Ecological Entomology

*Ecological Entomology* (2020), **45**, 373–385 DOI: 10.1111/een.12831

INVITED REVIEW

# A practical guide to DNA metabarcoding for entomological ecologists

MINGXIN LIU,[1,2] LAURENCE J. CLARKE,[3] SUSAN C. BAKER,[1,2] GREGORY J. JORDAN[1] and CHRISTOPHER P. BURRIDGE[1] [1]School of Natural Sciences, University of Tasmania, Hobart, Australia, [2]ARC Centre for Forest Value, University of Tasmania, Hobart, Australia and [3]Antarctic Climate and Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Australia

**Abstract.** 1. DNA metabarcoding is a cost-effective species identification approach with great potential to assist entomological ecologists. This review presents a practical guide to help entomological ecologists design their own DNA metabarcoding studies and ensure that sound ecological conclusions can be obtained.

2. The review considers approaches to field sampling, laboratory work, and bioinformatic analyses, with the aim of providing the background knowledge needed to make decisions at each step of a DNA metabarcoding workflow.

3. Although most conventional sampling methods can be adapted to DNA metabarcoding, this review highlights techniques that will ensure suitable DNA preservation during field sampling and laboratory storage. The review also calls for a greater understanding of the occurrence, transportation, and deposition of environmental DNA when applying DNA metabarcoding approaches for different ecosystems.

4. Accurate species detection with DNA metabarcoding needs to consider biases introduced during DNA extraction and PCR amplification, cross-contamination resulting from inappropriate amplicon library preparation, and downstream bioinformatic analyses. Quantifying species abundance with DNA metabarcoding is in its infancy, yet recent studies demonstrate promise for estimating relative species abundance from DNA sequencing reads.

5. Given that bioinformatics is one of the biggest hurdles for researchers new to DNA metabarcoding, several useful graphical user interface programs are recommended for sequence data processing, and the application of emerging sequencing technologies is discussed.

**Key words.** Biodiversity monitoring, DNA preservation, ecological inferences, environmental DNA, high-throughput sequencing, invertebrates, metabarcoding.

## Introduction

DNA metabarcoding is a developing approach that identifies multiple species from a mixed sample (bulk DNA or eDNA) based on high-throughput sequencing (HTS) of a specific DNA marker. It differs from conventional DNA barcoding (usually based on Sanger DNA sequencing of individual specimens) because the amount of DNA sequence data derived by HTS allows taxonomy to be rapidly assigned to many species present in a sample. DNA metabarcoding is rapidly emerging as a cost-effective approach for large-scale studies sampling environments where constraints of conventional morphology-based species identification is logistically or financially impractical.

Invertebrates are among the most studied taxa in DNA metabarcoding studies. These include studying the responses of terrestrial biodiversity to land-use change (Wood *et al.*, 2017; Schmidt *et al.*, 2018), monitoring ecological dynamics of lake macroinvertebrates (Bista *et al.*, 2018), and biodiversity assessment in marine arthropods (Fonseca *et al.*, 2010; Sinniger *et al.*, 2016). Shifting conventional field surveys of

Correspondence: Mingxin Liu, School of Natural Sciences, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia. E-mail: mingxin.liu@utas.edu.au

invertebrates to DNA metabarcoding is often feasible but is not yet widely applied. However, both bulk sample and eDNA studies provide substantial research opportunities, the former because of potential cost efficiencies and the latter because of its capacity to retrieve species DNA from samples of soil, water, sediments or other material, in the physical absence of the species itself.

DNA metabarcoding workflows are intrinsically different from conventional invertebrate identification approaches in that they use DNA as a proxy for species detection. Understanding key characteristics of these workflows is essential to guide decisions about methodology. First, DNA metabarcoding depends on sample collection and handling methods that ensure DNA preservation. Second, because routine DNA metabarcoding workflows consist of multiple laboratory steps (i.e. DNA extraction, PCR amplification, and DNA sequencing), sufficient technical knowledge and informed choice at each step are required for reliable species detection. Third, interpreting HTS data requires specialised bioinformatics skills. For entomologists without the skills or facilities to conduct genetic analyses in-house, most aspects of the DNA metabarcoding workflow can be outsourced to commercial laboratories.

This review is a practical guide for entomological ecologists wishing to design their own DNA metabarcoding studies. First, based on a literature review and experience, we discuss sampling methods, highlight their characteristics pertinent to DNA metabarcoding and provide guidance for conducting invertebrate collection from terrestrial, aquatic and sedimentary environments. Second, we discuss practical sample handling procedures and laboratory experimental considerations. We also consider recent research efforts in mitigating species detection biases. This section addresses important methodological issues that may be overlooked by entomological ecologist end-users. Third, we describe useful graphical user interface (GUI) programs for bioinformatic analysis that forego the need for command-line proficiency. Finally, we briefly discuss potential future applications of state-of-the-art HTS platforms/technologies that could expand the current capability of DNA metabarcoding approaches.

A glossary of terms and definitions is available in Box 1.

## Sample collection for DNA metabarcoding studies

DNA metabarcoding studies inevitably start with either field sampling or assessment of the suitability of existing samples. These processes involve critical decisions to ensure that the samples will reliably answer the study questions. Careful consideration of sampling and curation procedures is needed to avoid DNA contamination and to ensure DNA preservation.

### *Contamination*

Although reusing equipment can be acceptable for conventional entomological sampling, the sensitivity of HTS to detect trace amounts of DNA creates substantial risks of sample contamination through transfer of DNA. The use of new equipment for each sample is appropriate in some cases, but decontamination is the best solution for many studies. For example, during eDNA sampling, a trowel or soil corer is often decontaminated and reused repeatedly (e.g. Ritter *et al.*, 2018).

Washing equipment between each sample is a common and practical way to minimise DNA contamination, and is particularly important for eDNA studies and less vital when processing bulk community samples. Protocols range from a simple wash with bleach, soapy water, distilled water or ethanol, to a stringent stepwise procedure consisting of washing with soapy water, rinsing with distilled water, rinsing with ethanol, and eliminating RNases and DNases (Erdozain *et al.*, 2019). Nevertheless, soaking equipment in 10% bleach for at least 10 min is

**Box 1.** Glossary.

**Bioinformatics.** Applying computational approaches for processing complex biological data.

**Bioinformatics pipeline.** Set of data-processing elements connected in series, where the output of one element is the input for the next.

**Bulk sample DNA**. The DNA mixture extracted from multiple specimens simultaneously.

**DNA barcode.** A DNA marker routinely used for identifying species – e.g. the mitochondrial cytochrome *c* oxidase I (COI) gene is the standard DNA barcode for identifying animal species.

**DNA marker.** Any gene or DNA fragment that is used to identify a species, individual or genotype.

**DNA metabarcoding.** Methods that identify species from bulk sample DNA or environmental DNA via high-throughput sequencing of a DNA marker.

**DNA taphonomy.** The natural processes of preservation, transportation, and deposition of environmental DNA.

**Environmental DNA (eDNA).** DNA extracted from environmental samples, such as soil, water, or sediment, without prior isolation of target organisms (Taberlet *et al.*, 2012).

**High-throughput sequencing (HTS).** The simultaneous sequencing of millions of DNA fragments.

**Operational taxonomic unit (OTU).** OTUs are often used as a proxy for taxonomic species in DNA metabarcoding studies when it is not possible to accurately determine species boundaries; they are typically constructed on the basis of DNA sequence similarity.

**PCR amplification.** This is the use of PCR to create many copies of a target DNA fragment. The fragment is targeted using a pair of primers (forward and reverse) that bind to the DNA flanking fragment of interest.

