample can be challenging. In this study, we quantified the effectiveness of
28 metabarcoding at detecting all species in six different mock communities of xylophagous
29 cerambycid beetles. No significant differences in the number of species detected were
30 observed between MinION, Illumina, and IonTorrent sequencing technologies. However, a
31 greater number of individuals was detected and identified to species using MinION. In
32 addition, the proportion of reads assigned to the species level was higher with Illumina
33 technology. The three sequencing technologies also showed similar results in detecting and
34 identifying closely related species and species at low abundance. The capture method greatly
35 influences sample preservation and detection. Indeed, individuals captured using
36 monopropylene and water had both lower DNA concentration and species detection rates
37 compared to individuals killed using just an insecticide without any collection medium.

38

39 **Keywords:** Biological invasions, Biosecurity, *Cerambycidae*, Illumina®, IonTorrent®,
40 Metabarcoding, Oxford Nanopore®, Xylophagous.

41

42 **Introduction**

43 Over the last few years, there has been an exponential increase in biological invasions that is
44 expected to persist over the next decades (Seebens et al. 2021). This is primarily due to factors
45 such as globalization, tourism, and global warming (Chown et al. 2015). Among the species
46 introduced beyond their native range by human activities, insects are the most prevalent
47 group (Seebens et al. 2018) and can cause a wide range of impacts. Non-native insects can
48 affect native flora, fauna and ecosystems in various ways (Kenis et al. 2009) and can transmit
49 pathogens and diseases, thus posing risks to public health (Mazza et al. 2014). Economic
50 implications are also to be considered since many invasive insects are significant pests for
51 agricultural crops and plantation forests, inducing huge costs for their management (Bradshaw
52 et al. 2016).

53

54 Among these non-native insects, species associated with woody plants are more and more
55 dominating, probably because of the growing trade of ornamental plants and wooden
56 packaging material transported with international cargo shipments (Roques et al. 2016;
57 Aukema et al. 2010). One of these important families of xylophagous beetles is the long-
58 horned Cerambycidae, with more than 200 species affecting forestry, horticulture, and
59 agriculture (Rossa and Goczał 2021) resulting in multimillion-dollar losses every year (Wang
60 2017). To detect potential new invasions of Cerambycids, biomonitoring programs have been
61 set up over large geographical areas with intensive trapping campaigns extending over several
62 years (Roques et al. 2023; Mas et al. 2023). However, rapidly evolving trades lead to changes
63 in trade routes and imported goods which results in an increasing arrival of new non-native
64 species. Many of these species have not been previously reported as invaders, some are not
65 considered to be pests in their native ranges, and some could even be unknown to science
66 (Seebens et al. 2018). As part of the European project HOMED (<https://homed-project.eu/>)
67 244 traps were set up across Europe (France, Italy, Spain, Switzerland, Portugal, Austria,

68 England, Greece, Slovenia, Netherlands, Bulgaria, Czech Republic, and in Sweden), 38 in Asia
69 (China, Siberia, Russia), 11 in North America (USA, Canada), five in the Caribbean (Martinique)
70 and four in Australia baited with generic lures, for simultaneous detection of multiple species
71 (Roques et al. 2023). Thousands of captured cerambycids had to be sorted out and identified
72 by expert taxonomists. This identification step is time-consuming and labor-intensive, thus
73 limiting the rapid detection of non-native individuals among large numbers of native ones
74 (Piper et al. 2019; Chua et al. 2023; Abeynayake et al. 2021). It is, however, essential that non-
75 native species are identified as quickly as possible to allow their eradication before
76 establishment and dispersal (Richardson et al. 2000; Blackburn et al. 2011; Giovani et al. 2020).
77

78 Traditional DNA barcoding, which allows taxonomic assignment of an individual based on the
79 sequencing of a short fragment of the Cytochrome Oxidase 1 (COI) gene (658bp) (Hebert et al.
80 2003) has been successfully used to accurately identify cerambycid pest species for
81 biomonitoring (Hodgetts et al. 2016, Wu et al. 2017, Kelnarova et al. 2019, Javal et al. 2021).
82 Despite its numerous advantages, individual DNA barcoding remains a laborious and time-
83 consuming approach in the context of mass trapped insects as it requires individual sorting of
84 thousands of specimens, tissue sampling (often legs), extracting and amplifying DNA and
85 finally sequencing each sample individually. But the recent application of high-throughput
86 sequencing (HTS) technologies to DNA barcoding allows to expedite the production of
87 thousands of DNA barcodes (deWaard et al. 2019; Srivathsan et al. 2021).

88
89 Metabarcoding is also based on high-throughput sequencing (HTS) technologies which
90 generates a large number of short DNA sequences (reads), allowing the identification of
91 multiple individuals simultaneously from a single mixed sample (hereafter called “bulks”) (Liu
92 et al. 2020), such as all the individual insects captured in a single biomonitoring trap. Using DNA
93 as a proxy for species detection and considering sequence variation within and among taxa,
94 the metabarcoding approach is constrained by the completeness of the reference databases
95 to accurately assign sequences to correctly identified taxa (Liu et al. 2020).
96

97 Although metabarcoding has several advantages, it still suffers from a number of
 98 methodological limitations that can make it unfitting for rapid biosecurity detection. In
 99 particular, the time required to process samples can be an issue (long delays between
 100 capturing individuals and obtaining sequencing results), particularly when the sampling sites
 101 are located far away from laboratories and transporting samples may require specific permits
 102 for certain species or when external providers are slow to sequence samples, which may
 103 hinder biomonitoring projects (Krehenwinkel et al. 2019; Egeter et al. 2022). Although MiSeq
 104 is generally recommended due to its lower error rate and well-established bioinformatic
 105 procedures, but Braukmann et al. (2019) demonstrated similar performance in sequence
 106 quality and insect species recovery compared to IonTorrent platforms (Ion Torrent PGM, and
 107 Ion Torrent S5).

108

109 In recent years, Oxford Nanopore Technologies® have released a portable sequencing
 110 platform, the MinION. This small sequencer can be connected via USB to a laptop to perform
 111 sequencing (Krehenwinkel et al. 2019) and the all-in-one version (Mk1C) even includes a
 112 screen and a computer in a portable format. The use of the MinION for a metabarcoding
 113 application offers the possibility of performing DNA sequencing of bulk samples directly on
 114 site without the need for transport or relying on external sequencing providers.

115

116 The main aim of our study was to assess the use of metabarcoding for the biosurveillance of
 117 Cerambycid wood-boring beetles. We evaluate the capability of three sequencing technologies
 118 Minion Nanopore, Illumina and IonTorrent to differentiate closely related cerambycid species
 119 and detect species present at low abundances in trap samples. Additionally, we examine the
 120 effect of various factors on species detection, such as the collection types (dry versus wet
 121 methods), DNA sample quality/quantity, and primer pair selection.

122

123 **Materials and Methods**

124 *Taxa sampling*

125 Specimens used in our experiments originated from Europe (France, Greece, Portugal, Spain),
 126 China (Beijing and Zhejiang Province) and USA (Michigan) (**Tab. S1**). Nearly all of them were

127 captured as part of a worldwide trapping experiment using multi-funnel traps baited with a
 128 generic attractant blend including eight Cerambycid pheromones (see details of the blend
 129 composition and trapping methods in Roques et al. 2023). To test for the impact of the
 130 collection methods on DNA preservation, we selected specimens killed according to two
 131 different procedures (**TabS S1, S2**). The “dry” procedure involved placing a section of mesh
 132 impregnated with α -cypermethrin insecticide (Storanet®, BASF Pflanzenschutz Deutschland,
 133 Germany) into the trap basins, of which the bottom had been replaced with a wire mesh to
 134 allow drainage and keep specimens dry. In the “wet” procedure, trap basins were filled
 135 monopropylene glycol (MPG) and water (H₂O) in a 50:50 ratio to act as a surfactant and
 136 preservative. The tested specimens were collected between summer 2018 and summer 2021,
 137 with the exception of two specimens that were hand-captured in 1987 and 2012, respectively
 138 (**TabS S1, S2**). Following collection, the beetles were preserved in ethanol 95°C and stored in
 139 at -20°C, except for the two hand-collected specimens, which were dried and pinned in insect
 140 boxes.

