INVITED TECHNICAL REVIEW

Strategies for sample labelling and library preparation in DNA metabarcoding studies

Kristine Bohmann[1] | Vasco Elbrecht[2] | Christian Carøe[1] | Iliana Bista[3,4] | Florian Leese[5] | Michael Bunce[6] | Douglas W. Yu[7,8,9] | Mathew Seymour[10] | Alex J. Dumbrell[11] | Simon Creer[12]

[1]Faculty of Health and Medical Sciences, Section for Evolutionary Genomics, Globe Institute, University of Copenhagen, Copenhagen, Denmark

[2]Department of Environmental Systems Science, ETH Zurich, Zürich, Switzerland

[3]Department of Genetics, University of Cambridge, Cambridge, UK

[4]Tree of Life, Wellcome Sanger Institute, Hinxton, UK

[5]Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Essen, Germany

[6]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, WA, Australia

[7]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[8]School of Biological Sciences, Norwich Research Park, University of East Anglia, Norwich, UK

[9]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming Yunnan, China

[10]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[11]School of Life Sciences, University of Essex, Colchester, UK

[12]Molecular Ecology and Evolution Group, School of Natural Sciences, Bangor University, Gwynedd, UK

Correspondence
Kristine Bohmann, Section for Evolutionary Genomics, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
Email: kbohmann@sund.ku.dk

Funding information
NERC Biomolecular Analysis Facility, Grant/Award Number: NE/M02086X/1, NE/N003756/1, NE/N006216/1, NE/S000291/1, NE/S005560/1 and NE/S006958/1; Carlsbergfondet, Grant/Award Number: CF18-1110; H2020 European Research Council, Grant/Award Number: 856488; Det Frie Forskningsråd, Grant/Award Number: 5051-00140; Wellcome, Grant/Award Number: WT207492, 104640/Z/14/Z and 092096/Z/10/Z

Abstract

Metabarcoding of DNA extracted from environmental or bulk specimen samples is increasingly used to profile biota in basic and applied biodiversity research because of its targeted nature that allows sequencing of genetic markers from many samples in parallel. To achieve this, PCR amplification is carried out with primers designed to target a taxonomically informative marker within a taxonomic group, and sample-specific nucleotide identifiers are added to the amplicons prior to sequencing. The latter enables assignment of the sequences back to the samples they originated from. Nucleotide identifiers can be added during the metabarcoding PCR and during "library preparation", that is, when amplicons are prepared for sequencing. Different strategies to achieve this labelling exist. All have advantages, challenges and limitations, some of which can lead to misleading results, and in the worst case compromise the fidelity of the metabarcoding data. Given the range of questions addressed using metabarcoding, ensuring that data generation is robust and fit for the chosen purpose is critically important for practitioners seeking to employ metabarcoding for biodiversity assessments. Here, we present an overview of the three main workflows for sample-specific labelling and library preparation in metabarcoding studies on Illumina sequencing platforms; one-step PCR, two-step PCR, and tagged PCR. Further, we

distill the key considerations for researchers seeking to select an appropriate meta-barcoding strategy for their specific study. Ultimately, by gaining insights into the consequences of different metabarcoding workflows, we hope to further consolidate the power of metabarcoding as a tool to assess biodiversity across a range of applications.

1 | INTRODUCTION

In recent years, the analysis of environmental DNA (eDNA) and DNA extracted from bulk specimen samples has experienced an enormous surge in popularity in basic and applied biodiversity studies seeking to detect e.g., animal, plant, algae, fungi, and bacteria (Bálint et al., 2016; Compson et al., 2020; Creer et al., 2016; Jarman et al., 2018; Lindahl et al., 2013; Taberlet et al., 2012). Within the field of genetic biodiversity assessment, DNA metabarcoding is currently the most widely used approach, as it allows targeted, parallel, and as such relatively cost-effective, identification of multiple taxa from environmental samples, such as soil, water, and faeces, as well as from bulk samples of organisms (Taberlet, Coissac, Pompanon, et al., 2012). Here, the applications of metabarcoding range widely; for example, detection of invasive species (e.g., Pochon et al., 2013); assessment of water quality via identification of freshwater invertebrates in bulk specimen samples (e.g., Elbrecht et al., 2017) and environmental samples (e.g., Seymour et al., 2020); identification of plant-pollinator interactions (e.g, Gous et al., 2019; Lucas et al., 2018); detection of vertebrate wildlife via invertebrate "samplers" of vertebrate blood or faeces (e.g., Calvignac-Spencer et al., 2013), and assessment of for example, niche partitioning (e.g., Razgour et al., 2011) and ecosystem services (e.g., Aizpurua et al., 2017) through detection of diet items. Furthermore, metabarcoding is explored for implementation in routine biomonitoring around the world (Aylagas et al., 2018; Li et al., 2018, 2019; Pont et al., 2018, 2021; Zizka et al., 2020; www.danubesurvey.org; www.syke.fi), and is an integral component of the proposals for the Next Generation of Biomonitoring programmes (Bohan et al., 2017).

Metabarcoding relies on PCR amplification of extracted DNA with primers designed to target a taxonomically informative marker for a selected taxonomic group (Taberlet, Coissac, Pompanon, et al., 2012) (Figure 1). The backbone of metabarcoding analyses is the addition of sample-specific nucleotide identifiers to amplicons and the use of these to assign metabarcoding sequences back to the samples they originated from ("demultiplexing"). This allows pooling of hundreds to thousands of samples for sequencing and utilisation of the capacity of high-throughput sequencing platforms (Figure 1). Amplicon labelling can be achieved at two stages during a metabarcoding workflow: prior to library build, as 5′ nucleotide "tags" on metabarcoding primers, and during library build as library indices. The strategies to achieve this labelling can be categorised into three

main approaches; one-step PCR, two-step PCR, and tagged PCR (Figure 2). All three approaches have advantages, challenges, and limitations which, if not considered, can result in misleading data interpretation, and in the very worst case can lead to unusable data and considerable wasted time and money, as for instance in the case of the so-called "tag-jumps" (Carøe & Bohmann, 2020; Esling et al., 2015; Schnell et al., 2015). Despite this, in contrast to discussions on metabarcoding substrate selection, DNA extraction, and data processing, the strategies for amplicon labelling and library preparation workflows have received little systematic attention in the metabarcoding literature (although see Murray et al., 2015).

Here, we present an overview of the three most commonly used workflows with which to achieve sample-specific labelling and library preparation in metabarcoding studies, and how they can potentially influence the resulting data. For the sake of simplicity, we mainly focus on metabarcoding of plants and animals in basic and applied biodiversity studies with sequencing on arguably the most used high-throughput sequencing platform series today, the Illumina sequencing platforms. Note that points raised will be relevant for metabarcoding of other organisms and to high-throughput sequencing platforms with similar labelling structures to Illumina platforms, such as Ion Torrent (Thermo Fischer Scientific), BGI platforms (BGI Genomics), Oxford Nanopore Technologies MinION, and PacBio (Pacific Biosciences). In the present article, we provide critical considerations for researchers to choose the optimal metabarcoding strategy for generating reliable data tailored to their individual study; for example, regarding sample type and number, research question, speed of laboratory processing, contamination risk, budget, and whether similar studies are to be carried out in the laboratory. Ultimately, by gaining detailed and critical insights into the consequences of choosing different metabarcoding workflows, we hope to further increase the potential of metabarcoding as a reliable tool for use across a wide range of applications.