**Sequencing depth.** The number of DNA sequences recovered (typically per sample) during high-throughput sequencing.

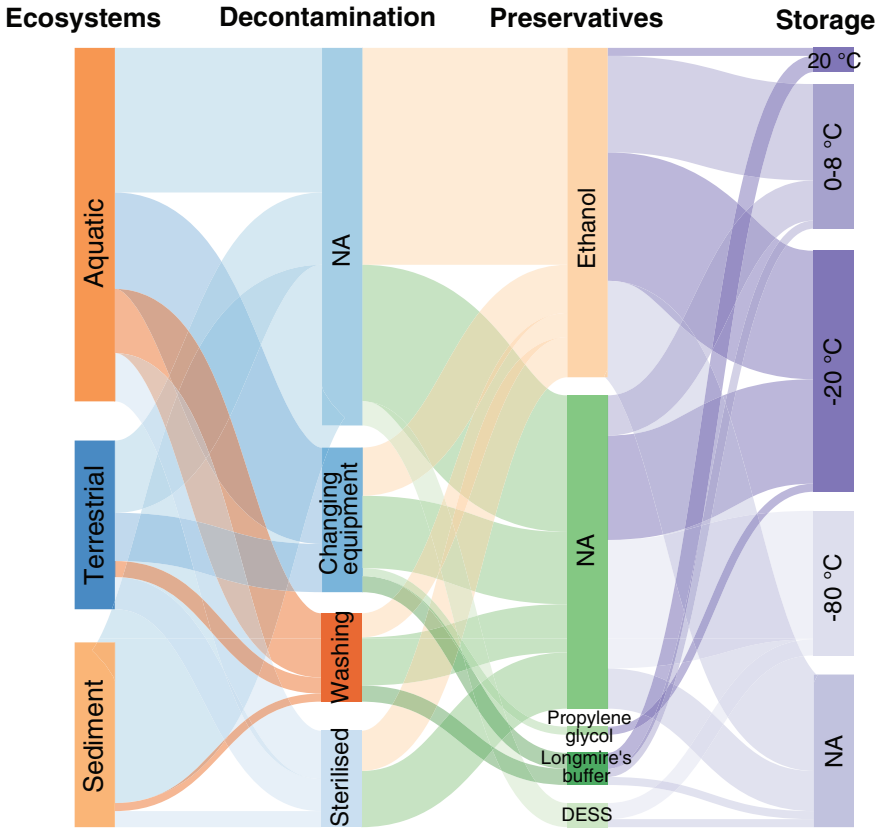


**Fig. 1.** Relative frequencies of four stages of sampling practices from 84 empirical DNA metabarcoding studies published between 2015 and 2019, including ecosystem sampled and procedures of sample preservation, storage, and steps to minimise sample contamination. The thickness of the lines linking categories were weighted by the number of papers. 'NA' represents studies where the corresponding treatment was not reported or was not needed. For example, 'NA' in preservatives for aquatic samples means that no preservative was used. The diagram was created with SANKEYMATIC (http://sankeymatic.com/build/).

a simple but effective method for DNA decontamination (Prince & Andrus, 1992; Kemp & Smith, 2005; Fernandez *et al.*, 2018; Jeunen *et al.*, 2018). Further, we recommend applying this method in the field together with a rinse treatment in ethanol or sterilised water for removing remaining bleach. Regardless of decontamination procedure, including a field negative control (e.g. a swab of the sampling tool) is strongly recommended for quantifying potential contamination (Dickie *et al.*, 2018). Use of sterile plastic bags or vials is acceptable for sample transfer, and wearing sterile gloves during field sampling reduces the amount of human DNA introduced into samples. When possible, single-use sterilised equipment minimises the potential for cross-contamination (e.g. water filters, vials, tubes, bottles). If not purchased sterile and decontaminated, plasticware (e.g. tubes and bottles) can be decontaminated beforehand by UV exposure (30 min), bleach washing, or autoclaving in the laboratory.

### *DNA preservation*

Preservation methods strongly impact the quality of DNA. Protocols that prevent DNA degradation during sampling and curation vary according to the source of the samples (e.g. terrestrial, freshwater, marine, and sediments) and whether bulk samples or eDNA are targeted. The choice of trapping/trappant in studies where invertebrates accumulate in the trap over long periods (e.g. pitfall trap studies) needs particular care to ensure preservation. Both the type of preservative and the conditions of transportation and storage are important. For trapping terrestrial invertebrates, most studies used > 95% molecular grade ethanol as preservative (Fig. 1), although food-grade propylene glycol is a good alternative due to its non-toxicity, non-flammability, and low rates of evaporation (Ferro & Park, 2013; Patrick *et al.*, 2016). Certain other preservatives, such as ethylene glycol, more dilute ethanol (< 70%), and dimethyl sulphoxide, are unsuitable for DNA preservation (Vink *et al.*, 2005; Moreau *et al.*, 2013), and formaldehyde must be avoided as it actively degrades DNA. Water samples were usually vacuum-pumped and filtered through a membrane to collect aquatic eDNA. Preserving the filter membrane in ethanol (Spens *et al.*, 2017; Macher *et al.*, 2018; Serrana *et al.*, 2018) or Longmire's buffer (Longmire *et al.*, 1997; Grey *et al.*, 2018; Lacoursiere-Roussel *et al.*, 2018), is effective. Temperature is a key factor for DNA preservation. Ambient or room temperature is only suitable for very short-term storage; most studies store samples below room

temperature to reduce DNA degradation (Fig. 1). Nevertheless, depending on ambient climatic conditions and choice of killing agent and preservative, traps such as pitfall traps and intercept traps can be operated for several days to weeks in the field before sample collection and processing (Barsoum *et al.*, 2019; Dopheide *et al.*, 2019). We recommend creating cool conditions for temporary storage and transportation from the field using an ice box or dry ice. Most studies reported storage conditions of −20 or −80°C after transporting samples to the laboratory.

*Special considerations for sampling eDNA*

The widely used approach for retrieving eDNA from water samples involves filtration, ethanol precipitation, and centrifugation (Tsuji *et al.*, 2019). As a result, sampling biodiversity is constrained by the volume of processed water due to limitations of filter membranes (e.g. pore size). These constraints may be responsible for observed heterogeneity of biodiversity among samples (Grey *et al.*, 2018; Beentjes *et al.*, 2019), and therefore a large number of replicate samples may be required for accurate species detection (Ficetola *et al.*, 2014). For better estimation of biodiversity using a filtration method, three to nine sampling replicates (*c*. 300 ml water per replicate) within each study site are recommended (Grey *et al.*, 2018; Beentjes *et al.*, 2019). Using ethanol precipitation is suitable for small volume water samples (e.g. < 90 ml; Harper *et al.*, 2019) in areas where access to infrastructure for filtration is difficult. It also creates good consistency of water volume across samples. Although undenatured ethanol is preferred for precipitation (Creer *et al.*, 2016), its use may be constrained logistically or financially.

Many eDNA studies have adopted quantitative procedures for field sampling, such as soil sampling with corers (Ritter *et al.*, 2018; Zinger *et al.*, 2018), filtering fixed volumes of water (Lacoursiere-Roussel *et al.*, 2018; Macher *et al.*, 2018), and using artificial substrate units for macroinvertebrate sampling (Cahill *et al.*, 2018). Standardised and robust protocols that provide quantitative and reproducible results will increase the uptake of eDNA metabarcoding (Dickie *et al.*, 2018). There have been only a few standardised sampling protocols for quantitative sampling, such as for soil invertebrates (e.g. ISO 23611, parts 1–6), aquatic macroinvertebrates (Carew *et al.*, 2018), and marine sediments (Aylagas *et al.*, 2016). The DNAqua-Net initiative has played an important role in establishing standard and routine protocols of eDNA for aquatic biomonitoring under the European framework (Leese *et al.*, 2018). Future eDNA metabarcoding studies should also focus on developing standardised and quantitative sampling practices (Dickie *et al.*, 2018).

