ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 18/12/2023

DOI: https://doi.org/10.3897/arphapreprints.e117433

DNA metabarcoding, an efficient way to detect non-native cerambycid beetles in trapping collections?

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

To be submitted to Neobiota

DNA metabarcoding, an efficient way to detect non-native cerambycid beetles in trapping collections ?

Loïs Veillat[1], Stéphane Boyer[2], Marina Querejeta[2,4], Emmanuelle Magnoux[1], Alain Roques[1], Carlos Lopez-Vaamonde[1,2*], Geraldine Roux[1,3*]

[1]INRAE, UR633, Zoologie Forestière, F-45075 Orléans, France
[2]Institut de Recherche sur la Biologie de l'Insecte, UMR7261 CNRS - Université de Tours, France
[3]Laboratoire Physiologie, Ecologie et Environnement P2E, Université d'Orléans, 1 rue de Chartres, B.P. 6759, 45067 Orléans cedex 2, France
[4]Department of Functional Biology, University of Oviedo, Oviedo, Asturias, Spain

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

*These authors contributed equally to this work

Abstract

Individual sorting and identification of thousands of insects collected in mass trapping biosurveillance programs is a labor intensive and time-consuming process. Metabarcoding, which allows for the simultaneous identification of multiple individuals in a single mixed sample, has the potential to expedite this process. However, detecting all the species present in a bulk sample can be challenging, especially when under-represented non-native specimens were intercepted. In this study, we quantified the effectiveness of metabarcoding at detecting exotic species within six different mock communities including or not native and non-native species of European xylophagous cerambycid beetles. Although we did not observe significant differences in the total number of species detected between MinION, Illumina, and IonTorrent sequencing technologies, a greater number of individuals was detected and identified to species

using MinION, including the detection of three non-native cerambycids. The three sequencing technologies also showed similar results in detecting and identifying closely related species and species at low abundance. The capture method appears to greatly influence sample preservation and detection. Indeed, individuals captured in traps containing monopropylene and water had both lower DNA concentration leading to lower species detection rates compared to individuals killed using just an insecticide without any collection medium.

Keywords: Biological invasions, Biosecurity, *Cerambycidae*, Illumina®, IonTorrent®, Oxford Nanopore®, xylophagous, exotic, alien.

Introduction

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

Among these non-native insects, species associated with woody plants are increasingly dominating, accounting for 76.5% of all herbivore species newly recorded in Europe from 2000 to 2014 probably because of the growing trade of ornamental plants and wooden packaging material transported in international cargo shipments (Aukema et al. 2010; Roques et al. 2016). One of these important families of xylophagous beetles is the long-horned Cerambycidae, with more than 200 species affecting forestry, horticulture, and agriculture (Rossa and Goczał 2021) resulting in multimillion-dollar losses every year (Wang 2017). To detect potential new invasions of Cerambycids, biomonitoring programs have been set up over large geographical areas with intensive trapping campaigns extending over several years (Roques et al. 2023; Mas et al. 2023). However, rapidly evolving trades lead to changes in trade routes and imported goods which results in an increasing arrival of new non-native species. Many of these species have not been previously reported as invaders, some are not considered to be pests in their native ranges, and some could even be unknown to science (Seebens et al. 2018). As part of the European project HOMED (https://homed-project.eu/) 244 Cerambycid traps were set up across Europe (France, Italy, Spain, Switzerland, Portugal, Austria, England, Greece, Slovenia, Netherlands, Bulgaria, Czech Republic, and in Sweden), 38 in Asia (China, Siberia, Russia), 11 in North America (USA, Canada), five in the Caribbean (Martinique) and four in Australia, all baited with generic lures, for simultaneous detection of multiple species (Roques et al. 2023). In such large-scale trapping campaign, thousands of captured insects must to be sorted out and identified by expert taxonomists. This identification step is time-consuming and labor-intensive, thus limiting the rapid detection of non-native individuals among large numbers of native ones (Piper et al. 2019; Abeynayake et al. 2021; Chua et al. 2023). It is essential that those non-native

species are detected as quickly as possible to allow their eradication before establishment and dispersal (Richardson et al. 2000; Blackburn et al. 2011; Giovani et al. 2020).

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

Metabarcoding is also based on high-throughput sequencing (HTS) technologies which generates a large number of short DNA sequences (reads), allowing the accurate identification of multiple individuals simultaneously from a single mixed sample (hereafter called “bulk”) (Liu et al. 2020), such as all the individual insects captured in a single biomonitoring trap. Moreover, compared to traditional morphological identification, metabarcoding offers a significant reduction in costs (Batovska et al. 2021), providing equivalent or better detection and identifying a much wider spectrum of taxa (Elbrecht et al. 2017; Andújar et al. 2018). Using DNA as a proxy for species

detection and considering sequence variation within and among taxa, the metabarcoding approach is constrained by the completeness of the reference databases to accurately assign sequences to correctly identified taxa (Liu et al. 2020).

Although metabarcoding has several advantages, ensuring the accuracy of detections is paramount as erroneous detections of pest species can lead to severe environmental and economic consequences (Batovska et al. 2021). Metabarcoding approaches still suffers from a number of methodological limitations that can make it unfitting for rapid biosecurity detection. In particular, the time required to process samples can be an issue (long delays between capturing individuals and obtaining sequencing results), particularly when the sampling sites are located far away from laboratories and transporting samples may require specific permits for certain species or when external providers are slow to sequence samples. These limitations, may hinder biomonitoring projects and thus slow down the detection of potential invasive species, preventing the implementation of measures to mitigate their impacts. (Krehenwinkel et al. 2019; Egeter et al. 2022). Although MiSeq is generally recommended due to its lower error rate and well-established bioinformatic procedures, but Braukmann et al. (2019) demonstrated similar performance in sequence quality and insect species recovery compared to IonTorrent platforms (Ion Torrent PGM, and Ion Torrent S5).

In recent years, Oxford Nanopore Technologies® have released a portable sequencing platform, the MinION. This small sequencer can be connected via USB to a laptop to perform sequencing (Krehenwinkel et al. 2019) in the field and obtain sequencing data in real time conditions. Indeed, the MinION for a metabarcoding application offers the

possibility of performing DNA sequencing of bulk samples directly on site without the need for transport or relying on external sequencing providers. At present, although the MinION does not seem suitable for the characterization of complex communities, it is already suitable for the analysis of metabarcoding data when the species diversity per sample is low and the target species are well represented in public databases (Ho et al. 2020).

The main aim of our study was to popularize a metabarcoding approach for the biosurveillance of Cerambycid wood-boring beetles. We evaluated the effectiveness of three Next Generation Sequencing technologies, i.e. the portable Nanopore sequencer MinION, the Illumina MiSeq and the Ion GeneStudio S5 (IonTorrent®) to detect invasive specimen from different mock communities. We specifically examined the accuracy to differentiate between closely related cerambycid species and to detect low-abundance species in mixed trap samples. We furthermore analysed several metabarcoding primer pairs to evaluate the accuracy for species identification. We finally highlighted the importance of field sampling protocol, especially trapping methods (dry versus monopropylene glycol) on species detection.

Materials and Methods

Taxa sampling

Mock communities were constructed using field-trapped specimens from different countries in Europe (France, Greece, Portugal, Spain), China (Beijing and Zhejiang Province) and USA (Michigan) (**Tab. 1**), as part of a worldwide trapping experiment using multi-funnel traps baited with a generic attractant blend including eight

Cerambycid pheromones (see details of the blend composition and trapping methods in Roques et al. 2023). Most of the specimens were caught using α-cypermethrin insecticide (Storanet®, BASF Pflanzenschutz Deutschland, Germany) into the trap basins, of which the bottom had been replaced with a wire mesh to allow drainage and keep specimens dry (hereafter call "dry" method). Other specimens were captured using a 50:50 ratio of monopropylene glycol (MPG) and water (hereafter call "wet" method). Cerambycides collected from field-traps were stored in ethanol 95° and kept at -20°C until experiment. Date, country of collection, type of trap and specimens used in mock communities are detailed in Table 1.