2 | TAGGING AND INDEXING APPROACHES IN METABARCODING STUDIES

Today, the most commonly used high-throughput sequencing platform for metabarcoding studies is the Illumina series, where for example the MiSeq, iSeq, HiSeq, NextSeq, and NovaSeq have
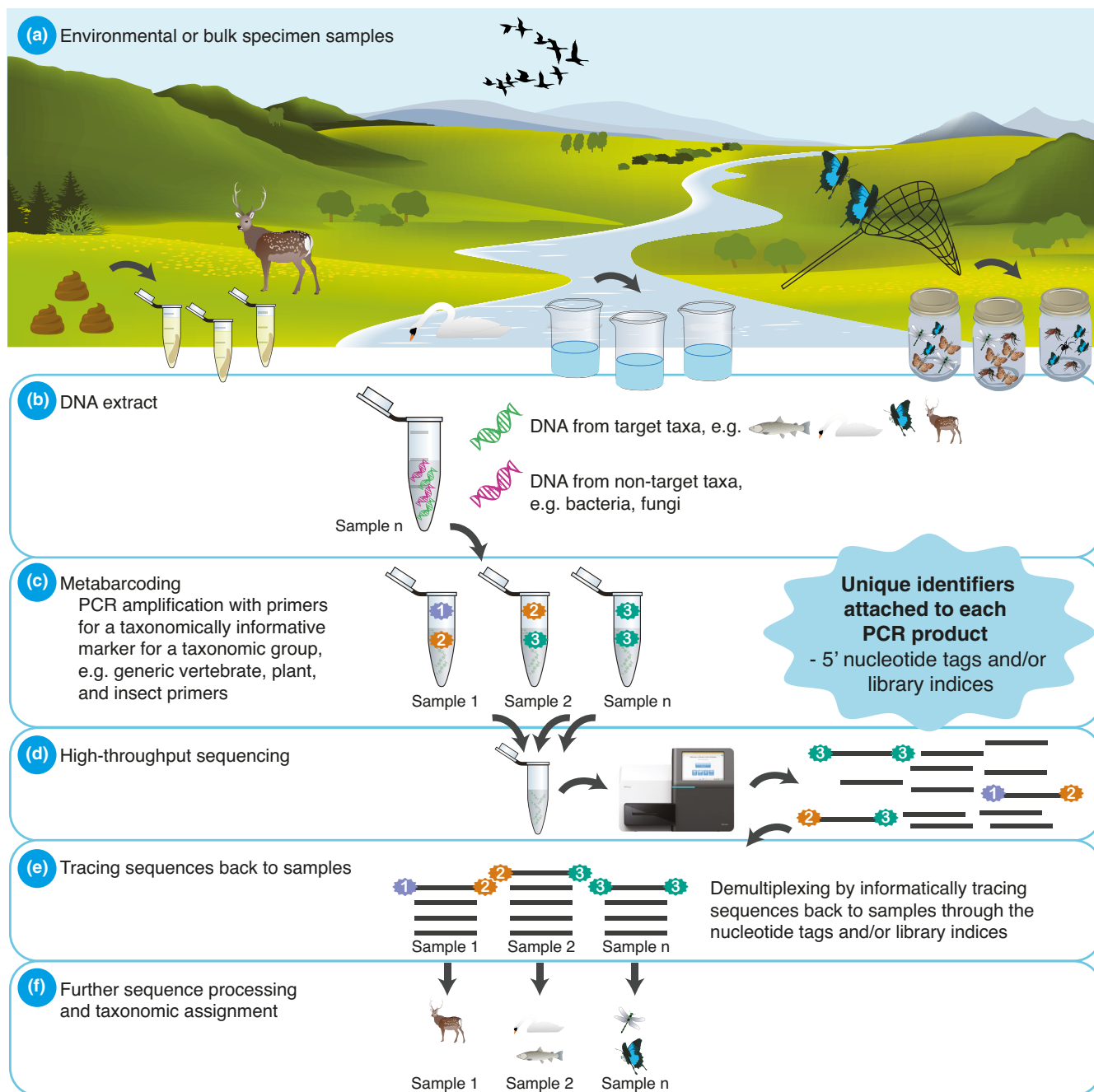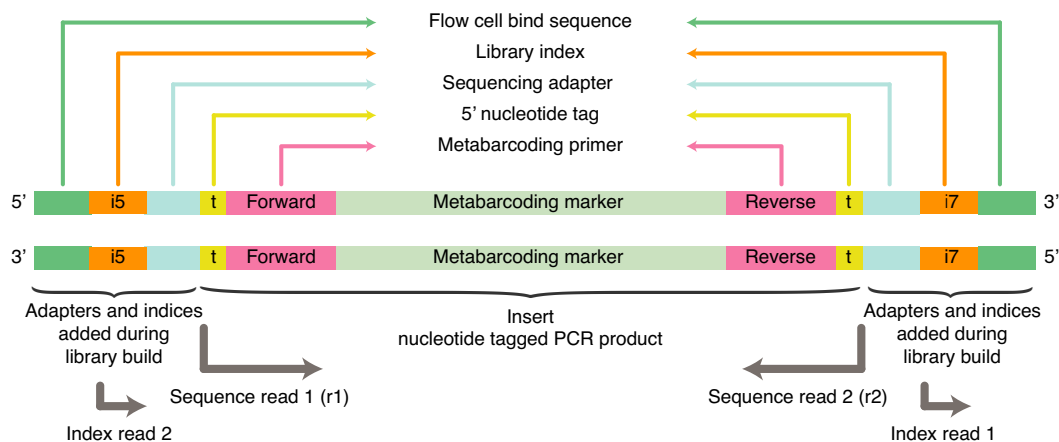
FIGURE 1 Simplified overview of a metabarcoding workflow. (a–b) DNA extracted from environmental samples such as soil, water, and faeces or from bulk specimen samples. The DNA extracts are typically a complex mix of DNA from target and nontarget organisms. (c) DNA extracts are PCR-amplified with metabarcoding primers that target a taxonomically informative marker for a taxonomic group. Importantly, identifiers unique to each PCR product are added in the form of 5′ nucleotide tags on primers and/or as indices added to sequence libraries during library build. (d) The taxonomic markers of hundreds to thousands of samples are sequenced in parallel on a high-throughput sequencing platform producing millions of sequence reads. (e) The sequences can be traced back to the samples they originated from through the nucleotide tags and/or library indices, and (f) can be further analysed. Images courtesy of the Integration and Application Network, University of Maryland Centre for Environmental Science (ian.umces.edu/symbols/) and Illumina.com

been employed (Jarman et al., 2018). These platforms offer high throughput, relatively low error rates, and relatively long paired-end reads, typically up to 150 bp of each paired read on the iSeq100, NextSeq550/1000/2000, HiSeq 3000/4000, and NovaSeq (up to 250 bp on SP flow cell), and 300 bp of each paired read on the MiSeq platform (www.illumina.com, applied in e.g., Elbrecht et al.,

2017; Hope et al., 2014; Quéméré et al., 2013; Shehzad, Riaz, et al., 2012; Singer et al., 2019; Stoeck et al., 2018).

The sequencing depth required per sample is commonly much lower in metabarcoding studies than in shotgun sequencing studies (e.g., Srivathsan et al., 2015; Stat et al., 2017), and in metabarcoding studies it is (economically) feasible to sequence tens, hundreds, or

(a) Structure of dual-tagged and dual-indexed Illumina metabarcoding library sequence

Flow cell bind sequence
Library index
Sequencing adapter
5' nucleotide tag
Metabarcoding primer

5' [i5] [t] [Forward] Metabarcoding marker [Reverse] [t] [i7] 3'

3' [i5] [t] [Forward] Metabarcoding marker [Reverse] [t] [i7] 5'

Adapters and indices added during library build

Insert nucleotide tagged PCR product

Adapters and indices added during library build

Sequence read 1 (r1)

Sequence read 2 (r2)

Index read 2

Index read 1

Strategies for addition of sample-specific identifiers and library build in metabarcoding studies

(b) One-step PCR with fusion primers

DNA extract
Sample 1

t F   R t
Fusion primers

PCR

F   R
Library sequence
Sample 1

Library sample 1   Library sample 2

Pooling

(c) Two-step PCR - untagged

DNA extract
Sample 1

F R
Primers with sequence overhangs

PCR 1

F   R
PCR amplicon with sequence overhangs

Fusion primers with sequence complementary to overhang, adapters and indexes

i5   i7

PCR 2

i5 F   R i7
Library sequence
Sample 1

Library sample 1   Library sample 2

Pooling

(d) Two-step PCR - tagged

DNA extract
Sample 1

t F   R t
Primers with sequence overhangs and tags

PCR 1

t F   R t
PCR amplicon with sequence overhangs and tags

Fusion primers with sequence complementary to overhang, adapters and indexes

i5   i7

PCR 2

i5 t F   R t i7
Library sequence
Sample 1

Library sample 1   Library sample 2

Pooling

(e) Tagged PCR and library build on amplicon pool

DNA extract
Sample 1

t F   R t
Primers with tags

PCR

t F   R t
Tagged PCR amplicon

Tagged PCR product sample 1   Tagged PCR product sample 2

Pooling

Library build on amplicon pool

Incl. index PCR

Adapter ligation

Index primers
i5   i7

PCR

i5 t F   R t i7
Library sequence

OR

Excl. index PCR

Adapter and index ligation

i5 t F   R t i7
Library sequence
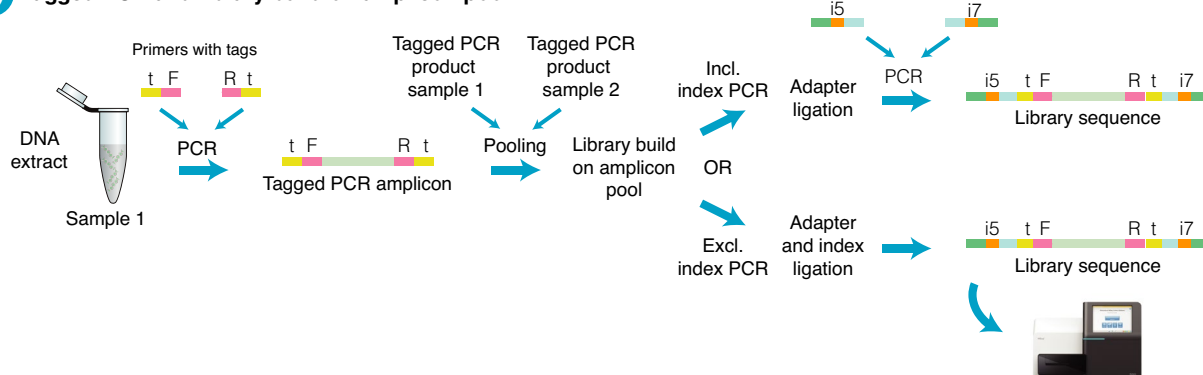
FIGURE 2 Metabarcoding approaches can be divided into three overall strategies for adding nucleotide tags and library indices. (a) The composition of a dual-tagged and dual-indexed metabarcoding Illumina library sequence. Note that the metabarcoding marker, primers, and tags are sequenced as Illumina read 1 and read 2, while index reads are sequenced separately as i7 and i5 reads and used to multiplex sequencing libraries. (b–e) Strategies for adding nucleotide tags and indices to metabarcoding markers. The one-step PCR (b) is depicted with the use of nucleotide tags, which eliminates the need for indices

even thousands of samples per sequencing run. To allow pooling and parallel sequencing of this magnitude, different molecular labelling systems have been developed. For metabarcoding studies, the addition of sample-specific identifiers to PCR amplicons can be achieved either as nucleotide tags during the metabarcoding PCR, or as library indices when converting amplicons into sequencing libraries, that is, as part of the workflow of adding sequencing adapters to amplicons. A metabarcoding sequencing library consists of amplicons carrying sequencing adapters and indices and can consist of one or more PCR products from one or more samples as outlined below. Note that given the inconsistent use of terminology in the metabarcoding literature, for clarity, we use the original term for nucleotide tags in amplicon sequencing as used by Binladen et al., (2007), and Illumina's terminology to describe the nucleotide reads that are used to demultiplex sequencing libraries, the i5 and i7 index reads. That is, 5′ nucleotide tags are sequenced with the metabarcoding marker and primers in the Illumina sequencing read 1 (and read 2 for paired-end sequencing), while library indices are sequenced as separate index reads, i.e., if dual-indexing is performed as i5 and i7 reads (Figure 2a) (https://support.illumina.com).