141

142 *Mock community construction and DNA extraction*

143 Six mock communities were constructed as follows:

144 Test 1: Identifying closely-related species and the impact of the capture method.

145 To assess the efficiency of the different sequencing technologies and primers to differentiate
 146 between species, bulks 1 and 2 were composed of closely-related species (**Tab. S1**). Five of
 147 seven specimens in Bulk 1 were captured using the "dry" method, while the eight of nine of
 148 specimens in Bulk 2 were captured using the "wet" method. One (bulk 2) or two (bulk 1)
 149 exceptions (wet *versus* dry) condition were added to the bulks as controls. Two legs were
 150 collected from each individual and pooled to constitute the bulks. The whole set of legs was
 151 then ground using flame-sterilized metal pestles to limit the risk of contamination. DNA from
 152 the ground material was extracted using the Qiagen DNeasy Blood and Tissue Kit following the
 153 manufacturer’s instructions. Two other legs were taken from the specimens, from which we
 154 assessed the quantity and quality of DNA for each specimen individually (**Fig. 1a**).

155

156 Test 2: Detecting low abundance specimens.

157 Bults 3 and 4 were composed of six species represented by heterogeneous DNA
 158 concentrations (**Tab. S1**) to assess the ability of the sequencing technologies and primers to
 159 detect species present in a very low abundance. The DNA of each individual was extracted
 160 using two legs that were ground as above and processed using the Qiagen DNeasy Blood and
 161 Tissue Kit. To construct bults 3 and 4, individual DNA extracts were quantified using a
 162 fluorometer (Nanodrop™, Thermo Fisher Scientific) and mixed together according to their
 163 concentration to achieve the desired proportions of DNA for each individual (6 individuals of
 164 different species ranging from 41% to 3% for Bulk 3 and 6 individuals of different species
 165 ranging from 50% to 0.5% for bulk 4). All individuals in bults 3 and 4 were captured using the
 166 "dry" collection method. (**Tab. S1, Fig. 1b**).

167

168 Test 3: Mimicking field trap content on species composition.

169 Bults 5 and 6 comprised individuals from a number of species native to Europe usually found
 170 in the traps deployed there, with the addition of non-native species which have either already
 171 been introduced or are still not present in Europe (Bulk 5: 22 individuals of eight species,
 172 including two older specimens from 2012 and 1987 and including one non-native; Bulk 6: 41
 173 individuals of 12 species including two non-native ones). Six individuals in bulk 5 were captured
 174 using the "dry" collection method and two individuals were hand-collected. Two specimens
 175 were captured using the "wet" method in bulk 6 (**Tab. S1**). The DNA was extracted following
 176 the same protocol as for bults 1 and 2 (**Fig. 1a**).

177

178 *PCR amplification*

179 All bulk samples were amplified with the following two primer pairs: BF3/BR2 (call hereafter
 180 "B") (CCHGAYATRGCHTTYCCHCG / TCDGGRTGNCCRAARAAYCA (Elbrecht and Leese 2017;
 181 Elbrecht et al. 2019), which generates a 458 bp amplicon; and fwhF2/fwhR2n (call hereafter
 182 "F") (GGDACWGGWTGAACWGTWTAYCCHCC / GTRATWGCHCCDGCTARWACWGG), which
 183 generates a 254 bp amplicon (Vamos et al. 2017). Each PCR comprised 15.3 µl H₂O, 2.5 µl 10X
 184 PCR buffer, 2.5 µl dNTP [1mM], 1 µl of each primer [0,4mM], 0.2 µl Dream Taq (Thermo Fisher
 185 Scientific), 0.5 µl Betaine [100mM], 2 µl DNA and H₂O for a total of 25 µl per reaction. PCR was
 186 performed using the following program: 95°C for 5 min, 29 cycles of 95°C for 30 s, 48°C for 30

187 s, and 72°C for 50 s and 72°C for 5 min (Elbrecht et al. 2019). PCR products were then run on
 188 a 2% agarose gel stained with ethidium bromide and visualised by UV transilluminator. The PCR
 189 products were then purified with the NucleoFast 96 PCR plate purification kit (Macherey-
 190 Nagel).

191

192 *Illumina® library preparation*

193 A second PCR was performed on the products of the first PCR to add Illumina® tags and
 194 adapters, prepared by ligating Nextera XT indices through an eight cycle PCR. The second PCR
 195 was carried out with the same conditions as for the initial PCR. Reactions comprised PCR
 196 amplification reactions (25 µl) contained the following: 2 µl of template DNA, 1.5 µl of each
 197 primer [10 µM], 5 µl of 5X GoTaq (Promega) reaction buffer, 1 µl of MgCl₂ [25 mM], 1 µl of BSA
 198 [1 mg/ml], 0.5 µl of dNTPs [5 mM], 13.87 µl of molecular-grade water and 0.13 µl of GoTaq G2
 199 Polymerase (Promega), 5 µl of the purified products from the first PCR, and 2 µl of barcodes.
 200 The PCR conditions were the same as for the first PCR, with eight cycles. The products of the
 201 second PCR were verified on a 2% agarose gel. PCR products were then equimolarly pooled
 202 into two different pools (one pool per primer pair used) and purified using the GeneJET Gel
 203 Extraction kit on an agarose gel, following manufacturer's instructions. This library was
 204 sequenced in Illumina MiSeq using V2 chemistry (300 × 300 bp, 500 cycles) in the Sequencing
 205 Center within the Biozentrum of the Ludwig-Maximilian University in Munich (Germany).

206

207 *MinION library preparation*

208 Libraries were prepared according to the Oxford Nanopore Technologies® protocol: "PCR
 209 barcoding (96) amplicons (SQK-LSK110) (version: PBAC96_9114_v110_revF_10Nov2020)"
 210 with the following specifications. The PCR barcoding expansion Pack 1-96 (EXP-PBC096) was
 211 used to perform the second PCR to incorporate the Oxford Nanopore Technologies® barcode
 212 sequences on the amplicons generated in the first PCR. Final PCR products were then
 213 quantified using Qubit and equimolarly pooled before being purified with Agencourt AMPure
 214 XP beads (Beckam Coutler). The final pool was then sequenced on the MinION sequencer
 215 (Mk1c; Oxford Nanopore Technologies®, UK) using a R10.3 flowcell (MIN111) with 1331 pores
 216 available and the LSK110 ligation sequencing kit.

217

218 *IonTorrent® library preparation*

219 For the production of the libraries, we started with 5 ng of DNA (Qubit measurement). The
 220 Nextflex Cellfree DNaseq kit (PerkinElmer) was used for the process. The quality of the libraries
 221 was assessed using Qubit (for quantification) and Bioanalyzer (using the HighSensitivity kit
 222 from Agilent, for size verification). After quality control, each library was amplified by emulsion
 223 PCR on the Ion One Touch 2 instrument, with a concentration of 15 pg/μl. Subsequently, the
 224 libraries were sequenced on an Ion GeneStudio S5 system using a single-end sequencing
 225 protocol with a 300 bp read length. Sequencing was performed on an Ion 520 Chip by the GeT-
 226 BioPuces platform (Toulouse, France).

227

228 *Reference Barcode Dataset*

229 A dataset was built using all the public sequences of Cerambycidae available in BOLD systems
 230 (Ratnasingham and Hebert 2007). It was then verified whether all 33 species present in the
 231 bulk samples were represented by at least one sequence in the database. We found that three
 232 species were not present in BOLD and we therefore barcoded them by Sanger sequencing on
 233 an ABI 3500 genetic analyzer (Applied Biosystems) using the big-dye terminator sequencing
 234 V3.0 kit (Applied Biosystems). The three newly generated barcodes along with one barcode
 235 per species represented in the mock communities are available from BOLD in the dataset DS-
 236 MINION ([dx.doi.org/XX/ DS-MINION](https://dx.doi.org/10.26434/chemrxiv-2023-10)).

237

238 *Illumina® data processing*

239 The raw data was analysed using the *FROGS* pipeline, a standardized pipeline containing a set
 240 of tools that are used to process amplicon reads that have been produced from Illumina®
 241 sequencing (Escudié et al. 2018; Henrie et al. 2022). First, amplicons with a size between 408
 242 and 508 for the BF3/BR2 primer pair and 204 and 304 for the fwhF2/fwhR2n primer pair were
 243 retained. Sequence clustering was then performed using the SWARM algorithm (Mahé et al.
 244 2014) with a maximum sequence difference set at $d=1$, as recommended by SWARM. Chimeric
 245 sequences were then removed. Sequences were aligned to the same database used for the
 246 MinION and IonTorrent® data analysis. In order to remove all spurious detections, OTU

247 detections with less than 10 reads were removed. The identification was considered as ‘valid’
 248 at species-level from a similarity threshold $\geq 98\%$ (Alberdi et al. 2018). Below that threshold,
 249 OTUs were considered unidentified.