Mock community construction and DNA extraction

Six mock communities with varying species composition were assembled as follows:

Test 1: Identifying closely-related species.

To assess the efficiency of the different sequencing technologies and primers to differentiate between sister species, bulks 1 and 2 were composed of congeneric species (**Tab. 1**). Two legs from each individual were collected and pooled to constitute the bulks. The whole set of legs was then ground using flame-sterilized metal pestles to limit the risk of contamination. DNA from the ground material was extracted using the Qiagen DNeasy Blood and Tissue Kit following the manufacturer's instructions. Two other legs were taken from the specimens, from which we assessed the quantity and quality of DNA for each specimen individually (**Fig. 1a**).

Test 2: Detecting low abundance species.

Bulks 3 and 4 were composed of six species represented by heterogeneous DNA concentrations (**Tab. S1**) to assess the ability of the sequencing technologies and primers to detect species present in a very low abundance. DNA of each individual was previously extracted using two legs that were ground as above and processed using the Qiagen DNeasy Blood and Tissue Kit. To construct bulks 3 and 4, individual DNA extracts were quantified using a fluorometer (Nanodrop™, Thermo Fisher Scientific) and mixed together according to their concentration to achieve the needed proportions of DNA for each individual (six individuals of different species ranging from 41% to 3% for Bulk 3 and six individuals of different species ranging from 50% to 0.5% for bulk 4) (**Tab. 1, Fig. 1b**).

Test 3: Mimicking field trap content on species composition.

Bulks 5 and 6 comprised individuals from a number of species native to Europe usually found in the traps deployed there, with the addition of non-native species which have either already been introduced or at risk of being introduced in Europe (Bulk 5: 22 individuals of eight species, including one non-native (*Cordylomera spinicornis*); Bulk 6: 41 individuals of 12 species including two non-native ones (*Xylotrechus stebbingi* and *Xylotrechus chinensis*) (**Tab. 1**). The DNA was extracted following the same protocol as for bulks 1 and 2 (**Fig. 1a**).

Table 1 : Species, origin, date and condition of capture of the specimens used in the 6 bulks used. Species names in bold correspond to exotic species. We consider specimens that have been captured on a different continent than their place of origin as exotic.

Bulk	*Species*	Country of collection	Collection Year	Collection type
1	*Arhopalus ferus*	Portugal	2020	Cypermethrin insecticide

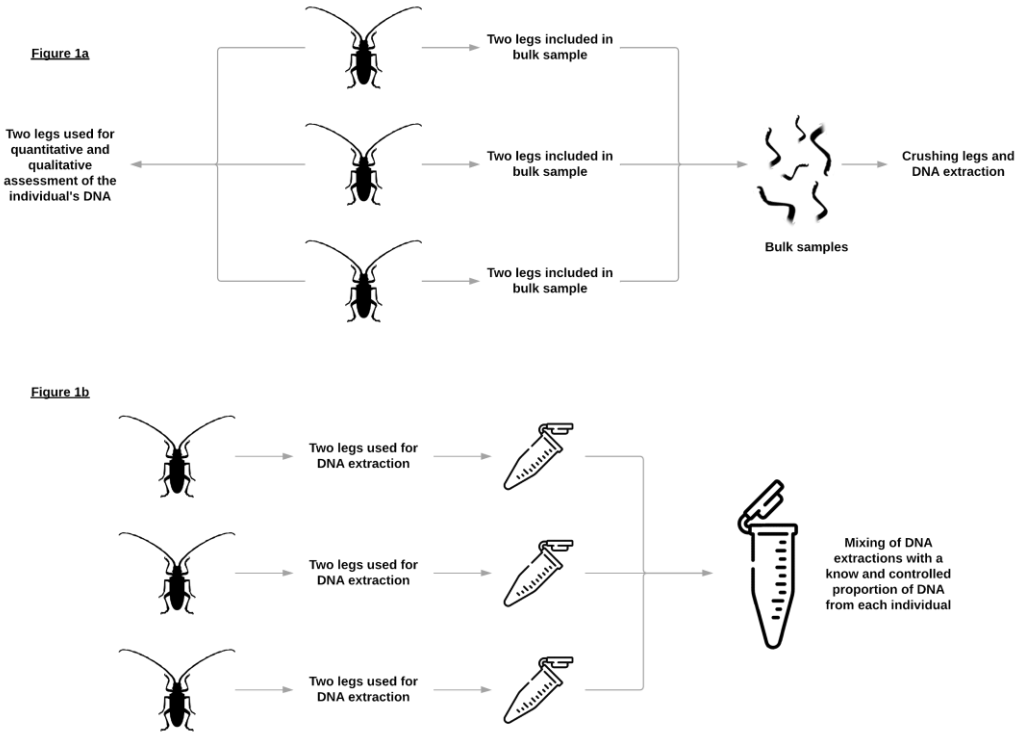
1	*Arhopalus rusticus*	France	2021	Cypermethrin insecticide
1	*Arhopalus syriacus*	Portugal	2019	Monopropylene glycol
1	*Xylotrechus arvicola*	Portugal	2021	Cypermethrin insecticide
1	***Xylotrechus chinensis***	Greece	2019	Cypermethrin insecticide
1	***Xylotrechus stebbingi***	Greece	2019	Cypermethrin insecticide
1	*Xylotrechus undulatus*	USA	2019	Monopropylene glycol
2	*Monochamus galloprovincialis*	Portugal	2019	Monopropylene glycol
2	*Monochamus sutor*	France	2019	Cypermethrin insecticide
2	*Monochamus carolinensis*	USA	2019	Monopropylene glycol
2	*Monochamus scutellatus*	USA	2019	Monopropylene glycol
2	*Phymatodes amoenus*	USA	2019	Monopropylene glycol
2	***Phymatodes testaceus***	USA	2019	Monopropylene glycol
2	*Phymatodes varius*	USA	2019	Monopropylene glycol
2	*Phymatodes aereus*	USA	2019	Monopropylene glycol
2	***Phymatodes dimidiatus***	USA	2019	Monopropylene glycol
3	*Pyrrhidium sanguineum*	France	2020	Cypermethrin insecticide
3	***Xylotrechus stebbingi***	Spain	2021	Cypermethrin insecticide
3	*Monochamus galloprovincialis*	Spain	2021	Cypermethrin insecticide
3	***Xylotrechus chinensis***	Greece	2019	Cypermethrin insecticide
3	*Chlorophorus glabromaculatus*	France	2020	Cypermethrin insecticide
3	*Phymatodes testaceus*	France	2020	Cypermethrin insecticide
4	*Arhopalus ferus*	France	2020	Cypermethrin insecticide
4	*Monochamus sutor*	France	2019	Cypermethrin insecticide
4	*Aegomorphus francottei*	France	2020	Cypermethrin insecticide