Metabarcoding approaches can be divided into three overall strategies for adding nucleotide tags and library indices (Taberlet et al., 2018) (Figure 2):

1. The "one-step PCR" approach in which sample DNA extracts are amplified and built into sequence libraries in one reaction. Here, metabarcoding primers carry sequencing adapters, nucleotide tags, and/or library indices, referred to as "fusion primers" (Figure 2b). This approach is used in for example, Kozich et al. (2013), Elbrecht and Leese (2015), Sickel et al. (2015), Grealy et al. (2016), Berry et al. (2017), Elbrecht et al. (2017), Hardy et al. (2017), Elbrecht and Steinke (2018), Seersholm et al. (2018), and Bessey et al. (2020). If indices are used, then each PCR replicate or sample is a sequencing library and as such is returned as a separate fastq file following sequencing. It should be noted that most studies add nucleotide tags next to the primers thereby eliminating the need for i5 and i7 "indexing".

2. The two-step PCR approach in which sample DNA extracts are PCR-amplified with two primer sets. In the primary reaction the metabarcoding primers carry 5′ sequence overhangs of c. 33–34 nucleotides in length. These can be with (Clarke et al., 2017; Griffiths et al., 2020; Kitson et al., 2018; Li et al., 2019; Vesterinen et al., 2018) or without (Bista et al., 2017; de Vere et al., 2017; Galan et al., 2017; Miya et al., 2015; Swift et al., 2018; Vesterinen et al., 2018) nucleotide tags (Figure 2c,d). The sequence overhangs allow the resulting amplicons to be targeted by the second round of primers, which carry sequencing adapters and indices.

Most commonly, two consecutive PCRs are carried out, such as in Miya et al. (2015), de Vere et al. (2017), Galan et al. (2017), Kaunisto et al. (2017), Swift et al. (2018), and Vesterinen et al. (2018). However, a few studies carry out only one reaction with the two primer sets, such as Clarke, Czechowski, et al. (2014). The two-step PCR approach is based on Illumina's 16S rRNA system originally developed for microbiome studies (www.illumina.com). If unique ndexing is used on PCR replicates in the two-step approach, each PCR replicate is an individual sequencing library and as such is returned as a separate fastq file following sequencing.

3. The "tagged PCR" approach, in which sample DNA extracts are PCR amplified with metabarcoding primers that carry 5′ nucleotide tags. Following PCR amplification, the individually tagged PCR products are pooled, and ligation-based library preparation is carried out on pools of 5′ tagged amplicons. The ligated adapters can themselves contain indices, which eliminates the need for a second PCR step (e.g., Carøe & Bohmann, 2020; Thomsen et al., 2016), or the adapter ligation can be followed by a PCR step with indexed primers (e.g., Bohmann et al., 2018; Hope et al., 2014). This approach was first demonstrated by Binladen et al. (2007) on the 454 FLX platform and has since been used in for example, Shehzad, McCarthy, et al. (2012), Hibert et al. (2013), Hope et al. (2014), Thomsen et al. (2016), Apothéloz-Perret-Gentil et al. (2017), Sigsgaard et al. (2017), Bakker et al. (2017), Kocher et al. (2017), Thomsen and Sigsgaard (2019), and Lynggaard et al. (2020) (Figure 2e). In this approach, each library pool of PCR replicates is a sequencing library and is returned as a separate fastq file, each of which can contain data from a large number of tagged PCR replicates.

For all three strategies, it is important to carefully design tags and indices to ensure that oligonucleotide synthesis, PCR, and sequencing error will not cause them to be unidentifiable or confused (Coissac, 2012; Faircloth & Glenn, 2012). Further, all three strategies offer the option to add extra nucleotides to shift PCR amplicons in relation to each other and thereby to increase sequence complexity on the flow cell ("heterogeneity spacers", see for example, Bohmann et al., 2018; De Barba et al., 2014; Elbrecht & Leese, 2015).

In this article, we discuss the three main metabarcoding strategies. One approach not mentioned here is library preparation on individual unlabelled PCR products through a ligation-based library preparation protocol with or without an index PCR step. However, such ligation based protocol would entail several protocol steps to be carried out on each PCR product, such as end-repair and ligation of adapters (e.g., carrying indices such as in Illumina's TruSeq Nano DNA Library Prep kit, see Zizka et al., 2019). The reason that we do not consider this approach a main metabarcoding strategy is due

to low reported use of this method, its high cost and workload and thereby limited throughput (Zizka et al., 2019).

3 | PROS AND CONS OF METABARCODING APPROACHES

The ability to tag and index amplicons to fully harvest the power of high-throughput sequencing comes at a price as the labelling and pooling of hundreds of PCR replicates is highly complex and entails costs associated with preventing, detecting, and eliminating errors and biases. None of the metabarcoding approaches presented here is perfect; rather each of them has pros and cons. Below, we outline the advantages and disadvantages, specifically addressing issues related to cross-contamination risk, PCR amplification efficiency, chimera formation, tag-jumping, index-misassignment, cost, and workload. The issues associated with each metabarcoding strategy are important to keep in mind for choosing a metabarcoding strategy and for designing laboratory workflows and interpreting results.

3.1 | Cross-contamination risk

During the metabarcoding PCR, here specified as the PCR in which the metabarcoding marker is targeted, relatively short DNA sequences (typically <350 bp) are enriched through amplification. Especially when targeting trace amounts of DNA, PCR amplification can be highly susceptible to contamination and thereby to false positives. The risk of contamination when preparing metabarcoding PCRs, that is from the surroundings or laboratory reagents, is the same no matter which of the three overall metabarcoding approaches is used. Moreover, regardless of the metabarcoding strategy employed, cross-contamination can happen between nucleotide tagged and indexed primer stocks (which are delivered at high molarity). The risk of this happening will be similar between the strategies and will depend on the number of samples and the chosen setup within the employed strategy. In the following, we will therefore focus on how the three main metabarcoding approaches differ in their ability to allow detection of cross-contamination between PCR products after the metabarcoding PCR.

PCR products are labelled during the metabarcoding PCR amplification in the one-step PCR approach (Figure 2b), the two-step PCR approach where tagging is carried out in the first PCR (Figure 2d), and the tagged PCR approach (Figure 2e). If the resulting PCR products carry different tag combinations then cross-contamination between them is obviously not of concern. However, if the same tag combinations occur across multiple samples, then cross-contamination between them can be an issue. A solution is to process them in separate batches to avoid cross-contamination. Some laboratories do not reuse tag-primer combinations to eliminate cross-contamination risk (see Murray et al., 2015).

In the two-step approach, sample-specific labelling is not necessarily carried out during the metabarcoding PCR (Figure 2c,d). If not labelled, there is a risk of cross-contamination between unlabelled PCR products when handling them prior to the second PCR (Zizka et al., 2019). Therefore, this metabarcoding approach has the greatest theoretical risk of cross-contamination between PCR products (Figure 2c, Table 1). The risk of this kind of cross-contamination is eliminated if tagging is carried out in the first PCR, see for example Kitson et al. (2018). If untagged metabarcoding primers are used in the two-step PCR approach (Figure 2c), then cross-contamination can be eliminated if the two PCRs are carried out in the same reaction, that is, both two primer sets are included, see for example Clarke, Czechowski, et al. (2014).

Irrespective of the chosen approach, cross-contamination can be detected and filtered out by including sample replicates, PCR replicates, and positive and negative controls. Thus, these should be included in the laboratory workflow and sequencing (e.g., Bista et al., 2017). An important measure that enables one to filter out potential contamination during data processing is to use different nucleotide tag or library index combinations on each sample's individual PCR replicates. This will allow for stringent sequence processing across each sample's PCR replicates, that is, a restrictive approach in which only sequences that are shared by a number of a sample's PCR replicates are retained (see Alberdi et al., 2018, applied in, for example, Giguet-Covex et al., 2014; De Barba et al., 2014; Hope et al., 2014; Cohen et al., 2020; Lynggaard et al., 2021; Yang et al., 2021).

3.2 | PCR amplification

PCR amplification introduces biases, such as primer biases, and errors, such as nucleotide substitutions and chimeras (Haas et al., 2011; Murray et al., 2015; Piñol et al., 2015; Polz & Cavanaugh, 1998). Two of the three main metabarcoding strategies allow practitioners to carry out only a single PCR step before sequencing, namely the one-step PCR approach and the tagged PCR approach in which PCR-free library building is carried out (Figure 2b,e, Table 1). Because an extra PCR step adds an additional risk of introducing errors, these two approaches offer an advantage over the two-step PCR method (Figure 2c,d) and the tagged PCR approach in which the workflow includes an index PCR step (Figure 2e). It should be noted that the number of cycles in the indexing PCR is typically kept low to minimize PCR errors (e.g., eight cycles: Bohmann et al., 2018). Throughout any of these workflows there is a need to keep PCR cycles to a minimum, which might be especially true of metabarcoding workflows with two PCR steps.