Sediment eDNA is a promising tool to probe community dynamics through time, provided that the taphonomic processes of preservation, transportation and deposition of eDNA are taken into account (Balint *et al.*, 2018). Periodically sampling the surface sediment layer is common for studies investigating contemporary communities, such as seasonal changes (Salonen *et al.*, 2019), community succession after perturbation (Brannock *et al.*, 2017; Xie *et al.*, 2018), and general monitoring (Chariton *et al.*, 2015; Lobo *et al.*, 2017). However, sampling sediments raises the issue of whether the sample represents the present community or a previously existing community, and there is little agreement regarding the boundary separating present-day and historical DNA in sediments (Thomsen & Willerslev, 2015). Further research is needed to understand the source of DNA in sediments in terms of how and when they were preserved and from where they were transported.

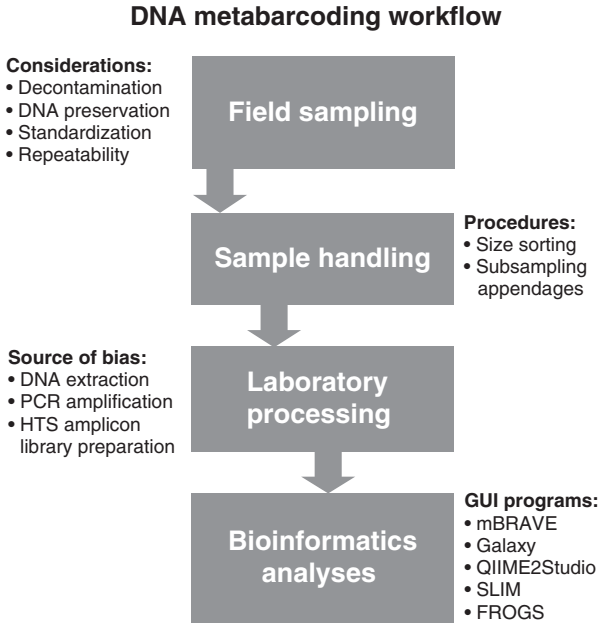


**Fig. 2.** Considerations and techniques for each of the key steps in DNA metabarcoding.

## Minimising bias in the laboratory

*Useful sample handling procedures*

Invertebrates in bulk samples are typically highly diverse in species composition, relative abundance, and biomass. Each of these sources of variation affects accurate estimation of community composition with DNA metabarcoding (Deiner *et al.*, 2015; Nichols *et al.*, 2018; Piñol *et al.*, 2018; Zinger *et al.*, 2019). However, several methods for pre-processing samples can improve species detection (Fig. 2). Size-sorting is often used because larger specimens tend to release more DNA and thus dominate the total sequence count in DNA metabarcoding studies (Deagle *et al.*, 2018a). Thus, sorting macroinvertebrates into multiple (e.g. two or three) size classes and then pooling the digested tissue according to the number of specimens in each class can reduce sequencing distortion by large taxa and better detect smaller taxa (Elbrecht *et al.*, 2017b). However, size-sorting is time-consuming; thus the decision of whether to size-sort samples should depend on the complexity of samples, sequencing depth, and the necessity of recovering small and/or rare taxa.

An alternative to size-sorting whole specimens is to use appendages from insect specimens, such as pulling a leg from

specimens, e.g. those as big as or bigger than a honeybee (Ji *et al.*, 2013), while the whole body is used if it is below the size threshold. Nevertheless, Creedy *et al.* (2019) showed that taxonomic composition complexity, rather than size differences between taxa, is the main factor affecting accurate recovery of community composition in DNA metabarcoding studies. Removing appendages from larger specimens is also useful when it is desired to retain specimens, e.g. for inclusion in arthropod collections.

### *Selecting molecular laboratory methods*

Biases in DNA metabarcoding species detection can also be introduced during DNA extraction and PCR amplification (Fig. 2). A certain degree of bias is unavoidable but can be reduced using appropriate molecular laboratory methods.

DNA extraction is most commonly performed with commercial DNA extraction kits. Dopheide *et al.* (2018) recovered incongruous community composition from eDNA when testing four DNA extraction protocols and different soil volumes, and suggested that studies should maintain a single DNA extraction method and employ large soil volume for invertebrate taxa. For water eDNA samples, the use of filtration and DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) is recommended by Deiner *et al.* (2015) for eukaryote biodiversity surveys, because this combination showed good detection rates for both lotic and rare species (see also Tsuji *et al.*, 2019). Although the Qiagen DNeasy Blood & Tissue kit is widely used for terrestrial bulk invertebrates, the PowerPlant Kit performs better for benthic macroinvertebrates compared with two other tested kits, because it removes PCR inhibitors (Majaneva *et al.*, 2018). Recently, a modular universal DNA extraction toolkit has provided an adaptable and inexpensive method that can replace widely used commercial kits for different sample types, such as tissue, soil and water samples (Sellers *et al.*, 2018).

The success of DNA metabarcoding relies on DNA marker selection, requiring careful consideration of both the DNA marker with respect to its ability to distinguish taxa and the availability of reference sequences for assigning taxonomy (Deagle *et al.*, 2014). Ideally, DNA metabarcoding markers should have sufficiently conserved flanking primer-binding sites to minimise taxonomic bias during PCR amplification, while the intervening sequence is sufficiently variable for species identification. There is also a trade-off related to the sequence length of the marker: longer DNA fragments should provide better taxonomic resolution, while shorter fragments (e.g. 100 bp or less) are more likely to be sequenced, especially in eDNA. PCR primer sequences determine the DNA fragment amplified, and, to varying extents, the taxa to be targeted by PCR amplification (Fig. 3). While choosing evolutionarily conserved primer-binding sites increases the taxonomic coverage of the DNA marker, degenerate bases can be used to increase coverage for taxa where the primer-binding sites are not conserved (Fig. 3), which generally improves taxon recovery (Ji *et al.*, 2013; Clarke *et al.*, 2017; Elbrecht & Leese, 2017). The primer database (http://boldsystems.org/index.php/Public_Primer_PrimerSearch) in the Barcode of Life Database (BOLD) system (Ratnasingham & Hebert, 2007) provides a searchable collection of published primers. Researchers can determine candidate primers by searching the database with studied taxa, but *in silico* and *in vivo* evaluations are also encouraged for validation, especially for less studied taxonomic groups.

The mitochondrial cytochrome *c* oxidase I (COI) gene is a useful DNA marker for many entomological studies. Its relatively high mutation rate ensures the ability to distinguish animal taxa, and several sufficiently conserved regions provide a range of potentially suitable primer-binding sites (Elbrecht & Leese, 2017; Elbrecht *et al.*, 2017a; Rennstam Rubbmark *et al.*, 2018). However, the co-amplification of nuclear mitochondrial pseudogenes (NUMTs) could result in overestimation of the number of true species in DNA metabarcoding studies (Song *et al.*, 2008; Wang *et al.*, 2018). Nuclear mitochondrial pseudogenes can be identified based on in-frame stop codons, insertions or deletions, but are otherwise challenging to detect (Song *et al.*, 2008). Given that COI is protein-coding, it also allows errors resulting from PCR amplification or HTS sequencing to be removed bioinformatically (Andujar *et al.*, 2018). Furthermore, COI is much better represented in sequence databases than are other markers (Box 2). However, it is more challenging to identify NUMTs that do not have any in-frame stop codons, insertions or deletions. The mitochondrial 16S and 12S ribosomal RNA genes have been used as alternative animal markers (Clarke *et al.*, 2014; Elbrecht *et al.*, 2016; Marquina *et al.*, 2018; Zhang *et al.*, 2018), but overall, we recommend using primers based on the COI barcode (Elbrecht *et al.*, 2019). The use of multiple or alternative markers does not necessarily improve species detection (Elbrecht *et al.*, 2019).