4	*Monochamus galloprovincialis*	France	2018	Cypermethrin insecticide
4	***Xylotrechus stebbingi***	Spain	2021	Cypermethrin insecticide
4	***Xylotrechus chinensis***	Greece	2019	Cypermethrin insecticide
5	*Pyrrhidium sanguineum*	France	2021	Cypermethrin insecticide
5	*Batocera rubus*	China	2012	Hand collected
5	*Cerambyx scopolii*	France	2020	Cypermethrin insecticide
5	***Cordylomera spinicornis***	France	2020	Cypermethrin insecticide
5	*Leiopus femoratus*	France	2021	Cypermethrin insecticide
5	*Leiopus nebulosus*	France	2020	Cypermethrin insecticide
5	*Pachyta bicuneata*	China	1987	Hand collected
5	*Stictoleptura cordigera*	France	2021	Cypermethrin insecticide
6	*Arhopalus rusticus*	France	2020	Cypermethrin insecticide
6	***Xylotrechus chinensis***	Greece	2019	Cypermethrin insecticide
6	*Plagionotus detritus*	France	2020	Cypermethrin insecticide
6	*Plagionotus arcuatus*	France	2020	Cypermethrin insecticide
6	***Xylotrechus stebbingi***	France	2020	Cypermethrin insecticide
6	*Arhopalus syriacus*	France	2020	Cypermethrin insecticide
6	*Arhopalus ferus*	France	2020	Cypermethrin insecticide
6	*Xylotrechus colonus*	USA	2019	Monopropylene glycol
6	*Chlorophorus ruficornis*	France	2021	Cypermethrin insecticide
6	*Phymatodes testaceus*	France	2021	Cypermethrin insecticide
6	*Prionus coriarius*	France	2010	Cypermethrin insecticide
6	*Phymatodes amoenus*	USA	2019	Monopropylene glycol

202

203

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



PCR amplification

All bulk samples were amplified with the following two primer pairs: BF3/BR2 (call hereafter "B") (CCHGAYATRGCHTTYCCHCG / TCDGGRTGNCCRAARAAYCA (Elbrecht and Leese 2017; Elbrecht et al. 2019), which generates a 458 bp amplicon; and fwhF2/fwhR2n (call hereafter "F") (GGDACWGGWTGAACWGTWTAYCCHCC / GTRATWGCHCCDGCTARWACWGG), which generates a shorter 254 bp amplicon (Vamos et al. 2017). Each PCR comprised 15.3 µl H2O, 2.5 µl 10X PCR buffer, 2.5 µl dNTPs [1mM], 1 µl of each primer [0,4mM], 0.2 µl Dream Taq (Thermo Fisher Scientific), 0.5 µl Betaine [100mM], 2 µl DNA and H2O for a total of 25 µl per reaction. PCR was performed using the following program: 95°C for 5 min, 29 cycles of 95°C for 30 s, 48°C for 30 s, and 72°C for 50 s and 72°C for 5 min (Elbrecht et al. 2019). PCR products were then run on a 2%

agarose gel stained with ethidium bromide and visualised by UV transluminator. The PCR products were then purified with the NucleoFast 96 PCR plate purification kit (Macherey-Nagel).

Illumina® library preparation

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

MinION library preparation

Libraries were prepared according to the Oxford Nanopore Technologies ® protocol: "PCR barcoding (96) amplicons (SQK-LSK110) (version:

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

IonTorrent® library preparation

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

Reference Barcode Dataset

A dataset was built using all the public sequences of all Cerambycidae species available in BOLD systems (Ratnasingham and Hebert 2007). It was then verified whether all 33 species present in the bulk samples were represented by at least one sequence in the database. Three species not previously included in the database were barcoded through

Sanger sequencing on an ABI 3500 genetic analyzer (Applied Biosystems) using the Big-Dye Terminator V3.1 sequencing kit (Applied Biosystems) and BF3/BR2 primer pair. They were finally added to our local database to ensure that they were represented by at least one barcode sequence. The number of sequences in the database for each species is shown in **Tab. S1.**

The three newly generated barcodes along with one barcode per species represented in the mock communities are available from BOLD in the dataset DS-MINION (dx.doi.org/10.5883/DS-MINION).

Illumina® data processing

The raw data was analysed using the *FROGS* pipeline, a standardized pipeline containing a set of tools that are used to process amplicon reads that have been produced from Illumina® sequencing (Escudié et al. 2018; Henrie et al. 2022). First, amplicons with a size between 408 and 508 for the BF3/BR2 primer pair and 204 and 304 for the fwhF2/fwhR2n primer pair were retained. Sequence clustering was then performed using the SWARM algorithm (Mahé et al. 2014) with a maximum sequence difference set at d=1, as recommended by SWARM. Chimeric sequences were then removed. Sequences were aligned to the same database used for the MinION and IonTorrent® data analysis. In order to remove all spurious detections, OTU detections with less than 10 reads were removed. In barcoding and metabarcoding studies of insects, the sequence similarity level for OTU identification usually ranges from 95% to 99% (e.g., Gibson et al., 2014; Zenker et al., 2016). We calculated the best threshold value for our dataset by applying the function *localMinima* from R package *spider* (Brown et al., 2012). Based on this analysis, we used a threshold of 98% to assign OTUs to species level.

MinION and IonTorrent® data processing

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

Statistical analysis

A two-sample test of proportions was used to compare and assess the significance of the proportion of reads assigned to the species levels for MinION, Illumina, and IonTorrent technologies using the "Social Science Statistics" website (https://www.socscistatistics.com/tests/anova/default2.aspx). The proportion of reads assigned to different taxonomic levels was calculated by summing the total reads from different bulk samples for each condition. To determine if the number of false positives was significantly different among the three technologies and the two primer pairs, we calculated the detection mean for each bulk under different conditions. We then performed an ANOVA test followed by a Tukey HSD test using the "Social Science Statistics" website. The Wilcoxon test was used in R to assess the significance of the impact of the conservation method (dry or wet) on sample DNA concentration, sample DNA quality, and associated detection counts.

Results

A total of 1,248,95 reads were sequenced for the MinION Nanopore® technology using the fwhF2/fwhR2n primer pair, with an average of 78,037 reads per sample. After quality filtering, removing reads of incorrect size or insufficient quality, 1,113,844 (89.2%) reads were retained, with an average of 69,615 reads per bulk. For the BF3/BR2 primer pair, a total of 1,132,604 reads were sequenced, with an average of 62,922 reads per sample. After quality filtering, a total of 948,832 (83.8%) reads were retained, with an average of 52,712 reads per bulk.

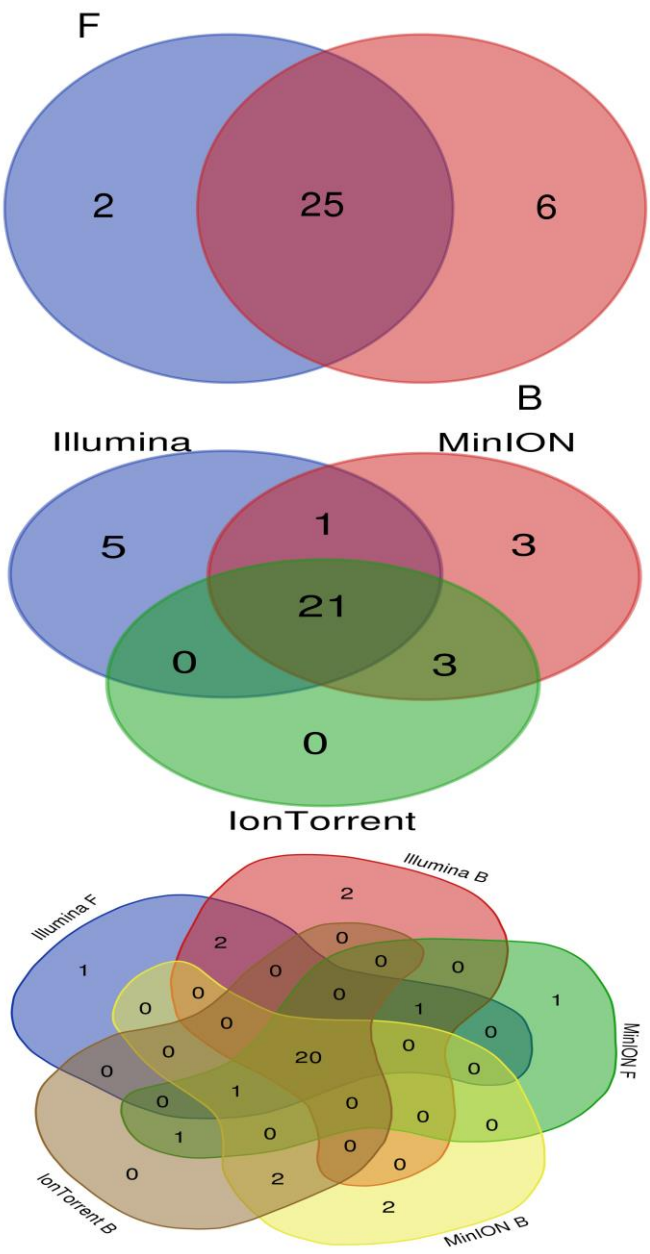
The Illumina® sequencing produced a total of 1,549,894 reads using the BF3/BR2 primer pair, with an average of 258,316 reads per bulk. After quality filtering, 1,025,637 (66.2%) reads were retained, with an average of 170,940 reads per bulk. For the fwhF2/fwhR2n primer pair, a total of 2,299,072 reads were sequenced, with an average of 383,179 reads per bulk. After quality filtering, 1,686,058 (73.3%) reads were retained, with an average of 281,010 reads per bulk.