250

251 *MinION and IonTorrent® data processing*

252 Bioinformatics analyses were performed on the Genotoul Bioinformatics Platform (INRAE,
 253 Toulouse, France). Basecalling and demultiplexing were performed for MinION data using
 254 Guppy v6.1.7; ONT; high accuracy base calling mode; parameters: -c
 255 dna_r10.3_450bps_hac.cfg --min_qscore 5 --trim_barcodes. Then, for MinION and
 256 IonTorrent® data, we used the *msi* data processing pipeline v0.3.6 (Egeter et al. 2022) to
 257 reduce the error rate of the reads by polishing them after the basecalling step. Reads smaller
 258 than 40bp were removed with cutadapt v4.0 (Martin 2011). The size range was set between
 259 408bp and 508bp for BF3/BR2 and between 204bp and 304bp for fwhF2/fwhR2n. The
 260 clustering step was carried out with *ISONCLUST* v0.0.6.1 (Sahlin and Medvedev 2020; with
 261 parameters: --mapped_threshold 0.825 and --aligned_threshold 0.55) and a consensus
 262 sequence per cluster was generated using RACON v1.5.0 (Vaser et al. 2017). The polished reads
 263 were then clustered at 97% sequence identity with *CD-HIT* v4.8.1 (Fu et al. 2012) and a
 264 representative sequence from each cluster (centroid) was selected. The polished reads were
 265 then aligned to the local database with BLAST. The following parameters were used: -word_size
 266 11 -perc_identity 95 -qcov_hsp_perc 98 -gapopen 0 -gapextend 2 -reward 1 -penalty 1 -
 267 max_target_seqs 100. Finally, a taxonomic assignment was performed for each query using a
 268 lowest common ancestor (LCA) approach with the bioinformatics package metabinkit (Chain
 269 et al. 2016; Egeter et al. 2018 Kitson et al. 2019) with the following parameters: 98% at species
 270 level, 97% at genus level, 95% at family level (Alberdi et al. 2018; Egeter et al. 2022). Similarly,
 271 to the Illumina® data processing, OTU detections with less than 10 reads were removed.

272

273 *Statistical analysis*

274 The Wilcoxon test was used in R to assess the significance of the impact of the conservation
 275 method (dry or wet) on sample DNA concentration, sample DNA quality, and associated
 276 detection counts. A two-sample test of proportions was used to compare and assess the
 277 significance of the proportion of reads assigned to the species levels for MinION, Illumina, and

278 IonTorrent technologies using the "Social Science Statistics" website
 279 (<https://www.socscistatistics.com/tests/anova/default2.aspx>). The proportion of reads
 280 assigned to different taxonomic levels was calculated by summing the total reads from
 281 different bulk samples for each condition. To determine if the number of false positives was
 282 significantly different among the 3 technologies and the two primer pairs, we calculated the
 283 detection mean for each bulk under different conditions. We then performed an ANOVA test
 284 followed by a Tukey HSD test using the "Social Science Statistics" website.

285

286 **Results**

287 A total of 1.248.95 reads were sequenced for the MinION Nanopore® technology using the
 288 fwhF2/fwhR2n primer pair, with an average of 78.037 reads per sample. After quality filtering,
 289 removing reads of incorrect size or insufficient quality, 1.113.844 (89.2%) reads were retained,
 290 with an average of 69.615 reads per bulk. For the BF3/BR2 primer pair, a total of 1.132.604
 291 reads were sequenced, with an average of 62.922 reads per sample. After quality filtering, a
 292 total of 948.832 (83.8%) reads were retained, with an average of 52.712 reads per bulk. The
 293 Illumina® sequencing produced a total of 1.549.894 reads using the BF3/BR2 primer pair, with
 294 an average of 258.316 reads per bulk. After quality filtering, 1.025.637 (66.2%) reads were
 295 retained, with an average of 170.940 reads per bulk. For the fwhF2/fwhR2n primer pair, a total
 296 of 2.299.072 reads were sequenced, with an average of 383.179 reads per bulk. After quality
 297 filtering, 1.686.058 (73.3%) reads were retained, with an average of 281.010 reads per bulk.
 298 For the IonTorrent® technology, 838.489 reads were sequenced, with an average of 139.748
 299 reads per bulks with the BF3/BR2 primer pair. After the quality filtering, 280.695 (33.5%) reads
 300 remains with an average of 46.782 reads per bulks.

301

302 *Benchmarking of sequencing technologies*

303 The MinION technology accurately identified 28 out of 48 specimens at the species level,
 304 Illumina® technology allowed specific identification of 27 specimens and IonTorrent®
 305 identified 24 specimens. The primer pair fwhF2/fwhR2n allowed specific identification of 27
 306 specimens with two OTU of its own while the primer pair BF3/BR2 allowed the identification
 307 of 31 specimens with six OTU of its own. Illumina® F, Illumina® B and MinION F allowed for 25

308 species-level identifications across all bulks, compared to 24 when considering the
 309 combination MinION F and IonTorrent® B. This difference was not significant. The number of
 310 identifications only obtained with one combination varied from zero (IonTorrent® B) to two
 311 (Minion B and Illumina® B) (**Fig. 2**).

312
 313 The proportion of reads assigned at the species taxonomic level was significantly higher with
 314 Illumina® technology (p.value < 0.00001) comparing with the MinION, all primers included,
 315 particularly when considering primer pair F. Nearly 97% of reads were assigned at the species
 316 level for the Illumina® F combination compared to 90% for the MinION F combination
 317 (p.value<0.0001). As for primer pair BF3/BR2, over 87.3% of reads were assigned at the species
 318 level for Illumina®, followed by over 79.7% for MinION technology and over 77.2% for
 319 IonTorrent® technology (**Fig. 3**). The couple of primer fwhF2/fwhR2n resulted in a significantly
 320 higher percentage of reads assigned at the species level (93.6%) (considering both Illumina®
 321 and MinION technologies) compared to couple of primers B (81.4%) (considering all three
 322 technologies) (p.value<0.00001).

323
 324 False positive detections were observed for each combination of primers and technology (**Fig.**
 325 **4**). Hence, an average of 13.5 false positives OTU were recorded for the primer pair
 326 fwhF2/fwhR2n, compared to an average of 4 false positives OTU when using the primer pair
 327 BF3/BR2, the difference being significant here (p.value = 0.00194). According to the technology
 328 used, regardless of the primers, an average of 10 / 7 and 6 false positives were recorded for
 329 Illumina®/MinION/IonTorrent® technologies respectively. There are no significant differences
 330 among the three sequencing technologies in terms of false positives.

331
 332 *Comparative study within bulks*

333 In total, 33 out of 48 individuals (68.8%) were detected at the species level by at least one
 334 experimental condition (**Fig. 5**).

335
 336 The objective of Bulks 1 and 2 was to compare the detection rates of closely related species
 337 under different conditions. No significant differences were observed among the different

338 methods used. Illumina® detected seven species out of 16 (43.75%), MinION also detected
 339 seven out of 16 (43.75%), and IonTorrent® detected six species out of 16 (37.5%). However,
 340 significant differences were observed among the studied taxonomic groups: three out of four
 341 species from the genus *Arhopalus* were detected, as well as for *Xylotrechus*. Only two out of
 342 four species were detected for the genus *Monochamus*, and one out of five species for the
 343 genus *Phymatodes*.

344

345 Metabarcoding of bulks 3 and 4 aimed at comparing the ability of different sequencing
 346 technologies to detect species present in low abundance in the traps. All sequencing
 347 technology/primer combinations allowed for the detection of minor species in bulks 3 and 4:
 348 *Phymatodes testaceus* with a presence of 3% in bulk 3 and *Xylotrechus chinensis* with a
 349 percentage of 0.5% in bulk 4. However, some species (although not in minority in the bulks)
 350 were not detected in one or several test (**Fig. 5**). In total, Illumina® was able to detect a higher
 351 number of individuals (11 out of 12 individuals detected) compared to MinION (nine of 12
 352 individuals detected) and IonTorrent® (nine of 12 individuals detected).

353

354 In bulk 5, the non-native species, *Cordylomera spinicornis* was detected only by Illumina B. For
 355 bulk 6, the non-native species *Xylotrechus chinensis* was detected by all conditions and
 356 *Xylotrechus stebbingi* by MinION B only. The results showed that MinION performs better to
 357 detect and identify trapped species (detecting eight out of 12 species (66.7%)) compared to
 358 Illumina® and IonTorrent® technologies (which detected five out of 12 species (41.7%)) for Bulk
 359 6. As for Bulk 5, all technologies detected the same number of species (four out of six (66.7%)).

360

361 *Impact of capture and storage conditions on individual detection*

362 Our results demonstrate a significative difference in the mean number of detections between
 363 samples that were collected using the “dry” method (α -cypermethrin insecticide) and the
 364 “wet” method (water-diluted propylene glycol) (Wilcoxon rank-sum test, $W = 74.5$, $p.value =$
 365 0.0006342) (**Fig. 6**). Almost all specimens (9 out of 12 - 75%) collected using the “wet” trapping
 366 procedure (water-diluted propylene glycol) could not be detected by any of the sequencing
 367 technologies, including MinION. Conversely, those collected using the “dry” trapping

368 procedure (using α -cypermethrin insecticide) had 30 out of 34 detected specimens (88.2%),
 369 by each of the sequencing technologies.