Aside from minimizing the number of PCR steps, the effect of 5′ nucleotide additions to metabarcoding primers should be considered as they are likely to decrease PCR efficiency (Murray et al., 2015; Schnell et al., 2015). Bulk sample and eDNA extracts consist of complex mixtures of DNA from a large number of organisms, which especially in the case of eDNA can be degraded (Taberlet et al., 2012). With DNA extracts, the primers are faced with the task

TABLE 1 Features of the three main metabarcoding strategies

Feature	Metabarcoding strategy				
	One-step PCR With 5′ nucleotide tags, without i5 and i7 indices	**Two-step PCR** Without 5′ nucleotide tags on metabarcoding primers	**Two-step PCR** With 5′ nucleotide tags on metabarcoding primers	**Tagged PCR** Library preparation with T4 DNA polymerase blunt-ending and post-ligation PCR	**Tagged PCR** Library preparation without T4 DNA polymerase blunt-ending and post-ligation PCR
Handling and workload	↓	↑	↑	↑	↑
Risk of cross-contamination between PCR products	↓	↑	↑	↓	↓
Tag-jumps	No	No	No	Yes	No
Potential for index misassignment/library bleeding on the flow cell	No (only if indices are used)	Yes	Yes (if indices are used)	Yes	Yes
Decrease in PCR efficiency due to nucleotide additions to metabarcoding primers	High	Potentially high	Potentially high	Low	Low
Cost of metabarcoding primers	↑	↓	↑	↓	↓
Number PCR steps prior to sequencing	1	2	2	2	1

of amplifying (trace copy number) target DNA from different taxa (Taberlet, Coissac, Pompanon, et al., 2012) potentially distorted by primer biases, inhibitors, and potentially abundant predator or host DNA (e.g., Clarke, Soubrier, et al., 2014; Deagle et al., 2014; Murray et al., 2015). Therefore, it is important to take the effect of 5′ nucleotide additions to metabarcoding primers into account.

The three main metabarcoding strategies have different lengths of nucleotide additions on the 5′-end of metabarcoding primers. The longest 5′-nucleotide additions are found in the one-step PCR approach where up to 60 nucleotides (sequence adapters and tags) are added to one or both of the primers, making the complete primer often over 80 bp long (e.g., Elbrecht & Leese, 2015). In the two-step PCR approach (Figure 2c,d), the sequence overhangs on the metabarcoding primers used in the first PCR are approximately half the length of the fusion primers, for example, 33–34 nucleotides if using Illumina Nextera Indices. The tagged PCR approach has the shortest nucleotide additions to the metabarcoding primers (Figure 2e) with tags of typically 5–10 nucleotides in length (e.g. Alberdi et al., 2018; Coissac, 2012; De Barba et al., 2014). The long additions to the metabarcoding primers in the one-step PCR approach cause a decrease in PCR efficiency, as witnessed by an increase in C_T values (Murray et al., 2015). A comparison of PCR efficiency to other metabarcoding strategies has not, to our knowledge, been formally assessed for the two-step PCR approach, but the two-step PCR approach has been shown to have higher consistency as compared to the one-step fusion primer approach (Zizka et al., 2019). Even the short nucleotide additions in the tagged PCR approach have been shown to decrease PCR efficiency, as witnessed by a significant increase in C_T values (Schnell et al., 2015). Thus, no method is free of decreased PCR efficiency caused by the nucleotide additions to 5′-end of metabarcoding primers. However, it has to our knowledge, not been formally tested whether - and to what extent - the shorter nucleotide tag additions in the tagged PCR approach offers greater PCR efficiency and taxonomic detection than the two other approaches, and thereby it can only be speculated that it is the most sensitive when it comes to detection of taxa in low abundance. It should be noted that increasing the cycle number in the PCR amplifications is not an acceptable solution to increase sensitivity, as increased cycle number will reduce taxonomic diversity (Kelly et al., 2019; Piñol et al., 2015). Regardless of metabarcoding strategy, we stress the importance of optimising PCR amplifications (usually by qPCR) to detect PCR inhibition, identify samples with low template quantity, and track PCR efficiency issues (Murray et al., 2015; Yang et al., 2021).

Theoretically, the reduced PCR efficiency in the one-step and two-step PCR approaches caused by the long overhangs on primers might be counteracted by spiking the PCRs with metabarcoding primers without any 5′ attachments (e.g., Murray et al., 2015). However, this has been shown to have modest PCR efficiency improvements for the one-step approach (e.g., Murray et al., 2015). Alternatively, a pre-enrichment can be carried out before the metabarcoding PCR. That is, running a PCR with metabarcoding primers with no nucleotide additions prior to the metabarcoding PCR, as done in Zizka et al. (2019) and Elbrecht and Steinke (2018) for

the one-step PCR approach. However, this not only introduces another PCR amplification step, but can increase the risk of cross-contamination between PCR products due to the initial unlabelled PCR amplification step (e.g., Murray et al., 2015). Note that adding such a pre-enrichment step to the one-step approach can cause it to be mistaken for a two-step PCR approach.

Apart from the length of the nucleotide additions, it has been investigated whether differences in nucleotide tag sequences can result in biases in the tagged PCR approach. Although some studies show that such tag bias is an issue (Berry et al., 2011; O'Donnell et al., 2016), other studies show that it is not (Leray & Knowlton, 2017; Yang et al., 2021). If tag bias does exist, it should theoretically be minimised if different tags are used on each sample's PCR replicates.

3.3 | Chimeras and tag-jumps

Chimeras can be formed during all PCR steps in any metabarcoding workflow (Figure 2b–e). Chimeras are amplicons which combine sequences from two or more different template molecules, and the majority are thought to result from incomplete primer extension during the elongation phase of the PCR cycle (Judo et al., 1998; Meyerhans et al., 1990; Shin et al., 2014; Wang & Wang, 1997). The probability of chimera formation increases when similar template sequences are amplified in the same PCR reaction (Judo et al., 1998, Smyth et al., 2010, but see also Fonseca et al., 2012), such as during the metabarcoding PCR or during the index PCR amplification of pools of tagged amplicons (Figure 2e). There are different consequences of chimeric sequences depending on where they arise. If they are created during a PCR amplification of a single sample's DNA extract, the chimeras will be intrasample chimeras, which can be falsely interpreted as novel taxa and erroneously inflate measures of diversity. On the other hand, if chimeras are created during a PCR amplification of pooled tagged amplicons, such as in the tagged PCR approach (Figure 2e), the chimeras may be intersample chimeras. Such intersample chimeras can result in tag-jumps and false attribution of amplicon sequences to samples, which can lead to false positives and inflation of diversity (Schnell et al., 2015).

All metabarcoding approaches are prone to intra-sample chimeras. However, as chimera formation increases when similar sequences are amplified in the same PCR reaction (e.g. Judo et al., 1998; Smyth et al., 2010), the use of metabarcoding primers with long 5′ overhangs, as in the one-step and two-step approaches, might be more prone to chimera formation since they carry long and similar sequences at the 5′ end of the primers. However, this hypothesis requires testing. Intrasample chimeras can be reduced by limiting the number of PCR cycles and extending elongation time (Haas et al., 2011; Qiu et al., 2001). Also, if samples are subjected to multiple, independent PCRs, chimeras can be filtered out by keeping only sequences that occur in multiple PCR replicates, the "restrictive approach" described in Alberdi et al., (2018). Chimera detection

programmes such as UCHIME (Edgar et al., 2011) can be used for further clean-up.

Inter-sample chimeras can cause havoc in metabarcoding studies. They can only occur in the tagged PCR approach where library build is carried out on pooled tagged amplicons from different samples (Figure 2e, Table 1). Here, tag-jumps can create sequences with new combinations of the nucleotide tags used in the amplicon pool (Schnell et al., 2015). If the new combinations of tags are already used in the amplicon pool, it will cause false assignment of sequences to samples, which should be avoided at all cost (Esling et al., 2015; Schnell et al., 2015). Such tag-jumps can cause negative controls to accumulate a number of sequences following bioinformatic sorting of sequences to samples, which makes sequencing of negative controls a valuable tool to detect tag-jumps.

The rate of tag-jumping has been estimated from ca. 2% to up to 49% of total sequences (Carøe & Bohmann, 2020; Esling et al., 2015; Schnell et al., 2015). This broad range can be caused by factors affecting intersample chimera formation during the index PCR. For example, DNA template and primer concentration, PCR cycle number, and sequence similarity (e.g., Carøe & Bohmann, 2020; Judo et al., 1998; Smyth et al., 2010). The range of tag-jump proportions highlights the unreliability of including an index PCR step in the tagged PCR approach. It should be noted that tag-jumps can also occur due to T4 DNA polymerase activity in the blunt-ending step during library preparation, as demonstrated in library building for the Roche/454 sequencing platform (van Orsouw et al., 2007; Palkopoulou et al., 2016) and for the Illumina sequencing platform (Carøe & Bohmann, 2020).

To avoid tag-jumps in the tagged PCR approach, and thereby prevent false assignment of sequences to samples, it is important to refine index PCR parameters to decrease the likelihood of chimera formation, or better yet, to omit the index PCR step (Figure 2e). Furthermore, blunt-ending using T4 DNA polymerase should be circumvented during library preparation (Carøe & Bohmann, 2020; Palkopoulou et al., 2016; Schnell et al., 2015). If both T4 DNA polymerase blunt-ending and index PCR are eliminated during library preparation of pools of tagged amplicons, tag-jumps can practically be eliminated (Carøe & Bohmann, 2020).

If the library preparation protocol contains a T4 DNA blunt-ending step and/or an index PCR step, and thereby can be assumed to generate tag-jumps, they can be detected and removed by using "twin-tags" during the original PCRs (e.g., F1-R1, F2-R2, etc.), because tag-jumped sequences would then produce nontwinned tag combinations not used in the set-up (e.g., F1-R2, F2-R3, etc.) (e.g. Schnell et al., 2015; Yang et al., 2021). However, using twin tags comes at the price of buying many more versions of tagged primers and building more libraries (Schnell et al., 2015). If twin tags are not used, chimera removal software can remove some chimeric sequences carrying false combinations of used tags (Schnell et al., 2015).