For the IonTorrent® technology, 838,489 reads were sequenced, with an average of 139,748 reads per bulks with the BF3/BR2 primer pair. After the quality filtering, 280,695 (33.5%) reads remains with an average of 46,782 reads per bulks.

Benchmarking of sequencing technologies

The MinION technology accurately identified 28 out of 48 specimens at the species level, Illumina® technology allowed specific identification of 27 specimens and IonTorrent® identified 24 specimens. The primer pair fwhF2/fwhR2n allowed specific identification of 27 specimens with two OTU of its own while the primer pair BF3/BR2 allowed the identification of 31 specimens with six OTU of its own. Illumina® F, Illumina® B and MinION F allowed for 25 species-level identifications across all bulks, compared to 24 when considering the combination MinION F and IonTorrent® B. This difference was not significant. The number of identifications only obtained with one combination varied from zero (IonTorrent® B) to two (Minion B and Illumina® B) (**Fig. 2**).

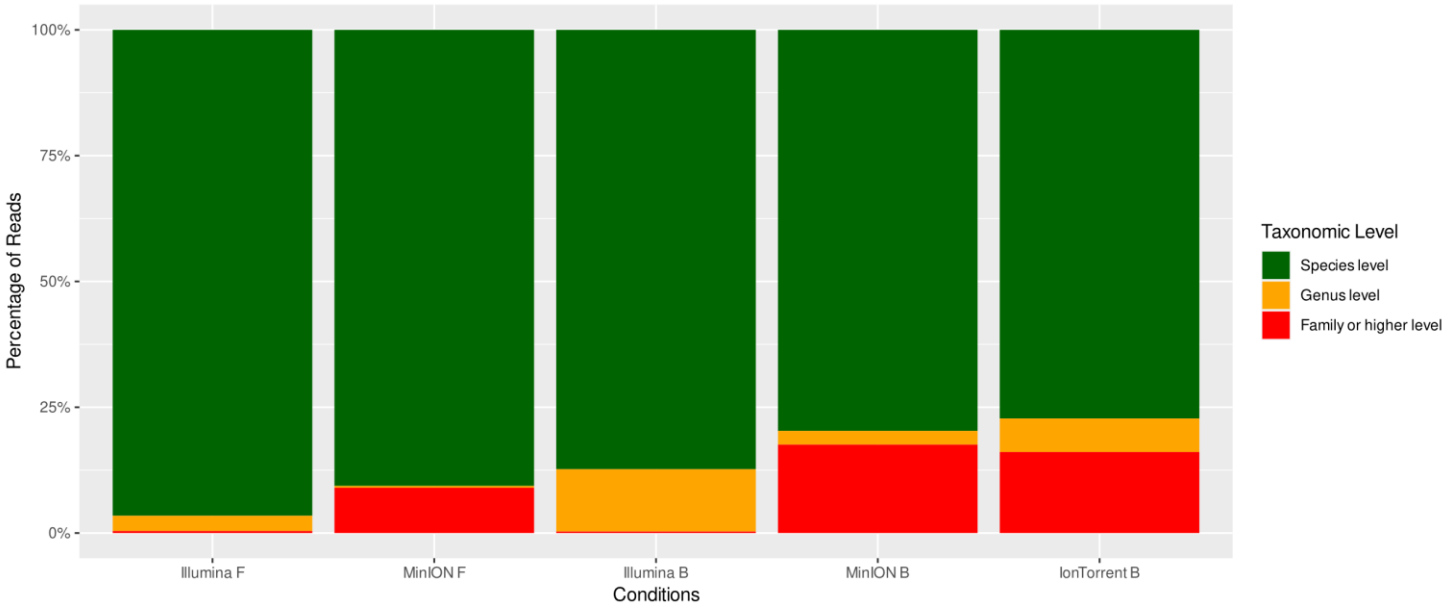
Figure 2: Venn diagrams showing the number of specific and shared species-level detections among primers used (upper, F=fwhF2/fwhR2n [254bp] and B=BF3/BR2 [458bp]), technology used (middle) and for primers and technologies used (lower).



The proportion of reads assigned at the species taxonomic level was significantly higher with Illumina® technology (p.value < 0.00001) comparing with the MinION, all primers included, particularly when considering primer pair F. Nearly 97% of reads were assigned at the species level for the Illumina® F combination compared to 90% for the MinION F combination (p.value<0.0001). As for primer pair BF3/BR2, over 87.3% of reads were assigned at the species level for Illumina®, followed by over 79.7% for

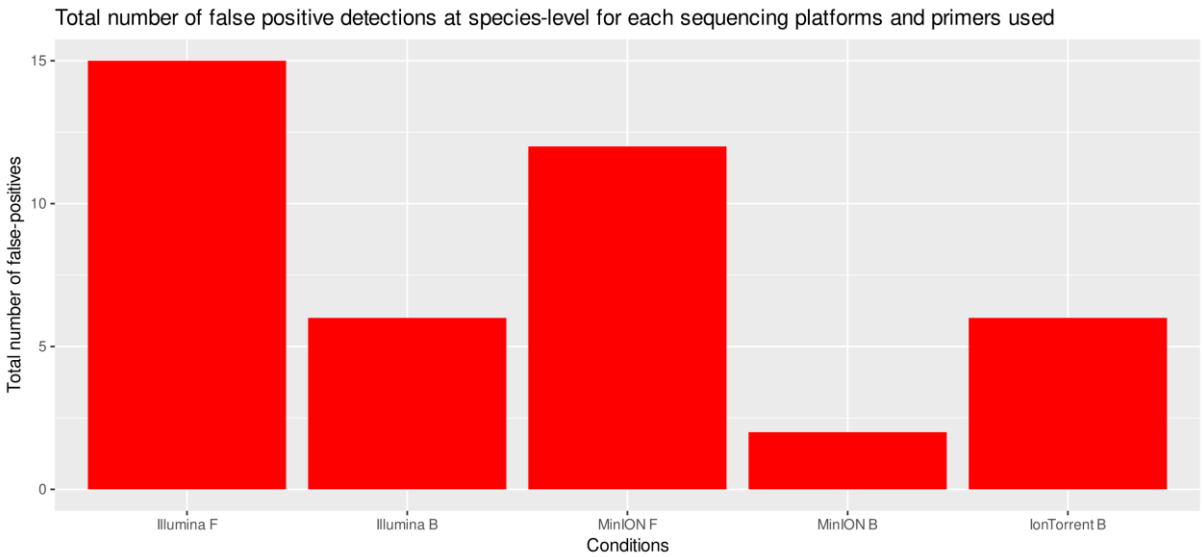
MinION technology and over 77.2% for IonTorrent® technology (**Fig. 3**). The primer pair fwhF2/fwhR2n resulted in a significantly higher percentage of reads assigned at the species level (93.6%) (considering both Illumina® and MinION technologies) compared to couple of primers B (81.4%) (considering all three technologies) (p.value<0.00001).