370

371 Regarding the impact of the collection type on the concentration and quality of DNA samples,
 372 individuals captured using the "dry" method had an average DNA concentration of 39 ng/ μ l
 373 and an average DNA concentration of 18.6 ng/ μ l for the "wet" method (Wilcoxon rank-sum
 374 test, $W = 123.5$, $p.value = 0.04533$). We observe an average A260/280 ratio of 1.9 for the "dry"
 375 method and an average of 2 for "wet" method. Finally, we observe an average A260/230 ratio
 376 of 0.8 for the "dry" method, while samples captured using the "wet" approach had an average
 377 A260/230 ratio of 0.5 (Wilcoxon rank-sum test, $W = 146$, $p.value = 0.1502$) (Fig. 6).

378

379 The specimens collected by hand and kept pinned dry were not detected by any of the
 380 sequencing technologies. Species with a "dry" collection type trapped in 2021 have a detection
 381 rate of 80% (detected by at least one condition), those trapped in 2020 have a detection rate
 382 of 86.7%, and those from 2019 have a detection rate of 100%. Both, individuals captured in
 383 2018 and 2010 were detected, unlike those captured in 1987 and 2012.

384

385 Discussion

386 *Benchmarking of sequencing technologies*

387 No significant differences in the number of species detected were observed between the three
 388 sequencing technologies, even if a greater number of individuals was detected and identified
 389 to species using MinION (28 specimens) compared to Illumina® (27 specimens) or IonTorrent®
 390 (24 specimens). These results are consistent with the findings of Srivathsan et al. (2021), which
 391 demonstrated that MinION barcodes are nearly identical to Sanger and Illumina barcodes for
 392 the same samples. It must be taken into account that we worked on a single pair of primers
 393 with the IonTorrent® technology, which may have reduced the number of detections.
 394 Detection rate was higher when using the primer pair generating a larger amplicon size,
 395 BF3/BR2. In addition, only BF3/BR2 allowed the species-level identification of *Xylotrechus*
 396 *stebbingi*. This difference may be due to the longer amplicon generated by this primer pair

397 (458bp), which has more genetic information and therefore more data to provide a reliable
 398 taxonomic assignment.

399

400 The Illumina® technology has a higher percentage of reads allowing species-level
 401 identification compared to MinION or IonTorrent®. The detection of specimens at only higher
 402 taxonomic level (genus or family), can be explained by sequencing errors that produce reads
 403 with less than 98% identity to the reference database. Thus our results suggest that Illumina®
 404 generates slightly less sequencing errors than MinION and IonTorrent®. It is also noted that
 405 the primer pair fwhF2/fwhR2n produced a higher percentage of reads allowing species-level
 406 identification than the primer pair BF3/BR2. However, as seen previously, fwhF2/fwhR2n
 407 generated a significant higher number of false positives than BF3/BR2 (**Fig. 4**). This may be due
 408 to the fact that fwhF2/fwhR2n is smaller in size than BF3/BR2 thus any loss of genetic
 409 information is more likely to lead to misidentification or false negatives.

410

411 The three technologies also showed similar results in detecting and identifying closely related
 412 species. Moreover, the results show that all three sequencing technologies (regardless of the
 413 associated primer pairs) enabled the detection and identification of species whose DNA
 414 represented a very low percentage of the DNA extraction (**Fig. 5**). Thus, all three technologies
 415 appear suitable for detecting and identifying species in low number in traps.

416

417 *Impacts of capture and storage conditions on DNA conservation*

418 Our results show an impact of sample capture conditions. Individuals captured using the “wet”
 419 method had lower DNA concentration and presented significantly much lower detection rates
 420 compared to individuals captured using the “dry” method. Ballare *et al.* (2019) also found that
 421 insects collected in propylene glycol traps produced lower-quality ddRADseq assemblages. On
 422 the contrary, Ferro and Park, 2013 reported that propylene glycol is an effective DNA
 423 preservative for molecular marker-based studies. However, in mentioned study, the insects
 424 were initially killed and preserved in 100% ethanol before being stored to glycol, unlike our
 425 study where the insects were initially killed in propylene glycol. The use of 100% ethanol as
 426 the initial agent for destruction and preservation may have resulted in better initial

427 preservation of specimens than if the specimens had been directly exposed to propylene
428 glycol.

429

430 The conservation method may also play a role in species detection, as the two specimens
431 conserved using the “pinned dry” method were not detected. However, we have insufficient
432 “pinned dry” specimens in this study to confirm this hypothesis. But numerous studies have
433 already demonstrated that it is possible to use dry insect specimens for genetic analyses,
434 although such types of analyses are much more complex than when using fresh specimens
435 (Wandeler et al. 2007; Hebert et al. 2013; Nakahama et al. 2018).

436

437 The storage duration does not appear to significantly impact species detection. We did not
438 observe a decrease in detection rates between individuals collected after 2018. However, the
439 impact of storage duration seems more pronounced for older individuals (collected between
440 1987 and 2012). Once again, we have limited data to draw solid conclusions regarding these
441 older samples although it is known that the storage time plays an important role in DNA yield,
442 fragment size, and PCR success (Dean and Ballard 2001).

443

444 Ultimately, the DNA concentration of the samples appears to play an important role in their
445 detection as DNA extracts with higher concentration were more likely to be detected (DNA
446 concentration of 38.6 ng/μl for detected samples compared to 19.6 ng/μl for undetected
447 samples). The quality ratios A260/230 and A260/280 are similar between detected and
448 undetected samples, suggesting they do not contribute to the observed detection rates in our
449 analysis.

450

451 *False positives, negatives and unmatched OTUs*

452 Despite precautions taken, several false positives were detected for all tested conditions. The
453 number of false positives is significantly higher with the primer pair fwhF2/fwhR2n, which is
454 smaller in size, potentially leading to incorrect taxonomic identifications compared to
455 BF3/BR2, which is larger in size. Despite the fact that Illumina technology is known to have a
456 lower sequencing error rate compared to MinION, our study found ten false positives

457 generated by Illumina, while MinION had seven false positives and IonTorrent had six false
 458 positives. The sensitivity of HTS technologies allows for the detection of very small amounts
 459 of DNA, thus detecting even the slightest cross contamination between samples (Liu et al.
 460 2020). These contaminations may have occurred during sample collection in the field or in the
 461 laboratory through cross-contamination between samples from the same study.

462

463 The false negative detections for a certain number of individuals can primarily be explained by
 464 the highly heterogeneous DNA quality of the different sequenced individuals (**Tab. S1**). In fact,
 465 DNA quality can be impacted by numerous mainly abiotic factors (pH, UV radiation,
 466 temperature), degrading DNA quality in a matter of days/weeks (Strickler et al. 2015; Collins
 467 et al. 2018; Harrison et al. 2019). During field trapping using stationary traps, captured insects
 468 are sometimes exposed to such conditions (high temperatures in trap containers when
 469 exposed to the sun in summer, high humidity in the container during heavy rains, ...), which
 470 can greatly accelerate the speed of DNA degradation in captured individuals. This degraded
 471 DNA is more difficult to be amplified, thus generating false negatives, especially when
 472 attempting to detect low-abundance insects in the trap, such as an invasive species in the
 473 process of establishing (Preston et al. 2022). Another possible cause for the high number of
 474 false negatives is the bias induced by PCR, such as uneven amplification of the DNA of the
 475 different individuals present in one sample (Preston et al. 2022). To avoid potential bias arising
 476 from identification mistakes due to errors or missing species in the references databases, we
 477 decided to work on a local and curated BLAST database. However, when target species are
 478 partially unknown, as is the case in field conditions, analyses must rely on public reference
 479 databases. Yet, out of the 35,000 known species of Cerambycidae to date, only 2,884 species
 480 (8%) are recorded in BOLD with a barcode fragment (as of April 17th, 2023). Furthermore, as
 481 mentioned above, databases can contain errors such as misassignment of an DNA sequence
 482 to a wrong species (due to morphological identification errors). This was precisely the error
 483 encountered for the species *Monochamus sutor* who has been identified as *Monochamus*
 484 *sartor* or the species *Leiopus nebulosus* who has been identified as *Leiopus linnei*.