Several other factors related to PCR amplification can impact the accuracy of DNA metabarcoding. Using PCR replicates can reduce false-negative detections (Alberdi *et al.*, 2018), and five to 10 replicates are particularly effective in recovering rare taxa from a given sample (Ficetola *et al.*, 2014; Dopheide *et al.*, 2018). Furthermore, if each PCR amplicon replicate is separately dual-tagged (two unique tags; one at each end of fragment; see Fig. 3) for each sample (Zepeda-Mendoza *et al.*, 2016), potential biases and errors introduced from cross-contamination (Minich *et al.*, 2019) and tag-jumping events (the incorrect assignment of sequences back to samples based on the erroneous production of DNA tag combination identifiers) (Schnell *et al.*, 2015) can be detected. Last, use of high-fidelity DNA polymerases is recommended to reduce PCR error rates (Pan *et al.*, 2014; Nichols *et al.*, 2018; Sze & Schloss, 2019).

### *Quantifying species abundance or biomass*

DNA metabarcoding is mostly used to detect the presence of species in samples rather than their relative abundance, although sequence frequency is sometimes used as a proxy of species abundance (Aizpurua *et al.*, 2018; Deagle *et al.*, 2018b). Several studies have demonstrated positive relationships between input species biomass and output sequencing reads. For example, Bista *et al.* (2018) detected a strong ($R^2 = 0.83$) positive correlation between biomass input and HTS read abundance for the beetle *Gyrinus marinus*. However, only a weak correlation was detected for other arthropod species, and a recent meta-analysis

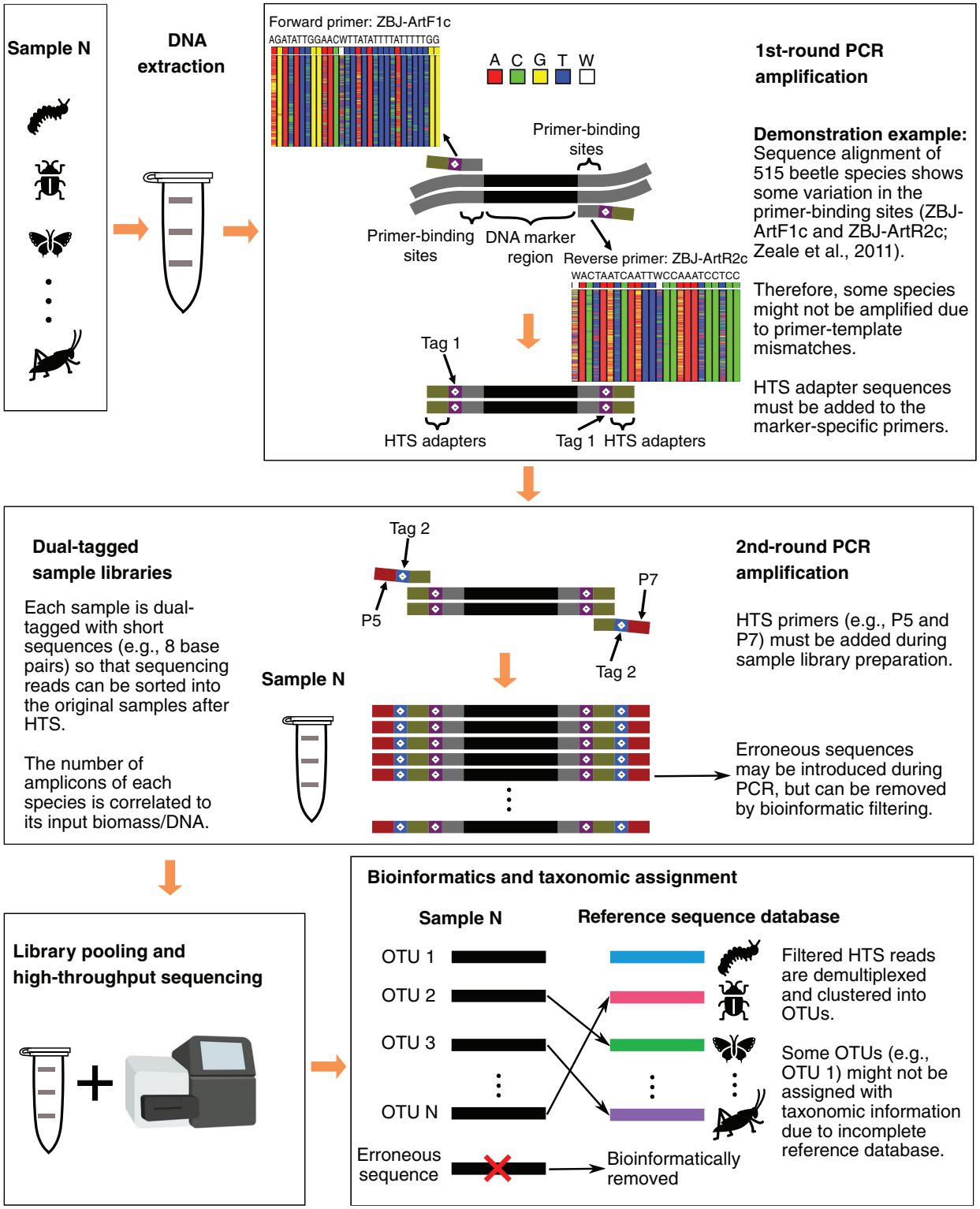


**Fig. 3.** Schematic diagram of methodological procedures involved in PCR-based DNA metabarcoding studies. HTS, high-throughput sequencing.

**Box 2.** Taxonomic assignment with reference sequence databases.

Reference sequence databases are generated by sequencing targeted DNA markers (e.g. COI) from curated morphological specimens. Although output DNA sequences in DNA metabarcoding can be bioinformatically curated and assigned to operational taxonomic units (OTUs), their information value is greatly enhanced if they can be matched to DNA sequences associated with physical reference collections. Notably, OTU curation methods based on sequence similarity (e.g. 97%), although they are widely used, are recognised to be arbitrary in that some correct biological sequences are clumped (Porter & Hajibabaei, 2018). This is because sequence variation within and among species are likely to vary across taxa. A better strategy is to incorporate single-nucleotide differences into OTU curation that probably retains all amplicon sequence variants (ASVs; Callahan *et al.*, 2017). Currently, the DADA2 pipeline (Callahan *et al.*, 2016) and UNOISE algorithm (Edgar, 2016) can produce such higher-resolution analogues of the traditional OTUs. The most widely used DNA sequence database, the Barcode of Life Data (BOLD) system, archives specimen classification, collection metadata, and images (Ratnasingham & Hebert, 2007). The GenBank database is also often used. However, it is worth noting that DNA metabarcoding can characterise biodiversity that is not represented in current reference sequence databases (Dopheide *et al.*, 2019).

Concerns about the limited taxonomic and geographical coverage of sequence databases for DNA metabarcoding studies are regularly raised. For example, Dopheide *et al.* (2019) found no representative sequence in the GenBank database for more than 900 invertebrate OTUs in one study. This is in line with our own experience in Tasmania, where most insect species are not represented in reference sequence databases. Vouchered museum specimens are useful resources for building local databases of reference sequences and this should be considered concurrently with DNA metabarcoding study design.

Extracting DNA from old museum specimens with non-destructive methods is often feasible, such as for 70% ethanol preserved spiders (Miller *et al.*, 2013) and pinned insect specimens, depending on how they were initially collected (Gilbert *et al.*, 2007; Thomsen *et al.*, 2009), although performance varies among taxa (Carew *et al.*, 2018). DNA preservation in museum specimens is dependent on taxon and storage conditions, but DNA extraction is generally difficult for specimens > 50 years old (Jinbo *et al.*, 2011). However, very short DNA fragments (e.g. 77–204 bp) have been extracted from 100- to 200-year-old dry insect specimens (Thomsen *et al.*, 2009).