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



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

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



Mock community analysis

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

Bulks 1 and 2 were assembled to compare the detection rates of closely related species under different sequencing and primers conditions. No significant differences were observed among the different methods used. Illumina® detected seven species out of 16 (43.75%), MinION also detected seven out of 16 (43.75%), and IonTorrent® detected six species out of 16 (37.5%). However, significant differences were observed among the studied taxonomic groups: three out of four species from the genus *Arhopalus* were detected, as well as for *Xylotrechus*. Two out of four species were detected for the genus *Monochamus*, and one out of five species for the genus *Phymatodes*. We also observed

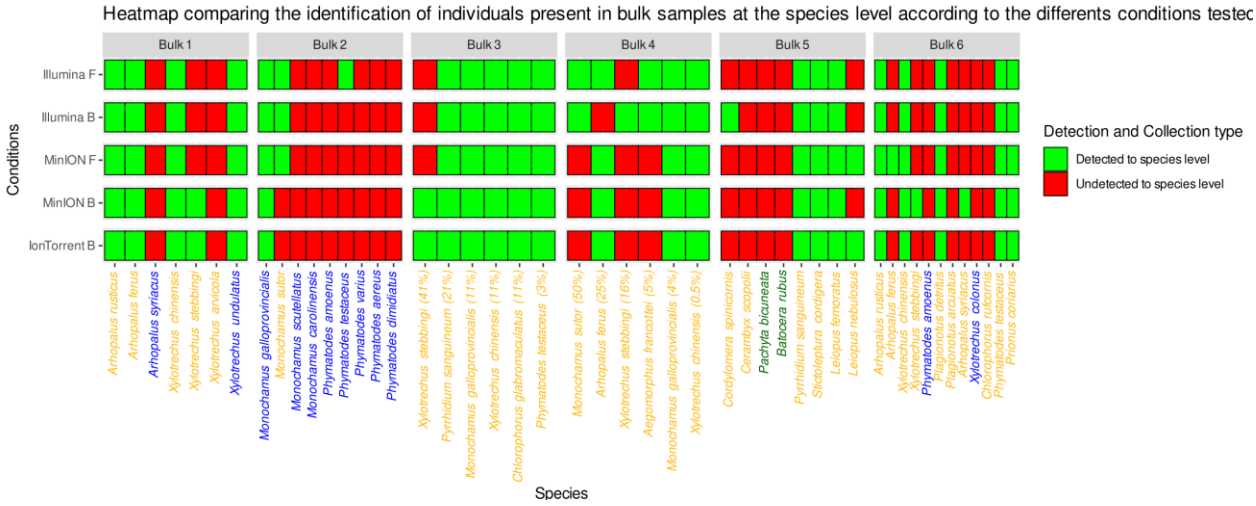
lower detection rates in Bulk 2, which was predominantly composed of specimens captured with 50% MPG in the trap.

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

Regarding the last two bulks mimicking the species composition in a field trap, MinION performed better to detect and identify specimens at species level (detecting eight out of 12 species (66.7%)) compared to Illumina® and IonTorrent® technologies (which detected five out of 12 species (41.7%)) for Bulk 6. As for Bulk 5, all technologies detected the same number of species (four out of six (66.7%)). In bulk 5, the non-native species, *Cordylomera spinicornis* was detected only by Illumina B. For bulk 6, the non-native species *Xylotrechus chinensis* was detected by all three technologies and *Xylotrechus stebbingi* by MinION B only.

Figure 5: Heatmap comparing the identification of individuals present in bulk samples at the species level (green square) or the absence of detection at the species level (red square) according to the sequencing technologies and primer pairs used. Species names

written in blue were collected using the wet method, those in yellow were collected using the dry method, and those in dark green were hand-captured.



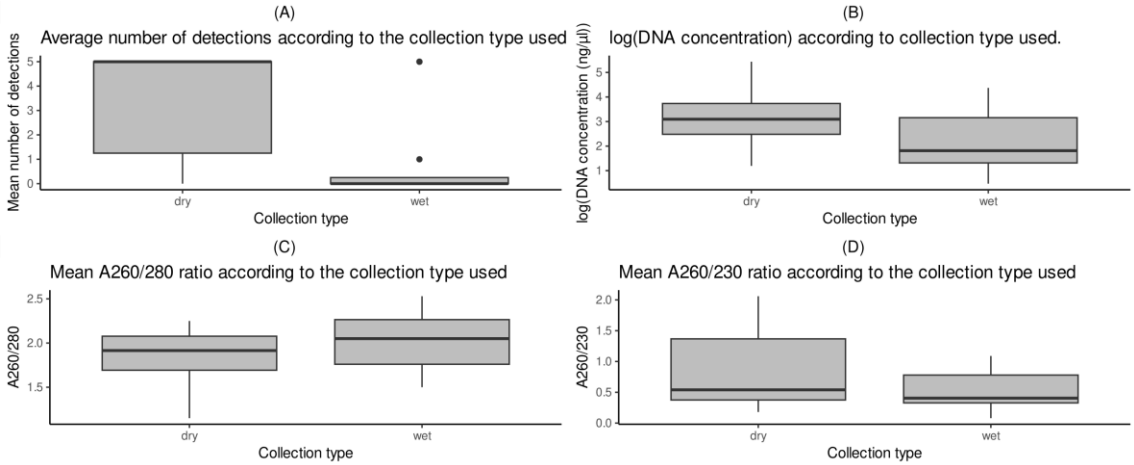
Impact of capture and storage conditions on individual detection

Our results demonstrate a significative difference in the mean number of detections between samples that were collected using the "dry" method (α-cypermethrin insecticide) and the "wet'" method (water-diluted propylene glycol) (Wilcoxon rank-sum test, W = 74.5, p.value = 0.0006342) (**Fig. 6**). Almost all of the specimens (9 out of 12 - 75%) collected using MPG trapping procedure were not detected by any of the sequencing technologies, including MinION. Conversely, those collected using the "dry" trapping procedure (using α-cypermethrin insecticide) had 30 out of 34 detected specimens (88.2%) across all technologies.

Regarding the quantity and quality of DNA extraction, Individuals captured using the "dry" method had higher DNA concentration (39 ng/µl on average) than MPG trapped specimens (18.6 ng/µl on average) (Wilcoxon rank-sum test, W = 123.5, p.value = 0.04533). The average A260/280 ratio was 1.9 for the "dry" method and 2 for MPG

method. However, the average A260/230 ratio of specimens trapped with the “dry” method (0.8) was higher than that of specimens captured with MPG (0.5) (Wilcoxon rank-sum test, W = 146, p.value = 0.1502) (**Fig. 6**).

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



Discussion

Benchmarking of sequencing technologies

A slightly higher number of individuals was detected and identified to species using MinION (28 specimens) compared to Illumina® (27 specimens) or IonTorrent® (24 specimens) even if the difference is not significant. It must be taken into account that we worked on a single pair of primers (BF3/BR2) with the IonTorrent® technology, which may have reduced the number of identifications, even if the detection rate was higher when using this primer pair generating a longer amplicon size (458bp). In fact, only BF3/BR2 allowed the species-level identification of the invasive species *Xylotrechus*

stebbingi. This difference may be due to the longer amplicon generated by this primer pair, which has more informative nucleotide sites to provide a reliable taxonomic assignment. It is also noted that the primer pair fwhF2/fwhR2n produced a higher percentage of reads allowing species-level identification than the primer pair BF3/BR2. However, as seen previously, fwhF2/fwhR2n generated a significant higher number of false positives than BF3/BR2 (**Fig. 4**). This may be due to the fact that fwhF2/fwhR2n is smaller in size than BF3/BR2 thus any loss of genetic information is more likely to lead to misidentification or false negatives.