The extent of tag-jumping and spillover of taxa between samples can be detected through inclusion of positive controls consisting of synthetic oligos or taxa not expected to occur in the data

set. However, note that such controls do not enable confident elimination of false positives caused by tag-jumps. The extent of tag-jumping can also be assessed by comparing all observed combinations of used tags to all originally used tag combinations (Schnell et al., 2015; Zepeda Mendoza et al., 2016).

3.4 | Misassignment of library indices

Incorrect assignment of indices between pooled libraries can cause sequence reads to be incorrectly assigned to libraries. Misassigned indices have been attributed to the formation of mixed clusters on the sequencing flow cell, that is, clusters originating from two different template molecules or clusters growing into each other, to low levels of free index primers present in the sequence library and to bulk amplification of pooled libraries (Costello et al., 2018; Nelson et al., 2014; Sinha et al., 2017; Valk et al., 2019; Vodak et al., 2018). Regardless of how index misassignment occurs, if it occurs in metabarcoding studies it can cause incorrect assignment of amplicon sequences to libraries, which can cause incorrect assignment of sequences to samples and false positives. This phenomenon can affect metabarcoding approaches that include indexing of libraries (Figure 2, Table 1). To avoid index misassignment it is recommended to dual-index libraries with unique library index combinations (Kircher et al., 2012; Sinha et al., 2017), www.illumina.com). Further, stringent bead purification (or size selection) can remove free adapters/primers from the libraries (Owens et al., 2018). The labelling in the different metabarcoding approaches further allows for accounting for potential incorrect assignment of sequences to libraries. In the tagged PCR approach, unique tagging of PCR replicates across all pooled libraries can be used to account for (and detect) index misassignment. However, this can be costly. In the one-step PCR approach, it is common to eliminate the use of i7 and i5 library indices, instead relying on 5′ nucleotide tags, which creates a single library that is free of index misassignment (Table 1). As with tag-jumping, the extent of incorrect assignment of indices and spillover of taxa between samples can be detected through inclusion of positive controls consisting of taxa not expected to occur in the data set and by comparing all observed to all used combinations of used indices when demultiplexing libraries.

It is important not to mistake tag-jumping, index misassignment, or cross-contamination between PCR products with cross-contamination of the primers themselves. Due to the high concentration of primers upon synthesis, cross-contamination (e.g., by aerosols) can manifest itself as low numbers of sequence reads and could be misinterpreted as tag-jumps or index-bleeding. Due to the risk of primer cross-contamination, some laboratories avoid ordering primers in 96-well plates. As mentioned, the risk of cross-contamination between nucleotide tagged primer stocks and indexed primer stocks, which could for example occur during resuspension of primers, will generally be the same no matter which of the three overall metabarcoding approaches is used. If the first PCR step in the two-step PCR approach is carried out without tags (Figure 2c),

the primers are unlabelled and any cross-contamination between the primers will not have consequences.

3.5 | Cost

Metabarcoding primers in the tagged and one-step PCR approaches are labelled, whereas the metabarcoding primers in the two-step approach can be either labelled or not (Figure 2). Due to the different labelling systems in the three primary metabarcoding approaches, there are different costs associated with them.

The fusion primers for the one-step PCR approach are the most expensive metabarcoding primers amongst the three approaches. This is because differently labelled versions are purchased for each metabarcoding primer set and because the increased oligo length results in lower yield of the full length product. If indexing is used instead of tagging and unique matching indices are used to account for index misassignment, one-step PCR can become increasingly expensive for larger scale studies. However all of this needs to be factored against the potential cost of repeating runs due to artefacts and contamination, and the fact that only a single PCR step is needed to go from sample extract to library. The tagged two-step PCR primers will be the second-most expensive (Figure 2d) due to their length and individual labelling.

In the tagged PCR approach (Figure 2e), the metabarcoding primers are relatively inexpensive as they only add 5′ tags of 5–10 nucleotides in length. However, these need to be purchased in many tagged versions for each metabarcoding primer set. Furthermore, if tag-jumping is to be taken into account by only using each tag once in a library amplicon pool, for example, by only amplifying with twin forward and reverse tags, then metabarcoding primer sets have to be ordered in many differently labelled versions (Schnell et al., 2015). To keep costs down, this twin-tagging needs to be balanced by pooling fewer PCR products into each library and thereby creating more sequence libraries, but this then increases expenses to library preparation (Figure 2e). However, if a library preparation protocol is used that does not create tag-jumps, tags can be freely combined, which lowers the number of tagged primers that must be purchased (Carøe & Bohmann, 2020; Schnell et al., 2015). In contrast to the other two metabarcoding approaches, the tagged PCR approach includes ligation-based library preparation of pools of amplicons, and the cost of this therefore has to be taken into account. The cost can be kept low if a protocol that does not generate tag-jumps is used and only a few libraries have to be made.

If a large number of metabarcoding primer sets are used, the two-step approach, where primers in the first PCR do not carry tags (Figure 2c), offers a relatively inexpensive solution. This means that the same primer set can be used across multiple samples and projects. This has the benefit that trying out new metabarcoding primer sets does not entail buying many labelled versions of the metabarcoding primer sets, as it does in the other metabarcoding approaches (Figure 2b,d,e). However, the second primer set in the two-step PCR approach is costly as it has to include both the

sequence complementary to the sequence overhang, the sequence adapters, and the library indices (Figure 2c). It is worth noting that many labelled index primers will have to be purchased if twin dual-indices are used to account for incorrect assignment of indices to libraries. This second primer set is, however, applicable across different metabarcoding primer sets and can thereby be used across many metabarcoding studies. For all three approaches, cost-effectiveness will be increased if the purchased primers are depleted effectively, that is, if they are not only to be used in one small study. The two primer sets in the untagged two-step PCR approach (Figure 2c) have good potential for being used up, as the first unlabelled metabarcoding primer set can be used across many samples and the second primer set can be used across different metabarcoding primer sets.

3.6 | Laboratory workload

The one-step PCR approach is without doubt the quickest method for generating sequence-ready libraries, as it only requires a single PCR-step to achieve both amplification and library preparation of the metabarcoding amplicons (Figure 2b). Researchers have used this method in research and commercial scenarios to turn around sequence data in 12–24 h in the field on the iSeq platform (Bunce, unpublished data). In some applications, especially requiring timely interventions, the rapid turnaround time of the one-step PCR approach may be a consideration. The workload for the two-step PCR approach and the tagged PCR approach depends, to some extent, on how many sample extracts and PCR replicates are to be processed. If it is a relatively high number, the tagged PCR approach is the quickest due to the library build being performed on pooled amplicons rather than through a PCR step on individual PCR products. However, as with all molecular biological workflows, carefully organised liquid handling and automation provide solutions to high-throughput studies.

4 | CHOOSING A METABARCODING APPROACH

It is clear that there is no such thing as a perfect metabarcoding sample-labelling approach, and that choosing which one is right for a given study or laboratory should be an informed trade-off of pros and cons balanced to the needs (Table 1). Within metabarcoding studies, those needs can range widely.

Metabarcoding studies range from those that look for one or a few taxa within sample units (Bohmann et al., 2018) to studies that look for many taxa within sample units (Seersholm et al., 2018), and sample numbers can range from tens (Elbrecht et al., 2017), to hundreds (Galan et al., 2017; Rodgers et al., 2017) or even thousands (Ji et al., 2021; Schnell et al., 2018). The research question and experimental set-up can require taxonomic identifications to be made within individual samples (Coghlan et al., 2012), while in other studies, taxonomic identifications from pools of individual samples or

from a number of samples within, for example, a geographic location is the goal (Grealy et al., 2016; Schnell et al., 2018). Sample types can range from bulk specimen samples consisting of high quality DNA from pools of entire organisms (Tang et al., 2015) to environmental samples in which DNA from target organisms can be fragmented and scarce (Stat et al., 2017). Furthermore, studies differ in how many metabarcoding primer sets are used - from only one (Bohmann et al., 2011; Drinkwater et al., 2018) to several (De Barba et al., 2014; Drummond et al., 2015; Zhang et al., 2018). Furthermore, the budget for a metabarcoding project will differ between studies, as will whether the metabarcoding primers are to be used in future studies. Lastly, some applications of metabarcoding, such as biosecurity or forensics, will necessitate a "high bar" for data fidelity and controls.

A multitude of combinations of the above metabarcoding study parameters exist, and as demonstrated by this article, the significance of the pros and cons of the metabarcoding approaches will differ with them. For example, while the tagged PCR approach (Figure 2e) may excel in amplifying low abundance templates given the shorter nucleotide additions to the metabarcoding primers than the one-step PCR primers (Murray et al., 2015; Zizka et al., 2019), the one-step PCR offers a quicker turnaround (Figure 2b). However, the one-step PCR strategy comes at the cost of buying long fusion primers, and is only worthwhile if the metabarcoding primers are to be used again.

When choosing a metabarcoding approach, the need for future multiplexing of the metabarcoding primers should be considered. That is, to use several metabarcoding primer sets that target different markers and taxonomic groups within the same PCR reaction to simultaneously screen for many taxonomic groups and thereby keep costs and work load at a minimum (e.g., De Barba et al., 2014). For this, the nucleotide tagged primers in the tagged PCR approach should theoretically be the most applicable, whereas the long additions to the metabarcoding primers in the one-step and two-step PCR approaches might be less conducive to multiplexing due to the extensive sequence homology.