While DNA metabarcoding may be constrained by the availability of reference data, the laboratory methods that facilitate DNA metabarcoding can also expedite construction of DNA sequence reference databases. For example, Liu *et al.* (2017)'s HIFI-Barcode pipeline used individually indexed primers during PCR and high-throughput sequencing such that DNA sequences for multiple specimens could be bioinformatically assigned back to their source. A similar procedure is Wang *et al.* (2018)'s reverse verification workflow, which allows later identification with source specimens for OTUs without taxonomic assignment. The first workflow is advantageous for previously classified specimens, while the latter is suitable for unclassified specimens. However, the efficiency of the reverse verification workflow is outstanding when relatively few taxa in the samples are absent from current reference sequence databases. While in-house reference sequence databases are common, recent initiatives aim to increase the taxonomic coverage of universal databases (Hobern & Hebert, 2019; Weigand *et al.*, 2019). However, even when in-house reference sequence databases are lacking, OTUs can be assigned to higher taxonomic levels such as order or family, if not to genus or species.

found limited quantitative ability of DNA metabarcoding across 22 studies (Lamb *et al.*, 2018).

Our understanding of the factors affecting the performance of quantitative DNA metabarcoding is still limited, but some issues have been addressed. Piñol *et al.* (2018) found that primers with fewer template–primer mismatches were better for quantitative DNA metabarcoding, especially for species of higher relative abundance in a sample. Similarly, there have been attempts to create calibrations to correct biases (e.g. primer and polymerase biases). Krehenwinkel *et al.* (2017) demonstrated significant correlation ($R^2 = 0.82$) between the proportion of input DNA and recovered HTS reads with taxon-specific correction factors derived from 10 mock arthropod communities. The correction factors can be established by fitting a regression line with mock communities to calibrate the relationship between DNA reads and abundance or biomass, and these calibrations are potentially similar between closely related taxa, enabling broader application (Krehenwinkel *et al.*, 2017). A strategy using PCR-free sequencing of mitochondrial DNA demonstrated high accuracy ($R^2 = 0.95$) of abundance estimation from arthropod samples (Ji *et al.*, 2019). In that study, many fixed-amount internal-standard DNA mixtures were sequenced along with test samples so that species occurrences and abundances could be calibrated with the sequencing reads derived from internal-standard DNA. For DNA metabarcoding studies, frequency of occurrence among replicate samples is also commonly used as a proxy for species abundance (Aizpurua *et al.*, 2018; Deagle *et al.*, 2018b).

### *Embracing state-of-the-art high-throughput DNA sequencing*

Currently, the Illumina MiSeq (San Diego, California) is the most popular HTS platform for DNA metabarcoding studies, mainly because MiSeq provides reasonable sequencing depth and low sequencing error rates with affordable cost. To date, no comprehensive benchmarking of sequencing platforms has been established for DNA metabarcoding studies. However, Braukmann *et al.* (2019) demonstrated similar performance of three platforms (MiSeq, Ion Torrent PGM, and Ion Torrent S5) for species recovery, although MiSeq is generally recommended due to its lower error rate and well-established bioinformatic procedures.

There is no consensus on the sequencing depth needed to recover all taxa in a given sample. A sequencing depth of $60\,000 \pm 55\,000$ reads per amplicon per sample is reported by many studies (Singer *et al.*, 2019). Increasing sequencing depth can increase the detection rate of low-abundance taxa to some extent (Braukmann *et al.*, 2019), but it is still subject to sequencing instrument limitations. Due to improvements in sequencing technology, the Illumina NovaSeq consistently detected more metazoan taxa from seawater samples than did the MiSeq with an equivalent sequencing depth (Singer *et al.*, 2019), but it is more expensive. We also note here that MiSeq sequencing and bioinformatics can be outsourced to commercial companies or dedicated laboratories at reasonable (and decreasing) cost (Brandon-Mong *et al.*, 2015; Hernandez-Triana *et al.*, 2017; Kerley *et al.*, 2018). Overall, we recommend that researchers

**Table 1.** Some suggested graphical user interface bioinformatics programs for processing high-throughput sequencing (HTS) data in DNA metabarcoding studies.

| Platforms/software | Description | Features |
|---|---|---|
| MBRAVE (Multiplex Barcode Research And Visualisation Environment) (http://mbrave.net/) | Cloud-based platform supporting the storage, validation, analysis, and publication of highly multiplexed projects based on HTS instruments | Allows end-users to follow a default analytical parameter settings or configure a specific pipeline. It now has a module for merging paired-end reads but only for Illumina sequencers. It will incorporate sequencing reads from other platforms (e.g. Nanopore sequencers) in the future. |
| GALAXY (https://usegalaxy.org/) (Afgan *et al.*, 2018) | Open source, web-based platform for data intensive biomedical research | Interface contains modular tools that can be easily deployed into an individual workflow by users without programming experience. |
| QIIME2STUDIO (Q2STUDIO) (https://docs.qiime2.org/2019.4/interfaces/q2studio/) (Bolyen *et al.*, 2018) | Functional prototype of a graphical user interface version for QIIME2 | Bioinformatics tools are adapted to plugins in Q2STUDIO. Users can configure available parameters for each plugin. |
| SLIM (https://trtcrd.github.io/SLIM/) (Dufresne *et al.*, 2019) | Web-based app that simplifies the creation and deployment of a processing pipeline | Provides an integrative toolkit that allow step-by-step data processing. |
| FROGS (Find, Rapidly, Otus with Galaxy Solution) (http://frogs.toulouse.inra.fr/) (Escudie *et al.*, 2018) | Galaxy-supported pipeline that facilitates analysis of large DNA amplicon sequencing datasets | Supports HTS paired sequence merging, cleaning, dereplication, and downstream statistical result visualisation. |

start with the MiSeq platform as it is suitable for most DNA metabarcoding studies with respect to sequencing accuracy, output and cost.

Advances in HTS technologies are creating new opportunities for DNA metabarcoding studies. Although long-read sequencing (e.g. MinION; Oxford Nanopore Technologies, Oxford, U.K.) is not used for eDNA studies targeting short DNA fragments, it has great scope for scaling up DNA metabarcoding to phylogenetic inferences for fresh bulk samples. Sequencing long amplicons (e.g. 4000 bp) provides an opportunity to accurately quantify the phylogenetic diversity of a sample, rather than just species diversity (Krehenwinkel *et al.*, 2019). We envision the future possibility for on-site monitoring with miniature and affordable equipment [e.g. Bento portable laboratory for DNA extraction and library preparation (Gilbert, 2017); and miniPCR device (Krehenwinkel *et al.*, 2019; Pomerantz *et al.*, 2018)] available for routine DNA metabarcoding.

### *Graphical user interface bioinformatics programs*

Regardless of sequencing platform, bioinformatic analyses can be daunting. For DNA metabarcoding, bioinformatics typically involves a pipeline that converts HTS data into an OTU table for downstream analysis comparing diversity and community composition through space or time. Detailed methodological procedures and bioinformatics processes have been discussed in previous reviews (Bik *et al.*, 2012; Coissac *et al.*, 2012; Deiner *et al.*, 2017, section 4.3, boxes 2 and 3; also see Piper *et al.*, 2019, fig. 2). We recommend beginners consult bioinformatic experts whenever possible to avoid mistakes throughout their data analyses. Although there are several advantages of advanced command-line-based methods to bioinformatics, this can be daunting for entomologists new to such approaches. Therefore, we also consider user-friendly GUI programs (Table 1) for ecological entomologists who might not be confident with command-line programs. Alternatively, the bioinformatics can be outsourced to a commercial provider.