The Illumina® technology has a higher percentage of reads allowing species-level identification compared to MinION or IonTorrent®. The detection of specimens at only higher taxonomic level (genus or family), can be explained by sequencing errors that produce reads with less than 98% identity to the reference database. Thus, our results suggest that Illumina® generates slightly less sequencing errors than MinION and IonTorrent®.

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

Impacts of capture and storage conditions on DNA conservation

Both the conditions of capture (wet versus dry methods) and storage (ie time lag between collection and lab processing) have an impact on both DNA concentration and quality and subsequently on the rate of species detection. Thus, the number of species detected is highly variable between bulks 1 and 2, which can be explained by the fact that both bulks are composed of species belonging to different genera (*Arhopalus* spp. and *Xylotrechus* spp. for bulk 1, *Monochamus* spp. and *Phymatodes* spp. for Bulk 2). A second, more likely, hypothesis would be that this difference in species detection is due to the capture methods used ('dry,' where individuals were captured without preservative fluid, and 'wet,' where individuals were preserved in 50% MPG until trap collection). For instance, the species *Phymatodes testaceus* for which we had both wet and dry specimens was always detected (10 detections out of 10 assays) when dry specimens were present, even in low concentration (3% in Bulk 3) (Bulk 3 and 6). On the other hand, wet specimens were rarely detected (one detection out of 5 assays). Individuals captured using MPG method had lower DNA concentration and presented significantly much lower detection rates compared to individuals captured using the “dry” method. Ballare *et al.* (2019) also found that insects collected in propylene glycol traps produced lower-quality ddRADseq assemblages. On the contrary, Ferro and Park, 2013 reported that propylene glycol is an effective DNA preservative for molecular marker-based studies on Coleoptera species. However, in that study, insects were initially killed and preserved in 100% ethanol before being stored to glycol, unlike our study where insects were initially killed in propylene glycol. The use of 100% ethanol as the initial killing agent may have resulted in better initial preservation of specimens than if the specimens had been directly exposed to propylene glycol.

The conservation method may also play a role in species detection, as the two specimens conserved using the “pinned dry” method were not detected. However, we have insufficient “pinned dry” specimens in this study to confirm this hypothesis. But numerous studies have already demonstrated that it is possible to use dry insect specimens for genetic analyses, although such types of analyses are more complex than when using fresh specimens (Wandeler et al. 2007; Hebert et al. 2013; Nakahama et al. 2018; Sire et al. 2019).

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

Ultimately, the DNA concentration of the samples appears to play an important role in their detection as DNA extracts with higher concentration were more likely to be detected (DNA concentration of 38.6 ng/µl for detected samples compared to 19.6 ng/µl for undetected samples).

False positives, negatives and unmatched OTUs

Despite precautions taken, several false positives were detected for all tested conditions. The number of false positives is significantly higher with the primer pair fwhF2/fwhR2n,

581 which is smaller in size, potentially leading to incorrect taxonomic identifications
582 compared to BF3/BR2, which is larger in size. Despite the fact that Illumina technology
583 is known to have a lower sequencing error rate compared to MinION (Silvestre-Ryan and
584 Holmes 2021), our study found ten false positives generated by Illumina, while MinION
585 had seven false positives and IonTorrent had six false positives. The sensitivity of HTS
586 technologies allows for the detection of very small amounts of DNA, thus detecting even
587 the slightest cross contamination between samples (Liu et al. 2020). These
588 contaminations may have occurred during sample collection in the field or in the
589 laboratory through cross-contamination between samples from the same study.
590
591 The false negative detections for some individuals may primarily be explained by the
592 highly heterogeneous DNA quality of the different sequenced individuals (**Tab. S1**). In
593 fact, DNA quality can be impacted by numerous mainly abiotic factors (pH, UV radiation,
594 temperature), degrading DNA quality in a matter of days/weeks (Strickler et al. 2015;
595 Collins et al. 2018; Harrison et al. 2019). During field trapping using stationary traps,
596 captured insects are sometimes exposed to such conditions (high temperatures in trap
597 containers when exposed to the sun in summer, high humidity in the container during
598 heavy rains, ...), which can greatly accelerate the speed of DNA degradation in captured
599 individuals. This degraded DNA is more difficult to be amplified, thus generating false
600 negatives, especially when attempting to detect low-abundance insects in the trap, such
601 as an invasive species in the process of establishing (Preston et al. 2022). Another
602 possible cause for the high number of false negatives is the bias induced by PCR, such as
603 uneven amplification of the DNA of the different individuals present in one sample
604 (Preston et al. 2022). To avoid potential bias arising from identification mistakes due to

errors or missing species in the references databases, we decided to work on a local and curated BLAST database. However, when target species are partially unknown, as is the case in field conditions, analyses must rely on public reference databases. Yet, out of the 35,000 known species of Cerambycidae to date, only 2,926 species (8.4%) are recorded in BOLD with a barcode fragment (as of November 16th, 2023). Furthermore, as mentioned above, databases can contain errors such as misassignment of an DNA sequence to a wrong species due to morphological identification errors. This was precisely the error encountered for the species *Monochamus sutor* who has been genetically identified as *Monochamus sartor* or the species *Leiopus nebulosus* who has been genetically identified as *Leiopus linnei* by our BOLD database.

Lack of morphological differentiation and/or identifications by non-specialists can lead to errors in databases, hindering their identification at species level. One also needs to pay attention to synonymy whereby a species is appears in the database under multiple names. We encountered this problem in our analysis with *Arhopalus ferus* (Bulks 1, 3, and 6) which was detected but under the name of *Arhopalus tristis*. Finally, mitochondrial paralogues such as Numts (non-functional copies of mitochondrial genes transported into the nuclear genome) present in databases can also bias results, making it impossible to identify correctly the species concerned (Bensasson et al. 2001). Numts are numerous in many organisms, including some cerambycids such as *Monochamus galloprovincialis (*Koutroumpa et al. 2009; Haran et al. 2015).

Biases

Based on the results obtained, it appears that the main biases observed during metabarcoding analyses on trap contents come from the degradation of DNA from

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

Conclusion

By comparing the accuracy and detection capacity of three metabarcoding strategies, this study contributes to improve our toolkit for non-native insect invasion monitoring.

All three sequencing technologies performed equally well and showed similar results for the detection and identification of exotic Cerambycid species in the traps, but as a portable, easy-to-use and cost-effective sequencer, the MinION has the potential to become an essential tool for biodiversity monitoring projects. Indeed, using the MinION reduces laboratory handling time compared with Illumina and avoids outsourcing sample sequencing, saving considerable time in detecting invasive species. This technology is precise enough to detect non-native species present at low abundances in traps and allows for accurate identifications as long as there is a sufficiently complete high-quality reference database to avoid identification errors or false positives/negatives. It is also crucial to pay close attention to issues of contamination and insect preservation during and after individual capture to work with the least degraded DNA possible.