485 Morphological similarities or identifications from non-specialists can lead to errors in
 486 databases, hindering their identification at species level. One also needs to pay attention to

487 synonymy when a species is called by multiple names. We encountered this problem in our
488 analysis where *Arhopalus fesus* (Bulks 1, 3, and 6) was detected but under the name of
489 *Arhopalus tristis*. Finally, mitochondrial paralogues such as Numts (non-functional copies of
490 mitochondrial genes transported into the nuclear genome) present in databases can also bias
491 results, making it impossible to identify correctly the species concerned (Bensasson et al.
492 2001). Numts are numerous in many organisms, including some cerambycids such as
493 *Monochamus galloprovincialis* (Koutroumpa et al. 2009; Haran et al. 2015).

494

495 *Biases*

496 Based on the results obtained, it appears that the main biases observed during metabarcoding
497 analyses on trap contents come from the degradation of DNA from individuals, which
498 generates false negatives. To limit the biases induced by the degradation of DNA samples, it is
499 important, when possible, to favour a "dry" rather than a "wet" trap and to plan for the
500 collection, transportation, and processing of captured individuals. This includes regularly
501 checking the traps as frequently as possible, thus avoiding excessively long exposure of the
502 individuals to unfavourable environmental conditions. Once individuals are brought back to
503 the laboratory and if DNA cannot be extracted straight away, it is important to limit any further
504 degradation by keeping samples in a -20°C freezer in 95% ethanol. DNA extractions, on the
505 other hand, should be stored in the preservation buffer provided with the extraction kits and
506 kept at -20°C (Preston et al. 2022). We also recommend limiting the use of primer pairs that
507 generate short amplicons, which can favour the amplification of non-target taxa and lead to
508 identification errors. The quality and completeness of the databases is also a very important
509 bias factor. To limit this bias, it is recommended to restrict the database used to targeted
510 species in order to minimize the risk of false positives due to contamination (Egeter et al.
511 2022). Limited taxonomic and geographical coverage of sequence databases is a huge
512 limitation in metabarcoding studies (for example, Dopheide et al. (2019) found no
513 representative sequence in the GenBank database for more than 900 invertebrate OTUs in one
514 study). Additionally, identification errors of species and cases of synonymy lead to false
515 negatives or cases of multiple affiliations.

516

517

518 **Conclusion**

519 By comparing the accuracy and detection capacity of three metabarcoding strategies, this
520 study contributes to improve our toolkit for insect invasion monitoring. All three sequencing
521 technologies performed equally well and showed similar results for the detection and
522 identification of invasives species in the traps, but as a portable, easy-to-use and cost-effective
523 sequencer, the MinION has the potential to become an essential tool for biodiversity
524 monitoring projects. Indeed, using the MinION saves laboratory handling time compared with
525 Illumina and avoids outsourcing sample sequencing, saving considerable time. This technology
526 is precise enough to detect species present at low abundances in traps and allows for accurate
527 identifications as long as there is a sufficiently high-quality reference database to avoid
528 identification errors or false positives/negatives. It is also crucial to pay close attention to
529 issues of contamination and insect preservation during and after individual capture to work
530 with the least degraded DNA possible.

531

532 **References**

533 Abeynayake SW, Fiorito S, Dinsdale A, Whattam M, Crowe B, Sparks K, Campbell PR, Gambley
534 C (2021) A Rapid and Cost-Effective Identification of Invertebrate Pests at the Borders Using
535 MinION Sequencing of DNA Barcodes. *Genes* 12: 1138.
536 <https://doi.org/10.3390/genes12081138>

537

538 Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable
539 metabarcoding of environmental samples. Mahon A (Ed.). *Methods in Ecology and Evolution*
540 9: 134–147. <https://doi.org/10.1111/2041-210X.12849>

541

542 Aukema JE, McCullough DG, Von Holle B, Liebhold AM, Britton K, Frankel SJ (2010) Historical
543 Accumulation of Nonindigenous Forest Pests in the Continental United States. *BioScience* 60:
544 886–897. <https://doi.org/10.1525/bio.2010.60.11.5>

545

- 546 Ballare KM, Pope NS, Castilla AR, Cusser S, Metz RP, Jha S (2019) Utilizing field collected insects
547 for next generation sequencing: Effects of sampling, storage, and DNA extraction methods.
548 Ecology and Evolution 9: 13690–13705. <https://doi.org/10.1002/ece3.5756>
549
- 550 Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's
551 misplaced witnesses. Trends in Ecology & Evolution 16: 314–321.
552 [https://doi.org/10.1016/s0169-5347\(01\)02151-6](https://doi.org/10.1016/s0169-5347(01)02151-6)
553
- 554 Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM
555 (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution
556 26: 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
557
- 558 Bradshaw CJA, Leroy B, Bellard C, Roiz D, Albert C, Fournier A, Barbet-Massin M, Salles J-M,
559 Simard F, Courchamp F (2016) Massive yet grossly underestimated global costs of invasive
560 insects. Nature Communications 7: 12986. <https://doi.org/10.1038/ncomms12986>
561
- 562 Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, De Waard
563 JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock
564 community. Molecular Ecology Resources 19: 711–727. [https://doi.org/10.1111/1755-
565 0998.13008](https://doi.org/10.1111/1755-0998.13008)
566
- 567 Chain FJJ, Brown EA, Maclsaac HJ, Cristescu ME (2016) Metabarcoding reveals strong spatial
568 structure and temporal turnover of zooplankton communities among marine and freshwater
569 ports. Cowie R (Ed.). Diversity and Distributions 22: 493–504.
570 <https://doi.org/10.1111/ddi.12427>
571
- 572 Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA (2015) Biological
573 invasions, climate change and genomics. Evolutionary Applications 8: 23–46.
574 <https://doi.org/10.1111/eva.12234>
575

- 576 Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN (2023)
577 Future of DNA-based insect monitoring. *Trends in Genetics* 39: 531–544.
578 <https://doi.org/10.1016/j.tig.2023.02.012>
579
- 580 Collins RA, Wangensteen OS, O’Gorman EJ, Mariani S, Sims DW, Genner MJ (2018) Persistence
581 of environmental DNA in marine systems. *Communications Biology* 1: 185.
582 <https://doi.org/10.1038/s42003-018-0192-6>
583
- 584 Dean MD, Ballard JWO (2001) Factors affecting mitochondrial DNA quality from museum
585 preserved *Drosophila simulans*. *Entomologia Experimentalis et Applicata* 98: 279–283.
586 <https://doi.org/10.1046/j.1570-7458.2001.00784.x>
587
- 588 deWaard JR, Levesque-Beaudin V, deWaard SL, Ivanova NV, McKeown JTA, Miskie R, Naik S,
589 Perez KHJ, Ratnasingham S, Sobel CN, Sones JE, Steinke C, Telfer AC, Young AD, Young MR,
590 Zakharov EV, Hebert PDN (2019) Expedited assessment of terrestrial arthropod diversity by
591 coupling Malaise traps with DNA barcoding 1. *Genome* 62: 85–95.
592 <https://doi.org/10.1139/gen-2018-0093>
593
- 594 Dopheide A, Tooman LK, Grosser S, Agabiti B, Rhode B, Xie D, Stevens MI, Nelson N, Buckley
595 TR, Drummond AJ, Newcomb RD (2019) Estimating the biodiversity of terrestrial invertebrates
596 on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications* 29:
597 1–14. Available from: <https://www.jstor.org/stable/26669248> (June 9, 2023).
598
- 599 Egeter B, Peixoto S, Brito JC, Jarman S, Puppo P, Velo-Antón G (2018) Challenges for assessing
600 vertebrate diversity in turbid Saharan water-bodies using environmental DNA. *Genome* 61:
601 807–814. <https://doi.org/10.1139/gen-2018-0071>
602
- 603 Egeter B, Veríssimo J, Lopes-Lima M, Chaves C, Pinto J, Riccardi N, Beja P, Fonseca NA (2022)
604 Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore

605 and Illumina sequencing comparison. *Molecular Ecology Resources* 22: 2232–2247.
606 <https://doi.org/10.1111/1755-0998.13610>

607

608 Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for
609 Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science* 5. Available
610 from: <https://www.frontiersin.org/articles/10.3389/fenvs.2017.00011> (July 2, 2023).