A generic bioinformatics workflow for DNA metabarcoding consists of five core steps (i.e. demultiplexing samples, merging pair-end reads, quality filtering, OTU curation, and taxonomic assignment), which can require several separate bioinformatics programs. Flexible and powerful GUI programs break down the hurdle of bioinformatics expertise and allows users to quickly design their own metabarcoding pipeline. Typically, GUI programs simplify the deployment of a processing pipeline into modular tools where users can customise parameters to fit their own data. Online tutorials and example datasets are usually provided along with GUI programs to enable non-specialists to gain hands-on experience and develop bioinformatics skills for their own analyses. While beginners may be intimidated by the complexity of bioinformatic analyses, we recommend that non-specialists start with these resources to analyse their data.

## Future research and challenges

The use of DNA metabarcoding as a sensitive and efficient method for species detection has been enthusiastically embraced by molecular ecologists, and has developed to a stage where non-molecular ecologists can, and should, take advantage of its benefits. DNA metabarcoding provides unparalleled opportunities for understanding complex ecological networks and their response to environmental change (Houadria *et al.*, 2018; Thierry *et al.*, 2019). Although DNA metabarcoding may not successfully identify all species in a sample, and the ecological inferences may not perfectly match those based on morphological specimen identification, information recovered by

DNA metabarcoding can provide equivalent ecological inferences (Ji *et al.*, 2013; Barsoum *et al.*, 2019). However, there are still challenges to the application of DNA metabarcoding. The knowledge of DNA taphonomy across ecosystems is so limited that sampling eDNA to investigate local invertebrate communities remains in the early stages of development. Further research is needed to understand the potential of sedimentary DNA metabarcoding for analysing ancient samples. Epp *et al.* (2012) observed low amplification of beetles from sediment, while Thomsen *et al.* (2009) had 20% detection success from 14 ancient beetle chitin remains. It is noted that DNA metabarcoding is also advantageous for its compatibility of non-invasive protocols (i.e. without destroying target specimens). Diverse arthropod species ($n > 135$) were uncovered from the environmental DNA of sampled flowers (Thomsen & Sigsgaard, 2019), while metabarcoding of extracted DNA from sample fixative has proved successful (e.g. denatured 96% ethanol; Zizka *et al.*, 2019). As DNA metabarcoding approaches are rapidly evolving, new applications for (e)DNA metabarcoding will probably be developed with further research and advances in technology.

DNA metabarcoding has great potential for monitoring insect pests in agriculture (Evans *et al.*, 2017; Morales-Hojas, 2017) and will be greatly aided by expanding reference sequence databases for conclusive detection of pest species (Piper *et al.*, 2019). In addition, DNA metabarcoding has been tested as an effective alternative for vector and arbovirus surveillance, which is an issue that concerns public health (Schneider *et al.*, 2016). For example, early detection of mosquito vectors using water sample eDNA presented a comparable or higher detection probability to traditional surveys (Schneider *et al.*, 2016; Boerlijst *et al.*, 2019). There have also been some attempts at using DNA metabarcoding to conduct population genetics studies (Elbrecht *et al.*, 2018; Marshall & Stepien, 2019). However, metabarcoding-based genotyping shows at least as many experimental biases as DNA metabarcoding for species identification, and is prone to the recovery of erroneous haplotypes (Elbrecht *et al.*, 2018). On the technical front, a future improvement would be to perform independent replicates (Ficetola *et al.*, 2014; Dopheide *et al.*, 2018), increase sequencing depth for higher detection rate (Braukmann *et al.*, 2019; Singer *et al.*, 2019), develop PCR-free strategies (e.g. genome-skimming; Bista *et al.*, 2018), and use single-cell sequencing for approximating unique individuals (Adams *et al.*, 2019). For instance, debris cells in environmental samples could be collected by flow cytometry and microfluidic technology, and followed by DNA extraction and sequencing for screening individual-level genetic variation (Adams *et al.*, 2019).

While most ecological entomologists are still confident and comfortable with conventional biodiversity surveys, DNA metabarcoding is already a practical tool to assist environmental management. DNA metabarcoding can rapidly characterise the species present in a sample, and, combined with the capacity to analyse hundreds of samples on a single HTS run, can increase the number of samples analysed whilst reducing the time needed to do so, and the associated cost. We are now in an era of biodiversity crisis, and using DNA metabarcoding to rapidly analyse more samples will help to inform better management. Currently, the main limitations of DNA metabarcoding for invertebrate identification include challenges around abundance estimation, incompleteness of barcode reference sequences and unsuitability of some pinned reference collections for producing these references, and destructive sampling methods. However, the rapid rate of progress and declining costs mean that DNA metabarcoding is likely to become a standard tool in the future (Deiner *et al.*, 2017). An example of an emerging initiative in aquatic ecosystems is DNAqua-Net, which is a permanent Working Group as part of the European Standards Committee (Leese *et al.*, 2018). This group of researchers, managers, politicians and other stakeholders jointly develops routine application for biodiversity assessment and provides standard protocols for all steps from sampling until analysis using genomic tools such as DNA metabarcoding.

## Acknowledgements

This work was funded by the Australian Research Council Linkage Grant LP140100075. ML was also supported by the China Scholarship Council (CSC no. 201608360109). The authors have no conflicts of interest to declare.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Doc. S1.** Input data for creating the Sankey diagram.

**Doc. S2.** DNA sequence alignment of the primer-binding sites of the primer ZBJ-ArtF1c for 515 beetle species.

**Doc. S3.** DNA sequence alignment of the primer-binding sites of the primer ZBJ-ArtR2c for 515 beetle species.

**Appendix S1.** Surveyed empirical DNA studies with respect to the ecosystem sampled, and methods employed for sample preservation, storage, and minimisation of contamination.

## References

Adams, C.I.M., Knapp, M., Gemmell, N.J., Jeunen, G.J., Bunce, M., Lamare, M.D. *et al.* (2019) Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool? *Genes*, **10**, 192.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M. *et al.* (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, **46**, W537–W544.

Aizpurua, O., Budinski, I., Georgiakakis, P., Gopalakrishnan, S., Ibañez, C., Mata, V. *et al.* (2018) Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: evidence from DNA metabarcoding. *Molecular Ecology*, **27**, 815–825.

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. & Bohmann, K. (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, **9**, 134–147.

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

Aylagas, E., Mendibil, I., Borja, Á. & Rodríguez-Ezpeleta, N. (2016) Marine sediment sample pre-processing for macroinvertebrates

metabarcoding: mechanical enrichment and homogenization. *Frontiers in Marine Science*, **3**, 203.

Balint, M., Pfenninger, M., Grossart, H.P., Taberlet, P., Vellend, M., Leibold, M.A. *et al.* (2018) Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, **33**, 945–957.

Barsoum, N., Bruce, C., Forster, J., Ji, Y. & Yu, D.W. (2019) The devil is in the detail: metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity. *Ecological Indicators*, **101**, 313–323.

Beentjes, K.K., Speksnijder, A.G.C.L., Schilthuizen, M., Hoogeveen, M. & van der Hoorn, B.B. (2019) The effects of spatial and temporal replicate sampling on eDNA metabarcoding. *PeerJ*, **7**, e7335.

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R. & Thomas, W.K. (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, **27**, 233–243.

Bista, I., Carvalho, G.R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M. *et al.* (2018) Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, **18**, 1020–1034.

Boerlijst, S.P., Trimbos, K.B., Van der Beek, J.G., Dijkstra, K.D.B., Van der Hoorn, B.B. & Schrama, M. (2019) Field evaluation of DNA based biodiversity monitoring of Caribbean mosquitoes. *Frontiers in Ecology and Evolution*, **7**, 240.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Al-Ghalith, G.A. *et al.* (2018) QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Preprints*, **6**, e27295v2.

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

Brannock, P.M., Sharma, J., Bik, H.M., Thomas, W.K. & Halanych, K.M. (2017) Spatial and temporal variation of intertidal nematodes in the northern gulf of Mexico after the deepwater horizon oil spill. *Marine Environmental Research*, **130**, 200–212.

Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Elbrecht, V., Steinke, D., Ratnasingham, S. *et al.* (2019) Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, **19**, 711–727.

Cahill, A.E., Pearman, J.K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R. *et al.* (2018) A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, **8**, 8908–8920.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**, 581–583.

Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **11**, 2639–2643.

Carew, M.E., Coleman, R.A. & Hoffman, A.A. (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*, **6**, e4980.

Chariton, A.A., Stephenson, S., Morgan, M.J., Steven, A.D.L., Colloff, M.J., Court, L.N. *et al.* (2015) Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environmental Pollution*, **203**, 165–174.

Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160–1170.

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

Coissac, E., Riaz, T. & Puillandre, N. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, **21**, 1834–1847.

Creedy, T.J., Ng, W.S. & Vogler, A.P. (2019) Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, **9**, 3105–3116.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K. *et al.* (2016) The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, **7**, 1008–1018.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, 20140562.

Deagle, B.E., Clarke, L.J., Kitchener, J.A., Polanowski, A.M. & Davidson, A.T. (2018a) Genetic monitoring of open ocean biodiversity: an evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources*, **18**, 391–406.

Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L. *et al.* (2018b) Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Molecular Ecology*, **28**, 391–406.

Deiner, K., Walser, J.-C., Mächler, E. & Altermatt, F. (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, **183**, 53–63.

Deiner, K., Bik, H.M., Machler, E., Seymour, M., Lacoursiere-Roussel, A., Altermatt, F. *et al.* (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*, **26**, 5872–5895.

Dickie, I.A., Boyer, S., Buckley, H.L., Duncan, R.P., Gardner, P.P., Hogg, I.D. *et al.* (2018) Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources*, **18**, 940–952.

Dopheide, A., Xie, D., Buckley, T.R., Drummond, A.J. & Newcomb, R.D. (2018) Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and Evolution*, **10**, 120–133.

Dopheide, A., Tooman, L.K., Grosser, S., Agabiti, B., Rhode, B., Xie, D. *et al.* (2019) Estimating the biodiversity of terrestrial invertebrates on a forested Island using DNA barcodes and metabarcoding data. *Ecological Appllications*, **29**, e01877.

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L.A., Pawlowski, J. & Cordier, T. (2019) SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinformatics*, **20**, 88.

Edgar, R.C. (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 081257.

Elbrecht, V. & Leese, F. (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, **5**, 11.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.N. *et al.* (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, **4**, e1966.

Elbrecht, V., Leese, F. & Bunce, M. (2017a) PrimerMiner: an r package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, **8**, 622–626.

Elbrecht, V., Peinert, B. & Leese, F. (2017b) Sorting things out: assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, **7**, 6918–6926.

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

Elbrecht, V., Braukmann, T.W.A., Ivanova, N.V., Prosser, S.W.J., Hajibabaei, M., Wright, M. *et al.* (2019) Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, **7**, e7745.

Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T. *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology*, **21**, 1821–1833.

Erdozain, M., Thompson, D.G., Porter, T.M., Kidd, K.A., Kreutzweiser, D.P., Sibley, P.K. *et al.* (2019) Metabarcoding of storage ethanol vs. conventional morphometric identification in relation to the use of stream macroinvertebrates as ecological indicators in forest management. *Ecological Indicators*, **101**, 173–184.

Escudie, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K. *et al.* (2018) FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics*, **34**, 1287–1294.

Evans, D.M., Gilbert, J.D.J. & Port, G.R. (2017) Everything is connected: network thinking in entomology. *Ecological Entomology*, **42**, 1–3.

Fernandez, S., Rodriguez, S., Martinez, J.L., Borrell, Y.J., Ardura, A. & Garcia-Vazquez, E. (2018) Evaluating freshwater macroinvertebrates from eDNA metabarcoding: a river Nalon case study. *PLoS One*, **13**, e0201741.

Ferro, M.L. & Park, J.-S. (2013) Effect of propylene glycol concentration on mid-term DNA preservation of Coleoptera. *The Coleopterists Bulletin*, **67**, 581–586.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M. *et al.* (2014) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, **15**, 543–556.

Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P. *et al.* (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, **1**, 98.

Gilbert, M.T., Moore, W., Melchior, L. & Worobey, M. (2007) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One*, **2**, e272.

Gilbert, M.T.P. (2017) Documenting DNA in the dust. *Molecular Ecology*, **26**, 969–971.

Grey, E.K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K.L. *et al.* (2018) Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, **8**, 8843.

Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D. *et al.* (2019) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, **826**, 25–41.

Hernandez-Triana, L.M., Brugman, V.A., Prosser, S.W.J., Weland, C., Nikolova, N., Thorne, L. *et al.* (2017) Molecular approaches for blood meal analysis and species identification of mosquitoes (Insecta: Diptera: Culicidae) in rural locations in southern England, United Kingdom. *Zootaxa*, **4250**, 67–76.

Hobern, D. & Hebert, P. (2019) BIOSCAN-revealing eukaryote diversity, dynamics, and interactions. *Biodiversity Information Science and Standards*, **3**, e37333.

Houadria, M.Y.I., Klimes, P., Fayle, T.M. & Gullan, P.J. (2018) Host-plant dissections reveal contrasting distributions of *Crematogaster* ants and their symbionts in two myrmecophytic *Macaranga* species. *Ecological Entomology*, **43**, 601–611.

Jeunen, G.-J., Knapp, M., Spencer, H.G., Lamare, M.D., Taylor, H.R., Stat, M. *et al.* (2018) Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, **19**, 426–438.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A. *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, **16**, 1245–1257.

Ji, Y., Huotari, T., Roslin, T., Schmidt, N.M., Wang, J., Yu, D.W. *et al.* (2019) SPIKEPIPE: a metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Molecular Ecology Resources*, **00**, 1–12.

Jinbo, U., Kato, T. & Ito, M. (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, **14**, 107–124.

Kemp, B.M. & Smith, D.G. (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Science International*, **154**, 53–61.

Kerley, G.I.H., Landman, M., Ficetola, G.F., Boyer, F., Bonin, A., Rioux, D. *et al.* (2018) Diet shifts by adult flightless dung beetles *Circellium bacchus*, revealed using DNA metabarcoding, reflect complex life histories. *Oecologia*, **188**, 107–115.

Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B. & Gillespie, R.G. (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, **7**, 17668.

Krehenwinkel, H., Pomerantz, A., Henderson, J.B., Kennedy, S.R., Lim, J.Y., Swamy, V. *et al.* (2019) Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience*, **8**, giz006.

Lacoursiere-Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K. *et al.* (2018) eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, **8**, 7763–7777.

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G. & Taylor, M.I. (2018) How quantitative is metabarcoding: a meta-analytical approach. *Molecular Ecology*, **28**, 420–430.

Leese, F., Bouchez, A., Abarenkov, K., Altermatt, F., Borja, Á., Bruce, K. *et al.* (2018) Why we need sustainable networks bridging countries, disciplines, cultures and generations for aquatic biomonitoring 2.0: a perspective derived from the DNAqua-Net COST action. In *Advances in Ecological Research*, Vol **58**, pp. 63–99. Academic Press, Oxford, U.K.

Liu, S., Yang, C., Zhou, C. & Zhou, X. (2017) Filling reference gaps via assembling DNA barcodes using high-throughput sequencing–moving toward barcoding the world. *GigaScience*, **6**, gix104.

Lobo, J., Shokralla, S., Costa, M.H., Hajibabaei, M. & Costa, F.O. (2017) DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports*, **7**, 15618.

Longmire, J.L., Maltbie, M. & Baker, R.J. (1997) *Use of "lysis buffer" in DNA Isolation and its Implication for Museum Collections*. *Museum of Texas Tech University*, **163**, 1–3.

Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. & Leese, F. (2018) Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate COI primers. *Molecular Ecology Resources*, **18**, 1456–1468.

Majaneva, M., Diserud, O.H., Eagle, S.H.C., Hajibabaei, M. & Ekrem, T. (2018) Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding and Metagenomics*, **2**, e26664.

Marquina, D., Andersson, A.F. & Ronquist, F. (2018) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Molecular Ecology Resources*, **19**, 90–104.

Marshall, N.T. & Stepien, C.A. (2019) Invasion genetics from eDNA and thousands of larvae: a targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. *Ecology and Evolution*, **9**, 3515–3538.

Miller, J.A., Beentjes, K.K., van Helsdingen, P. & Ijland, S. (2013) Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. *Zookeys*, **365**, 245–261.

Minich, J.J., Sanders, J.G., Amir, A., Humphrey, G., Gilbert, J.A. & Knight, R. (2019) Quantifying and understanding well-to-well contamination in microbiome research. *mSystems*, **4**, e00186–19.

Morales-Hojas, R. (2017) Molecular ecology of insect pests of agricultural importance: the case of aphids. *Ecological Entomology*, **42**, 18–27.

Moreau, C.S., Wray, B.D., Czekanski-Moir, J.E. & Rubin, B.E.R. (2013) DNA preservation: a test of commonly used preservatives for insects. *Invertebrate Systematics*, **27**, 81.

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M. *et al.* (2018) Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, **18**, 927–939.

Pan, W., Byrne-Steele, M., Wang, C., Lu, S., Clemmons, S., Zahorchak, R.J. *et al.* (2014) DNA polymerase preference determines PCR priming efficiency. *BMC Biotechnology*, **14**, 10.

Patrick, H.J., Chomic, A. & Armstrong, K.F. (2016) Cooled propylene glycol as a pragmatic choice for preservation of DNA from remote field-collected diptera for next-generation sequence analysis. *Journal of Economic Entomology*, **109**, 1469–1473.

Piñol, J., Senar, M.A. & Symondson, W.O.C. (2018) The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, **28**, 407–419.

Piper, A.M., Batovska, J., Cogan, N.O.I., Weiss, J., Cunningham, J.P., Rodoni, B.C. *et al.* (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*, **8**, giz092.

Pomerantz, A., Penafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A. *et al.* (2018) Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, **7**, giy033.

Porter, T.M. & Hajibabaei, M. (2018) Scaling up: a guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, **27**, 313–338.

Prince, A.M. & Andrus, L. (1992) PCR: how to kill unwanted DNA. *BioTechniques*, **12**, 358–360.

Ratnasingham, S. & Hebert, P.D. (2007) BOLD: the barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Rennstam Rubbmark, O., Sint, D., Horngacher, N. & Traugott, M. (2018) A broadly applicable COI primer pair and an efficient single-tube amplicon library preparation protocol for metabarcoding. *Ecology and Evolution*, **8**, 12335–12350.

Ritter, C.D., Zizka, A., Barnes, C., Nilsson, R.H., Roger, F. & Antonelli, A. (2018) Locality or habitat? Exploring predictors of biodiversity in Amazonia. *Ecography*, **42**, 321–333.

Salonen, I.S., Chronopoulou, P.M., Leskinen, E. & Koho, K.A. (2019) Metabarcoding successfully tracks temporal changes in eukaryotic communities in coastal sediments. *FEMS Microbiology Ecology*, **95**, fiy22.

Schmidt, N.M., Mosbacher, J.B., Eitzinger, B., Vesterinen, E.J. & Roslin, T. (2018) High resistance towards herbivore-induced habitat change in a high Arctic arthropod community. *Biology Letters*, **14**, 20180054.

Schneider, J., Valentini, A., Dejean, T., Montarsi, F., Taberlet, P., Glaizot, O. *et al.* (2016) Detection of invasive mosquito vectors using environmental DNA (eDNA) from water samples. *PLoS One*, **11**, e0162493.

Schnell, I.B., Bohmann, K. & Gilbert, M.T.P. (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, **15**, 1289–1303.

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

Serrana, J.M., Yaegashi, S., Kondoh, S., Li, B., Robinson, C.T. & Watanabe, K. (2018) Ecological influence of sediment bypass tunnels on macroinvertebrates in dam-fragmented rivers by DNA metabarcoding. *Scientific Reports*, **8**, 10185.

Singer, G.A.C., Fahner, N.A., Barnes, J.G., McCarthy, A. & Hajibabaei, M. (2019) Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. *Scientific Reports*, **9**, 5991.

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A.J., Yamamoto, H., Chevaldonné, P. *et al.* (2016) Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. *Frontiers in Marine Science*, **3**, 92.

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, **105**, 13486–13491.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T. *et al.* (2017) Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, **8**, 635–645.

Sze, M.A. & Schloss, P.D. (2019) The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere*, **4**, e00163–19.

Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L.H. (2012) Environmental DNA. *Molecular Ecology*, **21**, 1789–1793.

Thierry, M., Hrček, J. & Lewis, O.T. (2019) Mechanisms structuring host–parasitoid networks in a global warming context: a review. *Ecological Entomology*, **44**, 581–592.

Thomsen, P.F. & Sigsgaard, E.E. (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution*, **9**, 1665–1679.

Thomsen, P.F. & Willerslev, E. (2015) Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, **183**, 4–18.

Thomsen, P.F., Elias, S., Gilbert, M.T.P., Haile, J., Munch, K., Kuzmina, S. *et al.* (2009) Non-destructive sampling of ancient insect DNA. *PLoS One*, **4**, e5048.

Tsuji, S., Takahara, T., Doi, H., Shibata, N. & Yamanaka, H. (2019) The detection of aquatic macroorganisms using environmental DNA analysis – a review of methods for collection, extraction, and detection. *Environmental DNA*, **1**, 99–108.

Vink, C.J., Thomas, S.M., Paquin, P., Hayashi, C.Y. & Hedin, M. (2005) The effects of preservatives and temperatures on arachnid DNA. *Invertebrate Systematics*, **19**, 99–104.

Wang, W.Y., Srivathsan, A., Foo, M., Yamane, S.K. & Meier, R. (2018) Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: validating a reverse workflow for specimen processing. *Molecular Ecology Resources*, **18**, 490–501.

Weigand, H., Beermann, A.J., Ciampor, F., Costa, F.O., Csabai, Z., Duarte, S. *et al.* (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: gap-analysis and recommendations for future work. *Science of the Total Environment*, **678**, 499–524.

Wood, J.R., Holdaway, R.J., Orwin, K.H., Morse, C., Bonner, K.I., Davis, C. *et al.* (2017) No single driver of biodiversity: divergent

responses of multiple taxa across land use types. *Ecosphere*, **8**, e01997.

Xie, Y., Zhang, X., Yang, J., Kim, S., Hong, S., Giesy, J.P. *et al.* (2018) eDNA-based bioassessment of coastal sediments impacted by an oil spill. *Environmental Pollution*, **238**, 739–748.

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

Zepeda-Mendoza, M.L., Bohmann, K., Carmona Baez, A. & Gilbert, M.T. (2016) DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. *BMC Research Notes*, **9**, 255.

Zhang, G.K., Chain, F.J.J., Abbott, C.L. & Cristescu, M.E. (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary Applications*, **11**, 1901–1914.

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M. *et al.* (2018) Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, **28**, 528–543.

Zinger, L., Bonin, A., Alsos, I.G., Balint, M., Bik, H., Boyer, F. *et al.* (2019) DNA metabarcoding – need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, **28**, 1857–1862.

Zizka, V.M.A., Leese, F., Peinert, B. & Geiger, M.F. (2019) DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome*, **62**, 122–136.

Accepted 3 December 2019
First published online 30 December 2019
Associate Editor: Alex Córdoba-Aguilar