References

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

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

676 Andújar C, Arribas P, Gray C, Bruce C, Woodward G, Yu DW, Vogler AP (2018)
677 Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill.
678 Molecular Ecology 27: 146–166. https://doi.org/10.1111/mec.14410
679
680 Aukema JE, McCullough DG, Von Holle B, Liebhold AM, Britton K, Frankel SJ (2010)
681 Historical Accumulation of Nonindigenous Forest Pests in the Continental United States.
682 BioScience 60: 886–897. https://doi.org/10.1525/bio.2010.60.11.5
683
684 Ballare KM, Pope NS, Castilla AR, Cusser S, Metz RP, Jha S (2019) Utilizing field collected
685 insects for next generation sequencing: Effects of sampling, storage, and DNA extraction
686 methods. Ecology and Evolution 9: 13690–13705. https://doi.org/10.1002/ece3.5756
687
688 Batovska J, Piper AM, Valenzuela I, Cunningham JP, Blacket MJ (2021) Developing a non-
689 destructive metabarcoding protocol for detection of pest insects in bulk trap catches.
690 Scientific Reports 11: 7946. https://doi.org/10.1038/s41598-021-85855-6
691
692 Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes:
693 evolution's misplaced witnesses. Trends in Ecology & Evolution 16: 314–321.
694 https://doi.org/10.1016/s0169-5347(01)02151-6
695
696 Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU,
697 Richardson DM (2011) A proposed unified framework for biological invasions. Trends in
698 Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023
699

Bradshaw CJA, Leroy B, Bellard C, Roiz D, Albert C, Fournier A, Barbet-Massin M, Salles J-M, Simard F, Courchamp F (2016) Massive yet grossly underestimated global costs of invasive insects. Nature Communications 7: 12986. https://doi.org/10.1038/ncomms12986

Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, De Waard JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources 19: 711–727. https://doi.org/10.1111/1755-0998.13008

Brown SDJ, Collins RA, Boyer S, Lefort M, Malumbres-Olarte J, Vink CJ, Cruickshank RH (2012) S PIDER : An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources 12: 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x

Chain FJJ, Brown EA, MacIsaac HJ, Cristescu ME (2016) Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Cowie R (Ed.). Diversity and Distributions 22: 493–504. https://doi.org/10.1111/ddi.12427

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA (2015) Biological invasions, climate change and genomics. Evolutionary Applications 8: 23–46. https://doi.org/10.1111/eva.12234

Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN (2023) Future of DNA-based insect monitoring. Trends in Genetics 39: 531–544. https://doi.org/10.1016/j.tig.2023.02.012

Collins RA, Wangensteen OS, O'Gorman EJ, Mariani S, Sims DW, Genner MJ (2018) Persistence of environmental DNA in marine systems. Communications Biology 1: 185. https://doi.org/10.1038/s42003-018-0192-6

Dean MD, Ballard JWO (2001) Factors affecting mitochondrial DNA quality from museum preserved Drosophila simulans. Entomologia Experimentalis et Applicata 98: 279–283. https://doi.org/10.1046/j.1570-7458.2001.00784.x

deWaard JR, Levesque-Beaudin V, deWaard SL, Ivanova NV, McKeown JTA, Miskie R, Naik S, Perez KHJ, Ratnasingham S, Sobel CN, Sones JE, Steinke C, Telfer AC, Young AD, Young MR, Zakharov EV, Hebert PDN (2019) Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding 1. Genome 62: 85–95. https://doi.org/10.1139/gen-2018-0093

Dopheide A, Tooman LK, Grosser S, Agabiti B, Rhode B, Xie D, Stevens MI, Nelson N, Buckley TR, Drummond AJ, Newcomb RD (2019) Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. Ecological Applications 29: 1–14. Available from: https://www.jstor.org/stable/26669248 (June 9, 2023).

748 Egeter B, Peixoto S, Brito JC, Jarman S, Puppo P, Velo-Antón G (2018) Challenges for
749 assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA.
750 Genome 61: 807–814. https://doi.org/10.1139/gen-2018-0071
751
752 Egeter B, Veríssimo J, Lopes-Lima M, Chaves C, Pinto J, Riccardi N, Beja P, Fonseca NA
753 (2022) Speeding up the detection of invasive bivalve species using environmental DNA:
754 A Nanopore and Illumina sequencing comparison. Molecular Ecology Resources 22:
755 2232–2247. https://doi.org/10.1111/1755-0998.13610
756
757 Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers
758 for Freshwater Macroinvertebrate Bioassessment. Frontiers in Environmental Science 5.
759 Available from: https://www.frontiersin.org/articles/10.3389/fenvs.2017.00011 (July 2,
760 2023).
761
762 Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and
763 weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine
764 stream monitoring. Yu D (Ed.). Methods in Ecology and Evolution 8: 1265–1275.
765 https://doi.org/10.1111/2041-210X.12789
766
767 Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov
768 EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for
769 terrestrial arthropods. PeerJ 7: e7745. https://doi.org/10.7717/peerj.7745
770

771 Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-
772 Raquet G, Combes S, Pascal G (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution.
773 Berger B (Ed.). Bioinformatics 34: 1287–1294.
774 https://doi.org/10.1093/bioinformatics/btx791
775
776 Ferro ML, Park J-S (2013) Effect of Propylene Glycol Concentration on Mid-Term DNA
777 Preservation of Coleoptera. In: The Coleopterists Bulletin. , 581–586.
778 https://doi.org/10.1649/0010-065X-67.4.581
779
780 Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-
781 generation sequencing data. Bioinformatics (Oxford, England) 28: 3150–3152.
782 https://doi.org/10.1093/bioinformatics/bts565
783
784 Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, Hallwachs W,
785 Hajibabaei M (2014) Simultaneous assessment of the macrobiome and microbiome in a
786 bulk sample of tropical arthropods through DNA metasystematics. Proceedings of the
787 National Academy of Sciences of the United States of America 111: 8007–8012.
788 https://doi.org/10.1073/pnas.1406468111
789
790 Giovani B, Blümel S, Lopian R, Teulon D, Bloem S, Galeano Martínez C, Beltrán Montoya C,
791 Urias Morales CR, Dharmapuri S, Timote V, Horn N, Chouibani M, Mezui M'Ella JG,
792 Herrera V, Castinel A, Goletsos C, Moeller C, Naumann I, Stancanelli G, Bronzwaer S,
793 Tramontini S, MacDonald P, Matheson L, Anthoine G, De Jonghe K, Schenk M, Steinmöller
794 S, Rodriguez E, Cruz ML, Luck J, Fraser G, Brunel S, Montuori M, Fedchock C, Steel E,

795 Pennington HG, Day R, Rossi JP, Xia J (2020) Science diplomacy for plant health. Nature
796 Plants 6: 902–905. https://doi.org/10.1038/s41477-020-0744-x
797
798 Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G (2015) Ghost mtDNA haplotypes
799 generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and
800 phylogeographic pattern. Journal of Zoological Systematics and Evolutionary Research
801 53: 109–115. https://doi.org/10.1111/jzs.12095
802
803 Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the
804 environment and implications for studying biodiversity. Proceedings of the Royal Society
805 B: Biological Sciences 286: 20191409. https://doi.org/10.1098/rspb.2019.1409
806
807 Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c
808 oxidase subunit 1 divergences among closely related species. Proceedings of the Royal
809 Society B: Biological Sciences 270: S96–S99. https://doi.org/10.1098/rsbl.2003.0025
810
811 Hebert PDN, deWaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, Mantle B,
812 La Salle J (2013) A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural
813 History Collection. Kolokotronis S-O (Ed.). PLoS ONE 8: e68535.
814 https://doi.org/10.1371/journal.pone.0068535
815
816 Henrie JR, Thomson BM, Yungfleisch AA, Kerr M, Leavitt SD (2022) Characterizing
817 Crustose Lichen Communities—DNA Metabarcoding Reveals More than Meets the Eye.
818 Diversity 14: 766. https://doi.org/10.3390/d14090766

819

820 Ho JKI, Puniamoorthy J, Srivathsan A, Meier R (2020) MinION sequencing of seafood in
821 Singapore reveals creatively labelled flatfishes, confused roe, pig DNA in squid balls, and
822 phantom crustaceans. Food Control 112: 107144.
823 https://doi.org/10.1016/j.foodcont.2020.107144