611

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

615

616 Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet
617 G, Combes S, Pascal G (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution. Berger B (Ed.).
618 *Bioinformatics* 34: 1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>

619

620 Ferro ML, Park J-S (2013) Effect of Propylene Glycol Concentration on Mid-Term DNA
621 Preservation of Coleoptera. In: *The Coleopterists Bulletin*. , 581–586.
622 <https://doi.org/10.1649/0010-065X-67.4.581>

623

624 Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation
625 sequencing data. *Bioinformatics* (Oxford, England) 28: 3150–3152.
626 <https://doi.org/10.1093/bioinformatics/bts565>

627

628 Giovani B, Blümel S, Lopian R, Teulon D, Bloem S, Galeano Martínez C, Beltrán Montoya C, Urias
629 Morales CR, Dharmapuri S, Timote V, Horn N, Chouibani M, Mezui M'Ella JG, Herrera V, Castinel
630 A, Goletsos C, Moeller C, Naumann I, Stancanelli G, Bronzwaer S, Tramontini S, MacDonald P,
631 Matheson L, Anthoine G, De Jonghe K, Schenk M, Steinmüller S, Rodriguez E, Cruz ML, Luck J,
632 Fraser G, Brunel S, Montuori M, Fedchock C, Steel E, Pennington HG, Day R, Rossi JP, Xia J
633 (2020) Science diplomacy for plant health. *Nature Plants* 6: 902–905.
634 <https://doi.org/10.1038/s41477-020-0744-x>

635

636 Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G (2015) Ghost mtDNA haplotypes
637 generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and
638 phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research* 53:
639 109–115. <https://doi.org/10.1111/jzs.12095>

640

641 Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and
642 implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*
643 286: 20191409. <https://doi.org/10.1098/rspb.2019.1409>

644

645 Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase
646 subunit 1 divergences among closely related species. *Proceedings of the Royal Society B:
647 Biological Sciences* 270: S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>

648

649 Hebert PDN, deWaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, Mantle B, La Salle
650 J (2013) A DNA ‘Barcode Blitz’: Rapid Digitization and Sequencing of a Natural History
651 Collection. Kolokotronis S-O (Ed.). *PLoS ONE* 8: e68535.
652 <https://doi.org/10.1371/journal.pone.0068535>

653

654 Henrie JR, Thomson BM, Yungfleisch AA, Kerr M, Leavitt SD (2022) Characterizing Crustose
655 Lichen Communities—DNA Metabarcoding Reveals More than Meets the Eye. *Diversity* 14:
656 766. <https://doi.org/10.3390/d14090766>

657

658 Hodgetts J, Ostojá-Starzewski JC, Prior T, Lawson R, Hall J, Boonham N (2016) DNA barcoding
659 for biosecurity: case studies from the UK plant protection program. Wilson J-J (Ed.). *Genome*
660 59: 1033–1048. <https://doi.org/10.1139/gen-2016-0010>

661

662 Javal M, Terblanche J, Conlong D, Delahaye N, Grobbelaar E, Benoit L, Lopez-Vaamonde C,
663 Haran J (2021) DNA barcoding for bio-surveillance of emerging pests and species identification

664 in Afrotropical Prioninae (Coleoptera, Cerambycidae). Biodiversity Data Journal 9: e64499.
665 <https://doi.org/10.3897/BDJ.9.e64499>

666

667 Kelnarova I, Jendek E, Grebennikov VV, Bocak L (2019) First molecular phylogeny of *Agrilus*
668 (Coleoptera: Buprestidae), the largest genus on Earth, with DNA barcode database for forestry
669 pest diagnostics. Bulletin of Entomological Research 109: 200–211.
670 <https://doi.org/10.1017/S0007485318000330>

671

672 Kenis M, Auger-Rozenberg M-A, Roques A, Timms L, Péré C, Cock MJW, Settele J, Augustin S,
673 Lopez-Vaamonde C (2009) Ecological effects of invasive alien insects. Biological Invasions 11:
674 21–45. <https://doi.org/10.1007/s10530-008-9318-y>

675

676 Kitson JJN, Hahn C, Sands RJ, Straw NA, Evans DM, Lunt DH (2019) Detecting host–parasitoid
677 interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. Molecular
678 Ecology 28: 471–483. <https://doi.org/10.1111/mec.14518>

679

680 Koutroumpa FA, Lieutier F, Roux-Morabito G (2009) Incorporation of mitochondrial fragments
681 in the nuclear genome (*Numts*) of the longhorned beetle *Monochamus galloprovincialis*
682 (Coleoptera, Cerambycidae). Journal of Zoological Systematics and Evolutionary Research 47:
683 141–148. <https://doi.org/10.1111/j.1439-0469.2008.00492.x>

684

685 Krehenwinkel H, Pomerantz A, Prost S (2019) Genetic Biomonitoring and Biodiversity
686 Assessment Using Portable Sequencing Technologies: Current Uses and Future Directions.
687 Genes 10: 858. <https://doi.org/10.3390/genes10110858>

688

689 Liu M, Clarke LJ, Baker SC, Jordan GJ, Burrridge CP (2020) A practical guide to DNA
690 metabarcoding for entomological ecologists. Ecological Entomology 45: 373–385.
691 <https://doi.org/10.1111/een.12831>

692

- 693 Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014) Swarm: robust and fast clustering
694 method for amplicon-based studies. *PeerJ* 2: e593. <https://doi.org/10.7717/peerj.593>
695
- 696 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing
697 reads. *EMBnet.journal* 17: 10. <https://doi.org/10.14806/ej.17.1.200>
698
- 699 Mas H, Santoiemma G, Lencina JL, Gallego D, Pérez-Laorga E, Ruzzier E, Rassati D (2023)
700 Investigating beetle communities in and around entry points can improve surveillance at
701 national and international scale. *NeoBiota* 85: 145–165.
702 <https://doi.org/10.3897/neobiota.85.103904>
703
- 704 Mazza G, Tricarico E, Genovesi P, Gherardi F (2014) Biological invaders are threats to human
705 health: an overview. *Ethology Ecology & Evolution* 26: 112–129.
706 <https://doi.org/10.1080/03949370.2013.863225>
707
- 708 Nakahama N, Uchida K, Ushimaru A, Isagi Y (2018) Historical changes in grassland area
709 determined the demography of semi-natural grassland butterflies in Japan. *Heredity* 121: 155–
710 168. <https://doi.org/10.1038/s41437-018-0057-2>
711
- 712 Petec G, Vîlcu R, Uluitu M, Petec M (1975) Relationship between bovine serum albumin
713 structure and its chemical equilibria with hydrogen and 5-hydroxytryptamine ions. *Physiologie*
714 (Bucarest) 12: 77–80.
715
- 716 Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019)
717 Prospects and challenges of implementing DNA metabarcoding for high-throughput insect
718 surveillance. *GigaScience* 8: giz092. <https://doi.org/10.1093/gigascience/giz092>
719
- 720 Ratnasingham S, Hebert PDN (2007) BARCODING: bold: The Barcode of Life Data System
721 (<http://www.barcodinglife.org>): BARCODING. *Molecular Ecology Notes* 7: 355–364.
722 <https://doi.org/10.1111/j.1471-8286.2007.01678.x>

723

724 Preston, M., Fritzsche, M. & Woodcock, P. 2022. Understanding and mitigating errors and
725 biases in metabarcoding: an introduction for non-specialists. JNCC Report No. 699, JNCC,
726 Peterborough, ISSN 0963-8091.

727

728 Richardson DM, Pysek P, Rejmanek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization
729 and invasion of alien plants: concepts and definitions. *Diversity* <html_ent glyph="@amp;"
730 *asci="&";"/> Distributions 6: 93–107. <https://doi.org/10.1046/j.1472-4642.2000.00083.x>*

731

732 Roques A, Auger-Rozenberg M-A, Blackburn TM, Garnas J, Pyšek P, Rabitsch W, Richardson DM,
733 Wingfield MJ, Liebhold AM, Duncan RP (2016) Temporal and interspecific variation in rates of
734 spread for insect species invading Europe during the last 200 years. *Biological Invasions* 18:
735 907–920. <https://doi.org/10.1007/s10530-016-1080-y>

736

737 Roques A, Ren L, Rassati D, Shi J, Akulov E, Audsley N, Auger-Rozenberg M-A, Avtzis D, Battisti
738 A, Bellanger R, Bernard A, Bernadinelli I, Branco M, Cavaletto G, Cocquempot C, Contarini M,
739 Courtial B, Courtin C, Denux O, Dvořák M, Fan J, Feddern N, Francese J, Franzen EKL, Garcia A,
740 Georgiev G, Georgieva M, Giarruzzo F, Gossner M, Gross L, Guarneri D, Hoch G, Hölling D,
741 Jonsell M, Kirichenko N, Loomans A, Luo Y, McCullough D, Maddox C, Magnoux E, Marchioro
742 M, Martinek P, Mas H, Mériguet B, Pan Y, Phélut R, Pineau P, Ray AM, Roques O, Ruiz M-C,
743 Monteys VS i, Speranza S, Sun J, Sweeney JD, Touroult J, Valladares L, Veillat L, Yuan Y, Zalucki
744 MP, Zou Y, Žunič-Kosi A, Hanks LM, Millar JG (2023) Worldwide tests of generic attractants, a
745 promising tool for early detection of non-native cerambycid species. *NeoBiota* 84: 169–209.
746 <https://doi.org/10.3897/neobiota.84.91096>