Lastly, it should be noted that whatever metabarcoding strategy is chosen, it should be clear from the present article that one should not change workflows within an experiment. Moreover, there is some justified concern within the metabarcoding community that the nuances in metabarcoding workflows makes interlaboratory comparison difficult (Blackman et al., 2019; Murray et al., 2015; Zizka et al., 2019).

5 | APPLICATIONS ON OTHER SEQUENCING PLATFORMS

Although to a more limited extent, other second generation sequencing technologies than Illumina are used in metabarcoding. For example, Ion Torrent (Thermo Fischer Scientific) and BGI platforms (BGI Genomics) (Braukmann et al., 2019; Forin-Wiart et al., 2018; Schnell et al., 2018; Yang et al., 2020). These technologies require the addition of sequencing adapters similar to Illumina platforms

and have similar labelling structure. Therefore, discussions regarding labelling strategies in the present article are largely applicable to metabarcoding on these other platforms. For example, the one-step (Schnell et al., 2018) and the two-step PCR approach (Braukmann et al., 2019; Nota et al., 2019) have been used on the Ion Torrent platform, and the tagged PCR approach has been used on BGI's MGISEQ platform (Yang et al., 2020). Further, third generation technologies yielding long reads have been employed in metabarcoding; Pacific Biosciences (PacBio) (James et al., 2016; Tedersoo et al., 2018) and the portable Oxford Nanopore Technologies MinION sequencer (Karst et al., 2021). These platforms also rely on the addition of sequencing adapters. The high error rate of these platforms (Dohm et al., 2020) compared to Illumina platforms (Stoler & Nekrutenko, 2021) makes correct taxa identification and sample specific labelling difficult. However, solutions to this are being developed (Karst et al., 2021). It is likely that metabarcoding applications will probably follow the platform with the highest sequencing fidelity although in some applications speed and portability may also increasingly become factors in platform choice.

6 | PERSPECTIVES

All metabarcoding strategies can generate robust data. However, like all laboratory workflows if they are not executed well or are inappropriate for the application, they may lead to flawed data. We advocate that just because PCR is a relatively simple method it does not mean that metabarcoding is simple, and there are many traps in metabarcoding workflows that can trip-up new users. Here, we have presented an overview of the three main metabarcoding strategies for assessment of biodiversity on Illumina sequencing platforms, and the downstream consequences for the resulting data with regards to cross-contamination risk, PCR amplification efficiency, chimera formation, tag-jumping, index-misassignment, as well as cost and workload. In doing so we wish to enable researchers and practitioners to make an informed choice of which metabarcoding strategy is best suited for their specific study. Ultimately, this is to avoid the worst case scenario: generation of unusable data and wasting a considerable amount of time and money, or even worse making wrong conclusions due to flawed data.

Metabarcoding of environmental DNA has some commonalities with the field of ancient DNA in which low quality and quantity of target DNA is also targeted amongst nontarget, and potentially more abundant, templates. In the early days of ancient DNA studies, PCR-based techniques, including amplifying already amplified DNA to enhance signals, were used, which caused authentication issues, as amplification of modern templates was mistaken for true ancient signals. This was followed by urgent calls for precautions to ensure reliability and authenticity of ancient DNA sequences (Cooper & Poinar, 2000; Pääbo et al., 2004). Also similarly to the field of ancient DNA, the take-home message should be that metabarcoding is becoming a self-critical and self-correcting field in which technical reliability is promoted and rewarded, with the long-term benefit

of uptake by stakeholders who will employ metabarcoding for environmental management. Reputational setbacks as the result of practitioners not executing their metabarcoding workflows well will probably resonate across a variety of biomonitoring, forensic, and bioseurity applications.

We thus stress the importance of being informed about the pros and cons of the chosen metabarcoding approach with regards to cross-contamination risk, PCR amplification efficiency, chimera formation, tag-jumping, index-misassignment, cost, and workload and to include appropriate quality assurance and quality control measures. This will help ensure that the generated data will facilitate informed data analysis and interpretation. We advocate that metabarcoding publications should include detailed information about the metabarcoding strategy and how its challenges have been taken into account in the laboratory, data processing, and interpretation of results. Furthermore, it may be appropriate to eventually develop a set of metabarcoding guidelines similar to the MIQE guidelines for qPCR (Bustin et al., 2009) to establish standard reporting practises, which would ultimately further increase the power and reliability of metabarcoding.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Kristine Bohmann https://orcid.org/0000-0001-7907-064X
Vasco Elbrecht https://orcid.org/0000-0003-4672-7099
Christian Carøe https://orcid.org/0000-0001-9601-6768
Florian Leese https://orcid.org/0000-0002-5465-913X
Michael Bunce https://orcid.org/0000-0002-0302-4206

Douglas W. Yu (ID) https://orcid.org/0000-0001-8551-5609
Mathew Seymour (ID) https://orcid.org/0000-0002-3654-4857
Alex J. Dumbrell (ID) https://orcid.org/0000-0001-6282-3043
Simon Creer (ID) https://orcid.org/0000-0003-3124-3550

REFERENCES

Aizpurua, O., Budinski, I., Georgiakakis, P., Gopalakrishnan, S., Ibañez, C., Mata, V., Rebelo, H., Russo, D., Szodoray-Parádi, F., Zhelyazkova, V., Zrncic, V., Gilbert, M. T. P., & Alberdi, A. (2017). Agriculture shapes the trophic niche of a bat preying on multiple pest arthropods across Europe: Evidence from DNA metabarcoding. *Molecular Ecology*, 27, 815–825. https://doi.org/10.1111/mec.14474

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 / British Ecological Society*, 9(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., & Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources*, 17(6), 1231–1242. https://doi.org/10.1111/1755-0998.12668

Aylagas, E., Borja, Á., Muxika, I., & Rodríguez-Ezpeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, 95, 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044

Bakker, J., Wangensteen, O. S., Chapman, D. D., Boussarie, G., Buddo, D., Guttridge, T. L., Hertler, H., Mouillot, D., Vigliola, L., & Mariani, S. (2017). Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. *Scientific Reports*, 7(1), 16886. https://doi.org/10.1038/s41598-017-17150-2

Bálint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., O'Hara, R. B., Öpik, M., Sogin, M. L., Unterseher, M., & Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, 40, 686–700. https://doi.org/10.1093/femsre/fuw017

Berry, D., Ben Mahfoudh, K., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, 77(21), 7846–7849. https://doi.org/10.1128/AEM.05220-11

Berry, T. E., Osterrieder, S. K., Murray, D. C., Coghlan, M. L., Richardson, A. J., Grealy, A. K., Stat, M., Bejder, L., & Bunce, M. (2017). DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecology and Evolution*, 7(14), 5435–5453. https://doi.org/10.1002/ece3.3123

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, 2(4), 493–504. https://doi.org/10.1002/edn3.74

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*, 2(2), e197. https://doi.org/10.1371/journal.pone.0000197

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, 8, 14087. https://doi.org/10.1038/ncomms14087

Blackman, R. C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A. F., Jones, J. I., Macher, J., Majaneva, M., Martins, F. M. S., Múrria, C., Meissner, K., Pawlowski, J., Schmidt Yáñez, P. L., Zizka, V. M. A., Leese, F., ... Deiner, K. (2019). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – The need for calibration experiments. *Metabarcoding Metagenom*, 3, 49–57. https://doi.org/10.5167/UZH-174614

Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution*, 32(7), 477–487. https://doi.org/10.1016/j.tree.2017.03.001

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

Bohmann, K., Monadjem, A., Lehmkuhl Noer, C., Rasmussen, M., Zeale, M. R. K., Clare, E., Jones, G., Willerslev, E., & Gilbert, M. T. P. (2011). Molecular diet analysis of two African free-tailed bats (molossidae) using high throughput sequencing. *PLoS One*, 6(6), e21441. https://doi.org/10.1371/journal.pone.0021441

Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J. R., Sones, J. E., Zakharov, E. V., & Hebert, P. D. N. (2019). Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, 19(3), 711–727. https://doi.org/10.1111/1755-0998.13008

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J. O., & Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, 22(4), 915–924. https://doi.org/10.1111/mec.12183

Carøe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *Molecular Ecology Resources*, 20(6), 1620–1631. https://doi.org/10.1111/1755-0998.13227

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(3), 873–883. https://doi.org/10.1002/ece3.2667

Clarke, L. J., Czechowski, P., & Soubrier, J. (2014). Modular tagging of amplicons using a single PCR for high-throughput sequencing. *Molecular Ecology Resources*, 14(1), 117–121. https://doi.org/10.1111/1755-0998.12162

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(6), 1160–1170. https://doi.org/10.1111/1755-0998.12265

Coghlan, M. L., Haile, J., Houston, J., Murray, D. C., White, N. E., Moolhuijzen, P., Bellgard, M. I., & Bunce, M. (2012). Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLoS Genetics*, 8(4), e1002657. https://doi.org/10.1371/journal.pgen.1002657

Cohen, Y., Bar-David, S., Nielsen, M., Bohmann, K., & Korine, C. (2020). An appetite for pests: Synanthropic insectivorous bats exploit cotton pest irruptions and consume various deleterious arthropods. *Molecular Ecology*, 29, 1185–1198. https://doi.org/10.1111/mec.15393

Coissac, E. (2012). OligoTag: A program for designing sets of tags for next-generation sequencing of multiplexed samples. *Methods in Molecular Biology*, 888, 13–31. https://doi.org/10.1007/978-1-61779-870-2_2

Compson, Z. G., McClenaghan, B., Singer, G. A. C., Fahner, N. A., & Hajibabaei, M. (2020). Metabarcoding from microbes to mammals:

Comprehensive bioassessment on a global scale. *Frontiers in Ecology and Evolution*, 8, 379. https://doi.org/10.3389/fevo.2020.581835