824

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

828

829 Javal M, Terblanche J, Conlong D, Delahaye N, Grobbelaar E, Benoit L, Lopez-Vaamonde C,
830 Haran J (2021) DNA barcoding for bio-surveillance of emerging pests and species
831 identification in Afrotropical Prioninae (Coleoptera, Cerambycidae). Biodiversity Data
832 Journal 9: e64499. https://doi.org/10.3897/BDJ.9.e64499

833

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

838

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

842

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

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

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

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

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2: e593. https://doi.org/10.7717/peerj.593

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10. https://doi.org/10.14806/ej.17.1.200

867 Mas H, Santoiemma G, Lencina JL, Gallego D, Pérez-Laorga E, Ruzzier E, Rassati D (2023)
868 Investigating beetle communities in and around entry points can improve surveillance at
869 national and international scale. NeoBiota 85: 145–165.
870 https://doi.org/10.3897/neobiota.85.103904
871
872 Mazza G, Tricarico E, Genovesi P, Gherardi F (2014) Biological invaders are threats to
873 human health: an overview. Ethology Ecology & Evolution 26: 112–129.
874 https://doi.org/10.1080/03949370.2013.863225
875
876 Nakahama N, Uchida K, Ushimaru A, Isagi Y (2018) Historical changes in grassland area
877 determined the demography of semi-natural grassland butterflies in Japan. Heredity
878 121: 155–168. https://doi.org/10.1038/s41437-018-0057-2
879
880 Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019)
881 Prospects and challenges of implementing DNA metabarcoding for high-throughput
882 insect surveillance. GigaScience 8: giz092. https://doi.org/10.1093/gigascience/giz092
883
884 Preston, M., Fritzsche, M. & Woodcock, P. 2022. Understanding and mitigating errors and
885 biases in metabarcoding: an introduction for non-specialists. JNCC Report No. 699, JNCC,
886 Peterborough, ISSN 0963-8091.
887
888 Ratnasingham S, Hebert PDN (2007) BARCODING: bold: The Barcode of Life Data System
889 (http://www.barcodinglife.org): BARCODING. Molecular Ecology Notes 7: 355–364.
890 https://doi.org/10.1111/j.1471-8286.2007.01678.x

891

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

896 Roques A, Auger-Rozenberg M-A, Blackburn TM, Garnas J, Pyšek P, Rabitsch W,
897 Richardson DM, Wingfield MJ, Liebhold AM, Duncan RP (2016) Temporal and
898 interspecific variation in rates of spread for insect species invading Europe during the
899 last 200 years. Biological Invasions 18: 907–920. https://doi.org/10.1007/s10530-016-
900 1080-y

901

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

913 Rossa R, Goczał J (2021) Global diversity and distribution of longhorn beetles
914 (Coleoptera: Cerambycidae). The European Zoological Journal 88: 289–302.
915 https://doi.org/10.1080/24750263.2021.1883129
916
917 Sahlin K, Medvedev P (2020) De Novo Clustering of Long-Read Transcriptome Data
918 Using a Greedy, Quality Value-Based Algorithm. Journal of Computational Biology: A
919 Journal of Computational Molecular Cell Biology 27: 472–484.
920 https://doi.org/10.1089/cmb.2019.0299
921
922 Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P,
923 Hulme PE, Kleunen M, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J,
924 Pyšek P, Winter M, Essl F (2021) Projecting the continental accumulation of alien species
925 through to 2050. Global Change Biology 27: 970–982.
926 https://doi.org/10.1111/gcb.15333
927
928 Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P,
929 van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG,
930 Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP,
931 Fuentes N, Guénard B, Jäger H, Kartesz J, Kenis M, Kühn I, Lenzner B, Liebhold AM,
932 Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-
933 Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K,
934 Tokarska-Guzik B, Walker K, Ward DF, Yamanaka T, Essl F (2018) Global rise in emerging
935 alien species results from increased accessibility of new source pools. Proceedings of the
936 National Academy of Sciences 115. https://doi.org/10.1073/pnas.1719429115

Silvestre-Ryan J, Holmes I (2021) Pair consensus decoding improves accuracy of neural network basecallers for nanopore sequencing. Genome Biology 22: 38. https://doi.org/10.1186/s13059-020-02255-1

Sire L, Gey D, Debruyne R, Noblecourt T, Soldati F, Barnouin T, Parmain G, Bouget C, Lopez-Vaamonde C, Rougerie R (2019) The Challenge of DNA Barcoding Saproxylic Beetles in Natural History Collections—Exploring the Potential of Parallel Multiplex Sequencing With Illumina MiSeq. Frontiers in Ecology and Evolution 7. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2019.00495 (November 28, 2023).

Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R (2021) ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. BMC Biology 19: 217. https://doi.org/10.1186/s12915-021-01141-x

Strickler KM, Fremier AK, Goldberg CS (2015) Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biological Conservation 183: 85–92. https://doi.org/10.1016/j.biocon.2014.11.038

Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. https://doi.org/10.3897/mbmg.1.14625

961 Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly
962 from long uncorrected reads. Genome Research 27: 737–746.
963 https://doi.org/10.1101/gr.214270.116
964
965 Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in
966 population genetics. Trends in Ecology & Evolution 22: 634–642.
967 https://doi.org/10.1016/j.tree.2007.08.017
968
969 Wang Q (2017) Cerambycidae of the World: Biology and Pest Management. CRC Press,
970 643 pp.
971
972 Wu Y, Trepanowski NF, Molongoski JJ, Reagel PF, Lingafelter SW, Nadel H, Myers SW, Ray
973 AM (2017) Identification of wood-boring beetles (Cerambycidae and Buprestidae)
974 intercepted in trade-associated solid wood packaging material using DNA barcoding and
975 morphology. Scientific Reports 7: 40316. https://doi.org/10.1038/srep40316
976
977 Zenker MM, Rougerie R, Teston JA, Laguerre M, Pie MR, Freitas AVL (2016) Fast Census
978 of Moth Diversity in the Neotropics: A Comparison of Field-Assigned Morphospecies and
979 DNA Barcoding in Tiger Moths. Cimmaruta R (Ed.). PLOS ONE 11: e0148423.
980 https://doi.org/10.1371/journal.pone.0148423
981

982 **Acknowledgements**
983 We would like to thank all colleagues who participated in the taxa sampling (see Roques
984 et al. 2023). We are thankful to Lucas Sire for insightful discussion on primer choice. The

authors thank the GeT-Biopuces platform of INSA Toulouse for the IonTorrent sequencing study. This work was supported by the PORTRAP project “Test de l’efficacité de pièges génériques multicomposés pour la détection précoce d’insectes exotiques xylophages dans les sites potentiels d’entrée sur le territoire national” and HOMED project (HOlistic Management of Emerging Forest Pests and Diseases) which received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 771271 (https://homed-projecteu/). We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing help and computing storage resources. Loïs Veillat was supported by a PhD studentship from HOMED project and doctoral school SSBCV at the university of Orléans.

Data accessibility

Barcode data for the 33 species used in the mock community experiment are available from BOLD in the dataset DS-MINION (dx.doi.org/10.5883/DS-MINION). Raw sequence data for this project and analytical script and files are available on figshare (https://figshare.com/projects/DNA_metabarcoding_an_efficient_way_to_detect_non-native_cerambycid_beetles_in_trapping_collections_/171432).

Author contributions

Loïs Veillat, Géraldine Roux, Carlos Lopez-Vaamonde and Stéphane Boyer conceived the study. Alain Roques collected field samples. Stéphane Boyer, Marina Querejeta, Emmanuelle Magnoux and Loïs Veillat conducted the laboratory sample processing. Loïs Veillat analysed the data and wrote the first draft. All authors contributed to the

preparation of the manuscript. Both senior authors, Géraldine Roux and Carlos Lopez-Vaamonde, contributed equally to this study.