747

748 Rossa R, Goczał J (2021) Global diversity and distribution of longhorn beetles (Coleoptera:
749 Cerambycidae). *The European Zoological Journal* 88: 289–302.
750 <https://doi.org/10.1080/24750263.2021.1883129>

751

- 752 Sahlin K, Medvedev P (2020) De Novo Clustering of Long-Read Transcriptome Data Using a
753 Greedy, Quality Value-Based Algorithm. *Journal of Computational Biology: A Journal of*
754 *Computational Molecular Cell Biology* 27: 472–484. <https://doi.org/10.1089/cmb.2019.0299>
755
- 756 Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE,
757 Kleunen M, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J, Pyšek P, Winter M,
758 Essl F (2021) Projecting the continental accumulation of alien species through to 2050. *Global*
759 *Change Biology* 27: 970–982. <https://doi.org/10.1111/gcb.15333>
760
- 761 Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, van
762 Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu
763 G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N,
764 Guénard B, Jäger H, Kartesz J, Kenis M, Kühn I, Lenzner B, Liebhold AM, Mosen A, Moser D,
765 Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S,
766 Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, Walker K, Ward DF,
767 Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased
768 accessibility of new source pools. *Proceedings of the National Academy of Sciences* 115.
769 <https://doi.org/10.1073/pnas.1719429115>
770
- 771 Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R (2021) ONTbarcoder
772 and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone.
773 *BMC Biology* 19: 217. <https://doi.org/10.1186/s12915-021-01141-x>
774
- 775 Strickler KM, Fremier AK, Goldberg CS (2015) Quantifying effects of UV-B, temperature, and
776 pH on eDNA degradation in aquatic microcosms. *Biological Conservation* 183: 85–92.
777 <https://doi.org/10.1016/j.biocon.2014.11.038>
778
- 779 Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate
780 metabarcoding. *Metabarcoding and Metagenomics* 1: e14625.
781 <https://doi.org/10.3897/mbmg.1.14625>

782

783 Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from
784 long uncorrected reads. *Genome Research* 27: 737–746.
785 <https://doi.org/10.1101/gr.214270.116>

786

787 Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population
788 genetics. *Trends in Ecology & Evolution* 22: 634–642.
789 <https://doi.org/10.1016/j.tree.2007.08.017>

790

791 Wang Q (2017) *Cerambycidae of the World: Biology and Pest Management*. CRC Press, 643 pp.
792 Wu Y, Trepanowski NF, Molongoski JJ, Reagel PF, Lingafelter SW, Nadel H, Myers SW, Ray AM
793 (2017) Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in
794 trade-associated solid wood packaging material using DNA barcoding and morphology.
795 *Scientific Reports* 7: 40316. <https://doi.org/10.1038/srep40316>

796

797 **Acknowledgements**

798 We would like to thank all colleagues who participated in the taxa sampling (see Roques et al.
799 2023). We are thankful to Lucas Sire for insightful discussion on primer choice. The authors
800 thank the GeT-Biopuces platform of INSA Toulouse for the IonTorrent sequencing study. This
801 work was supported by the PORTRAP project “Test de l’efficacité de pièges génériques
802 multicomposés pour la détection précoce d’insectes exotiques xylophages dans les sites
803 potentiels d’entrée sur le territoire national” and HOMED project (HOListic Management of
804 Emerging Forest Pests and Diseases) which received funding from the European Union’s
805 Horizon 2020 research and innovation program under grant agreement No. 771271
806 (<https://homed-project.eu/>). We are grateful to the genotoul bioinformatics platform Toulouse
807 Midi-Pyrenees for providing help and computing storage resources. Loïs Veillat was supported
808 by a PhD studentship from HOMED project and doctoral school SSBCV at the university of
809 Orléans.

810

811

812 **Data accessibility**

813 Barcode data for the 32 species used in the mock community experiment are available from
814 BOLD in the dataset DS-MINION ([dx.doi.org/XX/DS-MINION](https://dx.doi.org/10.3897/XX/DS-MINION)). Raw sequence data for this
815 project and analytical script and files are available on figshare
816 ([https://figshare.com/projects/Molecular_biosurveillance_of_wood-
817 boring_cerambycid_beetles_using_DNA_metabarcoding/171432](https://figshare.com/projects/Molecular_biosurveillance_of_wood-boring_cerambycid_beetles_using_DNA_metabarcoding/171432)).

818

819 **Author contributions**

820 Loïs Veillat, Géraldine Roux, Carlos Lopez-Vaamonde and Stéphane Boyer conceived the study.
821 Alain Roques collected field samples. Stéphane Boyer, Marina Querejeta, Emmanuelle
822 Magnoux and Loïs Veillat conducted the laboratory sample processing. Loïs Veillat analysed
823 the data and wrote the first draft. All authors contributed to the preparation of the manuscript.
824 Both senior authors, Géraldine Roux and Carlos Lopez-Vaamonde, contributed equally to this
825 study.