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do it right or not at all. *Science*, 289(5482), 1139. https://doi.org/10.1126/science.289.5482.1139b

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., Vicente, G., Dasilva, M., Brodeur, W., DeSmet, T., Dodge, S., Lennon, N. J., & Gabriel, S. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics*, 19(1), 332. https://doi.org/10.1186/s12864-018-4703-0

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution/British Ecological Society*, 7(9), 1008–1018. https://doi.org/10.1111/2041-210X.12574

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14(2), 306–323. https://doi.org/10.1111/1755-0998.12188

de Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M. J., Creer, S., & Ford, C. R. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports*, 7, 42838. https://doi.org/10.1038/srep42838

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(9). https://doi.org/10.1098/rsbl.2014.0562

Dohm, J. C., Peters, P., Stralis-Pavese, N., & Himmelbauer, H. (2020). Benchmarking of long-read correction methods. *NAR Genomics and Bioinformatics*, 2(2), lqaa037. https://doi.org/10.1093/nargab/lqaa037

Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., Gilbert, M. T. P., & Rossiter, S. J. (2018). Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources*, 19(1), 105–117. https://doi.org/10.1111/1755-0998.12943

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., Heled, J., Ross, H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C., Anderson, S. H., Carter, A., & Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4, 46. https://doi.org/10.1186/s13742-015-0086-1

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 10(7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., & Steinke, D. (2018). Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. *Freshwater Biology*, 5, 1. https://doi.org/10.1111/fwb.13220

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution/British Ecological Society*, 8(10), 1265–1275. https://doi.org/10.1111/2041-210x.12789

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, 43(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Faircloth, B. C., & Glenn, T. C. (2012). Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS One*, 7(8), e42543. https://doi.org/10.1371/journal.pone.0042543

Fonseca, V. G., Nichols, B., Lallias, D., Quince, C., Carvalho, G. R., Power, D. M., & Creer, S. (2012). Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research*, 40(9), e66. https://doi.org/10.1093/nar/gks002

Forin-Wiart, M.-A., Poulle, M.-L., Piry, S., Cosson, J.-F., Larose, C., & Galan, M. (2018). Evaluating metabarcoding to analyse diet composition of species foraging in anthropogenic landscapes using Ion Torrent and Illumina sequencing. *Scientific Reports*, 8. https://doi.org/10.1038/s41598-018-34430-7

Galan, M., Pons, J.-B., Tournayre, O., Pierre, É., Leuchtmann, M., Pontier, D., & Charbonnel, N. (2017). Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Molecular Ecology Resources*, 18(3), 474–489. https://doi.org/10.1111/1755-0998.12749

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J., & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, 5, 3211. https://doi.org/10.1038/ncomms4211

Gous, A., Swanevelder, D. Z. H., Eardley, C. D., & Willows-Munro, S. (2019). Plant–pollinator interactions over time: Pollen metabarcoding from bees in a historic collection. *Evolutionary Applications*, 12(2), 187–197. https://doi.org/10.1111/eva.12707

Grealy, A., Douglass, K., Haile, J., Bruwer, C., Gough, C., & Bunce, M. (2016). Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar. *Journal of Archaeological Science*, 75, 82–88. https://doi.org/10.1016/j.jas.2016.10.001

Griffiths, N. P., Bolland, J. D., Wright, R. M., Murphy, L. A., Donnelly, R. K., Watson, H. V., & Hänfling, B. (2020). Environmental DNA metabarcoding provides enhanced detection of the European eel Anguilla anguilla and fish community structure in pumped river catchments. *Journal of Fish Biology*, 97(5), 1375–1384. https://doi.org/10.1111/jfb.14497

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methe, B., DeSantis, T. Z., Petrosino, J. F., Knight, R., & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, 21(3), 494–504. https://doi.org/10.1101/gr.112730.110

Hardy, N., Berry, T., Kelaher, B. P., Goldsworthy, S. D., Bunce, M., Coleman, M. A., Gillanders, B. M., Connell, S. D., Blewitt, M., & Figueira, W. (2017). Assessing the trophic ecology of top predators across a recolonisation frontier using DNA metabarcoding of diets. *Marine Ecology Progress Series*, 573, 237–254. https://doi.org/10.3354/meps12165

Hibert, F., Taberlet, P., Chave, J., Scotti-Saintagne, C., Sabatier, D., & Richard-Hansen, C. (2013). Unveiling the diet of elusive rainforest herbivores in next generation sequencing era? The tapir as a case study. *PLoS One*, 8(4), e60799. https://doi.org/10.1371/journal.pone.0060799

Hope, P. R., Bohmann, K., Gilbert, M. T. P., Zepeda-Mendoza, M. L., Razgour, O., & Jones, G. (2014). Second generation sequencing and morphological faecal analysis reveal unexpected foraging behaviour by Myotis nattereri (Chiroptera, Vespertilionidae) in winter. *Frontiers in Zoology*, 11(1), 39. https://doi.org/10.1186/1742-9994-11-39

James, T. Y., Marino, J. A., Perfecto, I., & Vandermeer, J. (2016). Identification of putative coffee rust mycoparasites via single-molecule DNA sequencing of infected pustules. *Applied*

and Environmental Microbiology, *82*(2), 631–639. https://doi.org/10.1128/AEM.02639-15

Jarman, S. N., Berry, O., & Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. *Nature Ecology & Evolution*, *2*, 1192–1193. https://doi.org/10.1038/s41559-018-0614-3

Ji, Y., Baker, C. C. M., Popescu, V. D., Wang, J., Wu, C., Wang, Z., Li, Y., Wang, L., Hua, C., Yang, Z, Xu, C. C. Y., Diana, A., Wen, Q., Pierce, N., & Yu, D. W. (2021). Measuring protected-area outcomes with leech iDNA: large-scale quantification of vertebrate biodiversity in Ailaoshan reserve. *bioRxiv*. https://doi.org/10.1101/2020.02.10.941336

Judo, M. S., Wedel, A. B., & Wilson, C. (1998). Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research*, *26*(7), 1819–1825. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9512558

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., Knight, R., & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nature Methods*, *18*(2), 165–169. https://doi.org/10.1038/s41592-020-01041-y

Kaunisto, K. M., Roslin, T., Sääksjärvi, I. E., & Vesterinen, E. J. (2017). Pellets of proof: First glimpse of the dietary composition of adult odonates as revealed by metabarcoding of feces. *Ecology and Evolution*, *7*(20), 8588–8598. https://doi.org/10.1002/ece3.3404

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*, *9*(1), 12133. https://doi.org/10.1038/s41598-019-48546-x

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3. https://doi.org/10.1093/nar/gkr771

Kitson, J. J. N., Hahn, C., Sands, R. J., Straw, N. A., Evans, D. M., & Lunt, D. H. (2018). Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. *Molecular Ecology*, *28*(2), 471–483. https://doi.org/10.1111/mec.14518

Kocher, A., Thoisy, B., Catzeflis, F., Huguin, M., Valière, S., Zinger, L., Bañuls, A.-L., & Murienne, J. (2017). Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. *Methods in Ecology and Evolution / British Ecological Society*, *8*(10), 1276–1283. https://doi.org/10.1111/2041-210X.12729

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ*, *5*, e3006. https://doi.org/10.7717/peerj.3006

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., & Zhang, X. (2018). Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers. *Environmental Science & Technology*, *52*(20), 11708–11719. https://doi.org/10.1021/acs.est.8b03869

Li, J., Hatton-Ellis, T. W., Lawson Handley, L., Kimbell, H. S., Benucci, M., Peirson, G., & Hänfling, B. (2019). Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. *The Journal of Applied Ecology*, *56*(5), 1232–1244. https://doi.org/10.1111/1365-2664.13352

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., & Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – A user's guide. *New Phytologist*, *199*, 288–299. https://doi.org/10.1111/nph.12243

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., Hegarty, M., Jones, L., Neyland, P. J., & de Vere, N. (2018). Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports*, *8*(1), 5133. https://doi.org/10.1038/s41598-018-23103-0

Lynggaard, C., Bertelsen, M. F., Jensen, C. V., & Johnson, M. S. (2021). Airborne environmental DNA for terrestrial vertebrate community monitoring. *bioRxiv*. https://doi.org/10.1101/2021.07.16.452634.abstract

Lynggaard, C., Yu, D. W., Oliveira, G., Caldeira, C. F., Ramos, S. J., Ellegaard, M. R., Gilbert, M. T. P., Gastauer, M., & Bohmann, K. (2020). DNA-based arthropod diversity assessment in amazonian iron mine lands show ecological succession towards undisturbed reference sites. *Frontiers in Ecology and Evolution*, *8*. https://doi.org/10.3389/fevo.2020.590976

Meyerhans, A., Vartanian, J. P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research*, *18*(7), 1687–1691. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2186361

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088. https://doi.org/10.1098/rsos.150088

Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLoS One*, *10*(4), e0124671. https://doi.org/10.1371/journal.pone.0124671

Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L., & Graf, J. (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One*, *9*(4), e94249. https://doi.org/10.1371/journal.pone.0094249

Nota, K., Downing, S., & Iyengar, A. (2019). Metabarcoding-based dietary analysis of hen harrier (*Circus cyaneus*) in Great Britain using buccal swabs from chicks. *Conservation Genetics*, *20*(6), 1389–1404. https://doi.org/10.1007/s10592-019-01215-y

O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS One*, *11*(3), e0148698. https://doi.org/10.1371/journal.pone.0148698

Owens, G. L., Todesco, M., Drummond, E. B. M., Yeaman, S., & Rieseberg, L. H. (2018). A novel post hoc method for detecting index switching finds no evidence for increased switching on the Illumina HiSeq X. *Molecular Ecology Resources*, *18*(1), 169–175. https://doi.org/10.1111/1755-0998.12713

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Déspres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Palkopoulou, E., Baca, M., Abramson, N. I., Sablin, M., Socha, P., Nadachowski, A., Prost, S., Germonpré, M., Kosintsev, P., Smirnov, N. G., Vartanyan, S., Ponomarev, D., Nyström, J., Nikolskiy, P., Jass, C. N., Litvinov, Y. N., Kalthoff, D. C., Grigoriev, S., Fadeeva, T., … Dalén, L. (2016). Synchronous genetic turnovers across Western Eurasia in Late Pleistocene collared lemmings. *Global Change Biology*, *22*(5), 1710–1721. https://doi.org/10.1111/gcb.13214

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*(4), 819–830. https://doi.org/10.1111/1755-0998.12355

Pochon, X., Bott, N. J., Smith, K. F., & Wood, S. A. (2013). Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS One*, *8*(9), e73935. https://doi.org/10.1371/journal.pone.0073935

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, *64*(10), 3724–3730. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9758791

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H., & Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, *8*(1), 10361. https://doi.org/10.1038/s41598-018-28424-8

Pont, D., Valentini, A., Rocle, M., Maire, A., Delaigue, O., Jean, P., & Dejean, T. (2021). The future of fish-based ecological assessment of European rivers: from traditional EU Water Framework Directive compliant methods to eDNA metabarcoding-based approaches. *Journal of Fish Biology*, *98*(2), 354–366. https://doi.org/10.1111/jfb.14176

Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A. V., Tiedje, J. M., & Zhou, J. (2001). Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, *67*(2), 880–887. https://doi.org/10.1128/AEM.67.2.880-887.2001

Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., Ranirison, P., Crouau-Roy, B., Taberlet, P., & Chikhi, L. (2013). A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PLoS One*, *8*(3), e58971. https://doi.org/10.1371/journal.pone.0058971

Razgour, O., Clare, E. L., Zeale, M. R. K., Hanmer, J., Schnell, I. B., Rasmussen, M., Gilbert, T. P., & Jones, G. (2011). High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and Evolution*, *1*(4), 556–570. https://doi.org/10.1002/ece3.49/full.

Rodgers, T. W., Xu, C. C. Y., & Giacalone, J. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology*. https://doi.org/10.1111/1755-0998.12701

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*(6), 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C., Sinding, M.-H.- S., Bass, D., Cadle, J. E., Campbell, M. J., Dolch, R., Edwards, D. P., Gray, T. N. E., Hansen, T., Hoa, A. N. Q., Noer, C. L., Heise-Pavlov, S., Sander Pedersen, A. F., Ramamonjisoa, J. C., Siddall, M. E., … Gilbert, M. T. P. (2018). Debugging diversity-a pancontinental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool. *Molecular Ecology Resources*, *18*(6), 1282–1298. https://doi.org/10.1111/1755-0998.12912

Seersholm, F. V., Cole, T. L., Grealy, A., Rawlence, N. J., Greig, K., Knapp, M., Stat, M., Hansen, A. J., Easton, L. J., Shepherd, L., Tennyson, A. J. D., Scofield, R. P., Walter, R., & Bunce, M. (2018). Subsistence practices, past biodiversity, and anthropogenic impacts revealed by New Zealand-wide ancient DNA survey. *Proceedings of the National Academy of Sciences*, *115*(30), 7771–7776. https://doi.org/10.1073/pnas.1803573115

Seymour, M., Edwards, F. K., Cosby, B. J., Kelly, M. G., de Bruyn, M., Carvalho, G. R., & Creer, S. (2020). Executing multi-taxa eDNA ecological assessment via traditional metrics and interactive networks. *The Science of the Total Environment*, *729*, 138801. https://doi.org/10.1016/j.scitotenv.2020.138801

Shehzad, W., McCarthy, T. M., Pompanon, F., Purevjav, L., Coissac, E., Riaz, T., & Taberlet, P. (2012). Prey preference of snow leopard (*Panthera uncia*) in South Gobi, Mongolia. *PLoS One*, *7*(2), e32104. https://doi.org/10.1371/journal.pone.0032104

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., Pompanon, F., Coissac, E., & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard

cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, *21*(8), 1951–1965. https://doi.org/10.1111/j.1365-294X.2011.05424.x/full

Shin, S., Lee, T. K., Han, J. M., & Park, J. (2014). Regional effects on chimera formation in 454 pyrosequenced amplicons from a mock community. *Journal of Microbiology*, *52*(7), 566–573. https://doi.org/10.1007/s12275-014-3485-6

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., & Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, *15*, 20. https://doi.org/10.1186/s12898-015-0051-y

Sigsgaard, E. E., Nielsen, I. B., Carl, H., Krag, M. A., Knudsen, S. W., Xing, Y., Holm-Hansen, T. H., Møller, P. R., & Thomsen, P. F. (2017). Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, *164*(6), 128. https://doi.org/10.1007/s00227-017-3147-4

Singer, G., Fahner, N. A., Barnes, J., McCarthy, A., & Hajibabaei, M. (2019). *Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: A case study of eDNA metabarcoding seawater* (p. 515890). https://doi.org/10.1101/515890

Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., Travaglini, K. J., Wei, E., … Weissman, I. L. (2017). *Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing* (p. 125724). https://doi.org/10.1101/125724

Smyth, R. P., Schlub, T. E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M. P., & Mak, J. (2010). Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene*, *469*(1–2), 45–51. https://doi.org/10.1016/j.gene.2010.08.009

Srivathsan, A., Sha, J., Vogler, A. P., & Meier, R. (2015). Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources*, *15*(2), 250–261. https://doi.org/10.1111/1755-0998.12302/full

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 12240. https://doi.org/10.1038/s41598-017-12501-5

Stoeck, T., Pan, H., Dully, V., Forster, D., & Jung, T. (2018). Towards an eDNA metabarcode-based performance indicator for full-scale municipal wastewater treatment plants. *Water Research*, *144*, 322–331. https://doi.org/10.1016/j.watres.2018.07.051

Stoler, N., & Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics*, *3*(1), lqab019. https://doi.org/10.1093/nargab/lqab019

Swift, J. F., Lance, R. F., Guan, X., Britzke, E. R., Lindsay, D. L., & Edwards, C. E. (2018). Multifaceted DNA metabarcoding: Validation of a noninvasive, next-generation approach to studying bat populations. *Evolutionary Applications*, *12*, 2175. https://doi.org/10.1111/eva.12644

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press. Retrieved from https://play.google.com/store/books/details?id=1e9IDwAAQBAJ

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E. D., Wang, J., Yang, C., Bruce, C., Nevard, T., Potts, S. G., Zhou, X., & Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and*

Evolution / British Ecological Society, 6(9), 1034–1043. https://doi.org/10.1111/2041-210X.12416

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *The New Phytologist, 217*(3), 1370–1385. https://doi.org/10.1111/nph.14776

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, Deepwater Fishes. *PLoS One, 11*(11), e0165252. https://doi.org/10.1371/journal.pone.0165252

Thomsen, P. F., & Sigsgaard, E. E. (2019). Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution, 9*(4), 1665–1679. https://doi.org/10.1002/ece3.4809

van der Valk, T., van der Valk, T., Vezzi, F., Ormestad, M., Dalen, L., & Guschanski, K. (2019). *Index hopping on the illumina HiseqX platform and its consequences for ancient DNA studies.* https://doi.org/10.1101/179028

van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., & van Eijk, M. J. T. (2007). Complexity reduction of polymorphic sequences (CRoPS™): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One, 2*(11), e1172. https://doi.org/10.1371/journal.pone.0001172

Vesterinen, E. J., Puisto, A. I. E., Blomberg, A. S., & Lilley, T. M. (2018). Table for five, please: Dietary partitioning in boreal bats. *Ecology and Evolution, 8*(22), 10914–10937. https://doi.org/10.1002/ece3.4559

Vodak, D., Lorenz, S., Nakken, S., Aasheim, L. B., Holte, H., Bai, B., Hovig, E. (2018). *Sample-index misassignment impacts tumor exome sequencing.* https://doi.org/10.1101/182659

Wang, G. C., & Wang, Y. (1997). Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Applied and Environmental Microbiology, 63*(12), 4645–4650. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9406382

Yang, C., Bohmann, K., Wang, X., Cai, W., Wales, N., Ding, Z., Gopalakrishnan, S., & Yu, D. W. (2021). Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. *Methods in Ecology and Evolution, 12*(7), 1252–1264. https://doi.org/10.1111/2041-210x.13602

Yang, C., Zheng, Y., Tan, S., Meng, G., Rao, W., Yang, C., Bourne, D. G., O'Brien, P. A., Xu, J., Liao, S., Chen, A., Chen, X., Jia, X., Zhang, A. B., & Liu, S. (2020). Efficient COI barcoding using high throughput single-end 400 bp sequencing. *BMC Genomics, 21*(1), 862. https://doi.org/10.1186/s12864-020-07255-w

Zepeda Mendoza, M. L., Bohmann, K., Carmona Baez, A., & Gilbert, M. T. P. (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*(1), 255. https://doi.org/10.1186/s13104-016-2064-9

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*(10), 1901–1914. https://doi.org/10.1111/eva.12694

Zizka, V. M. A., Elbrecht, V., Macher, J.-N., & Leese, F. (2019). Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Molecular Ecology Resources, 19*(4), 893–899. https://doi.org/10.1111/1755-0998.13018

Zizka, V. M. A., Geiger, M. F., & Leese, F. (2020). DNA metabarcoding of stream invertebrates reveals spatio-temporal variation but consistent status class assessments in a natural and urban river. *Ecological Indicators, 115,* 106383. https://doi.org/10.1016/j.ecolind.2020.106383