826

827

828

829

830

831

832

833

834

835

836

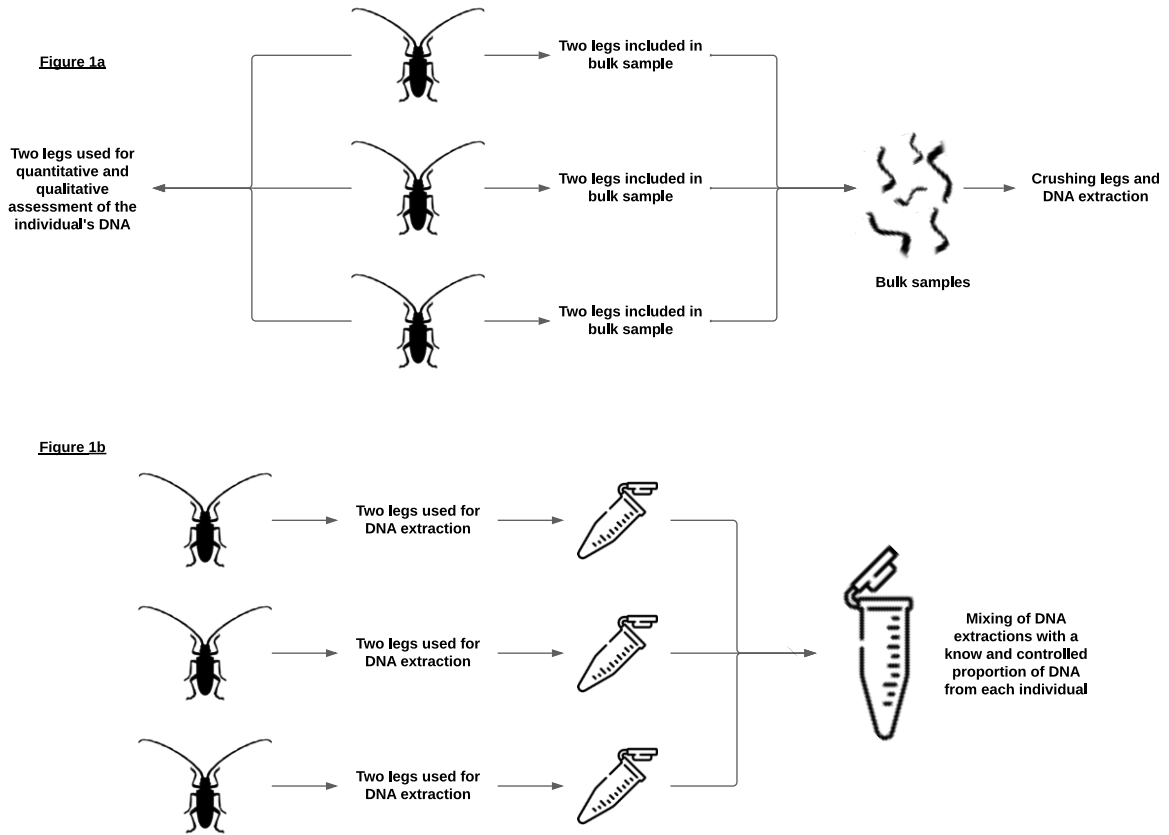
837

838

839

840

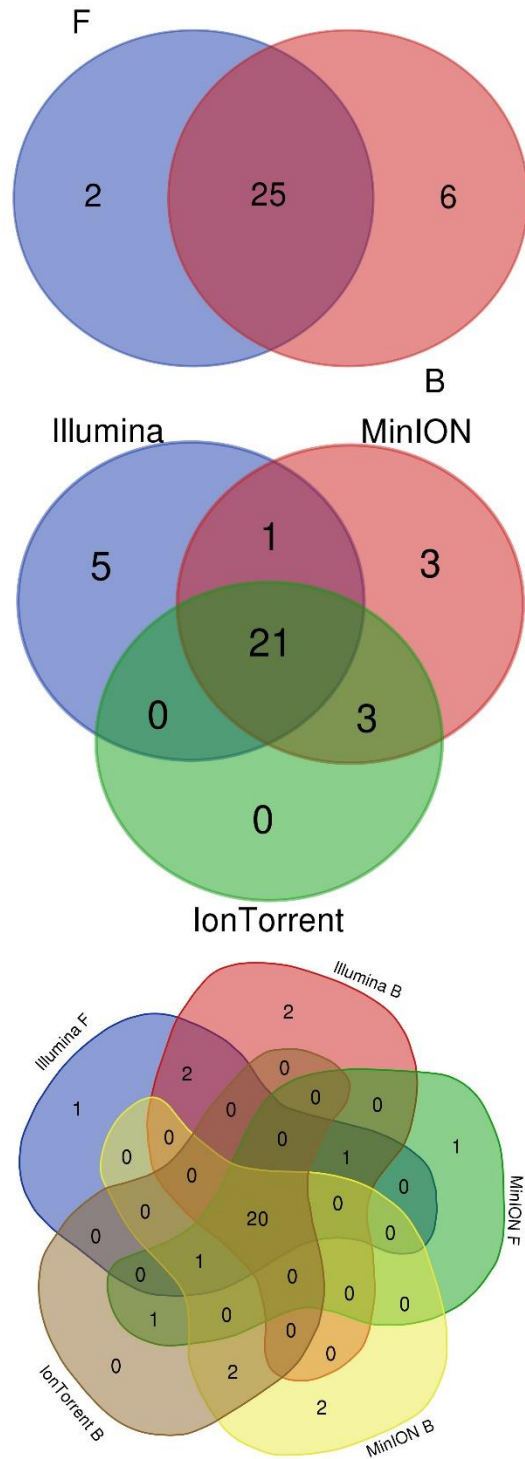
841 Figure 1: Overview of the DNA extraction protocol for tests 1 and 3 (Fig. 1a) and for test 2 (Fig.
842 1b).



843
844
845
846
847
848
849
850
851
852
853
854

855 Figure 2: Venn diagrams showing the number of specific and shared species-level detections
 856 among primers used (upper, F=fwhF2/fwhR2n [254bp] and B=BF3/BR2 [458bp]), technology
 857 used (middle) and for primers and technologies used (lower).

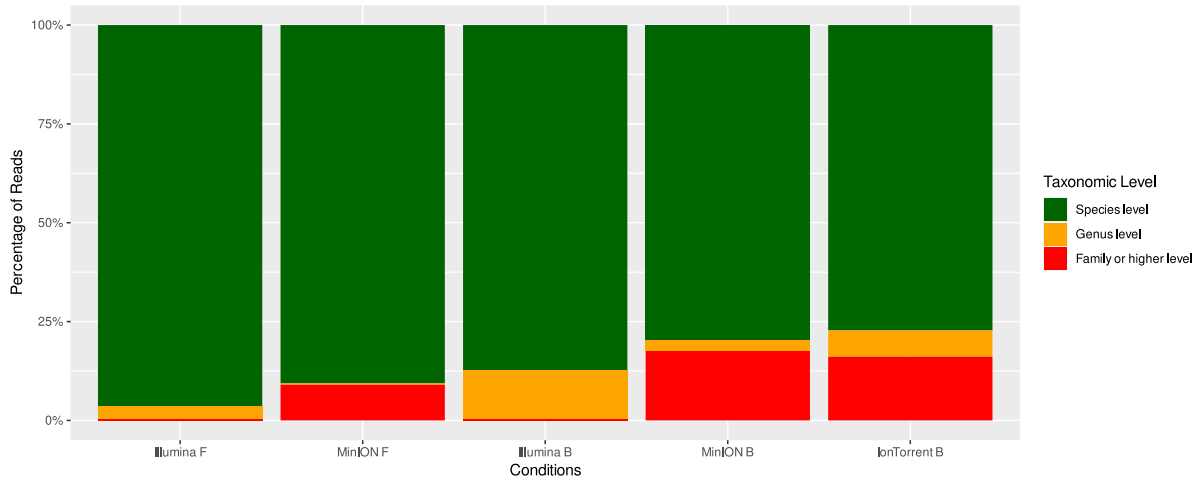
858



859

860

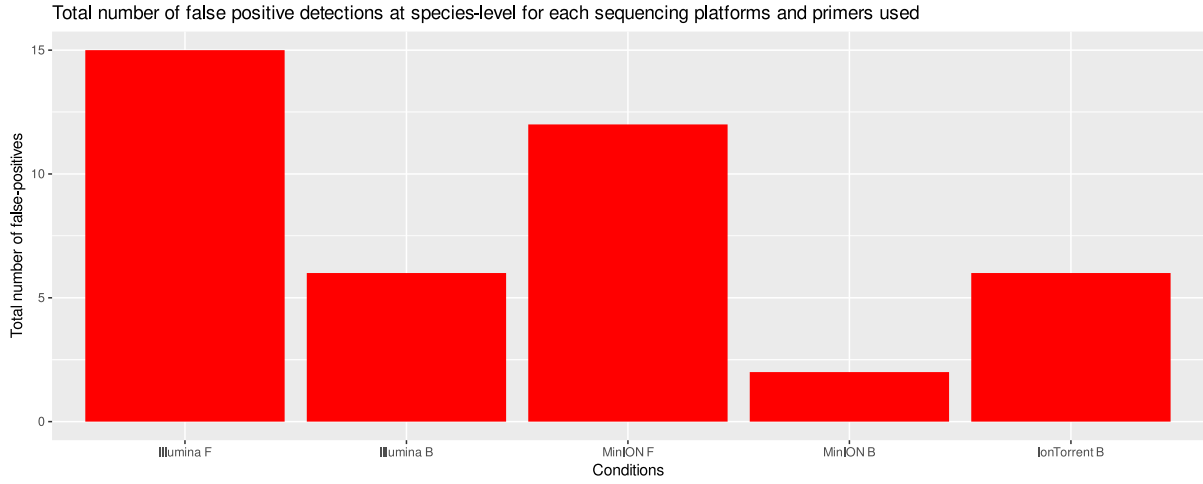
861 Figure 3: Proportion of reads assigned to each taxonomic level for each combination of
 862 sequencing technology and pair of primers (F: fwhF2/fwhR2n; B: BF3/BR2).
 863



864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883

884 Figure 4: Number of false positive detections at species-level for each sequencing platforms
885 and primers used (F=fwhF2/fwhR2n [254bp] and B=BF3/BR2 [458bp]).

886



887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

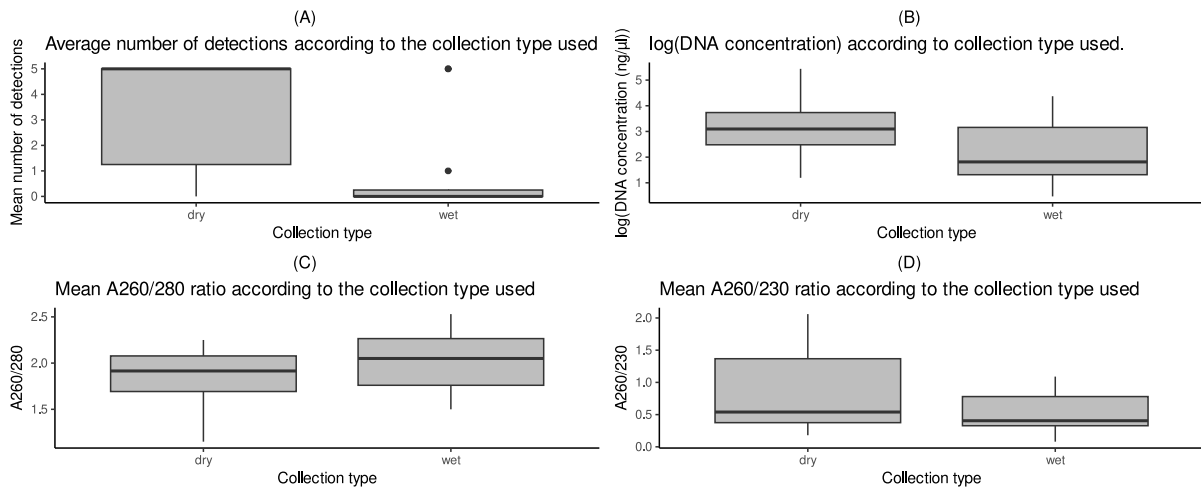
904

905

906

930 Figure 6: Boxplots representing (A) the average number of detections according to the type of
 931 preservation used, (B) the logarithm of the average DNA concentration according to the type
 932 of preservation used, (C) the A260/280 quality ratio according to the type of preservation used,
 933 and (D) the A260/230 quality ratio according to the type of preservation used. The black dots
 934 represent the extreme values.

935



936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952